State-dependent Modeling of Default Rates

by

Bowen Hu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this proposal. The decision of the Examining Committee is by majority vote.

Supervisor(s): Adam Kolkiewicz
Associate Professor, Dept. of Statistics and Actuarial Science,
University of Waterloo

Adam Metzler
Associate Professor, Dept. of Mathematics,
Wilfrid Laurier University

Committee Members: Tony Wirjanto
Professor, Dept. of Statistics and Actuarial Science,
University of Waterloo

David Saunders
Associate Professor, Dept. of Statistics and Actuarial Science,
University of Waterloo

Internal-External Member: Ken Vetzal
Associate Professor, School of Accounting and Finance,
University of Waterloo

External Member: Kenneth R. Jackson
Professor, Dept. of Computer Science,
University of Toronto

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Risk-weight function is the most popular formula for banking regulations used to calculate the amount of backup deposit that banks need to hold in order to bear extraordinary losses. The model behind the formula was introduced by Vasicek in 2002. In that paper, there are several intuitively appealing assumptions which are oversimplified. The most unrealistic assumption made by Vasicek is that correlations among each unit do not depend on the overall market environment.

Metzler (2020) has developed a generalized version of the Vasicek model to relax this assumption, which is called the state-dependent model. The model includes a parameter to allow the market correlations to change in a systematic way based on the overall economic level. We apply an EM algorithm that produces consistent estimates of the model parameters proposed by Metzler (2000). We also explore some properties of the model.

The model involves an independence assumption, which assumes that the default rate for each time is independent with each other. But according to the plots of the historical data, that assumption is obviously violated. In order to relax the independence assumption, we bring a dependence structure to the model with respect to time by using time series to model the so-called systematic risk factor M. By doing so, we bring the forecasting ability to the model and verify its accuracy in the empirical study.

The results suggest that the model we proposed shows some advantages compared with the classic auto-regression models. We also demonstrate that the model we proposed can be treated as a general extension of the classic auto-regression models.

In the last part, we try to overcome the other well-know problem of the Vasicek model. Both the Vasicek model and SDM model fall into the family of the Gaussian copula. Although the Gaussian copula is widely used in the industry for its nice properties, the 2008 financial crisis warned researchers that tail independence can lead to some fatal results. In order to solve this problem, we change the underlying distribution from normal distribution to t-distribution.

Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance.

I would first like to thank my supervisors, Professor Adam Kolkiewicz and Professor Adam Metzler, whose expertise was invaluable in helping me to finish my research. I am proud of and grateful for being your student.

I would also like to thank all of my committee members and examiners, Professor David Saunders, Professor Tony Wirjanto, Professor Ken Vetzal and Professor Kenneth R. Jackson. Your suggestions and comments benefit me not only for this dissertation but also my future career.

I am indebted also to all members of Department of Statistics and Actuarial Science. Thanks for all professors who taught me or worked with me. Thanks to Mary Lou Dufton for her warm care and prompt arrangement.

It is also important for me to express my gratefulness to the Professors taught during my Master and undergraduate stage. Thanks to Professor Roman Makarov for being my Master supervisor. Thanks to Professor Zilin Wang for helping me to start my research path.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You are always there for me. Finally, I could not have completed this dissertation without the support of my friends,especially, the girl I met.

A big thanks to my cat. Without her, this dissertation could be finished several months earlier.

Dedication

This dissertation is dedicated to my beloved people who have meant and continue to mean so much to me. Although they are no longer of this world, their memories continue to regulate my life.

Thank you to my supervisors who guided my not only in my study but also my life. Thank you to the committee who directed me on the right track.

For my parents who helped me in all things great and small.

This dissertation is dedicated to my girl who created so much unforgettable and valuable memory with me together.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

To let the banks survive during the worst economic scenarios, financial regulators require banks and other financial institutions to hold specific capital, which is called Regulatory Capital (RC). In mathematical terms, the regulatory capital is defined as the difference between the 99.9 percentile of the portfolio's loss and the expected portfolio loss. For the loans extended by financial institutions, Basel II requires the institution to calculate the Regulatory Capital (RC) based on the following formula:

$$RC = \sum_{i=1}^{N} EAD_i \cdot LGD_i \left[\Phi \left(\frac{\Phi^{-1}(PD_i) - a_i \Phi^{-1}(0.001)}{\sqrt{(1 - a_i^2)}} \right) - PD_i \right], \tag{1.1}$$

where EAD_i is the exposure at default, LGD_i is the loss given default, PD_i is the default probability, a_i is the factor loading, which describes the correlation among the risk units in the market, and Φ is the cumulative distribution function of the standard normal distribution. More details on this formula can be found in Basel Committee on Banking Supervision (2005).

Formula 1.1 was introduced and justified by Gordy (2003) in the Asymptotic Single Risk Factor (ASRF) framework as an easy approximation to the accurate value of RC within the model developed by Vasicek (2002). However, Kupiec (2009) documented empirically that the value given by Formula 1.1 always underestimates the probability that these high default rates happen. As a result, the RC derived from Formula 1.1 may be insufficient. Breitung and Eickmeier (2011) also support this claim.

There are several potential reasons for the underestimation in the Vasicek default model:

- Implied correlations among the loans (i.e., factor loadings) do not depend on the overall state of the economy and remain constant over time.
- Lack of time dynamics in the model, which is also supported by empirical evidence.
- The Underlying Gaussian distribution, which is well-known for its thin tail property, and is also criticized by researchers for computing the quantile of the default rate.
- Another serious shortcoming of the traditional Vasicek model is the Absence of randomness of LGD.

For each individual problem, researchers have proposed new models to fix them respectively. In our thesis, we propose a new model, which combines the ideas from existing models, to improve the first three problems identified in the above list.

For the constant factor loading problem, Cheng et al. (2016) suggests an econometric procedure that can identify a small number of significant breaks in the factor loadings. Meanwhile, Pelger and Xiong (2019) presents a state-varying factor model of large dimensions that assumes that the state factor loading changes according to some other observable data. Recently, Metzler (2020) proposes a State Dependent model (SDM) that generalizes factor loading from a constant to a function of a state variable. The model assumes that the state process which controls the factor loadings can not be observed directly. The author applied a maximum likelihood method to estimate the model parameters and demonstrated the significant impact of the market correlation on the amount of RC.

Absence of time dynamics is supported by the plots of the Federal Reserve Data's historical observations[1]. We can tell that there exists a strong correlation of the underlying exposures over time periods. In order to capture the dependence structure and improve the model's predictive ability, we introduce additional parameters to control the temporal dependence in the market. To simplify the computational workload, we assume that each time period is conditionally independent given the systematic risk factor. As a reasonable starting point, we adopt an AR process to model the underlying systematic risk factor. Since the AR(1) process may not be accurate enough to describe the risk factor dynamics, we plan to replace it with other structures later. We name this model as SDM-AR model.

In addition, the implicit usage of the Gaussian copula in the traditional Vasicek model is another potential source of the underestimation. The consequences of the 2008 financial crisis warned the researchers that zero tail dependence of Gaussian copula could lead

[1] Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks. https://www.federalreserve.gov/releases/chargeoff/deltop100sa.htm

to a fatal result. To overcome this disadvantage, we further extend SDM-AR model by introducing additional parameter to replace the Gaussian copula with a Student t-copula. We call the new model t-SDM-AR model.

The most straightforward (and the most natural) extension of the Vasicek model for LGD is proposed in Frye (2000), which assumes that LGD is a second risk indicator driving credit losses, and LGD is a function of collateral. We leave this extension as a potential future research direction.

Nevertheless, when the value of the state process is unobservable as it is in the Vasicek model, we have to deal with a model that contains some latent variables. Due to our assumption about the temporal dependence structure, the EM-algorithm we apply in the independent case is hard to apply here. Therefore, we need to find some other approaches to estimate the model. After implementing the model, we found that it is challenging to ensure that our results are robust with respect to the choice of the initial value because there exist some local maximum points. The problem of initial points selection will also be discussed in the part of empirical study.

We go through some necessary background knowledge in the following several subsections before presenting and discussing the SDM, SDM-AR and t-SDM-AR models.

1.1 Asymptotic single risk factor (ASRF) framework

Consider a portfolio that consists of N exposures. Let EAD_i and LGD_i denote the exposure at default and loss given default for the ith loan. Under the ASRF, both EAD_i and LGD_i are assumed to be constant and known. Then we use $\omega_i = LGD_i / \sum EAD_i$ to denote the relative size of loss when the default of ith loan occurs. So we can represent the total percentage loss of the portfolio in the following way:

$$L_N = \sum_{i=1}^{N} \omega_i \cdot B_i, \tag{1.2}$$

where B_i is an indicator variable that equals one when the ith loan is in default and 0 otherwise. The default probability for the ith loan is $PD_i = P(B_i = 1)$. Then the expected total loss is

$$E[L_N] = \sum_{i=1}^{N} \omega_i \cdot PD_i. \tag{1.3}$$

In the ASRF model proposed by Gordy (2003), there exists a random variable M to represent the overall economic performance. This variable is used as a systematic risk factor. Larger values of M represent a better state of the economy at the current time. It is natural for us to assume that the default rates of the loans are negatively related to M.

1.2 Vasicek model

The Vasicek Model assigns a value X_i, called credit quality, to the ith loan, where

$$X_i = a_i M + \sqrt{1 - a_i^2} \cdot \epsilon_i \quad i = 1, 2. \dots, n, \tag{1.4}$$

and

- n is the number of loans in the portfolio.
- M, and $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are iid $N(0, 1)$ random variables.
- $\{a_i\}_{1,\dots,n}$ is a sequence of constants that belong to $[0, 1]$.

Under this setting, the random variable X_i follows a standard normal distribution and $Cov(X_i, X_j) = a_i a_j$ for any $i \neq j$. As it considers a single common factor, and both common and idiosyncratic factors follow the normal distribution, the Vasicek single-factor model is equivalent to a single-factor Gaussian copula.

The ith loan defaults if $X_i \leqslant \Phi^{-1}(PD_i)$, where Φ^{-1} is the inverse of the cumulative distribution function (CDF) of a standard normal distribution, and PD_i is a parameter of the model, which denotes the probability of default for the ith loan.

M is the common systematic risk factor, and ϵ_i is the independent idiosyncratic risk factor associated with the ith loan. It follows that $P(B_i = 1) = P(X_i \leqslant \Phi^{-1}(PD_i)) = PD_i$.

The parameter a_i controls a trade-off between systematic and idiosyncratic risk factors. It is easy to see that X_i also follows a standard normal distribution. The correlations among the $\{X_i\}_{1,\dots,n}$ come from the common systematic risk factor M. Then we can calculate the conditional default probability given the systematic risk factor:

$$\begin{aligned} P(B_i = 1|M) &= P(X_i \leqslant \Phi^{-1}(PD_i)|M) \\ &= P(a_i M + \sqrt{1 - a_i^2}\epsilon_i \leqslant \Phi^{-1}(PD_i)|M) \\ &= \Phi\left(\frac{\Phi^{-1}(PD_i) - a_i M}{\sqrt{1 - a_i^2}}\right). \end{aligned}$$

1.2.1 Homogeneous Vasicek model

In both Metzler's (2020) and our study, we deal with the homogeneous version of the Vasicek model. In this case we let $a_i = a$ and $PD_i = PD$ for all i, where a and PD are constants. Since the portfolio is homogeneous and large, all loans are of the same size, and they are conditionally independent with each other when M is known, based on the law of large number (LLN), the following result holds (i.e. if the portfolio is large, LLNs ensures that the fraction of obligors that actually defaults is almost surely equal to the individual default probability.):

$$D = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} B_i = \Phi\left(\frac{\Phi^{-1}(PD) - aM}{\sqrt{1-a^2}}\right),$$

where D is the default rate. The thing we need to notice is that the systematic risk factor determines the value of D. Since D is a function of M, D is also a random variable. Our first goal is to find a limiting marginal distribution of the default rate, which is

$$\begin{aligned} P(D > d) &= P\left(\Phi\left(\frac{\Phi^{-1}(PD) - aM}{\sqrt{1-a^2}}\right) > d\right) \\ &= P\left(\frac{\Phi^{-1}(PD) - aM}{\sqrt{1-a^2}} > \Phi^{-1}(d)\right) \\ &= P\left(M < \frac{\Phi^{-1}(PD) - \sqrt{1-a^2} \cdot \Phi^{-1}(d)}{a}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{1-a^2} \cdot \Phi^{-1}(d)}{a}\right). \end{aligned} \tag{1.5}$$

As we can see from Equation 1.5, the loan portfolio loss distribution is fully characterized by two parameters: the probability of default, PD, and the asset correlation coefficient, a. The former parameter fixes the expected loss rate of the portfolio, while the latter controls the shape of the loss distribution.

By differentiating $1 - P(D > d)$, we get the probability density function of the default rate for a large portfolio in the Vasicek Model. It is:

$$f_{x,d}(d) = \sqrt{1-a^2}\frac{\phi(\sqrt{1-a^2}\Phi^{-1}(d); x, a^2)}{\phi(\Phi^{-1}(d))}. \tag{1.6}$$

1.2.2 Drawbacks of the Vasicek model

Kupiec (2009) illustrated that the value given by Formula 1.1 always underestimates the probability these high default rates happen. As a result, the RC may be insufficient if we use that formula. In the study of Breitung and Eickmeier (2011), they found the factor loadings can change either smoothly or abruptly. So, we will first talk about some well-known drawbacks of the traditional Vasicek model. After that, we present a brief introduction about the Vasicek model's extensions to overcome those problems.

Absence of randomness of LGD

In the traditional Vasicek model presented in Section 1.2.1 and Formula 1.1, we assume the LGD is a constant, which does not depend on the systematic risk factor M. However, the empirical evidence strongly suggests that the LGD should be a random variable and correlates with the overall market level. The literature supporting this observation is large and growing, such as Acharya et al. (2003), Frye (2000) and Andersen et al. (2004). In the paper of Andersen (2004), they proposed a model that split the recovery rate into a systematic factor term and an idiosyncratic one.

Flat correlation

The other well-known issue is the assumption made by Vasicek that the implied correlations among the loans (i.e. factor loadings) do not depend on the overall market behavior and remain constant all the time. Overcoming this problem has attracted a lot of attention. Burtschell et al. (2005) proposed two models named as stochastic correlation model and state-dependent correlation model respectively.

Lack of tail dependence

Last but not least, most of the models applied in the industry have an implicit assumption of a joint Gaussian distribution of the obligors' asset value. With more and more efforts have been devoted into the area of credit modelling, researchers have noticed the limitations of the Gaussian distribution, especially tail independence. Salmon (2012) and Balla (2014) suggest that lacking tail dependence is the main reason for the financial crisis in 2008. Schloegl, O'Kane (2005) and Pimbley (2008) proposed similar methodologies to replace the Gaussian copula with a Student t-copula.

1.2.3 Extension of the Vasicek model

For the purpose of overcoming the problems mentioned in the previous section, researchers have proposing new models basically from two main streams, the factor load model and copula based model. Although the paper published by Daivd Li (2000) realized that those two methods have a strong relation with each other, it is easy for us to understand the background information in this area by treating them separately. In the following part, we talk about some well-developed extensions of the Vasicek model.

Student-t and Archimedean copula

Using a multivariate normal distribution to describe the obligors' asset value leads to a so-called Gaussian Copula. Copulas concentrate on the dependency among the random variables, and the marginal distribution is irrelevant. Nevertheless, Gaussian Copula is just one of many copulas. Frey and McNeil (2001) replaced it with a student-t copula. There are two benefits by doing so. First, a t-distribution converges to the Gaussian distribution as the degree of freedom goes to infinity. Secondly, the t-copula has the property of tail dependence to capture extreme events with higher probabilities than the Gaussian copula.

Except for the t-copula, Schonbucher (2002) compared the difference between Gaussian copula and some other Archimedean copulas: Clayton, Gumbel and Frank copula. In this research work, they fixed individual default probabilities, exposure sizes and even the pairwise default correlations in order to separate the effects of the dependency structure. The result suggests that modelling the credit risk by adopting other copula is feasible. Also, the Gaussian copula and the Clayton copula imply almost identical loss distribution, while the Gumbel copula are considerably different from the two aforementioned copulas.

Factor loading model

A couple of models proposed in the paper by Burtschell, Gregory and Laurent (2005) concentrate on overcoming the constant correlation coefficient assumption. The main idea is to change the constant factor loading in Formula 1.4 by some random variables, which can be either independent or dependent with the systematic risk factor.

$$X_i = \tilde{a}_i M + \sqrt{1 - \tilde{a}_i{}^2} \cdot \epsilon_i \quad i = 1, 2\ldots, n, \tag{1.7}$$

where M, ϵ_i are Gaussian random variables, all these being jointly independent, $\tilde{a}_i$ are some random variables taking values in $[-1, 1]$. There are two concrete examples in this paper.

The first one is the case that the factor loading is independent with the systematic risk factor M.

$$X_i = ((1 - B_i)a_l + B_i a_h)M + \sqrt{1 - ((1 - B_i)a_l + B_i a_h)^2} \cdot \epsilon_i \quad i = 1, 2. \dots, n,$$

where B_i are independent Bernoulli random variables. Depending on B_i equaling to 0 or 1, we have a factor loading equal to a_l or a_h. The second one introduces dependence between factor loading and systematic risk factor.

$$X_i = \alpha + (a_l \mathbb{1}_{\{M \leqslant e\}} + a_h \mathbb{1}_{\{M > e\}})M + v \cdot \epsilon_i \quad i = 1, 2. \dots, n,$$

where a_l, a_h, e are some input parameters, $a_l, a_h \in [-1, 1]$. α and v are some constants to ensure that $E[X_i] = 0$ and $E[X_i^2] = 1$.

The paper shows that a simple class of stochastic correlation models can provide a reasonably good fit to the default rate. Later, we will show that all those two models are actually two special cases of the model we proposed.

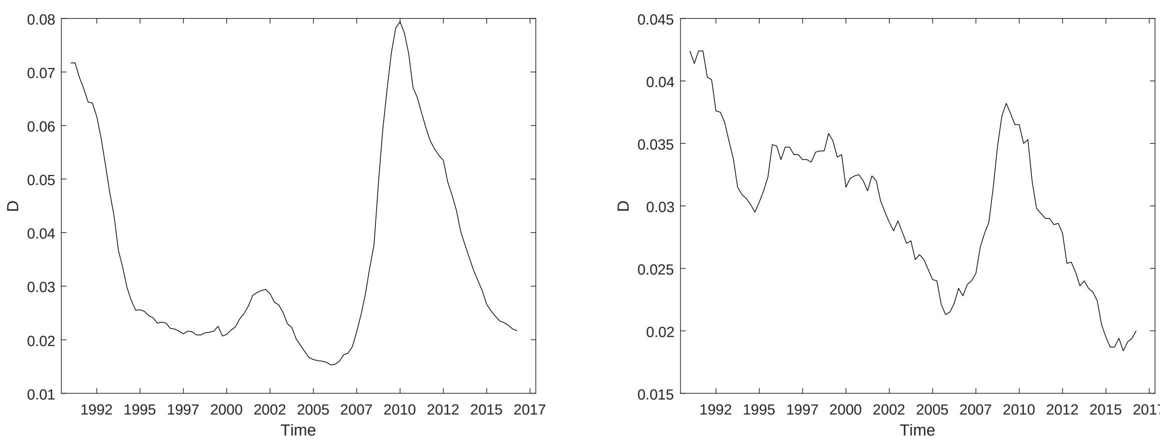
1.3 Federal reserve delinquency rates

In this section, we briefly demonstrate the historical data we will study in this thesis. The data set is called Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks. It is available on the website of the Board of Governors of the Federal Reserve System. The 100 largest banks are measured by consolidated foreign and domestic assets. The data set consists of quarterly delinquency rates for 11 different categories. The time period is from the first quarter of 1991 to the fourth quarter of 2016. The following table shows the 11 categories we used in our study.

Series	Abbreviation	Mean(%)	Std. Dev(%)	Kurtosis	Skewness
All	All	3.49	1.85	2.62	1.01
Business	B	2.26	1.42	3.52	1.15
Consumer	C	3.48	0.79	2.55	-0.04
Credit Card	CC	4.16	1.15	2.75	-0.05
Other Consumer	OC	2.99	0.6	2.29	-0.06
Agricultural	AG	3.3	1.74	3.11	0.90
Large Format Retail	LFR	1.36	0.54	2.29	0.66
Secured By Real Estate	SRE	4.77	3.19	1.99	0.70
Farmland	F	3.87	1.88	3.28	0.92
Mortgages	M	4.58	3.59	2.63	1.11
Commercial Real Estate	CRE	4.64	4.53	4.10	1.44

Table 1.1: Data categories

The following plots show the default rate for All and Other Consumer series.



(a) All series (b) Other Consumer series

Figure 1.1: Historical data plot

As we can see from the above figures, the data shows extreme persistence over time. The data shows strong visual evidence that temporal dependence structure is important for modelling default rate.

1.4 Expectation-maximization algorithm

In our thesis, since the market factor is latent, the EM-algorithm will be a key tool for us to estimate the model parameters. The expectation-maximization (EM) algorithm, introduced and named by Dempster, Laird and Rubin (1977), is an iterative method to find maximum likelihood estimates of parameters in a model with latent variables. Given a statistical model, a set of observations X, parameter θ, and a set of latent data Z that can not be observed directly, then we can write the joint likelihood function $L(\theta; X, Z) = p(X, Z|\theta)$ and the marginal likelihood of the observed data $L(\theta|X) = \int p(X, Z|\theta)dZ$.

The EM-algorithm is divided into two steps. The first step is the Expectation step. Define the expected value of the log-likelihood function of θ with respect to the current conditional distribution of Z given X and the current estimates of the parameters $\theta^{(t)}$,

$$Q(\theta|\theta^{(t)}) = E_{Z|X,\theta^{(t)}}[\log L(\theta; X, Z)].$$

The second step is the Maximization step, which is finding the parameter that maximizes Q.

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)}).$$

So first, we initialize the parameters θ by assigning an initial value. Then we repeat the E-step and M-step until the sequence $\theta^{(n)}$ is deemed to converge. As proven by Roderick and Rubin (1987), this algorithm will converge to a local maximum point.

Chapter 2

Static State-Dependent Model

In this chapter, we discuss the model proposed by Metzler (2020). The author has provided empirical evidence that correlations between financial assets change through time in some systematic ways. The explanation for this is illustrated by Campbell, Koedijk (2002) and Melkuev (2014) who show that the correlations usually rise dramatically during a financial crisis. In Metzler's paper, the author proposes a generalized version of the homogeneous Vasicek model, the so-called State-dependent Model (SDM), that allows us to capture this stylized fact. In this Chapter, we first apply the EM algorithm to estimate the model parameters. After that, we further generalize the factor loading of the SDM model to be more flexible.

The main idea employed in the SDM model is to introduce a standard normal latent variable T. First, T is correlated with the systematic risk factor M. Secondly, the factor loading becomes a function of T. Then the credit score for the i^{th} loan can be written as

$$X_i = a(T)M + \sqrt{1 - a(T)^2} \cdot \epsilon_i, \tag{2.1}$$

where $a : \mathbb{R} \to [0, 1]$ is a function. The correlation, β, between M and T is a new parameter in the model. It is natural to treat β as a measure of how closely the factor loading changes with respect to the overall market. The other notations remain the same as in Model 1.4.

For simplicity, we begin with a simple function $a()$ of the form:

$$a(t) = \sum_{k=1}^{K} a_k \cdot \mathbb{1}(t_{k-1} < t \leqslant t_k),$$

where $0 \leqslant a_1 < a_2 < \cdots < a_K \leqslant 1$ are the possible values for the factors loadings and $-\infty = t_0 < t_1 < \cdots < t_K = \infty$. Under this setting, we bring another notation to indicate the market regime determined by T:

$$R = \sum_{k=1}^{K} k \cdot \mathbb{1}(t_{k-1} < T \leqslant t_k). \tag{2.2}$$

Since $\{t_1, t_2, \ldots, t_K\}$ is a partition of the real line, the market would be in only one regime at a time.

2.1 Properties of the model

In this section, we first motivate further Metzler's (2020) model and then present some of its properties proved by the author. Compared to the traditional Vasicek model, the factor loading is linked to the systematic risk factor M via the random variable T. Since M and T are correlated with each other, and the factor loading $a(T)$ is a linear function of T, it is easy to show that $a(T)$ and M are correlated as well. As we have explained earlier, the correlation, β, between T and M works as a measure of the state dependence. For example, if β is large, then there exists a very high probability that the correlation among the loans is high when the overall economic level is low. The model includes both the Gaussian mixture model and the Random Factor loading model as special cases when $\beta = 0$ and $\beta = 1$ respectively. The Vasicek model mentioned in Chapter 1 is also a special case when the function $a(\cdot)$ is constant.

2.1.1 Conditional market correlation for each regime

Since the goal of SDM is to capture the phenomenon that market correlation is higher in the stressed market scenario, the first thing presented by Metzler (2020) is how the probability distribution for different regimes changes based on the given value of the systematic risk factor M. We can calculate the conditional probability of the k^{th} regime given the realized value of the systematic risk factor as:

$$p_k(m) = P(R = k | M = m) = \Phi(t_k; \beta m, 1 - \beta^2) - \Phi(t_{k-1}; \beta m, 1 - \beta^2), \tag{2.3}$$

where R is defined in Equation 2.2. The market correlation given that we are in regime k, $Corr(X_i, X_j | R = k)$, can be calculated by the following procedure.

1. We need the conditional variance of the credit quality given M:

$$Var(X_i | R = k) = a_k^2 Var(M | R = k) + (1 - a_k^2) Var(\epsilon_i | R = k).$$

$Corr(M, R) \neq 0$ since R is fully determined by T, which is correlated with M.

2. The conditional variance of M given regime k, $\sigma_k^2 = Var(M|R = k)$, can be found in the following theorem.

   Theorem 1. *Suppose that* $\binom{X_1}{X_2}$ *follows a bivariate normal distribution with mean* $\binom{\mu_1}{\mu_2}$ *and variance* $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$*. Let* $\alpha = (a - \mu_1)/\sigma_1$ *and* $\beta = (b - \mu_1)/\sigma_1$*. Then we have*

   $$E[X_2|a \leqslant X_1 < b] = \mu_2 - \rho\sigma_1 \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right) \tag{2.4}$$

   $$Var[X_2|a \leqslant X_1 < b] = \sigma_2^2 + \rho^2\sigma_1^2 \left[-\frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right]. \tag{2.5}$$

   The details of the proof can be found in the paper of Metzler (2020). By applying Theorem 1, we have

   $$\sigma_k^2 = 1 - \beta^2 \frac{t_k\phi(t_k) - t_{k-1}\phi(t_{k-1})}{\Phi(t_k) - \Phi(t_{k-1})} + \left(-\beta \frac{\phi(t_k) - \phi(t_{k-1})}{\Phi(t_k) - \Phi(t_{k-1})} \right)^2. \tag{2.6}$$

3. Since ϵ_i and R are independent of each other, $Var(\epsilon_i|R = k) = Var(\epsilon_i) = 1$. As a result,

   $$Var(X_i|R = k) = a_k^2\sigma_k^2 + (1 - a_k^2)$$

   and

   $$Cov(X_i, X_j|R = k) = a_k^2 Var(M|R = k) = a_k^2\sigma_k^2.$$

2.1.2 Regime probabilities given a systematic risk factor

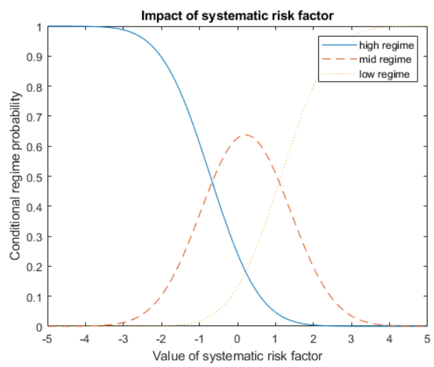
We present some properties of the regime changes given the value of the systematic risk factor M. The unconditional regime probability, p_k, is determined by the value of t_k

$$p_k = \Phi(t_k) - \Phi(t_{k-1}).$$

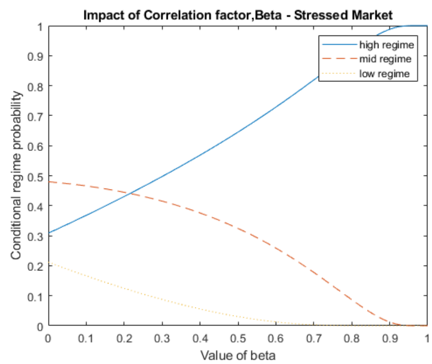
But the conditional regime probability given the systematic risk factor M is also determined by the value of both M and β. In Metzler's paper, the author shows that

$$\begin{aligned} p_k(m) &= P(R = k|M = m) \\ &= \Phi(t_k; \beta m, 1 - \beta^2) - \Phi(t_{k-1}; \beta m, 1 - \beta^2), \end{aligned} \tag{2.7}$$

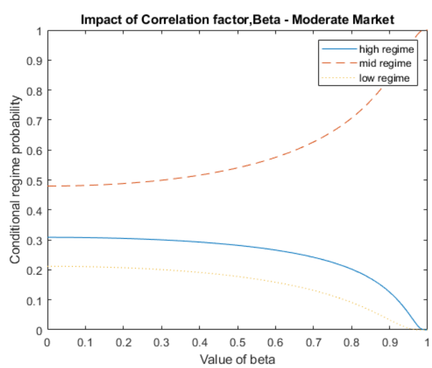
where $\Phi(\cdot, m, \sigma^2)$ is the CDF of the normal distribution with mean m and variance σ^2. Due to the fact that $t_0 = -\infty$ and $t_K = \infty$, it follows that when the overall market is extremely bearish, the market will be in a high correlation state almost surely, $p_1(-\infty) = 1$, and vice versa. But it is hard to find a general way to infer the impact of both M and β on $p_k(m)$ for the intermediate regimes. The following plots demonstrate some interesting properties of the relations between $p_k(m)$, M and β.



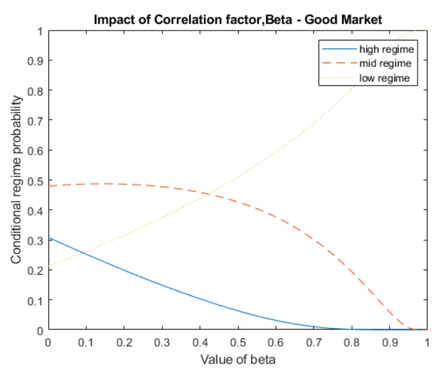
(a) Impact of systematic risk factor



(b) Impact of Beta when overall market is stressed (i.e. $M = \Phi^{-1}(0.05)$)



(c) Impact of Beta when overall market is moderate (i.e. $M = \Phi^{-1}(0)$)



(d) Impact of Beta when overall market is good (i.e. $M = \Phi^{-1}(0.95)$)

Figure 2.1: Impact of M and β on Conditional Regime Probability Property. All plots are drawn under the framework of three regimes. The values of $\{t_i\}_{0,\ldots,3} = [-\infty,\ -0.5,\ 0.8,\ \infty]$. (a) represents the impact of M on the conditional regime probability with $\beta = 0.7$. (b),(c) and (d) show the impact of β on the conditional regime probability with different market levels.

All of the plots in Figure 2.1 are drawn under the assumption that the market can be in one of the three regimes: high, moderate and low correlation regimes. The values of $\{t_i\}_{0,\ldots,3}$ are $[-\infty,\ -0.5,\ 0.8,\ \infty]$. Graph (a) shows the impact of M on the conditional regime probability. We set $\beta = 0.7$. As we can see, the probability of the high correlation regime is decreasing with respect to M, and conversely, the probability of the low correlation regime is increasing with respect to M. But, the probability that the market is in the regime with a moderate correlation is bell-shaped. Parts (b), (c) and (d) demonstrate the impact of β on the conditional regime probability when the overall market is under stressed, moderate and good scenarios, respectively. It is easy to see that the regime is fully determined by the systematic risk factor M when $|\beta|$ is large enough. That means when β equals 1, the conditional regime probabilities become binary (either 1 or 0) depending on the realized value of the systematic risk factor.

2.1.3 Default threshold of credit quality X_i

Due to the fact that the SDM model relax the constant factor loading assumption, the distribution of credit score also changes. So it is necessary for us to find the marginal distribution and default threshold of the credit scores defined in the formula (2.1). In order to do so, we calculate the conditional cumulative distribution function of credit score first. It has been shown in Metzler's paper (2020) that:

$$P(X_i \leqslant x | M = m) = \sum_{k=1}^{K} [\Phi(x; a_k m, 1 - a_k^2) \cdot p_k(m)]. \tag{2.8}$$

Then, the unconditional CDF of X_i can be found by:

$$P(X_i \leqslant x) = \int_{-\infty}^{\infty} \left[\sum_{k=1}^{K} \Phi(x; a_k m, 1 - a_k^2) \cdot p_k(m) \right] \cdot \phi(m) dm. \tag{2.9}$$

The term inside the large brackets in Equation 2.9 can be rewritten further in the form of:

$$\begin{aligned} \Phi(x; a_k m, 1 - a_k^2) \cdot p_k(m) =& \Phi(x; a_k m, 1 - a_k^2) \cdot (\Phi(t_k; \beta m, 1 - \beta^2) - \Phi(t_{k-1}; \beta m, 1 - \beta^2)) \\ =& \Phi(x; a_k m, 1 - a_k^2) \cdot \Phi(t_k; \beta m, 1 - \beta^2) \\ & - \Phi(x; a_k m, 1 - a_k^2) \cdot \Phi(t_{k-1}; \beta m, 1 - \beta^2). \end{aligned} \tag{2.10}$$

The product of two normal CDFs can be treated as a 2-dimensional normal CDF with correlation 0, original means and variances, respectively:

$$\Phi(x_1; \mu_1, \sigma_1^2) \cdot \Phi(x_2; \mu_2, \sigma_2^2) = \Phi_2 \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right), \tag{2.11}$$

where $\Phi_2(x;\mu,\Sigma)$ is the CDF of the 2-dimension normal distribution with mean μ and covariance matrix Σ. By applying Equation 2.11 to the last line of Equation 2.10, we can get:

$$\begin{aligned}\Phi(x;a_km,1-a_k^2)\cdot p_k(m) =&\Phi_2\left(\begin{bmatrix}x\\t_k\end{bmatrix};\begin{bmatrix}a_km\\\beta m\end{bmatrix},\begin{bmatrix}1-a_k^2 & 0\\0 & 1-\beta^2\end{bmatrix}\right)\\&-\Phi_2\left(\begin{bmatrix}x\\t_{k-1}\end{bmatrix};\begin{bmatrix}a_km\\\beta m\end{bmatrix},\begin{bmatrix}1-a_k^2 & 0\\0 & 1-\beta^2\end{bmatrix}\right).\end{aligned} \tag{2.12}$$

After switching the order of summation and integral in Equation 2.9, we can represent the unconditional CDF in the following form:

$$P(X_i<x)=\sum_{k=1}^{K}E_M\left[\Phi_2\left(\begin{bmatrix}x\\t_k\end{bmatrix};\begin{bmatrix}a_kM\\\beta M\end{bmatrix},\begin{bmatrix}1-a_k^2 & 0\\0 & 1-\beta^2\end{bmatrix}\right)-\Phi_2\left(\begin{bmatrix}x\\t_{k-1}\end{bmatrix};\begin{bmatrix}a_kM\\\beta M\end{bmatrix},\begin{bmatrix}1-a_k^2 & 0\\0 & 1-\beta^2\end{bmatrix}\right)\right]$$

The following theorem proven by Metzler (2020) can be used to further simplify the above equation.

Theorem 2. *Let M have a standard normal distribution and let $a=[a_1,\ldots,a_n]^T$, $b=[b_1,\ldots,b_n]^T$. Then,*

$$E[\Phi_n(a+bM,0,\Sigma)]=\Phi_n(a,0,\Sigma+bb^T), \tag{2.13}$$

where $\Phi_n(x;\mu,\Sigma)$ is the n-dimension normal CDF with mean μ and covariance matrix Σ.

According to the theorem, we have

$$P(X_i<x)=\sum_{k=1}^{K}\left[\Phi_2\left(\begin{bmatrix}x\\t_k\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \beta a_k\\\beta a_k & 1\end{bmatrix}\right)-\Phi_2\left(\begin{bmatrix}x\\t_{k-1}\end{bmatrix};\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \beta a_k\\\beta a_k & 1\end{bmatrix}\right)\right]. \tag{2.14}$$

Although the CDF of the credit score can be expressed in closed form, it may be too time-consuming or even impossible to find the closed form for its inverse CDF. As a result, we can find the credit threshold x_{pd} by solving the equation $P(X_i<x_{pd})=PD$ numerically. Since the CDF of the credit quality is monotone increasing and also bounded between 0 and 1, there exists exactly one solution to that equation.

2.1.4 Distribution of large portfolio default rate

In this section, we take a look at the distribution of the default rate D. Since all the details of the calculations are available in Metzler's paper (2020), we skip the details and only present the results in here. The distribution is similar to the homogeneous Vasicek Model mentioned in Section 1.2.1. If we assume there exist N loans, then the default rate can be calculated as:

$$\begin{aligned} D &= \sum_{i=1}^{N} \frac{B_i}{N} \\ &= \sum_{i=1}^{N} \frac{\mathbb{1}_{(X_i < x_{PD})}}{N}. \end{aligned}$$

Once we are given the condition that $M = m$ and $R = k$, then $\{X_i\}_{1,\ldots,N}$ become independent of each other. Based on the LLNs, we can approximate the default rate by the following formula when we have a large enough number of loans:

$$\begin{aligned} (D|M = m, R = k) &= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{\mathbb{1}_{(X_i < x_{PD}|M=m,R=k)}}{N} \\ &= E[\mathbb{1}_{(X_i < x_{PD}|M=m,R=k)}] \\ &= P(X_i < x_{pd}|M = m, R = k). \end{aligned} \tag{2.15}$$

It is easy to see that once we are given the condition that $M = m$ and $R = k$, we can treat the state-dependent model as a classic Vasicek model with the same systematic risk factor value of $M = m$ and factor loading $a = a_k$. So

$$D = \sum_{k=1}^{K} \left[v_k(M) \cdot \mathbb{1}_{(R=k)} \right], \tag{2.16}$$

where x_{pd} is the default threshold corresponding to the default probability PD and $v_k(M) = \Phi\left(\frac{x_{pd} - a_k M}{\sqrt{1-a_k^2}}\right)$ is the Vasicek curve corresponding to a default threshold x_{pd} and the factor loading a_k. The intuition is straightforward. Once the regime is fixed, the relation between the systematic risk factor M and the default rate D is the same as the one in the Vasicek model with the factor loading a_k.

The next thing we need to determine is the distribution function of D. For each regime, the default rate will exceed a certain threshold value d only when $v_k(M) > d$. According to the definition of the model and the fact that events are mutually exclusive, we have:

$$\mathbb{1}_{(D>d)} = \sum_{k=1}^{K} \mathbb{1}_{(v_k(M)>d,R=k)}.$$

After taking expectation on both sides of the above equation, we get:

$$P(D > d) = \sum_{k=1}^{K} P(M < v_k^{-1}(d), T \in [t_{k-1}, t_k)),$$

where $v_k^{-1}()$ is the inverse function of $v_k()$ as shown above.

We know that M and T follow a 2-dimensional standard normal distribution with a correlation β. So

$$\begin{aligned} P(D > d) &= \sum_{k=1}^{K} P(M < v_k^{-1}(d), T \in [t_{k-1}, t_k)) \\ &= \sum_{k=1}^{K} \left[\Phi_2 \left(\begin{bmatrix} v_k^{-1}(d) \\ t_k \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix} \right) - \Phi_2 \left(\begin{bmatrix} v_k^{-1}(d) \\ t_{k-1} \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix} \right) \right]. \end{aligned}$$

By differentiating the function $1 - P(D > d)$, we can obtain the density of the default rate, denoted by $f_D(d)$:

$$f_D(d) = \sum_{k=1}^{K} p_k(v_k^{-1}(d)) \cdot f_k(d), \tag{2.17}$$

where $f_k(d) = f_{x_{pd},a_k}(d)$ is the Vasicek density function, defined in Equation 1.6 with a default threshold x_{pd} and a factor loading a_k. $p_k()$ is the conditional probability of regime k defined in Equation 2.3.

2.1.5 Joint PDF of D and R

It is natural to consider the joint distribution function of D, M and R. But the interesting fact is that under the assumption that the number of loans is infinite, the default rate, D, is totally determined by the values of M and R. For example, if we know which regime we are in and the value of the systematic risk factor, then D can be calculated by Equation 2.16. This is equivalent to saying that D is a deterministic function of M and R. In its original formulation, the EM algorithm needs the joint PDF of the observation data D and all latent variables M, R. But in our model, we only need the joint PDF of D and R since D is no longer a random variable, but a fixed value once M and R are given. Using a basic definition of conditional probability, we have:

$$\begin{aligned} f_{D,R}(d, k) &= P(R = k | D = d) \cdot f_D(d) \\ &= P(R = k | M = v_k^{-1}(d)) \cdot f_D(d) \\ &= p_k(v_k^{-1}(d)) f_k(d), \end{aligned}$$

where $p_k()$ is defined in Formula 2.3. We can also write this in a more general way as:

$$f_{D,R}(d,k) = \sum_{r=1}^{K} \left[p_r(v_r^{-1}(d)) f_r(d) \mathbb{1}_{(k=r)} \right]. \tag{2.18}$$

It is instructive to consider the posterior regime probability, $P(R = k|D = d)$, which is also necessary for the implementation of the EM-algorithm.

$$\begin{aligned} P(R = k|D = d) &= \frac{f_{D,R}(d,k)}{f_D(d)} \\ &= \frac{\sum_{r=1}^{K} \left[p_k(v_k^{-1}(d)) f_k(d) \mathbb{1}_{(r=k)} \right]}{f_D(d)}. \end{aligned} \tag{2.19}$$

2.2 Parameter estimation

In this section, we apply the EM-algorithm to estimate the state-dependent model on the Federal Reserve Data on quarterly delinquency rates for various types of loan from 1991Q1 to 2016Q4[1]. Loans are treated as default if they are 30 or more days overdue. Eleven loan categories are available. In our model, the default rate D is the observed data, and the regime indicator R is the latent variable. In addition to the assumptions mentioned in previous sections, we add the following assumptions when we apply the EM algorithm:

1. There only exist two different market regimes, low and high correlation regimes. (i.e. $K = 2$)

2. The quarterly data are independent and identically distributed.

For the state-dependent model with $K = 2$, the model involves five parameters, which are $\theta = \{a_h, a_l, \beta, t_1, PD\}$. a_l and a_h are the factor loadings in the low and high correlation regimes respectively. β is the correlation between M and T. t_0 and t_2 are $-\infty$ and ∞ respectively. t_1 is the value that determines the probability of the market regime. The low correlation regime will occur when $T \leqslant t_1$ and the high correlation regime will occur when $T > t_1$. PD is the probability of default.

In the first step of the EM algorithm, we need to calculate the expected value of the log-likelihood function, with respect to the conditional distribution of the latent variable

[1]Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks. https://www.federalreserve.gov/releases/chargeoff/deltop100sa.htm

R given the observed data D under the current estimate of the parameters, $\theta^{(t)}$. The following equation is used to calculate the expected value:

$$V(\theta|\theta^{(t)}) = E_{R|D,\theta^{(i)}}[\log L(\theta; D, R)], \tag{2.20}$$

where $L(\theta; D, R)$ is the likelihood function of D and R:

$$L(\theta; D, R) = \Pi_{i=1}^{n} f_{D,R}(d_i, r_i) = \Pi_{i=1}^{n} \left(\sum_{k=1}^{2} p_k(v_k^{-1}(d)) f_k(d) \mathbb{1}_{(r_i=k)} \right), \tag{2.21}$$

where $p_k()$ can be found in Equation 2.7. $v_k^{-1}(d)$ is the inverse function of $v_k(m)$, which defined by Equation 2.16. $f_k(d)$ is defined in Equation 2.17. Then, Equation 2.20 can be written in the following form:

$$V(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \left(\sum_{k=1}^{2} \log(p_k(v_k^{-1}(d_i)) f_k(d_i)) \cdot \mathbb{P}(R = k|D = d_i; \theta^{(t)}) \right). \tag{2.22}$$

Next, we proceed to the maximization step, which involves finding the parameters that maximize the expected value with respect to θ:

$$\theta^{(t+1)} = \arg\max_{\theta} V(\theta|\theta^{(t)}). \tag{2.23}$$

The entire algorithm can be summarized as follows:

1. Initialize the parameters $\theta^{(1)}$ to a starting value.
2. Calculate the expected value as a function of θ, $V(\theta|\theta^{(t)})$, with the probability of each possible result of R.
3. Maximize the expected value w.r.t θ and set $\theta^{(t+1)}$ equal to it.
4. Repeat steps 2 and 3 until $\theta^{(t)}$ converges.

In our application of the EM algorithm in software R, we have encountered the following issues. First, overflow is a problem, which needs to be fixed. Theoretically, the probability density function of the standard normal distribution, $\phi(x)$, should always be greater than 0 for any $x \in R$. But $\phi(x)$ is treated as 0 on the computer when x is far away from its mean. As a result, when we calculated the log-likelihood value, the log-likelihood function returns NaN, which causes the whole algorithm to end up with NaN. In order to solve this problem, we add an IF statement to check if $\phi(x)$ is 0 or not. If so, we change the value to 10^{-320}, which is close to the machine epsilon. The bias created from this method is minimal.

The other problem is generated from the optimization step. As mentioned in the previous section, $V(\theta|\theta^{(t)})$ does not exist in closed form, so we adopt a Nelder-Mead method to find an optimal solution. Since we apply the numerical method in five-dimensional optimization, accuracy and robustness are hard to guarantee. Therefore, we use the long-run average of the observed data D as the estimator for PD,

$$\widehat{PD} = \frac{1}{N}\sum_{i=1}^{N} D_i. \tag{2.24}$$

This helps us to lower the dimension of the optimization problem from five to four. Even so, the numerical method still performs strangely when the estimators are moving closer to the boundary. For example, when the correlation parameter β tends to 1, i.e. $\beta \to 1$, the values of other parameters start to fluctuate within a relatively small region. By choosing the results with the largest likelihood value, we can sidestep this problem.

2.2.1 Simulation study of EM algorithm

In this section, we demonstrate the accuracy of EM-algorithm before we apply it to the Federal Reserve Data. We first apply it to simulated data generated by the SDM with known parameters. The data are generated in the following procedure:

1. Select values for all the parameters $a_1, \ldots, a_k$, β, ts and PD.
2. Generate N pairs of M and T based on a multivariate standard normal distribution with a correlation β.
3. Use the simulated data M and T to calculate the value of D by Equation 2.16.

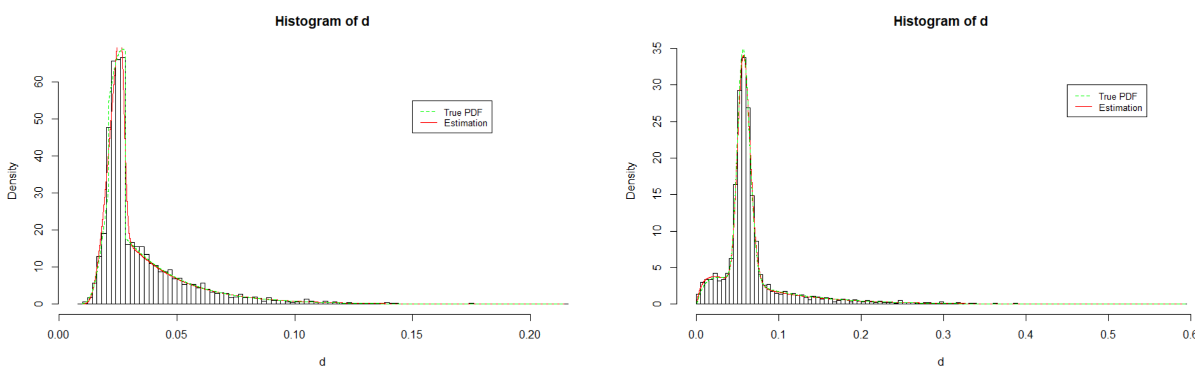
We set the number of regimes $K = 2$ in this section in order to reduce the computational workload. Later on, we will discuss the effect of the number of regimes. So, in order to reduce the estimation error caused by a low number of data points, we make $N = 5000$, which is a reasonably large number. We choose the actual value of the parameters in the following way:

	a_1	a_2	β	t_1	PD
Actual Value	0.32	0.12	1	0.174	0.035
Estimator	0.3274	0.1186	0.9803	0.1587	0.0351

	a_1	a_2	β	t_1	PD
Actual Value	0.46	0.07	0.5	-0.24	0.07
Estimator	0.4748	0.0708	0.4686	-0.2438	0.06983

Table 2.1: Simulation test of the EM-algorithm with 5000 simulated data points and two regimes. The results shown in the table are generated based on one simulation.

As we can see from Table 2.1, the EM algorithm appears to work well, and the methodology we applied to estimate PD is also accurate enough. The plots in Figure 2.2 also verify the accuracy of the EM algorithm. The initial starting points we have used here are middle points among the possible space, i.e., $a_1 = 0.51, a_2 = 0.49, \beta = 0, t_1 = 0$.



(a) Estimation on the simulation data based on the upper table in Table 2.1

(b) Estimation on the simulation data based on the lower table in Table 2.1

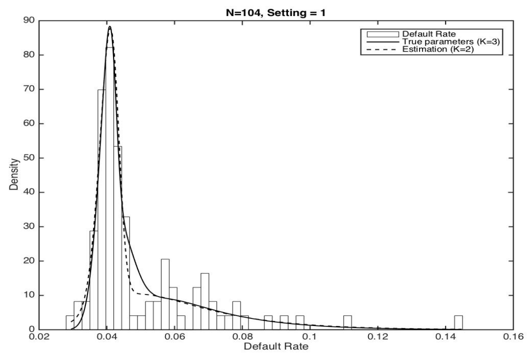
Figure 2.2: EM algorithm on simulation data: The number of points is 5000. The values of the parameters are presented in Table 2.1. The green dashed line represents the theoretical PDF of default rate, and the red solid line stands for the estimated PDF.

2.2.2 The effect of different numbers of regimes

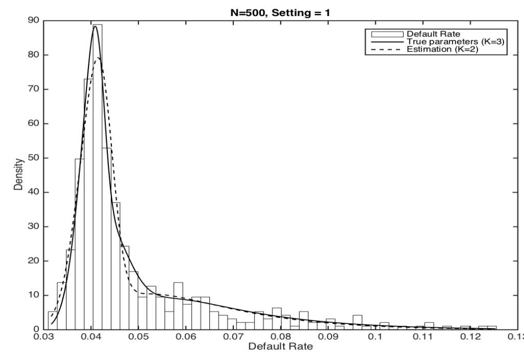
In this section, we wish to take a deeper look at the number of regimes' impact on the estimation accuracy. We examine the 2-regimes model estimation results based on the simulated data generated from a 3-regimes model. If the estimated probability density is not far away from the true one, we can reduce the computational workload by working with the 2-regimes state-dependent model in future research. The Federal Reserve Data's estimation results in the next section also suggest that the difference of the fitting results between the 2-regimes model and the 3-regimes model is negligible. We test 6 different parameter settings presented in Table 2.2 and 2 different numbers of samples are simulated. Based on the histogram plots of those fittings, it turns out that even when we underestimate the number of regimes, the model can still provide a good approximation to the true distribution of the default rate.

	a_1	a_2	a_3	β	t_1	t_2	PD
1	0.23	0.12	0.05	0.95	-0.44	0.24	0.05
2	0.18	0.12	0.09	0.95	-0.44	0.24	0.05
3	0.18	0.12	0.09	0.35	-0.44	0.24	0.05
4	0.23	0.12	0.05	0.35	-0.44	0.24	0.05
5	0.18	0.12	0.09	0.95	-0.1	0.18	0.05
6	0.18	0.12	0.09	0.35	-0.1	0.18	0.05

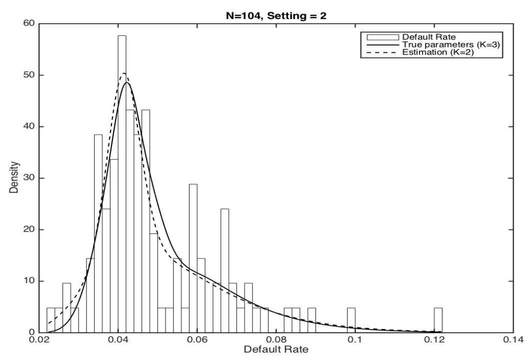
Table 2.2: Parameter Setting for simulating data from the SDM model with 3-regimes. Then, we estimate the 2-regime model to the simulated data to check the impact of fitting data with a misspecified model.



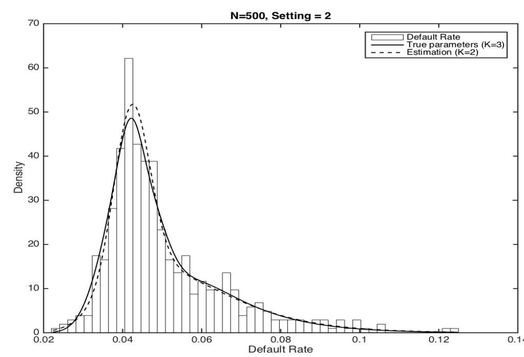
(a) Under parameter setting 1 with 104 data points



(b) Under parameter setting 1 with 500 data points



(c) Under parameter setting 2 with 104 data points



(d) Under parameter setting 2 with 500 data points

Figure 2.3: Simulation Fitting 2 regime model on 3 regime data: The solid line represents the theoretical PDF of the default rate under the 3-regime model. The dashed line shows the estimated PDF under the 2-regime model.

2.2.3 Calibration results

This section demonstrates some empirical results for the state-dependent model under the assumption that there exist 2 or 3 different market regimes, $K = 2$ or 3. The following section shows Federal Reserve Data histograms on quarterly delinquency rates for various types of loans. We also superimpose the classic Vasicek density and the density curve derived from the state-dependent model in the same histogram to compare them directly. The time period is from the first quarter of 1991 to the fourth quarter of 2016.

	ρ_h	ρ_l	β	$P(R=1)$	PD	$d_{0.001,SD}$	$d_{0.001,Vas}$	$(d_1-d_2)/d_2$	aic_{SD}	aic_{Vas}
1. All	9.94%	1.02%	1	72.57%	3.49%	17.45%	11.90%	46.60%	-610.54	-572.83
2. Bus	9.74%	4.32%	0.9928	53.75%	2.26%	12.98%	9.78%	32.73%	-636.01	-631.81
3. Cons	1.65%	0.79%	-0.9776	14.68%	3.48%	6.20%	6.81%	-8.94%	-709.81	-702.70
4. CC	3.04%	0.98%	-0.9519	18.75%	4.16%	7.81%	9.36%	-16.50%	-642.34	-623.24
5. OC	1.18%	0.42%	-0.8917	59.49%	2.99%	4.75%	5.42%	-12.31%	-765.07	-762.36
6. AG	5.65%	3.20%	1	96.35%	3.30%	12.74%	12.24%	4.13%	-575.74	-571.04
7. LFR	2.66%	0.90%	0.9964	94.21%	1.36%	4.17%	3.87%	7.70%	-809.94	-806.53
8. SRE	13.27%	4.37%	1	86.51%	4.77%	27.43%	22.34%	22.77%	-484.79	-459.76
9. F										
10. M	26.39%	3.77%	1	56.29%	4.58%	41.40%	22.51%	83.87%	-517.19	-464.51
11. CRE	18.87%	4.90%	1	88.27%	4.64%	34.67%	30.27%	14.54%	-464.93	-440.60

Table 2.3: Estimation results for Federal Reserve Data when $K = 2$. For each series we estimate the parameters for the Vasicek model and the five parameters for the state-dependent model. ρ_h and ρ_l are calculated based on the formulas in Section 2.1.1. $P(R = 1)$ refers to the probability that the market is in the high-correlation regime. $(d_1 - d_2)/d_2$ is the percentage difference in 99% quantile of Vasicek density and state-dependent model.

For each data set, we calculate the Akaike information criterion (AIC) value as well. AIC value works as a measurement of model quality. A model with a lower AIC value is usually better than the one with a higher AIC value. It turns out that the state-dependent model always has a lower value on AIC than the classic Vasicek model when $K = 2$. Also, the absolute value of the correlation term β for each series is close to 1. Most interestingly, the difference of the quantile heavily relies on the sign of the correlation term β. We can instantly notice that negative correlation brings us a lighter tail than positive correlation. These results indicate that dependence between the systematic risk factor and the factor loading has a considerable impact on the default rate distribution.

	ρ_1	ρ_2	ρ_3	β	$P(R=1)$	$P(R=2)$	$P(R=3)$
1. All	0.03583449	0.17918289	0.02122849	1	10.138%	53.105%	36.757%
2. Bus	0.10137856	0	0.00710649	0.9479	82.343%	0.258%	17.399%
3. Cons	0.01221025	0.01385329	0	0.8565	9.133%	89.708%	1.159%
4. CC	0.03964081	0.01505529	0.00000324	-0.9626	23.377%	71.883%	4.740%
5. OC	0.007056	0.01352569	0.00087025	-1	3.822%	87.185%	8.993%
6. AG	0.00160801	0.09381969	0.00160801	-1	8.219%	85.202%	6.579%
7. LFR	0.02819041	0.20223009	0.00968256	-1	41.379%	39.062%	19.558%
8.SRE	0.06140484	0.14622976	0.05803281	1	9.402%	76.372%	14.226%
9. F	0.05466244	0	0.04301476	-0.2405	34.853%	0.378%	64.769%
10. M	0.27394756	0.08491396	0.03129361	1	49.238%	35.409%	15.353%
11. CRE	0.20187049	0.05564881	0	0.9988	86.516%	13.368%	0.116%

Table 2.4: Estimation results for Federal Reserve Data when $K = 3$.

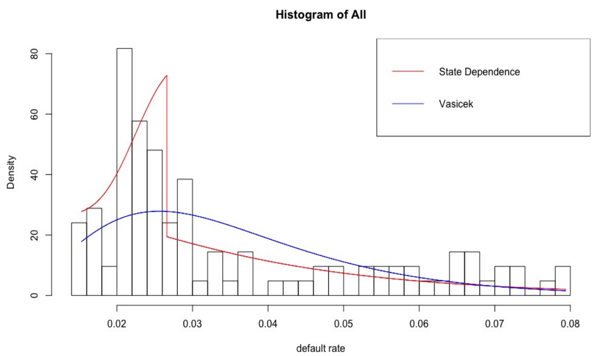
As we can see from Table 2.4, some estimation results suggest that the market may have only 2 regimes even under the assumption that there are 3 regimes. We evaluate the performance of the model based on the AIC score as before. The following table provides the AIC values for the model with $K = 2$ and $K = 3$.

	AIC ($K = 2$)	AIC($K = 3$)
1. All	-610.54	**-625.2234**
2. Bus	**-636.01**	-632.5814
3. Cons	**-709.81**	-690.0734
4. CC	-642.34	**-649.1723**
5. OC	**-765.07**	-720.7929
6. OP	**-575.74**	-516.3857
7. LFR	-809.94	**-823.4793**
8.SRE	-484.79	**-504.1335**
9. F		
10. M	-517.19	**-530.1692**
11. CRE	**-464.93**	-458.4619

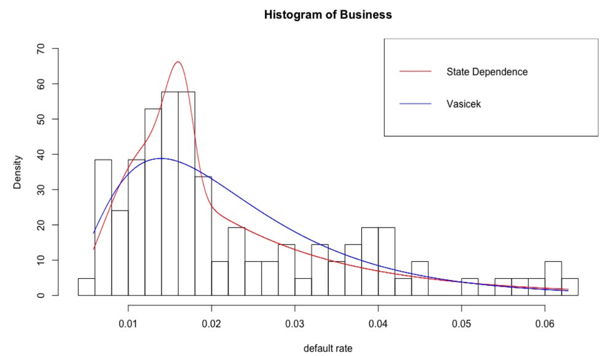
Table 2.5: AIC score comparison between $K = 2$ and $K = 3$ state-dependent model.

Based on the AIC score, we can hardly conclude that more regimes always bring us a better model. In order to reduce the computational workload and stabilize the estimator, we will mainly focus on the model with 2 regimes in future research.

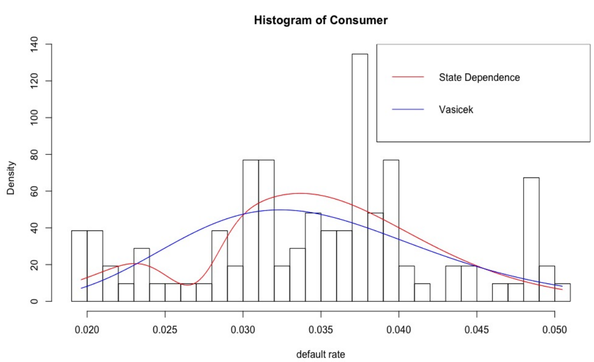
2.2.4 Histograms with the density curves



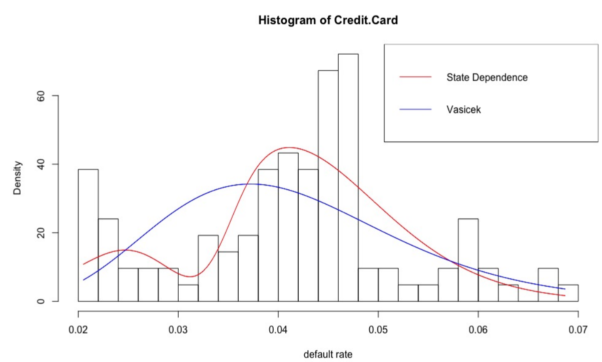
(a) All



(b) Business



(c) Consumer



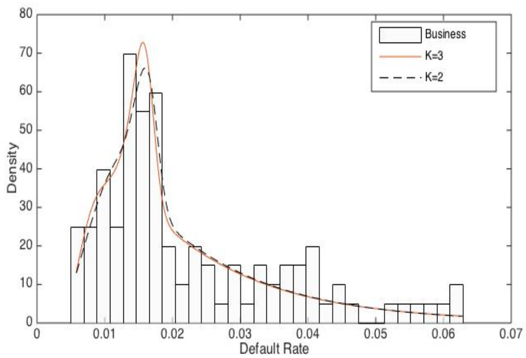
(d) Credit card

Figure 2.4: The histograms of different series with Vasicek density curve and state-dependent density curve: We can see from the histograms that the data presents some bi-mode feature. By checking some properties about the default rate time series in Section 4.6, we realize that this is caused by the extreme persistence in the data.

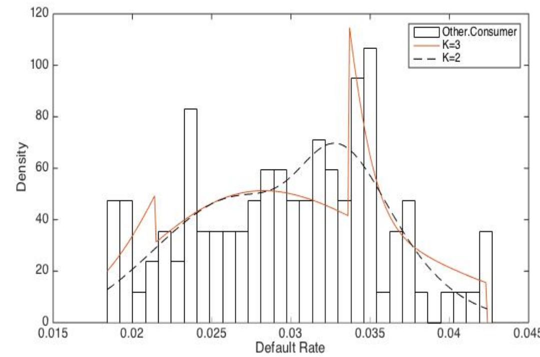
The above figure presents a subset of series in our data set. We selected four most representative histograms to demonstrate our findings. The rest of the figures can be found in Appendix. This setting is also applied to the rest of figures in this thesis.

The density fittings strongly suggest that the state-dependent model captures more information than the traditional Vasicek model does. When the absolute value of the state correlation β approaches 1, the density curve becomes discontinuous at certain points.

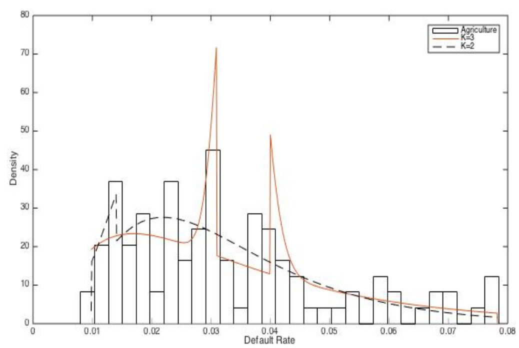
We also present the density curves of models with $K = 2$ and $K = 3$ together to have a more direct comparison.



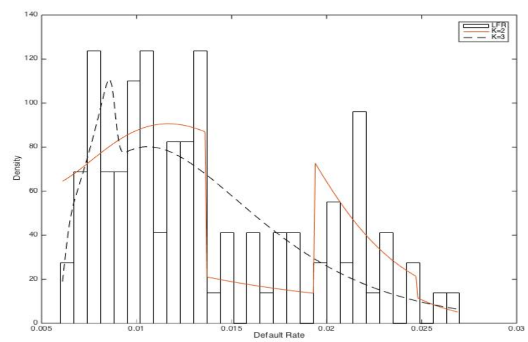
(a) Business



(b) Other.Consumer



(c) Agriculture



(d) LFR

Figure 2.5: The histograms of different series with state-dependent density curve when $K = 2$ and $K = 3$.

As we can see, sometimes the 3-regimes model fits the data more closely than the 2-regimes model does. But it is hard to conclude that the 3-regimes model is uniformly superior to the 2-regimes one.

2.3 Continuous factor loading

In this section, we relax the assumption that $a(\cdot)$ is a simple discrete function in the form of

$$a(t) = \sum_{k=1}^{K} a_k \cdot \mathbb{1}(t_{k-1} < t \leq t_k).$$

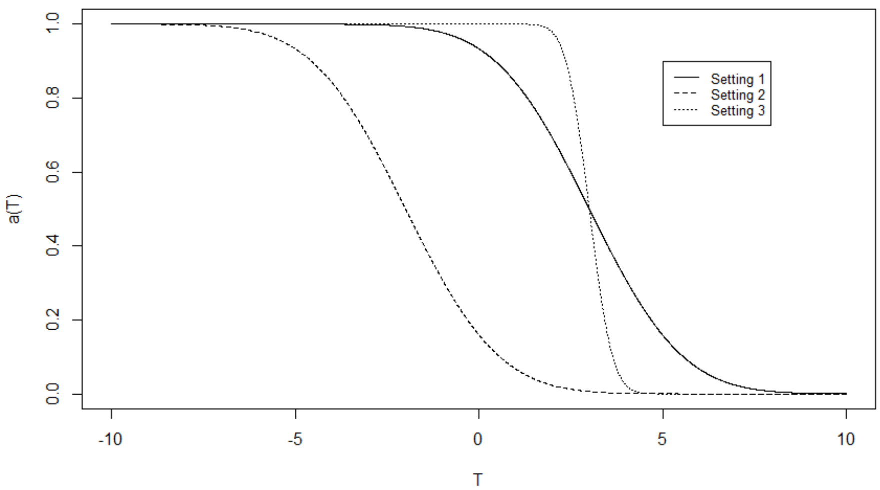
We want to find a suitable continuous function that satisfies the following condition:

1. Monotone, decreasing, and continuous function of t.
2. Bounded between 0 and 1.
3. The sensitivity with respect to t can be controlled in a straightforward way by a small number of parameters.

As a result, the following function is a good starting point for this section:

$$a_{\alpha_1,\alpha_2}(T) = \Phi\left(\frac{\alpha_1 - T}{\alpha_2}\right), \tag{2.25}$$

where α_1 and α_2 are two unknown parameters, which need to be estimated, and $\Phi()$ is the CDF of the standard normal. It is clear that the function is bounded between 0 and 1. Because we use $-T$ in the Function 2.25, $a_{\alpha_1,\alpha_2}(T)$ is a decreasing function with respect to T if $\alpha_2 > 0$. The parameters α_1 and α_2 control the sensitivity of the factor loading with respect to T.



(a) Setting 1 is $\alpha_1 = 3$ and $\alpha_2 = 2$. Setting 2 is $\alpha_1 = -2$ and $\alpha_2 = 2$. Setting 3 is $\alpha_1 = 3$ and $\alpha_2 = 0.5$.

Figure 2.6: Effect of α_1 and α_2 on the realized value of $a(t)$ defined in Eqaution 2.25

Figure 2.6 demonstrates the impact of α_1 and α_2 on the value of $a(T)$. It is readily seen that α_1 shifts the value horizontally and α_2 controls the flatness of the function. The line is flatter when α_2 increases.

2.3.1 Conditional regime probabilities

In Section 2.1.2, we defined the function $p_k(m) = \mathbb{P}(R = k|M = m)$. So we need to change it according to the modification of $a(t)$. Under the continuous factor loading setting, we do not have the regime indicator variable R anymore since there are infinitely many regimes. Then $p_k(m)$ should be replaced by

$$p_t(m) = f_{T|M}(t|m) = \phi(t, \beta m, 1 - \beta^2),$$

which is the conditional PDF of T given M.

2.3.2 Conditional correlations

In this section we find the relation between β and market correlation ρ_t under the continuous factor loading setting.

$$\begin{aligned} Var(X_i|T=t) &= a(t)^2 \cdot Var(M|T=t) + (1-a(t)^2) \cdot Var(Z|T=t) \\ &= a(t)^2(1-\beta^2) + (1-a(t)^2) \\ &= 1 - a(t)^2\beta^2 \end{aligned}$$

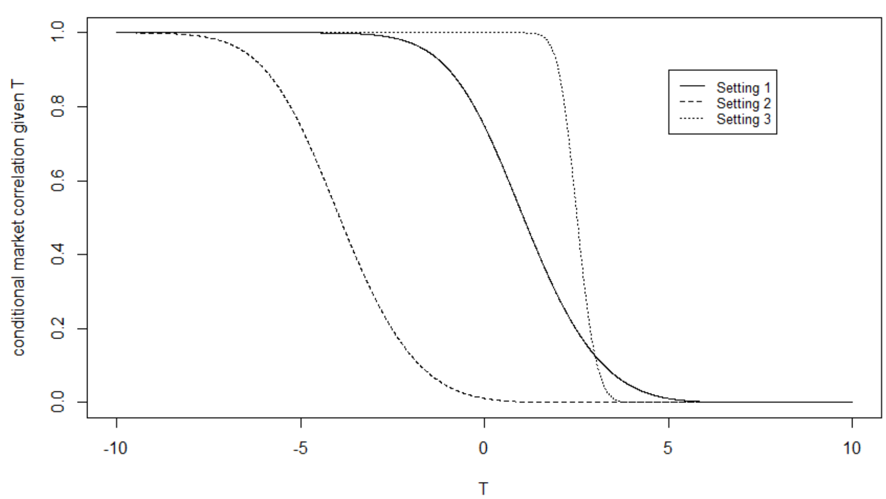
and

$$Cov(X_i, X_j|T=t) = a(t)^2(1-\beta^2).$$

So

$$\rho_t = \frac{a(t)^2(1-\beta^2)}{1-a(t)^2\beta^2}.$$

The correlation preserves the property that in the case of $\beta = 0$, squared factor loadings and asset correlations are identical. But when the factor loading is constant, the previous property does not hold, since once we condition on T, the distribution of M will also be changed.



(a) All the parameter settings are same as Figure 2.6

Figure 2.7: Conditional market correlation given T

Figure 2.7 demonstrates the behavior of the conditional market correlation for different values of T.

2.3.3 Marginal distribution of credit quality

Due to the fact that we change the factor loading function, the threshold for the credit quality also changes. As before, we need to look at the credit quality distribution function under the continuous factor loading function. This is given by:

$$
\begin{aligned}
\mathbb{P}(X \leqslant x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{P}(X \leqslant x|M=m, T=t)\phi_2(m,t)\ dm\ dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{P}\left(a(t)m + \sqrt{1-a(t)^2}Y \leqslant x|M=m, T=t\right)\phi_2(m,t)\ dm\ dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{P}\left(Y \leqslant \frac{x-a(t)m}{\sqrt{1-a(t)^2}}\middle| M=m, T=t\right)\phi_2(m,t)\ dm\ dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi\left(\frac{x-a(t)m}{\sqrt{1-a(t)^2}}\right)\phi_2(m,t)\ dm\ dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi\left(\frac{x-a(t)m}{\sqrt{1-a(t)^2}}\right)\phi_M(m|T=t)\phi_T(t)\ dm\ dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi\left(\frac{x-a(t)m}{\sqrt{1-a(t)^2}}\right)\phi_M(m|T=t)\ dm\ \phi_T(t)\ dt \\
&= \int_{-\infty}^{\infty} \mathbb{E}_M\left[\Phi\left(\frac{x-a(t)M}{\sqrt{1-a(t)^2}}\right)\middle| T=t\right]\phi_T(t)\ dt \\
&= \int_{-\infty}^{\infty} \mathbb{E}_M\left[\mathbb{P}\left(Z \leqslant \frac{x-a(t)M}{\sqrt{1-a(t)^2}}\right)\middle| T=t\right]\phi_T(t)\ dt \\
&= \int_{-\infty}^{\infty} \mathbb{E}_M\left[\mathbb{P}\left(\sqrt{1-a(t)^2}Z + a(t)M \leqslant x\right)\middle| T=t\right]\phi_T(t)\ dt,
\end{aligned}
$$

where $Z \sim N(0,1)$ and $M|T=t \sim N(\beta t, 1-\beta^2)$. In addition to that, Z and M are independent of each other. We can define $\tilde{Z} = \sqrt{1-a(t)^2}Z + a(t)M$. So $\tilde{Z}|T=t \sim N[a(t)\beta t,\ 1-a(t)^2\beta^2]$. Then we can further simplify the distribution function of the credit quality,

$$\begin{aligned}\mathbb{P}(X \leqslant x) &= \int_{-\infty}^{\infty} \mathbb{P}(\tilde{Z} \leqslant x | T = t)\phi_T(t)\ dt \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{x - a(t)\beta t}{\sqrt{1 - a(t)^2\beta^2}}\right)\phi_T(t)\ dt.\end{aligned}$$

To evaluate the right-hand side of the above formula, we need to use a numerical method to find out the corresponding default threshold x_{PD} given a default probability PD.

2.3.4 Probability distribution of a large portfolio default rate

Last but not least, we need to find out the density function of a large portfolio default rate. Because under the current setting, we can not use R to indicate which regime we are in, we will deal with the joint distribution function of M and T directly. First we have

$$P(D > d) = \int_{-\infty}^{\infty} P(T = t, v_t(M) > d)dt = \int_{-\infty}^{\infty} P(T = t, M < v_t^{-1}(d))dt,$$

where $P(T = t, M < v_t^{-1}(d))$ is not really a probability that $T = t$ and $M < v_t^{-1}(d)$ because T is a continuous random variable. Instead, it is actually a half-integrated joint PDF of M and T, $P(T = t, M < v_t^{-1}(d)) = \int_{-\infty}^{v_t^{-1}(d)} \phi_2(t, m; \beta)dm$ and $v_t(M) = \Phi\left(\frac{x_{pd} - a(t)M}{\sqrt{1 - a(t)^2}}\right)$.
After differentiating $1 - \int_{-\infty}^{\infty} P(T = t, M < v_t^{-1}(d))dt$ with respect to d, we obtain the PDF of the large default rate as follows:

$$f_D(d) = \int_{-\infty}^{\infty} p_t(v_t^{-1}(d)) \cdot f_t(d)dt, \tag{2.26}$$

where $f_t(d) = f_{x_{pd}, a(t)}(d)$.
The intuition is quite apparent. In order to have $D = d$, the value of M must equal $v_t^{-1}(d)$ if we are given the condition that $T = t$. For a fixed T the density of M at $v_t^{-1}(d)$ is $f_t(d)$, and the density that we have $T = t$ given this value of M is $p_t(v_k^{-1}(d))$.

Now, in order to apply the EM algorithm, we still need the joint PDF of D and T. It is just the part inside the integration of Equation 2.26.

$$f_{D,T}(d, t) = p_t(v_t^{-1}(d)) \cdot f_t(d).$$

The expected value in Step 1 of the EM algorithm can be calculated by the following equation:

$$V(\theta|\theta^{(t)}) = E_{T|D,\theta^{(i)}}[\log L(\theta; D, T)], \tag{2.27}$$

where $L(\theta; D, T)$ is the likelihood function of D and T:

$$\log L(\theta; D, T) = \sum_{i=1}^{n} \log f_{D,T}(d_i, t_i) = \sum_{i=1}^{n} \left[\log(p_{t_i}(v_{t_i}^{-1}(d_i)) \cdot f_{t_i}(d_i))\right]. \tag{2.28}$$

Now the Equation 2.27 can be written in the following form:

$$V(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \left(\int_{-\infty}^{\infty} (\log(p_{t_i}(v_{t_i}^{-1}(d_i)) \cdot f_{t_i}(d_i)) \cdot P(f_{T|D}(t|D = d_i; \theta^{(t)})dt \right). \tag{2.29}$$

As we can tell from equation 2.29, it may be hard to write $V(\theta|\theta^{(t)})$ in a closed form. So we need to apply some numerical method to calculate $V(\theta|\theta^{(t)})$.

2.3.5 Parameter estimation

We apply the EM algorithm based on the continuous factor loading state-dependent model to the same historical data we have used in Section 2.2. But the result is not desirable. The following table is the estimator for the historical data.

Data set	α_1	α_2	β
1	-2.57	3.17	0.76
2	-7.55	11.11	0.78
3	-1885.08	1518.91	0.41
4	-2661.26	2434.33	0.41
5	-1818.48	1362.43	0.41
6	-21.75	29.32	0.42
7	-12.08	11.83	0.88
8	-2.01	4.76	0.98
9	-35.76	45.73	0.47
10	-1.02	2.15	0.88
11	-0.84	4.75	0.97

Table 2.6: Estimates for continuous factor loading state dependent model

As we can see from the Table 2.6, the third, fourth and fifth data set have unreasonably large values for α_1 and α_2. With such large values and the fact that T has a standard normal distribution, the continuous factor loading function $a(T)$ is almost constant at a certain level. Due to the large values of both α_1 and α_2, the model actually reduces to the classical Vasicek model with a constant factor loading, and this renders β meaningless.

2.4 Regime filtration

When we estimate the parameters for the discrete 2-regimes model, we can draw inference about the latent variables R and M. In this section, we applied the following methodology to infer the values of the systematic risk factor and the regime indicator R. The inference we made here is based on the parameter estimates we got in Section 2.2.3. We keep the assumption that there are only 2 regimes in the market, low and high correlation regimes. We also assume that each time period is independent (an assumption that will be relaxed in subsequent sections). Then, we have that the following equation:

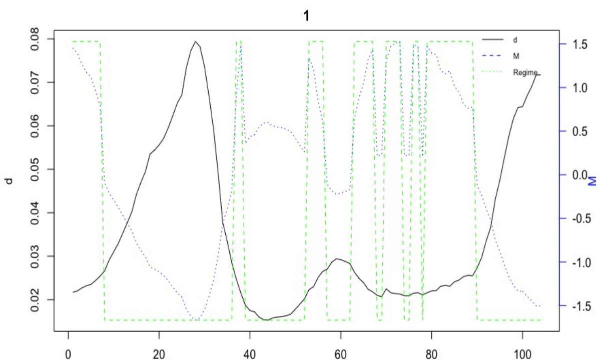
$$P(R_t = k|D_{1,\dots,t}) = P(R_t = k|D_t).$$

We calculate the regime with the maximum probability given the observation data at time i.

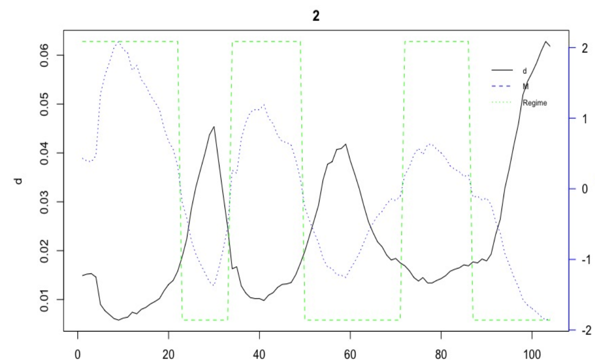
$$\hat{K}_i = \arg\max_k P(R_i = k|D_i = d_i). \tag{2.30}$$

Then we assume that the market is in $\hat{K}_i$ regime at time i. As a result, the default rate, D_i, can be used to calculate the implied systematic risk factor via the following equation:

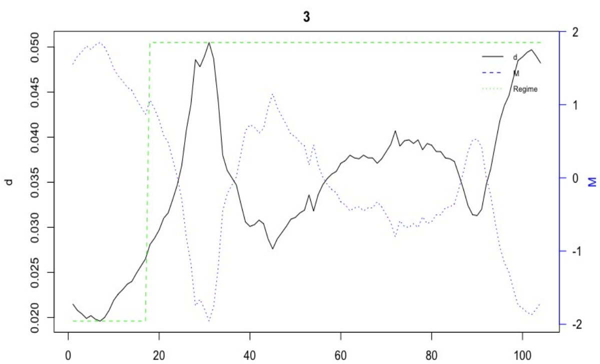
$$M_i = \frac{\frac{x_{PD}}{\sqrt{1-a^2_{\hat{K}_i}}} - \Phi^{-1}(D_i)}{\frac{a_{\hat{K}_i}}{\sqrt{1-a^2_{\hat{K}_i}}}} = \frac{x_{PD} - \sqrt{1-a^2_{\hat{K}_i}}\Phi^{-1}(D_i)}{a_{\hat{K}_i}}.$$



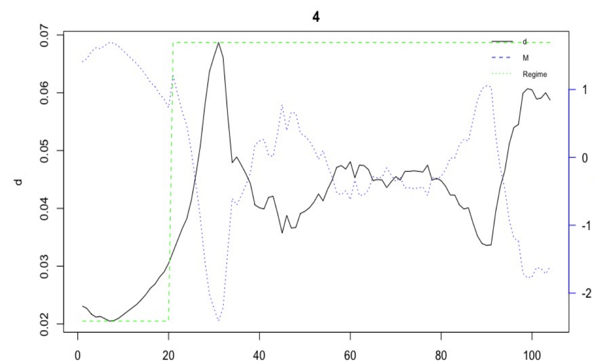
(a) All



(b) Business



(c) Consumer



(d) Credit card

Figure 2.8: The 2-regimes state-dependent model filtration plots. The green dashed line refers to the regimes we are in. The high position of the green dashed line stands for the high correlation regimes. The blue dotted line is the implied systematic risk factor M. The black solid line is the observed default rate.

The interesting fact is that although the regimes change from time to time, the changes of M and D are pretty similar to each other. We believe that we can take a look at the dynamics of the observation data D to find a suitable dependence structure for it, and the same structure should also be a good model for the systematic risk factor M.

Based on Figure 2.8, it is easy to see that both the time series of observed default rates and implied systematic risk factor strongly suggest the existence of time dynamic. We use the estimated parameters for the Business series to generate a set of simulated data under the independence assumption. After plotting the simulated data in Figure 2.9, we notice that the dynamic of the simulated default rate is quite different from the historical data. This fact motivates the approach that we take later in Chapter 3, where we explicitly model a temporal dependence for the systematic factors in the SDM.

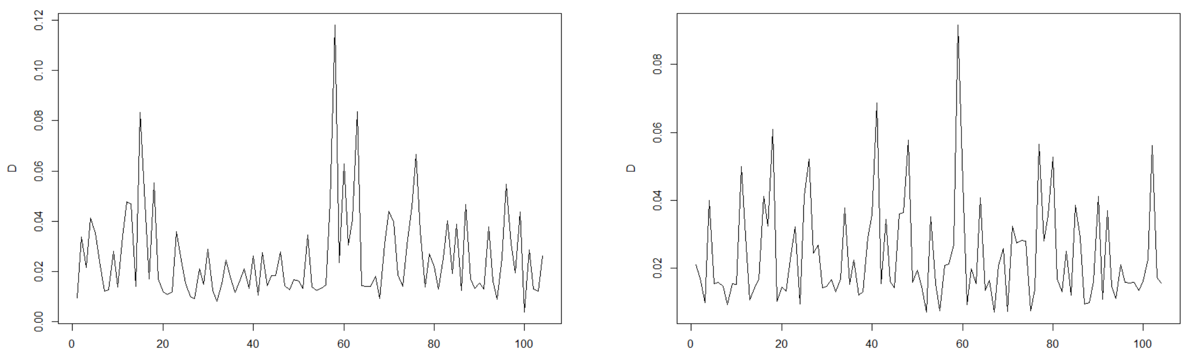


Figure 2.9: The simulation data based on the model we fitted for the second data set

Chapter 3

Dynamic State-Dependent Model

In this chapter, we propose a new model to introduce time dynamics to the static model from Chapter 2 based on the discrete factor loading. Based on the results we have generated in Section 2.4, we suspect that the independence among data points is strongly violated. In order to capture the dependence structure between different time periods, it is natural for us to propose a new model in Section 3.1 to capture this property. The model that we propose in this chapter is based on the SDM described in Chapter 2. But instead of assuming the data points are independent of each other, we wish to add a certain dependence structure to the model by adopting AR(1) process on the systematic risk factor, M. By doing so, the autocorrelation of the observation data can be explained by the dependence structure of the systematic risk factor. We will present some properties of the model in this Chapter. Then we will move to the model parameter estimation problem.

Due to the existence of latent variables and dependence structure, it is not a simple task to get a closed-form formula for the joint likelihood function for the observations. But by applying the classic filtering and smoothing procedures, we are able to get the one-step-ahead conditional predictive density function of the observations given the past path. As a result, the likelihood function of the observations can be written as a product of the one-step-ahead conditional density functions together. After that, the maximum likelihood method is applied to calibrate the model. Simulation tests and Monte Carlo tests are also applied to verify the accuracy of the estimation procedure.

From the empirical analysis that we conduct, the data implies that AR(1) performs poorly in forecasting the one-year-ahead market movement. The plots of both autocorrelation function and partial autocorrelation function also confirm this observation. In order to improve the forecasting ability, we change the model by switching the process of

M from AR(1) to AR(2). The same estimation method and tests are applied to the new model. The empirical results suggest that there is a marked improvement in the forecasting exercise.

3.1 Dynamic state-dependent model

In this section, we first present our new model and then discuss the temporal dependence structure that the model is designed to capture. The model is formulated in the following way:

$$X_{i,t} = a(T_t)M_t + \sqrt{1 - a(T_t)^2} e_{i,t} \tag{3.1}$$

$$M_t = \theta M_{t-1} + \epsilon_t \tag{3.2}$$

$$T_t = \beta M_t + \epsilon'_t, \tag{3.3}$$

and $a(\cdot)$ is the function defined at the beginning of Chapter 2:

$$a(x) = \sum_{k=1}^{K} a_k \cdot \mathbb{1}(t_k < x \leqslant t_{k+1}), \tag{3.4}$$

where K is the number of the market regimes. For each t, the error terms $e_{i,t}$, ϵ_t and ϵ'_t are independent with mean 0 and variances $\sigma_e^2 = 1$,$\sigma_\epsilon^2 = 1 - \theta^2$ and $\sigma_{\epsilon'}^2 = 1 - \beta^2$, respectively.

As we can notice from Equation 3.2, the formula we used to define the AR(1) structure of $\{M_t\}$ is different from the traditional way. We remove the constant term and force the idiosyncratic term's variance to be a function of θ instead of estimating it independently. There are several reasons for us to define it in this way. First, the systematic risk factor $\{M_t\}$ is used to represent the relative level of the market scenario. We use the mean value of its stationary distribution to denote the middle market level. Adding a constant term in Equation 3.2 only shifts the distribution horizontally without changing the shape of it. As a result, we eliminate the constant term in our model. Second, by forcing the variance of the idiosyncratic term to be a function of θ, the unit variance assumption about the systematic risk factor can be preserved for any value of θ.

We call this model the State-Dependent Model with AR(1) process, shortened as SDM-AR(1). In the rest of this thesis, $M_{1:N}$ and $T_{1:N}$ represent the time series of market level (systematic risk factor) and regime index. They are two latent variables that we cannot observe directly from the market. The only observable data are the quarterly default rate D_t for each time t. The relationship between the quarterly default rate D_t and the i^{th}

obligor credit score $X_{i,t}$ remains the same as in the Vasicek model described in Section 1.2 and is of the form

$$D_t := \sum_{i=1}^{N} \frac{\mathbb{1}_{(X_{i,t} < x_{PD})}}{N}, \tag{3.5}$$

where x_{PD} is the default threshold of the credit score, which satisfies the Equation

$$P(X_{i,t} \leqslant x_{PD}) = PD_i. \tag{3.6}$$

Here PD_i stands for the i^{th} obligor's probability of default, which is a parameter in our model. Since we are interested in the large portfolio under the homogeneous market assumption, $PD_i = PD$ and $N \to \infty$. As a result of the law of large numbers and the fact that for any i and j, $X_{i,t}$ and $X_{j,t}$ are independent with each other once M_t and T_t are given, we know that the distribution of D_t conditionally on M_t and T_t can be approximated as follows

$$\begin{aligned} D_t \approx & \lim_{N\to\infty} \sum_{i=1}^{N} \frac{\mathbb{1}_{(X_{i,t} < x_{PD})}}{N} \\ =& P(X_{i,t} \leqslant x_{PD} | M_t, T_t) \\ =& \Phi\left(\frac{x_{PD} - a(T_t)M_t}{\sqrt{1 - a(T_t)^2}}\right), \end{aligned} \tag{3.7}$$

where $\Phi(\cdot)$ is the CDF of standard normal distribution. When compared to the SDM in Chapter 2, the variables $\{M_t\}$ and $\{T_t\}$ are no longer independent here within each time period. The temporal dependence structure of series $\{M_t\}$ and $\{T_t\}$ are described by Formulas 3.2 and 3.3, which can be replaced by other processes in future research.

In order to simplify our presentation, we introduce the transferred default rate Y_t such that

$$Y_t = \Phi^{-1}(D_t).$$

If in this formula we substitute for D_t its approximation given by 3.7, then we obtain

$$\begin{aligned} Y_t &= \Phi^{-1}\left(\Phi\left(\frac{x_{PD} - a(T_t)M_t}{\sqrt{1 - a^2(T_t)}}\right)\right) \\ &= \frac{x_{PD}}{\sqrt{1 - a(T_t)^2}} - \frac{a(T_t)}{\sqrt{1 - a(T_t)^2}} M_t, \end{aligned} \tag{3.8}$$

where $a(\cdot)$ and X_{PD} are defined in Formulas 3.4 and 3.6, respectively.

In order to get the maximum likelihood estimators for the parameters, we need to determine the joint density function of $\{Y_1, \dots, Y_N\}$. From Equation 3.8, we should notice that once we condition on M_t, the conditional distribution of the random variable $Y_t|M_t$ will become discrete, with the number of possible values equal to the number of regimes we assume. We have the following representation of the conditional probability function of Y_t:

$$P(Y_t = y_t|M_{1:N} = m_{1:N}) = \sum_{k=1}^{K} P(T_t \in (t_k, t_{k+1})|M_t = m_t)\mathbb{1}\left(y_t = \frac{x_{PD} - a_k m_t}{\sqrt{1 - a_k^2}}\right).$$

As a result, the joint probability density function of the time series $M_{1:N}$ and $Y_{1:N}$ is

$$f_{Y\&M}(y_{1:N}, m_{1:N}) = \prod_{t=1}^{N}\left(\sum_{k=1}^{K} P(T_t \in (t_k, t_{k+1})|M_t = m_t)\mathbb{1}\left(y_t = \frac{x_{PD} - a_k m_t}{\sqrt{1 - a_k^2}}\right)\right) f_M(M_{1:N}), \tag{3.9}$$

where $f_M(\cdot)$ is the joint density function of the variables $M_{1:N}$ that follow the autoregressive process 3.2,

$$f_M(m_{1:N}) = \phi(m_1)\Pi_{t=2}^{N}\pi(m_t|m_{t-1})$$

with $\pi(m_t|m_{t-1})$ being the transition density of $\{M_t\}_{t=2,\dots}$. According to the Equation 3.2, we know

$$\begin{aligned} P(M_t \leqslant m_t|M_{t-1} = m_{t-1}) &= P(\theta M_{t-1} + \epsilon_t \leqslant m_t|M_{t-1} = m_{t-1}) \\ &= P(\epsilon_t \leqslant m_t - \theta m_{t-1}). \end{aligned}$$

As we mentioned at the beginning of Section 3.1, ϵ_t follows a normal distribution with mean 0 and variance $1 - \theta^2$. Then

$$\begin{aligned} P(M_t \leqslant m_t|M_{t-1} = m_{t-1}) &= \Phi(m_t - \theta m_{t-1}; 0, 1 - \theta^2) \\ &= \Phi(m_t; \theta m_{t-1}, 1 - \theta^2), \end{aligned}$$

where $\Phi(x; \mu; \sigma^2)$ is the CDF of the normal distribution with mean μ and variance σ^2. By taking the derivative with respect to m_t, we can get $\pi(m_t|m_{t-1})$ as

$$\pi(m_t|m_{t-1}) = \phi(m_t; \theta m_{t-1}, 1 - \theta^2).$$

Again, $\phi(x; \mu; \sigma^2)$ is the PDF of the normal distribution with mean μ and variance σ^2. But, due to the presence of an indicator function inside Equation 3.9, this joint likelihood function is actually ill-defined. The interval of the indicator function only contains one single point for each time $t = 1, \dots$ and $k = 1, 2$. Based on the fundamental property of integration, if we take integration of $f_{Y\&M}(y_{1:N}, m_{1:N})$ over all feasible regions of $Y_{1:N}$ and $M_{1:N}$, the answer will always be zero. That contradicts the definition of a density function. The proper method to find out the joint likelihood function of the observations will be discussed in Section 3.3.1.

3.2 Filtering, smoothing and forecasting

In this section, we describe the filtering and smoothing procedure for the model we have proposed. The one-step-ahead predictive density function of the observations plays an important role when we calculate the likelihood function.

The main challenge when using this approach is that we can only observe the overall default rate D_t for each time period, but the market risk factor M_t and the regime variable T_t are latent, meaning that we can not observe them directly. However, as we demonstrate below, we are able to use the filtering procedure to make inference about the market risk M_t and market regime index T_t based on the observations up to time t.

Following the exposition provided in Chapter 2.7 of the monograph by Petris, Petrone and Campagnoli (2009), for a general state space model, we assume that there is a latent process $M_{1:N}$, called state process, and the observations $Y_{1:N}$ can be viewed as an imprecise measurement of $M_{1:N}$. There are two critical assumptions for a state space model:

1. The state process $\{M_t\}_{t=1,\dots}$ is a Markov chain.
2. For any $t = 1, \dots, N$, conditionally on M_t, the Y_t are independent with its past and future values, and Y_t are fully determined by M_t and T_t only.

We can also treat $M_{1:N}$ as an auxiliary time series, which will determine the probability distribution of $Y_{1:N}$. So the purpose of the filtering procedure is to determine the distribution of M_t given the observations $Y_{1:t}$. We denote this filtering density function at time t by $f_{M_t|Y_{1:t}}(m_t|y_{1:t})$. In order to get the filtering density function $f_{M_t|Y_{1:t}}(m_t|y_{1:t})$ for $t = 1, \dots, N$, one can recursively repeat the following three steps from $t = 2$ to N.

(i) In the first step, we calculate the one-step-ahead predictive density function for the state variables, $f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})$. This function describes the probability distribution of M_t when we only have observations up to the previous time $Y_{1:t-1}$. It can be computed from the filtered density at time $t-1$, $f_{M_{t-1}|Y_{1:t-1}}(m_{t-1}|y_{1:t-1})$ and the conditional distribution function of M_t given M_{t-1}, $f_{M_t|M_{t-1}}(m_t|m_{t-1})$, according to

$$f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1}) = \int f_{M_t|M_{t-1}}(m_t|m_{t-1}) \cdot f_{M_{t-1}|Y_{1:t-1}}(m_{t-1}|y_{1:t-1}) dm_{t-1}. \quad (3.10)$$

(ii) Once we have the one-step-ahead predictive density function for the state variables, $f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})$ at time $t-1$, we are able to derive the one-step-ahead predictive density function for the observation Y_t, $f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})$.

$$f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) = \int f_{Y_t|M_t}(y_t|m_t) \cdot f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1}) dm_t.$$

This density function enables us to make a prediction about the distribution of the observation at time t when we only have observations up to time $t-1$.

(iii) Now, the filtering density $f_{M_t|Y_{1:t}}(m_t|y_{1:t})$ can be computed based on the Bayes rule with $f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})$ as the prior distribution and the likelihood $f_{Y_t|M_t}(y_t|m_t)$,

$$f_{M_t|Y_{1:t}}(m_t|y_{1:t}) = \frac{f_{Y_t|M_t}(y_t|m_t) \cdot f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})}. \quad (3.11)$$

In the following subsections, we will present more details about applying this general procedure to our model. In particular, we will first use Step 1 to determine the probability distribution for our market risk factor M_t given the observations from time 1 to t, $Y_{1:t}$. Then, we will derive the one-step-ahead predictive density for the observation at time t, Y_t, given $Y_{1:t-1}$. Finally, we will be able to update the distribution of the latent market risk factor M_t based on all available observations. Once the one-step-ahead predictive density function of the observation $f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})$ is derived for times $t = 1, \ldots, N$, we can obtain the joint likelihood function of the observations $f_{Y_{1:N}}(y_{1:N})$ according to

$$f_{Y_{1:N}}(y_{1:N}) = f_{Y_1}(y_1) \cdot \prod_{t=2}^{N} f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}).$$

3.2.1 One-step-ahead predictive density for the states

Compared to the general state space model, our model requires a different approach. This is due to the fact that once the value of the variable Y_t is known, the latent variable M_t

is no longer continuous but discrete, and vice versa. So the filtering density function 3.11 is not a probability density function but a probability mass function. This means that Equation 3.10 is actually a product of probability density functions and probability mass functions. As a result, Equation 3.10 takes the following form if we assume that there are only 2 potential regimes for the market, $K = 2$,

$$\begin{aligned} f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1}) &= \sum_{j=1}^{2} f_{M_t,M_{t-1}|Y_{1:t-1}}(m_t, M^{-1}(y_{t-1}, j)|y_{1:t-1}) \\ &= \sum_{j=1}^{2} f_{M_t|M_{t-1}}(m_t|M^{-1}(y_{t-1}, j)) \cdot P(M_{t-1} = M^{-1}(y_{t-1}, j)|Y_{1:t-1} = y_{1:t-1}), \end{aligned} \tag{3.12}$$

where

$$M^{-1}(y, i) = \frac{x_{PD} - \sqrt{1 - a_i^2}y}{a_i}. \tag{3.13}$$

In Equation 3.13, x_{PD} is the default threshold of the credit score introduced in Equation 3.6. Function 3.13 returns the possible value of M given the observation value of $Y = y$ and regime i^{th}. The first term in Equation 3.12,

$$f_{M_t|M_{t-1}}(m_t|M^{-1}(y_{t-1}, j))$$

is the transition density function of the AR(1) process $\{M_t\}$. According to the definition of $\{M_t\}$ in Equation 3.2, we know that the driving noise of the process $\{M_t\}$ follows the normal distribution with mean 0 and variance $1 - \theta^2$. Then we have

$$\begin{aligned} P(M_t \leqslant m_t|M_{t-1} = m_{t-1}) =& P(\theta M_{t-1} + \epsilon_t \leqslant m_t|M_{t-1} = m_{t-1}) \\ =& P(\epsilon_t \leqslant m_t - \theta m_{t-1}) \\ =& \Phi\left(\frac{m_t - \theta m_{t-1}}{\sqrt{1 - \theta^2}}\right). \end{aligned}$$

By taking the first derivative with respect to m_t, we have the transition density function in the following form

$$\begin{aligned} f_{M_t|M_{t-1}}(m_t|m_{t-1}) =& \frac{d}{dm_t} P(M_t \leqslant m_t|M_{t-1} = m_{t-1}) \\ =& \frac{d}{dm_t}\Phi\left(\frac{m_t - \theta m_{t-1}}{\sqrt{1 - \theta^2}}\right) \\ =& \frac{1}{\sqrt{1 - \theta^2}} \cdot \phi\left(\frac{m_t - \theta m_{t-1}}{\sqrt{1 - \theta^2}}\right). \end{aligned} \tag{3.14}$$

The second term in Equation 3.12,

$$P(M_{t-1} = M^{-1}(y_{t-1}, j) | Y_{1:t-1} = y_{1:t-1}),$$

represents the probability of $M_{t-1} = M^{-1}(y_{t-1}, j)$ given $Y_{1:t-1}$, the form of which will be derived in Section 3.2.3. Once the value of Y_t is fixed, the sample space of M_t becomes discrete with the same number of elements as the number of regimes.

3.2.2 One-step-ahead predictive density for the observations

In this section, we derive the one-step-ahead predictive density function of Y_t given $Y_{1:t-1}$. The result is stated in the theorem below.

Theorem 3. *The one-step-ahead predictive density function of Y_t given $Y_{1:t-1}$ is of the following form*

$$\begin{aligned} f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) = \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} P(T_t \in [t_i, t_{i+1}) | M_t = M^{-1}(y_t, i)) \cdot \\ f_{M_t|Y_{1:t-1}}(M^{-1}(y_t, i) | y_{1:t-1}). \end{aligned} \tag{3.15}$$

Proof. The first term in Equation 3.15, $P(T_t \in [t_i, t_{i+1}) | M_t = M^{-1}(y_t, i))$, is easy to calculate since M_t and T_t follow a bivariate standard normal distribution with correlation β. The second term, $f_{M_t|Y_{1:t-1}}(M^{-1}(y_t, i)|y_{1:t-1})$, is described in Section 3.2.1. We first start from the cumulative density function

$$P(Y_t \leqslant y_t | Y_{1:t-1} = y_{1:t-1}) = \sum_{i=1}^{2} P(T_t \in [t_i, t_{i+1}), M_t \geqslant M^{-1}(y_t, i) | Y_{1:t-1} = y_{1:t-1}).$$

Then we take the derivative of the above CDF with respect to y_t. We have

$$\begin{aligned} f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) &= \frac{d}{dy_t} P(Y_t \leqslant y_t | Y_{1:t-1} = y_{1:t-1}) \\ &= \frac{d}{dy_t} \sum_{i=1}^{2} P(T_t \in [t_i, t_{i+1}), M_t \geqslant M^{-1}(y_t, i) | Y_{1:t-1} = y_{1:t-1}) \\ &= \sum_{i=1}^{2} \frac{d}{dy_t} \int_{t_i}^{t_{i+1}} \int_{M^{-1}(y_t,i)}^{\infty} f_{T_t, M_t|Y_{1:t-1}}(a, b | y_{1:t-1}) \, db \, da \\ &= \sum_{i=1}^{2} \int_{t_i}^{t_{i+1}} \frac{d}{dy_t} \int_{M^{-1}(y_t,i)}^{\infty} f_{T_t, M_t|Y_{1:t-1}}(a, b | y_{1:t-1}) \, db \, da. \end{aligned}$$

According to the Leibniz integral rule,

$$\begin{aligned}
f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) &= \sum_{i=1}^{2} \int_{t_i}^{t_{i+1}} [-f_{T_t,M_t|Y_{1:t-1}}(a, M^{-1}(y_t,i)|y_{1:t-1})] \cdot \frac{d}{dy_t} M^{-1}(y_t,i)\ da \\
&= \sum_{i=1}^{2} \int_{t_i}^{t_{i+1}} [-f_{T_t,M_t|Y_{1:t-1}}(a, M^{-1}(y_t,i)|y_{1:t-1})] \cdot \left[-\frac{\sqrt{1-a_i^2}}{a_i}\right]\ da \\
&= \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} \int_{t_i}^{t_{i+1}} f_{T_t,M_t|Y_{1:t-1}}(a, M^{-1}(y_t,i)|y_{1:t-1})\ da.
\end{aligned}$$

Since we know that T_t is conditionally independent of the past variables once M_t is given, we have

$$\begin{aligned}
f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) &= \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} \int_{t_i}^{t_{i+1}} f_{T_t|M_t}(a|M^{-1}(y_t,i)) f_{M_t|Y_{1:t-1}}(M^{-1}(y_t,i)|y_{1:t-1})\ da \\
&= \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} \int_{t_i}^{t_{i+1}} f_{T_t|M_t}(a|M^{-1}(y_t,i))\ da \cdot f_{M_t|Y_{1:t-1}}(M^{-1}(y_t,i)|y_{1:t-1}) \\
&= \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} P(T_t \in [t_i, t_{i+1})|M_t = M^{-1}(y_t,i)) \cdot f_{M_t|Y_{1:t-1}}(M^{-1}(y_t,i)|y_{1:t-1}).
\end{aligned}$$

□

3.2.3 Filtering probability function

According to Equation 3.11, once we have the one-step-ahead predictive density for both the states and observations, we can easily calculate the filtering probability function $f_{M_t|Y_{1:t}}(m_t|y_{1:t})$. As we mentioned in the beginning of Section 3.2.1, once the observation Y_t is given, the variable M_t becomes discrete. Therefore, our filtering function is a probability mass function instead of a probability density function, which is in the following form

$$\begin{aligned}
&P(M_t = m_t|Y_{1:t} = y_{1:t}) = \frac{P(Y_t = y_t|M_t = m_t) \cdot f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})} \\
&= \sum_{i=1}^{2} \frac{\frac{\sqrt{1-a_i^2}}{a_i} P(T_t \in [t_i, t_{i+1})|M_t = M^{-1}(y_t,i)) \cdot f_{M_t|Y_{1:t-1}}(M^{-1}(y_t,i)|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})} \cdot \mathbb{1}(m_t = M^{-1}(y_t,i)).
\end{aligned} \tag{3.16}$$

The reason for the presence of an indicator function at the end of Equation 3.16 is that for any given value of Y_t, there exist only two potential values of M_t.

3.2.4 Filtering procedure for SDM-AR(1) model

In this section, we will summarize all the results from previous sections to formulate a complete filtering algorithm for the SDM-AR(1) model with 2 regimes:

1. First, we create a $2 \times N$ filtering matrix Ψ

   $$\Psi = \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} & \dots & \Psi_{1,N} \\ \Psi_{2,1} & \Psi_{2,2} & \dots & \Psi_{2,N} \end{bmatrix},$$

   such that

   $$\Psi_{i,j} = P(M_j = M^{-1}(y_j, i) | Y_{1:j} = y_{1:j}).$$

   We will keep updating the elements from left to right iteratively by the following steps.

2. In this step, we calculate the probability mass function of M_1 given Y_1. These are the elements in the first column of the matrix Ψ, $\Psi_{1,1}$ and $\Psi_{2,1}$. It is helpful to know that

   $$\begin{aligned} P(M_1 = m_1 | Y_1 = y_1) =& P(T_1 \in (-\infty\ t_1] | D_1 = \Phi(y_1)) \mathbb{1}_{\{m_1 = M^{-1}(y_1, 1)\}} \\ &+ P(T_1 \in (t_1\ \infty] | D_1 = \Phi(y_1)) \mathbb{1}_{\{m_1 = M^{-1}(y_1, 2)\}}, \end{aligned}$$

   where T_1 and D_1 are the market regime index and observed actual default rate at time $t = 1$, respectively. As we can see from the above equation, the probability of $M_1 = M^{-1}(y_1, i)$ given $Y_1 = y_1$ equals the probability of the market being in the i^{th} regime given the observation of the default rate at time t, $D_t = \Phi(y_1)$. Also, we only have one available observation, so the dependence structure brought by Equation 3.2 has no impact when we calculate such probability. As a result of those two facts, we can calculate $P(M_1 = m_1 | Y_1 = y_1)$ by Equation 2.19

   $$P(M_1 = m_1 | Y_1 = y_1) = \sum_{k=1}^{K} \frac{f_{D,R}(d, k)}{f_D(d)} \mathbb{1}_{\{m_1 = M^{-1}(y_1, k)\}},$$

   where $f_{D,R}(\cdot)$ and $f_D(\cdot)$ are defined in Equations 2.18 and 2.17, respectively.

Now, we repeatedly complete Steps 3 to 5 from time $t = 2$ to $t = N$.

3. In this step, we calculate the one-step-ahead predictive density function of the market risk factor M_t given $Y_{1:t-1}$ by the following equation

$$f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1}) = \sum_{j=1}^{2} f_{M_t|M_{t-1}}(m_t|M^{-1}(y_{t-1}, j)) \cdot P(M_{t-1} = M^{-1}(y_{t-1}, j)|Y_{1:t-1} = y_{1:t-1})$$

where $f_{M_t|M_{t-1}}(\cdot)$ is defined in Equation 3.14, and the last term in above equation $P(M_{t-1} = M^{-1}(y_{t-1}, j)|Y_{1:t-1} = y_{1:t-1})$ is already stored in the matrix Ψ as $\Psi_{j,t-1}$.

4. In this step, we calculate the one-step-ahead predictive density function of the observation Y_t given $Y_{1:t-1}$

$$f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) = \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} P(T_t \in [t_i, t_{i+1})|M_t = M^{-1}(y_t, i)) \cdot f_{M_t|Y_{1:t-1}}(M^{-1}(y_t, i)|y_{1:t-1}),$$

where $P(T_t \in [t_i, t_{i+1})|M_t = M^{-1}(y_t, i))$ is easy to calculate since M_t and T_t follow a joint standard normal distribution with correlation β. The second term, $f_{M_t|Y_{1:t-1}}(\cdot)$, is already derived from Step 3.

5. We update the tth column filtering matrix according to Equation 3.16

$$\Psi_{i,t} = P(M_t = M^{-1}(y_t, i)|Y_{1:t} = y_{1:t}) \tag{3.17}$$
$$= \frac{P(Y_t = y_t|M_t = M^{-1}(y_t, i)) \cdot f_{M_t|Y_{1:t-1}}(M^{-1}(y_t, i)|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})}. \tag{3.18}$$

As we can see from Equation 3.8, once given the value of the market risk factor M_t, the value of Y_t is totally dependent on the market regime index T_t. Then the first term in the numerator of Equation 3.18 can be computed as

$$P(Y_t = y_t|M_t = m_t) = P(T_t \in (-\infty\ t_1]|M_t = m_t)\mathbb{1}_{\{m_t = M^{-1}(y_t, 1)\}} + P(T_1 \in (t_1\ \infty]|M_t = m_t)\mathbb{1}_{\{m_t = M^{-1}(y_t, 2)\}}.$$

3.2.5 Smoothing

After we are able to filter the market risk factor $\{M_t\}$, we are interested in smoothing the values $M_1, \ldots, M_N$. It is worth explaining the difference between filtering and smoothing.

For the filtering problem, the observations are assumed to be available sequentially in time. In contrast, the problem of smoothing assumes that we already have observations on $\{Y_t\}$ for a certain period, and we wish to retrospectively study the behavior of the underlying market risk factor $\{M_t\}$. In general, we can use the following backward-recursive algorithm presented by Petris, Petrone and Campagnoli (2009) to compute the smoothing probability function $P(M_t = m_t|Y_{1:N} = y_{1:N})$ for $t = 1, \ldots, N$.

(i) Conditional on $Y_{1:N}$, the state process $\{M_1, M_2, \ldots, M_N\}$ has backward transition probabilities given by

$$P(M_t = m_t|M_{t-1} = m_{t-1}, Y_{1:N} = y_{1:N}) = \sum_{i=1}^{2} \frac{f_{M_{t+1}|M_t}(m_{t+1}|m_t)P(M_t = m_t|Y_{1:t} = y_{1:t})}{f_{M_{t+1}|Y_{1:t}}(m_{t+1}|y_{1:t})} \cdot \mathbb{1}(m_{t+1} = M^{-1}(y_{t+1}, i)).$$

(ii) The smoothing distribution of M_t given $Y_{1:N}$ can be computed according to the following backward recursion, $f_{M_t|Y_{1:N}}(m_t|y_{1:N})$, starting from $t = T, \ldots, 1$:

$$\begin{aligned}
P(M_t = m_t|Y_{1:N} = y_{1:N}) &= \sum_{i=1}^{2} P(M_t = m_t, M_{t+1} = M^{-1}(y_{t+1}, i)|Y_{1:N} = y_{1:N}) \\
&= \sum_{i=1}^{2} P(M_t = m_t|M_{t+1} = M^{-1}(y_{t+1}, i), Y_{1:N} = y_{1:N}) \cdot \\
&\quad P(M_{t+1} = M^{-1}(y_{t+1}, i)|Y_{1:N} = y_{1:N}) \\
&= \sum_{i=1}^{2} \frac{f_{M_{t+1}|M_t}(M^{-1}(y_{t+1}, i)|m_t)P(M_t = m_t|Y_{1:t} = y_{1:t})}{f_{M_{t+1}|Y_{1:t}}(M^{-1}(y_{t+1}, i)|y_{1:t})} \cdot \\
&\quad P(M_{t+1} = M^{-1}(y_{t+1}, i)|Y_{1:N} = y_{1:N}) \\
&= P(M_t = m_t|Y_{1:t} = y_{1:t}) \sum_{i=1}^{2} \frac{f_{M_{t+1}|M_t}(M^{-1}(y_{t+1}, i)|m_t)}{f_{M_{t+1}|Y_{1:t}}(M^{-1}(y_{t+1}, i)|y_{1:t})} \cdot \\
&\quad P(M_{t+1} = M^{-1}(y_{t+1}, i)|Y_{1:N} = y_{1:N}).
\end{aligned}$$

3.2.6 K-step ahead prediction

In Section 3.2.2, we present a formula for the one-step-ahead predictive function for the observations $\{Y_t\}$. However, in practice, it is also essential and necessary to make a prediction several time periods ahead. For this, we need to determine the K-step-ahead predictive density of Y_{t+K} given $Y_{1:t}$, which we denote by $f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t})$.

Theorem 4. *The K-step-ahead predictive density of Y_{t+K} given $Y_{1:t}$, $f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t})$, can be calculated in the following way*

$$\begin{aligned} f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t}) = &\sum_{i=1}^{2}\sum_{j=1}^{2} P(Y_{t+k} = y_{t+k}|M_{t+k} = M^{-1}(y_{t+k}, j)) \cdot \\ & f_{M_{t+k}|M_t}(M^{-1}(y_{t+k}, j)|M^{-1}(y_t, i)) \cdot \\ & P(M_t = M^{-1}(y_t, i)|Y_{1:t} = y_{1:t}), \end{aligned}$$

where $f_{M_{t+k}|M_t}(M^{-1}(y_{t+k}, j)|M^{-1}(y_t, i))$ is the K-step transition density function of the AR(1) model.

Proof. First, we know that according to the law of total probability we have

$$f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t}) = \sum_{i=1}^{2} f_{Y_{t+k},M_t|Y_{1:t}}(y_{t+k}, M^{-1}(y_t, i)|y_{1:t}).$$

Since for any integer $k \geq 0$, Y_{t+k} is conditionally independent of $Y_{1:t}$ once given M_t is given, we can rewrite the above equation in the following way

$$\begin{aligned} f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t}) = &\sum_{i=1}^{2} f_{Y_{t+k}|M_t}(y_{t+k}|M^{-1}(y_t, i)) \cdot P(M_t = M^{-1}(y_t, i)|Y_{1:t} = y_{1:t}) \\ = &\sum_{i=1}^{2}\sum_{j=1}^{2} f_{Y_{t+k},M_{t+k}|M_t}(y_{t+k}, M^{-1}(y_{t+k}, j)|M^{-1}(y_t, i)) \cdot \\ & P(M_t = M^{-1}(y_t, i)|Y_{1:t} = y_{1:t}) \\ = &\sum_{i=1}^{2}\sum_{j=1}^{2} P(Y_{t+k} = y_{t+k}|M_{t+k} = M^{-1}(y_{t+k}, j)) \cdot \\ & f_{M_{t+k}|M_t}(M^{-1}(y_{t+k}, j)|M^{-1}(y_t, i)) \cdot \\ & P(M_t = M^{-1}(y_t, i)|Y_{1:t} = y_{1:t}). \end{aligned}$$

□

We will use the formula presented in Theorem 4 in the empirical study to demonstrate the forecasting ability of our model.

3.3 Model estimation

Although our model is easy to understand, it is not a simple task to estimate the model parameters because of the existence of the latent processes $M_{1:N}$ and $T_{1:N}$. In this section, we will focus on the method of estimation of the model parameters. The main challenge is to determine a proper way of calculating the joint likelihood function of the observations. In the following sections, we present a method of computing the log-likelihood function of the observations $\{Y_t\}$. After that, we will talk about the MLE for the model parameters and perform simulation tests in order to verify the accuracy of such an estimation method. Results of our simulation test suggest that the estimation procedure works reasonably well.

3.3.1 Likelihood for the observations $\{Y_{1:N}\}$

In this section, we present a method to calculate the likelihood function for the observations $\{Y_{1:N}\}$. For any fixed values of parameters, we can easily execute the algorithm described in Section 3.2.4, which means that we are already able to compute the one-step-ahead predictive density function $f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})$ for each $t = 2, \dots, N$. After that, according to the definition of the conditional density function, we can derive the likelihood function of $\{Y_{1:N}\}$ by the following equation:

$$L(\delta; y_{1:N}) = \prod_{i=2}^{N} f_{Y_i|Y_{1:i-1}}(y_i|y_{1:i-1};\delta) f_{Y_1}(y_1;\delta), \tag{3.19}$$

where δ denotes the all the model parameters.

Then the log-likelihood function can be calculated as

$$\begin{aligned}\ell(\delta; y_{1:N}) &= \log L(\delta; y_{1:N}) \\ &= \sum_{i=2}^{N} \log f_{Y_i|Y_{1:i-1}}(y_i|y_{1:i-1};\delta) + \log f_{Y_1}(y_1;\delta),\end{aligned}$$

where $f_{Y_i|Y_{1:i-1}}(y_i|y_{1:i-1})$ is defined in Equation 3.17, and $f_{Y_1}(y_1)$ is the stationary distribution function of the observation series $\{Y_t\}$, which we derive now. According to the way we define the model in Section 3.1, the stationary distribution of the market risk series $\{M_t\}$ follows the standard normal distribution. As a result, the stationary distribution of the $\{Y_t\}$ and the distribution of D_t, which was defined in Equation 2.17, have the following relation:

$$\begin{aligned} P(Y_t \leqslant y_t) =& P(\Phi^{-1}(D_t) \leqslant y_t) \\ =& P(D_t \leqslant \Phi(y_t)). \end{aligned}$$

By taking the derivative with respect to y_t on both side, we will get

$$f_{Y_t}(y_t) = f_D(\Phi(y_t)) \cdot \phi(y_t),$$

where $f_D(\cdot)$ is defined in Equation 2.17, and $\phi(\cdot)$, $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution.

3.3.2 Maximum likelihood estimator for SDM-AR(1) model

Since we have determined a method to evaluate the log-likelihood for the observation series $\{Y_t\}$, we are now able to apply some numeric optimizer to get the MLE of the parameters, $a_1, a_2, \beta, \theta, t_1$ by maximizing the likelihood function with respect to those parameters

$$[\tilde{a_1}, \tilde{a_2}, \tilde{t_1}, \tilde{\beta}, \tilde{\theta}] = \arg \max_{a_1, a_2, t_1, \beta, \theta} \log(f_{Y_{1:N}}(y_{1:N}; a_1, a_2, t_1, \beta, \theta)),$$

where $f_{Y_{1:N}}(\cdot)$ is defined in Equation 3.19.

We have implemented our estimation procedure in Matlab with the numerical optimizer function `fmincon`. Since we wish to find the maximal value of the likelihood function but the `fmincon` provides the minimal value for the objective function, we applied the `fmincon` to $-\log(f_{Y_{1:N}}(y_{1:N}; a_1, a_2, t_1, \beta, \theta))$ with the following constraints:

$$\begin{aligned} 0 < a_2 < a_1 < 1 \\ -1 < \beta < 1 \\ -1 \leqslant \theta \leqslant 1. \end{aligned}$$

The `fmincon` function also requires an initial starting point as input. For different starting points, the `fmincon` function may eventually terminate at, and return, different local minima. In order to find the global maximum, we defined a set of values for each parameter and used all the possible combinations of those values to create a mesh over the space

$$\begin{aligned} a &= [0.5, 0.4, 0.3, 0.2, 0.1] \\ t &= [-1, -0.5, 0, 0.5, 1] \\ \beta &= [0.1, 0.3, 0.5, 0.7, 0.9] \\ \theta &= [-1, -0.5, 0, 0.5, 1]. \end{aligned}$$

For a_1 and a_2, we have chosen all the combinations from a such that $a_1 > a_2$. That means there are 10 different sets of values for $[a_1, a_2]$ and 5 sets for t, β and θ. This gives us 1250 different starting points in total. Then, we run the `fmincon` function with respect to those starting points and choose the parameters with the lowest objective function value as our estimates of the model parameters.

Test of our estimation method based on simulated data

We have conducted a small simulation study to assess some basic properties of our estimation method. First, we present the algorithm used to simulate data. For any given set of parameters, we have the following algorithm to generate the data:

1. At time $t = 1$, we generate M_1 and T_1 from the standard normal distribution with correlation β.
2. Calculate Y_1 from Equation 3.8.
3. For any time $t = 2, \ldots, N$, we simulate M_t based on the value of M_{t-1}, and T_t, based on the value of M_t, according to Equations 3.2 and 3.3 respectively.
4. Calculate Y_t according to Equation 3.8.

Now, we consider the accuracy of those estimators. We randomly chose seven sets of parameter values to cover some situations that the parameters may look like. For each set, we repeatedly simulate 2000 data points 200 times and separately estimate the parameters for each of our 200 data sets. The following tables provide some basic statistics about the estimators for each set.

	a_1	a_2	β	t_1	θ
True value	0.340	0.250	0.300	0.600	0.980
Average of the 200 runs	0.373	0.285	0.274	0.379	0.976
Std. Err of the 200 runs	0.096	0.072	0.401	0.832	0.014
P-value	0.73	0.63	0.94	0.79	0.77
True value	0.340	0.250	0.750	0.600	0.980
Average of the 200 runs	0.358	0.277	0.563	0.218	0.960
Std. Err of the 200 runs	0.078	0.067	0.561	1.150	0.121
P-value	0.81	0.68	0.73	0.73	0.86
True value	0.340	0.110	0.750	0.600	0.980
Average of the 200 runs	0.342	0.131	0.641	0.601	0.958
Std.dev of the 200 runs	0.136	0.080	0.385	0.852	0.036
P-value	0.98	0.79	0.77	0.987	0.54
True value	0.190	0.070	0.980	0.600	0.750
Average of the 200 runs	0.187	0.069	0.977	0.593	0.744
Std.dev of the 200 runs	0.007	0.006	0.008	0.084	0.017
P-value	0.66	0.86	0.71	0.93	0.72
True value	0.340	0.250	0.750	0	0.750
Average of the 200 runs	0.334	0.242	0.722	0.069	0.734
Std.dev of the 200 runs	0.017	0.045	0.317	0.631	0.093
P-value	0.72	0.85	0.92	0.91	0.86
True value	0.190	0.070	0.100	0.600	0.750
Average of the 200 runs	0.189	0.072	0.084	0.593	0.745
Std.dev of the 200 runs	0.007	0.014	0.170	0.301	0.026
P-value	0.88	0.88	0.92	0.98	0.84
True value	0.190	0.110	0.750	0	0.300
Average of the 200 runs	0.189	0.109	0.767	-0.037	0.297
Std.dev of the 200 runs	0.004	0.006	0.070	0.165	0.020
P-value	0.80	0.87	0.81	0.82	0.88

Table 3.1: SDM-AR(1) model estimators accuracy test with 200 sets of 2000 simulated data.

In Table 3.1, the row "true value" in each sub-table shows the values of the parameters we have used to simulate data. The average and standard error rows are calculated based on all the successfully finished tests. The last row is the P-value associated with the test of the null hypothesis that the estimators are unbiased. As we can see from the tables, there is no statistical evidence to reject the null hypothesis that the estimators are unbiased. The average values of the estimators are within one standard error of the true value. We also noticed that the accuracy of $\tilde{a}_1$ and $\tilde{a}_2$ is strongly negatively related to the value of θ. The estimates are more accurate when the value of θ is lower. The performance of $\tilde{\beta}$ is hard to conclude from those tables. But the estimator still provides a reasonable estimates of the true value. $\tilde{\theta}$ is the most accurate (as measured by standard error) among all. $\tilde{t}_1$ is the least accurate estimator. Since the value $\tilde{t}_1$ only works as a trigger for the regime probability, which means its value is not too sensitive compared to other parameters, this is still an acceptable estimator for our model.

It is also worth mentioning that for each test presented in Table 3.1, there is a possibility that the estimates do not correspond to the global maximum of the likelihood but rather to a local maximum. As a result, the accuracy of the estimators may suffer some negative effects from such problems.

In addition to the test described above, we also conduct a simulation test to check the overall accuracy of the estimators for more general parameter values. The parameters are set to be all possible combinations of the following values to ensure that the proposed method can be used in different situations

$$\begin{aligned}
a &= [0.34\ 0.25] \text{ or } [0.34\ 0.11] \text{ or } [0.19\ 0.11] \text{ or } [0.19\ 0.07] \\
t &= 0 \text{ or } 0.6 \\
\beta &= 0.1 \text{ or } 0.3 \text{ or } 0.75 \text{ or } 0.98 \\
\theta &= 0.1 \text{ or } 0.3 \text{ or } 0.75 \text{ or } 0.98 \\
PD &= 0.05 \text{ or } 0.1.
\end{aligned}$$

For each parameter set, we simulate one sample path. The number of simulated data points is $N = 2000$ in each test. As a result, there should be in total 256 simulation tests. The following table presents the estimation results.

In our study, 225 tests of the total 256 runs have successfully finished running, and the others were stopped due to some numerical issues of the Matlab optimizer. We used the `fmincon` function in Matlab with default setting. Table 3.2 is calculated based on those 225 tests. The first two rows show the average and standard error of the estimation errors, respectively. The last row is the P-value corresponding to the null hypothesis that the estimation error has a mean of 0. It is worth mentioning that this estimation procedure is now applied to different value settings of the parameters, so the distribution of the estimators should not be identical with each other. But we should expect that the mean of

	$a_1 - \hat{a}_1$	$a_2 - \hat{a}_2$	$\beta - \hat{\beta}$	$t_1 - \hat{t}_1$	$\theta - \hat{\theta}_1$
Average	-0.023	-0.007	-0.006	0.010	0.0006
Std. Err	0.0933	0.0562	0.1782	0.5013	0.0218
P-value	0.81	0.90	0.97	0.98	0.98

Table 3.2: Estimation error analysis based on the simulation data. We simulated one set of 2000 data points based on all 256 possible parameter combinations presented in the last paragraph. All the statistics are calculated based on the estimation error.

the combined estimation error is zero if the estimators were unbiased. In this test, we use the overall standard deviation as an approximation of the real one to calculate the P-value. As we can see, there is no evidence to reject the null hypothesis that they are unbiased for all estimators.

In conclusion, the estimation procedure works reasonably well and can be applied to historical data with extra caution about the local maximum problem.

3.4 Dependence structure of SDM-AR(1)

In this section, we study the autocorrelation structure for the observations $\{Y_t\}$ as induced by the model SDM-AR(1). Although, we know the expression of Y_t in terms of M_t and T_t, as given in Equation 3.8, it is still difficult to directly determine the autocorrelation between Y_t and Y_{t+c} for any $c = 1, \ldots$. For these reasons, we will first look at the conditional distribution of $Y_{1:N}$ given $T_{1:N}$. Then, we can apply the law of total variance to get the covariance matrix for the time series $\{Y_t\}$.

3.4.1 Conditional distribution of $\{Y_{1:N}\}$ given $\{T_{1:N}\}$

In this section, we present a method of finding the conditional distribution function of the observations $Y_{1:N}$ given the latent market regime variables $T_{1:N}$.

According to Equation 3.8, we can write the CDF of $\{Y_{1:N}\}$ conditional on $\{T_{1:N}\}$ in the following form

$$P(Y_1 \leqslant y_1, Y_2 \leqslant y_2, \dots Y_N \leqslant y_N | T_{1:N} = t_{1:N}) = P(\frac{x_{PD}}{\sqrt{1-a(t_1)^2}} - \frac{a(t_1)}{\sqrt{1-a(t_1)^2}} M_1 \leqslant y_1,$$

$$\frac{x_{PD}}{\sqrt{1-a(t_2)^2}} - \frac{a(t_2)}{\sqrt{1-a(t_2)^2}} M_2 \leqslant y_2, \dots, \frac{x_{PD}}{\sqrt{1-a(t_N)^2}} - \frac{a(t_N)}{\sqrt{1-a(t_N)^2}} M_N \leqslant y_N | T_{1:N} = t_{1:N}),$$

where X_{PD} is the default threshold of the credit score defined in Formula 3.6 and $a(t_i)$ is defined in Formula 3.4. Before we are able to calculate this conditional CDF of $\{Y_t\}$ given $\{T_t\}$ explicitly, we need to derive the distribution of $M_{1:N}$ given $T_{1:N}$ first.

Theorem 5. *The distribution of $M_{1:N}$ given $T_{1:N}$ is*

$$M_{1:N} | T_{1:N} = t_{1:N} \sim \mathcal{N}(\Sigma_{MT} \Sigma_{TT}^{-1}(t_{1:N}), \Sigma_{MM} - \Sigma_{MT} \Sigma_{TT}^{-1} \Sigma_{TM}),$$

where

$$\Sigma_{MM} = \begin{bmatrix} 1 & \theta & \theta^2 & \dots & \theta^{N-1} \\ \theta & 1 & \theta & \dots & \theta^{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{N-1} & \theta^{N-2} & \theta^{N-3} & \dots & 1 \end{bmatrix} \tag{3.20}$$

$$\Sigma_{MT} = \Sigma_{TM} = \begin{bmatrix} \beta & \theta\beta & \theta^2\beta & \dots & \theta^{N-1}\beta \\ \theta\beta & \beta & \theta\beta & \dots & \theta^{N-2}\beta \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{N-1}\beta & \theta^{N-2}\beta & \theta^{N-3}\beta & \dots & \beta \end{bmatrix} \tag{3.21}$$

$$\Sigma_{TT} = \begin{bmatrix} 1 & \theta\beta^2 & \theta^2\beta^2 & \dots & \theta^{N-1}\beta^2 \\ \theta\beta^2 & 1 & \theta\beta^2 & \dots & \theta^{N-2}\beta^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{N-1}\beta^2 & \theta^{N-2}\beta^2 & \theta^{N-3}\beta^2 & \dots & 1 \end{bmatrix}. \tag{3.22}$$

Proof. Since under our model, the time series $M_{1:N}$ follows an AR(1) process with normally distributed driving noise and T_i being a linear function of M_i plus a normally distributed driving noise, we can treat the series $M_{1:N}$ and $T_{1:N}$ as a $2N$-dimensional multi-variate normal distribution. In order to determine the dependence structure between the series $M_{1:N}$ and $T_{1:N}$, we will decompose the covariance matrix into three parts, Σ_{MM} is the covariance matrix for $M_{1:N}$, Σ_{TT} is the covariance matrix between $T_{1:N}$, and Σ_{MT} is the covariance matrix for $M_{1:N}$ and $T_{1:N}$. The dependence structure between M_i and M_{i+c} is easy to find due to the fact that M follows $AR(1)$ model, namely it can be directly derived from the autocorrelation function of $M_{1:N}$

$$cov(M_i, M_{i+c}) = \theta^c.$$

This implies the form of the covariance matrix $M_{1:N}$ given in 3.20.

Now we are going to find the covariance matrix of the time series $T_{1:N}$. We can see that the mean of $\{T_i\}$ is

$$E[T_i] = \beta E[M_i] + E[\epsilon'_i] = 0.$$

Since $\{M_i\}$ is second-order stationary with mean 0 and variance 1, the variance of T_i is

$$Var[T_i] = \beta^2 Var[M_i] + Var[\epsilon'_i] = \beta^2 + 1 - \beta^2 = 1.$$

In addition to that, $\{\epsilon'_i\}$ follows the distribution defined on Page 39. The covariance of T_i and T_{i+c} is

$$\begin{aligned} cov[T_i, T_{i+c}] &= E[T_i T_{i+c}] - E[T_i]E[T_{i+c}] \\ &= E[T_i T_{i+c}] \\ &= E[T_i * (\theta T_{i+c-1} - \theta \epsilon'_{i+c-1} + \beta \epsilon_{i+c} + \epsilon'_{i+c})]. \end{aligned}$$

Since T_i is independent of the future error term:

$$\begin{aligned} cov[T_i, T_{i+c}] &= \theta E[T_i T_{i+c-1}] \\ &= \theta^{c-1} E[T_i T_{i+1}] \\ &= \theta^{c-1} E[T_i * (\theta T_i - \theta \epsilon'_i + \beta \epsilon_i + \epsilon'_i)] \\ &= \theta^{c-1} (\theta E[T_i^2] - \theta \sigma^2_{\epsilon'}). \\ &= \theta^c \beta^2, \end{aligned}$$

and we also notice that $cov(T_i, T_i) = 1$. So the covariance matrix of $T_{1:N}$ is as given in 3.22. The covariance between M_{i+c} and T_i is

$$\begin{aligned} cov(M_{i+c}, T_i) &= E[(\theta M_{i+c-1} + \epsilon_{i+c})(\beta M_i + \epsilon'_i)] \\ &= E[\theta \beta M_{i+c-1} M_i] \\ &= \theta \beta * \theta^{c-1} \\ &= \theta^c \beta. \end{aligned}$$

So the covariance matrix for $M_{1:N}$ and $T_{1:N}$ is Matrix 3.21. As a result, we can treat $(M_{1:N}, T_{1:N})$ as jointly normally distributed variables with mean 0 and covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_{MM} & \Sigma_{MT} \\ \Sigma_{TM} & \Sigma_{TT} \end{bmatrix}.$$

Thus, from the well-known properties of a multivariate normal distribution

$$M_{1:N}|T_{1:N} = t_{1:N} \sim \mathcal{N}(\Sigma_{MT}\Sigma_{TT}^{-1}(t_{1:N}), \Sigma_{MM} - \Sigma_{MT}\Sigma_{TT}^{-1}\Sigma_{TM}).$$

□

As we can see from Equation 3.8, once the values of $T_{1:N}$ are given, $Y_{1:N}$ is just a linear transformation of $M_{1:N}$. We introduce the following notation:

$$\alpha(T_{1:N}) = \begin{bmatrix} \frac{X_{PD}}{\sqrt{1-a(T_1)}} \\ \frac{X_{PD}}{\sqrt{1-a(T_2)}} \\ \vdots \\ \frac{X_{PD}}{\sqrt{1-a(T_N)}} \end{bmatrix}, \ \chi(T_{1:N}) = \begin{bmatrix} \frac{a(T_1)}{\sqrt{1-a(T_1)}} & 0 & \dots & 0 \\ 0 & \frac{a(T_2)}{\sqrt{1-a(T_2)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{a(T_N)}{\sqrt{1-a(T_N)}} \end{bmatrix}.$$

Then the distribution of $Y_{1:N}$ given $T_{1:N}$ can be acquired by the following theorem.

Theorem 6. *The distribution of $Y_{1:N}$ given $T_{1:N}$ is*

$$Y_{1:N}|T_{1:N} \sim \mathcal{N}(\alpha(T_{1:N}) - \chi(T_{1:N})\Sigma_{MT}\Sigma_{TT}^{-1} \cdot (T_{1:N}),\ \chi(T_{1:N})(\Sigma_{MM} - \Sigma_{MT}\Sigma_{TT}^{-1}\Sigma_{TM})\chi(T_{1:N})^T).$$

We omit the proof, as it uses only basic properties of a multi-variate normal distribution and is similar to that for Theorem 5.

3.4.2 Autocorrelation of $\{Y_t\}$ in SDM-AR(1) model

In the last section, we have determined the conditional distribution of $Y_{1:N}$ given $T_{1:N}$. Now we can use the law of total variance to compute the unconditional variance matrix of $Y_{1:N}$.

Theorem 7. *The unconditional variance matrix of $Y_{1:N}$ can be calculated by the following equation*

$$Var(Y_{1:N}) = E_{T_{1:N}}[Var(Y_{1:N}|T_{1:N})] + Var_{T_{1:N}}(E[Y_{1:N}|T_{1:N}])$$

where $Var(Y_{1:N}|T_{1:N})$ and $E[Y_{1:N}|T_{1:N}]$ are $N \times N$ and $N \times 1$ matrices defined in Theorem 6 respectively.

Although, the covariance matrix of $Y_{1:N}$ admits an analytical representation, it is still difficult or even impossible to determine its explicit form. But we can use simulated values to generate numerical approximation of the covariance matrix by the following algorithm

1. Generate N^* series of $T_{1:N}$.
2. For each $i = 1, \ldots, N^*$, calculate the $Var(Y_{1:N}|T^i_{1:N})$ and $E[Y_{1:N}|T^i_{1:N}]$.
3. Evaluate the sample average of each elements in $Var(Y_{1:N}|T^i_{1:N})$ and the sample variance matrix of $E[Y_{1:N}|T^i_{1:N}]$ as an approximation for $E_{T_{1:N}}[Var(Y_{1:N}|T_{1:N})]$ and $Var_{T_{1:N}}(E[Y_{1:N}|T_{1:N}])$ respectively.

Let $Var(Y_{1:N})_{i,j}$ represent the element in the i^{th} row of the j^{th} column of the matrix $Var(Y_{1:N})$. We know that $Var(Y_{1:N})_{i,j} = Cov(Y_i, Y_j)$. Then the $Corr(Y_t, Y_{t+c})$ can be calculated by

$$Corr(Y_t, Y_{t+c}) = \frac{Var(Y_{1:N})_{t,t+c}}{\sqrt{Var(Y_{1:N})_{t,t} \cdot Var(Y_{1:N})_{t+c,t+c}}}.$$

Since the series $\{Y_t\}$ is stationary, $Var(Y_{1:N})_{t,t} = Var(Y_{1:N})_{t+c,t+c}$, and

$$Corr(Y_t, Y_{t+c}) = \frac{Var(Y_{1:N})_{t,t+c}}{Var(Y_{1:N})_{t,t}}.$$

The following ACF plots are calculated based on different values of the parameters presented in Table 3.3.

Figure	a_1	a_2	β	t_1	θ
3.1	0.5	0.2	1	0	0.95
3.2	0.35	0.2	1	0	0.95
3.3	0.5	0.2	1	0	0.5
3.4	0.5	0.2	0.5	0	0.95

Table 3.3: Value of parameters for the ACF plots

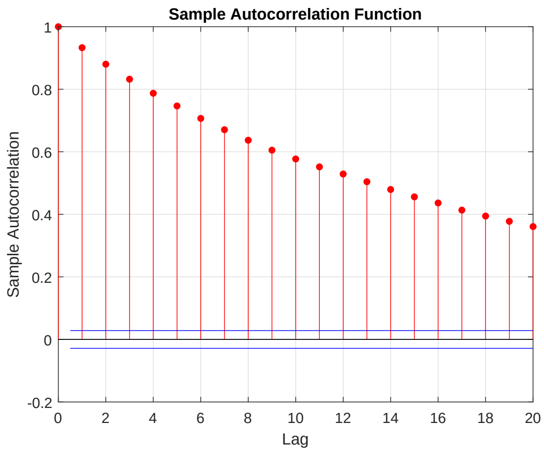


Figure 3.1: ACF plot based on SDM-AR(1) model with 1st set of parameters

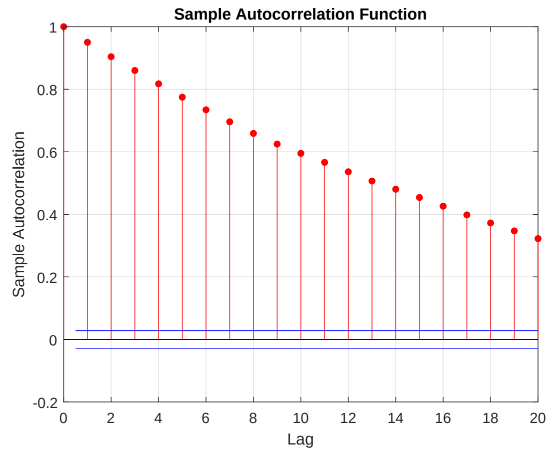


Figure 3.2: ACF plot based on SDM-AR(1) model with 2nd set of parameters

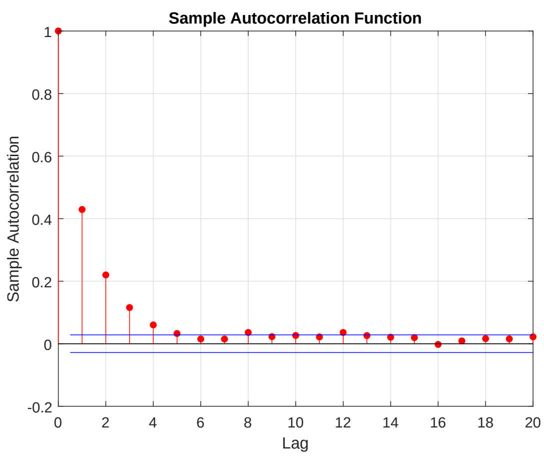


Figure 3.3: ACF plot based on SDM-AR(1) model with 3rd set of parameters

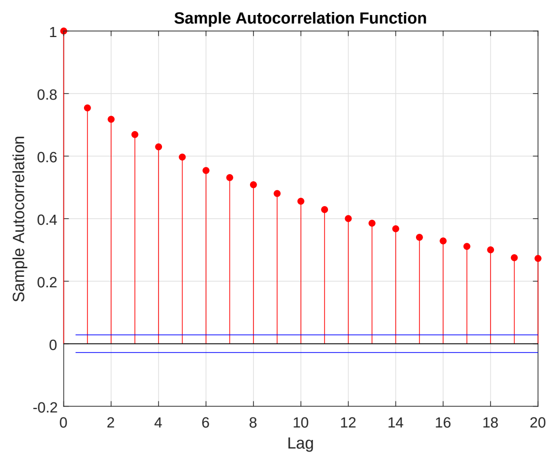


Figure 3.4: ACF plot based on SDM-AR(1) model with 4th set of parameters

As we can see from Figures 3.1 to 3.4, although the different values of the parameters have their own effect on the autocorrelations structure of the series $\{Y_t\}$, the ACF plots of the SDM-AR(1) are still close to the ACF of a classic AR(1) model. Since it is desirable to check if historical data has similar ACF, in next section we will present the ACF for some real historical data and compare it with the ACF of the SDM-AR(1) model.

3.4.3 ACF plots of real historical data

In this section, we will present sample autocorrelation and partial autocorrelation plots of the Federal Reserve Data. We only present six data sets in here since most of them have similar ACF. The following table shows the series whose ACF and PACF plots are presented in Figures 3.5 to 3.10.

a	b	c	d	e	f
All	Business	Consumer	Credit Card	Other Consumer	Agricultural

Table 3.4: Table of series

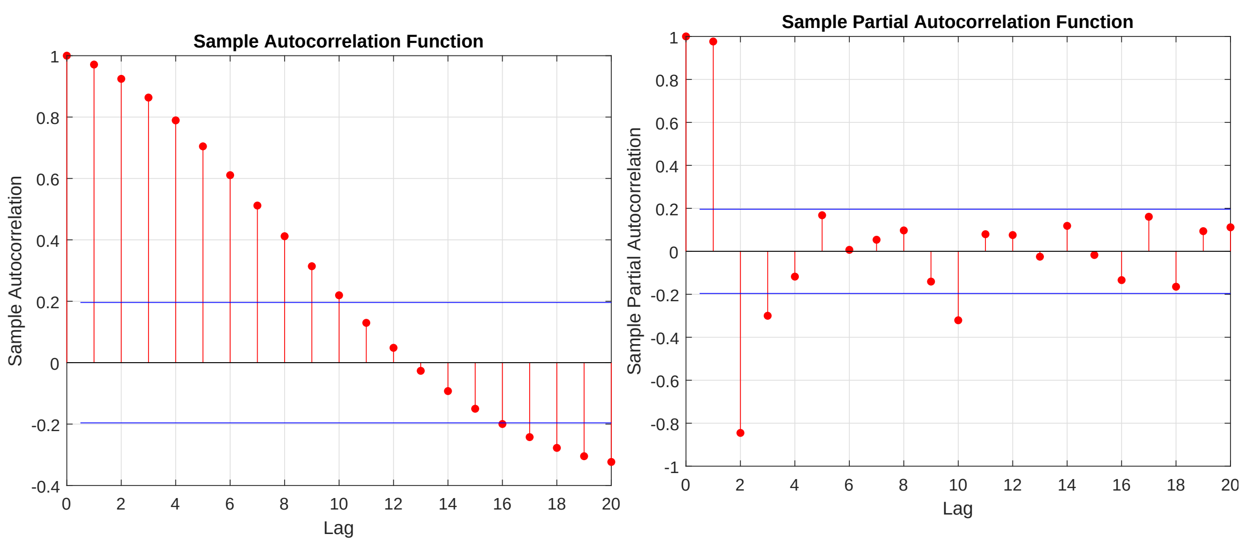


Figure 3.5: Autocorrelation function plots of the "All" historical data

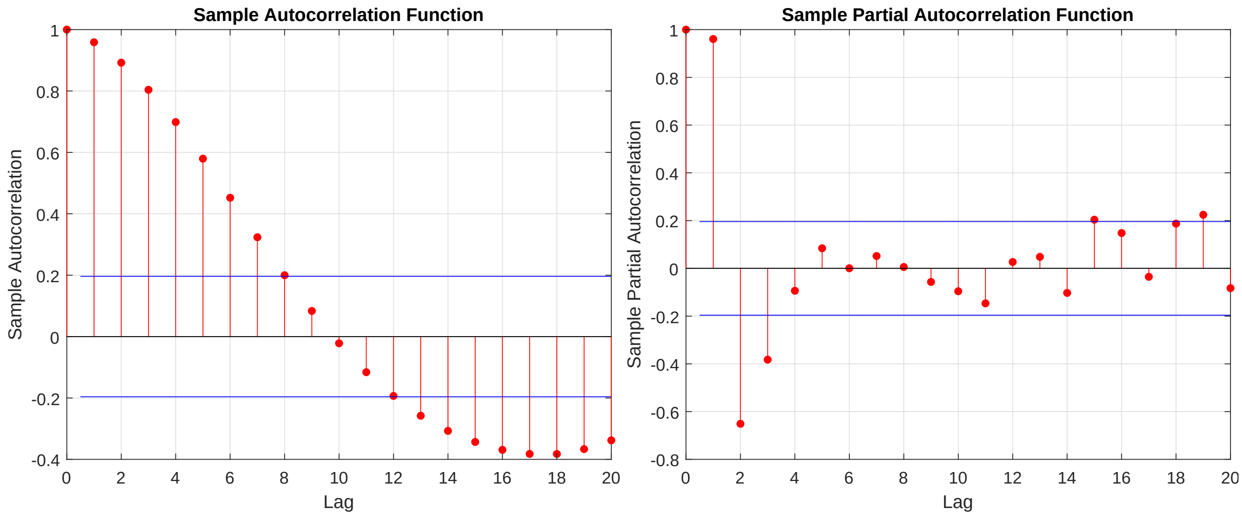


Figure 3.6: Autocorrelation function plots of the “Business” historical data

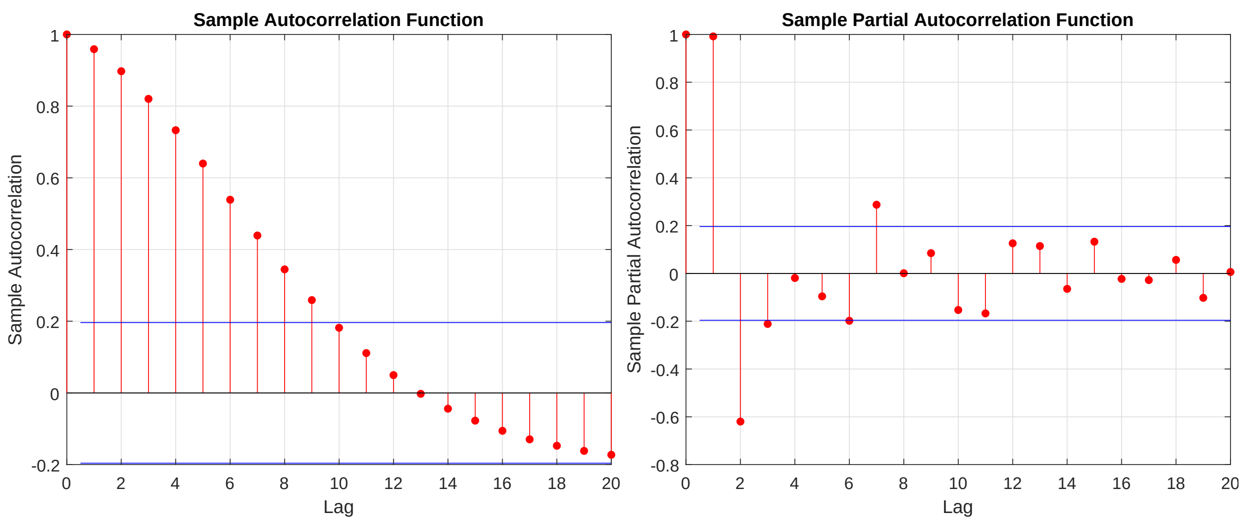


Figure 3.7: Autocorrelation function plots of the “Consumer” historical data

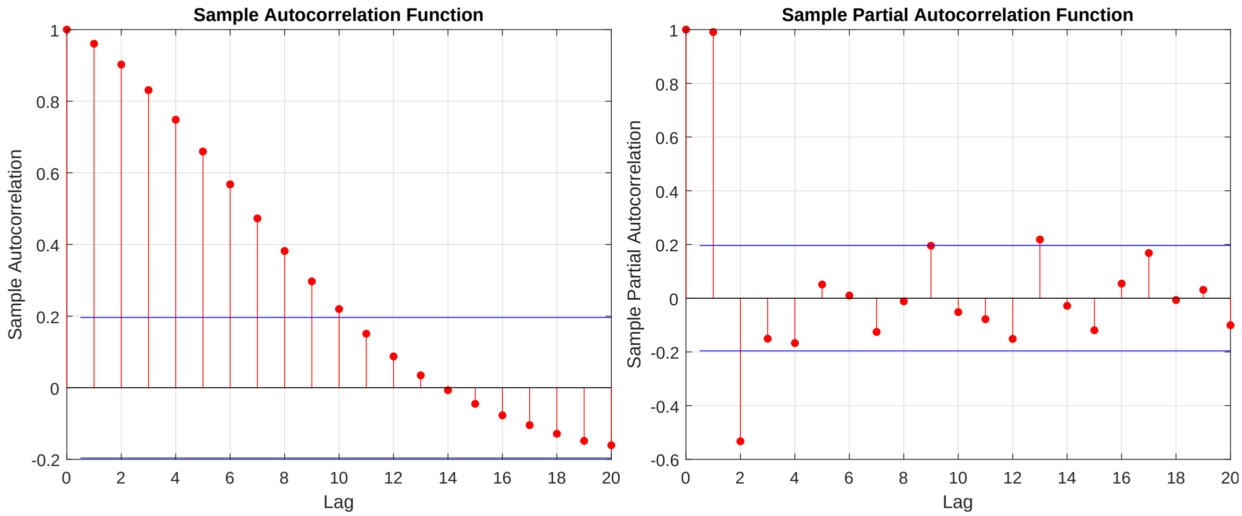


Figure 3.8: Autocorrelation function plots of the “Credit Card” historical data

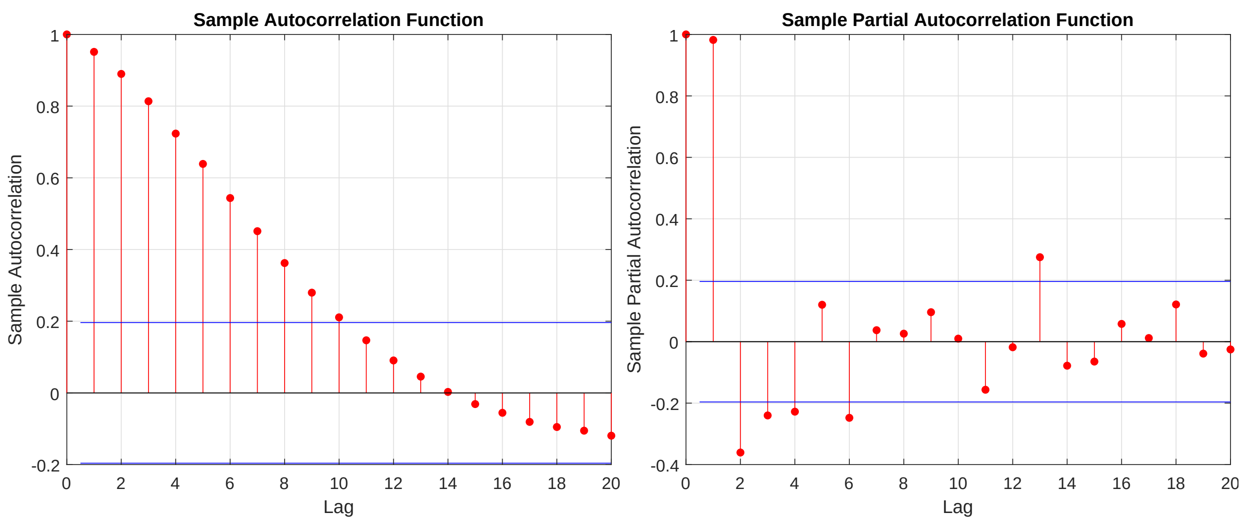


Figure 3.9: Autocorrelation function plots of the “Other Consumer” historical data

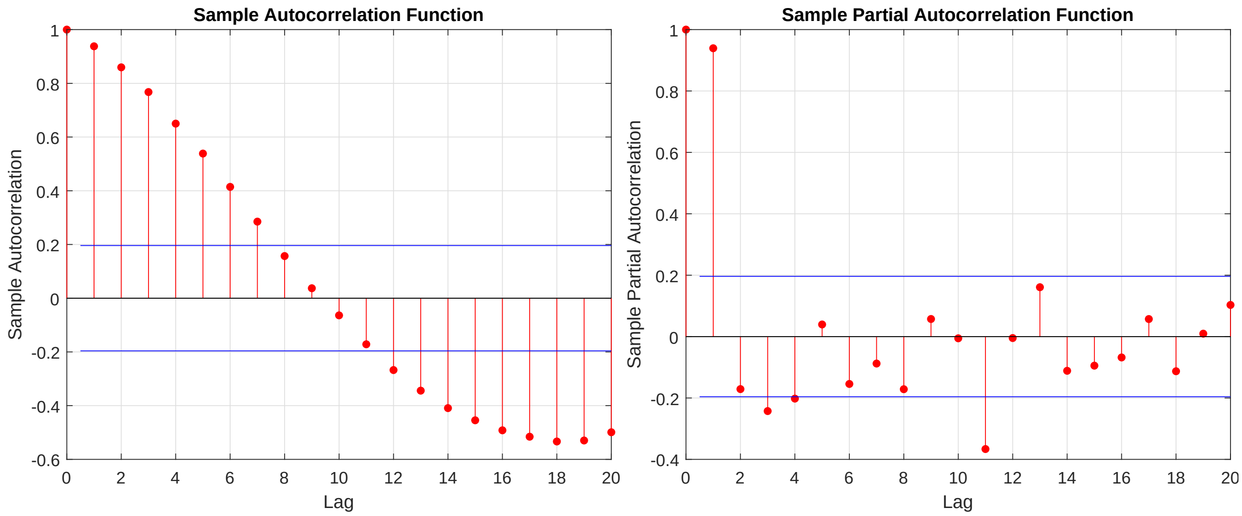


Figure 3.10: Autocorrelation function plots of the "Agricultural" historical data

As we can see from those ACF plots, the real historical data strongly disagree with the AR(1) process. By comparing Figures 3.5 to 3.10 with Figures 3.1 to 3.4, we can easily notice that the state-dependent model with AR(1) process can not accurately capture the autocorrelations for series $\{Y_t\}$. As we know, if the coefficient of an AR(1) process is positive, then the ACF should converge exponentially to 0 from the positive side. The ACF should appear on both sides of zero alternatively if the coefficient is negative. But it is impossible for AR(1) process to have an ACF of the same form as the one produced by historical data. Also the sample ACF plots of the real historical data suggest that AR(2) process should provide a better dependence structure for describing the behavior of real market. So, in the next section, we will introduce a state-dependent model with an AR(2) process.

3.5 Dynamic state-dependent model with AR(2)

As we have mentioned in the previous section, we noticed that the state-dependent model with an AR(1) underlying process can't properly describe the dependence structure for real observations $\{Y_t\}$. In order to improve the performance of the model, we would like to change the AR(1) to AR(2). So we have made the following changes to the model,

$$X_{i,t} = a(T_t)M_t + \sqrt{1 - a(T_t)^2} e_{i,t}$$
$$M_t = \theta_1 M_{t-1} + \theta_2 M_{t-2} + \epsilon_t \tag{3.23}$$
$$T_t = \beta M_t + \epsilon'_t, \tag{3.24}$$

where $a(\cdot)$ is the same function defined in beginning of Chapter 2

$$a(x) = \sum_{k=1}^{K} a_k \cdot \mathbb{1}(t_k < x \leqslant t_{k+1}),$$

and K is the number of market regimes. For $t = 1, \ldots,$ $e_{i,t}$ and ϵ'_t are independent error terms with mean 0 and variances $\sigma_e^2 = 1$ and $\sigma_{\epsilon'}^2 = 1 - \beta^2$ respectively. But the variance of ϵ_t is

$$\sigma_\epsilon^2 = 1 - \gamma_1\theta_1 - \gamma_2\theta_2, \tag{3.25}$$

where $\gamma_1 = \frac{\theta_1}{1-\theta_2}$ and $\gamma_2 = \theta_1\gamma_1 + \theta_2$. Also, the values of θ_1 and θ need to ensure the stationarity of $\{M_t\}$.

The main change we have made here is the process M_t. It was AR(1) in the previous model but AR(2) in here. In order to maintain the property that the stationary distribution of M_t is still standard normal, we change the variance of the error term in Equation 3.23 as well. By doing so, the relation between Y_t and M_t of the model SDM-AR(1) remains the same for the SDM-AR(2)

$$\begin{aligned} Y_t &= \Phi^{-1}\left(\Phi\left(\frac{x_{PD} - a(T_t)M_t}{\sqrt{1 - a^2(T_t)}}\right)\right) \\ &= \frac{x_{PD}}{\sqrt{1 - a(T_t)^2}} - \frac{a(T_t)}{\sqrt{1 - a(T_t)^2}} M_t, \end{aligned}$$

where x_{PD} is defined in Formula 3.6.

The stationary distribution of $\{M_t\}$, $\{T_t\}$ and the correlation between M_t and T_t for any $t = 1, \ldots, N$, remain the same as for the model SDM-AR(1). The only thing that is going to change is the dependence structures for the underlying market risk factor series $\{M_t\}$ and the market regime index $\{T_t\}$. This will also lead to a different dependence structure for the observations $\{Y_t\}$. In the following sections, we describe the new dependence structure of those series and the modification that we need to make for the filtering procedure.

3.6 Dependence structure of SDM-AR(2)

In this section, we determine the autocorrelation structure of the observed series $\{Y_t\}$ in the same manner as we have done in Section 3.4 but for the model based on the SDM-AR(2) process. So, we start from the joint distribution of $\{M_t\}$ and $\{T_t\}$ first. Then we derive the conditional distribution of $\{M_t\}$ given $\{T_t\}$. Also, due to the fact that the observations $\{Y_t\}$ are a linear transformation of $\{M_t\}$, if we give condition on $\{T_t\}$, we are able to determine the conditional distribution of $\{Y_t\}$ given $\{T_t\}$ from the conditional distribution of $\{M_t\}$ given $\{T_t\}$. As a last step, we will apply the law of total variance to get the unconditional variance matrix for $\{Y_t\}$.

Theorem 8. *The conditional distribution of $\{M_{1:N}\}$ given $\{T_{1:N}\}$ is*

$$M_{1:N}|T_{1:N} \sim \mathcal{N}(\Sigma_{MT}\Sigma_{TT}^{-1}(t_{1:N}), \Sigma_{MM} - \Sigma_{MT}\Sigma_{TT}^{-1}\Sigma_{TM}),$$

where

$$\Sigma_{MM} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(N-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(N-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(N-1) & \rho(N-2) & \rho(N-3) & \dots & 1 \end{bmatrix} \tag{3.26}$$

$$\Sigma_{MT} = \Sigma_{TM} = \begin{bmatrix} \beta & \rho(1)\beta & \rho(2)\beta & \dots & \rho(N-1)\beta \\ \rho(1)\beta & \beta & \rho(1)\beta & \dots & \rho(N-2)\beta \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(N-1)\beta & \rho(N-2)\beta & \rho(N-3)\beta & \dots & \beta \end{bmatrix} \tag{3.27}$$

$$\Sigma_{TT} = \begin{bmatrix} 1 & \rho(1)\beta^2 & \rho(2)\beta^2 & \dots & \rho(N-1)\beta^2 \\ \rho(1)\beta^2 & 1 & \rho(1)\beta^2 & \dots & \rho(N-2)\beta^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(N-1)\beta^2 & \rho(N-2)\beta^2 & \rho(N-3)\beta^2 & \dots & 1 \end{bmatrix} \tag{3.28}$$

$$\rho(1) = \frac{\theta_1}{1-\theta_2}, \quad \rho(2) = \frac{\theta_1^2 + (1-\theta_2)\theta_2}{1-\theta_2}$$

$$\rho(k) = \theta_1\rho(k-1) + \theta_2\rho(k-2) \text{ for } k = 3, 4, \dots$$

Proof. First, according to Equation 3.23, we know that $\{M_t\}$ follows an AR(2) process with a normally distributed driving noise. Also, based on Equation 3.3, $\{T_t\}$ is a linear transformation of $\{M_t\}$. As a result, we can treat the series $M_{1:N}$ and $T_{1:N}$ as a $2N$-dimension multi-variate normal distribution. So, our first step is to determine the variance matrix for $\{M_t\}$. Since $\{M_t\}$ follows an AR(2) process with constant 0 and coefficients θ_1 & θ_2, the autocorrelation function of it is

$$Corr(M_t, M_{t+k}) = \rho(k) = \theta_1\rho(k-1) + \theta_2\rho(k-2) \text{ for } k = 3, 4, \dots,$$

where

$$\rho(1) = \frac{\theta_1}{1-\theta_2}, \quad \rho(2) = \frac{\theta_1^2 + (1-\theta_2)\theta_1}{1-\theta_2}.$$

Due to the fact that the stationary distribution of $\{M_t\}$ is the standard normal distribution, we know that

$$Cov(M_t, M_{t+k}) = Corr(M_t, M_{t+k}),$$

and

$$Cov(M_t, M_t) = 1.$$

This implies the form of the covariance matrix for $M_{1:N}$ given in Matrix 3.26.

The next thing we need is the correlation between $\{M_t\}$ and $\{T_t\}$. We have

$$\begin{aligned}
Cov(M_t, T_{t+k}) =& E[M_t T_{t+k}] - E[M_t]E[T_{t+k}] \\
=& E[M_t T_{t+k}] \\
=& E[M_t(\beta M_{t+k} + \epsilon'_{t+k})] \\
=& E[\beta M_{t+k} M_t + \epsilon'_{t+k} M_t] \\
=& \beta E[M_{t+k} M_t] + E[\epsilon'_{t+k}]E[M_t] \\
=& \beta\rho(k).
\end{aligned}$$

This result suggests that the covariance matrix for $M_{1:N}$ and $T_{1:N}$ is in the form of Matrix 3.27.

Now, we find the covariance matrix for $T_{1:N}$:

$$\begin{aligned}
Cov(T_t, T_{t+k}) =& E[T_t T_{t+k}] - E[T_t]E[T_{t+k}] \\
=& E[(\beta M_t + \epsilon'_t)(\beta M_{t+k} + \epsilon'_{t+k})] \\
=& E[\beta^2 M_t M_{t+k} + \beta M_t \epsilon'_{t+k} + \beta M_{t+k} \epsilon'_t + \epsilon'_t \epsilon'_{t+k}] \\
=& \beta^2 E[M_t M_{t+k}] \\
=& \beta^2 \rho(k).
\end{aligned}$$

□

According to Theorem 6, we know that $Y_{1:N}$ given $T_{1:N}$ under the SDM-AR(2) model has the following distribution:

$$Y_{1:N}|T_{1:N} \sim \mathcal{N}(\alpha(T_{1:N}) - \chi(T_{1:N})\Sigma_{MT}\Sigma_{TT}^{-1} \cdot (T_{1:N}),\ \chi(T_{1:N})(\Sigma_{MM} - \Sigma_{MT}\Sigma_{TT}^{-1}\Sigma_{TM})\chi(T_{1:N})^T),$$

where $\alpha(T_{1:N})$ and $\chi(T_{1:N})$ are defined on Page 59.

Then the unconditional covariance matrix and the autocorrelations of $\{Y_t\}$ can be calculated by Theorem 7 and the numerical method mentioned on Page 59.

3.7 Likelihood function of Y_t in SDM-AR(2)

Before we calculate the joint likelihood function of $\{Y_t\}$, we need to determine the new one-step-ahead state predictive density, $f_{M_{t+1}|Y_{1:t}}(m_{t+1}|y_{1:t})$. Since we have changed the process of M_t from AR(1) to AR(2), we need the formula of $f_{M_{t+1},M_t|Y_{1:t}}(m_{t+1}, m_t|y_{1:t})$ instead of $f_{M_{t+1}|Y_{1:t}}(m_{t+1}|y_{1:t})$ to calculate the joint likelihood function of Y_t.

Theorem 9. *By the one-step-ahead state predictive density, $f_{M_{t+1},M_t|Y_{1:t}}(m_{t+1}, m_t|y_{1:t})$, can be calculated by following equation*

$$f_{M_{t+1},M_t|Y_{1:t}}(m_{t+1}, m_t|y_{1:t}) = \sum_{i=1}^{2} f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t, M^{-1}(y_{t-1}, i)) \cdot$$
$$P(M_t = m_t, M_{t-1} = M^{-1}(y_{t-1}, i)|Y_{1:t} = y_{1:t})$$

where $f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t, m_{t-1})$ is the transition density of an AR(2) process.

Proof. First, based on the law of total probability, we know that

$$f_{M_{t+1},M_t|Y_{1:t}}(m_{t+1}, m_t|y_{1:t}) = \sum_{i=1}^{2} f_{M_{t+1},M_t,M_{t-1}|Y_{1:t}}(m_{t+1}, m_t, M^{-1}(y_{t-1}, i)|y_{1:t}).$$

Then, since M_t follows an AR(2) process, M_{t+1} is independent of the past given M_t and

M_{t-1}. Therefore,

$$f_{M_{t+1},M_t|Y_{1:t}}(m_{t+1},m_t|y_{1:t}) = \sum_{i=1}^{2} f_{M_{t+1}|M_t,M_{t-1},Y_{1:t}}(m_{t+1}|m_t, M^{-1}(y_{t-1},i), y_{1:t}) \cdot$$
$$P(M_t = m_t, M_{t-1} = M^{-1}(y_{t-1},i)|Y_{1:t} = y_{1:t})$$
$$= \sum_{i=1}^{2} f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t, M^{-1}(y_{t-1},i)) \cdot$$
$$P(M_t = m_t, M_{t-1} = M^{-1}(y_{t-1},i)|Y_{1:t} = y_{1:t}), \tag{3.29}$$

where $f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t,m_{t-1})$ is the transition density of an AR(2) process. □

For the next step, we need to figure out how to calculate the new one-step-ahead predictive density $f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t})$.

Theorem 10. *The one-step ahead predictive density* $f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t})$ *is*

$$f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t}) = \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} P(Y_{t+1} = y_{t+1}|M_{t+1} = M^{-1}(y_{t+1},i))$$
$$\sum_{j=1}^{2} f_{M_{t+1},M_t|Y_{1:t}}(M^{-1}(y_{t+1},i), M^{-1}(y_t,j)|y_{1:t}),$$

where $P(Y_{t+1} = y_{t+1}|M_{t+1} = M^{-1}(y_{t+1},i))$ *is*

$$P(Y_{t+1} = y_{t+1}|M_{t+1} = M^{-1}(y_{t+1},i)) = P(T_{t+1} \in [t_i, t_{i+1})|M_{t+1} = M^{-1}(y_{t+1},i)).$$

Proof. We know that

$$f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t}) = \sum_{i=1}^{2} f_{Y_{t+1},M_{t+1}|Y_{1:t}}(y_{t+1}, M^{-1}(y_{t+1},i)|y_{1:t}).$$

Since Y_{t+1} will be independent of the previous values once M_{t+1} is given, we can keep writing the equation in the following way

$$
\begin{aligned}
f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t}) =& \sum_{i=1}^{2} f_{Y_{t+1}|M_{t+1}}(y_{t+1}|M^{-1}(y_{t+1},i)) \cdot f_{M_{t+1}|Y_{1:t}}(M^{-1}(y_{t+1},i)|y_{1:t}) \\
=& \sum_{i=1}^{2} \frac{\sqrt{1-a_i^2}}{a_i} P(Y_{t+1} = y_{t+1}|M_{t+1} = M^{-1}(y_{t+1},i)) \\
& \sum_{j=1}^{2} f_{M_{t+1},M_t|Y_{1:t}}(M^{-1}(y_{t+1},i), M^{-1}(y_t,j)|y_{1:t}).
\end{aligned}
$$

This gives us the result. □

After that, we can notice from the Equation 3.29 that we need to find a way to calculate the new filtering function $P(M_t = m_t, M_{t-1} = M^{-1}(y_{t-1}, i)|Y_{1:t} = y_{1:t})$.

Theorem 11. *We can calculate* $P(M_t = m_t, M_{t-1} = M^{-1}(y_{t-1}, i)|Y_{1:t} = y_{1:t})$ *in the following way:*

$$
\begin{aligned}
P(M_t = M^{-1}(y_t,i), M_{t-1} = M^{-1}(y_{t-1},j)|Y_{1:t} = y_{1:t}) =& \frac{P(Y_t = y_t|M_t = M^{-1}(y_t,i))}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})} \cdot \\
& f_{M_t,M_{t-1}|Y_{1:t-1}}(M^{-1}(y_t,i), M^{-1}(y_{t-1},j)|y_{1:t-1}).
\end{aligned}
$$

Proof. We have

$$
\begin{aligned}
P(M_t = M^{-1}(y_t,i), M_{t-1} = M^{-1}(y_{t-1},j)|Y_{1:t} = y_{1:t}) =& \frac{f_{M_t,M_{t-1},Y_t|Y_{1:t-1}}(M^{-1}(y_t,i), M^{-1}(y_{t-1},j), y_t|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})} \\
=& f_{Y_t|M_t,M_{t-1},Y_{1:t-1}}(y_t|M^{-1}(y_t,i), M^{-1}(y_{t-1},j), y_{1:t-1}) \cdot \frac{f_{M_t,M_{t-1}|Y_{1:t-1}}(M^{-1}(y_t,i), M^{-1}(y_{t-1},j)|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})} \\
=& P(Y_t = y_t|M_t = M^{-1}(y_t,i)) \cdot \frac{f_{M_t,M_{t-1}|Y_{1:t-1}}(M^{-1}(y_t,i), M^{-1}(y_{t-1},j)|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})}.
\end{aligned}
$$

□

Once we have those formulas, we are able to perform a similar filtering procedure as the one we have described in Section 3.2.4, and then to compute the values of $f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t})$ for all $t = 1 \ldots T-1$. Based on these, the log-likelihood function of $\{Y_t\}$ is

$$
\log L(\delta; y_{1:N}) = \log f_Y(y_1; \delta) + \sum_{i=2}^{T} \log(f_{Y_i|Y_{1:i-1}}(y_i|y_{1:i-1}; \delta)).
$$

3.7.1 K-step-ahead prediction of SDM-AR(2)

In this section, we describe a method of finding the K-step-ahead prediction for the observations $\{Y_t\}$ based on the SDM-AR(2) model. Although we have already derived procedure to the SDM-AR(1) model, we still need to modify some parts of the previous derivation since M_t is changed from AR(1) to AR(2). The following theorem presents an equation to calculate the K-step-ahead predictive function in SDM-AR(2).

Theorem 12. *The K-step-ahead predictive density of Y_{t+K} given $Y_{1:t}$, which is denoted by $f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t})$, can be calculated in the following way:*

$$\begin{aligned} f_{Y_{t+k}|Y_{1:t}}(y_{t+k}|y_{1:t}) = &\sum_{i=1}^{2}\sum_{j=1}^{2} P(Y_{t+k} = y_{t+k}|M_{t+k} = M^{-1}(y_{t+k}, j))\cdot \\ &f_{M_{t+k}|M_t,M_{t-1}}(M^{-1}(y_{t+k}, j)|M^{-1}(y_t, i), M^{-1}(y_{t-1}, i))\cdot \\ &P(M_t = M^{-1}(y_t, i), M_t = M^{-1}(y_{t-1}, i)|Y_{1:t} = y_{1:t}), \end{aligned}$$

where $f_{M_{t+k}|M_t,M_{t-1}}(M^{-1}(y_{t+k}, j)|M^{-1}(y_t, i), M^{-1}(y_{t-1}, i))$ is the K-step transition density function of the AR(2) model.

The proof for the K-step ahead predictive density of SDM-AR(2) is similar to the proof of Theorem 4, so we omit it here.

3.8 Special cases of the SDM-AR(2) model

In this section, we wish to demonstrate that the AR(1), AR(2) and SDM-AR(1) models are nested within the SDM-AR(2) model as special cases. It is trivial to see that we can set the parameter θ_2 in Equation 3.23 equal to zero to make the SDM-AR(2) model the same as the SDM-AR(1) model. Then, if we can prove that the classic AR(1) model in the following form

$$Y_t = C + \alpha Y_{t-1} + \epsilon'_t \tag{3.30}$$

can also be expressed as a special case of the SDM-AR(1) model, it would be enough for us to conclude that the AR(1) model also belongs to the subset of the SDM-AR(2) models, since the SDM-AR(1) model is obviously a subset of SDM-AR(2) model. After that, we will demonstrate that the AR(2) model defined as

$$Y_t = C + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \epsilon'_t$$

is a special case of the SDM-AR(2) model.

3.8.1 Expressing AR(1) model in terms of SDM-AR(1)

The following theorem demonstrates that AR(1) model can be expressed in terms of the SDM-AR(1) model.

Theorem 13. *By setting the parameters of SDM-AR(1) in the following forms,*

$$a_1 = a_2 = \sqrt{\frac{\sigma^2_{\epsilon'_t}}{(1-\alpha^2+\sigma^2_{\epsilon'_t})}} \tag{3.31}$$

$$\theta = \alpha$$

$$PD = \Phi\left(\frac{\sqrt{1-a_1^2}C}{(1-\alpha)}\right),$$

where PD is the sixth parameter for the SDM-AR(1) model, which controls the long-run average of the observations D_t. The SDM-AR(1) model is identical with the traditional AR(1) model in Equation 3.30.

Proof. According to Formulas 3.2 and 3.8 we have:

$$\begin{aligned} Y_t &= \frac{x_{PD}}{\sqrt{1-a(R_t)^2}} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}(\theta M_{t-1}+\epsilon_t) \\ &= \frac{x_{PD}}{\sqrt{1-a(R_t)^2}} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\theta M_{t-1} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\epsilon_t. \end{aligned} \tag{3.32}$$

Also based on Equation (3.8), we know that:

$$M_{t-1} = \frac{x_{PD}}{a(R_{t-1})} - \frac{\sqrt{1-a(R_{t-1})^2}}{a(R_{t-1})}Y_{t-1}.$$

By substituting this into Equation (3.32), we get

$$Y_t = \frac{x_{PD}}{\sqrt{1-a(R_t)^2}}(1-\theta\frac{a(R_T)}{a(R_{t-1})}) + \theta\frac{a(R_t)}{a(R_{t-1})}\frac{\sqrt{1-a(R_{t-1})^2}}{\sqrt{1-a(R_t)^2}}Y_{t-1} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\epsilon_t. \tag{3.33}$$

Then for any AR(1) model in the following form

$$Y_t = C + \alpha Y_{t-1} + \epsilon'_t,$$

we only need to set the parameters of SDM-AR(1) to meet the following conditions

$$
\begin{aligned}
Var(\frac{a_1}{\sqrt{1-a_1^2}}\epsilon_t) &= Var(\epsilon_t') \\
\frac{x_{PD}}{\sqrt{1-a_1^2}}(1-\theta) &= C \\
\theta &= \alpha \\
a_1 &= a_2.
\end{aligned}
$$

Then, from the condition $Var(\frac{a_1}{\sqrt{1-a_1^2}}\epsilon_t) = Var(\epsilon_t')$ and $\theta = \alpha$, we will have

$$
\begin{aligned}
\frac{a_1^2}{1-a_1^2}(1-\theta^2) &= \sigma_{\epsilon_t'}^2 \\
a_1^2(1-\theta^2) &= \sigma_{\epsilon_t'}^2 - \sigma_{\epsilon_t'}^2 a_1^2 \\
a_1^2(1-\alpha^2+\sigma_{\epsilon_t'}^2) &= \sigma_{\epsilon_t'}^2 \\
a_1 &= \sqrt{\frac{\sigma_{\epsilon_t'}^2}{(1-\alpha^2+\sigma_{\epsilon_t'}^2)}}.
\end{aligned}
$$

If we set $a_1 = a_2$, then we know that X_t defined in Equation 3.1 actually follows the standard normal distribution. Then according to the definition of x_{PD} in Section 2.1.3, we have

$$
x_{PD} = \Phi^{-1}(PD),
$$

where PD is the sixth parameter for the SDM-AR(1) model, which controls the long-run average of the observation D_t. Then as a result, we get

$$
\begin{aligned}
\frac{x_{PD}}{\sqrt{1-a_1^2}}(1-\theta) &= C \\
\frac{\Phi^{-1}(PD)}{C}(1-\theta) &= \sqrt{1-a_1^2}.
\end{aligned}
$$

We also have the condition that $\theta = \alpha$, so

$$\begin{aligned}
\frac{\Phi^{-1}(PD)}{C}(1-\alpha) &= \sqrt{1-a_1^2} \\
\Phi^{-1}(PD) &= \frac{\sqrt{1-a_1^2}C}{(1-\alpha)} \\
PD &= \Phi\left(\frac{\sqrt{1-a_1^2}C}{(1-\alpha)}\right).
\end{aligned}$$

In addition, as the result of $a_1 = a_2$, the impacts of β and t_1 are eliminated. Therefore, there is no constraint for those two parameters. □

3.8.2 Expressing AR(2) model in terms of SDM-AR(2)

In this section, we prove that the traditional AR(2) model defined in the following way

$$Y_t = C + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \epsilon'_t \tag{3.34}$$

can be expressed as a special case of the SDM-AR(2) model.

Theorem 14. *By setting the parameters of SDM-AR(2) model in the following form,*

$$\begin{aligned}
a_1 = a_2 &= \sqrt{\frac{\sigma^2_{\epsilon'_t}}{(\sigma^2_{\epsilon_t} + \sigma^2_{\epsilon'_t})}} \\
\theta_1 &= \alpha_1 \\
\theta_2 &= \alpha_2 \\
PD &= \Phi(\frac{\sqrt{1-a_1^2}C}{1-\alpha_1-\alpha_2}),
\end{aligned} \tag{3.35}$$

where $\sigma^2_{\epsilon'_t}$ is the variance of ϵ'_t, and $\sigma^2_{\epsilon_t}$ defined in 3.25 is the variance of the driving noise for series $\{M_t\}$. The SDM-AR(2) model is identical with the classic AR(2) model in Equation 3.34.

Proof. As we mentioned in Section 3.5, the relation between Y_t and M_t defined in 3.8 is still valid for SDM-AR(2) model. Then, according to Equations 3.23, we have

$$Y_t = \frac{x_{PD}}{\sqrt{1-a(R_t)^2}} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}(\theta_1 M_{t-1} + \theta_2 M_{t-2} + \epsilon_t)$$
$$Y_t = \frac{x_{PD}}{\sqrt{1-a(R_t)^2}} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\theta_1 M_{t-1} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\theta_2 M_{t-2} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\epsilon_t. \tag{3.36}$$

Also based on Equation (3.8), we know that:

$$M_{t-1} = \frac{x_{PD}}{a(R_{t-1})} - \frac{\sqrt{1-a(R_{t-1})^2}}{a(R_{t-1})}Y_{t-1}.$$
$$M_{t-2} = \frac{x_{PD}}{a(R_{t-2})} - \frac{\sqrt{1-a(R_{t-2})^2}}{a(R_{t-2})}Y_{t-2}.$$

By substituting this into Equation 3.36, we get

$$Y_t = \tilde{C} + \tilde{\alpha_1} Y_{t-1} + \tilde{\alpha_2} Y_{t-2} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\epsilon_t,$$

where

$$\tilde{C} = \frac{x_{PD}}{\sqrt{1-a(R_t)^2}}(1 - \theta_1 \frac{a(R_T)}{a(R_{t-1})} - \theta_2 \frac{a(R_T)}{a(R_{t-2})})$$
$$\tilde{\alpha_1} = \theta_1 \frac{a(R_t)}{a(R_{t-1})} \frac{\sqrt{1-a(R_{t-1})^2}}{\sqrt{1-a(R_t)^2}}$$
$$\tilde{\alpha_2} = \theta_2 \frac{a(R_t)}{a(R_{t-2})} \frac{\sqrt{1-a(R_{t-2})^2}}{\sqrt{1-a(R_t)^2}}.$$

Under the condition $a_1 = a_2$, we know that $a(T_t)$ is actually constant. Therefore, we can simplify the above equations

$$\tilde{C} = \frac{x_{PD}}{\sqrt{1-a_1^2}}(1 - \theta_1 - \theta_2) \tag{3.37}$$
$$\tilde{\alpha_1} = \theta_1$$
$$\tilde{\alpha_2} = \theta_2.$$

As a result, we know that we need to set θ_1 and θ_2 in SDM-AR(2) equal to α_1 and α_2 respectively to make those two models equivalent. Also due to the condition $a_1 = a_2$, we know that X_t defined in Equation 3.1 actually follows the standard normal distribution. Then according to the definition of x_{PD} in Section 2.1.3, we have

$$x_{PD} = \Phi^{-1}(PD),$$

where PD is the seventh parameter for the SDM-AR(2) model which controls the long-run average of the observation D_t. In order to have $C = \tilde{C}$, the following condition has to be valid

$$\begin{aligned}
C &= \frac{\Phi^{-1}(PD)}{\sqrt{1-a_1^2}}(1-\theta_1-\theta_2) \\
C &= \frac{\Phi^{-1}(PD)}{\sqrt{1-a_1^2}}(1-\alpha_1-\alpha_2) \\
\Phi^{-1}(PD) &= \frac{\sqrt{1-a_1^2}C}{1-\alpha_1-\alpha_2} \\
PD &= \Phi(\frac{\sqrt{1-a_1^2}C}{1-\alpha_1-\alpha_2}),
\end{aligned}$$

where the value of a_1 can be determined from the condition that $Var(\frac{a_1}{\sqrt{1-a_1^2}}\epsilon_t) = Var(\epsilon_t')$. That means

$$\frac{a_1^2}{1-a_1^2}\sigma_{\epsilon_t}^2 = \sigma_{\epsilon_t'}^2,$$

where $\sigma_{\epsilon_t}^2$ stands for the variance of ϵ_t which is defined in Equation 3.25. By solving above equation with respect to a_1, we have

$$\begin{aligned}
\frac{a_1^2}{1-a_1^2}\sigma_{\epsilon_t}^2 &= \sigma_{\epsilon_t'}^2 \\
a_1^2\sigma_{\epsilon_t}^2 &= \sigma_{\epsilon_t'}^2 - \sigma_{\epsilon_t'}^2 a_1^2 \\
a_1^2(\sigma_{\epsilon_t}^2 + \sigma_{\epsilon_t'}^2) &= \sigma_{\epsilon_t'}^2 \\
a_1 &= \sqrt{\frac{\sigma_{\epsilon_t'}^2}{(\sigma_{\epsilon_t}^2 + \sigma_{\epsilon_t'}^2)}}.
\end{aligned}$$

□

In conclusion, the SDM-AR(2) model is an unrestricted form of SDM-AR(1), AR(2) and AR(1) model. Then theoretically, for the same data set, the performance of the SDM-AR(2) should always be better than the others. We will compare those four models in the following chapter from the empirical perspective.

Chapter 4

Empirical Study of SDM-AR Model

In this chapter, we conduct an empirical study using historical data. The data we use is the same as the one used in Section 2.2.3. Let us recall that the data is obtained from the Federal Reserve and consists of quarterly delinquency rates for 11 different categories. The time period is from the first quarter of 1991 to the fourth quarter of 2016.

This Chapter is divided into the following sections. In Section 4.1, we estimate the parameters of the state dependent model with one and two lags in the autoregressive process driving the systematic risk factor. We find that the degree of state dependence, β, is relatively high in virtually all series. In Section 4.2, we assess the models' in-sample forecasting ability (point and 99.9% interval estimates, which are of the greatest interest to risk managers.). We find that the state-dependent model with two lags generates considerably more accurate in-sample forecasts through the financial crisis. In Section 4.3, we assess the models' out-of-sample one-step and four-step-ahead point forecasting ability. We realize that the state-dependent model with two lags can predict market behavior more accurately than the other models. In the last section, we evaluate the model's out-of-sample one-step and four-step-ahead interval forecasting. The results strongly agree that the state-dependent model with two lags shows strong advantages.

4.1 Parameter estimation

In order to estimate the parameters for each series appearing in Table **??**, we apply the methodology described in Sections 3.3 & 3.7. Let us recall that the log-likelihood functions for both SDM-AR(1) and SDM-AR(2) models take the form:

$$\ell(\xi; \vec{Y}) = \log(f_{Y_1;\xi}(y_1)) + \sum_{i=2}^{N} \log(f_{Y_i|Y_{1:i-1};\xi}(y_i|y_{1:i-1})), \tag{4.1}$$

where $\vec{Y}$ is the vector of transformed default rates that are calculated by

$$Y_i = \Phi^{-1}(D_i), \tag{4.2}$$

and ξ is the vector of parameters, which contains $\{a_1, a_2, \beta, t_2, \theta_1\}$ for the SDM-AR(1) model and $\{a_1, a_2, \beta, t_2, \theta_1\, \theta_2\}$ for the SDM-AR(2) model.

4.1.1 Selection of initial points

Further recall that the likelihood function must be optimized numerically. To find the global maximum, we systematically select different initial points, and for each initial point, we use the Matlab built-in optimizer, `fmincon`, to find a local minimum of the negative log-likelihood, $-\ell(\xi; \vec{Y})$. We then take that the local minimizer that produces the lowest value as the global minimum, i.e., maximum likelihood estimate.

For the initial points of the numerical optimizer for the SDM-AR(1) model, we defined a set of values for each parameter and used all the possible combinations of those values to create a mesh over the following space:

$$\begin{aligned} a &= [0.5, 0.4, 0.3, 0.2, 0.1] \\ t &= [-1, -0.5, 0, 0.5, 1] \\ \beta &= [0.1, 0.3, 0.5, 0.7, 0.9] \\ \theta &= [-1, -0.5, 0, 0.5, 1]. \end{aligned}$$

For a_1 and a_2, we chose all the combinations from a such that $a_1 > a_2$. That means there are 10 different sets of values for $[a_1, a_2]$ and 5 sets for t, β and θ, respectively. This gives us 1250 different initial points in total.

For the SDM-AR(2) model, we keep the space for a, t and β unchanged. As we can tell from the historical ACF plots in Section 3.4.3, the data sets suggest that there is a high correlation between each time. Then we set the initial points for θ_1 and θ_2 from the following combinations:

$$\{\theta_1, \theta_2\} = [\{0, 0\}, \{0.7, 0.29\}, \{0.8, 0.1\}, \{1.1, -0.2\}, \{1.5, -0.52\}].$$

As a result, there are 1250 different initial points for the SDM-AR(2) model.

4.1.2 Estimation results

After we finish running `fmincon` from different initial points for each model, we pick the results with the highest likelihood value as our estimate for the parameters. The following table shows the five largest log-likelihood values that we obtained from 1250 different starting points for the All series in the SDM-AR(1) model.

a_1	a_2	β	t_1	θ	Log likelihood
0.228	0.137	1.00	-0.634	0.984	219.809
0.226	0.136	1.00	-0.634	0.983	219.808
0.232	0.139	1.00	-0.634	0.984	219.808
0.232	0.139	1.00	-0.634	0.984	219.808
0.232	0.139	1.00	-0.634	0.984	219.808

Table 4.1: Calibration results of SDM-AR(1) model on All series

As we can see from Table 4.1, even if the estimation procedure starts from different initial points, the final results converge to the same point. This gives us confidence that if we systematically explore the parameter space, we are typically able to locate to the global maximum, i.e., that we are accurately computing the maximum likelihood estimates. The same method is applied to the rest of the data with the SDM-AR(1) and SDM-AR(2), respectively.

Series	Model	a_1	a_2	β	t_1	θ_1	θ_2	$\theta_1 + \theta_2$
All	SDM-AR(2)	0.232 (0.089)	0.162 (0.076)	1 (0.177)	-0.704 (1.123)	1.832 (0.047)	-0.853 (0.046)	0.979
	SDM-AR(1)	0.229 (0.082)	0.137 (0.053)	1 (0.056)	-0.635 (0.726)	0.984 (0.011)		
OC	SDM-AR(2)	0.115 (0.076)	0.085 (0.032)	1 (0.298)	-0.641 (1.435)	1.366 (0.088)	-0.386 (0.085)	0.98
	SDM-AR(1)	0.119 (0.071)	0.088 (0.075)	1 (0.137)	-0.631 (0.934)	0.986 (0.031)		
SRE	SDM-AR(2)	0.246 (0.071)	0.234 (0.079)	-0.303 (0.393)	0.672 (1.001)	1.878 (0.037)	-0.892 (0.035)	0.986
	SDM-AR(1)	0.236 (0.148)	0.221 (0.108)	1 (0.618)	1.323 (0.914)	0.987 (0.018)		
CRE	SDM-AR(2)	0.392 (0.128)	0.378 (0.109)	-1 (0.359)	-0.621 (1.399)	1.790 (0.051)	-0.802 (0.047)	0.988
	SDM-AR(1)	0.418 (0.176)	0.402 (0.156)	1 (0.298)	0.6943 (0.987)	0.993 (0.024)		

Table 4.2: Estimation results of SDM-AR(1) and SDM-AR(2) models

For a direct comparison of the SDM-AR(1) and SDM-AR(2) model estimation, Table 4.2 shows part of the results from these two models, with the remaining results presented in the Appendix. We can notice the following notable features of the data based on Table 4.2:

1. The factor loadings a_1 and a_2 in both the SDM-AR(1) and SDM-AR(2) models are similar.

2. The degree of state dependence, β, is extremely high for almost every series except SRE.

3. When the factor loadings in both regimes are similar, it appears that the state dependence parameter β and state change threshold t_1 are very hard to identify. This may be caused by the fact that a_1 and a_2 are close to each other in this series, which makes the effect of the different regimes negligible. As a result, the impacts of β and t_1 are also dramatically reduced.

For the parameters θ_1 and θ_2 in the SDM-AR(2) model, we can rewrite Formula 3.23 for M in the following way:

$$\begin{aligned} M_t &= \theta_1 M_{t-1} + \theta_2 M_{t-2} + \epsilon_t \\ &= (\theta_1 + \theta_2) M_{t-1} - \theta_2 (M_{t-1} - M_{t-2}) + \epsilon_t. \end{aligned}$$

As we can see from Table 4.2, the value of $\theta_1 + \theta_2$ in the SDM-AR(2) model is close to the value of θ_1 in the SDM-AR(1) model. Also the values of θ_2 are far away from zero. That implies that the dynamic of M_t not only heavily depends on the previous value M_{t-1} but also on the previous increment, $M_{t-1} - M_{t-2}$. Since $\theta_1 + \theta_2 \approx 1$, the next period's expected market level is the last period's market level, plus an adjustment based on momentum.

It is natural to ask if our model improves the forecasting ability of the classic AR model. In order to compare our models with the plain AR model, we also fitted the following AR model to our data with one or two lag terms:

$$Y_t = C + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \epsilon''_t$$

Again, $Y_t = \Phi^{-1}(D_t)$, where D_t is the observation at time t. C is a constant and $\{\epsilon''_t\}_{t=1,2,\ldots}$ follow the i.i.d. normal distribution with mean 0 and variance $\sigma^2_{\epsilon''}$. The setting for the AR model remains the same as in the rest of this study.

Series	Model	a_1	a_2	β	t_1	θ_1	$\sigma^2_{\epsilon''}$	C
All	SDM-AR(1)	0.229 (0.082)	0.137 (0.053)	1 (0.056)	-0.635 (0.726)	0.984 (0.011)		
	AR(1)	0.154				0.980 (0.016)	0.0094 (0.0001)	-0.0413 (0.0297)
OC	SDM-AR(1)	0.119 (0.071)	0.088 (0.075)	1 (0.137)	-0.631 (0.934)	0.986 (0.031)		
	AR(1)	0.099				0.986 (0.018)	0.00027 (0.00003)	-0.0291 (0.036)
SRE	SDM-AR(1)	0.236 (0.148)	0.221 (0.108)	1 (0.618)	1.323 (0.914)	0.987 (0.018)		
	AR(1)	0.209				0.985 (0.013)	0.00128 (0.00018)	-0.0297 (0.023)
CRE	SDM-AR(1)	0.418 (0.176)	0.402 (0.156)	1 (0.298)	0.6943 (0.987)	0.993 (0.024)		
	AR(1)	0.279				0.987 (0.013)	0.00216 (0.0002)	-0.0372 (0.023)

Table 4.3: In-sample estimation results of SDM-AR(1) and AR(1) models

Table 4.3 presents the results from both one-lag term models. The a_1 column for the AR(1) model is the implied factor loading based on Equation 3.31. From this table, we can make the following observations:

1. The coefficients of the AR process are almost the same for both models. This fact strongly suggests that the AR process plays an important role in default rate modeling.

2. The implied factor loading of the classic AR model is close to the average of SDM-AR(1) model's two-factor loadings.

When we compare estimates based on the SDM-AR(2) and AR(2) models presented in Table 4.4, we can observe the same phenomena as in Table 4.3.

Series	Model	a_1	a_2	β	t_1	θ_1	θ_2	$\sigma^2_{\epsilon''}$	C
All	SDM-AR(2)	0.232 (0.089)	0.162 (0.076)	1 (0.177)	-0.704 (1.123)	1.832 (0.047)	-0.853 (0.046)		
	AR(2)	0.189				1.824 (0.052)	-0.846 (0.049)	0.00024 (0.00002)	-0.0415 (0.018)
OC	SDM-AR(2)	0.115 (0.076)	0.085 (0.032)	1 (0.298)	-0.641 (1.435)	1.366 (0.088)	-0.386 (0.085)		
	AR(2)	0.089				1.340 (0.091)	-0.364 (0.089)	0.00023 (0.00003)	-0.0461 (0.033)
SRE	SDM-AR(2)	0.246 (0.071)	0.234 (0.079)	-0.303 (0.393)	0.672 (1.001)	1.878 (0.037)	-0.892 (0.035)		
	AR(2)	0.273				1.794 0.051)	-0.808 (0.051)	0.00042 0.00006)	-0.0253 (0.010)
CRE	SDM-AR(2)	0.392 (0.128)	0.378 (0.109)	-1 (0.359)	-0.621 (1.399)	1.790 (0.051)	-0.802 (0.047)		
	AR(2)	0.333				1.735 (0.060)	-0.750 (0.057)	0.0009 (0.0001)	-0.0305 (0.019)

Table 4.4: In-sample estimation results of SDM-AR(2) and AR(2) models

4.2 In-sample forecasting performance

In this section, we assess the accuracy of the models' in-sample point and interval predictions, at various horizons. First, we use the estimated parameters in Section 4.1.2 to generate the one-step-ahead prediction plots by the following method:

1. For each time $t = 1, \ldots, T-1$ for SDM-AR(1) and $t = 2, \ldots, T-1$ for SDM-AR(2), extract the observations $Y_{1:t}$ from the historical data.

2. Calculate the one-step-ahead predictive density $f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t})$ using the maximum likelihood estimates obtained in Section 4.1.2.

3. Evaluate the expected value of $\Phi(Y_{t+1})$ given $Y_{1:t}$ as the predictive value of the default rate for next time period.

$$\hat{D}_{t+1} = \int_{-\infty}^{\infty} \Phi(x) \cdot f_{Y_{t+1}|Y_{1:t}}(x|y_{1:t})dx.$$

We use the built-in integral function, `int`, in Matlab to evaluate this integral numerically.

4. For the $\alpha\%$ confidence interval of Y_{t+1} given $Y_{1:t} = y_{1:t}$, solve the following equation w.r.t x numerically to find out the lower and upper boundary, $Y_{t+1,\frac{1-\alpha\%}{2}}$ and $Y_{t+1,\frac{\alpha\%+1}{2}}$, respectively

$$\int_{-\infty}^{x} f_{Y_{t+1}|Y_{1:t}}(z|y_{1:t})dz = \frac{1-\alpha\%}{2} \text{ or } \frac{\alpha\%+1}{2}.$$

Since the relations between Y_t and D_t are monotone and bijective for any t, the confidence boundary of D_t equals to $\Phi(Y_{t+1,\frac{1-\alpha\%}{2}})$ and $\Phi(Y_{t+1,\frac{\alpha\%+1}{2}})$.

5. Repeat Steps 1 to 4 until $t = T - 1$.

After that, we plot the predictions we generate along with the historical data to have a direct picture of how the prediction evolves with respect to time. We also calculate some accuracy measures to help us quantify the performance of each model.

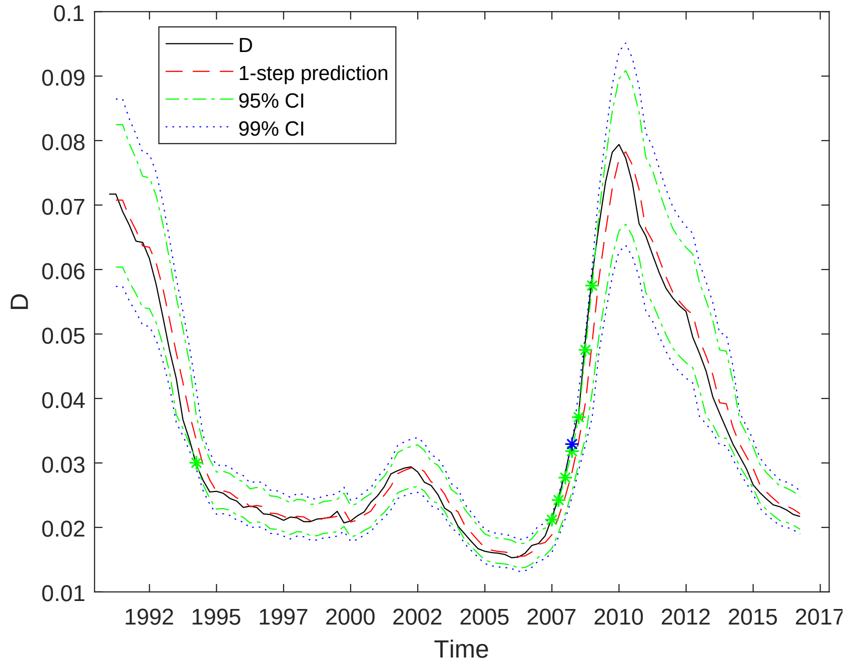
4.2.1 In-sample prediction (point and interval estimates)

Sub-graph (a) in Figure 4.1 presents the predictions based on the fitted SDM-AR(1) model and (b) based on the fitted SDM-AR(2) model. The red dash line in each plot is the expected value of the one-step-ahead prediction of $\Phi(Y)$, $E[\Phi(Y_t)|Y_{1:t-1}]$. The green dash-dotted and blue dotted lines represent the 95% and 99% CI, respectively. The lower lines are the 2.5% and 0.5% quantiles of $\Phi(Y_t)|Y_{1:t-1}$. The upper lines are 97.5% and 99.5% quantiles of $\Phi(Y_t)|Y_{1:t-1}$. The star markers in the plot represent the points at which the actual historical data breaches of the corresponding model's confidence interval.

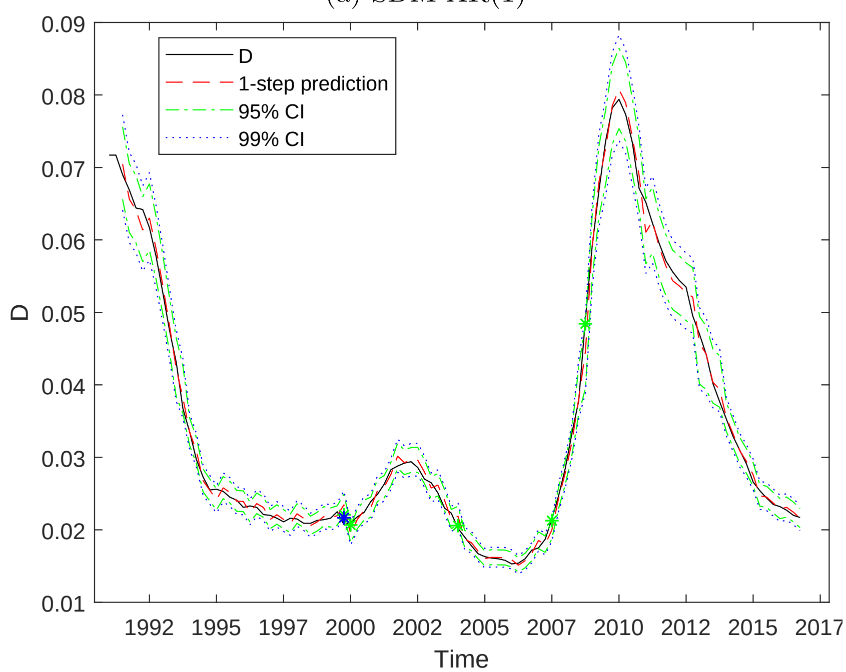
First, we can tell from the plots that the SDM-AR(2) model has better prediction during the financial crisis (2007 to 2010). Most of the time, the CI generated by the SDM-AR(2) shows distinct advantages. For example, in Figure 4.1, the 95% CI generated by the SDM-AR(1) model breached 7 times (6.8%) while the one based on SDM-AR(2) model only breached 4 times (4.9%).

It is also worth pointing out that the prediction made by SDM-AR(1) is pretty close to a naive prediction, where naive prediction is produced as equal to the last observed value. This result is mainly caused by the high values of θ_1, since all the values of θ_1 for these 3 series are at least 0.98. As a result, the prediction SDM-AR(1) made for the next time would be close to the current level. But when we look at the prediction made by SDM-AR(2), this phenomenon is less obvious.

Figure 4.2 shows a comparison between the fitting results from the SDM-AR(2) and AR(2) models. The plot legend and style remain the same as in Figure 4.1. We can tell from this comparison that the CI generated by the SDM-AR(2) model is more reliable compared with the one generated by the AR(2) model. As we can see from the plot, for the All series, both the 95% and 99% CI of SDM-AR(2) model breached one point less than the one of the AR(2) model.

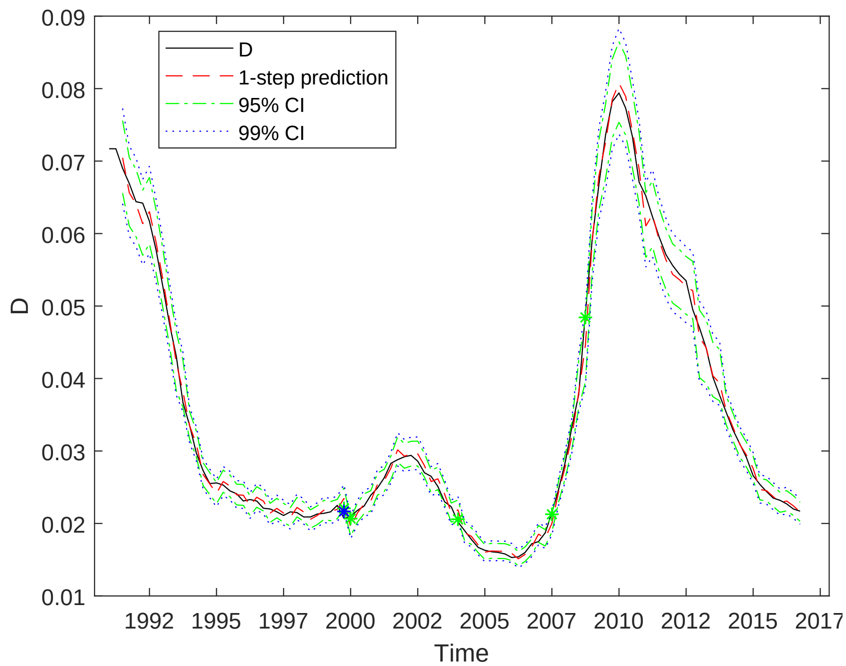


(a) SDM-AR(1)

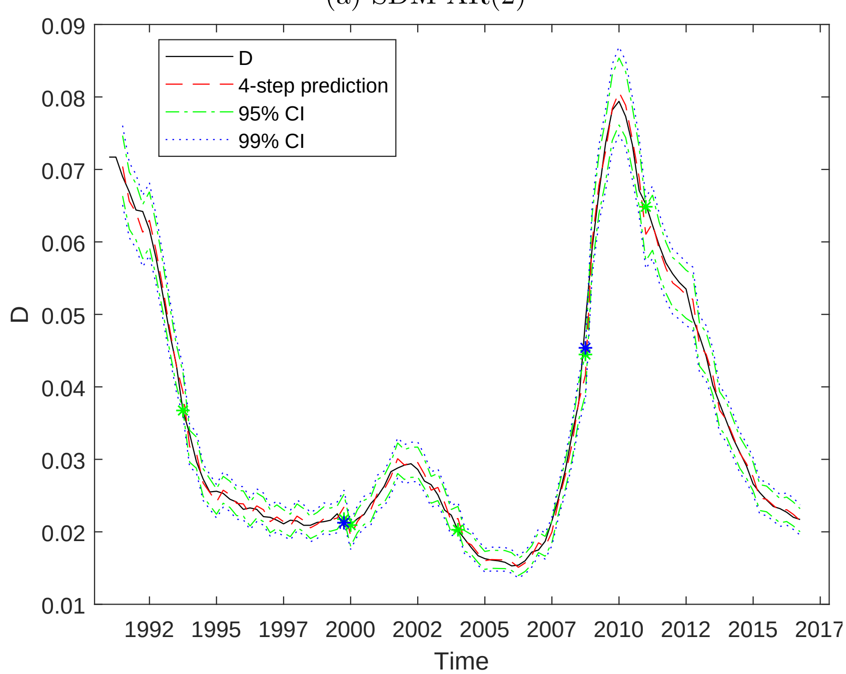


(b) SDM-AR(2)

Figure 4.1: In-sample predictive value and confidence intervals for historical "All" series based on fitted SDM-AR(1) and SDM-AR(2) models.



(a) SDM-AR(2)



(b) AR(2)

Figure 4.2: In-sample predictive value and confidence intervals for historical "All" series based on fitted SDM-AR(2) and AR(2) models.

However, for the point forecasting accuracy, it is hard to make a conclusion directly from the plots. So we have also calculate the following popular accuracy measurements, where MAE stands for the mean absolute error, RMSE is the root mean squared error, and MAPE is the mean absolute percentage error:

$$\text{MAE} = N^{-1} \sum_{i=1}^{N} |D_t - \hat{D}_t| \tag{4.3}$$

$$\text{RMSE} = \sqrt{N^{-1} \sum_{i=1}^{N} (D_t - \hat{D}_t)^2} \tag{4.4}$$

$$\text{MAPE} = N^{-1} \sum_{i=1}^{N} |D_t - \hat{D}_t| / |D_t|. \tag{4.5}$$

First, we compare the performance of the SDM-AR(1) and SDM-AR(2) models in the following table.

Series	model	MAE(%)	RMSE(%)	MAPE(%)
All	SDM-AR(2)	0.0802	0.1163	2.65
	SDM-AR(1)	0.2123	0.2913	5.88
OC	SDM-AR(2)	0.0854	0.1155	2.93
	SDM-AR(1)	0.0903	0.1208	3.13
SRE	SDM-AR(2)	0.1512	0.2258	3.36
	SDM-AR(1)	0.2924	0.4132	6.48
CRE	SDM-AR(2)	0.1731	0.2424	5.32
	SDM-AR(1)	0.3676	0.5138	9.59

Table 4.5: In-sample prediction error comparison between SDM-AR(1) and SDM-AR(2).

Based on the results presented in Table 4.5, we can quickly tell that the SDM-AR(2) model shows obvious advantages. In particular, when we look at the All series, the SDM-AR(2) model reduces those three accuracy measures by half. This result strongly suggests that the second-order lag term has a significant role in default rate prediction.

Series	model	MAE(%)	RMSE(%)	MAPE(%)
All	SDM-AR(2)	0.0802	0.1163	2.65
	AR(2)	0.0861	0.1342	2.73
OC	SDM-AR(2)	0.0854	0.1155	2.93
	AR(2)	0.0828	0.1143	2.84
SRE	SDM-AR(2)	0.1512	0.2258	3.36
	AR(2)	0.1684	0.2423	3.56
CRE	SDM-AR(2)	0.1731	0.2424	5.32
	AR2	0.1887	0.2497	5.41

Table 4.6: In-sample prediction error comparison between SDM-AR(2) and AR(2).

Now, we demonstrate the importance of state-dependence by comparing SDM-AR(2) with AR(2). From Table 4.6, the SDM-AR(2) model has a slight advantage when compared with the AR(2) model. Although the improvement brought by changing from AR(2) to SDM-AR(2) is not as significant as the one brought by changing from SDM-AR(1) to SDM-AR(2), the importance of different regimes is still non-negligible.

It is also important to notice that, in this section, we have estimated the models to the entire data set and used the result to make a prediction. This is unrealistic in the real world. In order to better measure the forecasting ability of the models, Sections 4.3, 4.4, and 4.5 consider out-of-sample performance. But before moving on, we consider the serial correlation of the data, which is another important diagnostic tool. Since we already pointed out in Section 3.4.3 that the first-order-lag models have obvious disadvantages in capturing serial correlation, we only make a comparison between the second-order-lag models in the next section.

4.2.2 Auto- and partial auto-correlation comparison between SDM-AR(2) and AR(2)

In this section, we study the performance of each model in capturing the autocorrelation of the observations $\{Y_t\}$. We compute the sample auto- and partial auto-correlation for each data set. We also calculate the auto-correlation for the SDM-AR(2) model according to the method described in Section 3.6. In addition to the ACF, we also use the simulated data to approximate the partial auto-correlation for the SDM-AR(2) model.

Figure 4.3 shows the sample ACF and PACF for the SRE series. The second and third rows of the figure are the ACF and PACF of the SDM-AR(2) and AR(2) models, with the estimated values presented in Table 4.4.

Here, by inspecting the ACF from each figure, the SDM-AR(2) model shows a distinct advantage, especially in terms of the PACF. We can easily notice that the autocorrelations for the historical data series are negative after the 18th lag term. Although the ACF of the SDM-AR(2) model becomes negative only after the 19th lag term, the ACF of the AR(2) model remains positive even at the 20th lag term. The similarity between the ACF of the data and the AR(2) model is lower than the one for the SDM-AR(2) model.

Also, from the PACF figures, it is well-known that the AR(2) model always has a zero value for the third-lag term, but the actual historical data series and the SDM-AR(2) model have values less than zero. The non-zero third lag term is an unavoidable drawback of the AR(2) model.

In conclusion, the SDM-AR(2) model can capture the dependence structure of historical data series much better than the AR(2) model.

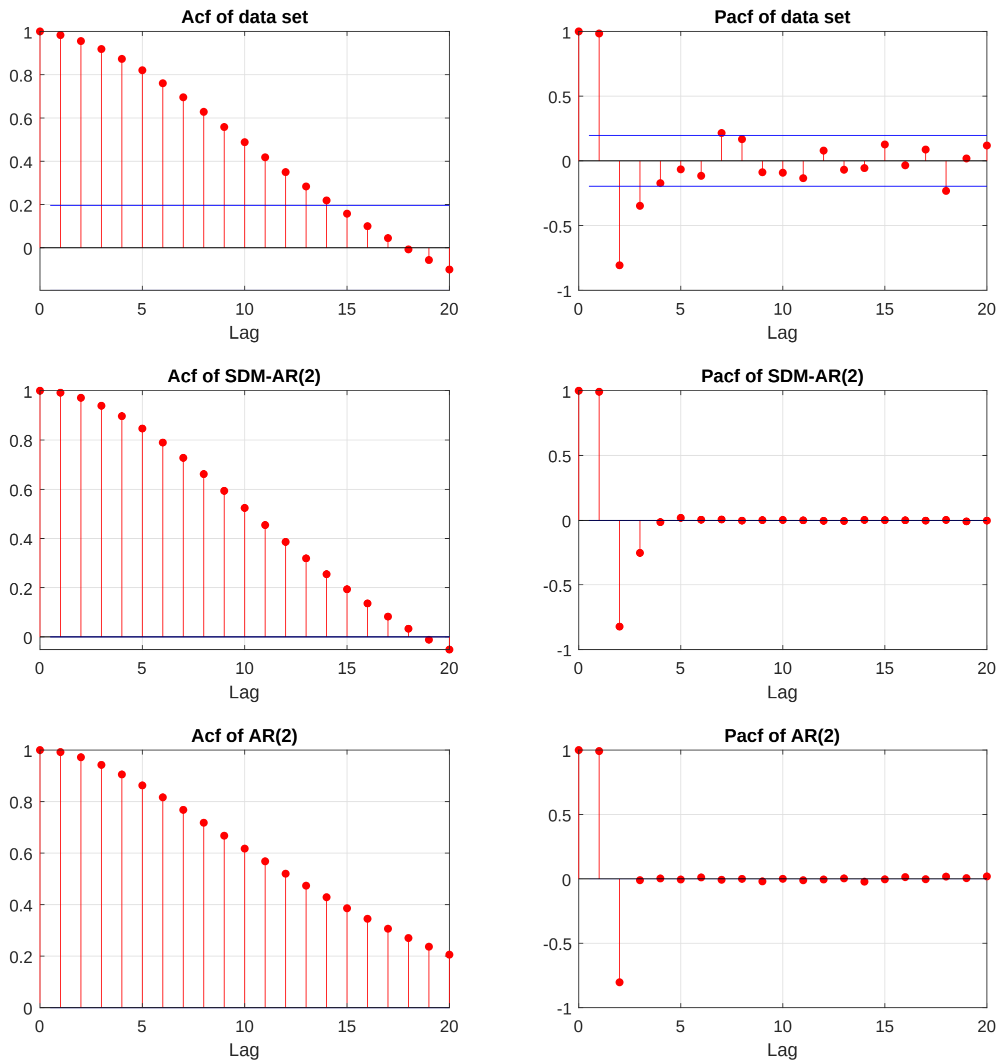


Figure 4.3: Sample ACF and PACF plots the of SRE Series from historical data and different models.

4.3 Out-of-sample prediction

In this section, we divide the historical data points of each series into two non-overlapping segments. The first part is used to estimate the model parameters, and the subsequent segment is used for evaluating the forecasting ability of the models. Let n_1 denote the number of points in the first segment and n_2 be the number of points in the second segment, so that $n_1 + n_2 = 104$. We consider both $n_1 = 60$, in which case the last point in the estimation window is the last quarter of 2005 (one year before the onset of the financial crisis), and $n_1 = 94$, in which case the estimation window is longer, and parameter estimates are ostensibly more reliable. Crucially, the shorter window allows us to see how the model would have performed during the crisis.

4.3.1 Classic AR models estimation results

We first estimate the classic AR(1) and AR(2) models with those two training segments. The following tables show the out-of-sample estimation result for the AR(1) and AR(2) models.

Series	Estimation Period	Implied a	θ_1	σ_ϵ^2 of AR	C of AR
All	$n_1 = 60$	0.009	0.975	0.000422	-0.0577
	$n_1 = 94$	0.152	0.979	0.000982	-0.0425
OC	$n_1 = 60$	0.211	0.998	0.000186	-0.00483
	$n_1 = 94$	0.117	0.990	0.000278	-0.0219
SRE	$n_1 = 60$	0.101	0.980	0.000404	-0.0503
	$n_1 = 94$	0.222	0.987	0.001341	-0.0254
CRE	$n_1 = 60$	0.188	0.987	0.000965	0.0451
	$n_1 = 94$	0.234	0.980	0.000229	-0.0471

Table 4.7: AR(1) model's estimation results based on $n_1 = 60\&94$ data points.

Series	Estimation Period	Implied a	θ_1	θ_2	σ_ϵ^2 of AR	C of AR
All	$n_1 = 60$	0.108	1.644	-0.670	0.000202	-0.052
	$n_1 = 94$	0.196	1.824	-0.846	0.000268	-0.0417
OC	$n_1 = 60$	0.056	1.467	-0.667	0.000406	-0.36797
	$n_1 = 94$	0.073	1.334	-0.369	0.000232	-0.0679
SRE	$n_1 = 60$	0.121	1.462	-0.482	0.000293	0.38799
	$n_1 = 94$	0.286	1.791	-0.804	0.000464	-0.2435
CRE	$n_1 = 60$	0.211	1.329	-0.342	0.000831	-0.0392
	$n_1 = 94$	0.319	1.726	-0.743	0.000978	-0.0338

Table 4.8: AR(2) model's estimation results based on $n_1 = 60\&94$ data points.

We can see from these results that the implied factor loadings of both models increase when we are extending the training segment to the end of the financial crisis. This is to be expected, since market correlations should be higher when the market is at a bearish level. Also, the variances of the noise term increase as well.

4.3.2 SDM-AR models estimation results

We also estimate the SDM-AR(1) and SDM-AR(2) models by using the method presented in Section 3.3. For a more direct comparison, the following two tables display some of the SDM-AR(1) and SDM-AR(2) models' estimation results based on 60 and 94 data points.

Series	Estimation Period	a_1	a_2	β	t	θ_1	σ_ϵ^2 of AR
All	$n_1 = 60$	0.387	0.189	1	-0.536	0.996	0.007984
	$n_1 = 94$	0.224	0.130	1	-0.615	0.982	0.035676
OC	$n_1 = 60$	0.225	0.169	-1	0.124	0.998	0.003996
	$n_1 = 94$	0.136	0.097	1	-0.563	0.989	0.021879
SRE	$n_1 = 60$	0.362	0.181	1	-0.673	0.995	0.010266
	$n_1 = 94$	0.282	0.265	1	1.012	0.991	0.017656
CRE	$n_1 = 60$	0.436	0.422	1	0.741	0.996	0.000831
	$n_1 = 94$	0.253	0.242	1	1.865	0.981	0.037237

Table 4.9: SDM-AR(1) model's estimation results based on $n_1 = 60\&94$ data points.

Series	Estimation Period	a_1	a_2	β	t	θ_1	θ_2	σ_ϵ^2 of AR
All	$n_1 = 60$	0.177	0.143	1	0.693	1.831	-0.848	0.005144
	$n_1 = 94$	0.235	0.162	1	-0.672	1.829	-0.851	0.006517
OC	$n_1 = 60$	0.420	0.386	-1	0.0298	1.176	-0.177	0.001645
	$n_1 = 94$	0.117	0.0839	1	-0.594	1.411	-0.433	0.024756
SRE	$n_1 = 60$	0.388	0.369	-0.351	1.121	1.861	-0.864	0.00096
	$n_1 = 94$	0.267	0.255	-0.349	0.645	1.881	-0.894	0.002873
CRE	$n_1 = 60$	0.930	0.928	-0.942	0.858	1.816	-0.816	$6.08e^{-10}$
	$n_1 = 94$	0.349	0.332	-1	-0.946	1.759	-0.775	0.006748

Table 4.10: SDM-AR(2) model's estimation results based on $n_1 = 60\&94$ data points.

The "σ_ϵ^2 of AR" term for the SDM-AR(2) and SDM-AR(1) models is the variance of the driving noise term for the process $\{M_t\}$. We can tell from both Tables 4.9 and 4.10 that the model parameters do suggest a change of market behavior, since the values of a_1, a_2 and t have changed. For both models, the variance of the driving noise term for the process $\{M_t\}$ increased dramatically. This is consistent with the fact we observed from the classic AR models.

The unforeseen phenomenon is the decrement of the factor loadings, a_1 and a_2. Theoretically, by including the financial crisis period, we expect the factor loadings to increase due to our original assumption that market correlation is higher in the bearish market. But the factor loadings of the SDM-AR(1) and SDM-AR(2) models changed in the opposite way compared with the ones of classic AR models. This may be caused by the increment of the variance of $\{M_t\}$'s driving noise.

4.4 Out-of-sample prediction (point estimates)

In this section, we discuss the out-of-sample point forecasting abilities of the four models, AR(1)&(2) and SDM-AR(1)&(2). The models are estimated with the two training data segments we used in Section 4.3. We use the parameters displayed in Section 4.3 to perform such predictions. By doing so, we are able to directly compare the prediction performance of different models based on pre- and post- financial crisis periods.

The prediction is made based on all available observations up to the current position. For example, the prediction we make at time t is evaluated based on

$$\hat{d}_{t+h} = E[\Phi(Y_{t+h})|Y_{1:t}],$$

where h is the length of the time interval we want to predict. In this section, we set $h = 1$ or 4, which are equivalent to one-quarter and one-year-ahead predictions. But the parameters we used to make such predictions remain the same as the value we obtained from the training segment with length n_1.

Then the prediction errors can be calculated ($\epsilon_t = d_t - \hat{d}_t$) for each time t after the training segment. Again, the same accuracy measures, MAE, RMSE and MAPE defined in Section 4.2.1 are evaluated in this section in order to quantify the prediction performance of each model.

4.4.1 Predicting one quarter ahead

In this subsection, we focus on the one-quarter-ahead prediction made by those four different models based on two different training segments. Tables 4.11 and 4.12 show the prediction accuracy results based on $n_1 = 60$ and 94, respectively. The results of the remaining series can be found in the Appendix.

Series	Model	MAE (%)	RMSE(%)	MAPE(%)
All	SDM-AR(2)	0.1107	0.1691	2.938
	AR(2)	0.1159	0.1846	2.905
	SDM-AR(1)	0.2723	0.3707	6.483
	AR(1)	0.2454	0.3984	5.761
OC	SDM-AR(2)	0.0914	0.1218	3.436
	AR(2)	0.1483	0.1765	6.295
	SDM-AR(1)	0.0974	0.1290	3.630
	AR(1)	0.1641	0.131	3.685
SRE	SDM-AR(2)	0.2162	0.3052	3.686
	AR(2)	0.2856	0.3992	4.917
	SDM-AR(1)	0.4205	0.5529	7.561
	AR(1)	0.4139	0.6199	7.286
CRE	SDM-AR(2)	0.1835	0.2722	4.782
	AR(2)	0.2686	0.3981	7.023
	SDM-AR(1)	0.4164	0.5404	11.45
	AR(1)	0.3916	0.5783	10.12

Table 4.11: Out-of-sample forecasting Error when $n_1 = 60$

Series	Model	MAE (%)	RMSE(%)	MAPE(%)
All	SDM-AR(2)	0.0375	0.0468	1.5074
	AR(2)	0.0364	0.0453	1.4648
	SDM-AR(1)	0.1216	0.1387	4.6910
	AR(1)	0.1028	0.1209	3.9423
OC	SDM-AR(2)	0.0695	0.0865	3.5412
	AR(2)	0.07	0.0882	3.5641
	SDM-AR(1)	0.0781	0.0930	3.9646
	AR(1)	0.0771	0.0920	3.9168
SRE	SDM-AR(2)	0.061	0.0787	1.4126
	AR(2)	0.0608	0.0821	1.4197
	SDM-AR(1)	0.2774	0.2961	6.4907
	AR(1)	0.2497	0.268	5.8429
CRE	SDM-AR(2)	0.0444	0.0562	4.2262
	AR(2)	0.0391	0.0505	3.7330
	SDM-AR(1)	0.1313	0.1466	12.726
	AR(1)	0.0915	0.1042	8.8034

Table 4.12: Out-of-sample forecasting Error when $n_1 = 94$

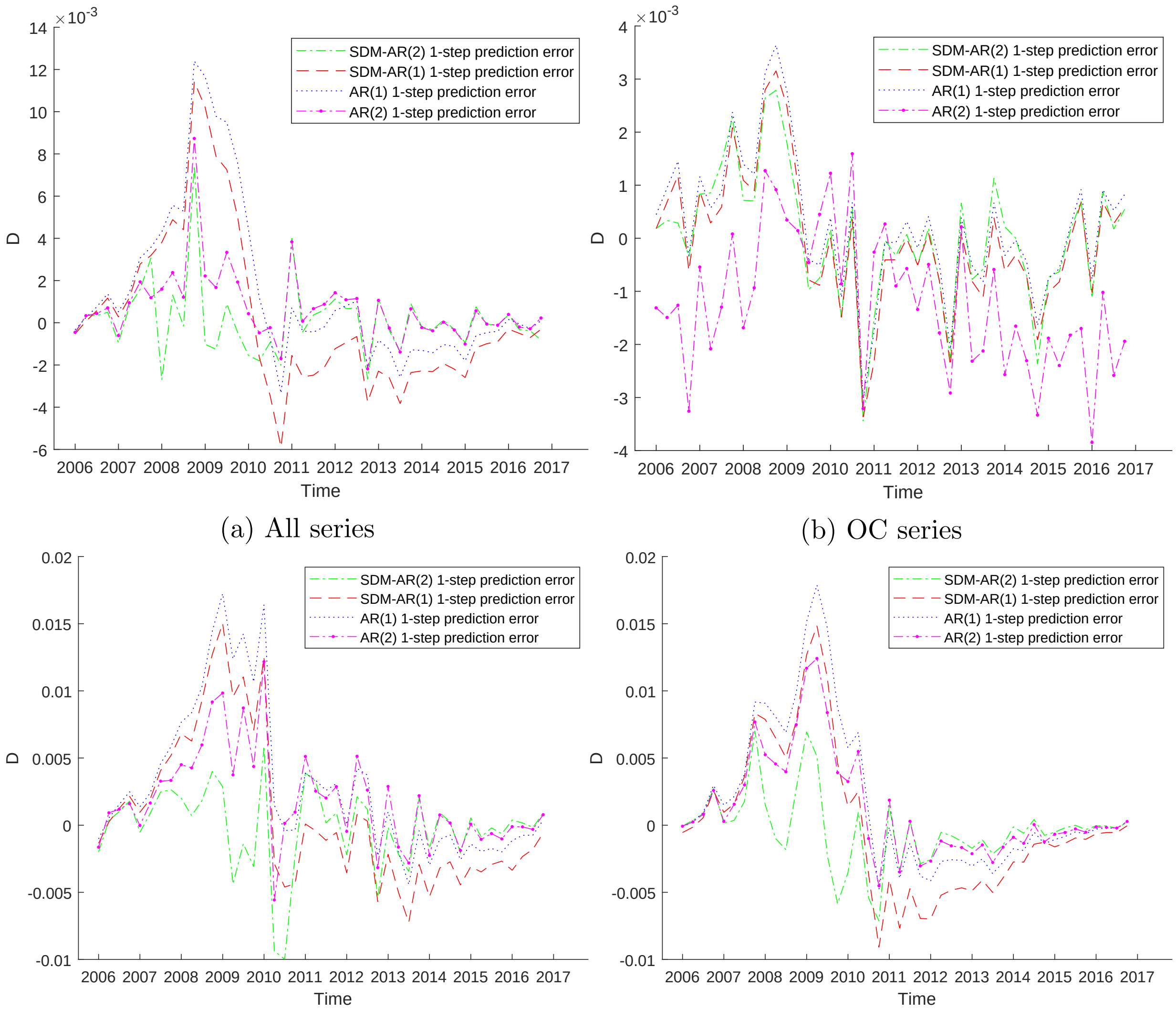
Based on Table 4.11 & 4.12, we can draw the following conclusions:

1. As we can see from both tables, by changing the one-lag structure (AR(1), SDM-AR(1)) to the two-lag structure (AR(2), SDM-AR(2)), most of the error measurements are reduced by 50% at least. This fact suggests that the two-lag structure plays an important role from the perspective of prediction.

2. When we only look at Table 4.11, there appears to be an evidence that state dependence structure (SDM-AR(1), SDM-AR(2)) is also essential. For example, we can notice that MAE of OC series is reduced dramatically by switching the model from classic AR models to the state-dependence models.

3. Overall, the SDM-AR(2) model presents clear advantages compared with other models. This is to be expected, since the other models can be written as a special case of the SDM-AR(2) model, which we have proved in Section 3.8.

These phenomena can also be found in the series we present in the Appendix. For a clear and direct comparison, we also show some forecasting error series of those four models. Due to the fact that there are only 10 points left as testing segments for the post-financial crisis training segments, we left the forecasting error series plots for $n_1 = 94$ in the Appendix and display some plots for $n_1 = 60$ in here.



(a) All series (b) OC series

(c) SRE series (d) CRE series

Figure 4.4: Out-sample one-step ahead predictive error based on $n_1 = 60$ training segment

First, it is easy to notice from most of those plots that when the financial crisis is not included in the training segment, three of the models systematically and consistently under-predict the default rates during the financial crisis (i.e. 2007 to 2010), whereas the SDM-AR(2) model does not. This conclusion is supported by the fact that the residual series produced by SDM-AR(1), AR(2) and AR(1) models are always greater than 0 during the financial crisis period, while the one of the SDM-AR2 model is more similar to white noise, which move up and down around 0 randomly.

When we look at the performance of the models after the crisis, the SDM-AR(2) and AR(2) models have similar accuracy. Those two models also show slight advantages when compared to SDM-AR(1) and AR(1). This result is consistent with the conclusion we drew from the accuracy measurement tables that the two-lag term structure is relatively more important for the post-crisis period. All those phenomena can also be found in most of the other series.

We have also checked the ACF plots for the predictive error series ($\epsilon_t = d_t - \hat{d}_t$) for each model. We only check the error ACF plots of the models estimated with the pre-crisis training segment.

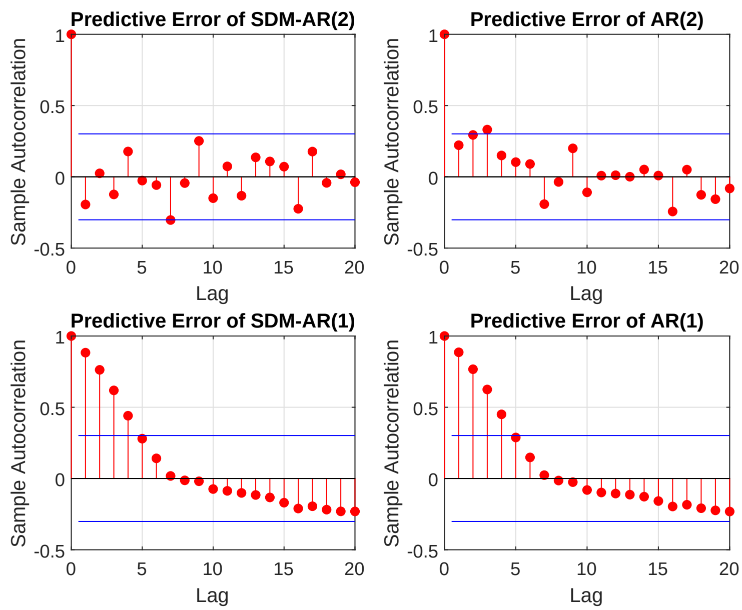


Figure 4.5: ACF plots of the prediction error series for the four models estimated with pre-crisis training segment - All Series

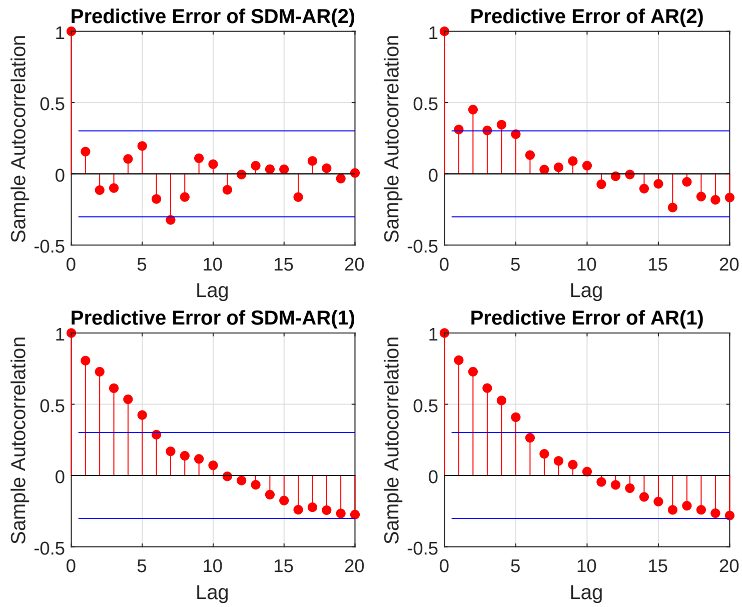


Figure 4.6: ACF plots of the prediction error series for the four models estimated with pre-crisis training segment - SRE Series

The blue horizontal lines represent autocorrelation confidence bounds consisting of 2 standard errors. From the perspective of predictive error series, the autocorrelations should be zero for the error series if the model fits the data well. But as we can see from the following plots, it is clear that one-lag structures violate this assumption. There also exists some terms exceeding the confidence bounds in the plot of the AR(2) model. The SDM-AR(2) model is the only model that can generate the predictive error series having zero autocorrelations.

4.4.2 Predicting four quarters ahead

Sometimes, the risk analyst may wish to forecast the default rate more than one quarter ahead. So, in this section, we will consider the four-step-ahead prediction, which is equivalent to one-year-ahead prediction. Since we are performing four-quarter-ahead prediction, we will only focus on the pre-crisis training segment. Otherwise, if we use 94 points to estimate the model, we will only have 6 points left for testing the forecasting accuracy, which is way too few to judge the performance of the model. There exist some numerical issues when calculating the four-quarter-ahead prediction of the OC series, so we exclude it in here.

Series	Model	MAE (%)	RMSE(%)	MAPE(%)
All	SDM-AR(2)	0.494	0.744	10.632
	AR(2)	0.645	1.027	13.390
	SDM-AR(1)	1.099	1.395	24.863
	AR(1)	0.896	1.433	19.233
SRE	SDM-AR(2)	0.873	1.149	13.743
	AR(2)	1.351	1.906	20.058
	SDM-AR(1)	1.612	2.039	26.651
	AR(1)	1.560	2.213	23.903
CRE	SDM-AR(2)	1.030	1.373	24.729
	AR(2)	1.325	1.88	32.734
	SDM-AR(1)	1.715	2.061	48.035
	AR(1)	1.501	2.085	36.900

Table 4.13: Four-quarter-ahead out-of-sample prediction results based on $n_1 = 60$ training observations.

The Table 4.13 shows the accuracy measurements. The SDM-AR(2) model shows more significant advantages in four-quarter-ahead prediction than it does in one-quarter-ahead. It is clear that all the accuracy measurements are improved dramatically by using the SDM-AR(2) model. But if we compare the SDM-AR(1) with AR(1) models, the result goes in the opposite way as we expect. The state dependence does not improve the prediction accuracy. Also, only changing the model from AR(1) to AR(2) does not bring us a significant improvement. This result suggests that when performing four-quarter-ahead prediction, either state-dependence or two-lag structure alone is not sufficient. We need both of them in order to forecast the market behavior accurately. For the remaining series we left in the Appendix, we can find that the SDM-AR(2) model provides at least the same accuracy as the other models do.

The following figure provides a close look at the prediction error series of the All series.

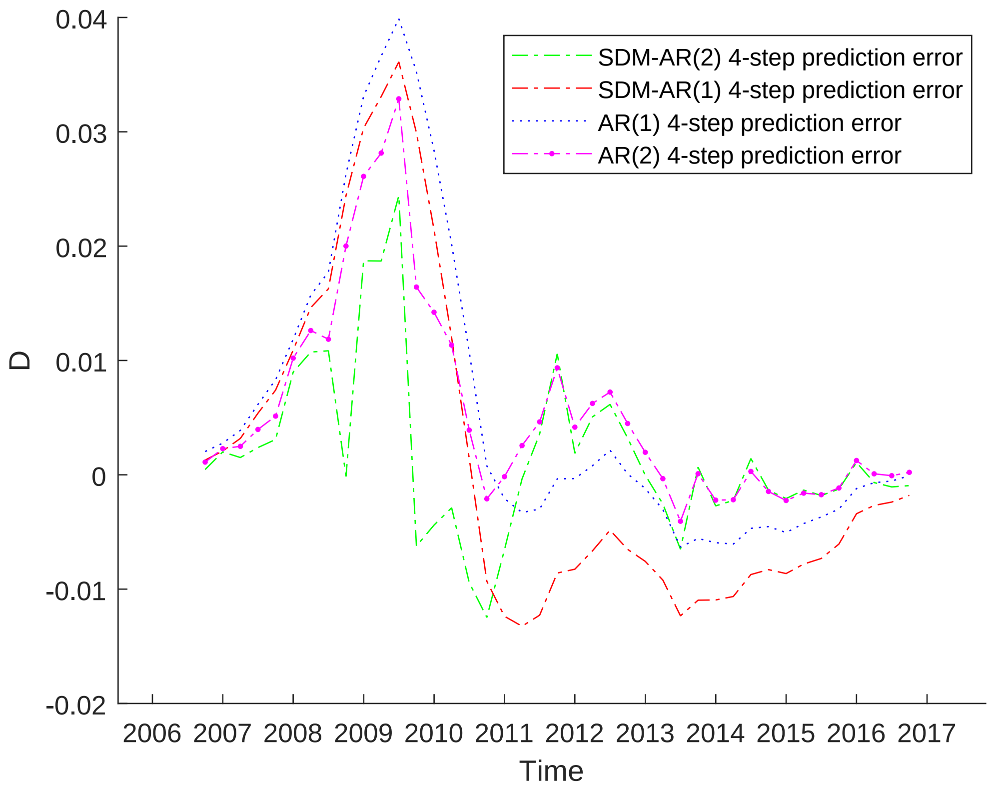


Figure 4.7: Four-step ahead prediction error series for All series when $n_1 = 60$

Although, it is less obvious that the models except SDM-AR(2) systematically under-predict the default rate during the financial crisis, the SDM-AR(2) is still able to show slight advantages. As we can see from Figure 4.7, the errors of SDM-AR(2) are closer to zero than the others. This is also supported by the accuracy measurements we have calculated in the previous table. Based on the remaining plots we have relegated to the Appendix, we are able to conclude that the SDM-AR(2) model can provide at least the same accuracy as the other three models do.

4.5 Out-of-sample prediction (interval estimates)

Besides the accurate point estimates, which is an important factor when choosing a model, interval estimates are more important than point estimates from a risk management perspective. Especially, after the financial crisis, banks put more efforts in developing models to foresee potential huge loss.

In this section we will focus on the one- and four-quarter-ahead one-sided 99.9% upper prediction confidence intervals based on the four fitted models. In order to quantify the performance of those intervals, we will count the number of points which fall outside the confidence interval, with the interpretation that the more points fall outside the interval, the worse the model is in capturing the potential risk for the next time period.

The out-of-sample predictive quantiles are calculated by conditioning on all the available observations up to the current position. For example, the h-quarter-ahead quantile at time t, $\hat{D}_{t+h,99.9\%}$ is $\Phi(x)$, where x is the solution of the following equation:

$$\int_{-\infty}^{x} f_{Y_{t+h}|Y_{1:t}}(z|y_{1:t})dz = 99.9\%,$$

where the integrand is the K-step-ahead prediction density for the observation defined by Theorem 4 and 12 for the SDM-AR(1) and SDM-AR(2) models, respectively. Then we iteratively solve this equation for each time after the training segment. The solutions form a curve, which presents how the 99.9% quantile evolves along the time. In this section, we focus on the one- and four-quarter-ahead interval prediction.

4.5.1 Interval predicting one quarter ahead

The following table presents the numbers of points that fall outside the 99.9% confidence intervals calculated by the above equations based on the models estimated with the pre-crisis training segment. The column n^* represents the number of points that fall outside the 99.9% CI.

Series	Model	n^*	Percentage
All	SDM-AR(2)	1	2.27%
	AR(2)	1	2.27%
	SDM-AR(1)	4	9.09%
	AR(1)	9	20.45%
OC	SDM-AR(2)	0	0%
	AR(2)	0	0%
	SDM-AR(1)	0	0%
	AR(1)	2	4.55%
SRE	SDM-AR(2)	0	0%
	AR(2)	6	13.63%
	SDM-AR(1)	3	6.81%
	AR(1)	10	22.72%
CRE	SDM-AR(2)	1	2.27%
	AR(2)	2	4.55%
	SDM-AR(1)	1	2.27%
	AR(1)	7	15.91%

Table 4.14: The numbers of points that fall outside of the 99.9% one-side confidence interval when $n_1 = 60$. The number of points in the testing set is 44. The percentage column presents the percentage of the points out of total test points.

Based on Table 4.14, we can easily notice several pieces of evidence to support the idea that the SDM-AR(2) model shows strong advantages compared with other models.

1. Changing from the one-lag structure (AR(1) and SDM-AR(1)) to the two-lag structure (AR(2) and SDM-AR(2)) brings a significant improvement to the interval estimates.

2. The state dependence also plays an important role in the one-quarter-ahead interval prediction. It is obvious that all the state-dependence models reduce the number of points that breach the interval by at least 50%.

3. The SDM-AR(2) model clearly outperforms the other models by having the least number of points breached the intervals.

The following figure presents the evolution of the model-implied 99.9% one-quarter-ahead quantile from the four models estimated with the pre-crisis training segment, along with the historical data. The star markers in the plot represent the points at which the historical data breached corresponding model's confidence interval. We only display the All series in here and relegated the rest to the Appendix.

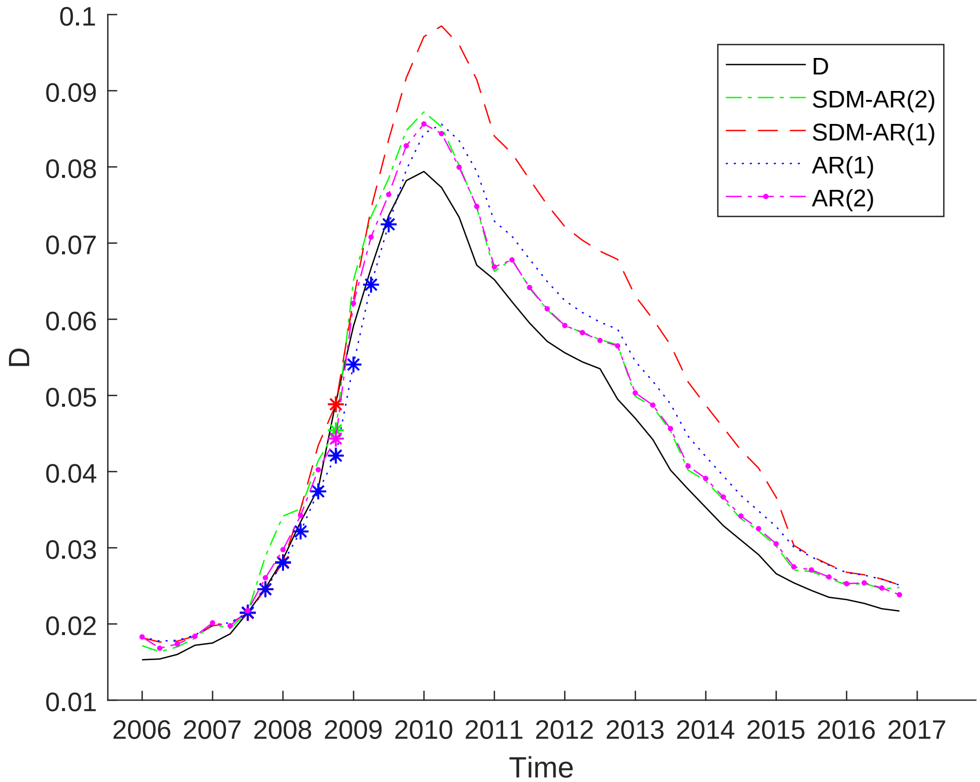


Figure 4.8: One-quarter-ahead prediction CI with 60 points estimation

It is hard to directly compare how the models behave during the financial crisis. But, during the recovery period after the financial crisis, the SDM-AR(2) and AR(2) models provide a lower value of CI. This may suggest that the other two models may overestimated the default rate for the next quarter. Especially, in 2010 which is the turning point of the financial crisis, the AR(2) and SDM-AR(2) models are able to lower their prediction for the next quarter faster than the other models. This result implies that the two-lag structure brings the ability to foresee the turning point when the market will change.

In order to have a clear assessment of the ability of the models in capturing the future risk, we zoom in the overall default rate (All series) from the beginning of 2007 to the end of 2011 in Figure 4.9. This is also the period which contains all the historical data which lie outside of the prediction confidence interval.

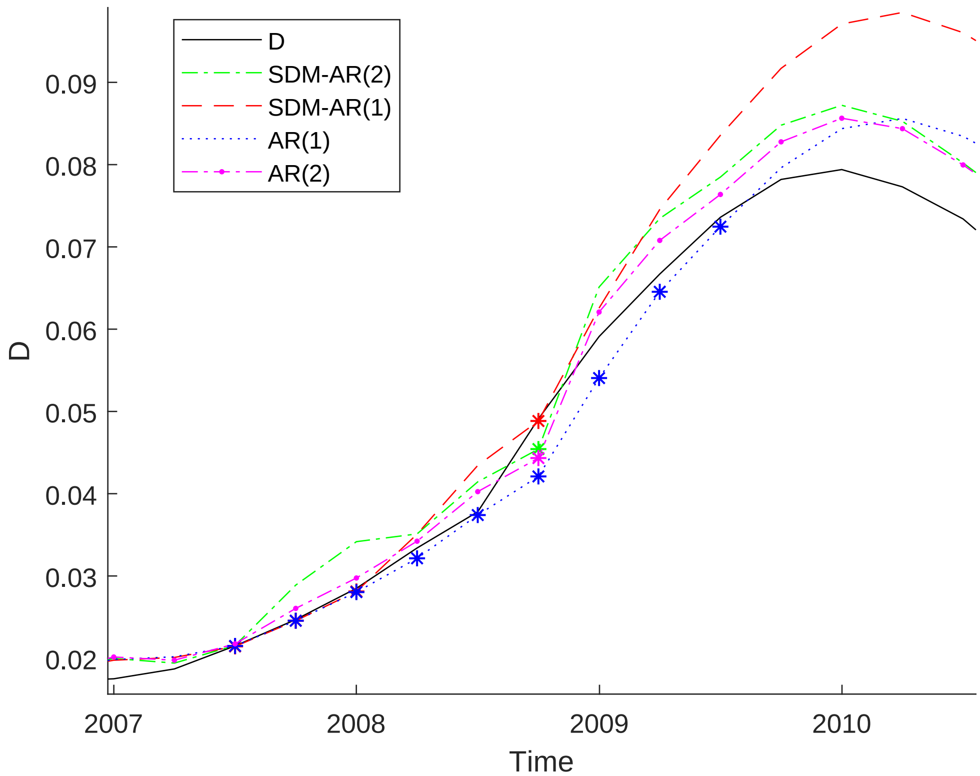


Figure 4.9: Zoom-in plot of the 99.9% CI during the financial crisis for All series generated by four models when $n_1 = 60$.

As we can see from this figure, all the intervals generated by those four models are breached in the third quarter of 2008. This suggests that the financial crisis is an extreme event which occurs with probability less than 0.1%. But this is also the only time when the data breaches the intervals generated by the SDM-AR(2) and AR(2) models.

Based on the figure, changing the model from AR(1) to AR(2) enables the prediction interval to cover the majority of the points. This result is consistent with the conclusion we drew from point estimates. Bringing the state-dependence to the AR(1) model also improves the result, but the improvement of changing AR(2) to SDM-AR(2) model is not substantial for All series.

In order to verify the importance of the state-dependence, we also zoom on the SRE series during the financial crisis in Figure 4.10.

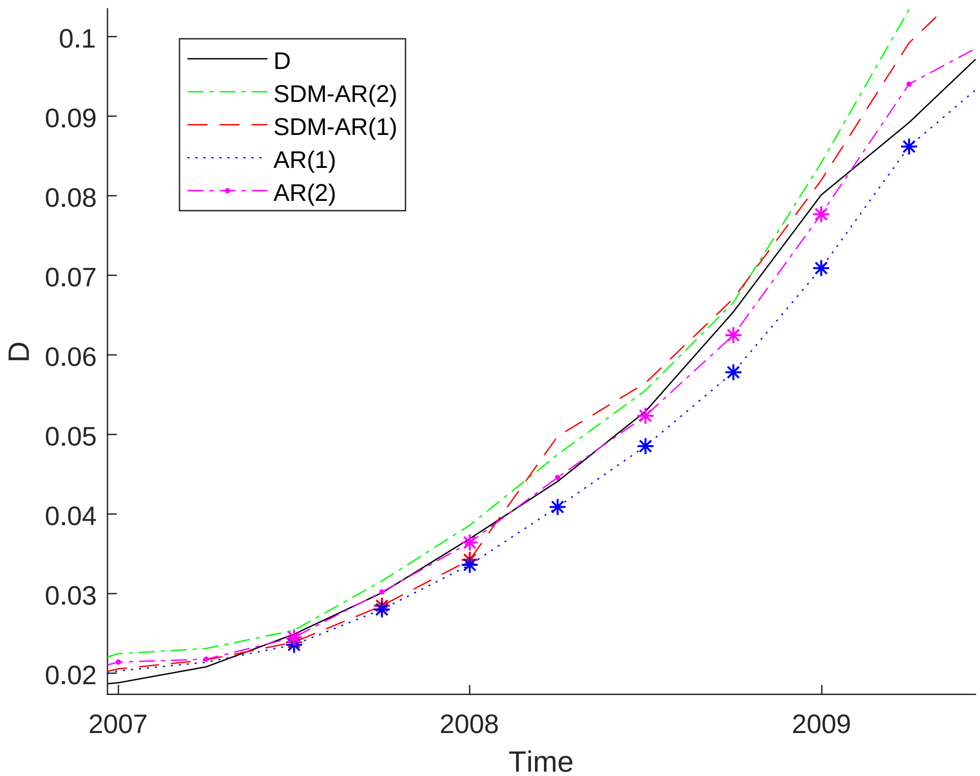


Figure 4.10: Zoom-in plot of the 99.9% CI during the financial crisis for SRE series generated by four models when $n_1 = 60$.

The improvment of introducing the state-dependence can be easily found in the middle of 2007 where the historical data breached all the confidence intervals of the other three models except the one of the SDM-AR(2) model. The same situation happens again at the beginning of 2008. This is a strong evidence to support the idea that using neither two-lag structure nor state-dependence alone is enough to predict the market behavior. We need both of them together in order to have an accurate result.

4.5.2 Interval predicting four-quarters ahead

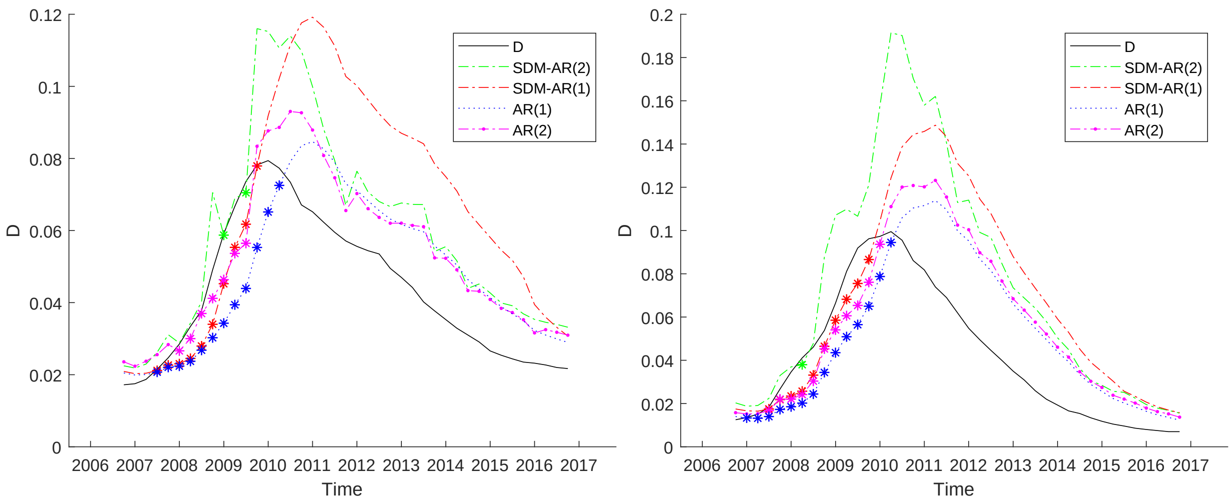
Agian, we also demonstrate the four-quarter-ahead interval prediction in this section. The same methodology as in the previous section is applied in here. The following table presents the numbers of points falling outside the 99.9% confidence intervals with pre-crisis training segment. As before, the column n^* represents the number of points falling outside the 99.9% CI.

Series	Model	n*	Percentage
All	SDM-AR(2)	2	5%
	AR(2)	7	17.5%
	SDM-AR(1)	10	25%
	AR(1)	12	30%
SRE	SDM-AR(2)	1	2.5%
	AR(2)	12	30%
	SDM-AR(1)	11	27.5%
	AR(1)	15	37.5%
CRE	SDM-AR(2)	1	2.5%
	AR(2)	11	27.5%
	SDM-AR(1)	10	25%
	AR(1)	14	35%

Table 4.15: The number of points that lie out of the 99.9% one-sided confidence interval when $n_1 = 60$. The number of back-testing points is 40. The percentage column presents the percentage of the points out of total test points.

When we focus on the four-quarter-ahead prediction intervals, the advantages of the SDM-AR(2) model are much more obvious. The table clearly reflects the fact that introducing the two-lag structures helps the model to accurately foresee more potential risks. The state-dependence is another important factor we need to include in the model we use. As we can see from the table, by introducing the state-dependence to the AR(2) model, we dramatically reduce the number of points breached the interval.

The following figures provide a direct comparison of the four-quarter-ahead prediction intervals from those four models estimated with the pre-crisis training segment. The star markers in the plot represent the points at which the historical data breached corresponding model's confidence interval. The rest of the plots can be found in the Appendix.



(a) All series (b) CRE series

Figure 4.11: Four-quarter-ahead prediction CI with 60 points estimation

We can easily tell from the figures, the SDM-AR(2) model provides us with the most reliable results. During the financial crisis, the CI of SDM-AR(2) is high enough to cover the majority of default rates in the following time period. It also reacts more rapidly than the other models do after the financial crisis. This is also a piece of evidence to support the idea that both the two-lag structure and state-dependence are important factors when we try to predict the default rate.

4.6 Inference about the systematic risk factor $\{M_t\}$

The purpose of this section is to make some inferences about the latent systematic risk factor $\{M_t\}$ based on the SDM-AR(2) model. In the first part of this section, we demonstrate the method we have used to select the best inferential value $\hat{M}_t$ of the underlying process $\{M_t\}$ for each time $t = 1, \ldots, T$. Then, by checking some fundamental characteristics of the series $\{\hat{M}_t\}$, we are able to verify if the AR(2) process defined in Equation 3.23 can adequately model the latent process $\{M_t\}$.

4.6.1 Systematic risk factor filtering

After we have successfully applied the maximum likelihood method for estimating the model parameters, we need to develop a procedure to estimate the current value of the systematic risk factor based on the observations up to time t (the current time). To solve such a filtering problem, we compute the filtering density function of the latent state $f_{M_t|Y_{1:t}}(m_t|y_{1:t})$.

In a classic dynamic linear model, the conditional variable $M_t|Y_{1:t} = y_{1:t}$ is usually continuous. But in the SDM-AR(2) model, we know that the distribution of M_t is discrete once given the value of Y_t and depends on the whole path of $\{Y_{1:t}\}$. As a result, the filtering function of the SDM-AR(2) model is a probability mass function.

According to Theorem 11 in Section 3.7 and the law of total probability, we can calculate the filtering function at time t, $P(M_t = m_t|Y_{1:t} = y_{1:t})$, by:

$$P(M_t = m_t|Y_{1:t} = y_{1:t}) = \sum_{k=1}^{K} P(M_t = m_t, M_{t-1} = M^{-1}(y_{t-1}, k)|Y_{1:t} = y_{1:t}), \tag{4.6}$$

where K is the number of regimes for the market, assumed equal to 2 in this study. The function $M^{-1}(y, k)$ is defined in Equation 3.13.

Then, for each time $t = 1 :, \ldots, T$, we choose $\hat{M}_t$ such that:

$$\hat{M}_t = \arg\max_{m_t} P(M_t = m_t|Y_{1:t} = y_{1:t}). \tag{4.7}$$

By doing so, we obtain an inferential series $\{\hat{M}_t\}$ about the latent process $\{M_t\}$. The following figure shows the series $\{\hat{M}_t\}$ for ALL series.

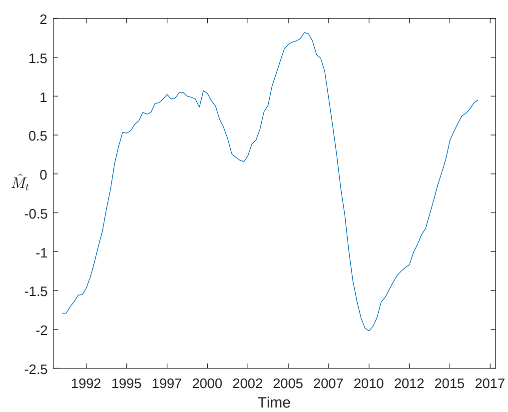


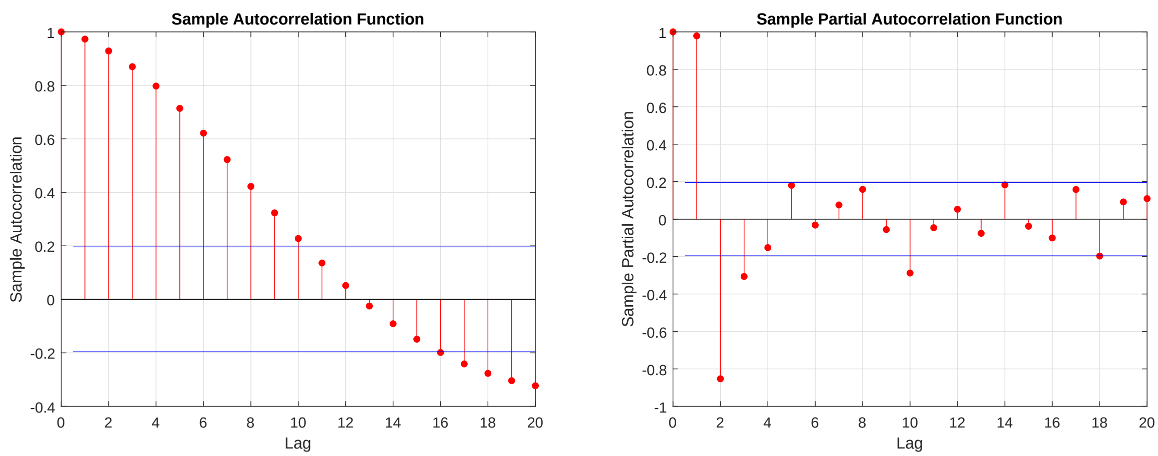
Figure 4.12: The inferential series $\{\hat{M}_t\}$ for ALL series.

4.6.2 AR(2) assumptions diagnostics

In this section, we want to verify if the following assumptions we have made for the latent process $\{M_t\}$ in the SDM-AR(2) model are possibly violated:

- The process $\{M_t\}$ follows the AR(2) process.
- The stationary distribution of $\{M_t\}$ is a standard normal.

By using the inferential series $\{\hat{M}_t\}$ we calculated in Section 4.6.1, we have obtained the following ACF and PACF plots:

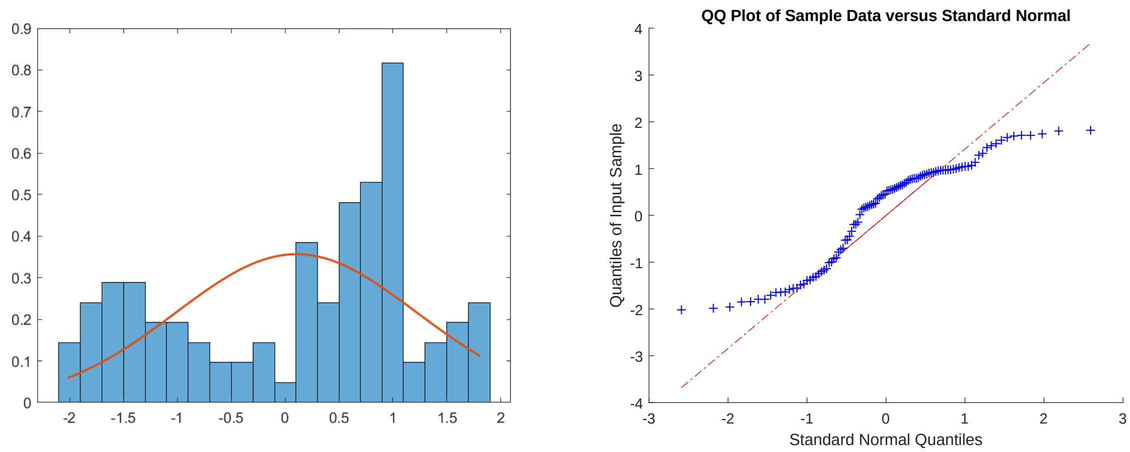


(a) ACF: inferential series $\{\hat{M}_t\}$ (b) PACF: inferential series $\{\hat{M}_t\}$

Figure 4.13: The ACF and PACF of the inferential series $\{\hat{M}_t\}$ for ALL series.

Figure 4.13 provides the ACF and PACF of the inferential series $\{\hat{M}_t\}$ for the ALL series. We only present one series here, but a similar behavior can also be found in other series. According to those two figures, there is no clear evidence to indicate the AR(2) assumption of the process $\{M_t\}$ is violated.

It is more challenging to verify the assumption that the stationary distribution of $\{M_t\}$ is standard normal. Because the Federal Reserve Data we studied in this chapter only contains 104 data points for each series, the number of data points is too small for a slowly mean-reverting series to check its stationary distribution. For example, if we directly check the histogram and QQ-plot of the inferential series $\{\hat{M}_t\}$ in Figure 4.12, the results are deceiving:



(a) Histogram: inferential series $\{\hat{M}_t\}$ (b) QQ-plot: inferential series $\{\hat{M}_t\}$

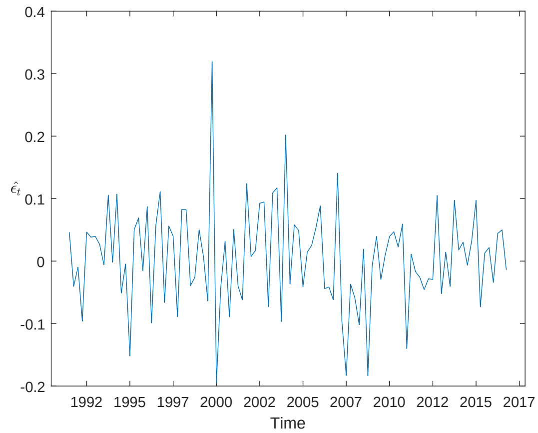
Figure 4.14: The histogram and QQ-plot of the inferential series $\{\hat{M}_t\}$ for ALL series. The red line in the histogram represents the normal density curve with the sample mean and variance.

The red line in Figure 4.14(a) represents the normal density with sample mean and variance of $\{\hat{M}_t\}$. The first thing we noticed is that the sample mean and variance are 0.11 and 1.22, which are close to 0 and 1. But the shape of the histogram is far away from the bell shape, which a normal distribution should have. The QQ-plot also supports that the normality is questionable. The other unexpected feature of these graphs is that the QQ-plot suggests the tail of $\{\hat{M}_t\}$ should be lighter than the standard normal distribution.

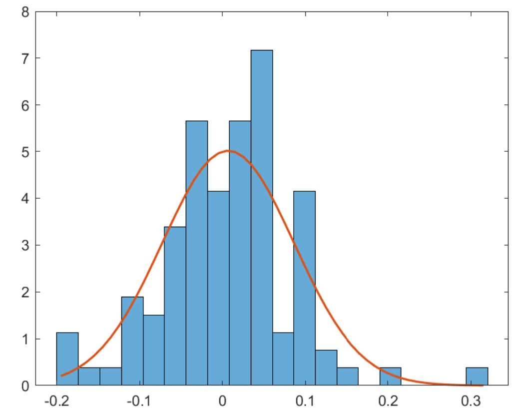
We know that the stationary distribution of an AR(2) process depends on the values of its lag coefficients and variance of the white-noise. So, instead of checking $\{\hat{M}_t\}$ directly, we calculate the residuals of $\{\hat{M}_t\}$ when fitting it to Equation 3.23, and denote them by $\{\hat{\epsilon}_t\}$:

$$\hat{\epsilon_t} = \hat{M_t} - \theta_1 \hat{M_{t-1}} - \theta_2 \hat{M_{t-2}}. \tag{4.8}$$

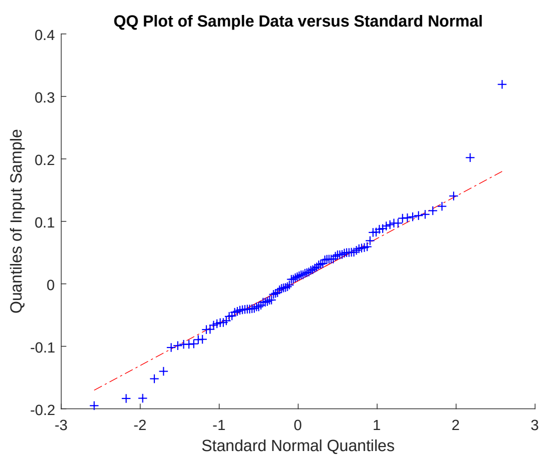
The following figures are corresponded to the four major components used to check the residuals $\{\hat{\epsilon}_t\}$:



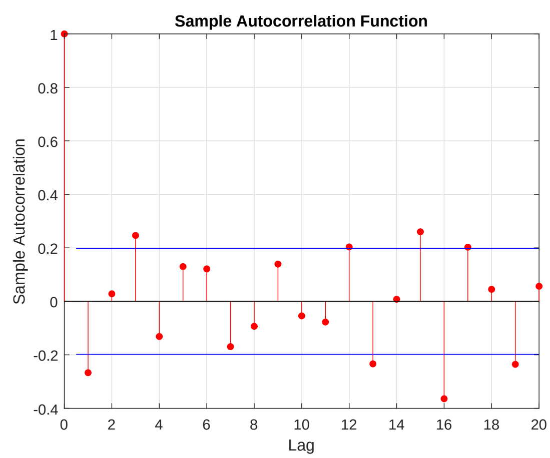
(a) Plot: The fitted residuals $\{\hat{\epsilon}_t\}$



(b) Histogram: The fitted residuals $\{\hat{\epsilon}_t\}$



(c) QQ-plot: The fitted residuals $\{\hat{\epsilon}_t\}$



(d) ACF: The fitted residuals $\{\hat{\epsilon}_t\}$

Figure 4.15: The fitted residuals $\{\hat{\epsilon}_t\}$ of inferential series $\{\hat{M}_t\}$ for ALL series.

As we can easily notice from the above figures, there is no compelling evidence to indicate a violation of the white noise assumptions. The sample mean and variance of $\{\hat{\epsilon}_t\}$ are 0.006 and 0.0063. Theoretically, to retain the standard normal stationary distribution of $\{M_t\}$ given $\theta_1 = 1.832$ and $\theta_2 = -0.853$, the variance of the white noise should be 0.0061, which is close to the sample variance we have. We also perform the Kolmogorov–Smirnov test on $\{\hat{\epsilon}_t\}$ to verify the normality assumption.

Series	KS test P-value
ALL	0.5741
OC	0.4765
SRE	0.5394
CRE	0.1995

Table 4.16: The Kolmogorov–Smirnov test

So, this result does imply that the stationary distribution of $\{\hat{M}_t\}$ is not inconsistent with the standard normal distribution.

We realize from the QQ-plot in Figure 4.15 that the distribution of $\{\hat{\epsilon}_t\}$ may have a heavier tail than the normal distribution. This phenomenon can also be found in most of the other series. As a result, we may suspect that using the model driven by a normal distribution to describe the default rate may underestimate the probability of the extreme event. In the next chapter, we propose a new model to replace the underlying driven distribution of the SDM-AR(2) model from a normal to t-distribution, which is well-known for its heavier tails.

Chapter 5

t-distributed Correlated State-Dependent Model

The "aftermath" of the 2008 financial crisis warned practitioners and academics about the shortcomings of the conventional models whose dependence structure is provided via the Gaussian copula. Under the Gaussian copula model, risk obligator defaults become asymptotically independent when the market goes far enough into the worst scenario. We can explain this phenomenon by using the concept of tail dependence. The main disadvantage of the Gaussian copula is that its tail dependence equals zero. This means that the Gaussian copula underestimates the default clustering that the market may face. In order to overcome this drawback, researchers like Schloegl and O'Kane (2005), and Pimbley (2018) used a similar method based on a chi-square random variable to propose an extension of the Vasicek model based on the t-distribution that preserves its analytical convenience but also provides a broader (fat tails) distribution of credit losses.

Let us recall that the credit score defined in the traditional Vasicek model is in the following form:

$$X_i = aM + \sqrt{1-a^2}\epsilon_i.$$

The variables M and ϵ_i are independent standard normal random variates. The component M is common to all obligors, while ϵ_i is the idiosyncratic term. The constant a is the factor loading that controls the level of market correlation. The model proposed by Schloegl, O'Kane (2005) and Pimbley (2018) similarly modifies the credit score by introducing another common factor S_ν in the following way:

$$X_i = S_\nu \left(aM + \sqrt{1-a^2}\epsilon_i\right) \tag{5.1}$$

$$S_\nu = \sqrt{\frac{\nu}{W}}, \tag{5.2}$$

where W is an independent chi-squared random variable with ν degrees of freedom. Hence, X_i is now changed to having a t-distribution with degrees of freedom ν. This model has been proved to be able to provide fat tails by adjusting the degrees of freedom ν. So, we use the name of "tail thickness factor" when referring to the new factor S_ν.

In this chapter, we adopt such a modification of the traditional Vasicek model to the SDM-AR(2) model we propose in Chapter 3. We call this new model the correlated state-dependent two-lag autoregressive model with a t-distribution systematic risk factor, abbreviated as t-SDM-AR(2).

When defining a model of the form 5.1, we need to pay some attention to the fact that the systematic risk factor M in Equation 5.1 is assumed to be uncorrelated through time. Based on our findings in Chapter 3, the systematic risk factor should exhibit some serial correlations. This difference raises the question of the effect of the tail thickness factor S_ν on the serial correlation.

To answer this question, in subsequent sections , we closely study the effect of including the tail thickness factor S_ν in the context of three possible extensions of the SDM-AR(2) model. One possible way is given by Equation 5.1. Both the systematic risk factor and the idiosyncratic term are affected by S_ν. The other possible alternative models are of the form:

$$X_i = aS_\nu M + \sqrt{1-a^2}\epsilon_i \tag{5.3}$$

$$X_i = aM + \sqrt{1-a^2}S_\nu\epsilon_i. \tag{5.4}$$

As we can tell from the above equations, the tail thickness factor S_ν only affects either the systematic risk factor or the idiosyncratic term. By comparing the formulas of the transformed default rate based on those three models, we will be able to get a clear picture of the effect of the tail thickness factor.

In the first part of this chapter, we take a closer look at some of the basic properties of the tail thickness factor S_ν. Then, we move on to study the effect brought by S_ν into the SDM-AR(2) model.

In the next part of this chapter, we bring the same modifications as in Equations 5.1, 5.3 and 5.4 to the SDM-AR(2) model. However, we also argue that these modifications change the variances of the systematic risk and the idiosyncratic factor, which make the effect of different distributions and variance indistinguishable. So, we introduce another scalar factor to force the variance to be at the same level as in the SDM-AR(2) model.

After that, we extend the methodology we have developed in Section 3.7 to calculate the likelihood function of the t-SDM-AR(2) models and perform some simulation tests to check the impact of the new parameter ν. In the last part of this chapter, we estimate the t-SDM-AR(2) models using historical data and compare the results with those for the SDM-AR(2) model.

The final results show no clear evidence that the t-SDM-AR(2) model fits the data better the SDM-AR(2) model does. However, the proposed t-SDM-AR(2) model can be treated as an extension of the SDM-AR(2). Also, the t-SDM-AR(2) model allows us to change the conservative degree of the predictive confidence interval by considering different values of ν.

5.1 Properties of the tail thickness factor S_ν

Before we study the new models, we take a closer look at some basic properties of S_ν. First, we calculate its expectation:

$$E(S_\nu) = \int_0^\infty \sqrt{\frac{\nu}{x}} f_\nu(x)\ dx,$$

where $f_\nu(x)$ is the probability density function of the chi-square distribution with ν degrees of freedom. We have,

$$\begin{aligned} E(S_\nu) &= \int_0^\infty \frac{\sqrt{\nu}}{\sqrt{x}} \frac{x^{\nu/2-1} e^{-x/2}}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\ dx \\ &= \frac{\sqrt{\nu}}{2^{\nu/2}\Gamma(\frac{\nu}{2})} 2^{\nu/2-1/2}\Gamma(\frac{\nu-1}{2}) \\ &= \sqrt{\frac{\nu}{2}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}. \end{aligned}$$

According to the Euler's definition of the Gamma function as an infinite product that

$$\lim_{\nu\to\infty} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})\frac{\nu}{2}^{-1/2}} = 1, \tag{5.5}$$

we have that $\lim_{\nu\to\infty} E(S_\nu) = 1$.

Using a similar method, we can calculate the second moment of S_ν:

$$\begin{aligned}
E(S_\nu^2) &= \int_0^\infty \frac{\nu}{x} \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\, dx \\
&= \frac{\nu}{2^{\nu/2}\Gamma(\frac{\nu}{2})} 2^{\nu/2-1}\Gamma(\nu/2-1) \\
&= \frac{\nu\Gamma(\frac{\nu}{2}-1)}{2\Gamma(\frac{\nu}{2})} \\
&= \frac{\nu}{\nu-2}.
\end{aligned}$$

Then, the variance of S_ν is

$$\begin{aligned}
Var(S_\nu) &= E(S_\nu^2) - E(S_\nu)^2 \\
&= \frac{\nu}{\nu-2} - \frac{\nu}{2}\left[\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right]^2.
\end{aligned}$$

By Equation 5.5, we have

$$\lim_{\nu\to\infty} Var(S_\nu) = 0.$$

By combining the facts that $\lim_{\nu\to\infty} E(S_\nu) = 1$ and $\lim_{\nu\to\infty} Var(S_\nu) = 0$, we can infer that as ν increases to infinity, the tail thickness factor S_ν converges to 1 in probability and the credit score X_i goes back to the original Vasicek model.

The density function of S_ν can be derived by finding the distribution function first,

$$P(S_\nu \leqslant s) = P(\sqrt{\frac{\nu}{W}} \leqslant s) = 1 - F_\nu\left(\frac{\nu}{s^2}\right),$$

where $F_\nu()$ is the CDF of the Chi-squared distribution with ν degrees of freedom. By taking the first derivative of $F_\nu()$ with respect to s, we find the PDF of S_ν,

$$f_{S_\nu}(s) = \frac{2\nu}{s^3} f_\nu\left(\frac{\nu}{s^2}\right) \quad \text{for} \quad s \in [0, \infty]. \tag{5.6}$$

5.2 The proposed t-SDM-AR(2) models

Let's first recall the SDM-AR(2) model we have proposed in Section 3.5

$$\begin{aligned} X_{i,t} &= a(T_t)M_t + \sqrt{1-a(T_t)^2}e_{i,t} \\ M_t &= \theta_1 M_{t-1} + \theta_2 M_{t-2} + \epsilon_t \\ T_t &= \beta M_t + \epsilon'_t, \end{aligned}$$

where

$$a(x) = \sum_{k=1}^{K} a_k \cdot \mathbb{1}(t_k < x \leqslant t_{k+1}),$$

and K is the number of market regimes. For $t = 1, \ldots, T$, $e_{i,t}$ and ϵ'_t are independent error terms with mean 0 and variances $\sigma_e^2 = 1$ and $\sigma_{\epsilon'}^2 = 1 - \beta^2$, respectively. But the variance of ϵ_t is

$$\sigma_\epsilon^2 = 1 - \gamma_1\theta_1 - \gamma_2\theta_2, \tag{5.7}$$

where $\gamma_1 = \frac{\theta_1}{1-\theta_2}$ and $\gamma_2 = \theta_1\gamma_1 + \theta_2$.

In this section, we retain all the other components of the SDM-AR(2) model besides the credit score $X_{i,t}$. We replace the formula for the credit score in three different ways:

$$X_{i,t} = \sqrt{\frac{\nu-2}{\nu}} S_{t,\nu} \left[a(T_t)M_t + \sqrt{1-a(T_t)^2}e_{i,t} \right] \tag{5.8}$$

$$X_{i,t} = a(T_t)\sqrt{\frac{\nu-2}{\nu}} S_{t,\nu} M_t + \sqrt{1-a(T_t)^2}e_{i,t} \tag{5.9}$$

$$X_{i,t} = a(T_t)M_t + \sqrt{1-a(T_t)^2}\sqrt{\frac{\nu-2}{\nu}} S_{t,\nu} e_{i,t}, \tag{5.10}$$

where $S_{t,\nu} = \sqrt{\nu/W_t}$ and $W_1, \ldots, W_N$ follow an independent Chi-squared distribution with a degree of freedom ν. We also introduce an additional scale factor $\sqrt{\frac{\nu-2}{\nu}}$ in the equations so that we can keep the variances of the systematic and idiosyncratic terms the same as in the SDM-AR(2) model

$$
\begin{aligned}
Var\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}M_t\right) &= \frac{\nu-2}{\nu}Var\left(S_{t,\nu}M_t\right) \\
&= \frac{\nu-2}{\nu}\left[E(S_{t,\nu}^2M_t^2) - E(S_{t,\nu})^2E(M_t)^2\right] \\
&= \frac{\nu-2}{\nu}E(S_{t,\nu}^2)E(M_t^2) \\
&= 1.
\end{aligned}
$$

The last equation is based on the facts that $E(S_{t,\nu}^2) = \frac{\nu}{\nu-2}$ and $E(M_t^2) = 1$. Similar result can also be derived for the idiosyncratic term. As a consequence of this scaling procedure, we are able to distinguish the effect of having distributions with different tails from the effect of different variances.

By combining one of the Equations 5.8-5.10 with the other components of the SDM-AR(2) model, we obtain three different model specifications. For ease of reference, in the rest of the chapter, we shall call the model with Equation 5.8 t-SDM-AR(2) with local and systematic fluctuation (t-SDM-AR(2)-LS). t-SDM-AR(2) with systematic fluctuation (t-SDM-AR(2)-S) and t-SDM-AR(2) with local fluctuation (t-SDM-AR(2)-L) will be used to denote the models defined in Equation 5.9 or 5.10. The reasons for naming the models in this way will be explained in Section 5.3.

5.3 The impact of $S_{t,\nu}$ in three t-SDM-AR(2) models

In this section, we study the effect of $S_{t,\nu}$ with given M_t and T_t under the assumption that the values of the parameters are identical for all models. To simplify the calculation, we introduce the transformed default rate

$$
Y_t = \Phi^{-1}(D_t). \tag{5.11}
$$

Under the large homogeneous portfolio assumption and the fact that for any i and j, $X_{i,t}$ and $X_{j,t}$ are conditionally independent with each other once given M_t, T_t and $S_{t,\nu}$, we know that the distribution of D_t converges to

$$
\begin{aligned}
D_t|M_t,T_t,S_{t,\nu} &\sim \lim_{N\to\infty}\sum_{i=1}^{N}\frac{\mathbb{1}_{(X_{i,t}<x_{PD})}}{N} \\
&\sim P(X_{i,t}\leqslant x_{PD}|M_t,T_t,S_{t,\nu}) \\
&= \Phi(Y_t).
\end{aligned}
$$

The transformed default rate Y_t for each model is of the form

$$Y_t = \begin{cases} [x_{PD}^{L,S}/\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}\right) - a(T_t)M_t] \cdot \kappa(T_t), & \text{for t-SDM-AR(2)-LS} \\ [x_{PD}^{S} - a(T_t)\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}\right)M_t] \cdot \kappa(T_t), & \text{for t-SDM-AR(2)-S} \\ \frac{\sqrt{\nu}}{\sqrt{\nu-2}S_{t,\nu}}[x_{PD}^{L} - a(T_t)M_t] \cdot \kappa(T_t), & \text{for t-SDM-AR(2)-L} \\ [x_{PD} - a(T_t)M_t] \cdot \kappa(T_t), & \text{for SDM-AR(2),} \end{cases}$$

where $\kappa(T_t) = \left[\sqrt{1 - a^2(T_t)}\right]^{-1}$. The values, $x_{PD}^{L,S}$, x_{PD}^{T}, x_{PD}^{L} and x_{PD} represent the default thresholds for each model. We use $Y_t^{L,S}$, Y_t^S, Y_t^L and Y_t to denote the transformed default rate for each model, respectively.

As we know, once M_t and T_t are given, Y_t is fixed and $Y_t^{L,S}$, Y_t^S, Y_t^L are functions of $S_{t,\nu}$ only. So, we use Y_t as a benchmark and divide $Y_t^{L,S}$, Y_t^S and Y_t^L by Y_t to compute relative changes, as compared to the SDM-AR(2) model.

In the situation where all parameters are identical across the four models, the numerical results suggest that differences among $x_{PD}^{L,S}$, x_{PD}^{S}, x_{PD}^{L} and x_{PD} are negligible:

$$x_{PD}^{L,S} \approx x_{PD}^{S} \approx x_{PD}^{L} \approx x_{PD}.$$

By making this assumption, we can concentrate on the impact of $S_{t,\nu}$ and ignore differences among the thresholds.

5.3.1 Systematic change ratio

We first divide Y_t^S by Y_t:

$$\begin{aligned} \frac{Y_t^S}{Y_t} &\approx \frac{[x_{PD} - a(T_t)\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}\right)M_t] \cdot \kappa(T_t)}{[x_{PD} - a(T_t)M_t] \cdot \kappa(T_t)} \\ &= \frac{x_{PD}}{x_{PD} - a(T_t)M_t} - \frac{a(T_t)M_t}{x_{PD} - a(T_t)M_t}\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}\right). \end{aligned}$$

We can notice that this ratio is a random variable whose value depends on $S_{t,\nu}$ when M_t and T_t are given. Some of the other properties of this ratio are easy to identify. First, the conditional expectation and variance of the ratio can be easily derived as

$$E_{S_{t,\nu}}\left[\frac{Y_t^S}{Y_t}|M_t,T_t\right] = \frac{x_{PD}}{x_{PD}-a(T_t)M_t} - \frac{a(T_t)M_t}{x_{PD}-a(T_t)M_t}\sqrt{\frac{\nu-2}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$$

$$Var_{S_{t,\nu}}\left[\frac{Y_t^S}{Y_t}|M_t,T_t\right] = \left[\frac{a(T_t)M_t}{x_{PD}-a(T_t)M_t}\right]^2\frac{\nu-2}{\nu}Var(S_{t,\nu}).$$

As we can see, both the expectation and variance are dependent on the value of M_t. For the given ν, the variance of the ratio increases when M_t decreases. From the above representations we can also conclude that

$$\lim_{\nu\to\infty} E_{S_{t,\nu}}\left[\frac{Y_t^S}{Y_t}|M_t,T_t\right] = 1. \tag{5.12}$$

$$\lim_{\nu\to\infty} Var_{S_{t,\nu}}\left[\frac{Y_t^S}{Y_t}|M_t,T_t\right] = 0. \tag{5.13}$$

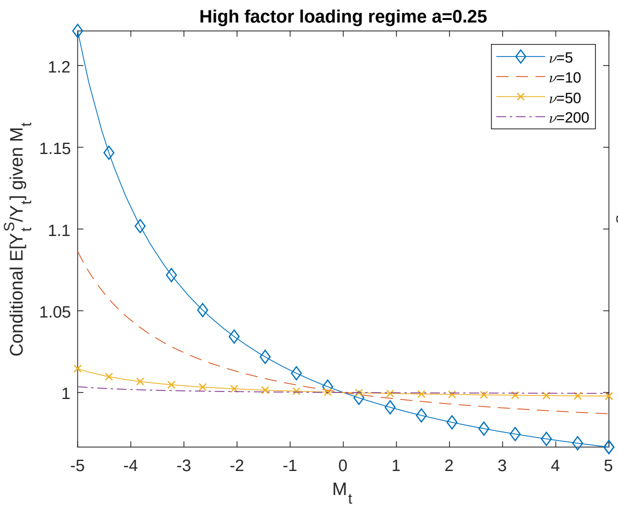
These results imply that the t-SDM-AR(2)-S model converges back to the SDM-AR(2) model as ν goes to infinity. In addition, by comparing the expectation of the ratio with 1, we have

$$1 - E_{S_{t,\nu}}\left[\frac{Y_t^S}{Y_t}|M_t,T_t\right] = \frac{a(T_t)M_t}{x_{PD}-a(T_t)M_t}\left[\sqrt{\frac{\nu-2}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} - 1\right].$$

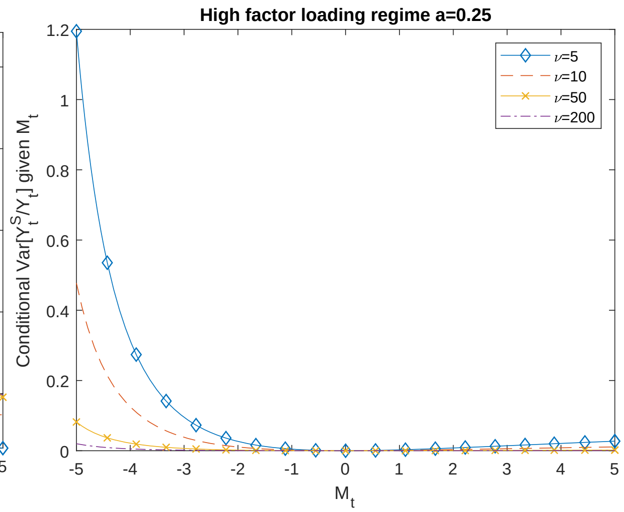
Note that $\sqrt{\frac{\nu-2}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} - 1$ is always less than 0 for any $\nu > 2$. For x_{PD} less than 0, the ratio $\frac{a(T_t)M_t}{x_{PD}-a(T_t)M_t}$ will be greater than 0 when $\frac{x_{PD}}{a(T_t)} < M_t < 0$. As a result, we can expect that Y_t^S should be larger than Y_t when $\frac{x_{PD}}{a(T_t)} < M_t < 0$. Then, according to Equation 5.11, the expected default rate of t-SDM-AR(2)-S should be larger than the one of SDM-AR(2) when $\frac{x_{PD}}{a(T_t)} < M_t < 0$.

We are able to draw the conclusion that by multiplying the tail thickness factor $S_{t,\nu}$ and the normalization scale factor $\sqrt{\frac{\nu-2}{\nu}}$ with the systematic risk factor M_t only, we should expect a broader distribution of the default rate than that for the SDM-AR(2) model when we are at a bearish market level. Also, the default rate distributions of both models are close to each other when the market is around its average. Thus, the systematic change ratio heavily depends on the level of the market.

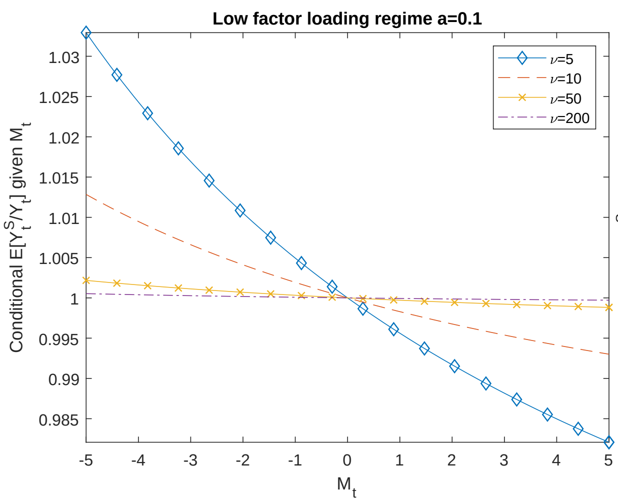
The following figures show the conditional expectation and variance of the systematic change ratio given M_t for different factor loadings and degree of freedoms.



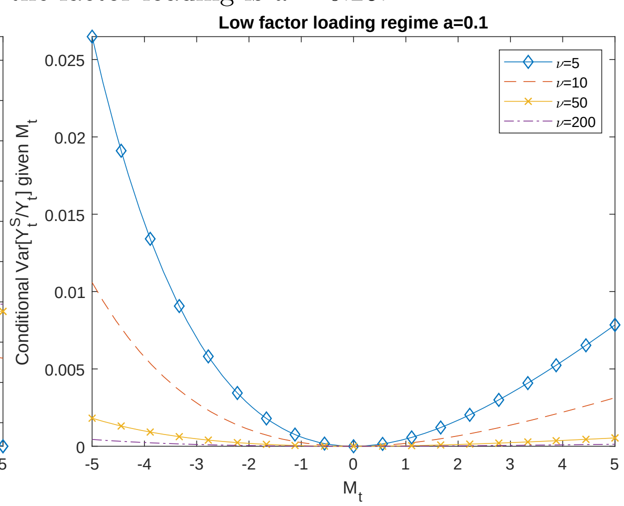
(a) Conditional expectation of $\frac{Y_t^S}{Y_t}$ given M_t when the factor loading is $a = 0.25$.



(b) Conditional variance of $\frac{Y_t^S}{Y_t}$ given M_t when the factor loading is $a = 0.25$.



(c) Conditional expectation of $\frac{Y_t^S}{Y_t}$ given M_t when the factor loading is $a = 0.1$.



(d) Conditional variance of $\frac{Y_t^S}{Y_t}$ given M_t when the factor loading is $a = 0.1$.

Figure 5.1: Effect of M_t on the conditional expectation and variance of $\frac{Y_t^S}{Y_t}$ with different factor loadings and degrees of freedom.

5.3.2 Local change ratio

In this section, we calculate the ratio between Y_t^L and Y_t, which we will call the "local change ratio":

$$\begin{aligned}\frac{Y_t^L}{Y_t} &\approx \frac{\frac{\sqrt{\nu}}{\sqrt{\nu-2}S_{t,\nu}}[x_{PD} - a(T_t)M_t] \cdot \kappa(T_t)}{[x_{PD} - a(T_t)M_t] \cdot \kappa(T_t)} \\ &= \frac{\sqrt{\nu}}{\sqrt{\nu-2}S_{t,\nu}}.\end{aligned}$$

As we can see, the interpretation of the local change ratio is quite simple. It is a random variable whose distribution is only dependent on ν and independent of M_t and T_t, even the values of other model parameters. Based on the study of $\frac{1}{S_{t,\nu}}$ made by Pimbley (2018), it is also easy to obtain:

$$\begin{aligned}E_{S_{t,\nu}}\left[\frac{Y_t^L}{Y_t}\right] &= \sqrt{\frac{2}{\nu-2}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \\ Var_{S_{t,\nu}}\left[\frac{Y_t^L}{Y_t}\right] &= \frac{\nu}{\nu-2} - \frac{2}{\nu-2}\left[\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\right]^2.\end{aligned}$$

As we can see, the local change ratio's conditional mean and variance remain the same regardless of the market level. Therefore, the default rate distribution should always be broader in the t-SDM-AR(2)-L model than the one in the SDM-AR(2) model under all market levels. As ν approaches infinity, the expected value and variance of the local change ratio converge to 1 and 0, respectively. This is also consistent with the fact that the t-SDM-AR(2)-L model converges to the SDM-AR(2) model.

5.3.3 Local-systematic change ratio

In this section, we compute the change ratio between the t-SDM-AR(2)-LS and the SDM-AR(2) model:

$$\begin{aligned}\frac{Y_t^{L,S}}{Y_t} &\approx \frac{[x_{PD}^{L,S}/\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}\right) - a(T_t)M_t] \cdot \kappa(T_t)}{[x_{PD} - a(T_t)M_t] \cdot \kappa(T_t)} \\ &= \frac{\sqrt{\nu}}{\sqrt{\nu-2}S_{t,\nu}}\frac{x_{PD}}{x_{PD} - a(T_t)M_t} - \frac{a(T_t)M_t}{x_{PD} - a(T_t)M_t}.\end{aligned}$$

Now, the change ratio is more complicated than both the systematic and local change ratios. By rearranging the above equation in the following form, we have:

$$\begin{aligned}\frac{Y_t^{L,S}}{Y_t} &= \frac{\sqrt{\nu}}{\sqrt{\nu-2}S_{t,\nu}}\left[\frac{x_{PD}}{x_{PD}-a(T_t)M_t} - \frac{a(T_t)M_t}{x_{PD}-a(T_t)M_t}\left(\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}\right)\right]\\ &= \frac{Y_t^L}{Y_t}\cdot\frac{Y_t^S}{Y_t}.\end{aligned}$$

This means that the change ratio between the t-SDM-AR(2)-LS and the SDM-AR(2) model can be decomposed into two parts: a systematic part and a local part. By keeping re-arranging the above equation, we can also find:

$$\begin{aligned}\frac{Y_t^{L,S}}{Y_t^S} &= \frac{Y_t^L}{Y_t}\\ \frac{Y_t^{L,S}}{Y_t^L} &= \frac{Y_t^S}{Y_t}.\end{aligned}$$

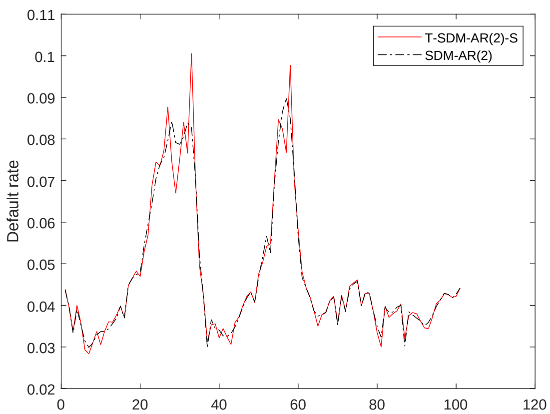
That means the t-SDM-AR(2)-LS model preserves the properties of both the t-SDM-AR(2)-L and the t-SDM-AR(2)-S model. The default rate distribution of the t-SDM-AR(2)-LS model should always be broader than the one of the SDM-AR(2) model at all market levels, and much broader when the market is far from its average.

5.3.4 Change ratio visualization

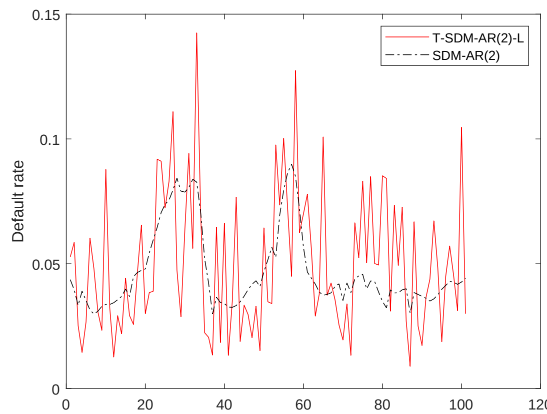
In this section, we depict some of the properties of the change ratios we have introduced in the last section by using simulated data from each model. We have chosen the following values of the parameters

$$a_1 = 0.25,\ a_2 = 0.1,\ \beta = 0.9,\ \theta = [1.8,\ -0.85],\ t_1 = 0,\ \nu = 20,\ PD = 0.05.$$

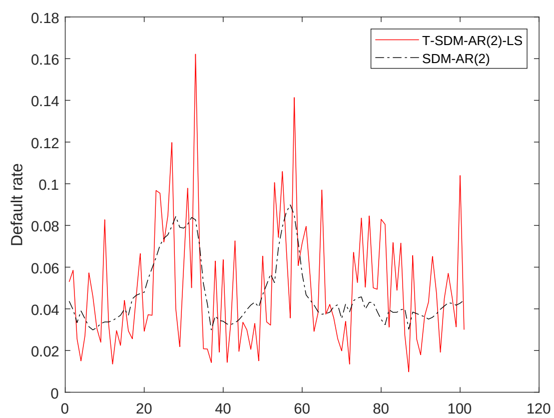
By fixing the seed of the random number generator, we use the same values of $\{M_t\}$, $\{T_t\}$ and $\{S_{t,\nu}\}$ to simulate the default rate based on t-SDM-AR(2)-LS, t-SDM-AR(2)-L, t-SDM-AR(2)-S and SDM-AR(2). The following figures show the 100 simulated default rate points from each model.



(a) t-SDM-AR(2)-S vs. SDM-AR(2)



(b) t-SDM-AR(2)-L vs. SDM-AR(2)

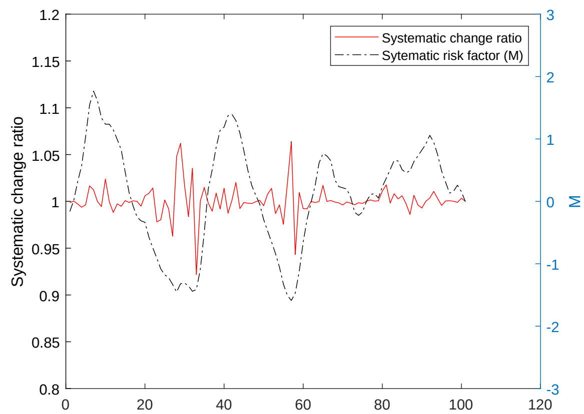


(c) t-SDM-AR(2)-LS vs. SDM-AR(2)

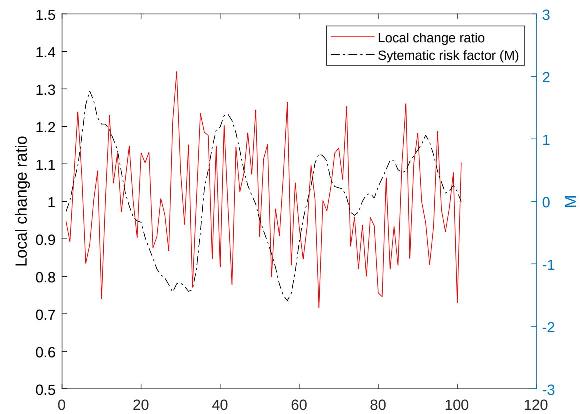
Figure 5.2: Simulated default rates from the t-SDM-AR(2)-LS, t-SDM-AR(2)-L, t-SDM-AR(2)-S, SDM-AR(2).

As we can observe from panel (a), the discrepancy between the two models is more pronounced when the default rate is relatively high. The discrepancies shown in the other two panels fluctuate randomly at all market levels.

The following figures show the systematic change ratio and local change ratio along with the systematic risk factor M_t.



(a) Systematic change ratio



(b) Local change ratio

Figure 5.3: Systematic and local change ratio vs. systematic risk factor

The above figures provide strong visual evidence to support the idea that the systematic change ratio has large variation when the market is bearish and the variation of the local change ratio is independent of the market level.

To summarize our findings in this section, we can interpret the t-SDM-AR(2)-LS, t-SDM-AR(2)-S, t-SDM-AR(2)-L and SDM-AR(2) models in the following way:

- The t-SDM-AR(2)-L model provides extra volatility in all market levels. This volatility is independent of the market level.
- The t-SDM-AR(2)-S model provides additional volatility when the market is bearish. Also, the additional volatility increases when the market declines.
- The t-SDM-AR(2)-LS model can be treated as a combination of t-SDM-AR(2)-L and t-SDM-AR(2)-S, and hence it provides additional volatility for all market scenarios. The additional volatility is amplified when the market declines.

Since we believe the overall market level should have some impact on the distribution of the default rate, we expect that the t-SDM-AR(2)-S and t-SDM-AR(2)-LS models should perform better than the other models in describing the Federal Reserve Data.

5.4 Credit threshold of the three t-SDM-AR(2) models

Since in the proposed models, we define the credit score $X_{i,t}$ differently than in the SDM-AR(2) model, it is essential to calculate the default threshold x_{PD} for each t-SDM-AR(2) model. In this section, we derive the formulas for each x_{PD}. First, by conditioning on $S_{t,\nu}$, M_t and T_t, we can find that the conditional distributions of $X_{i,t}$ are of the form:

$$P(X_{i,t} \leqslant x | S_{t,\nu}, M_t, T_t) = \begin{cases} \Phi([x - a(T_t)\zeta(\nu)S_{t,\nu}M_t]/[\kappa(T_t)\zeta(\nu)S_{t,\nu}]) & \text{for t-SDM-AR(2)-LS} \\ \Phi([x - a(T_t)\zeta(\nu)S_{t,\nu}M_t]/\kappa(T_t)) & \text{for t-SDM-AR(2)-S} \\ \Phi([x - a(T_t)M_t]/[\kappa(T_t)\zeta(\nu)S_{t,\nu}]) & \text{for t-SDM-AR(2)-L,} \end{cases} \tag{5.14}$$

where $\zeta(\nu) = \sqrt{\frac{\nu-2}{\nu}}$ and $\kappa(T_t) = \sqrt{1 - a(T_t)^2}$. We can also present the expressions in Equation 5.14 in the form of $\Phi(x; \mu(S_{t,\nu}, M_t, K), \sigma(K))$, where

$$\mu(S_{t,\nu}, M_t, K) = \begin{cases} a_K\zeta(\nu)S_{t,\nu}M_t & \text{for t-SDM-AR(2)-LS and t-SDM-AR(2)-S} \\ a_K M_t & \text{for t-SDM-AR(2)-L,} \end{cases}$$

$$\sigma(S_{t,\nu}, K) = \begin{cases} \sqrt{1 - a_K^2}\zeta(\nu)S_{t,\nu} & \text{for t-SDM-AR(2)-LS and t-SDM-AR(2)-L} \\ \sqrt{1 - a_K^2} & \text{for t-SDM-AR(2)-S.} \end{cases}$$

Then, we have:

$$P(X_{i,t} \leqslant x | S_{t,\nu}) = \int_{-\infty}^{\infty} \left[\sum_{k=1}^{K} \Phi\left(x; \mu(S_{t,\nu}, m_t, k), \sigma(S_{t,\nu}, k)\right) \cdot p_k(m) \right] \cdot \phi(m)\ dm, \tag{5.15}$$

where $p_k(m)$ is defined in Equation 2.7. By applying the same method we have used for Equation 2.12, we can further simplify Equation 5.15 as:

$$P(X_{i,t} \leqslant x | S_{t,\nu}) = \sum_{k=1}^{K} E_M \left[\Phi_2^* \left(\begin{bmatrix} x \\ t_{k+1} \end{bmatrix}, \begin{bmatrix} x \\ t_k \end{bmatrix}; \begin{bmatrix} \mu(S_{t,\nu}, M, k) \\ \beta M \end{bmatrix}, \begin{bmatrix} \sigma(S_{t,\nu}, k)^2 & 0 \\ 0 & 1 - \beta^2 \end{bmatrix} \right) \right],$$

where $\Phi_2^*(x_1, x_2; \mu, \Sigma) = \Phi_2(x_1; \mu, \Sigma) - \Phi_2(x_2; \mu, \Sigma)$.

To apply Theorem 2, we can set $a(k) = \begin{bmatrix} x \\ t_k \end{bmatrix}$, $\Sigma(S_{t,\nu}, k) = \begin{bmatrix} \sigma(S_{t,\nu}, k)^2 & 0 \\ 0 & 1-\beta^2 \end{bmatrix}$ and $b(S_{t,\nu}, k) = \begin{bmatrix} b^*(S_{t,\nu}, k) \\ \beta \end{bmatrix}$, where

$$b^*(S_{t,\nu}, k) = \begin{cases} a_k \zeta(\nu) S_{t,\nu} & \text{For t-SDM-AR(2)-LS and t-SDM-AR(2)-S} \\ a_k & \text{For t-SDM-AR(2)-L.} \end{cases}$$

Then we have:

$$P(X_{i,t} \leqslant x | S_{t,\nu}) = \sum_{k=1}^{K} \Phi_2^*(a(k+1), a(k); 0, \Sigma(S_{t,\nu}, k) + b(S_{t,\nu}, k) b(S_{t,\nu}, k)^T).$$

By integrating over $S_{t,\nu}$, we find the unconditional distribution function of $X_{i,t}$ to be

$$P(X_{i,t} \leqslant x) = \int_0^\infty P\left(X_i \leqslant x | s\right) f_{S_\nu}(s)\, ds,$$

where $f_{S_\nu}(s)$ is defined in Equation 5.6. The credit threshold x_{PD} can be found numerically by solving the following equation:

$$P(X_{i,t} \leqslant x_{PD}) = PD.$$

5.5 Filtering procedure and likelihood function

In this section, we develop a filtering procedure and likelihood function for the observations of the three t-SDM-AR(2) models.

We start this section by developing a filtering procedure. The first step is to determine the one-step-ahead state predictive density function. After that, we start to derive the one-step-ahead observation predictive density function. The last function we need is the filtering density function. By using the same method we have used to find the likelihood function of the SDM-AR(2) model, we will also be able to derive the likelihood function of the t-SDM-AR(2) models.

5.5.1 One-step-ahead predictive density for the states

The main idea behind the one-step-ahead predictive density function of the t-SDM-AR(2) models is similar to the one of the SDM-AR(2) model in Section 3.7. We first derive the one-step-ahead state predictive density by the following equation:

$$f_{M_{t+1},M_t|Y_{1:t}}(m_{t+1},m_t|y_{1:t}) = \int_{-\infty}^{\infty} f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t,m_{t-1}) f_{M_t,M_{t-1}|Y_{1:t}}(m_t,m_{t-1}|y_{1:t})\ dm_{t-1}, \tag{5.16}$$

where $f_{M_{t+1}|M_t,M_{t-1}}()$ is the transition density function of the AR(2) process $\{M_t\}$, given by

$$f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t,m_{t-1}) = \frac{1}{\sigma_\epsilon}\phi(\frac{m_{t+1}-\theta_1 m_t - \theta_2 m_{t-1}}{\sigma_\epsilon}),$$

with σ_ϵ defined in Equation 5.7.

The last term in Equation 5.16, $f_{M_t,M_{t-1}|Y_{1:t}}$, is the filtering density function that we will determine later in this section. A proof of Equation 5.16 is similar to that we have presented in Section 3.7.

5.5.2 One-step-ahead predictive density function for the transformed default rate

The one-step-ahead prediction density function of the observations can be derived by first applying the law of total probability,

$$\begin{aligned} f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) &= \int_{-\infty}^{\infty} f_{Y_t,M_t|Y_{1:t-1}}(y_t,m_t|y_{1:t-1})\ dm_t \\ &= \int_{-\infty}^{\infty} f_{Y_t|M_t,Y_{1:t-1}}(y_t|m_t,y_{1:t-1}) f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})\ dm_t. \end{aligned}$$

Since we know that Y_t is conditionally independent of its past path $\{Y_{1:t-1}\}$ once given M_t, we can simplify the above representation to the following form:

$$f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) = \int_{-\infty}^{\infty} f_{Y_t|M_t}(y_t|m_t) f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1}) dm_t, \tag{5.17}$$

where $f_{Y_t|M_t}()$ is defined in Section D for each t-SDM-AR(2) model, and $f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1})$ can be calculated by integrating the Equation 5.16 with respect to m_{t-1} over all real numbers

$$f_{M_t|Y_{1:t-1}}(m_t|y_{1:t-1}) = \int_{-\infty}^{\infty} f_{M_t,M_{t-1}|Y_{1:t-1}}(m_t,m_{t-1}|y_{1:t-1})\ dm_{t-1}.$$

5.5.3 Filtering density function

We are able to compute the filtering density function, $f_{M_t,M_{t-1}|Y_{1:t}}(m_t, m_{t-1}|y_{1:t})$, by using the definition of the conditional density first:

$$\begin{aligned} f_{M_t,M_{t-1}|Y_{1:t}}(m_t, m_{t-1}|y_{1:t}) &= \frac{f_{M_t,M_{t-1},Y_{1:t}}(m_t, m_{t-1}, y_{1:t})}{f_{Y_{1:t}}(y_{1:t})} \\ &= \frac{f_{M_t,M_{t-1},Y_{1:t}}(m_t, m_{t-1}, y_{1:t})/f_{Y_{1:t-1}}(y_{1:t-1})}{f_{Y_{1:t}}(y_{1:t})/f_{Y_{1:t-1}}(y_{1:t-1})} \\ &= \frac{f_{Y_t,M_t,M_{t-1}|Y_{1:t-1}}(y_t, m_t, m_{t-1}|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})} \\ &= \frac{f_{Y_t|M_t}(y_t|m_t) f_{M_t,M_{t-1}|Y_{1:t-1}}(m_t, m_{t-1}|y_{1:t-1})}{f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})}. \end{aligned}$$

The last step is derived by the property that Y_t is conditionally independent of both its past path $\{Y_{1:t-1}\}$ and M_{t-1} once the value of M_t is given.

5.5.4 Likelihood function of the transformed default rate

The likelihood function can be calculated by decomposing it into the following form:

$$f_{Y_{1:N}}(y_{1:N}) = f_{Y_1}(y_1) \prod_{i=2}^{N} f_{Y_i|Y_{1:i-1}}(y_i|y_{1:i-1}), \tag{5.18}$$

where $f_{Y_i|Y_{1:i-1}}$ has been determined in Section 5.5.2 and $f_{Y_t}()$ is the stationary density function of the transformed default rate, which can be derived by considering its stationary distribution function first,

$$P(Y_t \leqslant y_t) = \int_{-\infty}^{\infty} (1 - P(Y_t \geqslant y_t|M_t = m_t)) \cdot \phi(m_t)\ dm_t$$

By taking the derivatives, we can find the stationary density function of Y_t.

5.6 Model estimation

In this section, we discuss the problem of model estimation in the context of the three t-SDM-AR(2) models defined in Section 5.2. Although we have developed the formula for the likelihood function in Equation 5.18, the challenge is that the equation contains some three-dimensional integration. In our experience, the built-in `integral3` function in Matlab does not appear to be well-suited to this problem, in the sense that computational times for the integrands we are dealing with are excessively high. In order to speed up the calculation, we use the standard numerical method presented in Appendix E to approximate the integral.

We demonstrate the accuracy of the estimation procedure in Section 5.6.1. Then, we briefly discuss the method we used to select the initial points for the numeric optimizer. After that, we use the simulated data to demonstrate the proposed models' forecasting ability. In the last part, we apply the proposed models to the Federal Reserve Data we used in Chapter 4.

5.6.1 Simulation study

In this section we use simulated data to verify the accuracy of the estimation procedure for the proposed models.

In order to verify that the proposed estimation procedure works for different parameter settings, we first simulate one time series with 500 observations based on four different parameter settings. The initial point for the numerical optimizer is set at the true values of the parameters. By doing so, we are able to verify the performance of the estimation procedure in different scenarios given condition on having good starting values. Table 5.1 presents the results for this test.

	a_1	a_2	β	t	θ_1	θ_2	ν
True value	0.2	0.1	1	-0.3	1.8	-0.85	20
t-SDM-AR(2)-L	0.176	0.051	0.99	0.751	1.849	-0.903	19
t-SDM-AR(2)-S	0.193	0.085	0.99	-0.252	1.757	-0.811	16
t-SDM-AR(2)-LS	0.169	0.058	0.99	0.780	1.859	-0.912	19
True value	0.2	0.1	1	-0.3	1.8	-0.85	100
t-SDM-AR(2)-L	0.180	0.097	0.99	0.780	1.820	-0.877	96
t-SDM-AR(2)-S	0.189	0.102	0.99	0.339	1.787	-0.853	103
t-SDM-AR(2)-LS	0.178	0.097	0.99	0.810	1.824	-0.882	92
True value	0.2	0.1	1	-0.3	1.8	-0.85	300
t-SDM-AR(2)-L	0.193	0.105	0.99	0.536	1.799	-0.860	302
t-SDM-AR(2)-S	0.203	0.098	0.99	-0.10	1.777	-0.829	290
t-SDM-AR(2)-LS	0.193	0.106	0.99	0.540	1.803	-0.863	290
True value	0.2	0.1	0.5	-0.3	1.8	-0.85	20
t-SDM-AR(2)-L	0.214	0.115	0.24	-0.463	1.849	-0.913	23
t-SDM-AR(2)-S	0.205	0.104	0.713	-0.069	1.803	-0.860	17
t-SDM-AR(2)-LS	0.167	0.066	0.63	-0.135	1.838	-0.875	19

Table 5.1: The simulation test of t-SDM-AR(2)-L, t-SDM-AR(2)-S, and t-SDM-AR(2)-LS models based on four different parameter settings with 500 simulated data points.

As we can see from Table 5.1, the estimators for both θ_1 and θ_2 work reasonably well in all cases considered in this study. The pattern of results becomes more complicated when look at the parameters a_1 and a_2. The discrepancy between the true values and the estimates of a_1 and a_2 increases when the degree of freedom ν decreases. This result suggests that the effect of having different regimes can be difficult to distinguish from the effect of modeling the systematic risk factor based on distributions with different tails. This brings to our attention the well-known parameter identification problem, solution of which we leave as a possible future research direction.

In the second stage of our simulation study, we estimated some basic statistics regarding the estimators by repeating the simulation study mentioned in last paragraph 50 times. Because of the computational complexity of the problem, we only perform this study to the t-SDM-AR(2)-S model and reduce the number of observation for each simulated path to 300. The results are displayed in Table 5.2.

	a_1	a_2	t	β	θ_1	θ_2	ν
True Value	0.2	0.1	0	0.9	1.8	-0.85	100
Average of estimators	0.196	0.102	0.01	0.888	1.764	-0.815	83
Std.Err of estimators	0.019	0.014	0.09	0.04	0.04	0.04	17.28
T-stat	-0.199	0.126	0.076	-0.311	-0.981	0.953	-1.006
P-Value	0.84	0.90	0.94	0.76	0.33	0.34	0.319
True Value	0.3	0.1	0	0.5	1.5	-0.65	50
Average of estimators	0.302	0.101	-0.01	0.487	1.488	-0.643	44
Std.Err of estimators	0.025	0.008	0.10	0.08	0.06	0.05	35.78
T-stat	0.060	0.087	-0.067	-0.168	-0.220	0.138	-0.170
P-Value	0.95	0.93	0.95	0.87	0.83	0.89	0.87
True Value	0.3	0.1	0	0.5	1.5	-0.65	20
Average of estimators	0.297	0.099	-0.01	0.477	1.477	-0.635	21
Std.Err of estimators	0.020	0.008	0.08	0.07	0.06	0.06	6.29
T-stat	-0.124	-0.098	-0.137	-0.319	-0.368	0.247	0.154
P-Value	0.90	0.92	0.89	0.75	0.71	0.80	0.87

Table 5.2: t-SDM-AR(2)-S model estimators stability test. We repeat the estimation procedure 50 times for each parameter setting. In each iteration, 300 data points are simulated to perform the estimation.

From Table 5.2, we can notice that there is no obvious evidence of a bias of the proposed estimation procedure. Overall, all the estimators appear to perform well. This gives us the confidence about the accuracy of the proposed estimation method.

5.6.2 Initial point selection

The results presented in Section 5.6.1 suggest that the proposed estimation procedure works reasonably well. But there is still one problem we need to address. This is how to properly select the initial points for the numerical optimizer we used in Matlab to find the maximum likelihood estimates. Due to the computational complexity of the t-SDM-AR(2) models, we are not able to use the exhaustive search method as we did in Section 4.1.2. Therefore, We need to find out a reasonable and reliable initial point.

In the current stage, we have already derived the estimates for the SDM-AR(2) model in Section 4.1.2, so we suspect that the value of the parameters for the t-SDM-AR(2) models should be not too far away from the ones for the SDM-AR(2) model. In order to verify this idea, we perform the following simulation test. We first generate 500 data points from the t-SDM-AR(2)-S, t-SDM-AR(2)-L and t-SDM-AR(2)-LS models respectively, and then estimate the SDM-AR(2) model with the simulated data to check if the results are close enough to the true values. We test four different values of the parameters. Table 5.3 presents the results.

As we can tell from Table 5.3, the impact of ν is quite large. The estimators work well in the situations when the data is generated from the t-SDM-AR(2)-S model, regardless of the degrees of freedom. But for the t-SDM-AR(2)-L and t-SDM-AR(2)-LS models, the consequence of eliminating the $S_{t,\nu}$ brings a significant effect on the estimation values of the model parameters. As we can see, the discrepancy decreases when we increase the value of ν. This is because all three t-SDM-AR(2) models converge to the SDM-AR(2) model as ν approaches infinity. By combining all those results, we can notice that the SDM-AR(2) model is robust to small and modest deviations from its assumed framework. However, the presented results also suggest that for larger deviations it is important to use a proper model.

Model used to simulate data	a_1	a_2	t	β	θ_1	θ_2	ν
True Value	0.2	0.1	0	0.9	1.8	-0.85	20
t-SDM-AR(2)-S	0.188	0.116	-0.049	1	1.116	-0.201	
t-SDM-AR(2)-L	0.297	0.297	2.379	0.999	0.219	0.153	
t-SDM-AR(2)-LS	0.318	0.091	1.611	0.588	0.185	0.134	

(a) Estimation result of SDM-AR(2) model with simulated data from different t-SDM-AR(2) model.

Model used to simulate data	a_1	a_2	t	β	θ_1	θ_2	ν
True Value	0.2	0.1	0	0.9	1.8	-0.85	100
t-SDM-AR(2)-S	0.206	0.155	-0.067	0.96	1.626	-0.690	
t-SDM-AR(2)-L	0.241	0.178	-1.040	-0.33	0.461	0.240	
t-SDM-AR(2)-LS	0.201	0.189	-1.589	-0.97	0.429	0.246	

(b) Estimation result of SDM-AR(2) model with simulated data from different t-SDM-AR(2) model.

Model used to simulate data	a_1	a_2	t	β	θ_1	θ_2	ν
True Value	0.2	0.1	0	0.9	1.8	-0.85	300
t-SDM-AR(2)-S	0.197	0.110	0.225	0.927	1.711	-0.771	
t-SDM-AR(2)-L	0.215	0.160	-1.406	-0.29	0.659	0.179	
t-SDM-AR(2)-LS	0.169	0.160	0.942	0.998	0.607	0.215	

(c) Estimation result of SDM-AR(2) model with simulated data from different t-SDM-AR(2) model.

Model used to simulate data	a_1	a_2	t	β	θ_1	θ_2	ν
True Value	0.2	0.1	0	0.5	1.8	-0.85	20
t-SDM-AR(2)-S	0.214	0.104	0.071	0.608	1.224	-0.300	
t-SDM-AR(2)-L	0.312	0.286	-1.248	-0.995	0.200	0.145	
t-SDM-AR(2)-LS	0.302	0.000	2.432	0.232	0.171	0.147	

(d) Estimation result of SDM-AR(2) model with simulated data from different t-SDM-AR(2) model.

Table 5.3: Simulation test: Estimation result of SDM-AR(2) model to the simulated data from three t-SDM-AR(2) models respectively with the initial points of the numerical optimizer set at the true values.

5.6.3 Forecasting

In this section, we perform some studies on the forecasting ability of each of the t-SDM-AR(2) models. We first generate 104 simulation data points from each t-SDM-AR(2) model with the same random seed and the following parameter setting:

a_1	a_2	t	θ_1	θ_2	β	PD
0.2	0.1	0	1.8	-0.85	0.99	0.05

Table 5.4: Parameter settings for generating simulation data

For each parameter setting, we set the degree of freedom to be 10, 50 and 100 to check the effect of the degree of freedom on forecasting ability.

The one-step ahead forecasting value at time t, $\hat{D}_t$, is calculated by:

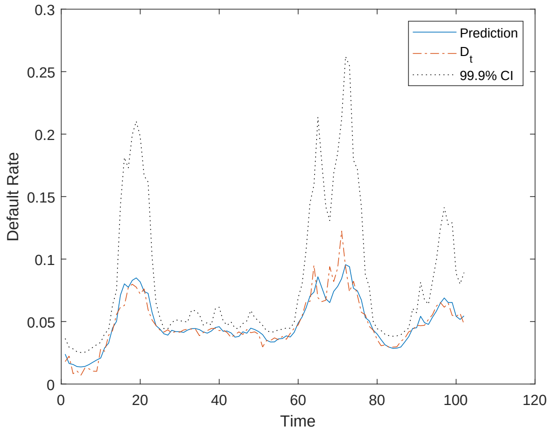
$$\hat{D}_t = E[\Phi(Y_t)|Y_{1:t-1} = y_{1:t-1}] \tag{5.19}$$

$$= \int_{-\infty}^{\infty} \Phi(y_t) f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})\ dy_t, \tag{5.20}$$

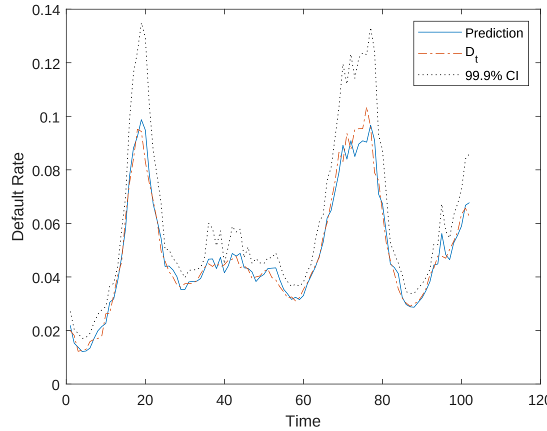
where $f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})$ is defined in Section 5.5.2 for each t-SDM-AR(2) model. We also calculate the one-step ahead predictive up-side 99.9% upper bound, $\hat{D}_{t,99.9\%} = \Phi(\hat{y}_{t,99.9\%})$, where $\hat{y}_{t,99.9\%}$ is found by solving the following equation:

$$\int_{-\infty}^{\hat{y}_{t,99.9\%}} f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})\ dy_t = 0.999. \tag{5.21}$$

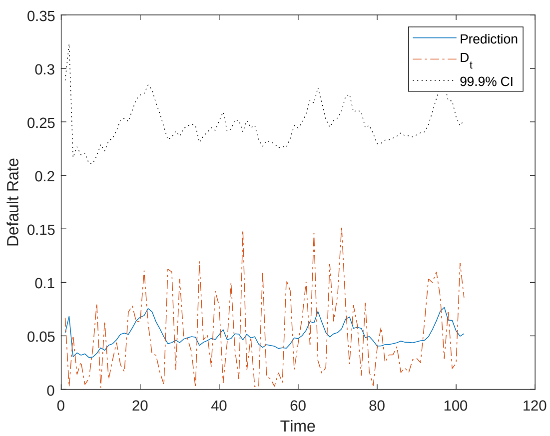
The following figures show the simulated data along with the forecasting series and the 99.9% CI.



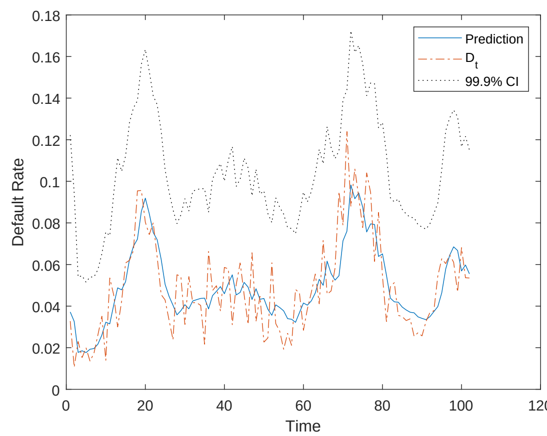
(a) t-SDM-AR(2)-S: $\nu = 10$



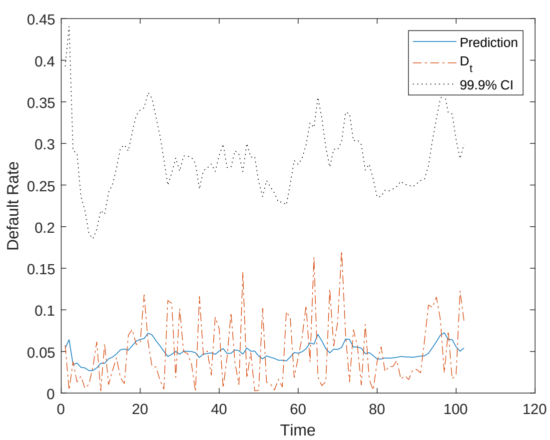
(b) t-SDM-AR(2)-S: $\nu = 100$



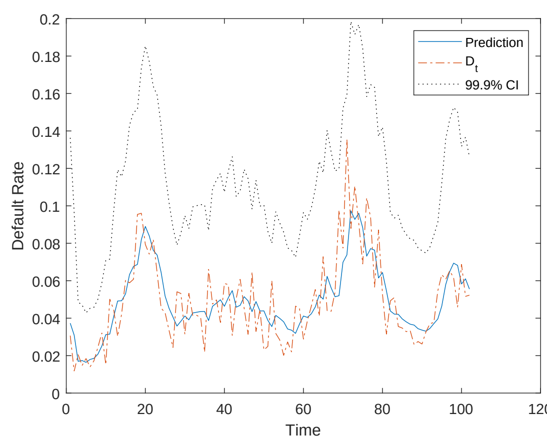
(c) t-SDM-AR(2)-L: $\nu = 10$



(d) t-SDM-AR(2)-L: $\nu = 100$



(e) t-SDM-AR(2)-LS: $\nu = 10$



(f) t-SDM-AR(2)-LS: $\nu = 100$

Figure 5.4: The one-step-ahead forecasting series and confidence interval of t-SDM-AR(2) models

First, we can notice that all the simulation data fall into the prediction confidence interval in all figures. The following tables provide a more quantitative measurement on the point prediction. The formulas for those measurements can be found in Equations 4.3-4.5.

Unit: %	t-SDM-AR(2)-S			t-SDM-AR(2)-L			t-SDM-AR(2)-LS		
ν	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
10	0.48	0.77	12.28	3.17	3.84	243.68	3.19	3.94	218.38
50	0.29	0.43	6.13	1.53	1.86	43.95	1.58	1.95	42.97
100	0.27	0.38	5.71	1.13	1.41	28.14	1.17	1.51	27.83

Table 5.5: The forecasting accuracy measurements of t-SDM-AR(2) models

As we can see from Table 5.5, point estimation is more accurate when the degree of freedom is larger. The results also suggest that the level of forecasting accuracy is relatively higher for the t-SDM-AR(2)-S model than the others.

5.7 Empirical study

Using simulated data, we have demonstrated that the estimated values of the parameters of the SDM-AR(2) give a reasonable starting point for the t-SDM-AR(2)-S model. Given that we can not apply the exhaustive searching method we applied in Section 4.1.2 due to the computational complexity, the values of the estimates based on the SDM-AR(2) model are reasonable initial points for the numerical optimizer. Under this assumption, we apply the maximum likelihood method to the t-SDM-AR(2)-S, t-SDM-AR(2)-L and t-SDM-AR(2)-LS models.

Let us recall that the data is obtained from the Federal Reserve and consists of quarterly delinquency rates for 11 different categories. A brief summary table for the data set can be found at the beginning of Chapter 4. In this section, we present the results based on the same categories we have used in Chapter 4.

5.7.1 Parameter estimation and in-sample prediction

In this section, we perform the estimation to check the models' fitting accuracy. The following table shows the results along with the estimates for the SDM-AR(2) model to compare with.

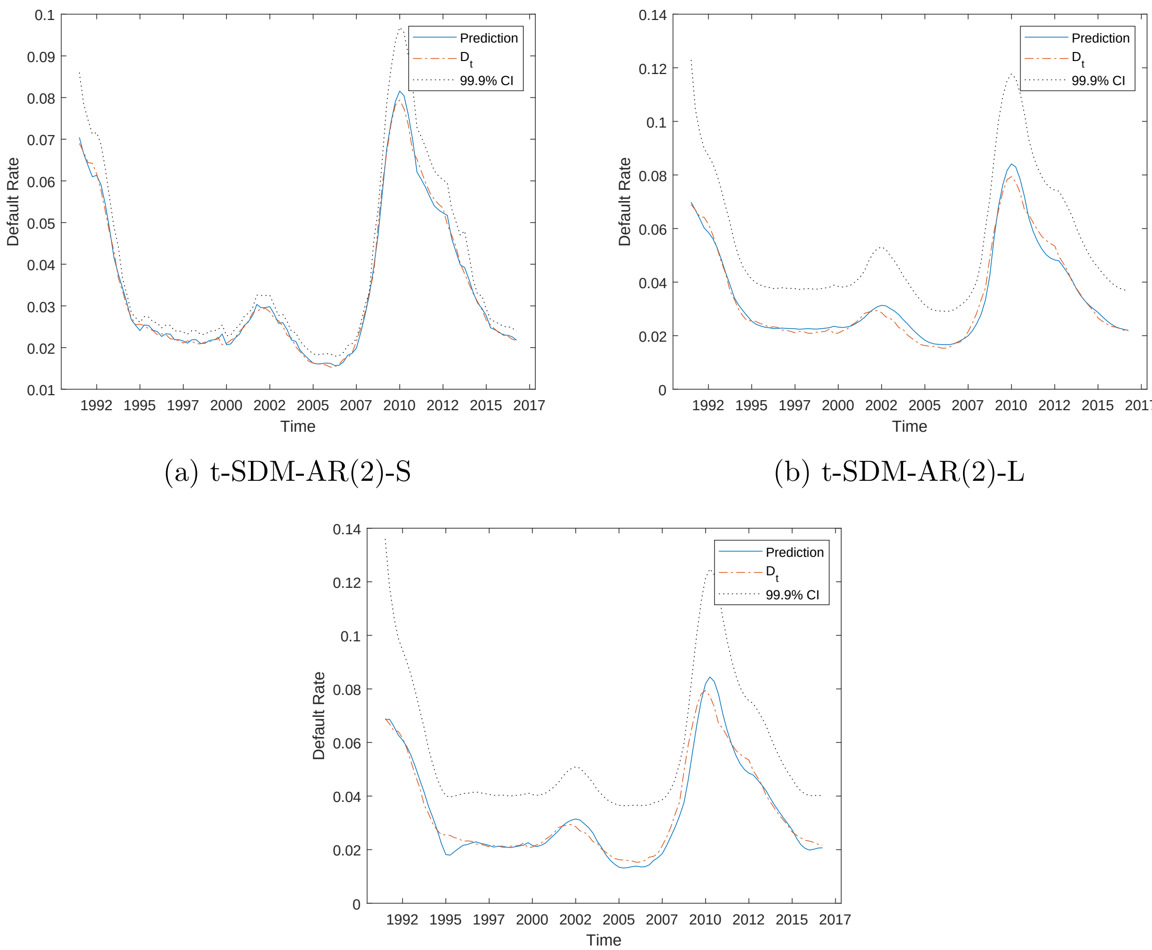
Series	Model	a_1	a_2	t	β	θ_1	θ_2	ν	$\ell(y_{1:T})$
All	t-SDM-AR(2)-S	0.204	0.126	-0.767	0.99	1.867	-0.889	500	-286.72
	t-SDM-AR(2)-L	0.322	0.144	-0.441	1.00	1.926	-0.946	500	-160.56
	t-SDM-AR(2)-LS	0.186	0.155	0.958	1.00	1.903	-0.924	500	-159.97
	SDM-AR(2)	0.232	0.162	-0.704	1.00	1.832	-0.853		-287.43
OC	t-SDM-AR(2)-S	0.104	0.097	-0.066	0.36	1.520	-0.539	252	-285.78
	t-SDM-AR(2)-L	0.058	0.048	0.998	0.99	-0.135	0.865	348	-396.05
	t-SDM-AR(2)-LS	0.042	0.013	-0.999	0.99	-0.015	0.985	70	-365.73
	SDM-AR(2)	0.115	0.085	-0.641	1.00	1.366	-0.386		-283.51
CRE	t-SDM-AR(2)-S	0.494	0.468	-0.901	1.00	1.819	-0.828	500	-206.95
	t-SDM-AR(2)-L	0.417	0.395	-0.117	-1.00	1.952	-0.963	500	-151.50
	t-SDM-AR(2)-LS	0.302	0.260	-0.450	-1.00	1.917	-0.930	500	-152.66
	SDM-AR(2)	0.392	0.378	-0.621	-1.00	1.790	-0.802		-217.80

Table 5.6: In-sample estimation result comparison between t-SDM-AR(2)-S, t-SDM-AR(2)-L, t-SDM-AR(2)-LS and SDM-AR(2).

We have encountered some numerical issues when estimating the t-SDM-AR(2) models for the SRE category, and hence we omit the results here.

Although we can see from the results that the t-SDM-AR(2)-S model is pretty close to the SDM-AR(2) model, we need to keep in mind that the initial point of the optimizer is set to be the value of SDM-AR(2) model. By checking on the values of ν, we realize that some of them are high enough to let us draw the conclusion that all three t-SDM-AR(2) models are close to the SDM-AR(2) model with the same parameters.

To assess the goodness of fit, we obtain the one-step-ahead predictions for each t-SDM-AR(2) model based on the values of the parameters in Table 5.6. Each prediction is calculated by the method discussed in Section 5.6.3. The following figures show the prediction series along with the 99.9% upside interval predictions from each t-SDM-AR(2) model.



(a) t-SDM-AR(2)-S

(b) t-SDM-AR(2)-L

(c) t-SDM-AR(2)-LS

Figure 5.5: t-SDM-AR(2): The one-step-ahead point and interval prediction for ALL series

By looking at the figures, we can see that all the t-SDM-AR(2) models provide some reasonable predictions. Also, the upper boundary of each model covers all the historical data. The table below shows the accuracy measurements for the fittings of the t-SDM-AR(2) models using a subset of the Federal Reserve Data:

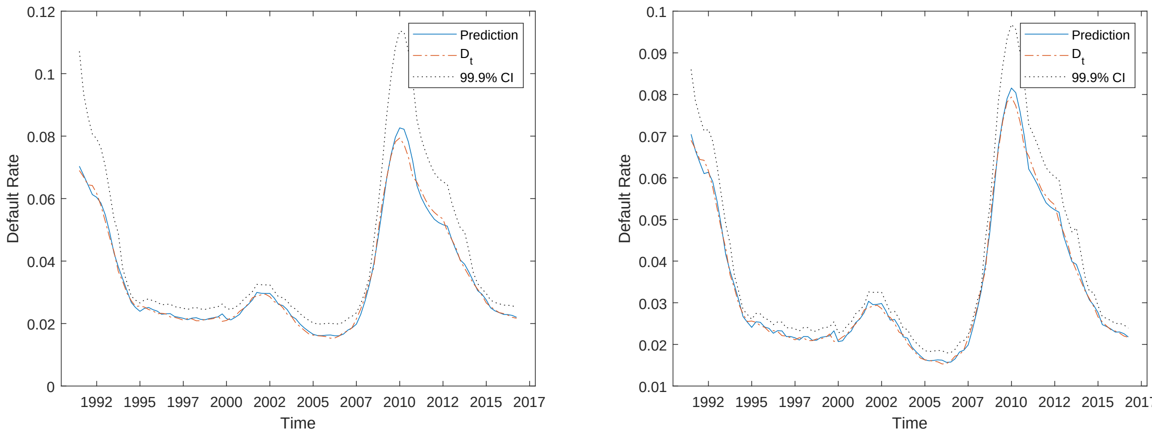
Series	Model	MAE(%)	RMSE(%)	MAPE(%)
ALL	t-SDM-AR(2)-S	0.09	0.11	2.55
	t-SDM-AR(2)-L	0.19	0.25	6.02
	t-SDM-AR(2)-LS	0.25	0.36	7.82
	SDM-AR(2)	0.08	0.11	2.51
OC	t-SDM-AR(2)-S	0.08	0.11	2.85
	t-SDM-AR(2)-L	0.45	0.53	16.8
	t-SDM-AR(2)-LS	0.41	0.52	16.11
	SDM-AR(2)	0.08	0.11	2.93
CRE	t-SDM-AR(2)-S	0.18	0.26	5.41
	t-SDM-AR(2)-L	0.27	0.38	7.93
	t-SDM-AR(2)-LS	0.29	0.42	8.79
	SDM-AR(2)	0.17	0.24	5.33

Table 5.7: The in-sample accuracy measurements of t-SDM-AR(2) models.

As we can easily tell from Table 5.7, the t-SDM-AR(2)-S and SDM-AR(2) models provide the most accurate predictions when compared with the other two models. The accuracies of the t-SDM-AR(2)-L and t-SDM-AR(2)-LS models are similar to each other.

5.7.2 Relation between degree of freedom and interval prediction

In this section, we demonstrate and discuss the relationship between the degree of freedom and the interval prediction. The following two figures show the predictions of the t-SDM-AR(2)-S model based on the same parameter values in Table 5.6 for ALL series with the degree of freedom set to be 100 and 500, respectively.



(a) Degree of freedom $\nu = 100$.

(b) Degree of freedom $\nu = 500$.

Figure 5.6: t-SDM-AR(2)-S for ALL series.

It is easy to notice that the interval prediction is more conservative when the degree of freedom is lower. The following table shows the impact of changing the degree of freedom on the point prediction.

Unit: %	t-SDM-AR(2)-S			t-SDM-AR(2)-L			t-SDM-AR(2)-LS		
ν	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
100	0.10	0.14	2.77	0.32	0.41	11.07	0.51	0.62	17.83
500	0.09	0.11	2.55	0.19	0.25	6.02	0.25	0.36	7.82

Table 5.8: The effect of changing the degree of freedom on the point prediction for ALL series.

Table 5.8 does suggest that lowering the degree of freedom makes the point prediction less accurate. The accuracy of the point prediction for t-SDM-AR(2)-L and t-SDM-AR(2)-LS is more sensitive than that for the t-SDM-AR(2)-S model. However, the reduction of the accuracy for the t-SDM-AR(2)-S model stands within a tolerable level.

5.7.3 Out-of-sample prediction for the t-SDM-AR(2) models

In this section, we perform the out-of-sample test in the same manner as we have done in Sections 4.4 and 4.5. The following table shows the estimation results of the three t-SDM-AR(2) models based on the training set that includes only the first 60 data points.

Series	Model	a_1	a_2	t	β	θ_1	θ_2	ν
ALL	t-SDM-AR(2)-S	0.117	0.09	-0.03	1	1.779	-0.805	319
	t-SDM-AR(2)-L	0.309	0.126	-0.6	1	1.931	-0.947	500
	t-SDM-AR(2)-LS	0.181	0.116	-0.662	1	1.88	-0.899	500
OC	t-SDM-AR(2)-S	0.255	0.203	-0.963	1	1.442	-0.445	72
	t-SDM-AR(2)-L	0.892	0.478	-0.301	-1	-0.059	0.939	500
	t-SDM-AR(2)-LS	0.664	0.013	0.619	-1	1.997	-0.997	86
CRE	t-SDM-AR(2)-S	0.399	0.377	-0.234	0.91	1.745	-0.752	363
	t-SDM-AR(2)-L	0.472	0.354	0.84	0.98	1.999	-0.999	27
	t-SDM-AR(2)-LS	0.325	0.282	-1	1	1.933	-0.939	500

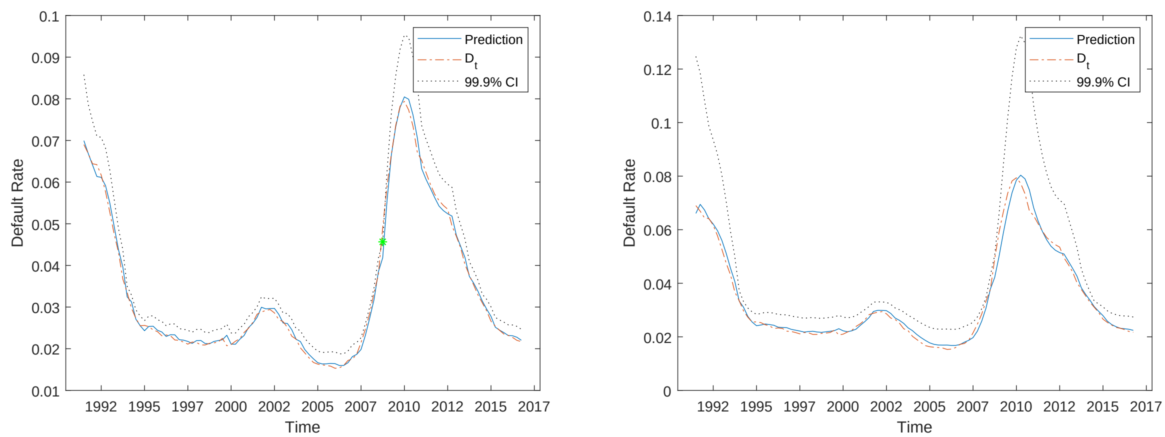
Table 5.9: Out-of-sample estimation result of t-SDM-AR(2)-LS, t-SDM-AR(2)-S and SDM-AR(2).

According to Table 5.9, there does not appear to be any relationship or pattern among estimates produced by the different t-SDM-AR(2) models. The following two tables show the accuracy measures for the training and test set.

Series	Model	Training Set			Testing Set		
Unit(%)		MAE	RMSE	MAPE	MAE	RMSE	MAPE
ALL	t-SDM-AR(2)-S	0.07	0.1	2.81	0.11	0.17	2.85
	t-SDM-AR(2)-L	0.17	0.22	7.05	0.27	0.37	6.37
	t-SDM-AR(2)-LS	0.2	0.24	7.97	0.29	0.4	7.07
	SDM-AR(2)	0.12	0.16	4.53	0.11	0.17	2.94
OC	t-SDM-AR(2)-S	0.09	0.12	2.84	0.09	0.12	3.5
	t-SDM-AR(2)-L	0.29	0.35	8.73	0.29	0.43	9.63
	t-SDM-AR(2)-LS	0.51	0.59	17.51	0.71	0.79	30.74
	SDM-AR(2)	0.08	0.1	2.64	0.09	0.12	3.44
CRE	t-SDM-AR(2)-S	0.21	0.31	5.92	0.2	0.3	5.48
	t-SDM-AR(2)-L	0.74	1.04	17.7	2.85	4.02	54.08
	t-SDM-AR(2)-LS	0.27	0.38	8.27	0.56	0.81	12.67
	SDM-AR(2)	0.17	0.26	5.35	0.18	0.27	4.78

Table 5.10: Accuracy measurements for the training and testing sets.

Based on Table 5.10, it is not surprising to notice that some predictions for the testing set perform worse than that for the training set. But the t-SDM-AR(2)-S and SDM-AR(2) models work consistently well for both training and testing sets. By combining the results from the In-sample test, it is easy to notice that the t-SDM-AR(2)-S and SDM-AR(2) models have better ability in describing the Federal Reserve Data. The following plots show the predictions based on the t-SDM-AR(2)-S model for the ALL series with different degrees of freedom.



(a) Degree of freedom $\nu = 319$. (b) Degree of freedom $\nu = 30$.

Figure 5.7: Prediction based on the t-SDM-AR(2)-S model for the ALL series

The green star mark in the sub-figure (a) of Figure 5.7 represents the point that breaches the prediction interval. The sub-figure (b) is generated with the same parameter values but the degree of freedom is set to be 30 to generate a broader forecasting interval. By doing so, we obtain a more conservative interval prediction so that the breached point is included.

ν	MAE(%)	RMSE(%)	MAPE(%)
319	0.11	0.17	2.85
30	0.21	0.31	4.6

Table 5.11: Accuracy measurements of t-SDM-AR(2)-S for the testing set of ALL series with different degrees of freedom.

As we can see from Table 5.11, lowering the degree of freedom does increase the prediction errors. This result shows that by adding the $S_{t,\nu}$ into the SDM-AR(2) model allows us to control the conservative level of the interval prediction.

5.8 Economic capital comparison

The traditional way of calculating the Economic Capital relies on the Vasicek model we have presented in Section 1.2. Banks and companies fit the model defined by Equation 1.2 to historical default rate series and then calculate the 99.9% quantile of the default rate distribution in the form of Equation 1.6. Since the calculation of the regulatory capital involves the value of LGD and EAD, which is absenct in our data set, we simply define the economic capital as the 99.9% quantile of the default rate which is also an important capital that banks usually pay attention to.

The underlying temporal independence assumption of the Vasicek model makes the economic capital a static value all the time. In our study, we have relaxed this assumption by assuming that the systematic risk factor has some time-dependence structure. As a result of introducing temporal dependence to the systematic risk factor, we can predict the 99.9% quantile of the default rate for next time period condition on current default rate. We name such prediction as dynamic economic capital in contrast to the static one that the Vasicek model provides, which is independent to current default rate. In addition to that, the models that we have proposed also can calculate an overall static economic capital, since the systematic risk factor process has a stationary distribution.

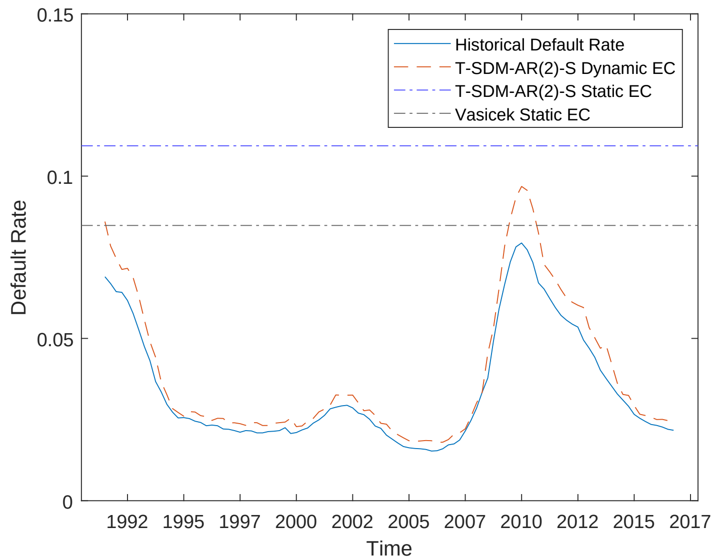


Figure 5.8: Economic capital comparison

For a more concrete example, Figure 5.8 shows the static economic capital of the Vasicek and t-SDM-AR(2)-S models along with the dynamic economic capital based on the t-SDM-AR(2)-S model for the ALL series in the Federal Reserve Data.

It is clear from Figure 5.8 that the economic capital based on dynamic models is lower than those based on a static model. This makes sense since when the market is in a relatively good scenario, it is not necessary to hold an excess of the economic capital. Also, we can notice from the plot that when the default rate is way above its average, the dynamic economic capital of the t-SDM-AR(2)-S exceeds the static economic capital of the Vasicek model.

For the dynamic EC, the traditional AR(2) model also possesses similar ability. The following figure directly compares the t-SDM-AR(2)-S and AR(2) models' dynamic EC.

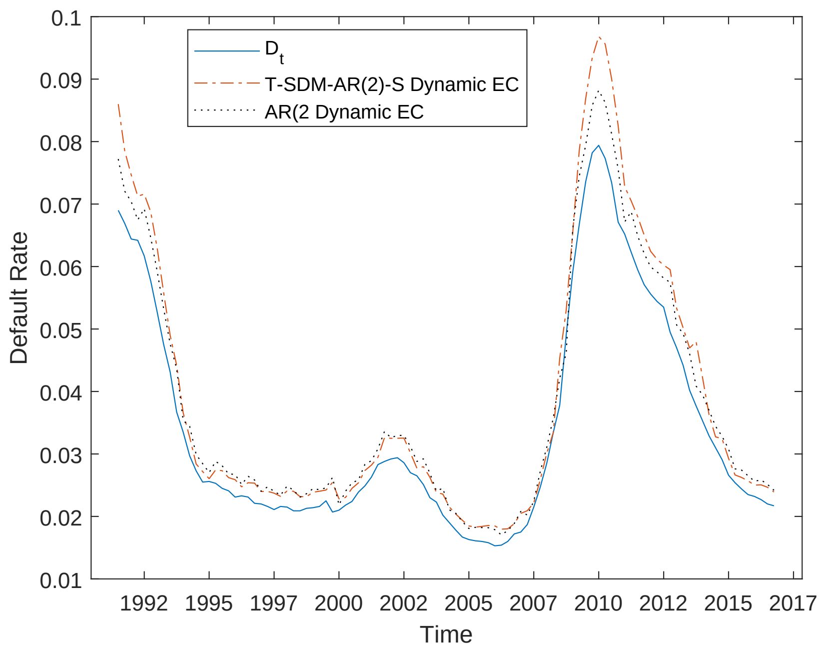


Figure 5.9: Dynamic Economic capital comparison between t-SDM-AR(2) and AR(2) models

Based on Figure 5.9, we can readily notice two features of the t-SDM-AR(2)-S's dynamic EC. First, the t-SDM-AR(2)-S's dynamic EC exceeds the one from the AR(2) model when the market is bearish. Secondly, both dynamics of the Economic Capital are close with each other when the default rate is relatively low. These two features suggest that the t-SDM-AR(2)-S model would be more conservative than the AR(2) model at a stressed market level, and its performance is similar to the AR(2) model when the market is moderate.

Chapter 6

Conclusion and Future Research Directions

In this chapter, we will summarize our contributions and talk about some potential further research directions. All in all, the conclusion can be divided into the following parts. In the first part, we investigate some basic properties of the State-dependent model (SDM) proposed by Metzler (2020), which is an extension of the classic Vasicek default model. In the main part of the thesis, we propose an extension of the SDM model, called the Correlated State-dependent model, abbreviated as SDM-AR, based on the SDM model by introducing an autoregressive structure into the dynamic of the latent systematic risk factor. This approach will bring the forecasting ability to the model. In our empirical study, we demonstrate that the proposed approach significantly improves the forecasting ability of the SDM model. In the last part, we further extend the SDM-AR model to replace the underlying distribution from a normal distribution to t-distribution.

Our research has raised some interesting questions, some of which we are planning to address in the future. Although we have given a quick look at the continuous factor loading, further work is needed to fully understand the implications of it. The other potential area of research is developing a proper methodology to deal with the non-stationarity of the Federal Reserve Data. Also, extending the AR(2) structure used to model the systematic risk factor $\{M_t\}$ is worthy of some endeavor. The other source of weakness of the proposed approach that could which could restrain the practicality of the model is its computational complexity.

6.1 Contribution summary

The main goal of this study was to address some well-known drawbacks of the Vasicek default model. The first one is the underlying assumption that the implied correlations among the loans do not depend on the overall market level and remain constant all the time. In order to relax this assumption, the state-dependence structure is introduced.

The main component of the Vasicek default model is the credit score defined as

$$X_i = aM + \sqrt{1 - a^2} \cdot \epsilon_i.$$

The factor loading a controls the relation between the individual credit score with the latent systematic risk factor M. The credit score is more sensitive to the overall economy when a is larger. The factor loading a remains constant for any level of the overall economy. This contradicts the existing empirical evidence, which shows that the market correlation tends to rise during the bearish market.

With the intention of making the factor loading vary systematically with the overall market level, we employ the idea of changing the factor loading a into a function of a standard normal random variable T, which correlates with the systematic risk factor M with a correlation coefficient β

$$X_i = a(T)M + \sqrt{1 - a(T)^2} \cdot \epsilon_i,$$

where the function of factor loading $a(T)$ has the following form:

$$\begin{aligned} a(t) &= \sum_{k=1}^{K} a_k \cdot \mathbb{1}(t_{k-1} < t \leqslant t_k) \\ 0 &\leqslant a_1 < a_2 < \ldots < a_K \leqslant 1 \\ -\infty &= t_0 < t_1 < \ldots < t_K = \infty. \end{aligned}$$

By defining the model in this form, it is straightforward to notice that the model falls into the Gaussian Mixture category when $\beta = 0$, and Random Factoring proposed by Burtschell, Gregory and Laurent (2005) when $\beta = \pm 1$. Also, by setting $a_K = a_{K-1} = \ldots = a_1$, the model regresses back to the traditional Vasicek model.

After developing the EM-algorithm in Chapter 2, we perform an empirical study using the Federal Reserve Data with the SDM and Vasicek models. By comparing the estimation results of those two models, we realize that the correlation between the systematic risk factor M and factor loading a plays an important role in capturing the probability of extreme events. For all ten historical data series that we consider in this study, seven of them indicate that the estimate of β is close to 1 and the rest of them have β close to -1. Furthermore, when we focus on the calculation of Regulatory Capital, the fat tail brought by the state-dependent structure suggests that Regulatory Capital should be more than the amount estimated based on the traditional Vasicek model.

In Chapters 3 and 4 of the thesis, we establish a framework for introducing the forecasting ability to the SDM model by adopting an autoregressive model for the systematic risk factor $\{M_t\}$. We call this model state-dependent model with an underlying autoregression process, shortened as SDM-AR. The model is defined in the following form:

$$\begin{aligned} X_{i,t} &= a(T_t)M_t + \sqrt{1 - a(T_t)^2} e_{i,t} \\ M_t &= \theta_1 M_{t-1} + \epsilon_t \text{ or } \theta_1 M_{t-1} + \theta_2 M_{t-2} + \epsilon_t \\ T_t &= \beta M_t + \epsilon'_t, \end{aligned}$$

where M_t is modeled by the AR(1) and AR(2) processes for the SDM-AR(1) and SDM-AR(2) models, respectively. Also, the variance of ϵ_t and ϵ'_t are defined such that remaining the stationary distribution of $\{M_t\}$ and $\{T_t\}$ follow standard normal distributions.

We also show the fact that the SDM-AR model can be treated as a general extension of the Vasicek, SDM and AR models. Through the filtering procedure that we have developed for the SDM-AR model, we are able to estimate the model using the maximum likelihood method and to forecast future observations.

The maximum likelihood method is applied to the Federal Reserve Data to estimate the models AR(1), AR(2), SDM-AR(1) and SDM-AR(2) models, separately. In addition to that, the in-sample and out-of-sample tests are also employed to verify the goodness-of-fit of each model. The empirical results suggest that the role of both two-lag and state-dependence structure are important.

For forecasting, we have concentrated on both the point and interval forecasts. The comparison of the out-of-sample point forecasts for the AR(2) and SDM-AR(2) models indicate that the improvement brought by the state-dependence structure is around 20% in lowering the mean absolute error and other accuracy measures. By comparing results based on the SDM-AR(1) and SDM-AR(2) models, we have found strong evidence in favor of using the two-lag structure. Using this structure is especially important during the financial crisis, as all other models consistently under-predict the market risk except the SDM-AR(2) model.

The advantage of the SDM-AR(2) model becomes more considerable when we focus on the interval forecasting. At the beginning of the financial crisis, there exist some observations which are only contained within the forecasting confidence interval generated by the SDM-AR(2) model.

In summary, these discoveries advocate the notable impact of state-dependence and the two-lag structure in enhancing the default rate forecasting.

The models that we propose in Chapters 3 and 4 are driven by a normal distribution. However, it is well-known that tails of the normal distribution are not heavy enough to properly describe the probabilities of extreme events. The main aim of the study presented in Chapter 5 is to overcome this drawback by changing the normal distribution to a t-distribution. To achieve this goal, we introduce a random variable called a tail thickness factor, $S_{t,\nu}$, into the credit score of the SDM-AR(2) model. We call this model T-driven State-dependent model with an underlying autoregression process, and denote it as t-SDM-AR(2). To be more specific, we define three versions of the model. The only differences are in the definition of the credit score in each model

$$X_{i,t} = \begin{cases} \sqrt{\frac{\nu-2}{\nu}} S_{t,\nu} \left[a(T_t) M_t + \sqrt{1 - a(T_t)^2} e_{i,t} \right] & \text{For t-SDM-AR(2)-LS} \\ a(T_t) \sqrt{\frac{\nu-2}{\nu}} S_{t,\nu} M_t + \sqrt{1 - a(T_t)^2} e_{i,t} & \text{For t-SDM-AR(2)-S} \\ a(T_t) M_t + \sqrt{1 - a(T_t)^2} \sqrt{\frac{\nu-2}{\nu}} S_{t,\nu} e_{i,t} & \text{For t-SDM-AR(2)-L,} \end{cases}$$

$$S_{t,\nu} = \sqrt{\frac{\nu}{W_t}}$$

where $\{W_t\}$ follows an independent Chi-squared distribution with the degree of freedom ν. All three t-SDM-AR(2) models can be considered as extensions for the SDM-AR(2) model due to the fact that the factor $\sqrt{\frac{\nu-2}{\nu}} S_{t,\nu}$ converges to 1 when $\nu \to \infty$.

The empirical studies of the t-SDM-AR(2) models show that the t-SDM-AR(2)-S model is the most suitable t-SDM-AR(2) model in modelling the Federal Reserve Data. But the improvements in point and interval predictions are still not substantial enough compared with the SDM-AR(2). However, t-SDM-AR(2) models enable us to control the conservative level of the predictive confidence interval by changing the value of the degree of freedom. Although the t-SDM-AR(2) models do not show notable advantages for the Federal Reserve Data, they may be useful to model other data with relatively higher variation and lower auto-correlation.

Last but not least, both the SDM-AR(2) and t-SDM-AR(2) models provide us with a framework to compute the regulatory capital dynamically. Also from the perspective of the regulators, our models can capture the pro-cyclicality of the default rate by introducing the temporal dependence.

6.2 Future research directions

There are still some unexplored aspects of the models that we have proposed.

- As we mentioned the in Section 1.2.2, the absence of randomness of LGD is also a potential research directions. A lot of researchers have developed varied models to capture the randomness of the LGD. How to properly absorb those models into our dynamic SDM-AR models shall also be an area which worth some effort to study.

- All the models in this study are proposed under the assumption that the factor loading is in the form of a simple discrete function with respect to a random variable correlated with the systematic risk factor. Although we give a quick peek about changing the factor loading function into a continuous form as

$$a_{\alpha_1,\alpha_2}(t) = \Phi\left(\frac{\alpha_1 - T}{\alpha_2}\right),$$

  where α_1 and α_2 are two parameters and $\Phi()$ is the cumulative distribution function of the standard normal distribution. But, estimated parameters of such a model on the Federal Reserve Data give fail to indicate the existence of different regimes. Considerably more work will need to be done to clearly determine the reason for the non-existence of regimes implied by the continuous factor loading function. Other forms of continuous factor loading also deserve some attention.

- The issue of non-stationary is a thought-provoking problem which should be carefully explored in further research as well. We can readily notice the non-stationarity of the historical data from the plot. This obviously violates the underlying assumption of the models we have proposed in this study that the systematic risk factor follows a stationary autoregression model. We also rely on the stationary assumption of the systematic risk factor to define the default threshold x_{PD} which works as the trigger value of default. How to properly define the new credit threshold if we adopt a non-stationary process to model the systematic risk factor is an essential problem we have to investigate.

- The other unsolved challenge we have faced in this study is the computational complexity of the t-SDM-AR(2) models. Under those models, the likelihood function involves some three-dimensional integration. It is infamous that the numerical solver for such a integration in Matlab is unstable and excessively time-consuming. In order to reduce the computational time, we create 3-D mesh grid over the space and evaluate the value of the integrand in each of the point. Then we use this result to approximate the integration. The side effect of reducing the computational time by such a method is at the expense of accuracy. Developing some efficient methodology to evaluate the likelihood value of t-SDM-AR(2) model should be a potential further study.

- In this study, we concentrate on the default rate of the Federal Reserve Data only. A natural approach to improve forecasting performance is to introduce additional explanatory variables that have high correlation with the default rate. How to suitably select those variables deserves a further study.

References

[1] Albanese, Claudio, Li, David, Lobachevskiy, Edgar, and Meissner, Gunter. A comparative analysis of correlation approaches in finance. *The Journal of Derivatives*, 21(2):42–66, 2013.

[2] Andersen, Leif and Sidenius, Jakob. Extensions to the Gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk Volume*, 1(1):05, 2004.

[3] Bai, Jushan and Ng, Serena. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

[4] Balla, Eliana, Ergen, Ibrahim, and Migueis, Marco. Tail dependence and indicators of systemic risk for large US depositories. *Journal of financial Stability*, 15:195–209, 2014.

[5] Basel., Committee. *An explanatory note on the Basel II IRB risk weight functions.* Technical report, Bank for International Settlements., 2005.

[6] Bassamboo, Achal, Juneja, Sandeep, and Zeevi, Assaf. Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*, 56(3):593–606, 2008.

[7] Batiz-Zuk, Enrique, Christodoulakis, George, and Poon, Ser-Huang. Structural Credit Loss Distributions under Non-Normality. *The Journal of Fixed Income*, 23(1):56–75, 2013.

[8] Bellotti, Tony and Crook, Jonathan. Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29(4):563–574, 2013.

[9] Bharath, Sreedhar T and Shumway, Tyler. Forecasting default with the KMV-Merton model. In *AFA 2006 Boston Meetings Paper*, 2004.

[10] Bharath, Sreedhar T and Shumway, Tyler. Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, 21(3):1339–1369, 2008.

[11] Breitung, Jörg and Eickmeier, Sandra. Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84, 2011.

[12] Buckley, Ian, Saunders, David, and Seco, Luis. Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research*, 185(3):1434–1461, 2008.

[13] Burtschell, Xavier, Gregory, Jonathan, and Laurent, Jean-Paul. A comparative analysis of CDO pricing models under the factor copula framework. *The Journal of Derivatives*, 16(4):9–37, 2009.

[14] Campbell, Rachel, Koedijk, Kees, and Kofman, Paul. Increased correlation in bear markets. *Financial Analysts Journal*, 58(1):87–94, 2002.

[15] Cespedes, Juan Carlos Garcia, de Juan Herrero, Juan Antonio, Kreinin, Alex, and Rosen, Dan. A simple multi-factor "factor adjustment" for the treatment of credit capital diversification. *Journal of Credit Risk*, 2(3):57–85, 2006.

[16] Chamroukhi, Faicel, Samé, Allou, Govaert, Gérard, and Aknin, Patrice. Time series modeling by a regression approach based on a latent process. *Neural Networks*, 22(5-6):593–602, 2009.

[17] Chatfield, Chris and Xing, Haipeng. *The Analysis Of Time Series: An Introduction With R*. CRC press, 2019.

[18] Chelimo, John Kigen. *Calibration of vasicek model in a hidden markov context: The case of Kenya*. PhD thesis, Strathmore University, 2017.

[19] Cheng, Xu, Liao, Zhipeng, and Schorfheide, Frank. Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4):1511–1543, 2016.

[20] Christoffersen, Peter and Pelletier, Denis. Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108, 2004.

[21] Crook, Jonathan and Moreira, Fernando. Checking for asymmetric default dependence in a credit card portfolio: A copula approach. *Journal of Empirical Finance*, 18(4):728–742, 2011.

[22] Duffie, Darrell, Eckner, Andreas, Horel, Guillaume, and Saita, Leandro. Frailty correlated default. *The Journal of Finance*, 64(5):2089–2123, 2009.

[23] Düllmann, Klaus and Gehde-Trapp, Monika. Systematic risk in recovery rates: An empirical analysis of US corporate credit exposures. *Bundesbank Series 2 Discussion Paper No.*, 2004.

[24] Dwyer, Douglas W. The distribution of defaults and Bayesian model validation. *Journal of Risk Model Validation*, 1(1):23–53, 2007.

[25] Elizalde, Abel and Repullo, Rafael. Economic and regulatory capital in banking: What is the difference? *Tenth issue (September 2007) of the International Journal of Central Banking*, 2018.

[26] Elouerkhaoui, Youssef. *Credit Correlation*. Springer, 2017.

[27] Escobar, Marcos, Frielingsdorf, Tobias, and Zagst, Rudi. Impact of factor models on portfolio risk measures: A structural approach. *Ryerson Applied Mathematics Laboratory. Technical Report*, 2010.

[28] Frey, Rüdiger and McNeil, Alexander J. Dependent defaults in models of portfolio credit risk. *Journal of Risk*, 6:59–92, 2003.

[29] Frey, Rüdiger, McNeil, Alexander J, and Nyfeler, Mark. Copulas and credit models. *Risk*, 10(111114.10), 2001.

[30] Ghahramani, Zoubin and Jordan, Michael I. Factorial hidden Markov models. *Machine learning*, 29(2):245–273, 1997.

[31] Giesecke, Kay. Credit risk modeling and valuation: An introduction. *Available at SSRN 479323*, 2004.

[32] Gordy, Michael B. A comparative anatomy of credit risk models. *Journal of Banking & Finance*, 24(1-2):119–149, 2000.

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

[34] Gregory, X Burtschell1 J and Laurent, JP. Beyond the Gaussian copula: Stochastic and local correlation. Technical report, Working Paper, BNP Paribas, 2005.

[35] Hallin, Marc and Lippi, Marco. Factor models in high-dimensional time series—a time-domain approach. *Stochastic processes and their applications*, 123(7):2678–2695, 2013.

[36] Hamilton, James D. Regime switching models. In *Macroeconometrics and time series analysis*, pages 202–209. Springer, 2010.

[37] Hamilton, James Douglas. *Time Series Analysis*. Princeton university press, 2020.

[38] Holzmann, Hajo, Munk, Axel, Suster, Max, and Zucchini, Walter. Hidden Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, 13(3):325–347, 2006.

[39] Hull, John C and White, Alan D. Valuing credit default swaps II: Modeling default correlations. *The Journal of derivatives*, 8(3):12–21, 2001.

[40] Jaworski, Piotr, Durante, Fabrizio, Hardle, Wolfgang Karl, and Rychlik, Tomasz. *Copula Theory And Its Applications*, volume 198. Springer, 2010.

[41] Jeanblanc, Monique and Rutkowski, Marek. Modeling of default risk: An overview. *Mathematical finance: Theory and practice*, pages 171–269, 2000.

[42] Jiménez, Gabriel and Mencia, Javier. Modelling the distribution of credit losses with observable and latent factors. *Journal of Empirical Finance*, 16(2):235–253, 2009.

[43] Kim, Chang-Jin. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22, 1994.

[44] Koopman, Siem Jan and Lucas, André. A non-Gaussian panel time series model for estimating and decomposing default risk. *Journal of Business & Economic Statistics*, 26(4):510–525, 2008.

[45] Kupiec, Paul. How Well Does the Vasicek-Basel Airb Model Fit the Data? Evidence from a Long Time Series of Corporate Credit Rating Data. *SSRN Electronic Journal*, 11 2009.

[46] Lamb, Robert and Perraudin, William. Dynamic default rates. *Manuscript, University of Minnesota*, 2008.

[47] Lando, David. Credit risk modeling: Theory and applications. 2009.

[48] Laurent, Jean-Paul and Gregory, Jon. Basket default swaps, CDOs and factor copulas. *Journal of risk*, 7(4):103–122, 2005.

[49] Levy, Moshe and Kaplanski, Guy. Portfolio selection in a two-regime world. *European Journal of Operational Research*, 242(2):514–524, 2015.

[50] Li, David X. On default correlation: A copula function approach. *The Journal of Fixed Income*, 9(4):43–54, 2000.

[51] Lopes, Hedibert Freitas and Carvalho, Carlos Marinho. Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137(10):3082–3091, 2007.

[52] McLachlan, Geoffrey J and Krishnan, Thriyambakam. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

[53] Melkuev, David. Asset return correlations in episodes of systemic crises. Master's thesis, University of Waterloo, 2014.

[54] Metzler, Adam. State dependent correlations in the Vasicek default model. *Dependence Modeling*, 8(1):298–329, 2020.

[55] Nelsen, Roger B. *An Introduction To Copulas.* Springer Science & Business Media, 2007.

[56] Pelger, Markus and Xiong, Ruoxuan. State-varying factor models of large dimensions. *arXiv preprint arXiv:1807.02248*, 2018.

[57] Petris, Giovanni and An, R. An R package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16, 2010.

[58] Petris, Giovanni, Petrone, Sonia, and Campagnoli, Patrizia. Dynamic linear models. In *Dynamic Linear Models with R*, pages 31–84. Springer, 2009.

[59] Pimbley, Joseph M. T-Vasicek credit portfolio loss distribution. *The Journal of Structured Finance*, 24(3):65–78, 2018.

[60] Prado, Raquel and West, Mike. *Time Series: Modeling, Computation, And Inference.* CRC Press, 2010.

[61] Puza, Borek. *Bayesian Methods For Statistical Analysis.* ANU Press, 2015.

[62] Repullo, Rafael and Suarez, Javier. The procyclical effects of bank capital regulation. *The Review of financial studies*, 26(2):452–490, 2013.

[63] Rösch, Daniel. An empirical comparison of default risk forecasts from alternative credit rating philosophies. *International Journal of Forecasting*, 21(1):37–51, 2005.

[64] Rutkowski, Marek and Tarca, Silvio. Regulatory capital modeling for credit risk. *International Journal of Theoretical and Applied Finance*, 18(05):1550034, 2015.

[65] Salmon, Felix. The formula that killed Wall Street. *Significance*, 9(1):16–20, 2012.

[66] Schloegl, Lutz and O'Kane, Dominic. A note on the large homogeneous portfolio approximation with the Student-t copula. *Finance and Stochastics*, 9(4):577–584, 2005.

[67] Schönbucher, Philipp. Taken to the limit: Simple and not-so-simple loan loss distributions. *The Best of Wilmott*, 1:143–160, 2002.

[68] Schönbucher, Philipp J. Factor models for portofolio credit risk. Technical report, Bonn Econ Discussion Papers, 2000.

[69] Shumway, Robert H and Stoffer, David S. *Time Series Analysis And Its Applications*, volume 3. Springer, 2000.

[70] Simons, Dietske and Rolwes, Ferdinand. Macroeconomic default modeling and stress testing. *Eighteenth issue (September 2009) of the International Journal of Central Banking*, 2018.

[71] Sinopoli, Bruno, Schenato, Luca, Franceschetti, Massimo, Poolla, Kameshwar, Jordan, Michael I, and Sastry, Shankar S. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–1464, 2004.

[72] Su, Liangjun and Wang, Xia. On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101, 2017.

[73] Suchintabandid, Sira. Modeling Term Structure of Default Correlation. *The Journal of Derivatives*, 22(4):26–36, 2015.

[74] Tarashev, Nikola and Zhu, Haibin. Specification and calibration errors in measures of portfolio credit risk: The case of the ASRF model. *Thirteenth issue (June 2008) of the International Journal of Central Banking*, 2018.

[75] Van Ravenzwaaij, Don, Cassey, Pete, and Brown, Scott D. A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic bulletin & review*, 25(1):143–154, 2018.

[76] Vasicek, Oldrich. The distribution of loan portfolio value. *Risk*, 15(12):160–162, 2002.

[77] Zucchini, Walter, MacDonald, Iain L, and Langrock, Roland. *Hidden Markov Models For Time Series: An Introduction Using R.* CRC press, 2017.

Appendix A

Estimation with assumption $\theta = 0$

In this section, we wish to introduce a new estimation method for the SDM-AR(1) model which can reduce the amount of the computational workload. We would like to apply an iterative procedure to get the estimators to approach to the true value step by step.

(i) The first thing we want to do is to get some rough idea about the parameters a_1, a_2, t_1, β.

(ii) After that, we can make some inference about the latent process M_t by the filtering procedure described at the beginning of Section 3.2 based on the guess of a_1, a_2, t_1, β.

(iii) Once we have a potential latent process M_t, it is easier for us to estimate the autocorrelation parameter θ of the process M_t.

(iv) Then we will re-estimate a_1, a_2, t_1, β by maximizing the joint likelihood function of the observable series Y_t but fixed the value of θ to be the estimation result from the last step.

(v) Keep repeating steps (ii) to (iv) until the estimators converges.

The idea behind this procedure is that after each iteration, we should be able to get a more accurate estimation of θ. With the improvement in estimating θ, the estimator of a_1, a_2, t_1, β should be better compared to the result obtained from a less accurate estimator of θ.

In order to have a rough idea about the parameters a_1, a_2, t_1, β first, we will start with the EM-algorithm estimation procedure under the assumption that $\theta = 0$. Since we have pre-fixed the stationary distribution of M_t and T_t to be standard normal distribution with the correlation coefficient β, we believe that after enough time, the stationary distribution of the observations Y_t will be only dependent on a_1, a_2, t_1, β. Given this property, we can first ignore the θ by assuming that $\theta = 0$, then the model goes back to the independent case so that the other parameters can be estimated by the EM-algorithm described in Chapter 1.

After obtaining the estimates $\hat{a_1}, \hat{a_2}, \hat{t_1}, \hat{\beta}$, we can use the smoothing procedure described in the Section 3.2.5 with $\theta = 0$. Then for each time t, we will choose the M_t^* which has the higher value of $P(M_t = m_t | Y_{1:t} = y_{1:t}; \hat{a_1}, \hat{a_2}, \hat{t_1}, \hat{\beta}, \theta = 0)$. After obtaining the time series $\{M_t^*\}$, we can get the MLE $\hat{\theta}$ based on the likelihood function of $f_{M_{T:2}|M_1}(m_{T:2}^* | m_1^*)$. The next step is to re-estimate the a_1, a_2, t_1, β by maximizing the likelihood function of Y_t with a fixed value of $\theta = \hat{\theta}$. Then we just keep repeating these two steps until the estimators converge.

A.1 Joint likelihood function of Y_t

In this section, we want to find out how to calculate the joint likelihood of the observable series Y_t without the assumption that $\theta = 0$. According to the law of total probability, we first decompose the conditional density function of Y_t given $Y_{1:t-1}$, $f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1})$, into the several parts so that it is easy to calculate:

$$
\begin{aligned}
f_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}) =& \sum_{i=1}^{2} \sum_{j=1}^{2} f(R_t = i, R_{t-1} = j, Y_t = y_t | Y_{1:t-1} = y_{1:t-1}) \\
=& \sum_{i=1}^{2} \sum_{j=1}^{2} f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}) \cdot \\
& P(R_t = i | R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}) \cdot P(R_{t-1} = j | Y_{1:t-1} = y_{1:t-1})
\end{aligned} \tag{A.1}
$$

Theorem 15. *The conditional density function of Y_t given R_t, R_{t-1} and $Y_{1:t-1}$, $f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{1:t-1} = y_{1:t-1})$, has the Markovian property with respect to $Y_{1:t-1}$:*

$$
f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}) = f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{t-1} = y_{t-1}) \tag{A.2}
$$

Proof. Before being able to calculate this density function, we need to figure out the relationship between Y_t and Y_{t-1} first. According to formulas 3.2 and 3.8 we have:

$$\begin{aligned} Y_t &= \frac{x_{PD}}{\sqrt{1-a(R_t)^2}} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}(\theta M_{t-1} + \epsilon_t) \\ &= \frac{x_{PD}}{\sqrt{1-a(R_t)^2}} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\theta M_{t-1} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\epsilon_t \end{aligned} \tag{A.3}$$

Also based on Equation (3.8), we know that:

$$M_{t-1} = \frac{x_{PD}}{a(R_{t-1})} - \frac{\sqrt{1-a(R_{t-1})^2}}{a(R_{t-1})}Y_{t-1}.$$

Then substituting this into Equation (A.3),

$$Y_t = \frac{x_{PD}}{\sqrt{1-a(R_t)^2}}(1-\theta\frac{a(R_T)}{a(R_{t-1})}) + \theta\frac{a(R_t)}{a(R_{t-1})}\frac{\sqrt{1-a(R_{t-1})^2}}{\sqrt{1-a(R_t)^2}}Y_{t-1} - \frac{a(R_t)}{\sqrt{1-a(R_t)^2}}\epsilon_t \tag{A.4}$$

As a result, Y_t is conditionally independent of $Y_{t-2:1}$ once given R_t, R_{t-1} and Y_{t-1} since the only random variable, ϵ_t, is the iid error term.

$$f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}) = f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{t-1} = y_{t-1})$$

□

After this we can easily calculate the conditional density function of Y_t given R_t, R_{t-1} and Y_{t-1}:

$$f(Y_t | R_t = i, R_{t-1} = j, Y_{t-1} = y_{t-1}) = \phi(\gamma; 0, 1-\theta^2)(\frac{\sqrt{1-a(R_t)^2}}{a(R_t)})$$

where

$$\gamma = \frac{\sqrt{1-a(R_t)^2}Y_t - x_{PD}(1-\theta\frac{a(R_t)}{a(R_{t-1})}) - \theta\frac{a(R_t)\sqrt{1-a(R_t)^2}}{a(R_{t-1})}Y_{t-1}}{-a(R_t)}.$$

Then, we can proceed to prove that the second term in Equation A.1, $P(R_t = i | R_{t-1} = j, Y_{1:t-1} = y_{1:t-1})$ also has the Markovian property.

Theorem 16. *The conditional probability mass function of* R_t *given* R_{t-1} *and* $Y_{1:t-1}$, $P(R_t = i|R_{t-1} = j, Y_{1:t-1} = y_{1:t-1})$ *has the Markovian property with respect to* $Y_{1:t-1}$.

$$P(R_t = i|R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}) = P(R_t = i|R_{t-1} = j, Y_{t-1} = y_{t-1}) \tag{A.5}$$

Proof. Based on Equation (3.2) and Equation(3.24), we know that:

$$T_t = \beta\theta M_{t-1} + \beta\epsilon_t + \epsilon'_t.$$

We can notice that once the value of the systemic risk factor M_{t-1} is given, then T_t is independent of the other T_i for $\forall i < t$ since both ϵ_t and ϵ'_t are iid error terms. Then we would like to simplify the conditional part of the left side of Equation A.5 into the following form:

$$\begin{aligned}\{R_{t-1} = j \cap Y_{t-1} = y_{t-1}\} &= \{\{T_{t-1} \in [t_j, t_{j+1})\} \cap [\cup_{k=1}^{2}\{T_{t-1} \in [t_k, t_{k+1}), M_{t-1} = M^{-1}(y_{t-1}, k)\}]\} \\ &= \{T_{t-1} \in [t_j, t_{j+1}), M_{t-1} = M^{-1}(y_{t-1}, j)\}\end{aligned}$$

As mentioned before, T_t is independent of the other T_i for $\forall i < t$ once the value of the systematic risk factor M_{t-1} is given. So

$$\begin{aligned}P(R_t = i|R_{t-1} = j, Y_{t-1} = y_{t-1}) =& P(T_t \in [t_i, t_{i+1})|\{T_{t-1} \in [t_j, t_{j+1}), M_{t-1} = M^{-1}(y_{t-1}, j)\}) \\ =& P(T_t \in [t_i, t_{i+1})|M_{t-1} = M^{-1}(y_{t-1}, j))\end{aligned} \tag{A.6}$$

□

The other thing that we need to pay attention is the conditional distribution of T_t given M_{t-1}. It is easy to see that

$$T_t|M_{t-1} = m_{t-1} \sim N(\beta\theta m_{t-1}, 1 - \beta^2\theta^2) \tag{A.7}$$

The last term in Equation A.1 is actually the filtering density function since

$$P(R_{t-1} = j|Y_{1:t-1} = y_{1:t-1}) = P(M_{t-1} = M^{-1}(y_{t-1}, j)|Y_{1:t-1} = y_{1:t-1}).$$

So it can be calculated based on the procedure provided in the previous section.

As a result, the joint likelihood function of Y_t can be calculated based on the following equation

$$f_{Y_{T:2}|Y_1}(y_{T:2}|y_1) = \prod_{i=2}^{T} f_{Y_i|Y_{i-1:1}}(y_i|y_{i-1:1}) \tag{A.8}$$

A.2 Iterative Estimation Algorithm

In this section, we would like to show the estimation procedure described in the Section A for our model in detail. Since we will start with assuming that $\theta = 0$, we can use the EM-algorithm mentioned in Section 2.2 to get a rough guess about a_1, a_2, t_1 and β. Then we use $\hat{a_1}, \hat{a_2}, \hat{t_1}, \hat{\beta}$ to denote the estimation results from the EM-algorithm under the assumption that $\theta = 0$. After that, we can use $\hat{a_1}, \hat{a_2}, \hat{t_1}, \hat{\beta}$ to compute all the possible values of M for each time period. We chose the M_t^* which has the higher value of $P(M_t = M_t^* | Y_{1:t} = y_{1:t}; \hat{a_1}, \hat{a_2}, \hat{t_1}, \hat{\beta}, \theta = 0)$ for each time t. Based on our choices of M_t^*, we can use this new time series to estimate the θ, and use $\hat{\theta}$ to denote the estimation result. After that, we can solve the following optimaiztion problem to re-estimate the other parameters in order to get a more accurate result.

$$\tilde{a_1}^{(1)}, \tilde{a_2}^{(1)}, \tilde{t_1}^{(1)}, \tilde{\beta}^{(1)} = \arg \max_{a_1, a_2, t_1, \beta} \sum_{i=2}^{T} \log f^*_{Y_i|Y_{i-1:1}}(y_i | y_{i-1:1}; a_1, a_2, t_1, \beta) \tag{A.9}$$

where

$$\begin{aligned} f^*_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}; a_1, a_2, t_1, \beta) = & \sum_{i=1}^{2} \sum_{j=1}^{2} f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}; a_1, a_2, t_1, \beta, \hat{\theta}) \cdot \\ & P(R_t = i | R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}; a_1, a_2, t_1, \beta, \hat{\theta}) \cdot \quad \text{(A.10)} \\ & P(R_{t-1} = j | Y_{1:t-1} = y_{1:t-1}; \hat{a_1}, \hat{a_2}, \hat{t_1}, \hat{\beta}, \hat{\theta}). \quad \text{(A.11)} \end{aligned}$$

Once we have the new estimator $\tilde{a_1}^{(1)}, \tilde{a_2}^{(1)}, \tilde{t_1}^{(1)}, \tilde{\beta}^{(1)}$, we can apply the method mentioned at the beginning of this section to get a new estimator of θ and denoted by $\tilde{\theta}^{(1)}$. Once we have $\tilde{\theta}^{(1)}$, we return to optimization problem A.9 to update the estimator to $\tilde{a_1}^{(2)}, \tilde{a_2}^{(2)}, \tilde{t_1}^{(2)}, \tilde{\beta}^{(2)}$ by changing the objective function, $f^*_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}; a_1, a_2, t_1, \beta)$ in the following way,

$$\begin{aligned} f^*_{Y_t|Y_{1:t-1}}(y_t|y_{1:t-1}; a_1, a_2, t_1, \beta) = & \sum_{i=1}^{2} \sum_{j=1}^{2} f(Y_t = y_t | R_t = i, R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}; a_1, a_2, t_1, \beta, \tilde{\theta}^{(1)}) \cdot \\ & P(R_t = i | R_{t-1} = j, Y_{1:t-1} = y_{1:t-1}; a_1, a_2, t_1, \beta, \tilde{\theta}^{(1)}) \cdot \\ & P(R_{t-1} = j | Y_{1:t-1} = y_{1:t-1}; \tilde{a_1}^{(1)}, \tilde{a_2}^{(1)}, \tilde{t_1}^{(1)}, \tilde{\beta}^{(1)}, \tilde{\theta}^{(1)}). \end{aligned}$$

The main idea is to repeat doing this iteration several rounds until the estimators converge.

A.3 Iterative Estimation result

In this section, we wish to use simulated data to verify if the estimation approach works in the way described in the last section. We simulated 2000 data points. Theoretically, we would expect that the estimators, $\tilde{a_1}^{(i)}, \tilde{a_2}^{(i)}, \tilde{t_1}^{(i)}, \tilde{\beta}^{(i)}$ and $\tilde{\theta}^{(i)}$ should move closer and closer to the true values after each iteration. But the simulation test suggest that this estimation approach does not work in the way we expect and actually moves in the opposite way. The following table shows the true values of the parameters used for generating the data.

	a_1	a_2	β	t_1	θ	Fval
True Value	0.38	0.19	0.7	-0.2	0.3	583.677

The next table displays the evaluation of the estimators for each iteration. The Fval column is the log-likelihood value. As we can see, the log-likelihood value also keeps decreasing after each iteration.

The table shows that the procedure provided a relatively good estimation results in the first round. But it quickly moved away from the true values and provided a lower log-likelihood value in the next iteration. After 30 iterations, the estimators do converge to a stable point but the log-likelihood value is obviously lower than the first iteration. The value of each estimators is also far away from the true point. Similar situation happens in another simulation tests. The result suggests that this iterative method does not work properly in calibrating the model. We suspect that the method involves too many rounds of numeric optimization and this is the main reason for the failure. It is hard to guarantee that each optimization can find the global maximum and once some of them are stuck in some local maximum, then the estimators may start to move away from the true value. So we need to find another estimation method in order to find the estimators for the parameters. Since we notice that we can calculate the $f_{Y_i|Y_{i-1:1}}(y_i|y_{i-1:1})$ for $i = 2 \dots T$, we should be able to calculate the joint likelihood of Y_t by using it. So, we will apply the direct ML method instead.

	a_1	a_2	β	t_1	θ	Fval
1	0.380	0.190	0.717	-0.226	0.286	583.979
2	0.353	0.209	0.751	-0.461	0.286	553.062
3	0.352	0.211	0.713	-0.479	0.274	556.265
4	0.352	0.210	0.727	-0.474	0.275	555.017
5	0.352	0.211	0.722	-0.477	0.275	555.453
6	0.352	0.211	0.724	-0.476	0.275	555.275
7	0.352	0.211	0.723	-0.476	0.275	555.347
8	0.352	0.211	0.724	-0.476	0.275	555.315
9	0.352	0.211	0.723	-0.477	0.275	555.314
10	0.352	0.211	0.723	-0.476	0.275	555.315
11	0.352	0.211	0.723	-0.476	0.275	555.318
12	0.352	0.211	0.723	-0.476	0.275	555.318
13	0.352	0.211	0.723	-0.476	0.275	555.319
14	0.352	0.211	0.723	-0.476	0.275	555.319
15	0.352	0.211	0.723	-0.476	0.275	555.319
16	0.352	0.211	0.723	-0.476	0.275	555.319
17	0.352	0.211	0.723	-0.476	0.275	555.319
18	0.352	0.211	0.723	-0.476	0.275	555.319
19	0.352	0.211	0.723	-0.476	0.275	555.319
20	0.352	0.211	0.723	-0.476	0.275	555.319
21	0.352	0.211	0.723	-0.476	0.275	555.319
22	0.352	0.211	0.723	-0.476	0.275	555.319
23	0.352	0.211	0.723	-0.476	0.275	555.319
24	0.352	0.211	0.723	-0.476	0.275	555.319
25	0.352	0.211	0.723	-0.476	0.275	555.319
26	0.352	0.211	0.723	-0.476	0.275	555.319
27	0.352	0.211	0.723	-0.476	0.275	555.319
28	0.352	0.211	0.723	-0.476	0.275	555.319
29	0.352	0.211	0.723	-0.476	0.275	555.319
30	0.352	0.211	0.723	-0.476	0.275	555.319
31	0.352	0.211	0.723	-0.476	0.275	555.319

Appendix B

Calibration results

In this section, we present the empirical results for the SDM-AR models along with the classic AR models.

Series	Model	a_1	a_2	β	t	θ_1	θ_2	σ^2 of AR	C of AR
1	SDM-AR2	0.232	0.162	1	-0.704	1.832	-0.853	0.00613	
	AR2	0.189				1.824	-0.846	0.00024	-0.0415
	SDM-AR1	0.228	0.137	1	-0.634	0.984		0.03135	
	AR1	0.154				0.980		0.00094	-0.0413
2	SDM-AR2	0.395	0.33	-0.918	0.626	1.873	-0.892	0.00408	
	AR2	0.210				1.620	-0.658	0.00120	-0.0818
	SDM-AR1	0.225	0.178	1	0.2988	0.971		0.05608	
	AR1	0.182				0.967		0.00219	-0.0722
3	SDM-AR2	0.341	0.239	0.988	-0.611	1.735	-0.738	0.00157	
	AR2	0.106				1.602	-0.625	0.00019	-0.0431
	SDM-AR1	0.409	0.259	1	-0.695	0.998		0.00310	
	AR1	0.161				0.994		0.00031	-0.0140
4	SDM-AR2	0.613	0.208	0.997	-0.705	1.779	-0.782	0.00130	
	AR2	0.142				1.515	-0.537	0.00042	-0.0409
	SDM-AR1	0.514	0.320	1	-0.717	0.997		0.00424	
	AR1	0.199				0.992		0.00059	-0.0164
5	SDM-AR2	0.115	0.085	1	-0.641	1.366	-0.386	0.02438	
	AR2	0.089				1.340	-0.364	0.00023	-0.0461
	SDM-AR1	0.119	0.088	1	-0.631	0.986		0.02703	
	AR1	0.099				0.986		0.00027	-0.0291
6	SDM-AR2	0.237	0.208	1	-1.614	1.145	-0.209	0.09856	
	AR2	0.216				1.111	-0.173	0.00488	-0.1200
	SDM-AR1	0.233	0.205	1	-1.630	0.945		0.10545	
	AR1	0.213				0.945		0.00503	-0.1064
7	SDM-AR2	0.142	0.132	-1	-1.578	1.210	-0.272	0.08742	
	AR2	0.136				1.189	-0.250	0.00169	-0.1372
	SDM-AR1	0.133	0.124	-1	-1.704	0.944		0.10748	
	AR1	0.132				0.948		0.00180	-0.1179
8	SDM-AR2	0.245	0.234	-0.303	0.6721	1.878	-0.892	0.00302	
	AR2	0.273				1.794	-0.808	0.00042	-0.0253
	SDM-AR1	0.236	0.221	1	1.3232	0.987		0.02452	
	AR1	0.209				0.985		0.00128	-0.0297
9	SDM-AR2	0.181	0.125	-1	2.2554	1.189	-0.233	0.06624	
	AR2	0.182				1.162	-0.205	0.00227	-0.0802
	SDM-AR1	0.177	0.154	-1	-0.897	0.956		0.08481	
	AR1	0.174				0.961		0.00237	-0.0743
10	SDM-AR2	0.334	0.304	-0.794	0.5719	1.820	-0.829	0.00295	
	AR2	0.316				1.658	-0.667	0.00068	-0.0161
	SDM-AR1	0.272	0.245	1	0.9424	0.990		0.01951	
	AR1	0.320				0.994		0.00122	-0.0075
11	SDM-AR2	0.392	0.378	-1	-0.621	1.790	-0.802	0.00477	
	AR2	0.333				1.735	-0.750	0.00090	-0.0305
	SDM-AR1	0.418	0.402	1	0.6943	0.993		0.01222	
	AR1	0.279				0.987		0.00216	-0.0372

Table B.1: Calibration results based on the whole sample set

B.1 SDM-AR(1) Calibration Plots

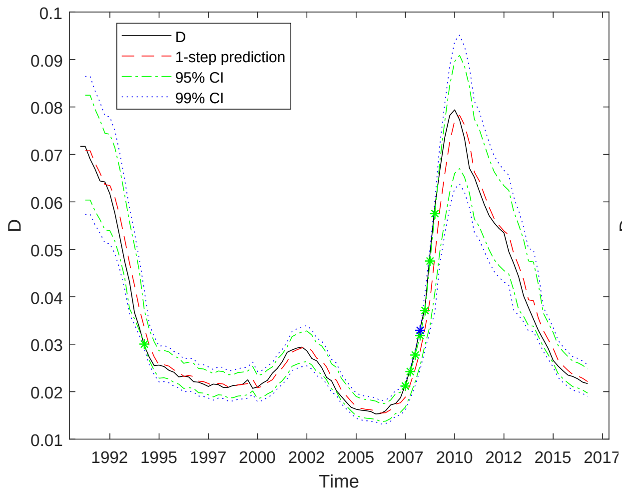


Figure B.1: Fitted value and confidence interval of SDM-AR(1) model for ALL series

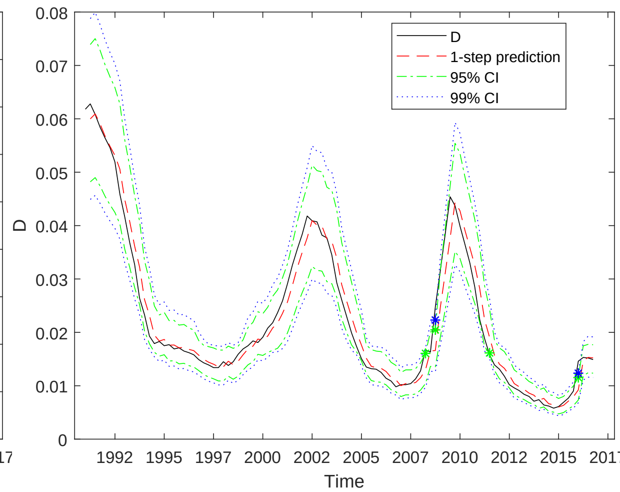


Figure B.2: Fitted value and confidence interval of SDM-AR(1) model for Business series

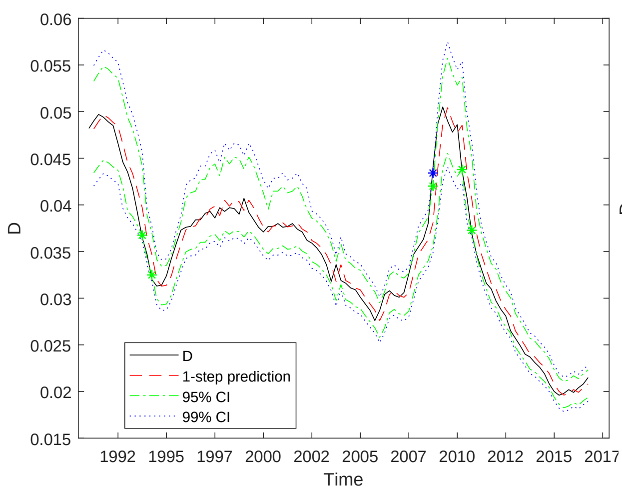


Figure B.3: Fitted value and confidence interval of SDM-AR(1) model for Consumer series

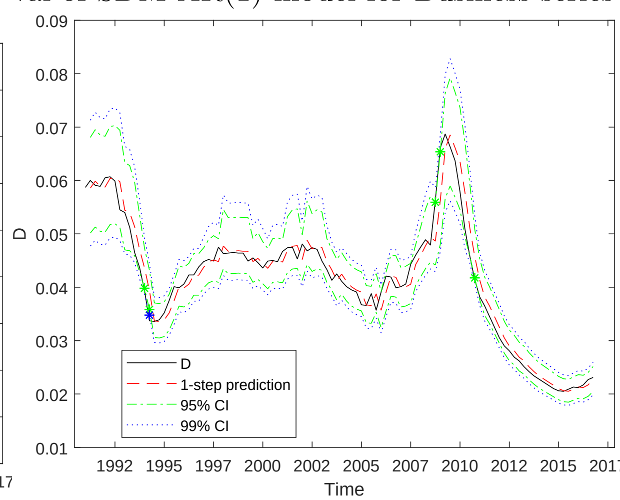


Figure B.4: Fitted value and confidence interval of SDM-AR(1) model for Credit Card series

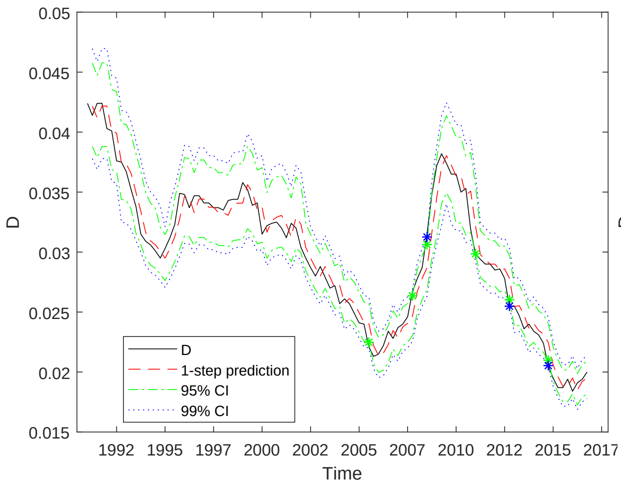


Figure B.5: Fitted value and confidence interval of SDM-AR(1) model for Other Consumer series

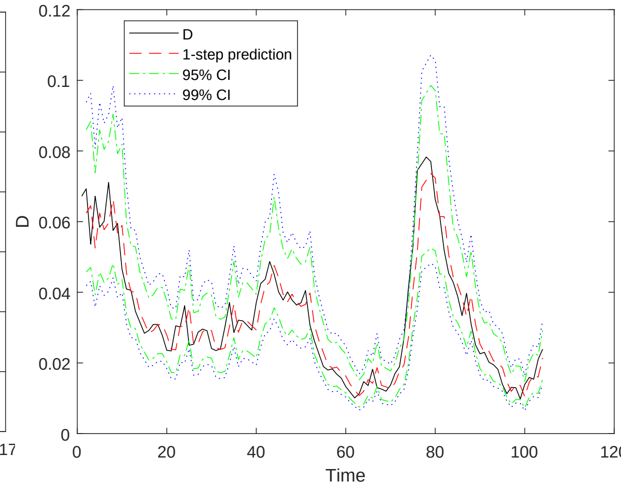


Figure B.6: Fitted value and confidence interval of SDM-AR(1) model for Agricultural series

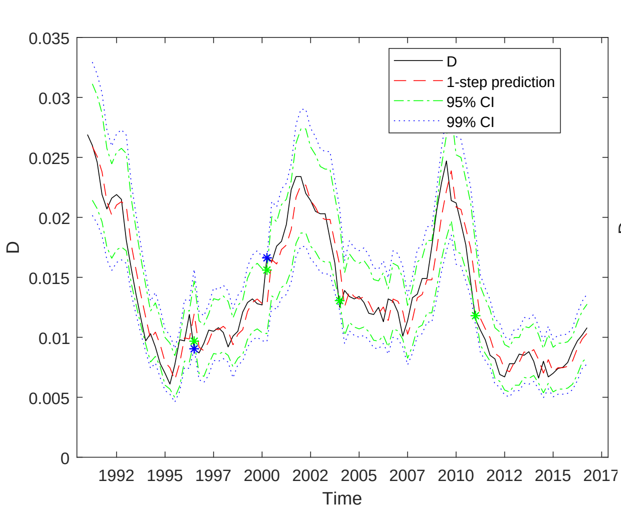


Figure B.7: Fitted value and confidence interval of SDM-AR(1) model for LFR series

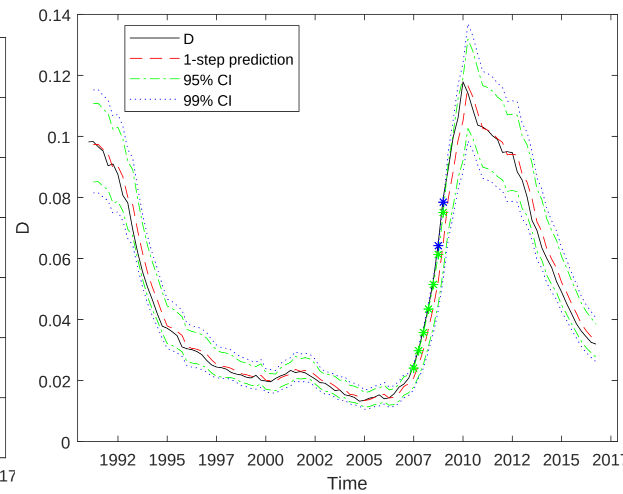


Figure B.8: Fitted value and confidence interval of SDM-AR(1) model for Secured By Real Estate series

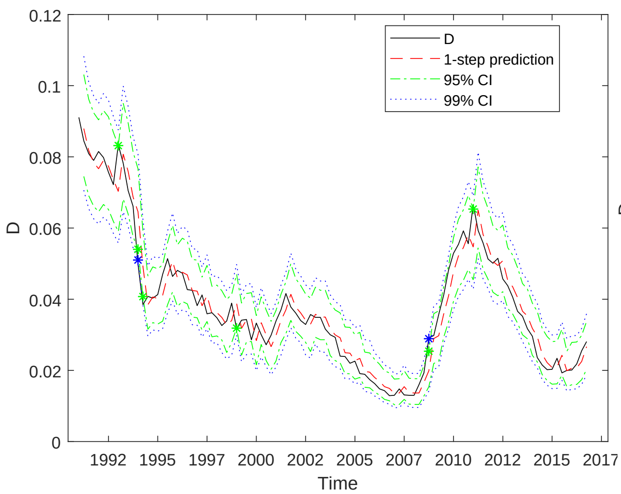


Figure B.9: Fitted value and confidence interval of SDM-AR(1) model for Farmland series

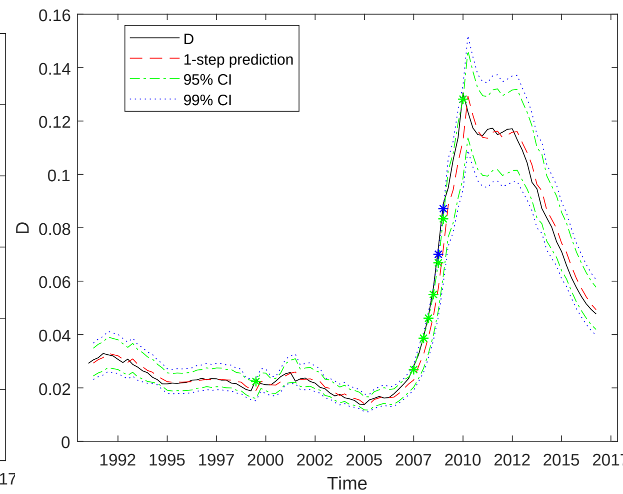


Figure B.10: Fitted value and confidence interval of SDM-AR(1) model for Mortgages series

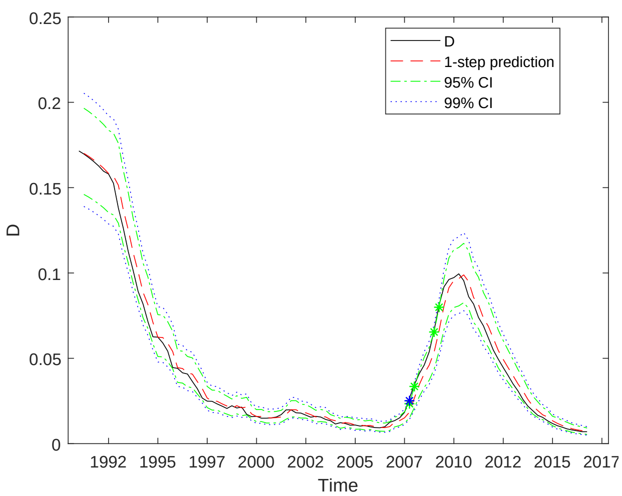


Figure B.11: Fitted value and confidence interval of SDM-AR(1) model for Commercial Real Estate series

B.2 SDM-AR(2) Calibration Plots

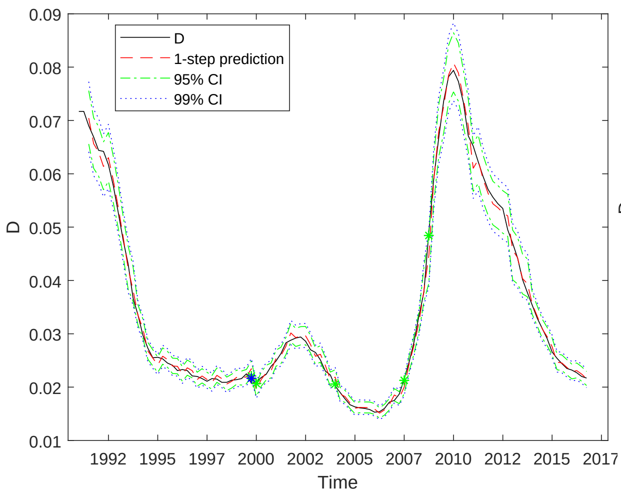


Figure B.12: Fitted value and confidence interval of SDM-AR(2) model for ALL series

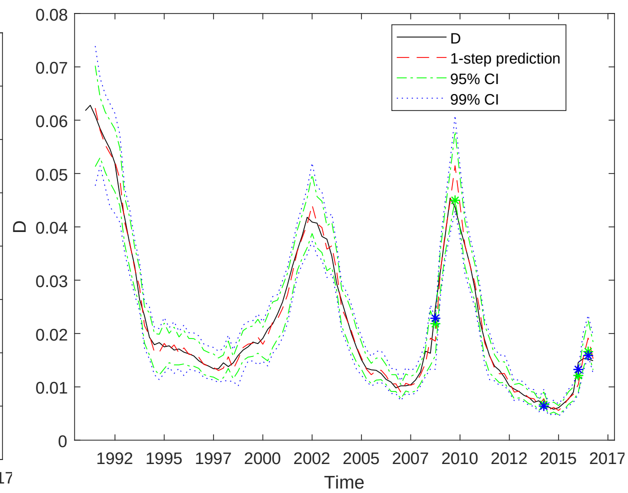


Figure B.13: Fitted value and confidence interval of SDM-AR(2) model for Business series

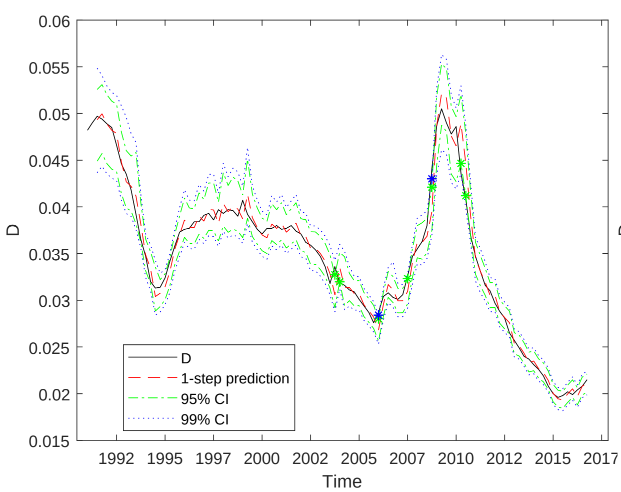


Figure B.14: Fitted value and confidence interval of SDM-AR(2) model for Consumer series

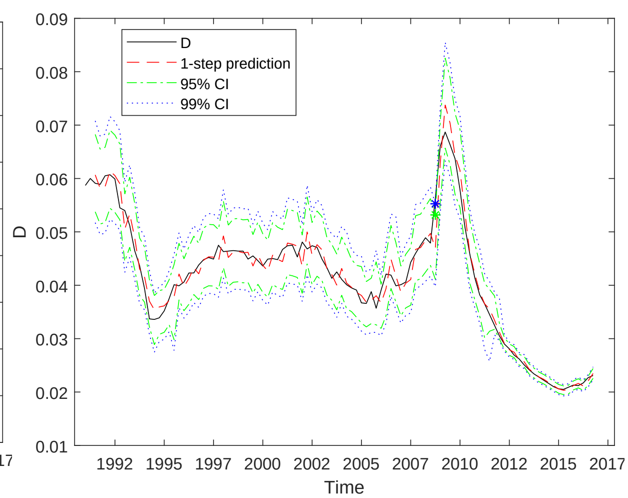


Figure B.15: Fitted value and confidence interval of SDM-AR(2) model for Credit Card series

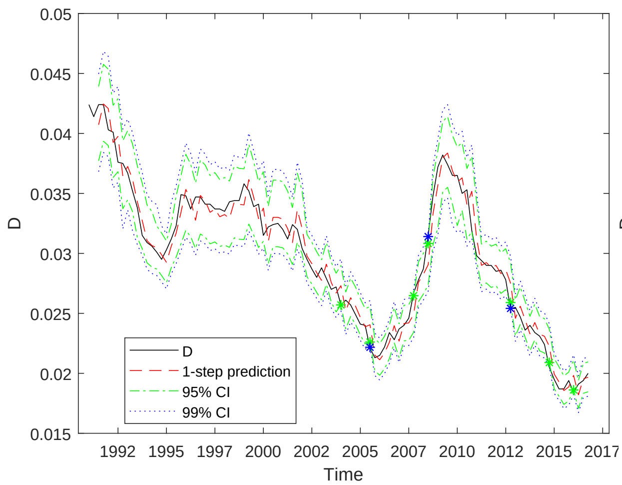


Figure B.16: Fitted value and confidence interval of SDM-AR(2) model for Other Consumer series

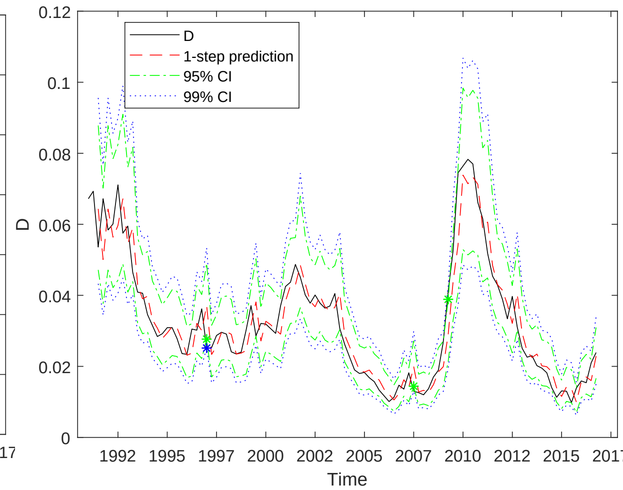


Figure B.17: Fitted value and confidence interval of SDM-AR(2) model for Agricultural series

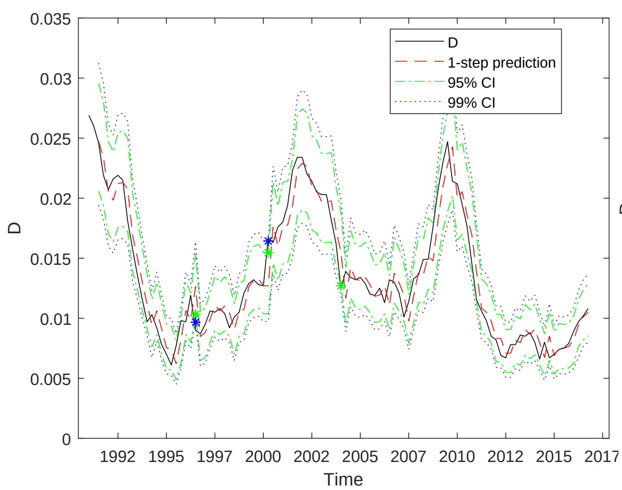


Figure B.18: Fitted value and confidence interval of SDM-AR(2) model for LFR series

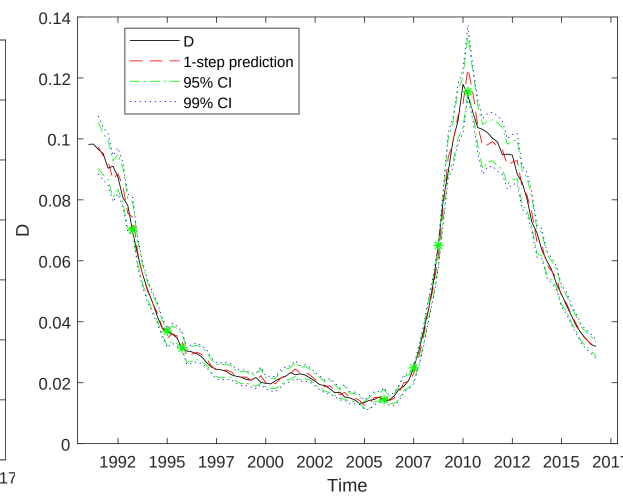


Figure B.19: Fitted value and confidence interval of SDM-AR(2) model for Secured By Real Estate series

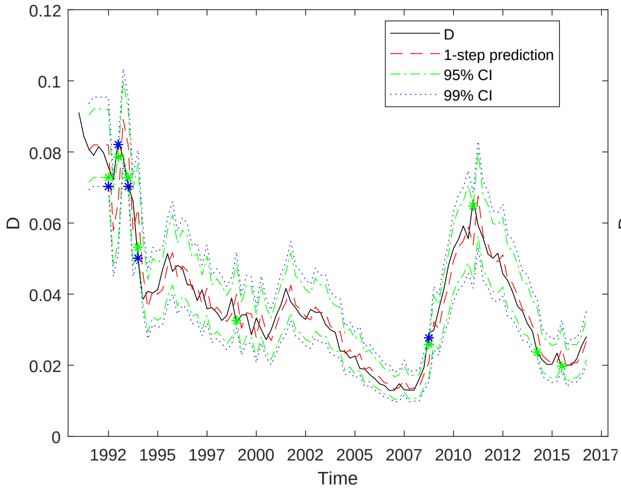


Figure B.20: Fitted value and confidence interval of SDM-AR(2) model for Farmland series

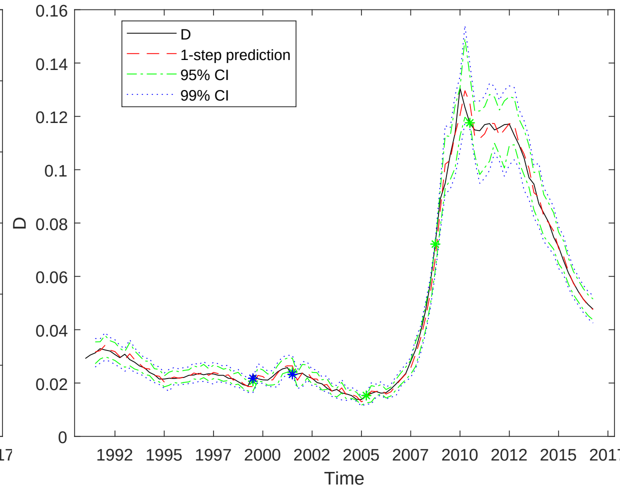


Figure B.21: Fitted value and confidence interval of SDM-AR(2) model for Mortgages series

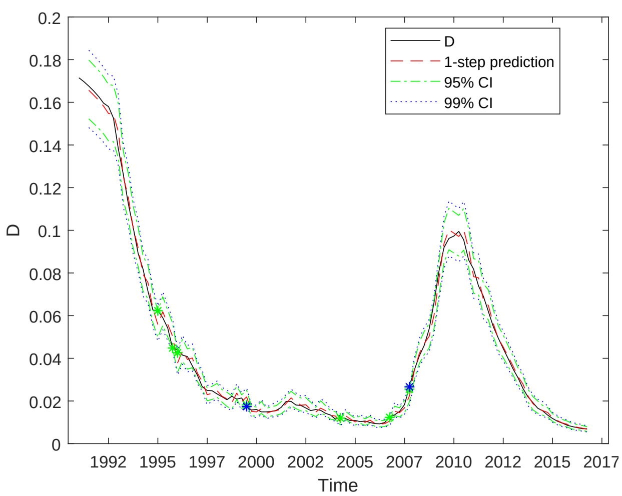


Figure B.22: Fitted value and confidence interval of SDM-AR(2) model for Commercial Real Estate series

Appendix C

Out-of-sample calibration results

In this section, we present the out-of-sample calibration and forecasting results for the SDM-AR models along with the classic AR models.

Series	Model	a_1	a_2	β	t	θ_1	θ_2	σ^2 of AR	C of AR
1	SDM-AR2	0.177	0.143	1	0.693	1.831	-0.848	0.00514	
	AR2	0.108				1.644	-0.67	0.00020	-0.052
	SDM-AR1	0.387	0.189	1	-0.536	0.996		0.00798	
	AR1	0.092				0.975		0.00042	-0.0577
2	SDM-AR2	0.191	0.173	1	1.303	1.819	-0.852	0.00968	
	AR2	0.164				1.757	-0.791	0.00039	-0.0698
	SDM-AR1	0.176	0.157	1	0.867	0.977		0.04547	
	AR1	0.154				0.976		0.00116	-0.0571
3	SDM-AR2	0.070	0.055	-1	-0.949	1.597	-0.633	0.02613	
	AR2	0.073				1.443	-0.467	0.00013	-0.046
	SDM-AR1	0.207	0.178	-1	-0.185	0.998		0.00399	
	AR1	0.130				0.995		0.00017	-0.0129
4	SDM-AR2	0.068	0.056	-1	-1.252	1.315	-0.382	0.08080	
	AR2	0.071				1.288	-0.342	0.00035	-0.0931
	SDM-AR1	0.065	0.054	-1	-1.302	0.947		0.10319	
	AR1	0.071				0.961		0.00039	-0.0695
5	SDM-AR2	0.42	0.386	-1	0.0298	1.176	-0.177	0.00164	
	AR2	0.056				1.467	-0.667	0.00040	-0.3679
	SDM-AR1	0.224	0.169	-1	0.124	0.998		0.00399	
	AR1	0.210				0.998		0.00018	-0.0048
6	SDM-AR2	0.358	0.194	-0.819	0.774	1.369	-0.381	0.01479	
	AR2	0.573				0.879	0.084	0.0386	-0.076
	SDM-AR1	0.368	0.195	-0.771	0.921	0.992		0.01593	
	AR1	0.210				0.957		0.00389	-0.0888
7	SDM-AR2	0.236	0.122	-1	-0.041	1.35243	-0.386	0.04128	
	AR2	0.128				1.1745	-0.241	0.00166	-0.1501
	SDM-AR1	0.349	0.249	-1	0.204	0.992679		0.01458	
	AR1	0.125				0.94297		0.00177	-0.1294
8	SDM-AR2	0.387	0.369	-0.351	1.120	1.860341	-0.863	0.00096	
	AR2	0.120				1.4622	-0.481	0.00029	0.38799
	SDM-AR1	0.361	0.181	1	-0.672	0.994854		0.01026	
	AR1	0.100				0.98013		0.00040	-0.0502
9	SDM-AR2	0.179	0.117	-1	1.961	1.086572	-0.123	0.06442	
	AR2	0.210				0.98507	-0.007	0.00201	-0.05
	SDM-AR1	0.151	0.100	-1	2.425	0.952981		0.09182	
	AR1	0.211				0.97828		0.00201	-0.0495
10	SDM-AR2	0.095	0.065	-1	-0.084	1.129254	-0.169	0.06592	
	AR2	0.093				1.0719	-0.103	0.00049	-0.0667
	SDM-AR1	0.107	0.077	-1	-0.054	0.973007		0.05325	
	AR1	0.096				0.97321		0.00050	-0.0572
11	SDM-AR2	0.930	0.927	-0.942	0.858	1.816297	-0.816	6.08E-05	
	AR2	0.211				1.3289	-0.342	0.00083	-0.0392
	SDM-AR1	0.436	0.421	1	0.740	0.996482		0.00702	
	AR1	0.188				0.98675		0.00096	-0.0450

Table C.1: Four models' estimation results based on $n_1 = 60$ data points, where C is the constant term and σ^2_ϵ is the variance of the error term of the AR(1) and AR(2) processes.

Series	Model	a_1	a_2	β	t	θ_1	θ_2	σ^2 of AR	C of AR
1	SDM-AR2	0.235	0.162	1	-0.672	1.829	-0.851	0.006517	
	AR2	0.195				1.824	-0.846	0.000268	-0.0417
	SDM-AR1	0.224	0.13	1	-0.615	0.982		0.035676	
	AR1	0.151				0.979		0.000982	-0.0425
2	SDM-AR2	0.372	0.235	0.004	-2.288	1.814	-0.846	0.009771	
	AR2	0.226				1.681	-0.713	0.00098	-0.0681
	SDM-AR1	0.255	0.182	1	0.183	0.978		0.043516	
	AR1	0.198				0.976		0.00195	-0.0568
3	SDM-AR2	0.321	0.224	0.986	-0.581	1.742	-0.746	0.00203	
	AR2	0.090				1.597	-0.631	0.000205	-0.0618
	SDM-AR1	0.382	0.245	1	-0.658	0.998		0.003996	
	AR1	0.370				0.999		0.000318	-0.0037
4	SDM-AR2	0.634	0.198	0.996	-0.711	1.828	-0.831	0.001013	
	AR2	0.129				1.515	-0.545	0.00046	-0.0539
	SDM-AR1	0.524	0.339	1	-0.676	0.998		0.003996	
	AR1	0.483				0.999		0.00061	-0.0046
5	SDM-AR2	0.117	0.083	1	-0.594	1.411	-0.433	0.024756	
	AR2	0.072				1.334	-0.369	0.000232	-0.0679
	SDM-AR1	0.136	0.096	1	-0.563	0.989		0.021879	
	AR1	0.117				0.99		0.000278	-0.0219
6	SDM-AR2	0.252	0.221	1	-1.516	1.149	-0.207	0.089778	
	AR2	0.212				1.124	-0.186	0.00466	-0.1189
	SDM-AR1	0.261	0.228	1	-1.48	0.957		0.084151	
	AR1	0.230				0.954		0.00505	-0.0926
7	SDM-AR2	0.220	0.131	-1	0.130	1.372678	-0.411	0.045444	
	AR2	0.136				1.2452	-0.311	0.001694	-0.1498
	SDM-AR1	0.307	0.239	-1	0.289	0.989665		0.020563	
	AR1	0.137				0.94989		0.001876	-0.1148
8	SDM-AR2	0.266	0.254	-0.348	0.645	1.880337	-0.893	0.002873	
	AR2	0.285				1.7902	-0.803	0.000464	-0.0243
	SDM-AR1	0.281	0.264	1	1.012	0.991133		0.017656	
	AR1	0.221				0.98698		0.001341	-0.0253
9	SDM-AR2	0.181	0.121	-1	2.103	1.177185	-0.221	0.067311	
	AR2	0.182				1.1485	-0.190	0.002311	-0.0786
	SDM-AR1	0.182	0.146	-1	-1.146	0.952495		0.092754	
	AR1	0.174				0.96097		0.002401	-0.0737
10	SDM-AR2	0.353	0.320	-0.806	0.694	1.821151	-0.829	0.002823	
	AR2								
	SDM-AR1	0.332	0.299	1	0.525	0.993143		0.013667	
	AR1								
11	SDM-AR2	0.349	0.331	-1	-0.945	1.759495	-0.774	0.006748	
	AR2	0.319				1.7261	-0.742	0.000978	-0.0337
	SDM-AR1	0.252	0.241	1	1.864	0.981205		0.037237	
	AR1	0.234				0.98009		0.002288	-0.0470

Table C.2: Four models' estimation results based on $n_1 = 94$ data points, where C is the constant term and σ_ϵ^2 is the variance of the error term of the AR(1) and AR(2) processes.

C.1 Out-of-sample one-step ahead prediction error

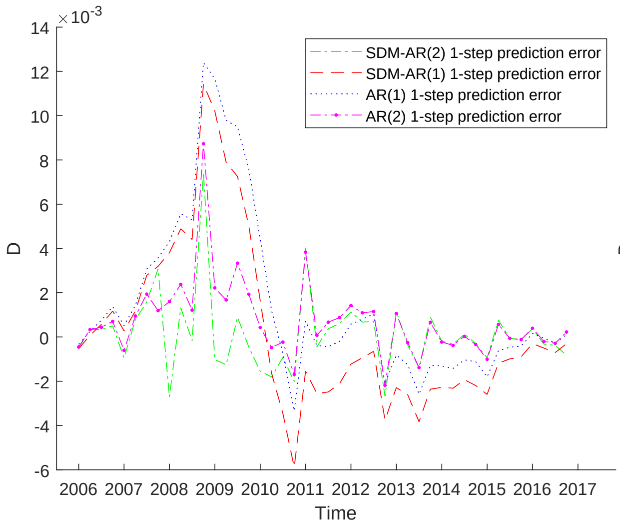
The following tables show the out-of-sample forecasting accuracy result for the SDM-AR and AR models.

Series	Model	MAE(%)	RMSE(%)	MAPE(%)
1	SDM-AR2	0.111	0.169	2.938
	AR2	0.115	0.184	2.905
	SDM-AR1	0.272	0.370	6.483
	AR1	0.245	0.398	5.761
2	SDM-AR2	0.153	0.267	9.548
	AR2	0.134	0.227	8.260
	SDM-AR1	0.204	0.301	11.38
	AR1	0.190	0.301	10.44
3	SDM-AR2	0.088	0.137	2.716
	AR2	0.088	0.141	2.653
	SDM-AR1	0.135	0.186	4.002
	AR1	0.130	0.186	3.820
4	SDM-AR2	0.183	0.244	5.131
	AR2	0.169	0.242	4.450
	SDM-AR1	0.229	0.309	6.001
	AR1	0.207	0.301	5.134
5	SDM-AR2	0.091	0.121	3.436
	AR2	0.148	0.176	6.295
	SDM-AR1	0.097	0.129	3.630
	AR1	0.164	0.131	3.685
6	SDM-AR2	0.411	0.572	15.02
	AR2	0.434	0.625	15.65
	SDM-AR1	0.481	0.673	16.61
	AR1	0.413	0.594	15.24
7	SDM-AR2	0.107	0.135	9.299
	AR2	0.106	0.134	9.089
	SDM-AR1	0.116	0.147	9.783
	AR1	0.113	0.145	9.495
8	SDM-AR2	0.216	0.305	3.686
	AR2	0.285	0.399	4.917
	SDM-AR1	0.420	0.552	7.560
	AR1	0.413	0.619	7.286
9	SDM-AR2	0.289	0.378	9.446
	AR2	0.293	0.398	9.195
	SDM-AR1	0.306	0.387	10.11
	AR1	0.294	0.399	9.227
10	SDM-AR2	0.461	0.671	6.199
	AR2	0.457	0.677	6.376
	SDM-AR1	0.456	0.672	6.601
	AR1	0.466	0.697	6.715
11	SDM-AR2	0.183	0.272	4.782
	AR2	0.268	0.398	7.023
	SDM-AR1	0.416	0.540	11.45
	AR1	0.391	0.578	10.12

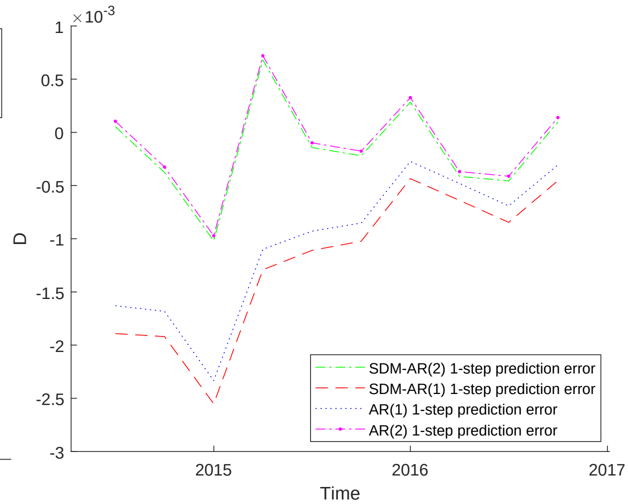
Table C.3: Out-of-sample Forecasting Error Comparison When $n_1 = 60$

Series	Model	MAE(%)	RMSE(%)	MAPE(%)
1	SDM-AR2	0.037	0.046	1.507
	AR2	0.036	0.045	1.464
	SDM-AR1	0.121	0.138	4.691
	AR1	0.102	0.120	3.942
2	SDM-AR2	0.235	0.413	17.383
	AR2	0.122	0.212	9.785
	SDM-AR1	0.100	0.180	9.245
	AR1	0.107	0.192	9.922
3	SDM-AR2	0.035	0.042	1.727
	AR2	0.041	0.051	2.015
	SDM-AR1	0.054	0.060	2.662
	AR1	0.054	0.059	2.641
4	SDM-AR2	0.031	0.037	1.429
	AR2	0.039	0.044	1.858
	SDM-AR1	0.045	0.051	2.078
	AR1	0.050	0.059	2.302
5	SDM-AR2	0.069	0.086	3.541
	AR2	0.070	0.088	3.564
	SDM-AR1	0.078	0.093	3.964
	AR1	0.077	0.092	3.916
6	SDM-AR2	0.257	0.302	18.56
	AR2	0.258	0.301	18.50
	SDM-AR1	0.271	0.309	19.03
	AR1	0.271	0.309	18.91
7	SDM-AR2	0.061	0.092	8.068
	AR2	0.061	0.086	7.945
	SDM-AR1	0.065	0.078	8.007
	AR1	0.063	0.076	7.730
8	SDM-AR2	0.061	0.078	1.412
	AR2	0.060	0.082	1.419
	SDM-AR1	0.277	0.296	6.490
	AR1	0.249	0.268	5.849
9	SDM-AR2	0.175	0.225	8.115
	AR2	0.178	0.226	8.122
	SDM-AR1	0.221	0.277	10.267
	AR1	0.190	0.234	8.4824
10	SDM-AR2	0.061	0.094	0.9134
	AR2	0.121	0.155	1.903
	SDM-AR1	0.333	0.351	5.381
	AR1	0.426	0.444	6.890
11	SDM-AR2	0.044	0.056	4.226
	AR2	0.039	0.050	3.733
	SDM-AR1	0.131	0.146	12.72
	AR1	0.091	0.104	8.803

Table C.4: Out-of-sample Forecasting Error Comparison When $n_1 = 94$

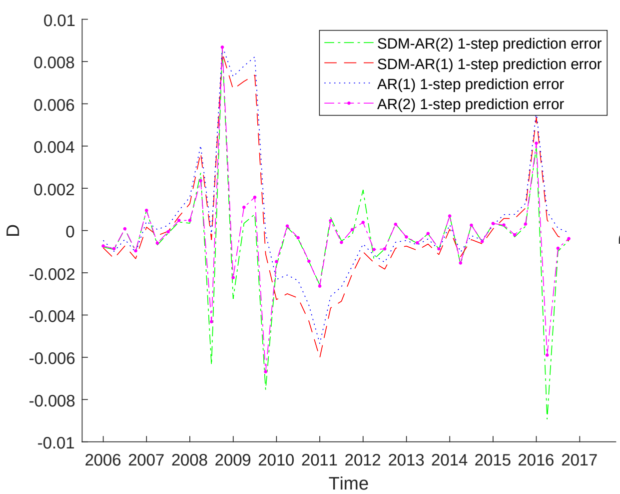


(a) One-step ahead prediction error with 60 points estimation

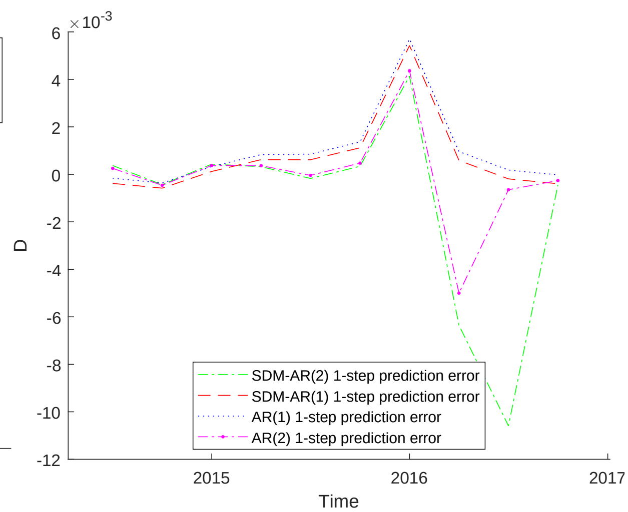


(b) One-step ahead prediction error with 94 points estimation

Figure C.1: Out-of-sample Forecasting Error Comparison for All series

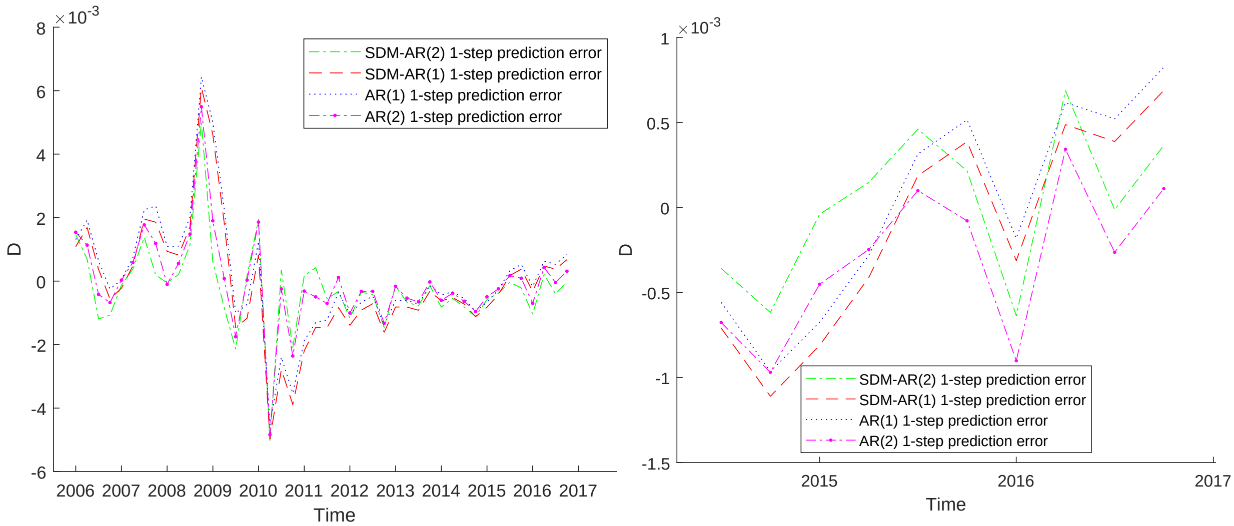


(a) One-step ahead prediction error with 60 points estimation



(b) One-step ahead prediction error with 94 points estimation

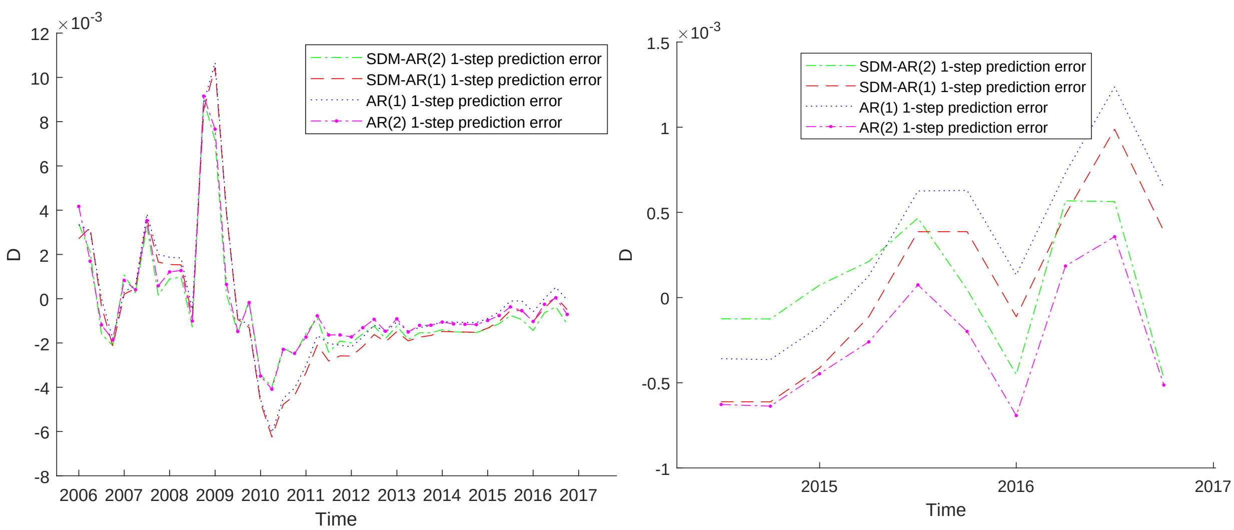
Figure C.2: Out-of-sample Forecasting Error Comparison for Business series



(a) One-step ahead prediction error with 60 points estimation

(b) One-step ahead prediction error with 94 points estimation

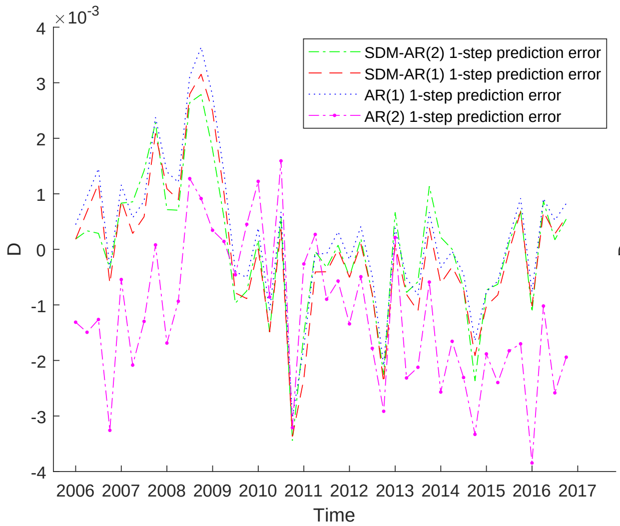
Figure C.3: Out-of-sample Forecasting Error Comparison for Consumer series



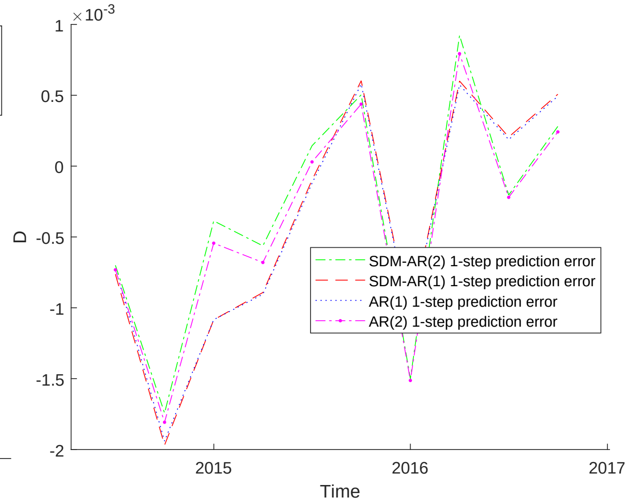
(a) One-step ahead prediction error with 60 points estimation

(b) One-step ahead prediction error with 94 points estimation

Figure C.4: Out-of-sample Forecasting Error Comparison for Credit Card series

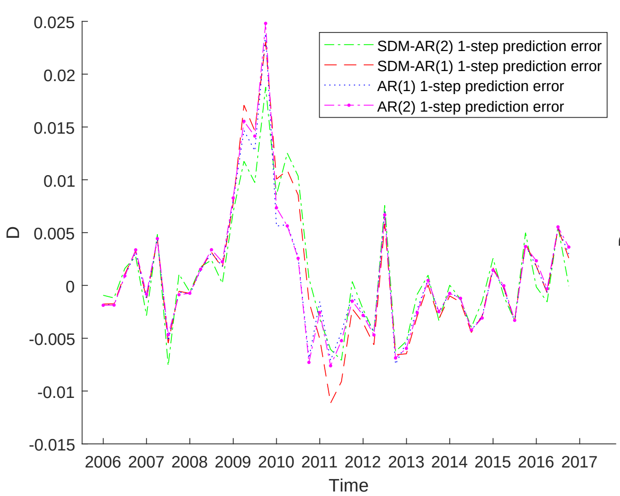


(a) One-step ahead prediction error with 60 points estimation

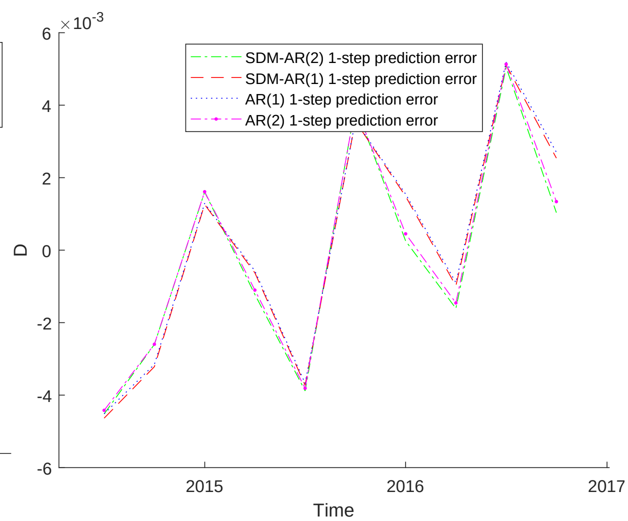


(b) One-step ahead prediction error with 94 points estimation

Figure C.5: Out-of-sample Forecasting Error Comparison for Other Consumer series

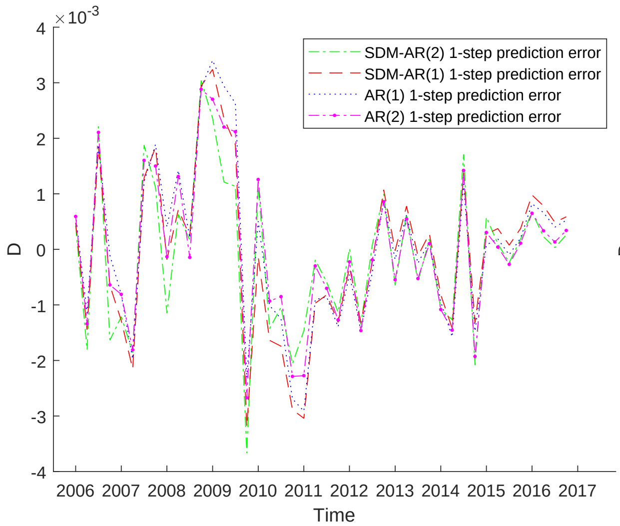


(a) One-step ahead prediction error with 60 points estimation

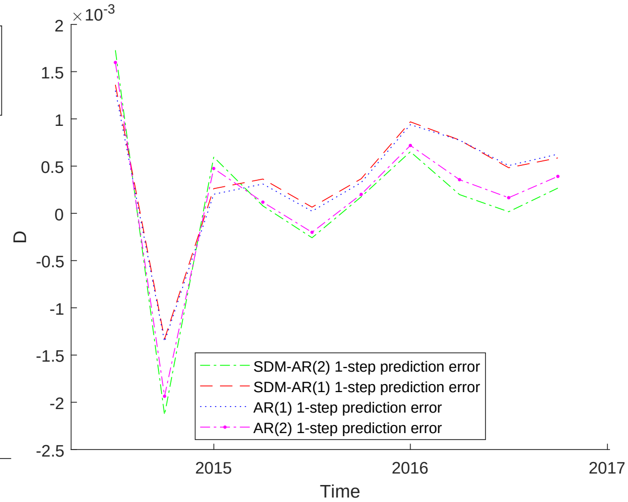


(b) One-step ahead prediction error with 94 points estimation

Figure C.6: Out-of-sample Forecasting Error Comparison for Agricultural series

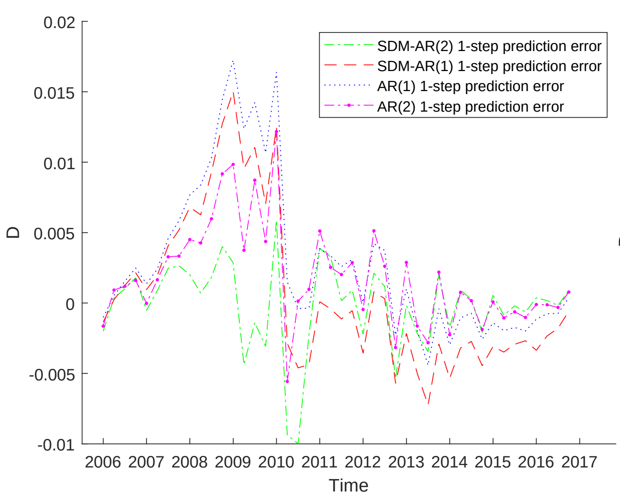


(a) One-step ahead prediction error with 60 points estimation

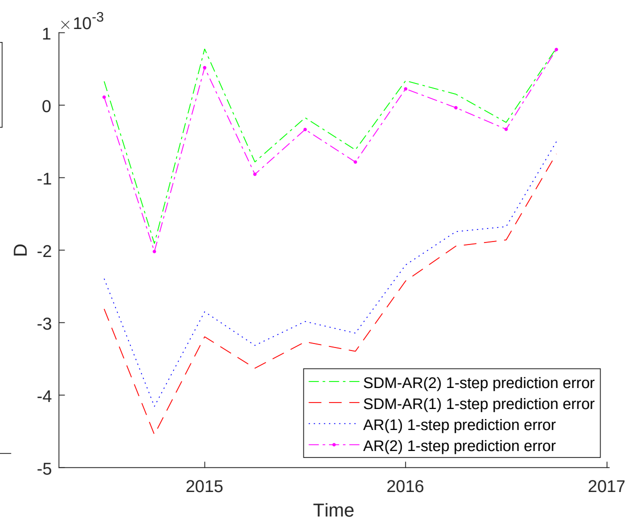


(b) One-step ahead prediction error with 94 points estimation

Figure C.7: Out-of-sample Forecasting Error Comparison for LFR series

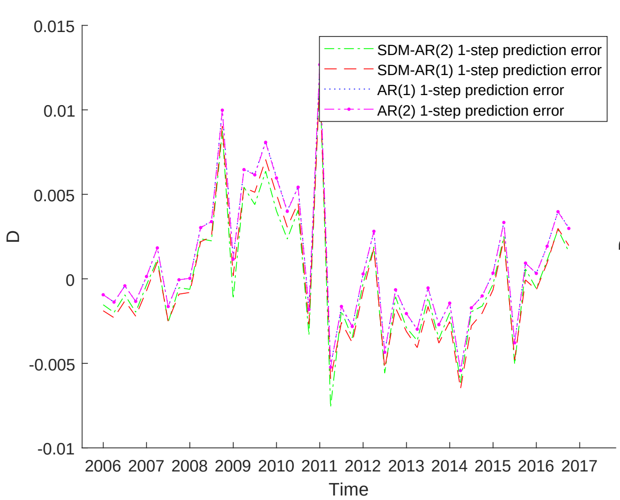


(a) One-step ahead prediction error with 60 points estimation

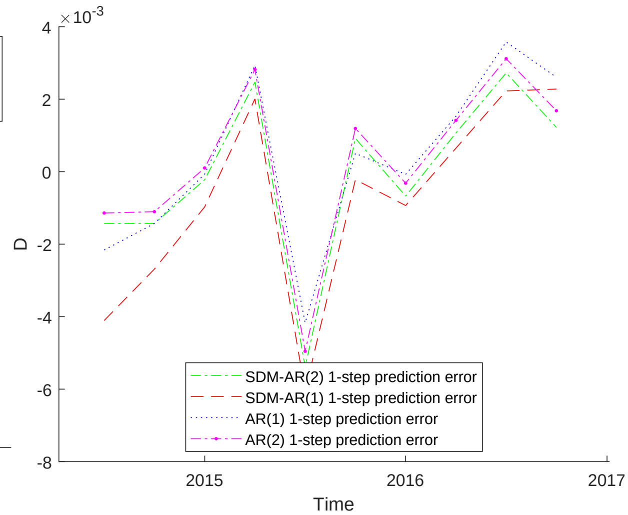


(b) One-step ahead prediction error with 94 points estimation

Figure C.8: Out-of-sample Forecasting Error Comparison for SRE series

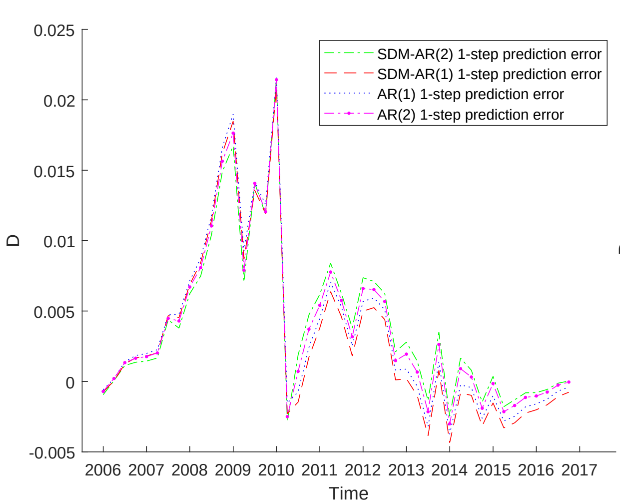


(a) One-step ahead prediction error with 60 points estimation

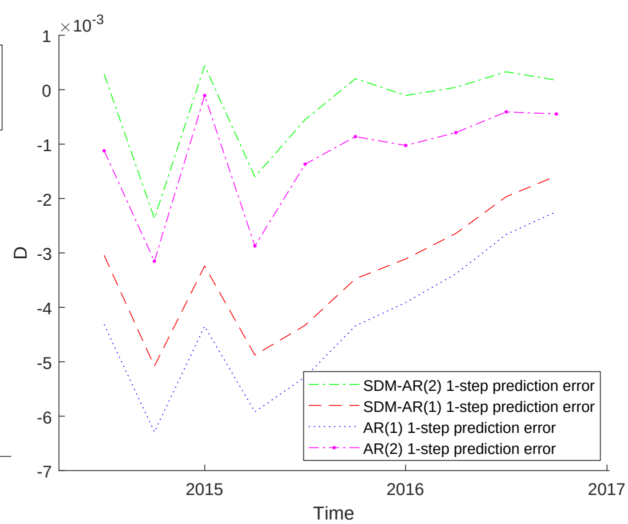


(b) One-step ahead prediction error with 94 points estimation

Figure C.9: Out-of-sample Forecasting Error Comparison for Farmland series

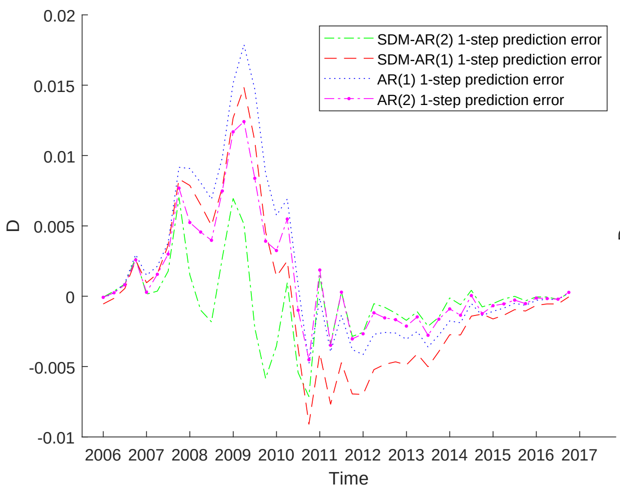


(a) One-step ahead prediction error with 60 points estimation

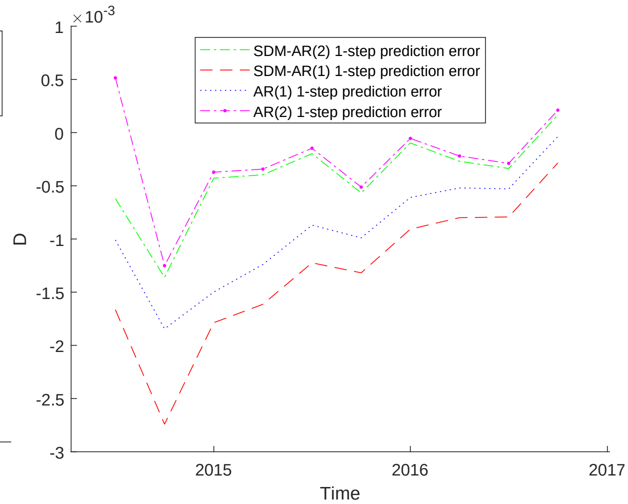


(b) One-step ahead prediction error with 94 points estimation

Figure C.10: Out-of-sample Forecasting Error Comparison for Mortgages series



(a) One-step ahead prediction error with 60 points estimation



(b) One-step ahead prediction error with 94 points estimation

Figure C.11: Out-of-sample Forecasting Error Comparison for CRE series

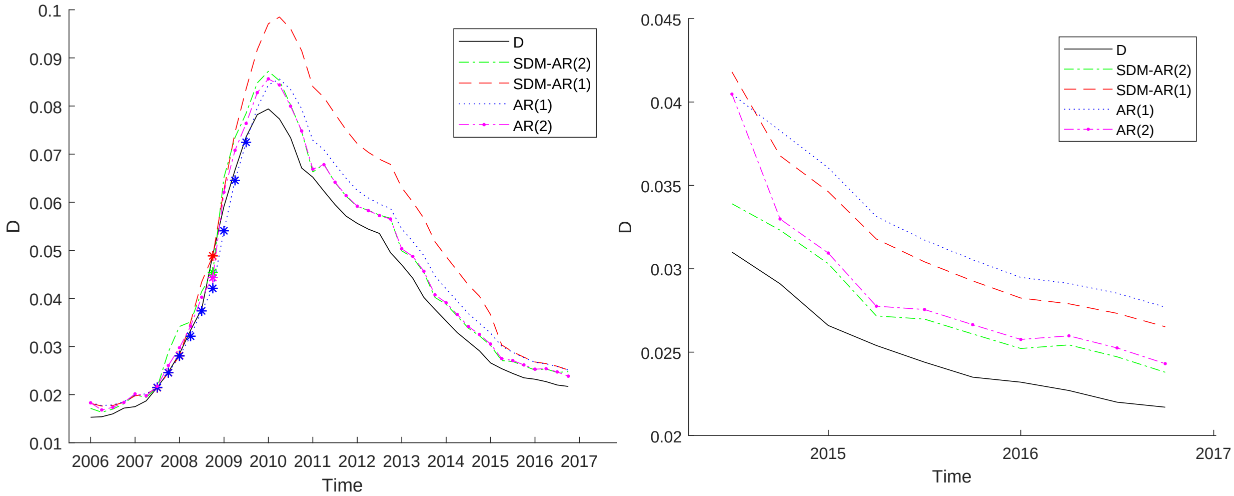
C.2 Out-of-sample one-step ahead prediction confidence interval

Series	Model	n*	Series	Model	n*
1	SDM-AR2	1	7	SDM-AR2	0
	AR2	1		AR2	0
	SDM-AR1	4		SDM-AR1	0
	AR1	9		AR1	0
2	SDM-AR2	2	8	SDM-AR2	0
	AR2	2		AR2	6
	SDM-AR1	2		SDM-AR1	3
	AR1	4		AR1	10
3	SDM-AR2	1	9	SDM-AR2	1
	AR2	1		AR2	1
	SDM-AR1	2		SDM-AR1	1
	AR1	2		AR1	1
4	SDM-AR2	2	10	SDM-AR2	
	AR2	2		AR2	
	SDM-AR1	2		SDM-AR1	
	AR1	2		AR1	
5	SDM-AR2	0	11	SDM-AR2	1
	AR2	0		AR2	2
	SDM-AR1	0		SDM-AR1	1
	AR1	2		AR1	7
6	SDM-AR2	1			
	AR2	2			
	SDM-AR1	2			
	AR1	2			

Table C.5: The number of points that lie out of the 99.9% one-sided confidence interval when $n_1 = 60$. The number of back-testing points is 44. The percentage column presents the percentage of the points out of total test points

Series	Model	n*	Series	Model	n*
1	SDM-AR2	0	7	SDM-AR2	0
	AR2	0		AR2	0
	SDM-AR1	0		SDM-AR1	0
	AR1	0		AR1	0
2	SDM-AR2	1	8	SDM-AR2	0
	AR2	1		AR2	0
	SDM-AR1	1		SDM-AR1	0
	AR1	1		AR1	0
3	SDM-AR2	0	9	SDM-AR2	0
	AR2	0		AR2	0
	SDM-AR1	0		SDM-AR1	0
	AR1	0		AR1	0
4	SDM-AR2	0	10	SDM-AR2	
	AR2	0		AR2	
	SDM-AR1	0		SDM-AR1	
	AR1	0		AR1	
5	SDM-AR2	0	11	SDM-AR2	0
	AR2	0		AR2	0
	SDM-AR1	0		SDM-AR1	0
	AR1	0		AR1	0
6	SDM-AR2	0			
	AR2	0			
	SDM-AR1	0			
	AR1	0			

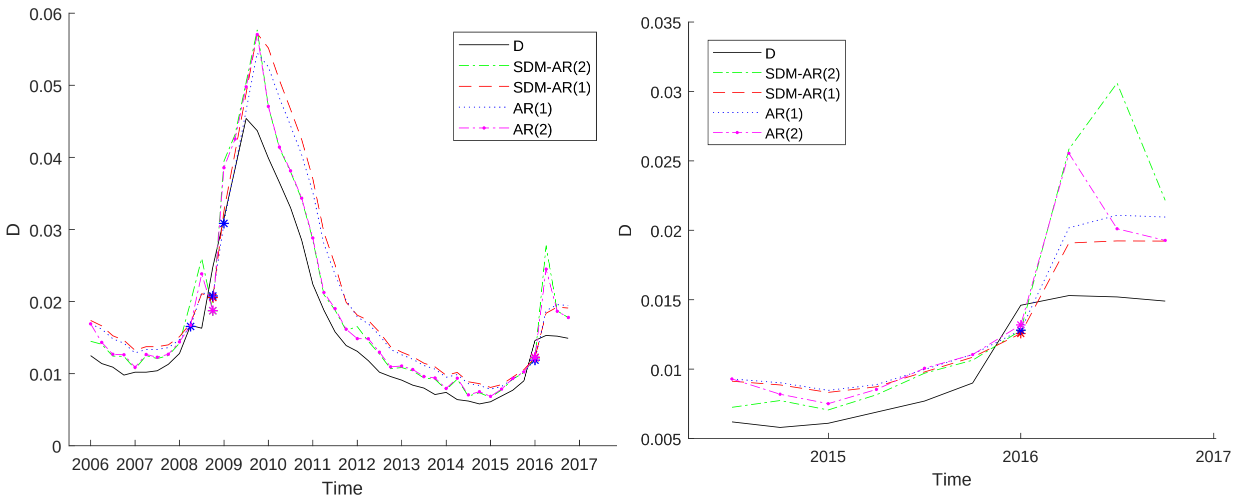
Table C.6: The number of points that lie out of the 99.9% one-sided confidence interval when $n_1 = 94$. The number of back-testing points is 10. The percentage column calculate the presents of the points out of total test points



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

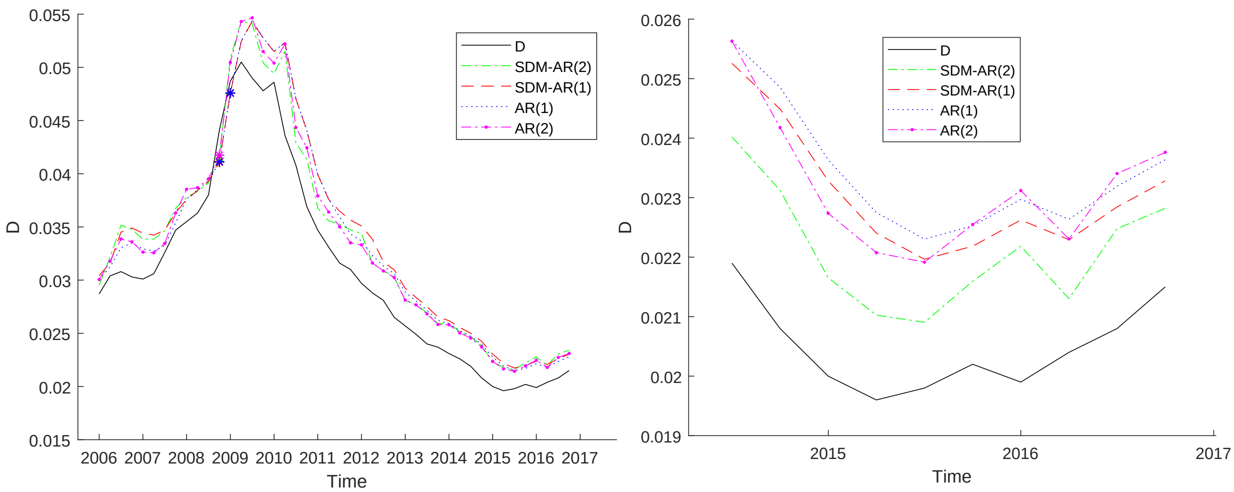
Figure C.12: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for ALL series



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

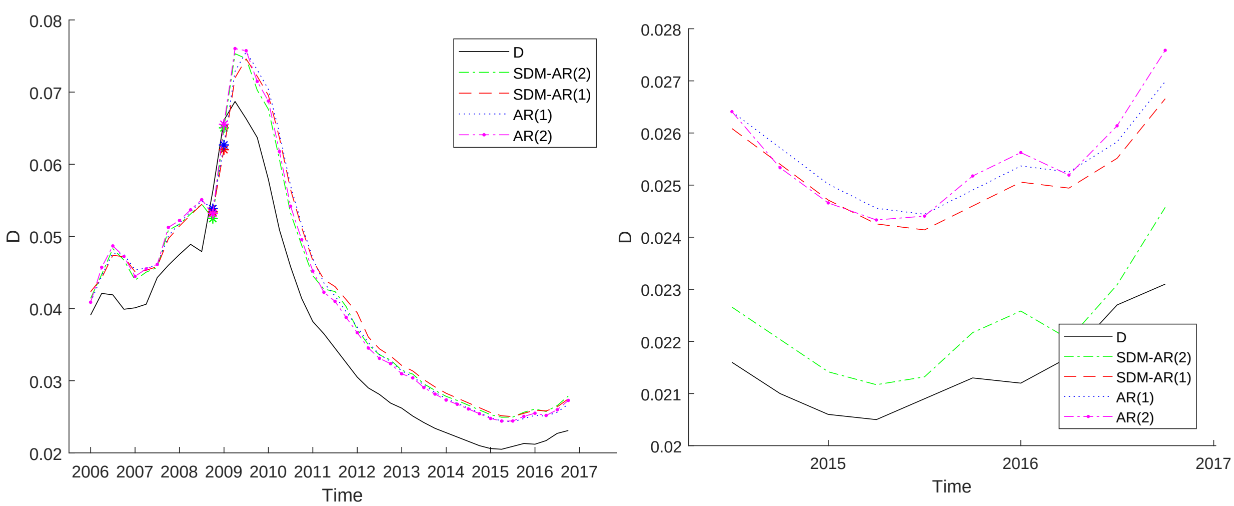
Figure C.13: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for Business series



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

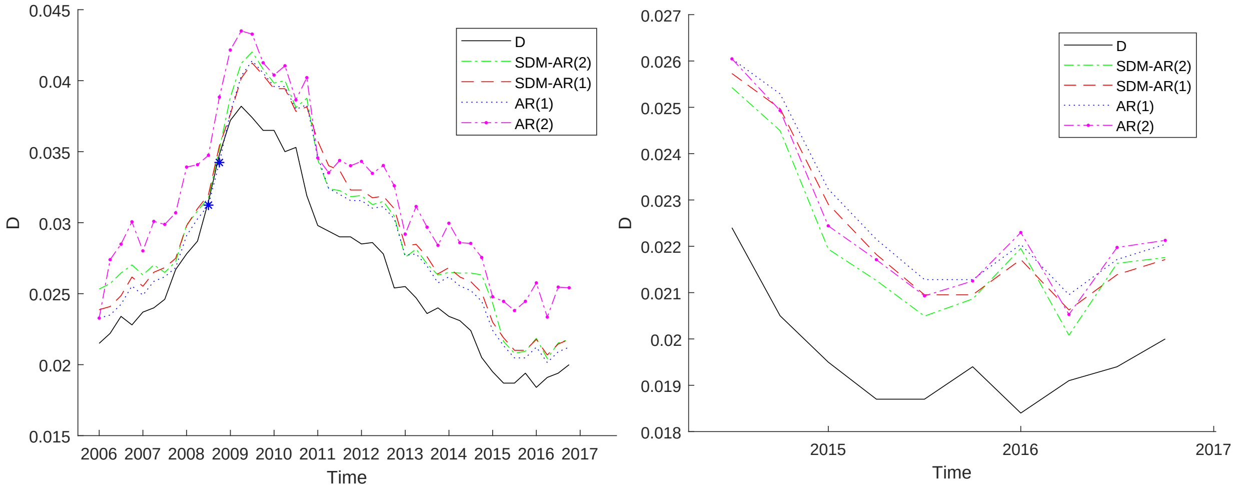
Figure C.14: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for Consumer series



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

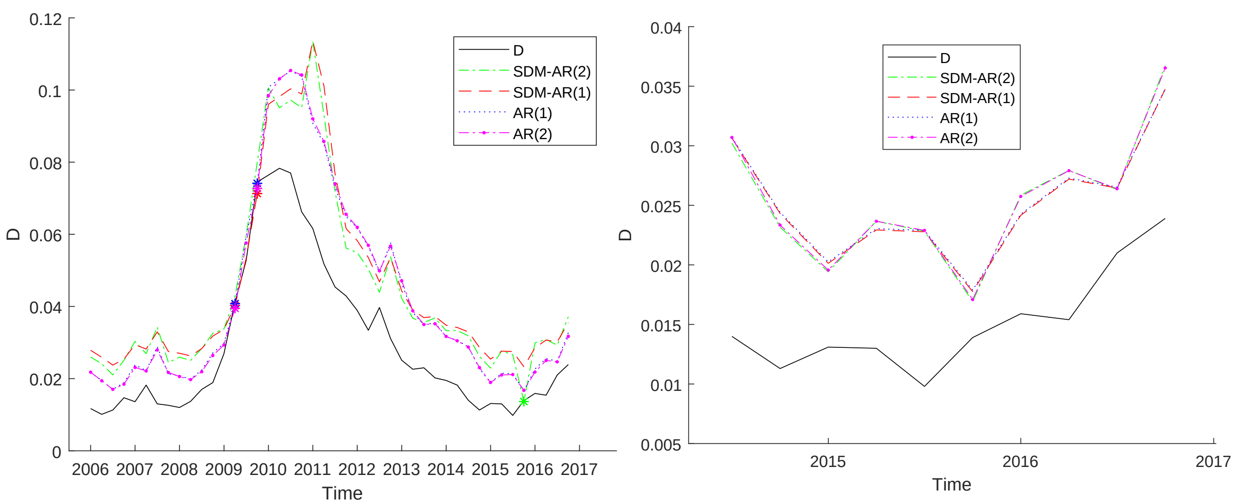
Figure C.15: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for Credit Card series



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

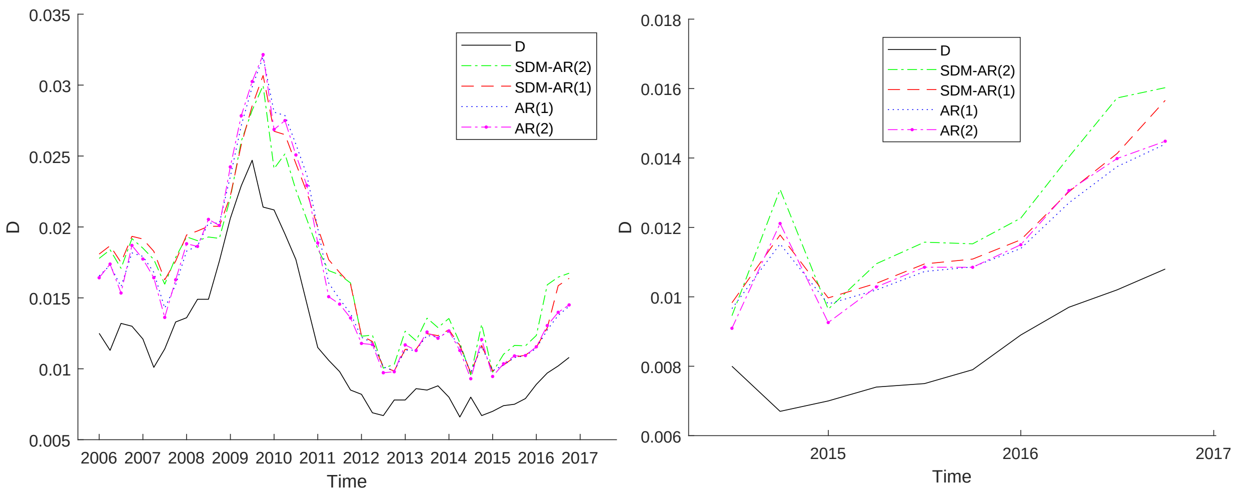
Figure C.16: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for Other Consumer series



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

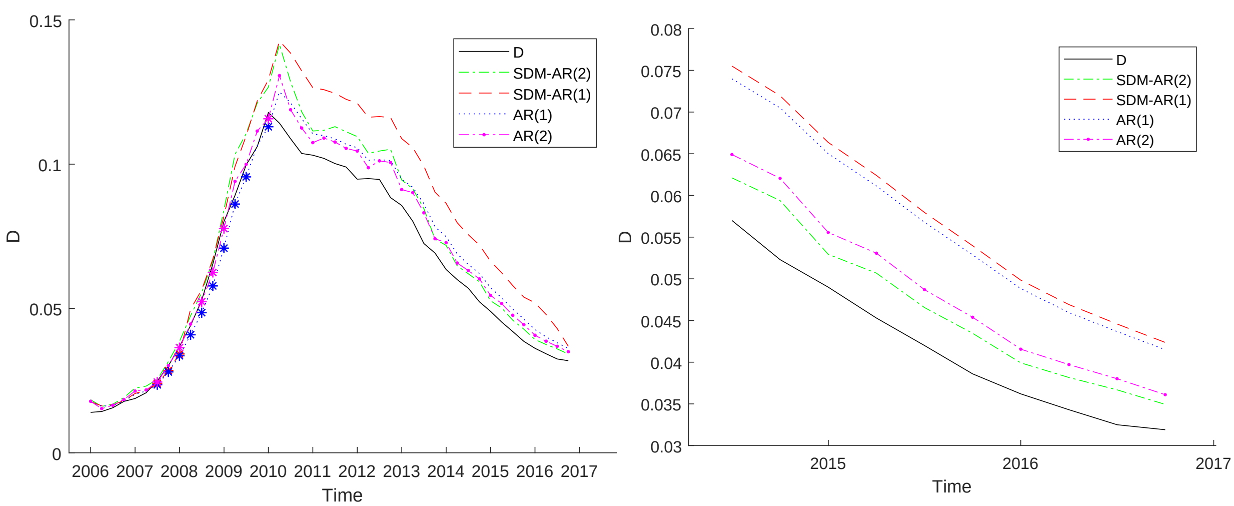
Figure C.17: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for Agriculture series



(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

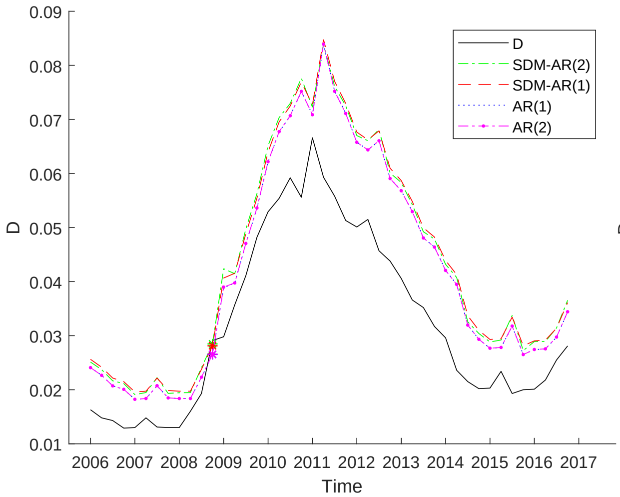
Figure C.18: Out-of-sample 99.9% upper-side forecasting confidence intervalcomparison for LFR series



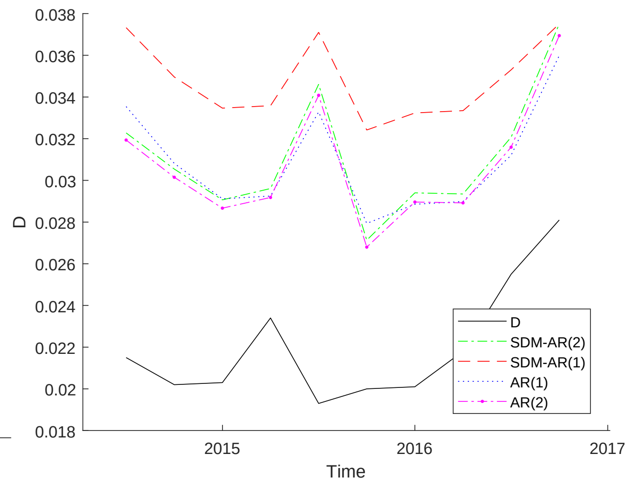
(a) One-step ahead prediction CI with 60 points estimation

(b) One-step ahead prediction CI with 94 points estimation

Figure C.19: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for SRE series

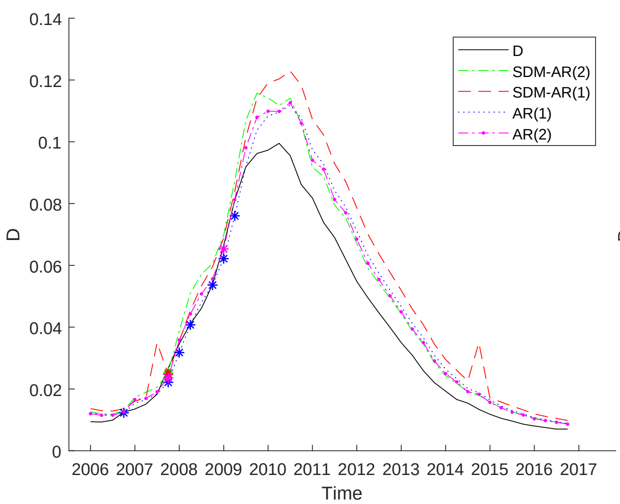


(a) One-step ahead prediction CI with 60 points estimation

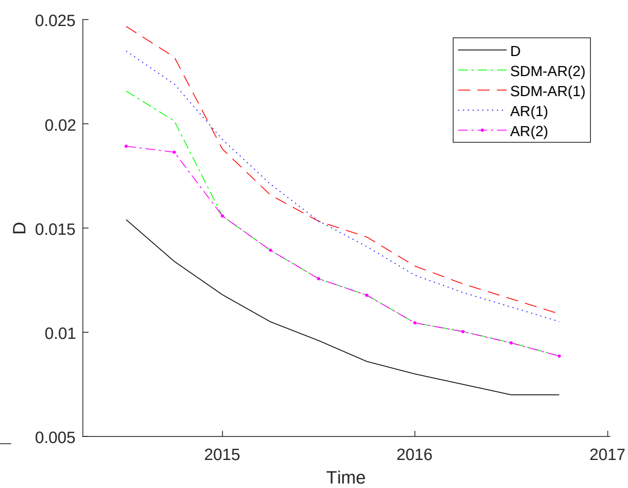


(b) One-step ahead prediction CI with 94 points estimation

Figure C.20: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for Farmland series



(a) One-step ahead prediction CI with 60 points estimation



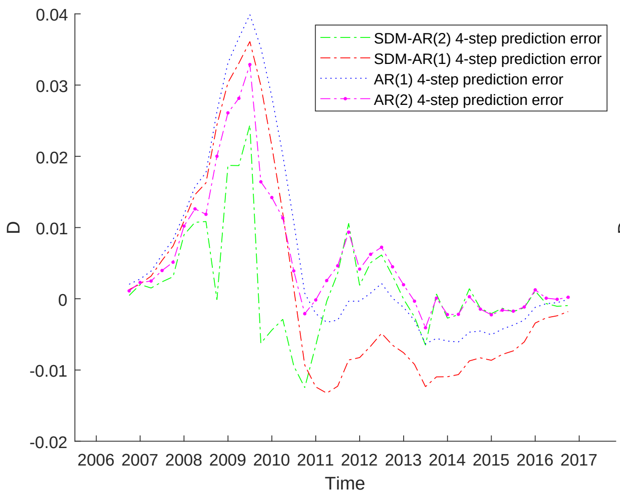
(b) One-step ahead prediction CI with 94 points estimation

Figure C.21: Out-of-sample 99.9% upper-side forecasting confidence interval comparison for CRE series

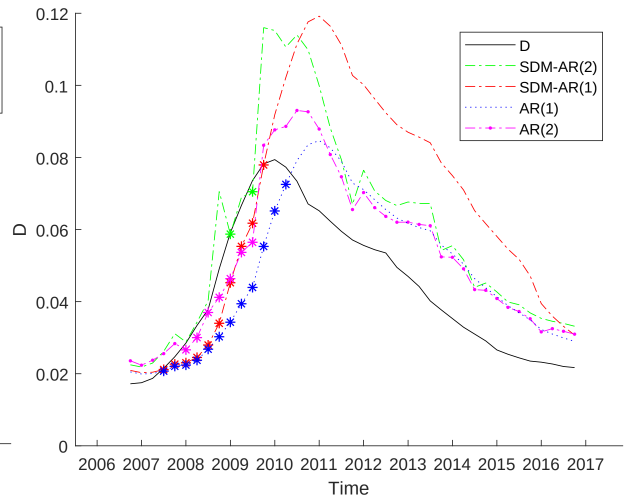
C.3 Out-of-sample 4-step ahead prediction

Series	Model	MAE(%)	RMSE(%)	MAPE(%)	n*
1	SDM-AR2	0.493	0.743	10.632	2
	AR2	0.645	1.027	13.390	7
	SDM-AR1	1.099	1.394	24.863	10
	AR1	0.896	1.433	19.233	12
2	SDM-AR2	0.550	0.847	30.536	1
	AR2	0.492	0.737	27.935	1
	SDM-AR1	0.746	0.995	43.616	8
	AR1	0.701	0.973	38.053	8
3	SDM-AR2	0.440	0.520	14.504	3
	AR2	0.396	0.522	12.018	3
	SDM-AR1	0.482	0.617	14.535	4
	AR1	0.468	0.603	13.679	7
4	SDM-AR2	0.812	0.917	24.547	3
	AR2	0.736	0.874	20.591	3
	SDM-AR1	0.788	0.948	22.176	3
	AR1	0.723	0.920	18.904	3
5	SDM-AR2	0.314	0.400	11.256	4
	AR2	0.738	0.875	32.730	0
	SDM-AR1	0.332	0.407	12.002	4
	AR1	0.306	0.413	10.804	8
6	SDM-AR2	1.135	1.641	36.050	4
	AR2	1.181	1.789	34.982	5
	SDM-AR1	1.286	1.839	39.459	5
	AR1	1.142	1.748	34.153	5
7	SDM-AR2	0.290	0.382	24.877	0
	AR2	0.275	0.359	23.713	0
	SDM-AR1	0.317	0.418	26.639	0
	AR1	0.288	0.378	24.356	0
8	SDM-AR2	0.873	1.149	13.743	1
	AR2	1.350	1.905	20.058	12
	SDM-AR1	1.611	2.039	26.651	11
	AR1	1.560	2.213	23.903	15
9	SDM-AR2	0.862	1.031	28.101	0
	AR2	0.837	1.134	23.865	6
	SDM-AR1	0.889	1.055	29.305	2
	AR1	0.817	1.108	23.357	5
10	SDM-AR2	1.861	2.528	22.180	21
	AR2	1.792	2.467	22.174	14
	SDM-AR1	1.712	2.304	22.147	13
	AR1	1.754	2.393	22.439	14
11	SDM-AR2	1.030	1.372	24.729	1
	AR2	1.325	1.88	32.734	11
	SDM-AR1	1.714	2.061	48.035	10
	AR1	1.501	2.084	36.900	14

Table C.7: Four-step ahead out-of-sample prediction results with 60 points training set

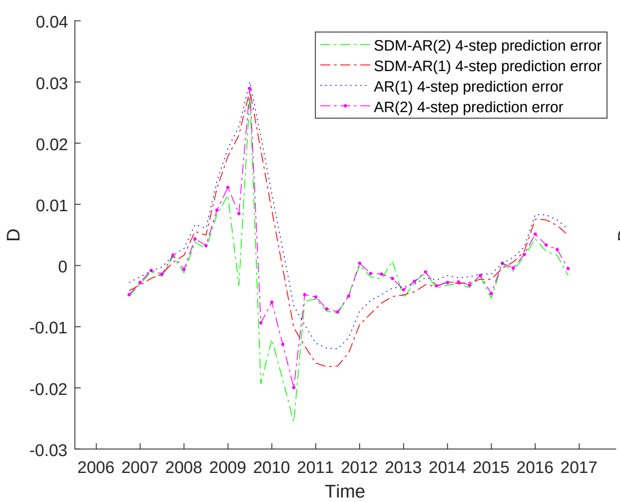


(a) Four-step ahead prediction error with 60 points training set

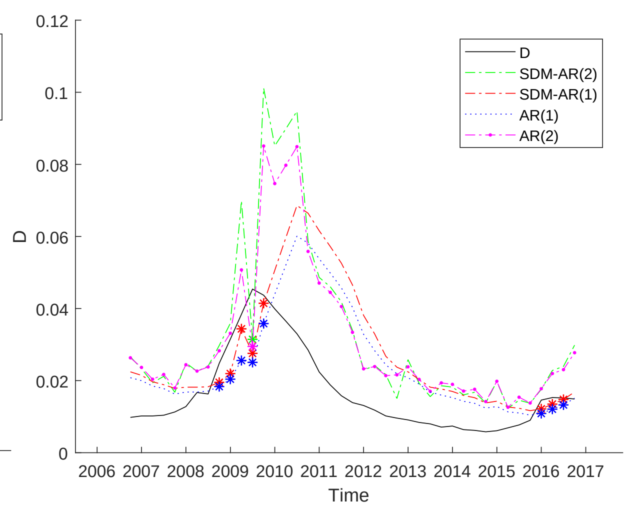


(b) Four-step ahead prediction CI with 60 points training set

Figure C.22: Out-of-sample four-step ahead prediction for ALL series

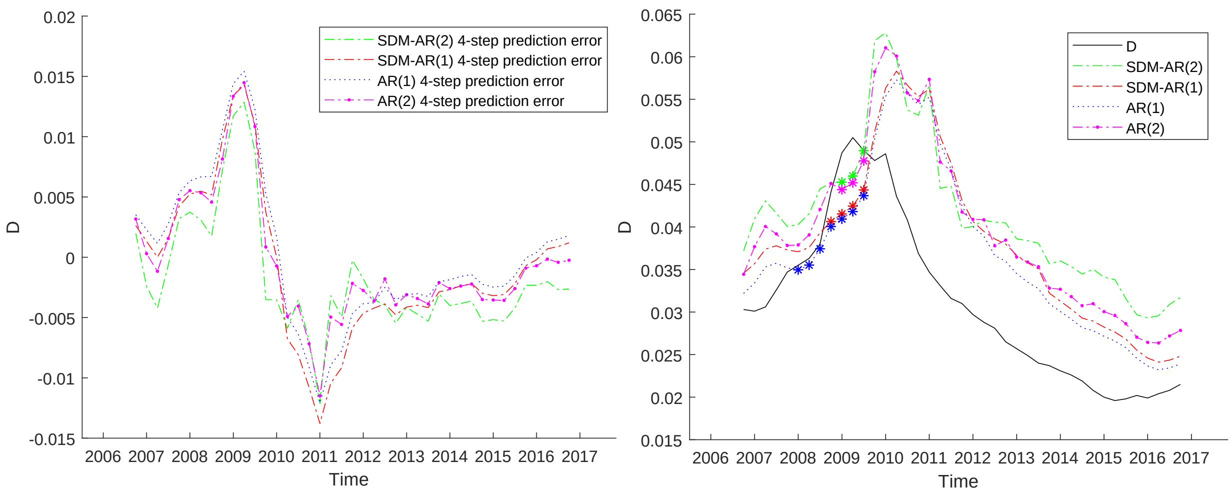


(a) Four-step ahead prediction error with 60 points training set



(b) Four-step ahead prediction CI with 60 points training set

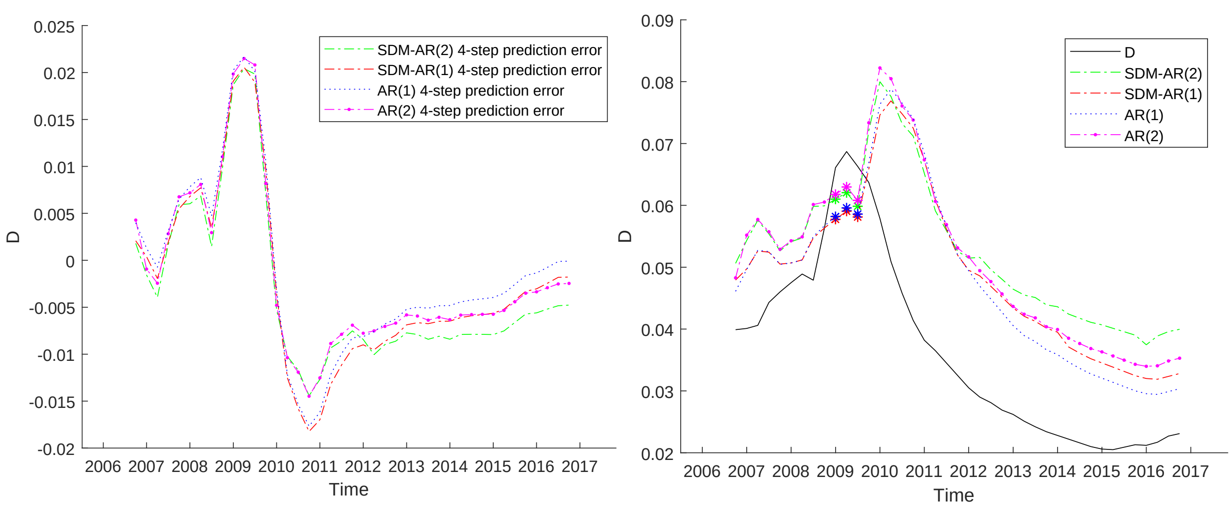
Figure C.23: Out-of-sample four-step ahead prediction for Business series



(a) Four-step ahead prediction error with 60 points training set

(b) Four-step ahead prediction CI with 60 points training set

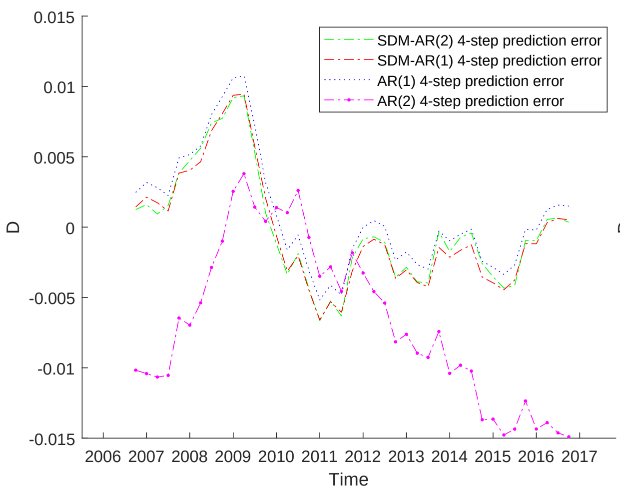
Figure C.24: Out-of-sample four-step ahead prediction for Consumer series



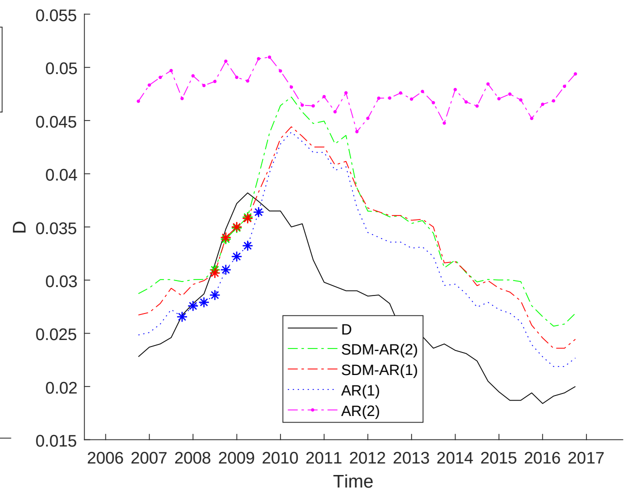
(a) Four-step ahead prediction error with 60 points training set

(b) Four-step ahead prediction CI with 60 points training set

Figure C.25: Out-of-sample four-step ahead prediction for Credit Card series

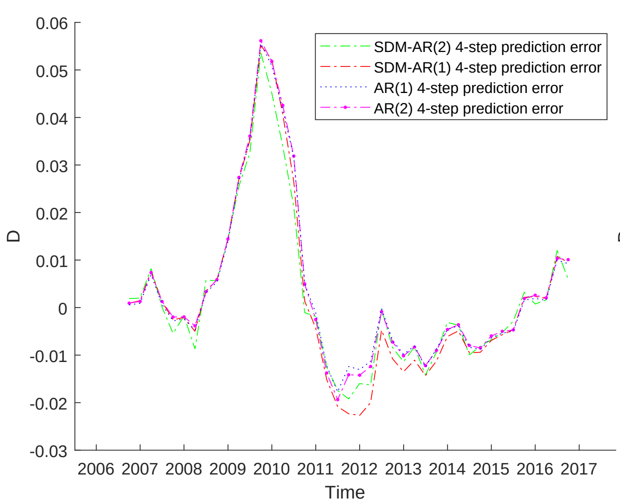


(a) Four-step ahead prediction error with 60 points training set

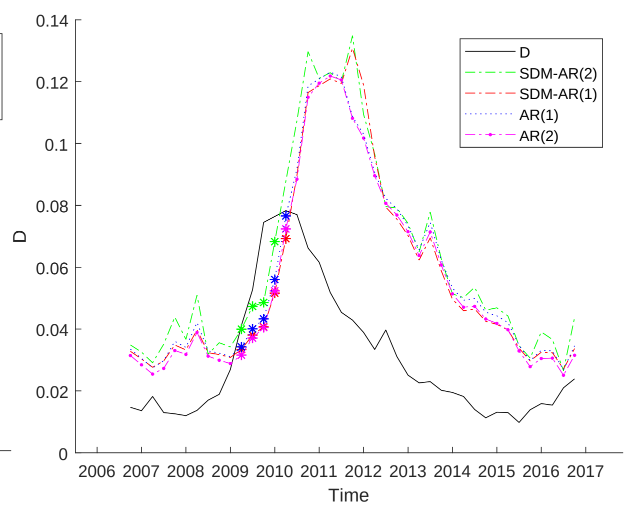


(b) Four-step ahead prediction CI with 60 points training set

Figure C.26: Out-of-sample four-step ahead prediction for Other Consumer series

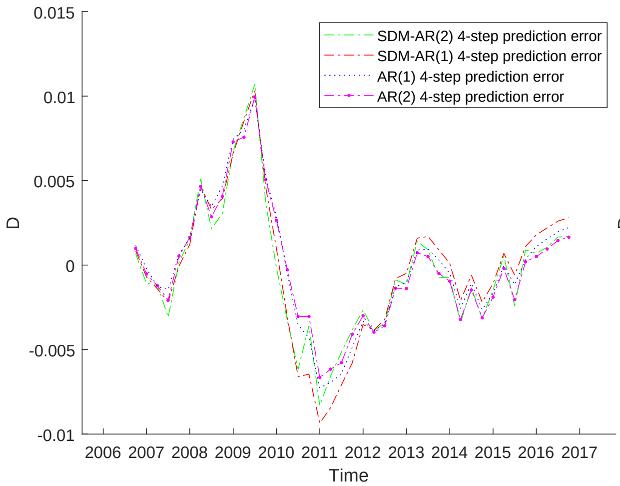


(a) Four-step ahead prediction error with 60 points training set

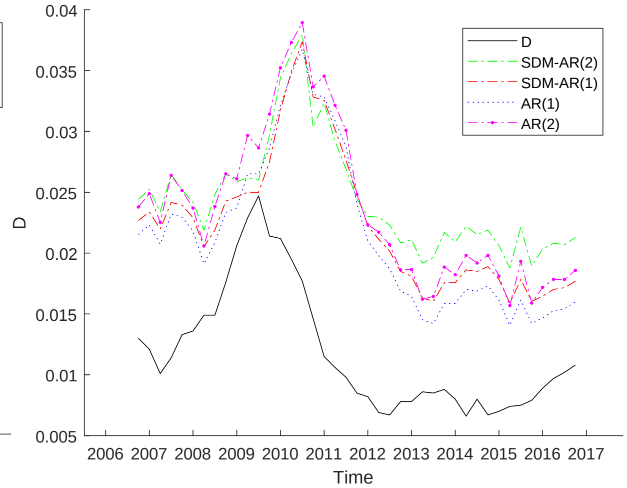


(b) Four-step ahead prediction CI with 60 points training set

Figure C.27: Out-of-sample four-step ahead prediction for Agriculture series

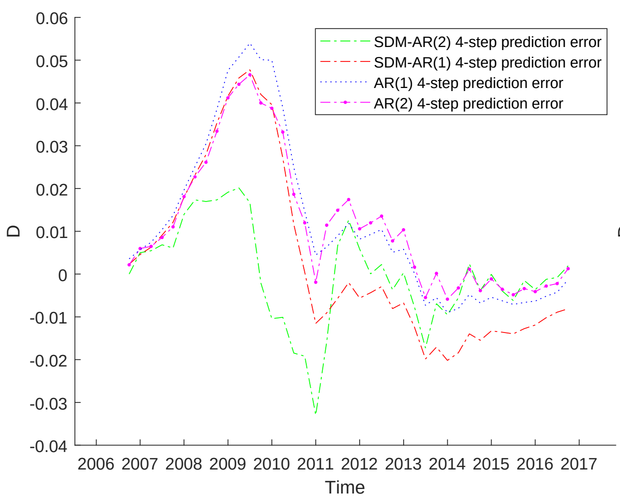


(a) Four-step ahead prediction error with 60 points training set

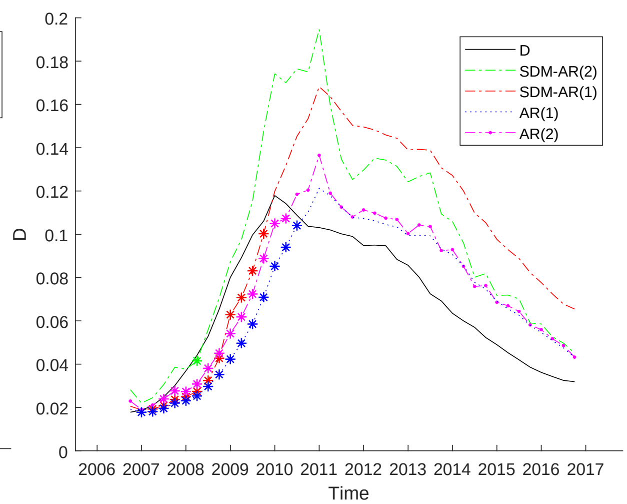


(b) Four-step ahead prediction CI with 60 points training set

Figure C.28: Out-of-sample four-step ahead prediction for LFR series

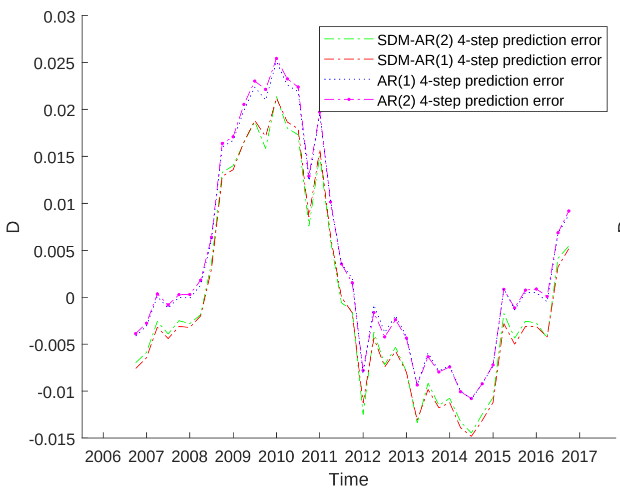


(a) Four-step ahead prediction error with 60 points training set

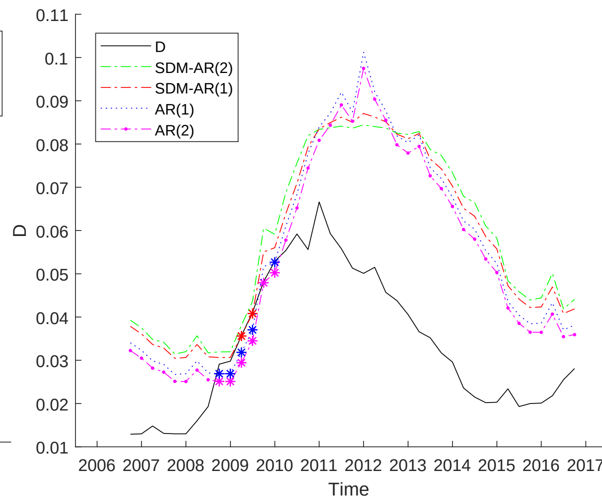


(b) Four-step ahead prediction CI with 60 points training set

Figure C.29: Out-of-sample four-step ahead prediction for SRE series

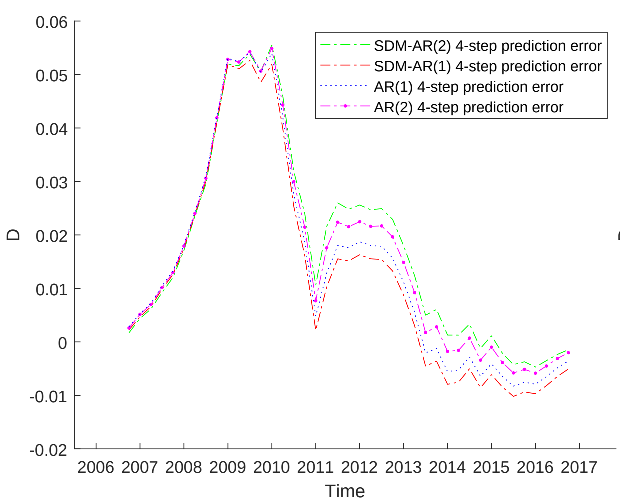


(a) Four-step ahead prediction error with 60 points training set

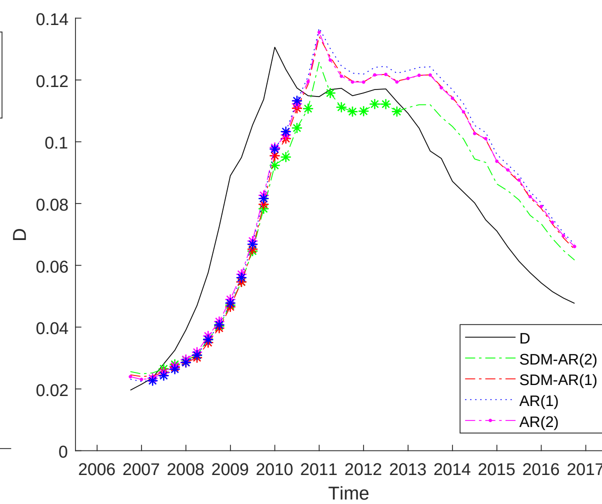


(b) Four-step ahead prediction CI with 60 points training set

Figure C.30: Out-of-sample four-step ahead prediction for Farmland series

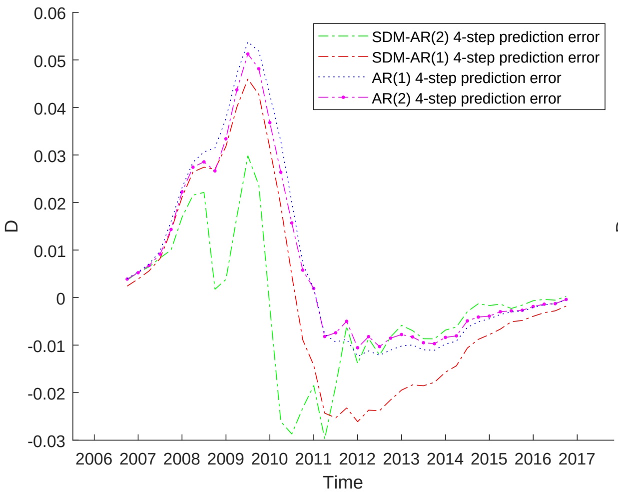


(a) Four-step ahead prediction error with 60 points training set

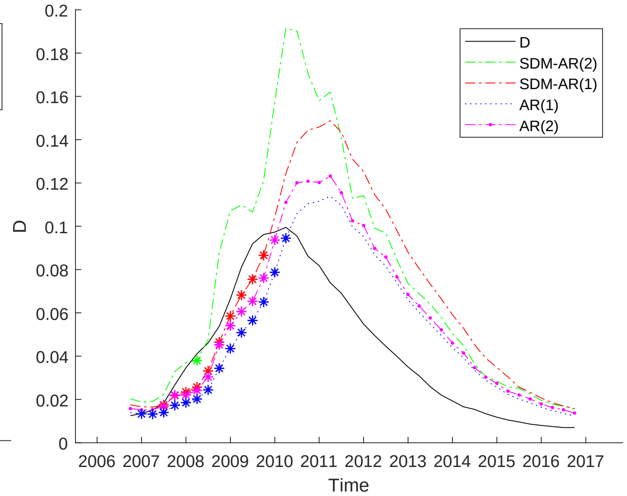


(b) Four-step ahead prediction CI with 60 points training set

Figure C.31: Out-of-sample four-step ahead prediction for Mortgages series



(a) Four-step ahead prediction error with 60 points training set



(b) Four-step ahead prediction CI with 60 points training set

Figure C.32: Out-of-sample four-step ahead prediction for CRE series

Appendix D

The conditional density function of Y_t given M_t.

In this section, we derive the conditional density function of Y_t given M_t, $f_{Y_t|M_t}(y_t|m_t)$, for the three t-SDM-AR(2) models defined in Section 5.2. The conditional density function $f_{Y_t|M_t}(y_t|m_t)$ is an essential part for calculating the one-step-ahead predictive density function for all three t-SDM-AR(2) models.

D.1 $f_{Y_t|M_t}(y_t|m_t)$ of t-SDM-AR(2)-S

To derive the conditional density function of Y_t given M_t, we start by looking at the conditional survival function first. We have:

$$
\begin{aligned}
P(Y_t \geqslant y_t|M_t = m_t) =& P(\frac{x_{PD} - a(T_t)\sqrt{\frac{\nu-2}{\nu}}S_{t,\nu}m_t}{\sqrt{1-a(T_t)^2}} \geqslant y_t|M_t = m_t) \\
=& P(S_{t,\nu} \leqslant \frac{x_{PD} - \sqrt{1-a(T_t)^2}y_t}{a(T_t)\sqrt{\frac{\nu-2}{\nu}}m_t}|M_t = m_t)\mathbb{1}_{\{m_t>0\}} \\
&+ P(S_{t,\nu} \geqslant \frac{x_{PD} - \sqrt{1-a(T_t)^2}y_t}{a(T_t)\sqrt{\frac{\nu-2}{\nu}}m_t}|M_t = m_t)\mathbb{1}_{\{m_t<0\}} \\
=& \sum_{i=1}^{2}\Bigg[P(S_{t,\nu} \leqslant \frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t}, T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{m_t>0\}} \\
&+ P(S_{t,\nu} \geqslant \frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t}, T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{m_t<0\}}\Bigg] \\
=& \sum_{i=1}^{2}\Bigg[P(S_{t,\nu} \leqslant \frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t})P(T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{m_t>0\}} \\
&+ P(S_{t,\nu} \geqslant \frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t})P(T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{m_t<0\}}\Bigg].
\end{aligned}
$$

Then, by taking the first derivative with respect to y_t, we find the conditional density function of Y_t given M_t,

$$
\begin{aligned}
f_{Y_t|M_t}(y_t|m_t) =& \frac{d}{dy_t}(1 - P(Y_t \geqslant y_t|M_t = m_t)) \\
=& \sum_{i=1}^{2} \left[\frac{\sqrt{1-a_i^2}}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t} f_{S_\nu}(\frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t}) P(T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{m_t>0\}} \right. \\
& \left. - \frac{\sqrt{1-a_i^2}}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t} f_{S_\nu}(\frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t}) P(T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{m_t<0\}} \right] \\
=& \sum_{i=1}^{2} \left[\frac{\sqrt{1-a_i^2}}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t} f_{S_\nu}(\frac{x_{PD} - \sqrt{1-a_i^2}y_t}{a_i\sqrt{\frac{\nu-2}{\nu}}m_t}) P(T_t \in [t_i, t_{i+1}]|M_t = m_t)(\mathbb{1}_{\{m_t>0\}} - \mathbb{1}_{\{m_t<0\}}) \right],
\end{aligned}
\tag{D.1}
$$

where f_{S_ν} is the density function of S_ν defined in Equation 5.6. Although, Function D.1 is undefined at the point $m_t = 0$, this is acceptable since M_t follows a continuous distribution.

D.2 $f_{Y_t|M_t}(y_t|m_t)$ of t-SDM-AR(2)-L

We use the same method as the one in Section D.1. The conditional survival function is of the form:

$$\begin{aligned}
P(Y_t \geqslant y_t|M_t = m_t) =& P(\frac{\sqrt{\nu}}{\sqrt{\nu-2}S_{t,\nu}}\frac{x_{PD} - a(T_t)m_t}{\sqrt{1-a(T_t)^2}} \geqslant y_t|M_t = m_t) \\
=& P(\frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a(T_t)m_t}{\sqrt{1-a(T_t)^2}} \geqslant S_{t,\nu}y_t|M_t = m_t) \\
=& P(S_{t,\nu} \leqslant \frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a(T_t)m_t}{\sqrt{1-a(T_t)^2}}\frac{1}{y_t}|M_t = m_t)\mathbb{1}_{\{y_t>0\}} \\
&+ P(S_{t,\nu} \geqslant \frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a(T_t)m_t}{\sqrt{1-a(T_t)^2}}\frac{1}{y_t}|M_t = m_t)\mathbb{1}_{\{y_t<0\}} \\
=& \sum_{i=1}^{K}\Bigg[P(S_{t,\nu} \leqslant \frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a_i m_t}{\sqrt{1-a_i^2}}\frac{1}{y_t})P(T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{y_t>0\}} \\
&+ P(S_{t,\nu} \geqslant \frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a_i m_t}{\sqrt{1-a_i^2}}\frac{1}{y_t})P(T_t \in [t_i, t_{i+1}]|M_t = m_t)\mathbb{1}_{\{y_t<0\}}\Bigg]
\end{aligned}$$

The conditional density function can be derived in the standard way:

$$\begin{aligned}
f_{Y_t|M_t}(y_t|m_t) =& \frac{d}{dy_t}(1 - P(Y_t \geqslant y_t|M_t = m_t)) \\
=& \sum_{i=1}^{K}\Bigg[\frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a_i m_t}{\sqrt{1-a_i^2}}\frac{1}{y_t^2}f_{S_\nu}(\frac{\sqrt{\nu}}{\sqrt{\nu-2}}\frac{x_{PD} - a_i m_t}{\sqrt{1-a_i^2}}\frac{1}{y_t})(\mathbb{1}_{\{y_t>0\}} - \mathbb{1}_{\{y_t<0\}})\Bigg]
\end{aligned}$$

D.3 $f_{Y_t|M_t}(y_t|m_t)$ of t-SDM-AR(2)-LS

We use the same method as the one in Section D.1. The conditional survival function is of the form:

$$\begin{aligned}
P(Y_t \geqslant y_t|M_t = m_t) &= P\left(\frac{x_{PD}\sqrt{\frac{\nu}{\nu-2}}/S_{t,\nu} - a(T_t)M_t}{\sqrt{1-a(T_t)^2}} \geqslant y_t|M_t = m_t\right) \\
&= P\left(x_{PD}\sqrt{\frac{\nu}{\nu-2}}/S_{t,\nu} \geqslant a(T_t)M_t + \sqrt{1-a(T_t)^2}y_t|M_t = m_t\right).
\end{aligned}$$

For $x_{PD} < 0$, we have

$$\begin{aligned}
P(Y_t \geqslant y_t | M_t = m_t) &= P\left(\frac{1}{S_{t,\nu}} \leqslant \frac{a(T_t)M_t + \sqrt{1 - a(T_t)^2} y_t}{x_{PD}\sqrt{\frac{\nu}{\nu-2}}} | M_t = m_t \right) \\
&= P\left(S_{t,\nu} \geqslant \frac{x_{PD}\sqrt{\frac{\nu}{\nu-2}}}{a(T_t)M_t + \sqrt{1 - a(T_t)^2} y_t} | M_t = m_t \right) \\
&= \sum_{i=1}^{K} \left[P\left(S_{t,\nu} \geqslant \frac{x_{PD}\sqrt{\frac{\nu}{\nu-2}}}{a_i M_t + \sqrt{1 - a_i^2} y_t} \right) P(T_t \in [t_k, t_{k+1}] | M_t = m_t) \right].
\end{aligned}$$

Then, by taking the first derivative, we find the conditional density function of Y_t given M_t,

$$\begin{aligned}
f_{Y_t|M_t}(y_t|m_t) =& \frac{d}{dy_t}(1 - P(Y_t \geqslant y_t | M_t = m_t)) \\
=& \sum_{i=1}^{K} \left[\frac{-x_{PD}\sqrt{\frac{\nu(1-a_i^2)}{\nu-2}}}{(a_i M_t + \sqrt{1 - a_i^2} y_t)^2} f_{S_\nu} \left(\frac{x_{PD}\sqrt{\frac{\nu}{\nu-2}}}{a_i M_t + \sqrt{1 - a_i^2} y_t} \right) P(T_t \in [t_k, t_{k+1}] | M_t = m_t) \right].
\end{aligned}$$

Appendix E

Likelihood approximation

As we can see from Equation 5.17, $f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t})$ is a three-dimensional integral over M_{t+1}, M_t, M_{t-1}. But the built-in function of Matlab is inefficient and unreliable when calculating such integral. In order to improve the computational efficiency, we adopt the following methodology to approximate the value of $f_{Y_{t+1}|Y_{1:t1}}(y_{t+1}|y_{1:t})$.

1. First, we create a three-dimensional mesh grid over $[-5\ 5]$ of the space of M_{t+1}, M_t, M_{t-1}. We use d to denote the grid size.

2. We evaluate the values of $f_{Y_{t+1}|M_{t+1}}(y_{t+1}|m_{t+1})$, $f_{M_{t+1}|M_t,M_{t-1}}(m_{t+1}|m_t, m_{t-1})$ and $f_{M_t,M_{t-1}|Y_{1:t}}(m_t, m_{t-1}|y_{1:t})$ for each point in the mesh grid we created in Step 1.

3. We time the values in Step 2 together along with the grid cube volume d^3 and sum them up to get an approximation for the value of $f_{Y_{t+1}|Y_{1:t}}(y_{t+1}|y_{1:t})$.

4. Then, the value of $f_{M_{t+1},M_t|Y_{1:t+1}}$ can also be approximated by the same manner.

In our study, we set the grid size d equal to 0.0175 to maintain the accuracy of approximation. The other reason for setting $d = 0.0175$ is because of the limit of the Matlab memory size constraint.