Modeling multi-period corporate defaults: macro, contagion and frailty effects in default clustering

by Tuohua Wu

MODELING MULTI-PERIOD CORPORATE DEFAULTS: MACRO, CONTAGION AND FRAILTY EFFECTS IN DEFAULT CLUSTERING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tuohua Wu

August 2010

MODELING MULTI-PERIOD CORPORATE DEFAULTS: MACRO, CONTAGION AND FRAILTY EFFECTS IN DEFAULT CLUSTERING

Tuohua Wu, Ph.D.

Cornell University 2010

This dissertation explores various channels for default clustering. The probability of extreme default losses in U.S. corporate portfolio is much greater than that estimated from model containing only observed macroeconomic variables. The additional sources of default clustering are provided by direct contagion and latent frailty factor. We build a top-down proportional hazard rate model with self-exciting specification. We develop efficient methods of moment for parameter estimation and goodness-of-fit tests for the default counting process. Our estimates are based on U.S. public firms between 1970 and 2008. We find strong evidence that contagion and frailty are equally important in capturing large portfolio losses. Our empirical findings can be used by banks and credit portfolio managers for economic capital calculations and dynamic risk management.

BIOGRAPHICAL SKETCH

Tuohua Wu was born in Wenzhou, China on July 28 1984. In July 2006, he received his Bachelor of Science degree in Mathematics from Tsinghua University in Beijing, China. He entered the doctoral program in Operations Research at Cornell University in Fall 2006. He will obtain his Ph.D. degree in August 2010 with a concentration in Applied Probability and minors in Finance and in Applied Mathematics. He will start working as an Associate in Global Markets Division at Citigroup in New York after graduation.

This document is dedicated to all Cornell graduate students.

ACKNOWLEDGEMENTS

I am truly grateful to my advisor Prof. Martin Wells for his guidance and support during my time at Cornell. I would also like to express my gratitude to Prof. Robert Jarrow and Prof. Sidney Resnick for serving on my committee, as well as the rest of the ORIE faculty for their help.

I sincerely appreciate the assistance of William Anderson from Statistical Science at Cornell University for his support in statistical methodology. I would also like to thank Prof. Stefan Weber from Leibniz Universitat Hannover for introducing me to the field of mathematical finance, and for continuing to provide excellent guidance. I would like to acknowledge the data support from Moody's Default Risk Service.

I would like to thank my parents for their support, encouragement and advice along these years. I am truly grateful to my fiancee Jinlu Cai for her commitment and companionship. I would also like to thank all my friends at Cornell who have made the last four years exciting and enjoyable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It is well accepted that corporate defaults are correlated and default rates vary significantly over time. In this paper, we build a top down aggregate hazard rate model for multi-period corporate defaults. We analyze macro, contagion and frailty effects in decomposing default risk among U.S. firms.

We are particularly interested in the sources for extreme portfolio losses. Including all macroeconomic, contagion and frailty factors provides an adequate assessment for tail distribution of portfolio losses. This methodology sheds light on application in portfolio credit risk in reality. For example, portfolio managers can use this to calculate the quantity of capitals he or she needs to set aside to withstand large losses at the high confidence level. Furthermore, it also provides key tools for estimating probabilities of losses to senior tranche collateralized debt obligations (CDOs). This senior tranche suffer losses only when the losses of underlying bond collateral pool exceed a high percentile.

There are various sources of default clustering in the correlated corporate default literature. One prevailing source is macroeconomic conditions. Examples of this can be found in [56], [21], [25] and many others. These business-cycle related variables affect default probabilities of most firms in the economy. In bottom-up doubly stochastic models, the conditional default hazard rate for each firm is determined by common macroeconomic factors and firm-specific variables. The default clustering effect is reinforced through movement of common observable factors. Defaults are independent conditioning on these common variables. In practice, it is convenient to specify the dynamics of corporate defaults in terms of its intensity. The doubly-stochastic approach makes the

reduced-form model easy to compute and calibrate to the real data.

The authors in [12] show that the defaults in the U.S. industrial firms between 1979 and 2003 can not be captured appropriately by doubly-stochastic models. They transform the time scale using the sum of individual default intensities and test whether defaults on this transformed time scale arrive according to a Poisson distribution. Their research indicates that conditioning on available macroeconomic and firm-specific information is not sufficient for the degree of observed default clustering, and hence rejects the joint hypothesis of the specification of individual default intensities and the conditional independence assumption. [45] argue that we can not reject the assumption of conditional independence if we use a different specification of firms' default intensities. They also indicate that the test procedure in [12] may not capture Hawkes' type contagion and contagion through covariates in the time transformation. From these findings, two other important sources of default clustering have emerged in the credit risk literature.

A first line of research shows that firms are exposed to a dynamic latent factor, also known as frailty. [18] highlight that the frailty factor can capture default clustering above and beyond macroeconomic and firm-specific variables observed in the market. The unobserved factor picks up additional components omitted in the model and hence can have a big impact on the conditional default rates of other surviving firms. The frailty models can be found in [14], [42] and [43].

A second way of inducing default correlation is to use direct contagion. More specifically, the default event of one firm either directly triggers the default of other firms or increases their default probabilities. Some examples of

early contagion models include [13], [39] and others. They specify default contagion as local interaction in networks. [2] and [1] discuss an aggregate contagion that feeds back into the hazard rate dynamics. They use the Hawkes model to explore the empirical role of direct contagion. Hawkes processes, or self-exciting processes, are a class of counting processes whose intensities depend on the arrival of past events. In this model, a default of one firm will raise the likelihood of failure of other firms. For example, the bankruptcy of auto parts manufacturer Delphi in 2005 jeopardized the financial condition of General Motors, Delphi's main purchaser.

These empirical findings imply that both contagion and frailty are important sources of default clustering. The two components generate similar effect on the conditional hazard rate but through different channels. In particular, they exhibit jumps when other firms default in the portfolio. Nevertheless, contagion and frailty have different economic foundations. The influence of contagion is channeled through the complex business relationships among firms. This is typical for firms in the same sector. For instance, the collapse of Penn Central caused 24 transportation companies to go bankrupt on the same day in 1970. On the other hand, the impact of frailty is less direct. The conditional probability of surviving firms is influenced by new information that updates the posterior distribution of latent variables. To illustrate, the collapse of Enron in 2001 revealed the faulty accounting practice that could have been used in other firms, and therefore increased the likelihood of failure of other firms. Both contagion and frailty can be present empirically.

Most of credit risk literature uses bottom up approach for modeling default correlation. While this method is ideal for analyzing default risk for individual

firm, it makes the model calibration computationally intensive if we want to do risk management from a portfolio perspective. [31] introduce the top down method and focus on the default hazard rate at aggregate level. This model has a parsimonious structure and hence facilitates analysis for portfolio credit derivatives such as credit default swaps and collateralized debt obligations. In this work, we adopt a top-down aggregate default intensity for the hypothetical portfolio. Specifically, the Cox proportional hazard rate is introduced, which includes both macro and frailty factors. Furthermore, a Hawkes specification as in [35] is also used in default intensity to account for the direct contagion effect. The empirical result suggests that the model can capture the dramatic fluctuation of U.S. corporate default rates during 1970-2008.

All these tools can be utilized to address the risk analysis and market valuation of collateralized debt obligations (CDO). A CDO is an asset-backed security whose underlying collateral is typically a portfolio of bonds or bank loans. A CDO allocates interest income and principal repayment from a collateral pool of different bonds, and repackage the cash flow according to a prioritized scheme. A standard approach is simple subordination and the prioritized collection of securities are called tranches. Senior CDO tranches get paid before mezzanine and lower subordinated tranches are paid, with any residual cash flow paid to an equity tranch.

CDOs are important to credit risk researchers for studying correlated default risk. CDOs motivate researchers to identify the joint distribution of default risk across firms since they are portfolio credit derivatives. The joint distribution cannot be inferred from the marginal distributions of single-name instruments. The clustering of defaults have a substantial impact on the market valuation

of CDOs. Researchers are also seeking computationally tractable methods to calculate the prices of CDOs effectively.

Both [18] and [2] use maximum likelihood estimation for estimating model parameters. The major challenge in doing so is that the likelihood function becomes intractable since the frailty factor is not adapted to observation filtration. In this project, we use a different estimation technique - Efficient Method of Moments (EMM) introduced by [29]. EMM is used particularly when estimating nonlinear non-Gaussian systems with latent variables. First, a semi-nonparametric model (SNP) is proposed for the approximation of transition densities. The SNP estimating equation is based on an expansion using Hermite functions. The leading term of expansion is considered as a parametric model that approximates the underlying process. A Method-of-Moment-type objective function is then constructed using the log-likelihood of the SNP score generator as moments. The parameters are estimated using a minimum chi-squared criterion. One important advantage of EMM is one obtains the goodness-of-fit test statistics directly from asymptotic results.

We also performed in-sample tests based on the above chi-squared statistics to assess model adequacy. Comparisons among different specifications suggests the full benchmark model is preferred to the model without contagion or frailty. In addition, the posterior mean of frailty is computed by the particle filtering method. Out-of-sample predictions indicate that including frailty in the model can better capture the tail distribution of extreme portfolio loss.

1.1 Related Literature

In the credit risk literature, the structural model of default timing is introduced by [50]. A firm defaults when its asset price drops to a sufficient low level relative to its liability. It is assumed that the firm's asset V follows a geometric Brownian motion:

$$dV_t = \mu V_t dt + \sigma V_t dW_t \tag{1.1}$$

where W is a standard Brownian motion in a fixed probability space. To better understand this model of a firm, they think assets are consisted of very liquid and tangible securities. The price of the asset is the price of these liquidly traded securities.

Now assume that the firm has issued two types of claims: debt and equity at time 0. Debt is a zero-coupon bond with a face value of D and maturity date T. The payoffs at date T to debt B_T and equity S_T, are given as

$$\begin{aligned} B_T &= \min(D, V_T) = D - \max(D - V_T, 0) \\ S_T &= \max(V_T - D, 0) \end{aligned}$$

The firm is considered to be run by the equity owners. At maturity of the debt, equity holders pay the face value of the debt if the asset value is higher than D. If assets are worth less than D, equity owners do not want to pay D since that have limited liability. Bond holders then get the remaining asset and take over the firm. From the structure of the payoffs, equity can be viewed as a call option on firm's asset. Debt can be viewed as the difference between a riskless bond and a put option on firm's asset. We can apply Black-Scholes formula ([50]) to price these options.

Under the structural model, [10] and [57] explain how to construct the distance to default. [21] argue that the distance to default covariate is the most statistically significant variable in their default prediction model. For a given firm, the distance to default is the number of standard deviations of asset growth by which a firm's market value exceeds a liability measure. Mathematically, this is defined as

$$D_t = \frac{\log(\frac{V_t}{L_t}) + (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \tag{1.2}$$

where L_t is a liability measure at time t, known as the "default point" in practice. Following the standard result by Moody's KMV ([10]), L_t is measured as the firm's book value of short-term debt, plus one half of its long-term debt based on its quarterly accounting balance sheet. The asset value V_t and volatility σ are not directly observed, and hence we need to estimate them according to a call option pricing formula.

[57] take the initial asset value V_t to be the sum of S_t (end-of-quarter stock price times number of shares outstanding), and the book value of total debt. They take the riskless rate r to be the one-year T-bill rate. They solve for the asset value V_t and volatility σ by iteratively applying the following equation

$$S_t = V_t\Phi(d_1) - L_t e^{-rT}\Phi(d_2) \tag{1.3}$$

$$\sigma = sdev(\ln(V_t) - \ln(V_{t-1})) \tag{1.4}$$

where

$$d_1 = \frac{\ln(\frac{V_t}{L_t}) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \tag{1.5}$$

and $d_2 = d_1 - \sigma\sqrt{T}$. $\Phi(\cdot)$ is the standard normal cumulative distribution function and $sdev(\cdot)$ represents sample standard deviation.

In the structural model, we need assumptions in the behavior of firm's asset and threshold value for triggering default. It is sometimes difficult to observe

the value of firm's asset in practice and the static assumption regarding the default trigger is also not reasonable. [38] and [22] proposed to use an alternative called reduced-form or intensity-based model. They directly model the conditional hazard rate based on the information filtration without knowledge of firm's asset and liability. The reduced-form model is easier to calibrate to the real default data.

In recent years, researchers have focused attention on portfolio credit risk models. Understanding the default correlation among constituent firms is essential for market evaluation and risk management of credit derivatives. Reduced-form models of portfolio credit risk can be distinguished by the way in which the intensity of the default process is specified. In a bottom up model, the portfolio intensity is an aggregate of the constituent intensities. [21], [12] and [45] provide a good summary of bottom up portfolio models. In a top down model, the portfolio intensity is specified without reference to individual constituents. This parsimonious setup can be found in [31], [4] and [48]. [30] provides an extensive survey for comparison among these two approaches.

Consider a portfolio of credit sensitive securities. The ordered portfolio default times are represented by a sequence of stopping times $T_1 < T_2 < ...$ that is strictly increasing and defined on a complete probability space $(\Omega, \mathcal{F}, P)$. The right-continuous and complete filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ represents the information flow. P can be the actual probability or a risk-neutral measure. The stopping time T_n denotes the n-th default time in the portfolio. Let N be the process that counts default events, given by

$$N_t = \sum_{n\geq 1} 1_{\{T_n \leq t\}}. \tag{1.6}$$

As is shown in [51], for a given filtration, the default process N can be specified

in terms of its compensator, which is the non decreasing predictable process A such that $N - A$ is a local martingale. The compensator contains the expected upward tendency of the default process. [51] shows the compensator in the limit form

$$A_t = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_0^t E[N_{s+\epsilon} - N_s | \mathcal{F}_s]. \tag{1.7}$$

The limit holds weakly in L^1. Equation (1.7) emphasizes the dependence between compensator and filtration. If the times are predictable, i.e. if an event is announced by a sequence of pre-default times, then A is equal to N. As an example, consider the familiar first passage credit models that derived from [3]. Here, a firm defaults if its continuous firm value process falls below a constant barrier. This definition of the default event generates a predictable default time. On the other hand, if the available information is insufficient to determine the precise firm value or default barrier, then the default times are totally inaccessible or unpredictable. In this case defaults come as a surprise and the compensator A is continuous. Unpredictable default times can conveniently be modeled in terms of a non negative, adapted intensity λ that satisfies

$$A_t = \int_0^t \lambda_s ds \tag{1.8}$$

almost surely. If the compensator is of the form (1.8), the credit model is intensity based.

In a top down model the process λ is specified directly. In a bottom up model, λ is an aggregate of constituent intensity processes. The structure of the information filtration $\mathbb{F}$ determines the key properties of a portfolio credit model. The filtration should always be fine enough to distinguish the arrival of events. Therefore the smallest filtration that supports a portfolio credit model is the filtration generated by the default process N itself. Bottom up and top down

models are based on distinct filtrations.

A bottom up model filtration usually contains much more information than the minimal filtration. It is always fine enough to distinguish the identity of each defaulter so that the constituent default times τ_k are stopping times. The filtration may contain additional information about systematic and idiosyncratic risk factors. A constituent default time τ_k generates a default process N^k that is zero before default and one afterwards. If the model is intensity-based, there is a strictly positive intensity process λ^k that represents firm k's conditional default rate. Mathematically, $N_t^k - \int_0^t (1 - N_s^k)\lambda_s^k ds$ is a local martingale.

Since the portfolio default process N is the sum over the constituent default processes N_k and defaults occur at distinct dates almost surely, the portfolio intensity is given by

$$\lambda = \sum_k (1 - N^k)\lambda_k, \tag{1.9}$$

see the references in [31]. The portfolio intensity λ is zero after all firms have defaulted.

In a top down model the researcher specifies the portfolio intensity λ without reference to the constituents. The dependence structure is implicit in this specification. The goal is to establish an intensity model that is more parsimonious than the bottom up portfolio intensity, which follows a complicated process that is driven by the constituent processes and depends on all single name parameters. This is achieved by choosing a model filtration $\mathbb{F}$ that is coarser than the bottom up model filtration. Typically, the top down model filtration is not fine enough to distinguish the identity of a defaulter. This means that an event arrival can be observed, but not the identity of the defaulted name.

Instead of describing the constituent intensities, we focus on the interarrival intensity for the portfolio. In other words, we change the perspective from the firm default times τ_k to the ordered default times T^k. The top down portfolio intensity coincides with the portfolio intensity generated by an exchangeable bottom up model for which all the constitute intensities are identical.

An intensity is usually specified as a stochastic process. We now concentrate on a family of multivariate point processes of correlated event times, whose arrival intensity is driven by an affine jump diffusion. Affine jump diffusion and its application in asset pricing are covered intensively in [20]. In their paper, ordinary differential equations characterize the transform of an affine point process and its probability distribution. The moments of an affine point process have a closed form, which yields computational tractability in applications.

Suppose the intensity is influenced by a set of risk factors that follow stochastic processes on their own. This setup facilitates the inclusion of exogenous jump diffusion risk factors that are related to event arrivals. [2] and [48] impose the presence of diffusive risk factors, which replicate the fluctuation of market prices. Suppose X is a Markov process and it represents a vector of stochastic risk factors. The intensity λ depends on both state vector X and the default counting process N. The transform of $(X, N)^T$ is computationally tractable if X is taken to be an affine jump diffusion.

We call a point process affine if its event arrival intensity is an affine function of an affine jump diffusion and its jump sizes are drawn from a fixed distribution. A Markov process X in a state space $D \subset \mathbb{R}^d \times \mathbb{R}_+$ is an affine jump diffusion

in the sense of [20] if X is a strong solution to the stochastic differential equation

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t + \sum_{i=1}^{m} \zeta^i dZ_t^i, X_0 \in \mathbb{R}^d \tag{1.10}$$

where W is an $\mathbb{R}^d$-valued standard Brownian motion, $\mu : D \to \mathbb{R}^d$ is the drift, $\sigma : D \to \mathbb{R}^{d\times d}$ is the volatility and each Z^i is a $\mathbb{R}^d$-valued point process. The component processes of each vector Z^i share event times and differ only in jump sizes, and we denote their common intensity by $\lambda^i(X_t, t)$ for some $\lambda^i : D \to \mathbb{R}_+$. The jump sizes are drawn from a distribution ν^i on $\mathbb{R}_+^d$. Each parameter ζ^i is a d-dimensional diagonal matrix. The drift, volatility and jump coefficient functions are bounded and continuous on $\mathbb{R}_+$, and they are assumed to have the following affine structure:

$$\begin{aligned}
\mu(x,t) &= K_0(t) + K_1(t)x, K_0(t) \in \mathbb{R}^d, K_1(t) \in \mathbb{R}^{d\times d} \\
(\sigma(x,t)\sigma(x,t)^T)_{jk} &= (H_0)_{jk}(t) + (H_1)(t) \cdot x, H_0(t) \in \mathbb{R}^{d\times d}, H_1(t) \in \mathbb{R}^{d\times d\times d} \\
\lambda^i(x,t) &= \Lambda_0^i(t) + \Lambda_1^i(t) \cdot x, \Lambda_0^i(t) \in \mathbb{R}, \Lambda_1^i(t) \in \mathbb{R}^d, i = 1, 2, ..., m.
\end{aligned}$$

To illustrate the affine point process, we consider a simple example. Suppose the intensity is driven by a one dimensional risk factor X with a single jump term L. L shares common event times with the default counter N. The jump sizes of L are governed by the distribution ν. Assume $K_0(t) = \kappa c$ for $\kappa \geq 0$ and $c > 0$, $K_1(t) = -\kappa$, $H_0(t)$ is a matrix of zeros, $H_1(t)$ is a tensor of zeros except $(H_1)_{111} = \sigma^2$, $X_0 = c$ and $\zeta = \delta \geq 0$. Let $\Lambda_0(t) = 0$ and $\Lambda_1(t) = 1$. Then the intensity λ satisfies the stochastic differential equation

$$d\lambda_t = \kappa(c - \lambda_t)dt + \sigma\sqrt{\lambda_t}dW_t + \delta dL_t. \tag{1.11}$$

The intensity drifts stochastically toward level c with diffusive fluctuations. Furthermore, the point process depends on the past events. This is exactly the same specification given in [2].

Proposition 1 in [20]lays down the foundation for conditional transform of an affine jump diffusion. This transformation is computationally tractable and useful in pricing credit derivatives as argued by [24]. Suppose T is the terminal date. Let $t \leq T$ and $u \in \mathbb{R}^{d+k}$. The conditional transform of $(d + k)$ dimensional affine jump diffusion $Y = (X, Z)^T$ is given by

$$\psi(u, Y_t, t, T) = E(\exp(u \cdot Y_T)|\mathcal{F}_t). \tag{1.12}$$

Under the technical conditions stated in [20], we have

$$\psi(u, Y_t, t, T) = \exp(\alpha(u, t, T) + \beta(u, t, T) \cdot Y_t) \tag{1.13}$$

where the coefficient functions $\beta(t) = \beta(u, t, T)$ and $\alpha(t) = \alpha(u, t, T)$ satisfy the ordinary differential equations

$$\partial_t \beta(t) = -K_1(t)^T \beta(t) - \frac{1}{2}\beta^T H_1(t)\beta(t) - \sum_{i=1}^{m} \Lambda_1^i(t)(\theta^i(\zeta^i \beta(t)) - 1) \tag{1.14}$$

$$\partial_t \alpha(t) = -K_0(t)^T \beta(t) - \frac{1}{2}\beta^T H_0(t)\beta(t) - \sum_{i=1}^{m} \Lambda_0^i(t)(\theta^i(\zeta^i \beta(t)) - 1) \tag{1.15}$$

with boundary conditions $\alpha(T) = 0$ and $\beta(T) = u$ and jump transform

$$\theta^i(c) = \int_{\mathbb{R}_+^{d+k}} e^{c \cdot z} d\nu^i(z) \tag{1.16}$$

[24] extends the work to differentiate the transform with respect to u to get an exponential transform under some regularity conditions. This extended transform embodies the joint distribution of the risk factor X and the point process N. For example, a tranche swap and an option on credit index can be priced.

From the above affine point process, it is indicated that the intensity can depend on the path of the underlying counting process. This process is called a self-exciting process, who is a special case of a self-affecting or path-dependent point process.

A basic example is a Hawkes process, introduced in [35], which is specified by a positive constant c and a nonnegative function $d(\cdot)$. The Hawkes intensity is given by the Stieltjes integral

$$h_t = c + \int_0^t d(t-u)dN_u = c + \sum_{k:T^k<t} d(t-T^k). \tag{1.17}$$

In this specification, the intensity of N is updated with default information along the path. The first default arrives according to a Poisson process with constant intensity c. In other words, T_1 is exponential distributed with parameter c. At the first default, the intensity is updated according to the function $d(\cdot)$. The second default arrives with deterministic intensity $c + d(t - T_1)$ at time t. The k-th default arrives with intensity $c + \sum_{i=1}^{k-1} d(t-T_i)$. The intensity of N at t is influenced by every default in the interval $[0, t]$ as specified by the function $d(\cdot)$.

[36] shows that the Hawkes process can be represented as a Poisson cluster process. Consider a homogeneous Poisson process with intensity c, whose arrivals represent macroeconomic shocks. A shock implies a default, which can have negative ripple effect on other surviving firms and trigger a cluster a new defaults. Given the originating default, the cluster is described by an independent nonhomogeneous Poisson process with deterministic intensity function $d(u)$. Each default can generate further clusters. The Hawkes process is the superposition of the original and the spawned Poisson processes.

Consider a simple example where $d(\cdot)$ is an exponential decay function:

$$d(u) = \sum_{j=1}^{k} \alpha_j e^{-\beta_j u} \tag{1.18}$$

with $\sum_{j=1}^{k} \frac{\alpha_j}{\beta_j} < 1$ for some $k \geq 1$. The constant parameters α_j control the jump size in the Hawkes intensity at the default times. The constant parameters β_j control the decay of the influence of a default on the current intensity.

In the statistical literature, it is well recognized some unobservable random variable can affect the underlying process. Such variable is called a frailty and cannot be observed. They summarize the effects of all the systemic or specific variables that were not put into the basic model. Frailty models have been well known for a long time in survival analysis. In the finance area, they have been proposed recently in the credit risk area. For instance, [7], [55] [14] and [18] impose an indirect influence on the surviving firms by Bayesian updating of the conditional distribution of the latent frailty variables. [55] introduces frailties as static multiplicative factors in the reduced-form model, which allows a closed form formula for the likelihood estimation. These latent factors are often assumed to be Gamma distributed and strong contagion effect can be imposed by the correlation of frailties.

It is not realistic to assume that the frailty over thirty years are static. The unobserved factors that drive the credit risk should be time-dependent. [18] extend the prior methodology to allow a frailty covariate to vary over time according to an autoregressive time-series specification. [26] models the information-driven default contagion where investor only observe noisy price processes of credit derivatives, which leads to a nonlinear filtering problem. [2] proposes affine jump diffusion model for analyzing U.S. corporate defaults. The contagion effect is worked through the self-exciting property of the default counting process. In their article, the diffusion process may not be adapted to the observation filtration. Default correlation is caused by exposure of firms to the diffusion risk factor and uncertainty about the risk level.

[19] allow for three types of credit risk: idiosyncratic or firm-specific default risk, industry-wide default risk in a specific sector and systematic risk in the

whole economy. They focus on individual names and sum up to a portfolio level. Their affine jump diffusion framework leads to exponential affine solutions for survival probabilities, and hence yields analytical tractability for CDO valuations.

[48] also decompose the default risk into three categories in their empirical analysis of CDOs. However, they model portfolio credit losses directly. They allow portfolio losses to occur as the realizations of three separate Poisson processes, each with a different jump size and intensity process. They conduct an extensive empirical analysis of CDO pricing, and provide direct estimates of the nature and degree of default clustering across firms expected by the markets.

1.2 Data

Data on default timing was provided by Moody's Default Risk Service. They contain detailed information on issuer's domicile, industry, rating, as well as default date and default type. The sample period is from January 1970 to March 2008. An issuer is included in our hypothetical portfolio if it is not sovereign, and its effective domicile is the United States and has a senior rating, which is a rating generated by Moody's senior rating algorithm described in [34]. As of March 2008, the data set contains a total of 5471 firms.

We use the same definition of default as Moody's Default Risk Service. Moody's definition of default includes three types of credit events:

1. A missed or delayed disbursement of interest and/or principal, including delayed payments made within a grace period;

2. Bankruptcy, administration, legal receivership, or other legal blocks (perhaps by regulators) to the timely payment of interest and/or principal;

3. A distressed exchange occurs where: (i) the issuer offers debt holders a new security or package of securities that amount to a diminished financial obligation (such as preferred or common stock, or debt with a lower coupon or par amount, lower seniority, or longer maturity); or (ii) the exchange had the apparent purpose of helping the borrower avoid default.

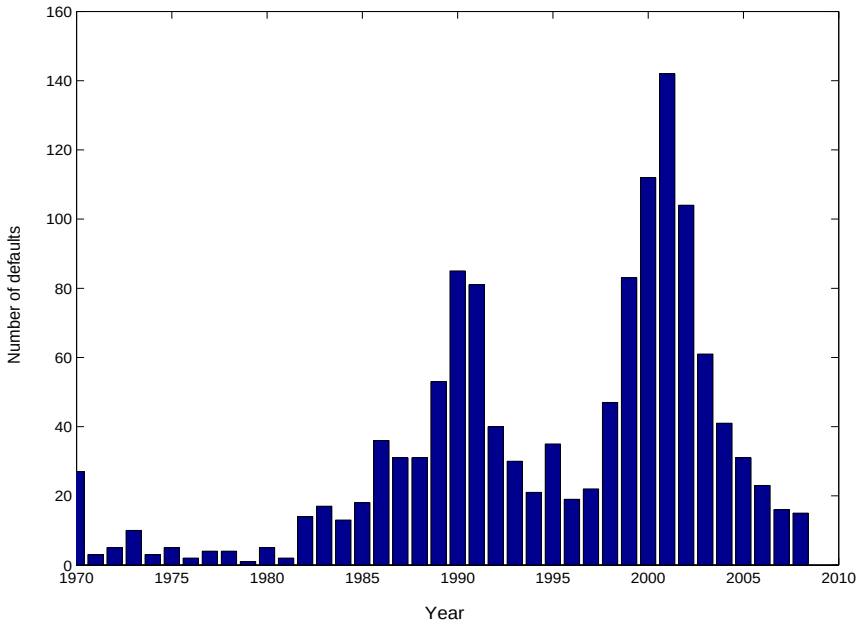

Figure 1.1: The number of defaults for each year between 1970 and 2008.

A repeated default by the same issuer is included in our default events if it is not within one year of the initial default and the issuer's rating is enhanced above Caa after the initial event. We exclude defaults within the same corporate family which occur less than one year apart. We do not treat exits such as merger and acquisition as default.

We observe 1253 defaults during the sample period in the economic wide portfolio. Let $\{T_k\}_{k=1}^n$ denote the ordered default dates in the portfolio. The resolution of events is one day and hence there exists multiple events on the same day whose exact default time can not be distinguished. Although the difference between simultaneous and nearly-timed default is not critical, it is an important distinction for the measurement and estimation procedure. It is not likely that two or more firms would default literally simultaneously unless there is a parent-subsidiary relationship, which is excluded in our default counting. Consequently, we take random perturbation to those dates with more than one default to avoid simultaneous events:

$$T'_k = T_k + U \tag{1.19}$$

where U is uniform distribution in $[-\frac{1}{365}\times 0.5, \frac{1}{365}\times 0.5]$. After this transformation, we can see the exact default time for each event and hence there is only one event at one default time. The new sequence of default events $T'_1 < T'_2 < ... < T'_n$ is strictly increasing and there is no default between two default dates. Figure 1.1 illustrates the total number of defaults in each year.

The macroeconomic time series are observed on a monthly basis. These data series are obtained from Moody's Economy.com.

CHAPTER 2

PORTFOLIO DEFAULT INTENSITY MODELING

This section provides a detailed specification of portfolio intensity model. We build a top-down default hazard rate model that incorporates the time-series dynamics of stochastic covariates. Our method can predict term structures of corporate default probabilities over multiple future periods.

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ is a complete information filtration and $\mathbb{P}$ is actual data-generating probability measure. Let $\mathbb{G} = \{\mathcal{G}_t, t \geq 0\}$ represent a market observation filtration. Since we treat all the firms in one hypothetical portfolio, these constitute firms are fixed in the beginning. In other words, although the number of active firms fluctuates over time, the total number of firms in the portfolio are considered as static. This treatment is different from that of bottom up approach in most literature, where intensities of individual firms become zero upon default and they exit the portfolio immediately. Let N denote the default counting process with increment one.

$$N_t = \sum_{k \geq 1}^{n} 1_{\{T'_k \leq t\}} \tag{2.1}$$

Define a non-negative progressively-measurable process λ to be the default intensity for the portfolio. The process $N_t - \int_0^t \lambda_s ds$ is a local martingale with respect to $\mathbb{F}$. We can interpret the intensity in such a way that conditional on $\mathcal{F}_t$, the probability of credit events between time t and $t+\Delta t$ is approximately $\lambda_t \Delta t$ for small Δt. This is measured in events per unit time, which is one year throughout the paper.

The severe default clustering during some years motivates us to use self-exciting specification following [35], where the portfolio intensity surges

quickly in response to default clusters. [11] and [44] provide a rigorous study of such path-dependent point processes. If we assume the whole default history starts on the first day of our sample period (January 1st 1970), then the self-exciting default intensity exists for our point process as proved in [36]. Mathematically, we define the portfolio default intensity as follows:

$$\lambda_t = \exp(\alpha + \beta \cdot X_t + \eta Y_t) + \int_0^t \delta e^{-\gamma(t-s)} dN_s. \tag{2.2}$$

Where X_t is a vector of macro-economic factors and Y_t is unobserved frailty. $\exp(\alpha)$ serves as a baseline hazard rate. As long as α is finite, the hazard rate defined in (2.2) is non-explosive. The market observation filtration $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ is generated by the observed variables

$$\{X_s, 0 \leq s \leq t\} \cup \{T_k' : \text{for such k that } T_k' \leq t\}.$$

X_t and N_t are hence adapted to the observation filtration $\mathcal{G}_t$. Because $\mathbb{G}$ is smaller than $\mathbb{F}$, the unobserved factor Y is not necessarily measurable with respect to $\mathbb{G}$. The self-exciting specification is similar as the analysis in [31].

If $\delta = 0$, then we have a top-down proportional hazard rate model which parallels the setup in [18]. If we let $\eta = 0$ and $\delta = 0$, then we obtain a point process that is doubly stochastic. More specifically, the top-down doubly stochastic model is defined in [17]: for each $t \geq 0, s > 0$

$$P(N_{t+s} - N_t = k | \mathcal{F}_t \vee \mathcal{G}_{t+s}) = \begin{cases} p(k, \int_t^{t+s} \lambda_u du), & k < m - N_t \\ \sum_{i=m-N_t}^{\infty} p(i, \int_t^{t+s} \lambda_u du), & k = m - N_t \\ 0 & k > m - N_t \end{cases} \tag{2.3}$$

where $p(i, .)$ is the probability mass function of a Possion variable. This is a simplified model that does not take default contagion and latent factor into account.

2.1 Macro Economic Factors

[21] built a forecasting model for credit risk that incorporates the time-series properties of a small group of macro factors. They also include firm-specific variables like KMV's "distance to default" which proves to have considerable explanatory power in their default models. In our top-down model, however, we concentrate primarily on macroeconomic conditions that are related to business cycle. We hence ignore the idiosyncratic risk factors. [25] have done extensive research on including multiple macroeconomic factors into hazard function of their Cox regression model. They specify three broad classes of macroeconomic variables: the overall level of economic activity (e.g., the unemployment rate, inflation rate), the direction in which the economy is moving (e.g., GDP and IP growth rate) and conditions in the financial markets (e.g. interest rates and stock market returns). Inspired by their empirical findings, we consider these variables in our regression models: Chicago Fed National Activity Index (CFNAI), growth rate of industrial production, 3-month Treasury bill rate, 10-year Treasury yield, slope, annual S&P 500 return, annual volatility of S&P 500 and TED spread. The CFNAI is a composite series that summarizes economic behavior including production and income, unemployment rate, and personal consumption. Slope is the difference between 10 year and 1 year Treasury bill rates, which reflects the term structure of yield curve. The TED spread is defined as the difference between 3-month LIBOR and 3-month T-bill rate, which serves as an indicator for credit spread.

Empirical default studies often introduce macroeconomic factors as contemporaneous variables. In practice, however, it is very likely that these macroeconomic variables would have a delayed effect on defaults. As suggested by [25],

we also allow lagged values of covariates to enter our model. We define each variable as a weighted average over a fixed window with exponentially declining weights. Let $\{x_t, t = 0, 1, ..., T\}$ be the raw monthly data series for a given variable. X_t represents the weighted value in the model. X_t is given by

$$X_t = \frac{\sum_{j=0}^{K} \omega^j x_{t-j}}{\sum_{j=0}^{K} \omega^j} \tag{2.4}$$

where K is the length of the lag window and ω is the exponential decay factor. The new covariate includes data in the current month. We identify our lagged period to one year and hence choose $K = 12$ months. We let $\omega = 0.83$ so that the impact of variable 12 months ago is roughly 10 % of its current counterpart. Note that there is no priori best choice for both parameters and the results are not sensitive to different choices of values. We also want our covariates to have the same magnitude so that we can compare their relative importance in determining the hazard rate. Therefore, we scale our weighted covariates by their standard deviations, and this procedure results in series of covariates with standard deviation one.

In order to pick fewer variables, we simply run regression model using monthly default number of firms against all the above macro factors. We use Mallows' C_P statistic as in [49] to choose the best regression model. If P predictors are selected from a total set of K variables ($P < K$), C_P statistic is defined as:

$$C_P = \frac{SSE_P}{SSE_{full}} - N + 2P \tag{2.5}$$

where SSE_p is the error sum of squares for the model with P predictors, SSE_{full} is the error sum of squares for the full model with K predictors and N is sample size. When a model has little bias, the expectation of C_P is close to $P + 1$. We prefer the model with smaller C_P statistic.

The smallest C_P gives a model with four statistically significant factors: 1-year S&P 500 return, 3-month T-bill rate, slope and annual industrial production growth rate. This resulting C_P is a little larger than 5 and the model turns out to be relative parsimonious specification.

2.2 Modeling Stochastic Covariates

We want to specify continuous-time dynamics for the macroeconomic factors and build a forecasting default hazard rate model. [21] use first-order Gaussian vector autoregression to model their covariates. We propose to apply stochastic differential equations to model the evolution of macroeconomic variables. The continuous-time specification can better characterize the risk factor dynamics in reality.

Let S_t be the trailing 1-year S&P 500 stock return. Let us consider a stochastic volatility model with a leverage effect (correlation between stock return and volatility):

$$\begin{aligned} dS_t &= a_1(a_0 - S_t)dt + \exp(u_t)dW_t^{(1)} \\ du_t &= b_1(b_0 - u_t)dt + \sigma_1(\rho_1 dW_t^{(1)} + \sqrt{1-\rho_1^2}dW_t^{(2)}) \end{aligned} \tag{2.6}$$

where $W^{(1)}$ and $W^{(2)}$ are independent Brownian motions with respect to $(\Omega, \mathcal{F}, \mathbb{P})$. In order to achieve parameter identifiability, we need to prefix some parameter values. We estimate a_0 from simple Ornstein-Uhlenbeck process instead of stochastic volatility model. After that, we assume the Brownian motion driving the volatility factor is independent of that in the stock return. We obtain the following result: $a_0 = 0, b_0 = -0.65, \sigma_1 = 0.50$ and fix them in our two-factor

stochastic volatility model. Stock return data is stationary and usually has mean close to zero which is also the case here.

Cox-Ingersoll-Ross (CIR) process is historically popular in modeling short term interest rates in academia. However the oversimplified specification can hardly fit the actual data well. We also proposed 2-factor stochastic volatility model for 3-month T-bill rates r_t.

$$\begin{aligned} dr_t &= \phi_1(\phi_0 - r_t)dt + \sqrt{r_t}\exp(v_t)dW_t^{(3)} \\ dv_t &= \kappa_1(\kappa_0 - v_t)dt + \sigma_2 dW_t^{(4)} \end{aligned} \tag{2.7}$$

where $W^{(3)}$ and $W^{(4)}$ are independent standard Brownian motions with respect to $(\Omega, \mathcal{F}, \mathbb{P})$. The interest rate model is known to be difficult to calibrate because of some huge spikes in 70s and 80s. Therefore, we detrend and smooth the time series. For the purpose of identifiability, we estimate $\phi_0 = 2.20$ and $\kappa_0 = -0.51$ from one-factor CIR process $dr_t = \phi_1(\phi_0 - r_t)dt + \sqrt{r_t}\exp(\kappa_0)dW_t^{(3)}$, and afterwards we prefix ϕ_0 and κ_0 in the stochastic volatility model.

The Ornstein-Uhlenbeck (OU) processes are specified for both slope C_t and annual growth rate of industrial production p_t. The OU process can be treated as continuous version for AR(1) model. Note that during some periods, slope could be negative and hence the OU model is appropriate in this case.

$$dC_t = \theta_1(\mu_1 - C_t) + \sigma_3(\rho_2 dW_t^{(3)} + \sqrt{1-\rho_2^2}dW_t^{(5)}) \tag{2.8}$$

$$dp_t = \theta_2(\mu_2 - p_t) + \sigma_4 dW_t^{(6)} \tag{2.9}$$

where $W^{(5)}$ and $W^{(6)}$ are independent standard Brownian motions with respect to $(\Omega, \mathcal{F}, \mathbb{P})$. We impose correlation between slope and 3-month T-bill rate. Notice that all the Brownian motions $W^{(i)}, i = 1, 2, ..., 6$ are independent. For the same identifiability reason, we estimate $\mu_1 = 1.30, \sigma_3 = 0.35$ and $\mu_2 = 0.60$ from

independent one-factor OU processes and fix them in our model.

2.3 Contagion and Frailty

The contagion is modeled as Hawkes process:

$$\int_0^t \delta e^{-\gamma(t-s)} dN_s = \sum_{k:T_k<t} \delta e^{-\gamma(t-T_k)}. \tag{2.10}$$

Mathematically, the intensity jumps up each time a firm defaults, and decreases exponentially afterwards. Whenever one firm defaults, it jeopardize the credit conditions of surviving firms. This effect is substantial within firms in the same industry sector and the same corporate family. Our extended model with direct contagion can better assess extreme default clustering in adverse environment.

Furthermore, the portfolio intensity can have frailty effect simultaneously although it has different economic mechanism. The frailty y_t is a latent variable that is not directly observed by market players. The impact of frailty is through information update and is less direct than contagion. We can view frailty as correction to the portfolio intensity besides macro covariates, it could be different macroeconomic variable at various time periods. For instance, it represented defection in mark-to-market methods and accounting rules among the corporate when Enron went bankrupt. It could be potential large consumer loan losses when the credit crunch accelerated in September 2008.

We assume that the frailty factor is orthogonal to the existing macroeconomic variables since it provides an additional channel for credit risk. We also assume the frailty is detrended in the sense that it is mean reverting around level 0. Each time there is latent shock in the economy, its effect would decay steadily as time

passes by. This motivates us to use the Ornstein-Uhlenbeck process as in [18] to model frailty Y_t:

$$\begin{aligned} dY_t &= -\kappa Y_t dt + dW_t^{(7)} \\ Y_0 &= 0 \end{aligned} \tag{2.11}$$

where $W_t^{(7)}$ is a standard Brownian motion with respect to $(\Omega, \mathcal{F}, \mathbb{P})$ and is independent of $W_t^{(i)}, i = 1, ..., 6$. The parameter κ is a non-negative mean reversion rate of Y. The transitional density of OU process is normal and hence facilitates the filtering procedure of its posterior path. Without loss of generality, we fix the volatility coefficient to be unity because the scaling parameter η before Y_t plays the same role in this model. Furthermore, the starting value and mean reverting level of Y are set to be 0 because the nonzero value can be absorbed in the constant term α in the aggregate intensity. The frailty model can capture the accumulative effect of various types of unobserved fundamental shocks to default intensity. The default contagion can be reinforced through two channels: direct ripple effect of other firms' defaults and unobserved economic factor.

CHAPTER 3

ESTIMATION VIA EMM

This section develops estimators for unknown parameters in the aggregate default intensity model. Often, simple and tractable models are proposed initially by researchers to describe observed phenomena. [18] and [2] both take advantage of maximum likelihood estimation for their time-series credit risk models. Nevertheless, the MLE approach becomes difficult when tractable likelihood function is not available. With the difficulty of obtaining the likelihood, researchers often turn to moment matching methods, such as minimum chi-square in the statistics literature or GMM in the economics literature, as a method to overcome computational difficulties. However, even moment matching methods such as the vanilla GMM, suffer computational difficulties especially when unobservable variables enter the model nonlinearly. These unobservable lead to multiple integrals in the criterion function in which standard numerical techniques are not applicable. Prominent examples are continuous time interest rate and stochastic volatility type models. The use of simulation, such as Monte Carlo, offers an avenue to overcome these otherwise intractable estimation problems.

[29] propose the efficient method of moments (EMM) for estimating challenging time series models in finance. This method becomes particularly useful when we are considering a nonlinear dynamic system with unobserved variables. The basic procedure of EMM consists of two steps. The first step is to estimate the joint density of the observed data. This approximating density is called the auxiliary model. The second step is to use the scores from the auxiliary model as moments to construct a minimum chi-squared estimator or GMM

type estimator. These scores are then evaluated using the simulation output of the so-called structural model, which is a model believed to have generated the observed data. The objective function is then minimized with respect to the parameters of the structural model. There are a whole host of diagnostic tests available under this approach. In particular, the standardized objective function asymptotically follows a chi-squared distribution with degrees of freedom equal to the number of scores less the number of model parameters. This provides an overall goodness-of-fit test.

More formally, let $\{x_t\}_{t=-\infty}^{+\infty} \in \mathbb{R}^M$ denote a M-dimensional discrete, stationary stochastic process. Denote $\{\tilde{x}_t\}_{t=-L}^{T}$ as realizations from the process x_t. Let π denote the vector of unknown parameters for a dynamic system. If the observed stochastic process $\{x_t\}_{t=1}^{T}$ is stationary, then a time invariant transition density exists for any finite lag L. Denote these transition densities by:

$$\{p_0(x_0|\pi), [p_t(x_t|x_{t-1}, ..., x_{t-L+1}, \pi)]_{t=1}^{T}\}$$

where L represents a finite lag. We estimate π by means of the score functions $\frac{\partial}{\partial\theta} \log f_t(x_t|x_{t-1}, ..., x_{t-L+1}, \theta)$ from a sequence of approximating densities:

$$\{f_0(x_0|\theta), [f_t(x_t|x_{t-1}, ...x_{t-L+1}, \theta)]_{t=1}^{T}\}$$

where θ is a vector of parameters in this auxiliary model. Write $y_{t-1} = (x_{t-1}^T, ...x_{t-L+1}^T)^T$, then the auxiliary densities can be expressed in a more compact way $[f_t(x_t|y_{t-1}, \theta)]_{t=1}^{T}\}$. For simplicity, we often omit the time subscript and write x and y for the contemporaneous value and lagged state vector.

3.1 Projection

In our application of the EMM, we use the SNP approach of [29] as the score generator. In essence SNP projects onto the image space of Hermite polynomials of degree K to approximate the data density function. Hermitian densities are easy to evaluate and differentiate. Also, the moments are easy to evaluate since they correspond to higher moments of the Gaussian and can be computed using standard recursions. The idea is the leading term of expansion is a parametric model that approximates the process and higher order terms capture deviation from this approximation. To be more precise, let $P(z)$ denote a multivariate polynomial of degree K_z:

$$P_{K_z}(z) = \sum_{|\alpha|=0}^{K_z} a_\alpha z^\alpha \tag{3.1}$$

where $z^\alpha = z_1^{\alpha_1} \cdot z_2^{\alpha_2} \cdots z_M^{\alpha_M}$ and $|\alpha| = \sum_{i=1}^{M} |\alpha_i|$.

A non-normalized Hermite function has the form

$$P_{K_z}(z)\sqrt{\phi(z)} \tag{3.2}$$

$\phi(z)$ is a M-dim multivariate Gaussian density:

$$\phi(z) \propto N(0, I_M) = \frac{1}{(2\pi)^{M/2}} \exp(-\frac{1}{2} z^T z).$$

Let z_t represent a stationary innovation process. Let $h(z)$ be its probability density, then $\sqrt{h(z)}$ is squared integrable. We know Hermite functions are orthogonal basis for L_2 space under a weighted Sobelov norm. Thus they are are dense in L_2. Consequently it can be shown that the following Hermite expansion exists for $\sqrt{h(z)}$:

$$\sqrt{h(z)} \;=\; \sum_{|\alpha|=0}^{\infty} a_\alpha z^\alpha \sqrt{\phi(z)}. \tag{3.3}$$

When K_z is large enough, the truncated expansion

$$\sqrt{h_{K_z}(z)} \;=\; \sum_{|\alpha|=0}^{K_z} a_\alpha z^\alpha \sqrt{\phi(z)} \tag{3.4}$$

provides a reasonable approximation for $\sqrt{h(z)}$. Thus the truncated density can be written as

$$h_{K_z}(z) = \frac{P_{K_z}(z)^2 \phi(z)}{\int P_{K_z}(s)^2 \phi(s) ds}. \tag{3.5}$$

The polynomial enters as a square to ensure positivity. $\int P_{K_z}(s)^2 \phi(s) ds$ is a normalizing constant since the integral of a density should equal to 1.

We reduce the stochastic process x_t to an innovation process z_t by introducing a location and a scale function. Thus, we define the following location-scale transformation:

$$x = Rz + \mu \tag{3.6}$$

where R is an upper triangular matrix and μ is a M-dimensional vector. Through this transformation, we generate a family of densities

$$f(x|\theta) \propto \frac{P_{K_z}[R^{-1}(x-\mu)]^2 \phi[[R^{-1}(x-\mu)]}{|det(R)| \int P_{K_z}(s)^2 \phi(s) ds}. \tag{3.7}$$

If we write $\Sigma = RR^T$, then

$$f(x|\theta) \propto P_{K_z}[R^{-1}(x-\mu)]^2 \cdot N_M(x|\mu, \Sigma) \tag{3.8}$$

where θ consists of coefficients of $P(z), \mu$ and R. If K_z equals zero, the result density is Gaussian. When K_z is positive, the corresponding density is Gaussian with modified shape, which can capture heavy-tail behavior.

To allow for heterogenous innovation, it proves to be more appropriate to let coefficients of polynomial depend on the lagged value y. More specifically, write $P(z, y)$ in a rectangular expansion:

$$P(z, y) = \sum_{|\alpha|=0}^{K_z} (\sum_{|\beta|=0}^{K_y} a_{\alpha\beta} y^\beta) z^\alpha \tag{3.9}$$

where $\beta = (\beta_1, \beta_2, \cdots, \beta_{ML})$, $|\beta| = \sum_{i=1}^{ML} |\beta_i|$ and $x^\beta = \Pi_{i=1}^{ML}(x_i)^{\beta_i}$. $P(z, y)$ can be assumed to depend on $L_P \leq L$ lags of y by putting some elements of matrix $A = [a_{\alpha\beta}]$ to 0. Because $P(z, y)^2 / \int P(s, y)^2 \phi(s) ds$ is a homogeneous function of the coefficients of the polynomial $P(z, y)$, $P(z, y)$ can only be determined to within a scaler multiple. To achieve a unique representation, we restrict the constant term $a_{00} = 1$

It is advantageous in applications to allow the scale R to depend on y because it reduces the degree K_y required to achieve an adequate approximation to the transition density $p(x|y, \pi)$. Therefore, we describe the conditional density $f(x|y, \theta)$ by the modified location-scale transformation

$$\begin{aligned} x_t &= R_{y_{t-1}} z_t + \mu_{y_{t-1}} \\ \mu_y &= b_0 + B y_{t-1} \\ \Sigma_{y_{t-1}} &= R_{y_{t-1}} R_{y_{t-1}}^T \end{aligned} \tag{3.10}$$

where b_0 and B are constant vector and matrix respectively. $\mu_{y_{t-1}}$ depends on $L_u \leq L$ lags, which is achieved by putting leading columns of B to 0.

Two choices of R_y that have good results in applications are an ARCH-like moving average and a GARCH-like ARMA specification. Let vech(R) denote a vector of length $M(M + 1)/2$ containing the elements of the upper triangle of R_y.

$$vech(R_{y_{t-1}}) = \rho_0 + \sum_{i=1}^{L_r} P_{(i)} |y_{t-1+L_r+i} - \mu_{y_{t-2-L_r+i}}| \tag{3.11}$$

where ρ_0 is a vector of length $M(M + 1)/2$, $P_{(1)}$ through $P_{(L_r)}$ are $M(M + 1)/2$ by M matrices. $|y - \mu|$ take element-wise absolute values of $y - \mu$. This scale function depends on L_r lagged innovations $y_t - \mu_{y_{t-1}}$ and $L_r + L_u \leq L$ lagged y_t in total.

For a GARCH specification, let

$$vech(R_{y_{t-1}}) = \rho_0 + \sum_{i=1}^{L_r} P_{(i)} |x_{t-1+L_r+i} - \mu_{y_{t-2-L_r+i}}| + \sum_{i=1}^{L_g} diag(G_{(i)}) R_{y_{t-2-L_g+i}} \tag{3.12}$$

where $G_{(i)}, i = 1, ..., L_g$ are vectors of length $M(M+1)/2$. The SNP version of GARCH is expressed in terms of absolute value of lagged residuals and standard deviations.

The SNP density in this case is

$$f(x_t|y_{t-1}, \theta) \propto P^2[R_y^{-1}(x_t - \mu_y), y_{t-1}] \cdot N_M(x_t|\mu_y, \Sigma_y). \tag{3.13}$$

We shall distinguish various lag values appearing in different components of the expansion. The number of lags in μ_y is denoted by L_u; the number of lags in R_y is L_r, and the number of lags in the y part of polynomial is L_p. Besides, large values of M can generate a large number of interactions (cross product terms) for even modest setting of K_z. Accordingly, we introduce two additional tuning parameters, I_z and I_y, to filter higher order of interactions. $I_z = 0$ means no interaction coefficients of z are set to 0. $I_z > 0$ indicates that coefficients corresponding to interactions z^α of order larger than $K_z - I_z$ are set to 0. Similar definition applies for y_β and I_y. In summary, L_u, L_r, L_g determine the location-scale transformation $x = R_y z + \mu_y$ and hence the nature of the leading term of the expansion. K_z, K_y, I_z and I_y determine the degree of $P(y, z)$ and hence the property of the innovation process z_t.

If K_z, K_y and L_r are put to 0, then the SNP model defines a Gaussian vector autoregression. If K_y and L_r are put to 0, then the model defines a non-Gaussian vector autoregression model with homogeneous innovation. If K_z and K_y are set to 0, then the model defines a Gaussian ARCH model. If $K_z > 0, K_y > 0, L_p > 0, L_u > 0$ and $L_r > 0$, then the model defines a general nonlinear process with heterogeneous innovation.

The tuning parameters of the SNP density is characterized by $(L_u, L_r, L_p,$

K_z, I_z, K_y, I_y). They may be determined by a statistical model selection criterion. One that works well is the Schwarz BIC criterion as in [8], which is computed as

$$\ell_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log f(\tilde{x}_t|\tilde{y}_{t-1}, \theta) \tag{3.14}$$

$$\hat{\theta} = arg\max_{\theta} \ell_T(\theta) \tag{3.15}$$

$$BIC = -\ell_T(\hat{\theta}) + (1/2)(p_\theta/T)\log(T) \tag{3.16}$$

where p_θ is the number of components in θ and $\tilde{x}_t$ is observed time series data. ℓ_T is the log likelihood for the sample. $\hat{\theta}$ is the quasi-maximum likelihood estimator, by fitting the SNP model to the realizations $\{\tilde{x}_t\}$. We choose $(L_u, L_r, L_p, K_z, I_z, K_y, I_y)$ to minimize the BIC. The criterion rewards good fits to actual data but penalize good fit obtained from excessive parameterizations.

A strategy found to work well is to move upward along a tree structure using the BIC criterion. Expand first in L_u with $L_r = L_p = K_z = K_y = 0$ until the BIC turns upward. Next, expand L_r with $L_p = K_z = K_y = 0$; then expand K_z with $K_x = 0$; and finally expand L_p and K_y.

3.2 Estimation

Our objectives are to estimate the vector of unknown parameters π and to test the hypothesis that the structural model under consideration generated the observed data $\{\tilde{x}_t\}_{t=-L}^{T}$. It is usual in practice that $p(x_t|y_{t-1}, \pi)$ does not have an analytic expression. However, the key feature of this data generating process is that it is relatively easy to compute the expectation of this nonlinear function given values for the structural parameters. The expectation

$E_\pi(\Psi) = \int \Psi(x_{-L}, \cdots x_0|\pi) p(x_{-L}, \cdots x_0|\pi) dy_{-L} \cdots dy_0$ can be calculated by simulation. For given π, we generate sample path $\{\hat{x}_t\}_{t=-L}^{N}$ from the system where N is sufficient large so that the Monte Carlo error is negligible.

We can construct a Method of Moments (MoM) estimator by specifying a moment function

$$\tilde{\Psi}(x_{-L}, \cdots x_0) = \begin{pmatrix} x_0 - \tilde{\mu}_1 \\ x_0^2 - \tilde{\mu}_2 \\ \vdots \\ x_0^k - \tilde{\mu_k} \\ x_{-1}x_0 - \tilde{\gamma}_1 \\ \vdots \\ x_{-L}x_0 - \tilde{\gamma}_L \end{pmatrix}$$

where $\tilde{\mu}_j = \frac{1}{T}\sum \tilde{x}_t^j$, and $\tilde{\gamma}_h = \frac{1}{T}\sum_{t=h+1}^{T} \tilde{x}_t \tilde{x}_{t-1}$. Compute the moment equation

$$m(\pi) = E_\pi(\tilde{\Psi}) = \int \tilde{\Psi}(x_{-L}, \cdots x_0|\pi) p(x_{-L}, \cdots x_0|\pi) dy_{-L} \cdots dy_0 \tag{3.17}$$

The moment function can be approximated by computing

$$m(\pi) = \frac{1}{N}\sum_{t=1}^{N} \tilde{\Psi}(\hat{x}_{t-L}(\pi), \cdots \hat{x}_t(\pi)) \tag{3.18}$$

when N is large enough. Here $\hat{x}(\pi)$ are drawn from the true data generating distribution $p(.|\pi)$, or known as the structural model defined by stochastic differential equations. The estimators are solved by equation $m(\pi) = 0$.

If a solution can not be found by the above estimating equations, for example when the number of moment equations is greater than the number of parameters p_π, we should resort to minimum chi-squared estimation. A strategy for minimum chi-squared estimation is to mimic the first order conditions (scores) of the quasi maximum likelihood estimator in (3.15). This method minimizes

the quadratic form of moment equations. As seen later, the optimal weighting matrix for forming the minimum chi-square criterion depend only on the auxiliary model and it is easily computed. The moment equations for EMM estimator are obtained from so-called score generator, which is the score function $\frac{\partial}{\partial\theta}\log f(x_t|y_{t-1},\theta)$ of the auxiliary model $f(x_t|y_{t-1},\theta)$.

Using the score generator and the actual data $\{\tilde{x}_t\}_{t=-L}^{N}$, recall the quasi-maximum likelihood estimation of the score generator

$$\hat{\theta} = arg\max_{\theta}\frac{1}{T}\sum_{t=1}^{T}\log f_t(\tilde{x}_t|\tilde{y}_{t-1},\theta). \tag{3.19}$$

Define the moment function

$$m(\pi,\theta) = E_{\pi}[\log\frac{\partial}{\partial\theta}f_t(x_t(\pi)|y_{t-1}(\pi),\theta)] \tag{3.20}$$

which is computed by averaging over a long simulation:

$$m(\pi,\hat{\theta}) = \frac{1}{N}\sum_{t=1}^{N}\log\frac{\partial}{\partial\theta}f_t(\hat{x}_t(\pi)|\hat{y}_{t-1}(\pi),\hat{\theta}). \tag{3.21}$$

Typically $p_\theta > p_\pi$. The minimum chi-squared estimator is defined

$$\hat{\pi}_N = arg\min_{\pi} m^T(\pi,\hat{\theta})\hat{\mathcal{I}}_N^{-1}m(\pi,\hat{\theta}) \tag{3.22}$$

where $\hat{\mathcal{I}}_N$ is an estimate of the variance $\sqrt{N}m(\pi,\hat{\theta})$, which you can consider as a weighting matrix.

If you believe the score estimator approximates the data generating process, a good estimator of variance is proposed by [29]

$$\hat{\mathcal{I}}_N = \frac{1}{N}\sum_{t=1}^{N}[\frac{\partial}{\partial\theta}\log f(\tilde{x}_t|\tilde{y}_{t-1},\hat{\theta})][\frac{\partial}{\partial\theta}\log f(\tilde{x}_t|\tilde{y}_{t-1},\hat{\theta})]^T. \tag{3.23}$$

If the belief is not as strong as above, then an estimator is

$$\hat{\mathcal{I}}_N = \sum_{\tau=-(n)^{1/5}}^{(n)^{1/5}}\omega(\frac{\tau}{[n^{1/5}]})\tilde{S}_{T\tau}$$

where

$$\omega(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & 0 \le x \le \frac{1}{2} \\ 2(1-|x|)^3 & \frac{1}{2} \le x \le 1 \end{cases}$$

and

$$\tilde{S}_{N\tau} = \begin{cases} \frac{1}{N}\sum_{t=1+\tau}^{N}[\frac{\partial}{\partial\theta}\log f(\tilde{x}_t|\tilde{y}_{t-1},\hat{\theta})][\frac{\partial}{\partial\theta}\log f(\tilde{x}_{t-\tau-1}|\tilde{y}_{t-\tau-1},\hat{\theta})]^T & \tau \ge 0 \\ \tilde{S}_{N,-\tau} & \tau < 0 \end{cases}$$

3.3 Asymptotics and Goodness-of-fit

Some asymptotic results were established for EMM estimators in [28] and [29]. If π^0 denotes the true value of π and θ^0 is an isolated solution of the moment equation $m(\pi^0, \theta) = 0$, then under regularity conditions specified by [27] and [29], Theorem 1 in [29] states that

$$\lim_{N\to\infty} \hat{\pi}_N = \pi^0 \quad a.s.$$

$$\sqrt{N}(\hat{\pi}_N - \pi^0) \to N(0, [(M^0)^T(\mathcal{I}^0)^{-1}(M^0)]^{-1})$$

$$\lim_{N\to\infty} \hat{M}_N = M^0 \quad a.s.$$

$$\lim_{N\to\infty} \hat{\mathcal{I}}_N = \mathcal{I}^0 \quad a.s.$$

where $\hat{M}_N = M(\hat{\pi}_N, \hat{\theta}_N)$ $M^0 = M(\pi^0, \theta^0)$, $M(\pi, \theta) = \frac{\partial}{\partial\pi} m(\pi, \theta)$ and

$$\mathcal{I}^0 = E_{\pi^0}[\frac{\partial}{\partial\theta}\log f(x_0|y_{-1},\theta^0)][\frac{\partial}{\partial\theta}\log f(x_0|y_{-1},\theta^0)]^T.$$

Under the null hypothesis that $\{p_0(x_0|\pi), [p(x_t|y_{t-1},\pi)]_{t=1}^T\}$ is correct model, then

$$L_0 = N \cdot m^T(\hat{\pi}_N, \hat{\theta})\hat{\mathcal{I}}^{-1} m(\hat{\pi}_N, \hat{\theta}) \tag{3.24}$$

is asymptotically chi-squared distributed with $p_\theta - p_\pi - 1$ degree of freedom.

The estimator is root-n consistent and asymptotically normal with an asymptotic distribution that depends on both the structural model and the score generator. [29] have shown that if there exists a local smooth mapping of the structural parameters into the parameters of the score generator (Assumption 3 in [29]), then the estimator has the same asymptotic distribution as the maximum likelihood. In other words, the estimator is as efficient as maximum likelihood estimator under the structural model. If the score generator closely approximates the structural model, even though it does not nest it, then the estimator is nearly efficient. [28] showed that if the score generator is the SNP density, then the efficiency of EMM estimator can be made close to that of MLE by making $K = K_z + K_y$ large enough.

Based on that, we can carry out goodness-of-fit test. A Wald confidence interval on an element π_i of π can by constructed in the usual way from an asymptotic standard error of $\sqrt{\hat{\sigma}_{ii}}$. A standard error can be computed numerically from the Jacobian $m(\pi, \theta)$ and we take the estimated asymptotic variance $\sqrt{\hat{\sigma}_{ii}}$ to be the ith diagonal element of $\hat{\Sigma} = \frac{1}{N}[\hat{m}^T(\hat{I}_N)^{-1}\hat{m}]^{-1}$. These intervals which are symmetric, are sometime misleading because they do not reflect the rapid increase in the EMM objective function when π_i approaches a value for which the system is explosive. Confidence interval obtained by inverting the criterion difference test L_0 can address this issue and are therefore more advantageous. To invert the test, one puts in the interval those π_i^* for which L_0 for the hypothesis $\pi_i^0 = \pi_i^*$ is less than the critical point of a chi-square on one degree of freedom. To avoid re-optimization one may use the approximation

$$\tilde{\pi}_N = \hat{\pi}_N + \frac{\rho_i^* - \hat{\rho}_{iN}}{\hat{\sigma}_{ii}}\hat{\Sigma}_{(i)} \tag{3.25}$$

in the formula for L_0 where $\hat{\Sigma}_{(i)}$ is the i-th column of $\hat{\Sigma}$.

When L_0 exceeds the chi-squared critical point, diagnostics that suggest improvements to the system are desirable. This is because that

$$\sqrt{N}m(\hat{\pi_N}, \hat{\theta}) \longrightarrow N(0, I^0 - M^0[(M^0)^T(I^0)^{-1}(M^0)]^{-1}(M^0)^T)$$

inspection of the t-ratios

$$T_N = S_N^{-1}\sqrt{N}m(\hat{\pi_N}, \hat{\theta})$$

where $S_N = (diag\{\tilde{I} - \hat{M}[(\hat{M})^T(\tilde{I})^{-1}(\hat{M})]^{-1}(\hat{M})^T\})$ and $\hat{M} = M(\hat{\pi}_N, \hat{\theta})$ can suggest reasons for failure. Different elements of the score correspond to different characteristics of the data and large t-ratios reveal those characteristics are not well approximated.

CHAPTER 4

PARTICLE FILTERING FOR POSTERIOR FRAILTY PATH

Since the frailty factor is unobserved, we need to infer its the posterior distribution. Through this procedure, we can interpret the model better and apply it to the computation of portfolio losses. The model under our consideration falls into the class of nonlinear, non-Gaussian state space models for which we apply particle filtering and smoothing algorithms. In this section, we first introduce the framework of Bayesian filtering, and then methods of particle filtering are illustrated.

4.1 Bayesian Filtering and Smoothing Framework

[6] provides a detailed survey for Bayesian filtering and its rich variations in the literature. State space models provide a convenient framework for analyzing a dynamic system via transition and measurement equations. The state variable x_t is not directly observed and carries information about the evolution of the system. The state also follows a first-order Markov process $p(x_t|x_{0:t-1}) = p(x_t|x_{t-1})$. It satisfies the transition equation

$$x_t = f(x_{t-1}, d_t). \tag{4.1}$$

The observations y_t are independent given the state variables x_t. The y_t are related to x_t through the measurement equation

$$y_t = g(x_t, v_t) \tag{4.2}$$

where d_t and v_t are i.i.d. random variables with known probability density functions. The functions $f(\cdot)$ and $g(\cdot)$ are known but can be nonlinear. The transition

density $p(x_t|x_{t-1})$ and measurement density $p(y_t|x_t)$ are determined by the densities of d_t and v_t. In general, the state variable x_t can be either discrete-valued, continuous-valued or mixture of the two.

All information about the latent states $\{x_t\}_{t=0}^T$ given the observations $\{y_t\}_{t=1}^T$ can be extracted from the joint posterior distribution $p(x_{0:T}|y_{1:T})$. Estimating recursively of the posterior distribution is the main goal of the optimal filtering and forecasting.

The first marginal we are interested in is the one-step ahead predictive density of the state variable

$$p(x_t|y_{1:t-1}) = p(x_t|y_1, ..., y_{t-1}) \tag{4.3}$$

which utilize information up to time $t-1$ to make one-step ahead forecast of the state. The second marginal is the filtering distribution

$$p(x_t|y_{1:t}) = p(x_t|y_1, ..., y_t) \tag{4.4}$$

which uses contemporaneous observations. Finally the smoothing distribution carries all the information in the sample to estimate the past realizations of the state

$$p(x_t|y_{1:T}) = p(x_t|y_1, ..., y_T). \tag{4.5}$$

In the following, we present a detailed derivation of recursive Bayesian estimation, which is the principle of sequential Bayesian filtering. The predictive density is computed through Chapman-Kolmogorov equation. We compute the posterior density for the current state via Bayes rule:

$$\begin{aligned} p(x_t|y_{1:t}) &= \frac{p(y_{1:t}|x_t)p(x_t)}{p(y_{1:t})} \\ &= \frac{p(y_t, y_{1:t-1}|x_t)p(x_t)}{p(y_t, y_{1:t-1})} \end{aligned}$$

$$
\begin{aligned}
&= \frac{p(y_t|y_{1:t-1}, x_t)p(y_{1:t-1}|x_t)p(x_t)}{p(y_t|y_{1:t-1})p(y_{1:t-1})} \\
&= \frac{p(y_t|y_{1:t-1}, x_t)p(x_t|y_{1:t-1})p(y_{1:t-1})}{p(y_t|y_{1:t-1})p(y_{1:t-1})} \\
&= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \qquad (4.6)
\end{aligned}
$$

As shown in the above formula, the posterior density $p(x_t|y_{1:t})$ is described by three terms: The prior $p(x_t|y_{1:t-1})$ defines the knowledge of the model

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \qquad (4.7)$$

where $p(x_t|x_{t-1})$ is the transition density of the state. The likelihood $p(y_t|x_t)$ determines the measurement equation. The denominator is the normalizing constant

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t. \qquad (4.8)$$

Calculation or approximation of these three terms are essential in the Bayesian filtering and smoothing.

The criterion of optimality used for Bayesian filtering is the Bayes risk of minimum mean-squared error (MMSE). It can be defined in terms of prediction of filtering error

$$E[\| x_t - \hat{x}_t \|^2 |y_{1:t}] = \int \| x_t - \hat{x}_t \|^2 p(x_t|y_{1:t})dx_t \qquad (4.9)$$

where the conditional mean $\hat{x}_t = E[x_t|y_{1:t}]$. Bayesian filtering is optimal in a sense that it explores the posterior distribution which integrates all available information expressed by probabilities. Optimal filters are analytically available in only a few cases. One is the Kalman filter, which computes the marginal density exactly when the functions $f(\cdot)$ and $g(\cdot)$ are linear and both d_t and v_t are Gaussian.

4.2 Particle Filters

The assumptions needed for Kalman filter may not hold in mant practical situations. Many dynamic systems are better assessed as nonlinear/non-Gaussian framework. Therefore, researchers, for example [9] and [15], have devoted to approximating the filtering and smoothing distribution when it is impossible to evaluate them analytically. We focus our attention on the sequential Monte Carlo approach for sequential state estimation. Sequential Monte Carlo technique, also called particle filtering, is a kind of recursive Bayesian filter based on a Monte Carlo simulation known as importance sampling.

The working mechanism of particle filters is as follows: The state space is partitioned into many parts, where the particles are filled according to some probability measure. The higher the probability, the denser the particles are concentrated. The particle system updates along the time via the state equation. The posterior density can be approximated by the empirical distribution of the state variables via random sampling. Since the posterior distribution is unknown or hard to sample from, we would rather choose alternative distribution called proposal distribution for efficient sampling.

To avoid integration in this Bayesian framework, the posterior density is empirically represented by a weighted sum of N_p samples drawn from the posterior distribution

$$p(x_t|y_{1:t}) \approx \frac{1}{N_p} \sum_{t=1}^{N_p} \delta(x_t - x_t^{(i)}) \equiv \hat{p}(x_t|y_{1:t}) \tag{4.10}$$

where $\{x_t^{(i)}, i = 1, ..., N_p\}$ are i.i.d. samples drawn from $p(x_t|y_{1:t})$. When N_p is sufficiently large,$\hat{p}(x_t|y_{1:t})$ approximates the true posterior $p(x_t|y_{1:t})$. Using this

approximating density, we can estimate the mean of a function of states

$$\begin{aligned} E[\phi(x_t)|y_{1:t}] &\approx \int \phi(x_t)\hat{p}(x_t|y_{1:t})dx_t \\ &= \frac{1}{N_p}\sum_{i=1}^{N_p}\int \phi(x_t)\delta(x_t - x_t^{(i)})dx_t \\ &= \frac{1}{N_p}\sum_{i=1}^{N_p}\phi(x_t^{(i)}). \end{aligned} \tag{4.11}$$

Because it is usually impossible to sample directly from the true posterior, it is common to sample from the so-called proposal distribution denoted by $q(x_t|y_{1:t})$. Therefore

$$\begin{aligned} E[\phi(x_t)|y_{1:t}] &= \int \phi(x_t)\frac{p(x_t|y_{1:t})}{q(x_t|y_{1:t})}q(x_t|y_{1:t})dx_t \\ &= \int \phi(x_t)\frac{p(y_{1:t}|x_t)p(x_t)}{p(y_{1:t})q(x_t|y_{1:t})}q(x_t|y_{1:t})dx_t \\ &= \frac{1}{p(y_{1:t})}\int \phi(x_t)W_t(x_t)q(x_t|y_{1:t})dx_t \end{aligned} \tag{4.12}$$

where

$$W_t(x_t) = \frac{p(y_{1:t}|x_t)p(x_t)}{q(x_t|y_{1:t})}. \tag{4.13}$$

Equation (4.12) can be rewritten as

$$\begin{aligned} E[\phi(x_t)|y_{1:t}] &= \frac{\int \phi(x_t)W_t(x_t)q(x_t|y_{1:t})dx_t}{\int p(y_{1:t}|x_t)p(x_t)dx_t} \\ &= \frac{\int \phi(x_t)W_t(x_t)q(x_t|y_{1:t})dx_t}{\int W_t(x_t)q(x_t|y_{1:t})dx_t} \\ &= \frac{E_{q(x_t|y_{1:t})}[W_t(x_t)\phi(x_t)]}{E_{q(x_t|y_{1:t})}[W_t(x_t)]}. \end{aligned} \tag{4.14}$$

By drawing i.i.d. samples $\{x_t^{(i)}\}$ from $q(x_t|y_{1:t})$, we can approximate $E[\phi(x_t)|y_{1:t}]$ by

$$\begin{aligned} E[\phi(x_t)|y_{1:t}] &\approx \frac{\frac{1}{N_p}\sum_{i=1}^{N_p} W_t(x_t^{(i)})\phi(x_t^{(i)})}{\frac{1}{N_p}\sum_{i=1}^{N_p} W_t(x_t^{(i)})} \\ &= \sum_{i=1}^{N_p}\tilde{W}_t(x_t^{(i)})\phi(x_t^{(i)}) \end{aligned} \tag{4.15}$$

where

$$\tilde{W}_t(x_t^{(i)}) = \frac{W_t(x_t^{(i)})}{\sum_{j=1}^{N_p} W_t(x_t^{(i)})}. \tag{4.16}$$

In sequential importance sampling (SIS) filtering, the importance weights $W_t^{(i)}$ can be updated recursively

$$\begin{aligned} W_t^{(i)} &\propto \frac{p(x_t^{(i)}|y_{1:t})}{q(x_t^{(i)}|y_{1:t})} \\ &\propto \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})p(x_{t-1}^{(i)}|y_{1:t-1})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)q(x_{t-1}^{(i)}|y_{1:t-1})} \\ &= W_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)}. \end{aligned} \tag{4.17}$$

Unfortunately, after a few iterations of the algorithm, the majority of the probability mass will be allocated to only a few particles. Hence the distribution of the importance weights become more and more skewed as time goes along. This phenomenon is called weight degeneracy in the literature. In order to solve this problem, [33] introduced a resampling step to the SIS algorithm which replicates the particles with high importance weights and discards the ones with low importance weights. The simplest resampling algorithm is multinomial resampling, which draws new particles from a multinomial distribution with probabilities equal to the normalized importance weights. Improvements have been made in [37] etc. It is not optimal to resample at each iteration as this increases the variation in the estimates. Instead, it should be conducted only when the variance of importance weights grows. [47] introduced a measure for degeneracy called effective sample size N_{eff}

$$\begin{aligned} N_{eff} &= \frac{N_p}{1 + Var_{q(.|y_{1:t})}[\tilde{W}(x_{0:t})]} \\ &= \frac{N_p}{E_{q(.|y_{1:t})}[(\tilde{W}(x_{0:t}))^2]} \leq N_p. \end{aligned} \tag{4.18}$$

In practice, the true N_{eff} is not available. Its estimate is given in

$$\hat{N}_{eff} = \frac{N_p}{\sum_{i=1}^{N_p}(\tilde{W}_t^{(i)})^2}. \tag{4.19}$$

When $\hat{N}_{eff}$ is below a predefined threshold N_T (say $N_p/2$ or $N_p/3$), the resampling procedure is conducted. When $\hat{N}_{eff} < N_T$, then each sample is accepted with probability $min\{1, W_t^{(i)}/N_T\}$ and the rejected samples are restarted and rechecked at all previously violated thresholds. The method is computationally expensive as t increases. The algorithm of SIS particle filter with resampling is summarized as follows:

For time steps $t = 0, 1, 2...$

1. For $i = 1, ..., N_p$, draw samples $x_t^{(i)} \sim q(x_t|x_{t-1}^{(i)}, y_{1:t})$ and set $x_{0:t}^{(i)} = \{x_{0:t-1}^{(i)}, x_t^{(i)}\}$.
2. For $i = 1, ..., N_p$, calculate the importance weights $W_t^{(i)}$ according to 4.17.
3. For $i = 1, ..., N_p$, normalize the importance weights $\tilde{W}_t^{(i)}$ according to 4.16.
4. Calculate $\hat{N}_{eff}$ according to 4.19, return if $\hat{N}_{eff} > N_T$, otherwise generate a new particle set $\{x_t^{(i)}\}$ by resampling with replacement N_p times from the previous set $\{x_{0:t}^{(i)}\}$ with probabilities $P(x_{0:t}^{(j)} = x_{0:t}^{(i)}) = \tilde{W}_{0:t}^{(i)}$. Reset the weights $\tilde{W}_t^{(i)} = \frac{1}{N_p}$.

The most important part of any particle filter is the choice of proposal distribution. The proposal density should be chosen to approximate the target density as closely as possible, and hence will keep the variance of the importance weights low. A convenient choice is the transition density of the state variables which is the prior distribution in a Bayesian framework

$$q(x_t|x_{t-1}^{(i)}, y_{1:t}) = p(x_t|x_{t-1}^{(i)}). \tag{4.20}$$

Substituting (4.20) into (4.17) leads to a simple update of the importance weights

$$W_t^{(i)} \propto W_{t-1}^{(i)} p(y_t|x_t^{(i)}). \tag{4.21}$$

If the particles are resampled at every iteration, the importance weight update simplifies further to $W_t^{(i)} \propto p(y_t|x_t^{(i)})$. This is the original particle filtering algorithm of [33] known as sampling importance resampling (SIR) filter or bootstrap filter.

Let us compare SIS and SIR filters. Both of the two filtering techniques use importance sampling. The difference is that in SIR filter, the resampling is always performed; whereas in SIS filter, importance weights are calculated sequentially, resampling is only taken when needed. The choice of proposal distributions in SIS and SIR plays a vital role in their final performance. Last, in both procedures, resampling is suggested to be done after filtering because resampling brings extra randomness to the current samples.

Although the transition density of the state variable is convenient, it does not include the current observation in the proposal density. Theoretically, it was shown ([58]) that the choice of proposal distribution $q(x_t|x_{t-1}^{(i)}, y_{1:t}) = p(x_t|x_{t-1}^{(i)}, y_t)$ minimizes the variance of importance weights $W_t^{(i)}$ conditioning on $x_{t-1}^{(i)}$ and $y_{1:t}$. By this argument, the importance weights can be recursively updated as $W_t^{(i)} \propto W_{t-1}^{(i)} p(y_t|x_{t-1}^{(i)})$. However, this optimal proposal distribution can only be calculated analytically in special cases, and the evaluation of the integral $p(y_t|x_{t-1}^{(i)}) = \int p(y_t|x_t)p(x_t|x_{t-1}^{(i)})dx_t$ can be tedious. There is no universal choice for proposal distribution, and choosing appropriate proposal density is usually problem dependent.

4.3 Auxiliary Particle Filter

Finding a proposal density that utilizes the current observation y_t in a computationally efficient manner can be very challenging. On the other hand, a potential weakness of generic particle filters is the approximation of filtered density is not sufficient to characterize the tail behavior of the true density. This effect is more significant when outliers exist. [53] and [54] introduced so-called auxiliary particle filter (APF). The main idea is to augment the existing particles with high importance weights in a sense that the predictive likelihood $p(y_t|x_{0:t-1}^{(i)})$ are large for these good particles. The APF varies from SIR in the way that it reverses the order of sampling and resampling, which is possible when the importance weights are dependent on x_t.

By inserting the likelihood inside the empirical distribution, we can rewrite the filtered density as

$$\begin{aligned} p(x_t|y_{1:t}) &\propto p(y_t|x_t) \int p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1} \\ &\propto \sum_{i=1}^{N_p} W_{t-1}^{(i)} p(y_t|x_t) p(x_t|x_{t-1}^{(i)}) \end{aligned} \tag{4.22}$$

where $p(x_{t-1}|y_{1:t-1}) = \sum_{i=1}^{N_p} W_{t-1}^{(i)} \delta(x_{t-1} - x_{t-1}^{(i)})$. By introducing an auxiliary variable $\xi \in \{1, ..., N_p\}$ that serves as index of the mixture component, the augmented joint density $p(x_t, \xi|y_{1:t})$ is updated as

$$\begin{aligned} p(x_t, \xi = i|y_{1:t}) &\propto p(y_t|x_t) p(x_t, \xi = i|y_{1:t-1}) \\ &= p(y_t|x_t) p(x_t|\xi = i, y_{1:t-1}) p(\xi = i|y_{1:t-1}) \\ &= p(y_t|x_t) p(x_t|x_{t-1}^{(i)}) W_{t-1}^{(i)}. \end{aligned} \tag{4.23}$$

A sample can be drawn from joint density (4.23) by neglecting the index ξ, by which a set of particles $\{x_t^{(i)}\}_{i=1}^{N_p}$ are drawn from the marginal density $p(x_t|y_{1:t})$ and

the index ξ is simulated with probabilities proportional to $p(\xi|y_{1:t})$. Therefore, (4.22) can be approximated by

$$p(x_t|y_{1:t}) \propto \sum_{i=1}^{N_p} W_{t-1}^{(i)} p(y_t|x_t^{(i)}, \xi^i) p(x_t|x_{t-1}^{(i)}) \tag{4.24}$$

where ξ^i denotes the index of the particle $x_t^{(i)}$ at time step $t-1$, namely $\xi^i = \{\xi = i\}$. The proposal distribution used to draw $\{x_t^{(i)}, \xi^i\}$ is chosen as a factorized form

$$q(x_t, \xi|y_{1:t}) \propto q(\xi|y_{1:t}) q(x_t|\xi, y_{1:t}) \tag{4.25}$$

where

$$q(\xi|y_{1:t}) \propto p(y_t|\mu_t^{(i)}) W_{t-1}^{(i)} \tag{4.26}$$

$$q(x_t|\xi, y_{1:t}) = p(x_t|x_{t-1}^{(i)}) \tag{4.27}$$

where $\mu^{(i)}$ is a value (e.g. mean or mode) associated with $p(x_t|x_{t-1}^{(i)})$. Thus the true posterior is further approximated by

$$p(x_t|y_{1:t}) \propto \sum_{i=1}^{N_p} W_{t-1}^{(i)} p(y_t|\mu_t^{\xi=i}) p(x_t|x_{t-1}^{\xi=i}). \tag{4.28}$$

From (4.26) and (4.27), the important weights are recursively calculated as

$$\begin{aligned} W_t^{(i)} &= W_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)}) p(x_t^{(i)}|x_{t-1}^{(\xi=i)})}{q(x_t^{(i)}, \xi^i|y_{1:t})} \\ &\propto \frac{p(y_t|x_t^{(i)})}{p(y_t|\mu_t^{(\xi=i)})}. \end{aligned} \tag{4.29}$$

The APF is basically a two-stage procedure: At the first stage, simulate particles with large predictive likelihoods; at the second stage, reweigh the particles and draw the augmented states. The auxiliary variable can be applied for SIS or SIR filters. An auxiliary SIR filtering algorithm is summarized below.

For time step $t = 1, 2, ...$:

1. For $i = 1, ..., N_p$, calculate $\mu_t^{(i)}$ (e.g. $\mu_t^{(i)} = E[p(x_t|x_{t-1}^{(i)})]$).
2. For $i = 1, ..., N_p$, calculate the first-stage weight $W_t^{(i)} = W_t^{(i-1)} p(y_t|\mu_t^{(i)})$ and normalize weights $\tilde{W}_t^{(i)} = \frac{W_t^{(i)}}{\sum_{j=1}^{N_p} W_t^{(j)}}$
3. Use the resampling procedure in SIR filter to obtain new $\{x_t^{(i)}, \xi^i\}_{i=1}^{N_p}$.
4. For $i = 1, ..., N_p$, sample $x_t^{(i)} \sim p(x_t^{(i)}|x_{t-1}^{(i)}, \xi^i)$, and update the second-stage weights $W_t^{(i)}$ according to (4.29).

In general, the second stage weights are much less variable than for the original SIR method. By making proposal density with high conditional likelihoods we reduce the sampling cost from particles with low likelihoods and so will not be resampled at the second stage. This improves the statistical efficiency of the sampling procedure. In generic particle filters, estimation is usually performed after the resampling step, which is less efficient because resampling introduces extra random variation. APF essentially overcomes this problem by conducting one-step ahead estimation based on the estimate $\mu_t^{(i)}$ that characterizes $p(x_t|x_{t-1}^{(i)})$.

4.4 Particle Filter for Aggregate Default Intensity

We apply the particle filtering technique to infer the posterior mean of frailty path $E[\eta Y_t|\mathcal{G}_t]$. The state variable is $x_t = (u_t, v_t, Y_t)$ where u_t and v_t are latent volatilities while Y_t is frailty factor. The observation vector is $y_t = (S_t, r_t, C_t, p_t, T_t)$. Since the SIR particle filter usually gives poor estimation results, we adopt auxiliary particle filtering method to calculate posterior paths. The proposal distribution we use is the transition density of the state variable: $q(x_t|x_{t-1}^{(i)}, y_{1:t}) = p(x_t|x_{t-1}^{(i)})$ and set $\mu_t^{(i)} = E[p(x_t|x_{t-1}^{(i)})]$.

Note that all the components of state vector x_t satisfy OU processes, and hence their transition densities are Gaussian. In particular, we have

$$p(u_t|u_{t-1}^{(i)}) \sim N(u_{t-1}^{(i)}e^{-b_1\Delta t} + b_0(1-e^{-b_1\Delta t}), \frac{\sigma_1^2}{2b_1}(1-e^{-2b_1\Delta t})) \tag{4.30}$$

$$p(v_t|v_{t-1}^{(i)}) \sim N(v_{t-1}^{(i)}e^{-\kappa_1\Delta t} + \kappa_0(1-e^{-\kappa_1\Delta t}), \frac{\sigma_2^2}{2\kappa_1}(1-e^{-2\kappa_1\Delta t})) \tag{4.31}$$

$$p(Y_t|Y_{t-1}^{(i)}) \sim N(Y_{t-1}^{(i)}e^{-\kappa_2\Delta t}, \frac{1}{2\kappa_2}(1-e^{-2\kappa_2\Delta t})) \tag{4.32}$$

where Δt is the time interval between time step t-1 and t. In our simulation, we take $\Delta t = \frac{1}{365}$. Next, we will calculate the likelihood $p(y_t|x_t^{(i)})$ in the importance weight updating procedure:

$$\begin{aligned} p(y_t|x_t^{(i)}) &= p(S_t, r_t, C_t, p_t, T_k|x_t^{(i)}) \\ &= p(T_k|S_t, r_t, C_t, p_t, x_t^{(i)})p(S_t, r_t, C_t, p_t|x_t^{(i)}) \\ &= p(T_k|S_t, r_t, C_t, p_t, Y_t^{(i)})p(S_t|x_t^{(i)})p(r_t|x_t^{(i)})p(C_t|x_t^{(i)})p(p_t|x_t^{(i)}) \\ &= p(T_k|S_t, r_t, C_t, p_t, x_t^{(i)})p(S_t|u_t^{(i)})p(r_t|v_t^{(i)})p(C_t)p(p_t) \end{aligned} \tag{4.33}$$

where the fourth equality holds because C_t and p_t are not latent. Let us calculate the three terms in 4.33 respectively. Given the volatility u_t, the stock return S_t satisfies OU process and hence its conditional density is Gaussian:

$$p(S_t|u_t^{(i)}) \sim N(S_0e^{-a_1t} + a_0(1-e^{-a_1t}), \frac{e^{2u_t^{(i)}}}{2a_1}(1-e^{-2a_1t})). \tag{4.34}$$

Conditioning on the volatility v_t, the 3-month T-bill rate is a CIR process. The transition density of the CIR process is known to be a noncentral chi-squared distribution. However, due to the complicated structure of this distribution, we do not use it for practical purpose. Alternatively, we approximate the density with a normal distribution after discretizing the SDE with an Euler scheme. The conditional density is evaluated as

$$p(r_t|v_t^{(i)}) \sim N(e^{-\phi_1\Delta t}r_{t-1} + \phi_0(1-e^{-\phi_1\Delta t}), e^{2v_t^{(i)}}r_{t-1}\Delta t). \tag{4.35}$$

The distributions of slope and annual IP growth rate are Gaussian since both satisfy OU processes:

$$\begin{aligned} p(C_t) &\sim N(C_0 e^{-\theta_1 t} + \mu_1(1 - e^{-\theta_2 t}), \frac{\sigma_3^2}{2\theta_1}(1 - e^{-2\theta_1 t})) \\ p(p_t) &\sim N(p_0 e^{-\theta_2 t} + \mu_2(1 - e^{-\theta_2 t}), \frac{\sigma_4^2}{2\theta_2}(1 - e^{-2\theta_2 t})). \end{aligned}$$

The conditional distribution for the default timing T is given as:

$$p(T_k > t | T_{k-1} = s, X, Y^{(i)}) = \exp(-\int_s^t (e^{\alpha + \beta \cdot X_t + \eta Y_u^{(i)}} + \int_0^u \delta e^{-\gamma} dN_v) du) \tag{4.36}$$

The density for this hazard rate model does not have an analytic form. The macro vector X_t and frailty Y_t are generated from stochastic equation system.

4.5 Bayesian Smoothing

Particle filtering techniques can be easily extended to the smoothing problem, where the future observations can be used to estimate the current state. In the Bayesian filtering framework, the goal is to estimate the posterior density $p(x_t|y_{1:t+\tau})$. There are three kinds of smoothing: fixed-point smoothing, fixed-lag smoothing and fixed-interval smoothing. Fixed-point smoothing is to obtain smoothed estimate of state x_t at a fixed point t for all $\tau \geq 1$. Fixed-lag smoothing is concerned with smoothing of data where there is a fixed delay τ. Fixed-interval smoothing deals with smoothing of a finite set of data. More specifically, the purpose is to achieve $p(x_t|y_{1:T})$ for fixed T and all t in the interval $[1, T]$.

In our default hazard problem, we want to compute the $\mathcal{F}_T$-conditional posterior distribution of the frailty process Y, where T is the final date of our sample. This is the conditional distribution of the latent state given all the available default and macroeconomic data through the end of the period.

In the fixed-interval smoothing, we first run a particle filter to obtain $p(x_t|y_{1:t})$ for $1 \leq t \leq T$. Secondly in the backward step, the smoothing process is recursively updated by

$$\begin{aligned} p(x_{t:T}|y_{1:T}) &= p(x_{t+1:T}|y_{1:T})p(x_t|x_{t+1:T}, y_{1:T}) \\ &= p(x_{t+1,T}|y_{1:T})p(x_t|x_{t+1}, y_{1:t}) \\ &= p(x_{t+1,T}|y_{1:T})\frac{p(x_{t+1}|x_t, y_{1:t})p(x_t|y_{1:t})}{p(x_{t+1}|y_{1:t})} \end{aligned} \tag{4.37}$$

where the second equality results from first-order Markov assumption. In above derivation, $p(x_{t:T}|y_{1:T})$ denotes current smoothed estimate, $p(x_{t+1:T}|y_{1:T})$ denotes future smoothed estimate, and $p(x_t|y_{1:t})$ is the current filtered estimate.

We have the following distribution

$$\hat{p}(x_{1:T}|y_{1:T}) = \sum_{i=1}^{N_p} \tilde{W}_T^{(i)} \delta(x_{1:T} - x_{1:T}^{(i)}) \tag{4.38}$$

where $\{\tilde{W}_T^{(i)}\}_{i=1}^{N_p}$ are the importance weights in time step T. By marginalization, we obtain the approximated fixed-interval smoothing distribution for any $1 \leq t \leq T$

$$\hat{p}(x_t|y_{1:T}) \approx \sum_{i=1}^{N_p} \tilde{W}_T^{(i)} \delta(x_t - x_t^{(i)}). \tag{4.39}$$

In practice, this method is usually infeasible because of the weight degeneracy problem. At final time step T, the state trajectories $\{x_{1:T}^{(i)}\}_{i=1}^{N_p}$ have been resampled many times, and hence there are only a few distinct trajectories at

time $t < T$. [16] proposed a new fixed-interval smoothing algorithm as follows. Their algorithm is based on the formula initiated by [40]

$$p(x_t|y_{1:T}) = p(x_t|y_{1:t}) \int \frac{p(x_{t+1}|y_{1:T})p(x_{t+1}|x_t)}{p(x_{t+1}|y_{1:t})} dx_{t+1}. \tag{4.40}$$

We use an alternative approximation of the smoothing distribution:

$$\hat{p}(x_t|y_{1:T}) = \sum_{i=1}^{N_p} \tilde{W}_{t|T}^{(i)} \delta(x_t - x_t^{(i)}). \tag{4.41}$$

The smoothing distribution has the same support $\{x_t^{(i)}\}_{i=1}^{N_p}$ as the filtering distribution, but the importance weights are different. The algorithm with new weights $\{\tilde{W}_{t|T}^{(i)}\}_{i=1}^{N_p}$ is as follows:

1. Initialization at time $t = T$

   For $i = 1, ..., N_p$, $\tilde{W}_{T|T}^{(i)} = \tilde{W}_T^{(i)}$;

2. For $k = T - 1, ..., 1$:

   For $i = 1, ..., N_p$, evaluate the importance weights

$$\tilde{W}_{t|T}^{(i)} = \sum_{j=1}^{N_p} \tilde{W}_{t+1|T}^{(j)} \frac{\tilde{W}_t^{(i)} p(x_{t+1}^{(j)}|x_t^{(i)})}{[\sum_{l=1}^{N_p} \tilde{W}_t^{(l)} p(x_{t+1}^{(j)}|x_t^{(l)})]}. \tag{4.42}$$

The algorithm is based on the following argument. Use approximation (4.41), we have

$$\int \frac{p(x_{t+1}|y_{1:T})p(x_{t+1}|x_t)}{p(x_{t+1}|y_{1:t})} dx_{t+1} \approx \sum_{i=1}^{N_p} \tilde{W}_{t+1|T}^{(i)} \frac{p(x_{t+1}^{(i)}|x_t)}{p(x_{t+1}^{(i)}|y_{1:t})} \tag{4.43}$$

$p(x_{t+1}^{(i)}|y_{1:t})$ can be approximated by

$$\begin{aligned} p(x_{t+1}^{(i)}|y_{1:t}) &= \int p(x_{t+1}^{(i)}|x_t)p(x_t|y_{0:t})dx_t \\ &\approx \sum_{j=1}^{N_p} \tilde{W}_t^{(j)} p(x_{t+1}^{(i)}|x_t^{(j)}). \end{aligned} \tag{4.44}$$

By substituting the above formula to (4.40), an approximation of $p(x_t|y_{1:T})$ is thus given

$$\begin{aligned}
\hat{p}(x_t|y_{1:T}) &= [\sum_{i=1}^{N_p} \tilde{W}_t^{(i)} \delta(x_t - x_t^{(i)})] \sum_{j=1}^{N} \tilde{W}_{t+1|T}^{(j)} \frac{p(x_{t+1}^{(j)}|x_t)}{[\sum_{l=1}^{N_p} \tilde{W}_t^{(l)} p(x_{t+1}^{(j)}|x_t^{(l)})]} \\
&= \sum_{i=1}^{N_p} [\sum_{j=1}^{N_p} \tilde{W}_{t+1|T}^{(j)} \frac{\tilde{W}_t^{(i)} p(x_{t+1}^{(j)}|x_t^{(i)})}{[\sum_{l=1}^{N_p} \tilde{W}_t^{(l)} p(x_{t+1}^{(j)}|x_t^{(l)})]}] \delta(x_t - x_t^{(i)}) \\
&= \sum_{i=1}^{N_p} \tilde{W}_{t|T}^{(i)} \delta(x_t - x_t^{(i)}). \qquad (4.45)
\end{aligned}$$

This algorithm requires storage of the marginal distributions $\hat{p}(x_t|y_{1:t})$ including weights and particles for all $t \in \{1, ..., T\}$. The memory requirement is $O(TN_p)$ and its complexity is $O(TN_p^2)$. This method is typically computationally intensive.

CHAPTER 5

STOCHASTIC POINT PROCESS SIMULATION

Point processes with stochastic intensity has wide application in credit risk. In this project, we need to simulate the default timing according to our aggregate intensity in order to make out-of-sample forecast and to assess portfolio losses.

5.1 Related Simulation Method

If the intensity is a deterministic function of time t, then N is a non-homogeneous Poisson process and we can generate the interarrival times by the inverse method from the inter-arrival time distribution, the order statistics property of the Poisson process, or the thinning scheme of [46]. In more general case, the intensity is state-dependent and is governed by random risk factors that follows stochastic processes. Some point processes are self-exciting, in which case the intensity depends on the past events. Under this circumstance, the time-scaling method is widely accepted in stochastic point process simulation. It is based on a result of [52], which implies that under mild conditions, any counting process can be transformed into a standard Poisson process by a change of time that is given by the counting process compensator, or cumulative intensity. Thus, the event times can be generated by re-scaling Poisson arrivals with the compensator.

Mathematically, if the default counting process has intensity λ, the compensator is $A_t = \int_0^t \lambda_s ds$ and it is continuous. Suppose the compensator A increases to ∞ almost surely. For $s > 0$, define $A_s^{-1} = \inf\{t : A_t > s\}$. The process A_s^{-1} is con-

tinuous and strictly increasing. The time change change theorem of [52] implies that process $N_{A^{-1}}$ is a standard Poisson process on $[0, \infty)$ in the right continuous, time-changed filtration $\mathcal{G}$ generated by the stopping time sigma-fields $\mathcal{F}_{A_s^{-1}}$ for $s \geq 0$. Let T_k denote the k-th default time, the random variables

$$S_k = A_{T_k} = \int_0^{T_k} \lambda_s ds \tag{5.1}$$

are the arrival times of a standard $\mathcal{G}$-Poisson process in $[0, \infty)$. Conversely, the k-th default time is the hitting time of A to the random variable S_k almost surely,

$$T_k = A_{S_k}^{-1} = \inf\{t : \int_0^t \lambda_s ds > S_k\}. \tag{5.2}$$

To simulate N by re-scaling of Poisson arrivals we need to generate a trajectory of the intensity λ. The continuous-time path of the compensator must be approximated on a discrete grid. However, the approximation of a continuous-time process by a discrete-time process introduces bias into the simulation estimator. The size of the bias is often unknown, and a very fine time discretization may be required to reduce the bias to an acceptable level. Even more computational effort may be required to verify that the bias is sufficiently small. All these issues make the time-change method computationally expensive.

This drawback leads to some exact simulation method that eliminates the need to discretize the compensator. [32] project the point process onto its own filtration, and then sample it in this coarser filtration. The projected intensity, which is the conditional expectation of the intensity in the reference filtration, is deterministic between event times. This property facilitates exact sampling of the point process by sequential thinning or the inverse transform scheme. This method is also related to the scheme developed by [5] for a jump-diffusion with state-dependent intensity that is almost surely bounded from above. The

boundedness property allows [5] to generate the jump times by state-dependent thinning.

5.2 Default Event Simulation

We follow the argument in [23] and [32] to simulate default counting process. We propose to project the default intensity λ onto the sub-filtration generated by the point process itself. We do not require exact simulation as in [32] since the stochastic system driving macro variables are complex and hence it is impossible to obtain an analytic solution for intensity projection.

We adopt discretization to approximate stochastic differential equations for macro covariates. The Milstein scheme as in [41] is used for the approximation. The observation window is one week, or namely $\Delta t = \frac{1}{52}$. The simulated macro covariates are held constant between two observation points. After conditioning on the sub-filtration, the projected intensity evolves deterministically between events and jumps at default times. Due to the special exponential decay structure for the intensity, the intensity is decreasing between two events if these two defaults happen within one week. The post-event intensity can serve as a dominating process before the next event in this case. This leads to the thinning procedure in simulation. If there is no default within one week, then the intensity jumps after one week because macro variables are updated every week. In this case, the counting process can be considered as an inhomogenous Poisson process.

The algorithm to simulate default events is listed below:

1. Simulate macro covariates on a weekly basis.
2. Simulate the first event time T_1 according to inhomogenous Poisson process, where the intensity updates every week, set $k = 1$.

   For k=1,2,... until the time horizon:
3. Draw $Z_k \sim Exp(\lambda_{T_k})$.
4. Set $V_k = T_k + Z_1 + ... + Z_k$.
5. Draw $U \sim U(0, 1)$, if $U \leq \frac{\lambda_{V_k}}{\lambda_{T_k}}$ and $Z_k \leq \Delta t$

   then set $T_{k+1} = V_k$.

   Else increase k by 1, go to step 3 and update the intensity.

This algorithm re-defines the dominating Poisson process after each rejection of a candidate time or weekly time grid. The acceptance probability is increased by the updated bound for λ.

CHAPTER 6

MAJOR EMPIRICAL RESULTS

This section shows the fitted model for the dynamic system via EMM. We also compare various models that contain different variables. In-sample tests and out-of-sample forecasts are performed respectively.

The macroeconomic vector under consideration is $X = (S_k, r_k, C_k, p_k)_{k=1}^{459}$ where each component consists of monthly data and satisfies the stochastic differential equations defined in 2.6-2.9. Recall that $\{T'_k\}_{k=1}^{n}$ are the revised default dates in the portfolio. Define interarrival times: $A_k = T'_k - T'_{k-1}, k = 1, ..., n$ and $T'_0 = 0$

The probability distribution of A_k is given

$$\begin{aligned} P(A_k > t | T'_{k-1} = s, Y) &= \exp(-\int_s^{s+t} (\exp(\alpha + \beta \cdot X_u + \eta Y_u) + \int_0^u \delta e^{-\gamma} dN_v) du) \\ dY_t &= -\kappa Y_t dt + dW_t^{(7)}. \end{aligned} \tag{6.1}$$

In our hazard rate model, the two latent volatility functions u and v and the frailty factor Y are not directly observed. Therefore, the likelihood function becomes intractable, which makes the maximum likelihood estimation impractical. This problem can be easily solved by applying efficient method of moments.

6.1 The Fitted Model

Let us look at the empirical results for macroeconomic factors. The preferred SNP model for the 4-dim macro vector is an AR(1)-GARCH(1,1)-Kz(4)-Iz(1), or in other words $(L_u, L_g, L_r, L_p, K_z, I_z, K_y, I_y) = (1, 1, 1, 1, 4, 1, 0, 0)$. The EMM estimation results for macroeconomic variables are reported in Table 6.1.

Table 6.1: EMM estimates for monthly trailing 1-year S&P 500 stock return, 3-month T-bill rate, slope and annual IP growth rate.

	Coefficient	Standard Error
a_1	0.2803	0.0426
b_1	0.8625	0.1337
ρ_1	-0.2128	0.0748
ϕ_1	0.3146	0.1268
κ_1	0.5629	0.0856
σ_2	1.2265	0.1274
θ_1	0.0878	0.0163
ρ_2	-0.6324	0.2016
θ_2	0.6405	0.1590
σ_4	1.1106	0.3176

Strong mean-reverting behavior was found for all macroeconomic variables. For the S&P 500 stock return, the leverage effect is well interpreted by the negative correlation between return and volatility. Both models for S&P 500 stock return and 3-month T-bill rate have no difficulty in capturing the volatility clustering in the actual data. We also find out that slope is negative correlated with short term interest rate. This is common because the long term interest rate is relatively stable. When the short term interest rate is raised, it increases more compared to long term rate, and hence the slope, which is just the difference, is reduced.

Denote $\pi = (\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \delta, \gamma, \eta, \kappa)$ and hence $p_\pi = 9$. Table 6.2 shows the estimated coefficient values for macro covariates as well as parameters for contagion and frailty. The parameters for both contagion and frailty are highly

Table 6.2: EMM estimates for coefficients of full default intensity model.

	Coefficient	Standard Error
constant α	2.1733	0.3842
SP 500 stock return β_1	0.8645	0.2072
3-month T-bill rate β_2	-0.1823	0.0470
slope β_3	0.3583	0.1170
annual IP growth rate β_4	-0.5275	0.2083
contagion loss size δ	1.6231	0.4472
contagion decay speed γ	2.3374	0.3825
frailty volatility η	0.3194	0.1274
frailty mean reversion κ	0.2286	0.0737

statistically significant.

The SNP model for the interarrival times A is $(L_u, L_g, L_r, L_p, K_z, I_z, K_y, I_y) = (1, 1, 3, 1, 6, 0, 0, 0)$ and the corresponding $p_\theta = 14$. The degree of freedom is $p_\theta - p_\pi - 1 = 4$. The chi-square value defined in equation 3.22 is 5.7961 and the corresponding p-value in the goodness-of-fit test is 0.2149. The p-value is much larger than significance level $\alpha = 0.05$, which suggests the model is well fit. Figure 6.1 illustrates the fitted portfolio default intensity measured in events per year. We can observe that the estimated default hazard rate can capture the dramatic fluctuation of annual default numbers.

Let us interpret the economic intuition behind this model. The positive coefficient of S&P 500 stock return is unexpected sign since people believe that when the broad stock market is strong, it would dramatically reduce the default hazard rate. However, it could be the case that the default risk increases when

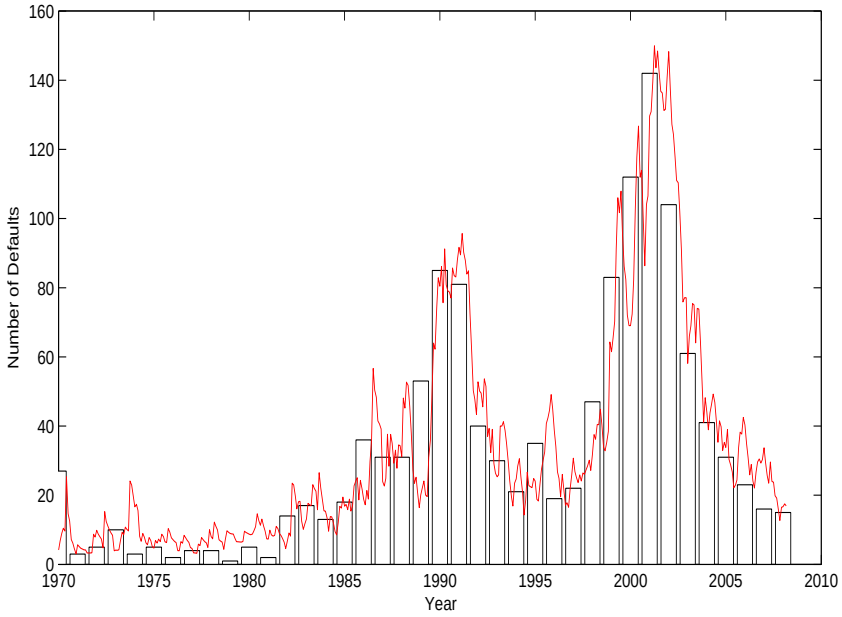

Figure 6.1: The estimated portfolio default intensity vs actual default number each year between 1970 and 2007.

the bubble bursts after boom years in stock market. Similar finding is also presented in [18]. The negative coefficient for 3-month T-bill rate implies that high interest rates decrease default risk. Central banks often raise interest rates to curb inflation when the economy is growing too fast and hence the hazard rate should be lower. The positivity of slope coefficient indicates that the default intensity is high when people are demanding more premium for holding long term instruments. The negative coefficient for annual IP growth rate is quite straightforward since the default hazard rate is lower when the whole economy is expanding. The parameter for the jump size is moderate for a single default event. Nevertheless, if many firms go bankrupt in a short period of time, the contagion has a great impact on the conditional hazard rate of surviving firms. This is well-known global ripple effect in the credit market. Default by one

firm tends to weaken the conditions of others in the same sector. The annual volatility for unobserved frailty is $\hat{\eta} = 31.94\%$, which is highly economically significant.

Table 6.3: EMM estimates for coefficients of aggregate intensity model without frailty.

	Coefficient	Standard Error
constant α	2.2945	0.3842
SP 500 stock return β_1	0.9031	0.2375
3-month T-bill rate β_2	-0.1546	0.0404
slope β_3	0.3078	0.1284
annual IP growth rate β_4	-0.6173	0.1735
contagion loss size δ	1.5791	0.3465
contagion decay speed γ	2.2569	0.4825

Table 6.4: EMM estimates for coefficients of aggregate intensity model without contagion.

	Coefficient	Standard Error
constant α	2.8276	0.3614
SP 500 stock return β_1	1.0358	0.1290
3-month T-bill rate β_2	-0.1713	0.0536
slope β_3	0.3685	0.1094
annual IP growth rate β_4	-0.5732	0.2046
frailty volatility η	0.3805	0.1427
frailty mean reversion κ	0.2639	0.0952

We are interested in the relative importance of each component in our intensity model. We assume here that magnitude of proportional hazard rate part is

twice as large as that of Hawkes-type contagion. A positive shock to stock return covariate by one standard deviation increases the default intensity roughly by $\frac{\exp(\alpha+\beta X_t+\eta_t Y_t)}{\lambda_t}(e^{\beta_1}-1) \approx \frac{2}{3}(e^{0.8645}-1) \approx 91.59\%$. When 3-month T-bill rate suffers a negative shock by one standard deviation, the default intensity is increased by about $\frac{\exp(\alpha+\beta X_t+\eta_t Y_t)}{\lambda_t}(e^{-\beta_2}-1) \approx \frac{2}{3}(e^{0.1823}-1) \approx 13.33\%$. Similar calculations yield that the one standard deviation impact to slope and IP growth rate will affect default intensity by 28.73 % and 46.31 % respectively. This implies that $S\&P$ 500 stock return is of greater importance among the macro factors. A positive jump in frailty by one standard deviation amplifies the default intensity roughly by $\frac{\exp(\alpha+\beta X_t+\eta_t Y_t)}{\lambda_t}(e^{\frac{\eta}{2\kappa}}-1) \approx \frac{2}{3}(e^{0.3194/(2*0.2286)}-1) \approx 67.40\%$. The influence of direct ripple effect can be analyzed under two different credit conditions. Under mild contagion environment when only one default occurs and also assume the default intensity is around 30, the default intensity is increased roughly by $\frac{\delta}{\lambda_t} \approx \frac{1.6231}{30} = 5.41\%$. On the other hand, if 10 defaults occur in a short period and assume the default intensity is around 80, the default intensity ramps up by about $\frac{10\delta}{\lambda} \approx \frac{10*1.6231}{80} = 20.23\%$.

Table 6.5: Goodness-of-fit test for default hazard models

	Full Model	Without Frailty	Without Contagion	Without Both
Degree of freedom	4	4	5	5
χ^2 statistics	5.7961	8.1947	10.0619	12.1453
p-value	0.2149	0.0847	0.0735	0.0328

Let us compare the full benchmark model to the alternative models after removing the frailty and contagion factor respectively. Table 6.3 and Table 6.4 report the estimated coefficients in the default intensity models without frailty

and contagion respectively. The signs and magnitudes of the macroeconomic covariate coefficients are similar to those in the full model. Note that the constant coefficient α in the model excluding contagion is greater than its counterpart in the full model. All these parameters remain statistically significant. Table 6.5 compares the goodness-of-fit test statistics for all these models. The models excluding frailty or contagion shows adequate fit, although their p-values are significantly lower. These two models have similar p-values which indicates that frailty and contagion play roughly equally important roles in our data period. For contrast, we also list the test result for model containing only macroeconomic covariates. The p-value 0.0328 is not significant at $\alpha = 0.05$ level. This implies that model without contagion and frailty can not depict properly the evolution of default intensity over the years, and hence is rejected.

6.2 The Posterior Distribution of Frailty

One interesting quantity is the mean function of posterior path of the frailty given the observable, i.e. $E[\eta Y_t|\mathcal{G}_t]$. This posterior distribution is important to interpret the frailty model and apply it to the portfolio loss calculation. We apply particle filtering algorithms covered in the previous section. The results demonstrate the effectiveness of particle filter for continuous time stochastic processes. The importance of recovering latent frailty is crucial for forecasting portfolio default intensity and risk management.

We compute the $\mathcal{G}_t$-conditional posterior distribution of the frailty process Y. This is the posterior distribution of frailty Y given contemporaneously available information of historical defaults and macro covariate data. Figure 6.2 shows

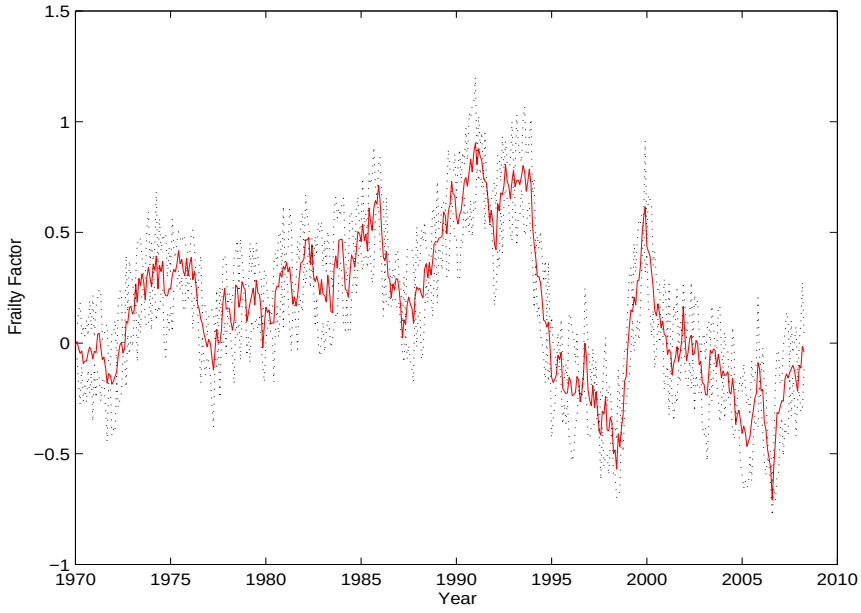

Figure 6.2: Posterior mean $E(\eta Y_t|\mathcal{G}_t)$ of the scaled frailty factor given contemporaneously available information $\mathcal{G}_t$

the posterior mean of the frailty factor accompanied by one standard deviation bands. The path is estimated by averaging 4000 samples of Y from auxiliary particle filter. We observe substantial fluctuation of frailty effect over time. The frailty factor can be positive or negative. In particular, the latent frailty level rises rapidly when there is negative shock in the economy. Global events such as 1990 gulf war and 2000 internet bubble drive the frailty up in a short time period. We also observe that the frailty path takes large negative values during years 2003-2006 before credit crunch. This indicates that model without frailty overestimate default probabilities given macroeconomic conditions at that time. However, the frailty variable moves quickly towards its mean value 0 in the second half of 2007 because the severe changes in macro factors can explain high default risk in global credit crisis.

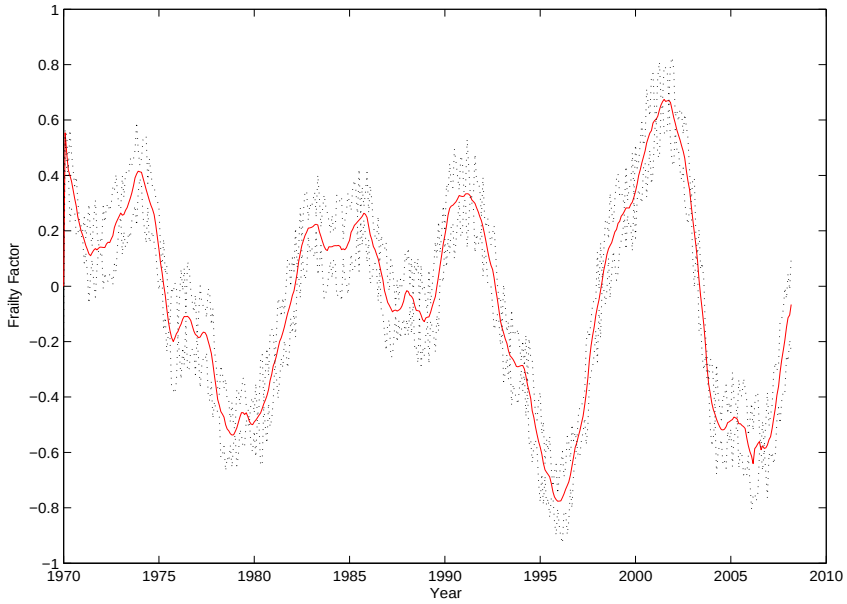

Figure 6.3: Posterior mean $E(\eta Y_t|\mathcal{G}_T)$ of the scaled frailty factor given all available information $\mathcal{G}_T$

While most credit risk managers need posterior distribution of Y_t given current information $\mathcal{G}_t$, we are also interested in obtaining frailty path given all information $\mathcal{G}_T$ at the final time T. We compute the posterior distribution of Y given all available information through the end of sample period by using particle smoothing algorithm. Figure 6.3 illustrates the $\mathcal{G}_T$-conditional posterior mean of the frailty process $E[\eta Y_t|\mathcal{G}_T]$ with one standard deviation bands. Similar fluctuation pattern is observed. The conditional mean is less choppy compared to that conditioning on contemporaneously information. This may be explained by the fact that the algorithm uses all available information and hence can smooth volatility of frailty path.

6.3 In-sample Test and Out-of-sample Prediction

We estimate the aggregate hazard rate model from a subset of the observation period and the goodness-of-fit tests are executed in each sub-period. All the observation windows start in January 1970 but end at different horizons in increment of one year. Table 6.6 shows the estimation results and in-sample test p-values. The p-values of all these sub-models exceed the 5% level, which indicates good fits overall.

Table 6.6: EMM estimation results for the benchmark model in different observation periods starting 01/70 and ending at various dates before 01/07. The p-values are from the goodness-of-fit tests in minimum chi-squared estimation.

End	01/01	01/02	01/03	01/04	01/05	01/06	01/07
α	2.0923	2.1785	2.1086	2.1662	2.1372	2.1758	2.1806
β_1	0.8403	0.9041	0.8738	0.8674	0.8961	0.8529	1.0945
β_2	-0.1702	-0.1692	-0.1538	-0.1784	-0.1683	-0.1812	-0.1840
β_3	0.3385	0.3729	0.3540	0.3638	0.3762	0.3629	0.3577
β_4	-0.5312	-0.5196	-0.5280	-0.5114	-0.5248	-0.5083	-0.5237
δ	1.6245	1.6863	1.5428	1.6046	1.6161	1.6229	1.6273
γ	2.2836	2.3195	2.2561	2.2762	2.3267	2.3638	2.3146
η	0.3762	0.3618	0.3894	0.3485	0.3547	0.3782	0.3612
κ	0.2573	0.2437	0.2802	0.2957	0.2782	0.2628	0.2739
p-value	0.1992	0.1628	0.1734	0.1835	0.2109	0.1936	0.2068

We use these fitted models to perform out-of-sample forecasts. Four different models are taken into account: (i) our benchmark full model, (ii) model without frailty, (iii) model without contagion, and (iv) model excluding both frailty and

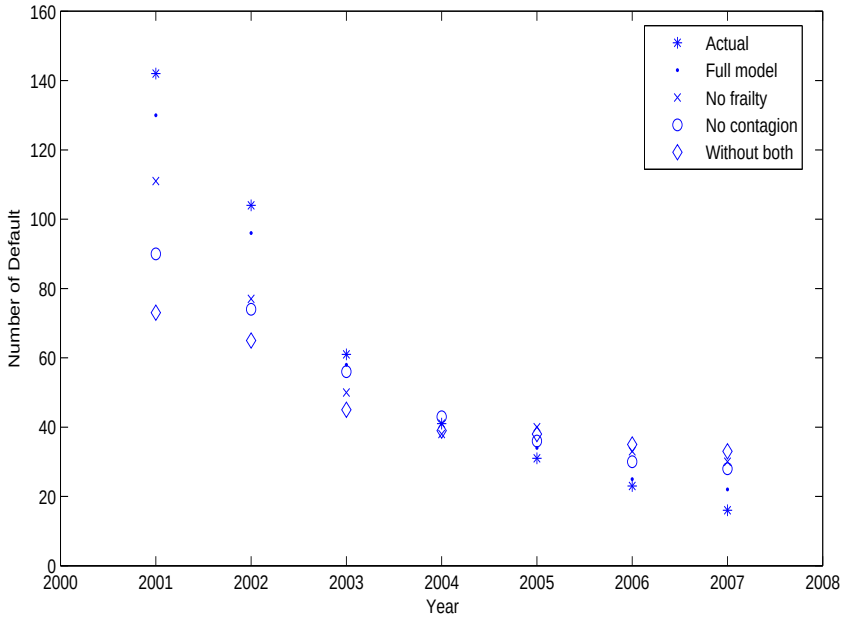

Figure 6.4: Predicted number of defaults from different models

contagion. In particular, we simulate credit event arrivals on one-year forecast window. For example, the model estimated from observation horizon 1970-2001 is used to predict the total number of defaults in 2002. The estimated model parameters are fixed in the subsequent year and the macroeconomic variables are generated from the stochastic differential equation system. Posterior frailty path is generated through particle filter conditioning on contemporaneous information. The arrival dates of defaults are generated from the hazard rate model. In each forecasting period, 10,000 paths were simulated. Figure 6.4 shows the predicted number of default events from different model specifications and realized default number in each year.

Some important observations are that all these models underestimate the number of credit events during dot-com bubbles. The full model that includes

macro factors as well as frailty and contagion stands out as the best fit. The model excluding contagion considerably underestimates the default number in year 2001 and 2002 because the ripple effect is dominant during these recession years. Models without frailty does not capture the tail loss distribution very well. In other words, these models underestimate the probabilities of unusually low portfolio losses and of unusually high portfolio losses. The model without frailty and the model without contagion can hardly surpass each other in the forecast competition. The model containing neither frailty or contagion is the worst fit and fails to capture the default clustering phenomenon.

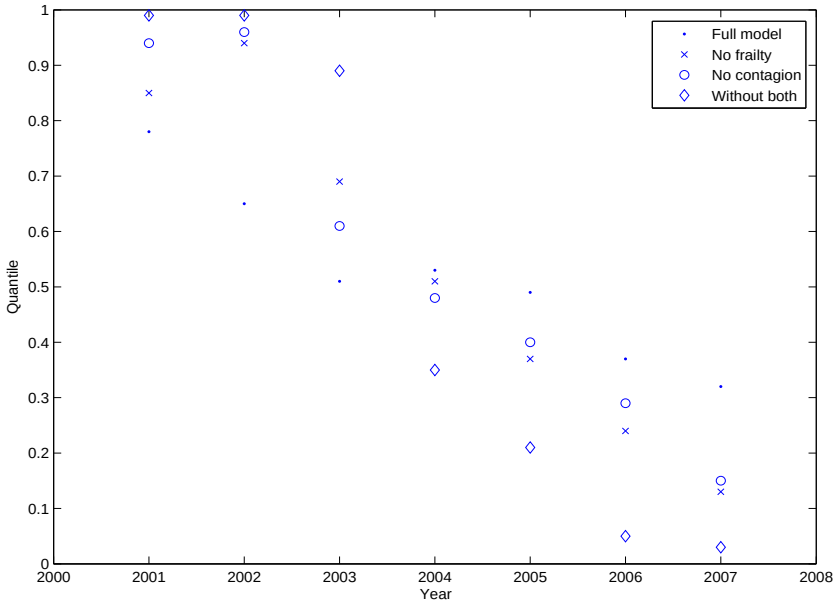

Figure 6.5: Quantiles of the realized number of defaults with respect to the predicted portfolio loss distribution as implied by different models

Let us focus on the distribution of portfolio losses from the above simulation. We calculate the quantile of the realized number of defaults with respect to these distributions. Figure 6.5 shows these quantiles in the same plot. The

quantiles for the full model are distributed relatively evenly in the interval [0.3, 0.8]. This indicates that the full model provides an accurate assessment of portfolio credit risk. On the other hand, the quantiles of the model without both variables cluster around 0 and 1, which suggests this model is a poor fit. It significantly underestimate the probabilities of extreme high and low losses. The quantiles for the rest two models lie in between and are slightly less extreme.

In order to further illustrate the role of frailty effect on the tail distribution of portfolio loss, we perform out-of-sample forecast based on the full benchmark model and model without frailty. The observation period is 01/70 to 01/02 and the prediction using fitted model is run in subsequent 6 years.

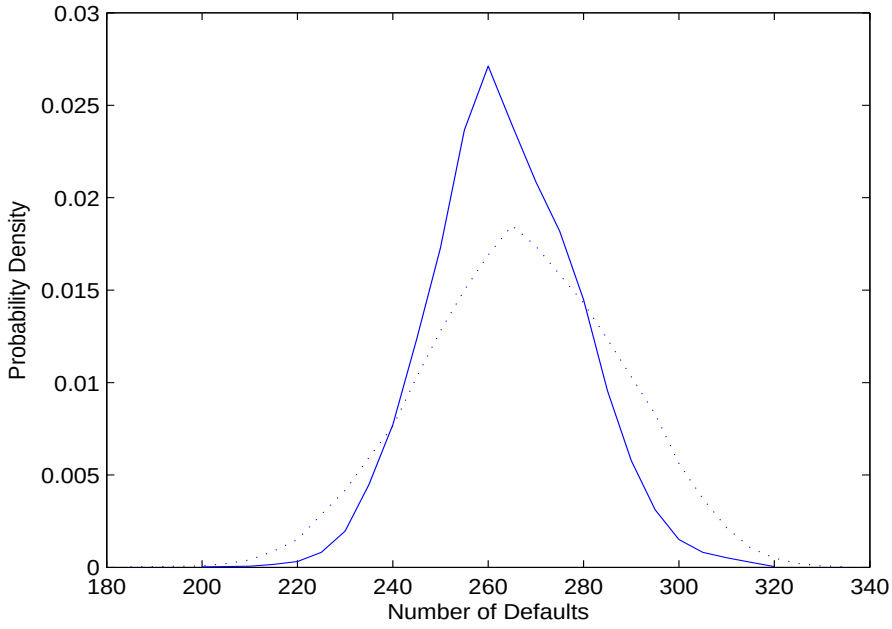

Figure 6.6: The probability density of the total number of defaults between 01/2002 and 12/2007. The solid line denotes the forecast made by the fitted full frailty model. The dotted line represents the prediction by the model without frailty.

Figure 6.6 shows the probability density plot for forecast default events be-

tween 2002 and 2007. The density plot for model with frailty is more heavy-tailed in both ends. The actual total number of defaults is 276 during the forecast horizon. This realized number of defaults roughly corresponds to 77% quantile of the distribution implied by the full frailty model, and it is around 90% quantile of the distribution implied by the model without frailty. Thus, the prediction made by hazard rate model without frailty significantly underestimates the probability of extreme events. The full frailty model has a 95% quantile and 99% quantile of 301 and 322 respectively, which exceeds the realized default number 276. This indicates this model can capture severe default clustering in the past few years. We conclude that the frailty effect is of great importance in analyzing the tail distribution of portfolio loss.

CHAPTER 7

APPLICATION IN CREDIT PORTFOLIO

7.1 Portfolio Intensity

We now apply the top-down framework to model the credit portfolios and bank loans. One thing that restricts the application of our approach is the large size of the hypothetical portfolio. In industry, only credit card loans or auto loans can have as many as 5000 individual constitutes that is of the similar size in our economic wide portfolio. However, most CDS/CDO portfolios have about 100 constitutes. In order to utilize our method in these credit portfolios, we have to narrow down the big portfolio to its small subset.

As usual, let λ_t represent the intensity for the original hypothetical portfolio. Let h_t denote intensity of a specific credit portfolio. The individual names in the small pool are within the economic wide portfolio. Suppose ρ is a company category index set where the category set can be various industry sectors or different credit ratings. Let Q_t^ρ denote the number of ρ-category firms in the big portfolio surviving up to time t. Let M_t^ρ denote the number of ρ-category firms in the small credit portfolio surviving up to time t. If Q_t is the total number of firms that have survived up to time t, then $Q_t = \sum_\rho Q_t^\rho$.

Recall that $T_1 < T_2 < ...$ are ordered default dates in the real economy. We model the intensity for ρ-category firms in the big portfolio as follows:

$$\lambda_t^\rho = \lambda_t \cdot \frac{Q_{t-}^\rho + \xi 1_{\{default \in \rho \text{ at time t}\}}}{Q_{t-} + \xi 1_{\{t=T_k \text{ for some k}\}}} \tag{7.1}$$

where $\xi > 0$ controls the impact of a new default at time t. Typically we choose $\xi \geq 1$. If there is no default at time t, the intensity simplifies to $\lambda_t^\rho = \lambda_t \frac{Q_t^\rho}{Q_t}$. Obvi-

ously, we have $\sum_\rho \lambda_t^\rho = \lambda_t$. The intensity for ρ-category firms in the small pool can be specified as:

$$h_t^\rho = \lambda_t^\rho \cdot \frac{M_t^\rho}{Q_t^\rho}. \tag{7.2}$$

In this case, the intensity is proportional to the number of surviving firms in the portfolios. Then, we can easily model the intensity for the small credit portfolio by summing up over all ρ categories:

$$h_t = \sum_\rho h_t^\rho = \lambda_t \sum_\rho \frac{Q_{t-}^\rho + \xi 1_{\{default \in \rho \text{ at time t}\}}}{Q_{t-} + \xi 1_{\{t=T_k \text{ for some k}\}}} \cdot \frac{M_t^\rho}{Q_t^\rho} \tag{7.3}$$

So far our argument is based on actual probability measure. In the next section, we will switch to risk-neutral measure and apply it in pricing portfolio credit derivatives. [19] indicates that there is little empirical evidence to distinguish the actual and risk-neutral dynamics of default intensities.

7.2 Portfolio Credit Derivatives

A portfolio credit derivative is considered as a contingent claim on the portfolio loss, which bears the correlated default risk in a reference portfolio such as loans and bonds. Index swaps are most popular portfolio credit derivatives. They are bilateral contracts, in which one party provides default protection on the reference portfolio, and the other party pays a premium for the protection. They can be considered as a form of insurance.

We adopt the valuation approach of [24] for index swaps. We assume that there are no arbitrage opportunities or market frictions. We fix a risk-neutral probability measure P^* with respect to a constant risk-free interest rate $r > 0$. The basic valuation problem is to determine the fair premium for the contract.

Suppose the index swap is based on a portfolio with n constituent names. These single-name credit swaps have common notional that normalizes to 1, common maturity date T and common quarterly swap premium payment dates $\{t_m\}$. Let $L_t = \sum_{k=1}^{n} l_k$ be a credit portfolio loss process, where l_k is the loss for the k-th individual name. The protection seller agrees to cover losses once there is a default in the reference portfolio. The protection buyer agrees to make a periodic payment at dates $\{t_m\}$. The value of default leg D_t is calculated through the discounted cumulative losses. By Fubini Theorem and integration by parts, we obtain

$$\begin{aligned} D_t &= E_t^*[\int_t^T e^{-r(s-t)} dL_s] \\ &= e^{-r(T-t)} E_t^*[L_T] - L_t + r \int_t^T e^{-r(s-t)} E_t^*[L_s] ds \end{aligned} \tag{7.4}$$

where r is risk-free interest rate and E_t^* denotes the risk-neutral expectation given available information up to time t . The cash flow that protection buyer pays at t_m is a fraction I of the portfolio notional that have survived until time t_m. The value of premium payment is given by

$$P_t(I) = I \sum_m e^{-r(t_m - t)} (n - E_t^*[\tilde{N}_{t_m}]) \tag{7.5}$$

where $\tilde{N}_t$ is default counter in the credit portfolio. Here we neglect premium accruals and day count revision. The fair index swap spread at time t is calculated by equating default leg and premium leg. In other words, we solve for I from equation $D_t = P_t(I)$. The spread only depends on expected defaults in the time horizon $(t, T]$.

We value index swap premium under the risk-neutral measure P^*. The default counting process and loss process are specified in terms of a risk-neutral intensity h^*. The risk neutral intensity h^* is the counterpart to the actual inten-

sity h in the last section. Let risk-neutral portfolio intensity h_t^* satisfy affine jump diffusion framework as in [2]:

$$dh_t^* = \kappa^*(c^* - h_t^*)dt + \sigma^* \sqrt{h_t^*} dW_t + \delta^* dL_t \tag{7.6}$$

where κ^*, c^*, σ^* and δ^* are risk-neutral parameters with condition $2\kappa^* c^* \geq (\sigma^*)^2$. For convenience, we assume the loss at each default is constant. In industry, the recovery value is widely accepted as 40%, and in other words the loss at default is 60%. Hence, the loss process is defined as:

$$L_t = \sum_{k=1}^{n} l_k = \sum_{k=1}^{n} 0.6 \cdot 1_{\{T_k \leq t\}}. \tag{7.7}$$

7.3 Estimation Results

We derive the risk neutral default intensity from market credit default swap spread. We use CDX North America 5-year generic series as index swap data. We have found similar results as that in [1].

We calibrate our model to two real index swap series. One is CDX North America Investment Grade 5Y market index and the other is CDX North America High Yield 5Y market index. 5-year maturity CDS contracts are the most liquid index swaps in the market. The CDX Investment Grade and High yield series consist of 125 and 100 equally weighted names respectively. Both series are taken from Bloomberg. The investment grade indices start on 10/21/2003 and end at 3/31/2008. The high yield series range from 11/4/2005 to 3/31/2008. Both indices are rolled over every 6 months and what we have are on the run series. We apply the calibration procedure proposed in [24]. We fit the data to our model on a weekly basis. More specifically, we use every Wednesday close data

during our sample period. If Wednesday data is not available, we use average number of Tuesday and Thursday data. The parameter vector $\theta = (\kappa^*, c^*, \sigma^*, \delta^*)$ and risk-free rate is $r = 5\%$. We want to numerically solve the nonlinear optimization problem

$$\min_{\theta} \sum_{m} (Market_m - Model_m(\theta))^2 \tag{7.8}$$

where $\{Market_m\}$ are market quotes and $\{Model_m(\theta)\}$ are index spread series derived from our model. We restrict the parameter space in a finite subspace and perform a grid search in that region. For example, we set the parameter space of CDX IG series $\Theta = [0, 5] \times (0, 10] \times [0, 5] \times [0, 5] \times (0, 20]$. The optimization is initialized at values drawn uniformly from the parameter space Θ, and is repeated for each of 100 independent draws. The optimal parameter θ^* is the minimum value of the optimization problem obtained by gradient-based method among all 100 runs. The estimation results are listed in Table 7.1.

Table 7.1: Grid search estimates of the risk-neutral parameters

Index	κ^*	c^*	σ^*	δ^*
IG	0.48	0.91	0.83	0.24
HY	1.73	1.56	1.85	1.32

We can calculate the risk-neutral intensity h^* for CDX portfolio from the above estimation results. In order to get the actual intensity, we take advantage of the full default history of economic wide portfolio. The intensity h for CDX portfolio is calculated using equation 7.3 where we set $\xi = 2$. Here ρ denotes Moody's credit rating category. Since defaults in investment grade portfolio are very rare, we treat Aaa,Aa,A and Baa as one category, represented by IG. Therefore $\rho = \{IG, Ba, B, Caa, Ca, C\}$.

We measure the default risk premium by the ratio of risk-neutral portfolio intensity h^* to actual intensity h. Figure 7.1 illustrates the fitted ratios $\frac{h^*}{h}$ for the 5-year IG and HY CDX portfolios.

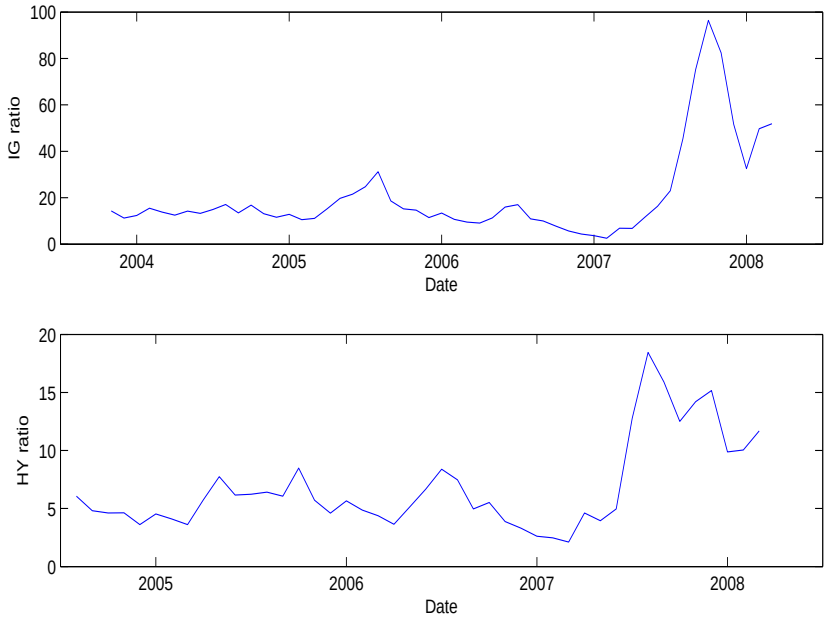

Figure 7.1: Ratio of risk-neutral portfolio intensity h^* to actual intensity h for 5 year CDX contracts. Upper panel: CDX NA Investment Grade portfolio; Lower panel: CDX NA High Yield portfolio

The risk-neutral intensity indicates the market price for protection against instantaneous defaults in the reference portfolio given current information. The actual intensity measures the expected number of instantaneous defaults in the portfolio given default history and macroeconomic factors. From the graph, we find that both ratios are significantly greater than 1 because investors demand substantial compensation for bearing correlated default risk. The ratio of high yield is smaller than that of investment grade for the reason that there is no default in CDX IG portfolio during sample period. We also find out the intensity ratios vary dramatically over time. The ratio increased quickly in May 2005

when credit ratings for General Motors and Ford are downgraded. It gradually declined to bottom in February 2007. It surged quickly to peak in August 2007 when subprime crisis hit the U.S. credit market.

CHAPTER 8

CONCLUDING REMARKS

8.1 Conclusion

We establish a top-down default intensity model for analyzing multi-period U.S. corporate defaults during 1970-2008. We carefully examine macro, contagion and frailty effects in explaining default clustering phenomenon. U.S. corporate default rates change over time well beyond levels of model that only includes observed macro covariates. A default event has a significant impact on the conditional default rate of surviving firms. The default clustering is either due to direct ripple effect or information update. These two sources, contagion and frailty, are equally important in capturing default clustering.

We use stochastic volatility models and Ornstein-Uhlenbeck processes to describe stochastic macroeconomic covariates. A Hawkes specification is utilized for modeling contagion effect. Since the volatility and frailty factors are not directly observed, the usual maximum likelihood estimation becomes intractable. We propose to apply efficient method of moments in parameter estimation. This approach is particularly efficient in nonlinear system with latent variables. The goodness-of-fit tests are naturally embedded in the estimation procedure. We use the chi-square statistics to measure model adequacy in various specifications. We find out that a model without frailty underestimates the probability of extreme events for large portfolio loss.

Although our model is based on a large hypothetical portfolio, we can apply proportional approach to model intensity in a much smaller subset. From the

index swap series, it is indicated that people demand much more risk premium for bearing correlated default risk under risk-neutral measure. These useful findings can be used by banks and asset managers of credit portfolios. They can estimate the capital needed to allocate in order to withstand default losses at high confidence level and perform dynamic risk management.

8.2 Future Research

In this dissertation, the frailty path is assumed to follow OU process. Although this parsimonious model facilitates the filtering step of posterior distribution, it may fail to capture the richness of frailty path over the years. More sophisticated specification could be introduced. For instance, more Brownian motion factors could be included in the model to provide additional shocks to the economy.

The top down approach considers all the firms in one single portfolio. On the other hand, we can also break down these firms into different sectors. Besides common macroeconomic covariates, we can add idiosyncratic variables that are present in a particular sector. Furthermore, the contagion factor should be more prevalent at a sector level than at an economy-wide level. As for frailty, we can have both common Brownian motion and sector-wise independent Brownian motions in the stochastic model.

It is also interesting to investigate frailty parameters from a Bayesian perspective. Since frailty is latent, uncertainty about the frailty mean reversion and volatility parameters could lead to an increase in the tail risk of portfolio losses. We can specify the prior distribution of these two parameters, and afterwards, the marginal posterior distributions of mean reversion and volatility can be in-

ferred. The influence of parameter uncertainty on portfolio losses can be studied accordingly.

BIBLIOGRAPHY

[1] S. Azizpour and K. Giesecke. Premia for correlated default risk. *Working paper, Stanford University*, 2008.

[2] S. Azizpour and K. Giesecke. Self-exciting corporate defaults: Contagion vs frailty. *Working Paper, Stanford University*, 2008.

[3] F. Black and J. C. Cox. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 31:351–367, 1976.

[4] D. Brigo, A. Pallavicini, and R. Torresetti. Calibration of cdo tranches with the dynamical generalized-poisson loss model. *Working paper,Banca IMI*, 2006.

[5] B. Casella and G. Roberts. Exact simulation of jump-diffusion processes with monte carlo applications. *Methodology and Computing in Applied Probability, forthcoming*, 2010.

[6] Z. Chen. Bayesian filtering: From kalman filters to particle filters, and beyond. *Manuscript, McMaster University*, 2003.

[7] P. Collin-Dufresne, R. Goldstein, and J. Helwege. Are jumps in corporate bond yields priced? modeling contagion via the updating of beliefs. *Working paper, Carnegie Mellon University*, 2003.

[8] M. Coppejans and A. Gallant. Cross validated snp density estimates. *Journal of Econometrics*, 110(1):27–65, 2002.

[9] D. Crisan. *"Particle filters - A theoretical perspective," in Sequential Monte Carlo Methods in Practice, ed. by A. Doucet, N. de Freitas, and N. Gordon*. Springer Verlag, New York, 2001.

[10] P. J. Crosbie and J. R. Bohn. Modeling default risk. *Technical Report, KMV, LLC*, 2002.

[11] D. J. Daley and D. Vere-Jones. A summary of the theory of point processes. *Paper presented to the stochastic point process conference, New York*, 1971.

[12] S. Das, D. Duffie, N. Kapadia, and L. Saita. Common failings: How corporate defaults are correlated. *Journal of Finance*, 62(1):93–117, 2007.

[13] M. Davis and V. Lo. Infectious defaults. *Quantitative Finance*, 1:382–387, 2001.

[14] C. Dellacherie, J. Fermanian, and M. Sbai. Estimation of a reduced-form credit portfolio model and extensions to dynamic frailties. *Working paper, BNP Paribas*, 2006.

[15] P. DelMoral and G. Salut. Particle interpretation of non-linear filtering and optimization. *Mathematical Physics*, 5(3):355–372, 1997.

[16] A. Doucet, N. de Freitas, K. Murphy, and S. Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. *UAI Proceedings*, pages 176–183, 2000.

[17] D. Duffie. Credit risk modeling with affine processes. *Journal of Banking and Finance*, 29:2751–2802, 2005.

[18] D. Duffie, A. Eckner, G. Horel, and L. Saita. Frailty correlated default. *Journal of Finance*, 64(5):2089–2123, 2009.

[19] D. Duffie and N. Gârleanu. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57(1):41–59, 2001.

[20] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376, 2000.

[21] D. Duffie, L. Saita, and K. Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665, 2007.

[22] D. Duffie and K.J. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12(4):687–720, 1999.

[23] D. Duffie and K.J. Singleton. Simulating correlated defaults. *Working paper, Graduate Sschool of Business, Stanford University*, 1999.

[24] E. Errais, K. Giesecke, and L. Goldberg. Pricing credit from the top down with affine point processes. *Working Paper, Stanford University*, 2006.

[25] S. Figlewski, H. Frydman, and W. Liang. Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *Working Paper, NYU Stern School of Business*, 2008.

[26] R. Frey and W. Runggaldier. Credit risk and incomplete information: a nonlinear-filtering approach. *Working paper, University of Leipzig*, 2007.

[27] A. Gallant. *Nonlinear Statistical Models*. Wiley, New York, 1987.

[28] A. Gallant and J. Long. Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, 84:125–141, 1997.

[29] A. Gallant and G. Tauchen. Which moments to match? *Econometric Theory*, 12:657–681, 1996.

[30] K. Giesecke. *"Portfolio Credit Risk: Top down vs. bottom up approaches", in Frontiers in Quantitative Finance: Credit Risk and Volatility Modeling Edited by R. Cont*. Wiley, 2008.

[31] K. Giesecke and L. Goldberg. A top down approach to multi-name credit. *Workin paper, Stanford University*, 2005.

[32] K. Giesecke, H. Kakavand, and M. Mousavi. Simulating point processes by intensity projection. *Proceedings of the 2008 Winter Simulation Conference*, 2008.

[33] N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings-F*, 140:107–113, 1993.

[34] D. Hamilton. Moodys senior ratings algorithm and estimated senior ratings. *Moodys Investors Service*, 2005.

[35] A. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Journal of Financial Economics*, 58(1):83–90, 1971.

[36] A. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[37] P. Clifford J. Carpenter and P. Fearnhead. Improved particle filter for nonlinear problems. *IEE Procedings-F*, 146(1):2–7, 1999.

[38] R. Jarrow and S. Turnbull. Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50(1):53–85, 1995.

[39] R. Jarrow and F. Yu. Counterparty risk and the pricing of defaultable securities. *Journal of Finance*, 56(5):1765–1799, 2001.

[40] G. Kitagawa. Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82:503–514, 1987.

[41] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Number 23 in Applications of Mathematics. Springer, New York, 1992.

[42] S. J. Koopman, A. Lucas, and A. Monteiro. The multi-stage latent factor intensity model for credit rating transitions. *Journal of Econometrics*, 142(1):399–424, 2008.

[43] S. J. Koopman, A. Lucas, and B. Schwaab. Forecasting cross-sections of frailty correlated default. *Tinbergen Institute Discussion Paper Series*, 029(04), 2008.

[44] A. Kwiecinski and R.Szekli. Some monotonicty and dependent propertyies of self-exciting point processes. *The Annals of Applied Probabilites*, 6(4):1211–1231, 1996.

[45] D. Lando and M. Nielsen. Correlation in coporate defaults: Contagion or conditional independence. *Working Paper,Copenhagen Business School*, 2008.

[46] P. Lewis and G. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Logistics Quarterly*, 26:403–413, 1979.

[47] J.S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of American Statistical Association*, 93:1032–1044, 1998.

[48] F. Longstaff and A. Rajan. An empirical analysis of collateralized debt obligations. *Journal of Finance*, 63(2):529–563, 2008.

[49] C.L. Mallows. Some comments on cp. *Technometrics*, 15:661–675, 1973.

[50] R.C. Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.

[51] P. Meyer. *Probability and Potentials*. Blaisdell, London, 1966.

[52] P. Meyer. Demonstration simplifee dun theoreme de knight, in seminaire de probabilites v. *Lecture Notes in Mathematics 191', Springer-Verlag Berlin*, pages 191–195, 1971.

[53] M. Pitt and N. Shephard. Filtering via simulation: Auxillary particle filter. *Journal of the American Statistical Association*, 94:590–599, 1999.

[54] M. Pitt and N. Shephard. *"Auxiliary variable based particle filters," in Sequential Monte Carlo Methods in Practice, ed. by A. Doucet, N. de Freitas, and N. Gordon*. Springer Verlag, New York, 2001.

[55] Schönbucher. Frailty models, contagion and information effects. *Working Paper, ETH Zürich*, 2005.

[56] T. Shumway. Forecasting bankruptcy more efficiently: A simple hazard model. *Journal of Business*, 74:101–124, 2001.

[57] M. Vassalou and Y. Xing. Default risk in equity returns. *Journal of Finance*, 59:831–868, 2004.

[58] V. S. Zaritskii, V. B. Svetnik, and L. I. Shimelevich. Monte carlo technique in problems of optimal data processing. *Automation and remote control*, 12:95–103, 1975.