

# **Dual-Frailty Default Intensity Model: An Application to Taiwan's Public Listed Firms**

**Gen-Yu Lin**

**Department of Finance  
National Chung Hsing University**

**Mei-Yuan Chen**

**Department of Finance  
National Chung Hsing University**

Aug. 10, 2015

---

Correspondence: Mei-Yuan Chen, Department of Finance, 250 Kuo Kuang  
Rd. Taichung 402, Taiwan, Tel: +886-4-22853323, Fax: +886-4-22856015,  
E-MAIL: mei\_yuan@dragon.nchu.edu.tw

# ABSTRACT

In this paper, a new default intensity model with dual frailties is proposed and an empirical evidence is provided to prove the model's adequacy. In addition to the conventional explanatory variables, Duffie, et al. (2009) introduces a time frailty and Chava et al. (2011) considers an industrial frailty to model the conditional default dependence. However, from an empirical study on Taiwan's public listed firms, Lin and Chen (2015) points out that the time frailty can only catch the default dependence along the time but not among the industries. On the other hand, the industrial frailty can only capture the default dependence from the industrial correlation but not from the time dependence. Hence, the default intensity model with time and industrial frailties is recommended. This paper introduces a dual frailty model to incorporate time and industrial frailties. Using the empirical data of Taiwan's public listed firms from January of 1995 to January of 2015, the dual frailty model is estimated and the Fisher's dispersion tests are conducted for bins from time and from industries. Our empirical findings indicate that the dual frailty model not only captures the time dependence of default clustering but also the industrial correlations. Therefore, our dual frailty model outperforms the intensity model with time frailty of Duffie, et al. (2009) and the industrial frailty of Chava, et al. (2011).

Keywords: Default clustering, default intensity model, time frailty, industrial frailty, dual-frailty model

JEL classification: C15, C33, G20

# 1 Introduction

Based on the Cox proportional hazard model of Cox (1972), default intensity model is derived by Lane, Looney, and Wansley (1986) and is widely applied to study the default probability of firms in finance. In financial literature, the considered conventional explanatory variables include Altman variables in Altman (1968), Campbell variables in Campbell et al. (2008), Duffie variables in Duffie et al. (2007), and Shumway variables in Shumway (2001). To establish the joint likelihood function and do the maximum likelihood estimation of a default intensity model, the defaults conditional on explanatory variables, called conditional defaults, are assumed to be independent. However, using Fisher's dispersion test and Chi-square test, Duffie et al. (2009) rejects the conditional independence of defaults and then recommends a time frailty specified with an OU stochastic process in the default model. The time frailty indeed is to model the default clustering from time dependence not captured by the macroeconomic variables considered in the intensity default model. Huang and Zhang (2011, in Chinese) applies an intensity default model with a time frailty to compare which sets of considered variables (Altman, Campbell, Duffie, and Shumway variables) have better model fitting with the empirical data from Taiwan public listed firms. Meanwhile, Chava et al. (2011) introduces an industrial frailty into the default intensity model to consider the default dependence caused by industrial momentum. Huang et al. (2012, in Chinese) applies the intensity default model with industrial frailty to compare model fittings with different sets of explanatory variables for Taiwan public listed firms.

However, from an empirical study on Taiwan's public listed firms, Lin and Chen (2015, in Chinese) shows that the default intensity model with a time frailty can only catch the default correlation from time but not from the

industrial momentum. On the other side, the default intensity model with an industrial frailty can only catch the default clustering caused by industrial momentum but not from time. The necessity of a default intensity model with both time frailty and industrial frailty is therefore raised by Lin and Chen (2015).

The purpose of this paper is to introduce a default intensity model with time frailty and industrial frailty. This model is called dual-frailty default intensity model. The remaining of this paper is organized as follows. Section 2 is to introduce the dual-frailty default intensity model and its estimation. An empirical study is conducted in section 3. Our empirical data consists of Taiwan's public listed firms from January, 1995 to January, 2015. Some conclusions and suggestions are provided in last section.

## 2 Dual-Frailty Default Intensity Model

The proportional hazard model of Cox (1972) assumes the hazard function as:

$$h_i(t) = h_0(t) \exp(\mathbf{X}'\boldsymbol{\beta}),$$

where  $h_i(t)$  indicates the hazard rate of individual firm  $i$  at time  $t$ ,  $\mathbf{X}$  denotes the explanatory variables, and  $h_0(t)$  is called the baseline hazard. Given assumption of the baseline hazard rate is constant,  $h_0(t) = \lambda$ , the hazard function can be rewritten as

$$h_i(t) = \lambda \exp(\mathbf{X}'\boldsymbol{\beta}) = \exp[\log(\lambda) + \mathbf{X}'\boldsymbol{\beta}] = \exp(\alpha + \mathbf{X}'\boldsymbol{\beta}) = \lambda_{it}.$$

By assuming the default occurrences following a Poisson distribution with parameter  $\lambda_{it}$ , the hazard function becomes the foundation of a default intensity model. The default intensity model suggested by Duffie et al. (2007)

assumes the defaults following a Poisson distribution with a known parameter  $\lambda$ . That is,  $D_{it}|\lambda \sim i.i.d.Poisson(\lambda \Delta t)$ , where  $D_{it}$  denotes the dummy variable for the  $i$  firm at time  $t$  with  $D_{it} = 1$  for the firm is default and  $D_{it} = 0$  for not default,  $\lambda$  is the parameter of the Poisson distribution, and  $\Delta t$  is the time interval. Thus, the default probability function is

$$Prob(D_{it} = x) = \frac{(\lambda \Delta t)^x e^{-\lambda \Delta t}}{x!}.$$

It is worth to note that if the parameter  $\lambda$  is constant, the default is a usual Poisson process. However,  $\lambda_{it} = \Lambda(\mathbf{X}_{it}; \boldsymbol{\beta}) = \exp(\mathbf{X}_{it}'\boldsymbol{\beta})$  is assumed in Duffie et al. (2007). Since the explanatory variables  $\mathbf{X}_{it}$  are stochastic and then  $\lambda_{it}$  is also stochastic, the default becomes a doubly stochastic processor Cox process (Cox, 1955). Lefebvre (2005) points that if  $\lambda_{it}$  is non-stochastic, a doubly stochastic process reduces to a usual Poisson process and then the default can be assumed to be independent. Given  $\lambda_{it} = \exp(\mathbf{X}_{it}'\boldsymbol{\beta})$  is correct specified and  $\mathbf{X}_{it}$  are observed,  $\lambda_{it}$  becomes non-stochastic. Thus,  $D_{it}|\mathbf{X}_{it} \sim i.i.d.Poisson(\lambda_{it} \Delta t)$  is valid. Since the independence holds only under conditions, it is called *conditional independence*. Given the conditional independence, the joint likelihood function of  $D_{it}|\mathbf{X}_{it}$  for all  $i$  and  $t$  in sample can be constructed and then the MLE (maximum likelihood estimator) are obtainable. However, Das et al. (2007) provides evidences that the conditional independence is rejected by Fisher dispersion test given all explanatory variables  $\mathbf{X}_{it}$  considered in existing literature. Instead of incorporating other new explanatory variables with the default intensity model, Duffie et al. (2009) and Chava et al. (2011) consider time frailty and industrial frailty to model to capture the default correlation in the intensity model.

## 2.1 Default Intensity Model with a Time Frailty

The default intensity model with a time frailty considered in Duffie et al. (2009) is:

$$\lambda_{it} = \exp(\mathbf{X}_{it}'\boldsymbol{\beta} + \eta Y_t), i = 1, 2, \dots, m, t = t_i, t_i + 1, \dots, T_i,$$

where  $m$  denotes the number of firms considered in the sample,  $t_i$  and  $T_i$  are the beginning and ending time of the  $i$ th firm. And,  $\lambda_{it}$  represents the default intensity of the  $i$ th firm at time  $t$ .  $Y_t$  is the time frailty which follows an Ornstein-Uhlenbeck (OU) stochastic process:  $dY_t = -\kappa Y_t dt + dB_t$ , where  $B$  is a standard Brownian motion and  $\kappa$  represents the mean-reversion rate of this process.  $\mathbf{X}_{it}$  denotes the explanatory variables observed for the  $i$ th firm at time  $t$  which consist of micro- and macro-economic variables. To do the default intensity forecasts,  $\mathbf{X}_{it}$  is assumed to follow a first-order AR process:  $\mathbf{X}_t = \mathbf{X}_{t-1}'\boldsymbol{\gamma} + u_t$ .

Denote a vector  $\mathbf{D}_i = (D_{it_i}, D_{i(t_i+1)}, \dots, D_{iT_i})$  to represent the status of default in the survival time period. To have  $D_{it}$  only with value 1 or 0, the time interval  $\Delta t$  is set to be small enough. This implies a firm commits a default only once within a time interval. To be simplified, all defaults of all firms in the sample are denoted as  $\mathbf{D} = (\mathbf{D}_1, \dots, \mathbf{D}_m)$ . Also denote  $\mathbf{X}_i = (\mathbf{X}_{it_i}, \dots, \mathbf{X}_{iT_i})$  as the sample values of explanatory variables for the  $i$ th firm. And then, all sample values of explanatory variables of all firms are represented as  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_m)$ . The vector of time frailty is denoted as  $\mathbf{Y} = (Y_0, Y_1, \dots, Y_T)$ , where  $T$  is the ending time of the sample period under study. Define the joint likelihood function of the defaults of the  $i$ th company as  $f(\mathbf{D}_i | \mathbf{X}_i, \mathbf{Y}; \boldsymbol{\beta}, \eta)$ . Under the condition which the default intensity is known, the defaults are i.i.d. Poisson distributed since the conditional

independence holds. Then the joint likelihood function becomes

$$\begin{aligned}
f(\mathbf{D}_i | \mathbf{X}_i, \mathbf{Y}; \boldsymbol{\beta}, \eta) &= f(D_{it_i}, D_{i(t_i+1)}, \dots, D_{iT_i} | \mathbf{X}_i, \mathbf{Y}; \boldsymbol{\beta}, \eta) \\
&= f(D_{it_i} | \mathbf{X}_i, \mathbf{Y}; \boldsymbol{\beta}, \eta) \times f(D_{i(t_i+1)} | \mathbf{X}_i, \mathbf{Y}; \boldsymbol{\beta}, \eta) \times \dots \times f(D_{iT_i} | \mathbf{X}_i, \mathbf{Y}; \boldsymbol{\beta}, \eta) \\
&= \prod_{t=t_i}^{T_i} \frac{(\lambda_{it} \Delta t)^{D_{it}} e^{-\lambda_{it} \Delta t}}{D_{it}!} \\
&= \prod_{t=t_i}^{T_i} (\lambda_{it} \Delta t)^{D_{it}} e^{-\lambda_{it} \Delta t} \\
&= e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} (\lambda_{it} \Delta t)^{D_{it}} \\
&= e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})].
\end{aligned}$$

Then, the joint likelihood function from all sample observations is

$$\begin{aligned}
\mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Y}; \kappa, \boldsymbol{\gamma}, \eta, \boldsymbol{\beta}) &= \mathcal{L}(\mathbf{D}_\infty, \mathbf{D}_\epsilon, \dots, \mathbf{D}_\uparrow | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta}) \times \mathcal{L}(\mathbf{X}, \mathbf{Y}; \kappa, \boldsymbol{\gamma}) \\
&= f(\mathbf{D}_1 | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta}) f(\mathbf{D}_2 | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta}) \dots f(\mathbf{D}_m | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta}) \mathcal{L}(\mathbf{X}; \boldsymbol{\gamma}) \mathcal{L}(\mathbf{Y}; \kappa) \\
&= \prod_{i=1}^m \left( e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right) \mathcal{L}(\mathbf{X}; \boldsymbol{\gamma}) \mathcal{L}(\mathbf{Y}; \kappa) \quad (1)
\end{aligned}$$

Since the time frailty  $\mathbf{Y}$  is unobservable and neither is the joint likelihood function, the MLE can not be obtained from the maximization of (1). To remove the unobserved time frailty from the likelihood function, a general solution is to take an expectation of the joint likelihood function and then obtain the MLE from maximization of the expected joint likelihood function. The expected joint log likelihood function is

$$\begin{aligned}
\log[\mathcal{L}(\mathbf{D}, \mathbf{X}; \kappa, \boldsymbol{\gamma}, \eta, \boldsymbol{\beta})] &= \int \log[\mathcal{L}(\mathbf{D}, \mathbf{X}, y; \kappa, \boldsymbol{\gamma}, \eta, \boldsymbol{\beta})] p_Y(y) dy
\end{aligned}$$

$$\begin{aligned}
&= \int [\log \mathcal{L}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_m | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta}) + \log \mathcal{L}(\mathbf{X}; \boldsymbol{\gamma}) + \log \mathcal{L}(y; \kappa)] p_Y(y) dy \\
&= E[\log \mathcal{L}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_m | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta})] + E[\log \mathcal{L}(\mathbf{Y}; \kappa)] + \log \mathcal{L}(\mathbf{X}; \boldsymbol{\gamma}), \quad (2)
\end{aligned}$$

where  $p_Y(\cdot)$  is the unconditional probability density function of the time frailty. Since the term  $\log \mathcal{L}(\mathbf{X}; \boldsymbol{\gamma})$  in (2) is independent to the time frailty, the parameter  $\boldsymbol{\gamma}$  can be estimated by regressing  $\mathbf{X}_t$  on  $\mathbf{X}_{t-1}$  and parameters  $\eta$ ,  $\boldsymbol{\beta}$ , and  $\kappa$  can be estimated by maximizing

$$\begin{aligned}
&E[\log \mathcal{L}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_m | \mathbf{X}, \mathbf{Y}; \eta, \boldsymbol{\beta})] + E[\log \mathcal{L}(\mathbf{Y}; \kappa)] \\
&= E \left\{ \log \left[ \prod_{i=1}^m \left( e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right) \right] \right\} \\
&E \left\{ \log \left[ \phi \left( y_1 | \mu = 0, \sigma^2 = \frac{1}{2\kappa} \right) \prod_{t=2}^T \phi \left( y_t | \mu = e^{-\kappa} y_{t-1}, \sigma^2 = \frac{1 - e^{-\kappa}}{2\kappa} \right) \right] \right\} \quad (3)
\end{aligned}$$

where  $\phi(\cdot)$  denotes the density function of a Gaussian distribution. Since the time frailty is still unobservable, MCEM (Monte Carlo Expectation Maximization) has to be applied to estimate the parameters  $\eta$ ,  $\boldsymbol{\beta}$ , and  $\kappa$ . The details of the MECM for the parameter estimations on the default intensity model with a time frailty are referred as to Duffie et al. (2009).

### 2.1.1 EM Algorithm for Model Estimation

The parameter vector,  $\boldsymbol{\theta}$  is estimated with the MLE by maximizing (3). As for the expectation in (3), it cannot be calculated but but can be approximated by Monte Carlo integration. This gives rise to the stochastic EM algorithm. Maximum likelihood estimation of the intensity parameter vector  $\boldsymbol{\theta}$  involves the following steps:

1. Initialize an estimate of  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \eta, \kappa)$  at  $\boldsymbol{\theta}^{(0)} = (\tilde{\boldsymbol{\beta}}_{noF}^{MLE}, 0.05, 0)$ , where  $\tilde{\boldsymbol{\beta}}_{noF}^{MLE}$  is the maximum likelihood estimator of  $\boldsymbol{\beta}$  in the model without frailty.



2. (E-step) Given the current parameter estimate  $\boldsymbol{\theta}^{(k)}$  and the observed covariate and default data  $\mathbf{X}$  and  $\mathbf{D}$ , respectively, draw  $m + n$  independent sample paths  $Y^{(1)}, \dots, Y^{(m+n)}$  from the conditional density  $p_Y(\cdot | \mathbf{X}, \mathbf{D}, \boldsymbol{\theta}^{(k)})$  of the latent OU frailty process  $Y$ . This is done with the Gibbs sampler described below. By the Law of Large Number,

$$\begin{aligned}\hat{Q}(\boldsymbol{\theta}^{(k)}) &= \frac{1}{n} \sum_{k=m+1}^{m+n} \log \mathcal{L}(\mathbf{D} | \mathbf{X}, Y^{(k)}; \boldsymbol{\beta}, \eta) \\ &\rightarrow \int \log \mathcal{L}(\mathbf{D} | \mathbf{X}, y; \boldsymbol{\beta}, \eta) p_Y(y) dy = E[\log \mathcal{L}(\mathbf{D} | \mathbf{X}, \mathbf{Y}; \boldsymbol{\eta}, \boldsymbol{\beta})]\end{aligned}$$

3. (M-step) Maximize  $\hat{Q}(\boldsymbol{\theta}^{(k)})$  with respect to the parameter vector  $\boldsymbol{\theta}$ , for example, by Newton–Raphson. The maximizing choice of  $\boldsymbol{\theta}$  is the new parameter estimate  $\boldsymbol{\theta}^{(k+1)}$ .
4. Replace  $k$  with  $k + 1$ , and return to Step 2, repeating the E-step and the M-step until reasonable numerical convergence is achieved.

### 2.1.2 Applying Gibbs Sampler with Frailty

The posterior density  $p_Y(\cdot | \mathbf{X}, \mathbf{D}, \boldsymbol{\theta})$  of the latent frailty process  $\mathbf{Y}$  is illustrated as follows. This is a complicated high-dimensional density. It is prohibitively computationally intensive to directly generate samples from this distribution. Nevertheless, MCMC methods can be used to explore this posterior distribution by generating a Markov chain over  $\mathbf{Y}$ , denoted  $\{\mathbf{Y}^{(n)}\}_{n \geq 1}^N$ , whose equilibrium density is  $p_Y(\cdot | \mathbf{X}, \mathbf{D}, \boldsymbol{\theta})$ .

The linchpin to MCMC is that the joint distribution of the frailty path  $\mathbf{Y} = \{Y_t : 0 \leq t \leq T\}$  can be broken down into a set of conditional distributions. A general version of the Clifford-Hammersley (CH) Theorem (Hammersley and Clifford (1970) and Besag (1974)) provides conditions under which a set of conditional distributions characterize a unique joint distribu-

tion. For example, in our setting, the CH Theorem indicates that the density  $p_Y(\cdot|\mathbf{X}, \mathbf{D}, \boldsymbol{\theta})$  is uniquely determined by the conditional distributions:

$$\begin{aligned} Y_0 &| Y_1, Y_2, \dots, Y_T, \mathbf{X}, \mathbf{D}, \boldsymbol{\theta} \\ Y_1 &| Y_0, Y_2, \dots, Y_T, \mathbf{X}, \mathbf{D}, \boldsymbol{\theta} \\ &\vdots \\ Y_T &| Y_1, Y_2, \dots, Y_{T-1}, \mathbf{X}, \mathbf{D}, \boldsymbol{\theta} \end{aligned}$$

or

$$Y_t | \mathbf{Y}_{(-t)}, \mathbf{X}, \mathbf{D}, \boldsymbol{\theta}.$$

It remains to show how to sample  $Y_t$  from its conditional distribution given  $\mathbf{Y}_{(-t)}$ . As, for  $0 \leq t \leq T$ ,

$$\begin{aligned} p(Y_t | \mathbf{X}, \mathbf{D}, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) &= \frac{p(\mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta})}{p(\mathbf{X}, \mathbf{D}, \mathbf{Y}_{(-t)}, \boldsymbol{\theta})} \\ &\propto p(\mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &= p(\mathbf{X}, \mathbf{D} | Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \times p(Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &= \frac{p(\mathbf{X}, \mathbf{D}, \boldsymbol{\theta} | Y_t, \mathbf{Y}_{(-t)})}{p(\boldsymbol{\theta} | Y_t, \mathbf{Y}_{(-t)})} \times p(Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &\propto p(\mathbf{X}, \mathbf{D}, \boldsymbol{\theta} | Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &= p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}) \times p(\mathbf{X}, \mathbf{D} | Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &\propto L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &= L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t | \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \times p(\mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\ &\propto L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t | \mathbf{Y}_{(-t)}, \boldsymbol{\theta}). \end{aligned}$$

Given the the Markov property,  $p(Y_t | \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) = p(Y_t | Y_{t-1}, Y_{t+1}, \boldsymbol{\theta})$ , then

$$\begin{aligned} p(Y_t | \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) &= p(Y_t | Y_{t-1}, Y_{t+1}, \boldsymbol{\theta}) \\ &= \frac{p(Y_{t-1}, Y_t, Y_{t+1} | \boldsymbol{\theta})}{p(Y_{t-1}, Y_{t+1} | \boldsymbol{\theta})} \end{aligned}$$

$$\begin{aligned}
&\propto p(Y_{t-1}, Y_t, Y_{t+1} | \boldsymbol{\theta}) \\
&= p(Y_{t-1}, Y_t | \boldsymbol{\theta}) \times p(Y_{t+1} | Y_{t-1}, Y_t, \boldsymbol{\theta}) \\
&= p(Y_{t-1}, Y_t | \boldsymbol{\theta}) \times \frac{p(Y_{t+1}, Y_{t-1} | Y_t, \boldsymbol{\theta})}{p(Y_{t-1} | \boldsymbol{\theta})} \\
&= \frac{p(Y_{t-1}, Y_t | \boldsymbol{\theta})}{p(Y_{t-1} | \boldsymbol{\theta})} \times p(Y_{t+1}, Y_{t-1} | Y_t, \boldsymbol{\theta}) \\
&= \frac{p(Y_{t-1}, Y_t | \boldsymbol{\theta})}{p(Y_{t-1} | \boldsymbol{\theta})} \times p(Y_{t+1} | Y_t, \boldsymbol{\theta}) \times p(Y_{t-1} | Y_t, \boldsymbol{\theta}) \\
&\propto \frac{p(Y_{t-1}, Y_t | \boldsymbol{\theta})}{p(Y_{t-1} | \boldsymbol{\theta})} \times p(Y_{t+1} | Y_t, \boldsymbol{\theta}) \\
&= p(Y_t | Y_{t-1}, \boldsymbol{\theta}) \times p(Y_{t+1} | Y_t, \boldsymbol{\theta}).
\end{aligned}$$

Hence

$$\begin{aligned}
p(Y_t | \mathbf{X}, \mathbf{D}, \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) &\propto L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t | \mathbf{Y}_{(-t)}, \boldsymbol{\theta}) \\
&\propto L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t, \mathbf{Y}_{(-t)}) \times p(Y_t | Y_{t-1}, \boldsymbol{\theta}) \times p(Y_{t+1} | Y_t, \boldsymbol{\theta}).
\end{aligned}$$

This equation determines the desired conditional density of  $Y_t$  given  $Y_{t-1}$  and  $Y_{t+1}$  in an implicit form. Although it is not possible to directly draw samples from this distribution, the Random Walk Metropolis-Hastings algorithm is applicable. The proposal density  $q(Y_t^{(n)} | \mathbf{X}, \mathbf{D}, Y_t^{(n-1)}, \boldsymbol{\theta}) = N(Y_t^{(n-1)}, 4)$ , that is, the mean is taken to be the value of  $Y_t$  from the previous iteration of the Gibbs sampler, and the variance to be twice the variance of the standard Brownian motion increments.

### 2.1.3 Random Walk Metropolis–Hastings Algorithm

The Random Walk Metropolis–Hastings algorithm to sample  $Y_t$  in the  $n$ th iteration of the Gibbs sampler is as follows.

1. Draw a candidate  $y \sim N(Y_t^{(n-1)}, 4)$

2. Compute

$$\alpha(y, Y_t^{(n)}) = \min \left( \frac{L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t = y, \mathbf{Y}_{(-t)}^{(n-1)})}{L(\boldsymbol{\theta} | \mathbf{X}, \mathbf{D}, Y_t^{(n-1)}, \mathbf{Y}_{(-t)}^{(n-1)})}, 1 \right)$$

3. Draw  $U$  with the uniform distribution on  $(0, 1)$ , and let

$$Y_t^{(n)} = \begin{cases} y & \text{if } U < \alpha(y, Y_t^{(n)}) \\ Y_t^{(n-1)} & \text{otherwise} \end{cases}$$

## 2.2 Default Intensity Model with an Industrial Frailty

Since the joint likelihood in the section (2.1) function of the default intensity is established under conditional independence among all time periods  $t$  and all firms  $i$ , the conditional independence will not hold when the default correlation exists among time  $t$  and/or among firms  $i$ . It is clear that the time frailty introduced by Duffie et al. (2009) is designed to capture the time dependence of defaults. However, the time frailty is not able to capture the default correlation caused from the dependence among firms. Since the industrial momentum is well documented to be the source of firm dependence, Chava et al. (2011) introduces an industrial frailty into the default intensity model. The default intensity model with an industrial frailty is modified as

$$\lambda_{jit} = Z_j \exp(\mathbf{X}_{jit}' \boldsymbol{\beta}),$$

with  $j = 1, \dots, S$  where  $S$  is the number of industries classified,  $i = 1, \dots, s_j$  where  $s_j$  is the number of firms in the  $j$ th industry,  $t = t_{ji}, t_{ji} + 1, \dots, T_{ji}$  where  $t_{ji}$  and  $T_{ji}$  denote the beginning and ending time of the  $i$ th firm in  $j$ th industry,  $\lambda_{jit}$  and  $\mathbf{X}_{jit}$  are the default intensity and sample values of explanatory variables of the  $i$ th firm in  $j$ th industry at time  $t$ , respectively, and  $Z_j$  denotes the industrial frailty of the  $j$ th industry. To have conjugate posterior distributions and closed form solutions for the parameters,  $Z_j \sim$

$i.i.d.$   $\Gamma(1/\theta, \theta)$  is assumed. To simplify the model, the dynamic system of AR(1) model for  $\mathbf{X}_t$  is ignored. In addition, by taking the time interval  $\Delta t$  small enough so that the number of default,  $D_{jit}$ , equals to 1 or 0 only.

Let  $\mathbf{D}_{ji} = (D_{jit_{ji}}, D_{ji(t_{ji}+1)}, \dots, D_{jiT_{ji}})$  represent the default status of  $i$ th firm which in the  $j$ th industry at all time intervals in its sample period  $t \in [t_{ji}, T_{ji}]$ . Denote  $\mathbf{D}_j = (D_{j1}, D_{j2}, \dots, D_{js_j})$  and  $\mathbf{X}_j = (\mathbf{X}_{j1}, \mathbf{X}_{j2}, \dots, \mathbf{X}_{js_j})$  as default status and sample values of explanatory variables, respectively, for all firms in the  $j$ th industry. Then,  $\mathbf{D} = (\mathbf{D}_1, \dots, \mathbf{D}_S)$  and  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_S)$  denote default status and sample values of explanatory variables for all firms in  $S$  industries, respectively. Also denote  $\mathbf{Z} = (Z_1, \dots, Z_S)$  as the industrial frailty vector for  $S$  industries. Since  $D_{jit}|\mathbf{X}, \mathbf{Z} \sim i.i.d. \text{Poisson}(\lambda_{jit})$ , the joint likelihood is

$$\begin{aligned} \mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \boldsymbol{\beta}) &= \mathcal{L}(\mathbf{Z}; \theta) \prod_{j=1}^S \prod_{i=1}^{s_j} \mathcal{L}(\mathbf{D}_{ji}|\mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) \\ &= \mathcal{L}(\mathbf{Z}; \theta) \prod_{j=1}^S \prod_{i=1}^{s_j} \left( e^{-\sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit} \Delta t} \prod_{t=t_{ji}}^{T_{ji}} (\lambda_{jit} \Delta t)^{D_{jit}} \right), \end{aligned}$$

where  $\mathcal{L}(\mathbf{Z}; \theta)$  represents the joint likelihood function of industrial frailties,  $Z_1, \dots, Z_S$ . As the time interval can be chosen as 1, i.e.,  $\Delta t = 1$ , the joint log likelihood function becomes

$$\begin{aligned} \log[\mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \boldsymbol{\beta})] &= \log[\mathcal{L}(\mathbf{Z}; \theta)] + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \log(\lambda_{jit}) - \sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit} \right) \\ &= \log[\mathcal{L}(\mathbf{Z}; \theta)] + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} \{D_{jit} [\log(Z_j) + \mathbf{X}'_{jit}\boldsymbol{\beta}] - Z_j \Lambda_{ji}\} \right) \quad (4) \end{aligned}$$

where  $\Lambda_{ji} = \sum_{t=t_{ji}}^{T_{ji}} \exp(\mathbf{X}'_{jit}\boldsymbol{\beta})$ .

Since  $Z_j \sim i.i.d.\Gamma(1/\theta, \theta)$ ,

$$\begin{aligned}\log[\mathcal{L}(\mathbf{Z}; \theta) &= \sum_{j=1}^S \log \left( \frac{1}{\theta^{1/\theta} \Gamma\left(\frac{1}{\theta}\right)} Z_j^{\frac{1}{\theta}-1} \exp\left(-\frac{Z_j}{\theta}\right) \right) \\ &= \sum_{j=1}^S -\frac{Z_j}{\theta} + \left(\frac{1}{\theta} - 1\right) \log(Z_j) - \left[ \log \Gamma\left(\frac{1}{\theta}\right) + \frac{1}{\theta} \log(\theta) \right] \quad (5)\end{aligned}$$

Then, substitute (5) into (4) and then

$$\begin{aligned}\log[\mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \boldsymbol{\beta})] &= \sum_{j=1}^S \left[ \left( \frac{1}{\theta} - 1 + D_j^* \right) \log(Z_j) - \frac{Z_j}{\theta} \right] - S \left[ \log \Gamma\left(\frac{1}{\theta}\right) + \left(\frac{1}{\theta}\right) \log(\theta) \right] \\ &\quad + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} (\mathbf{X}'_{jit} \boldsymbol{\beta}) - Z_j \Lambda_{ji} \right) \quad (6)\end{aligned}$$

Furthermore, the prior distribution of industrial frailty is assumed as  $Z_j \sim i.i.d.\Gamma(1/\theta, \theta)$  and then the posterior distribution of  $Z_j$  under the observed information  $\Omega = (\mathbf{D}, \mathbf{X}, \theta, \boldsymbol{\beta})$  is

*posterior*  $\propto$  *prior*  $\times$  *Likelihood*

$$\begin{aligned}&\propto \left[ \frac{Z_j^{\frac{1}{\theta}-1} \exp\left(-\frac{Z_j}{\theta}\right)}{\theta^{\frac{1}{\theta}} \Gamma\left(\frac{1}{\theta}\right)} \right] \times \prod_{i=1}^{s_j} \left( e^{-\sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit}} \prod_{t=t_{ji}}^{T_{ji}} [D_{jit} \lambda_{jit} + (1 - D_{jit})] \right) \\ &\propto \left[ \frac{Z_j^{\frac{1}{\theta}-1} \exp\left(-\frac{Z_j}{\theta}\right)}{\theta^{\frac{1}{\theta}} \Gamma\left(\frac{1}{\theta}\right)} \right] \times \prod_{i=1}^{s_j} \left[ \exp\left( \sum_{t=t_{ji}}^{T_{ji}} Z_j \exp(\mathbf{X}'_{jit} \boldsymbol{\beta}) \right) \prod_{t=t_{ji}}^{T_{ji}} [Z_j \exp(\mathbf{X}'_{jit} \boldsymbol{\beta})]^{D_{jit}} \right] \\ &\propto \left[ \frac{Z_j^{(\frac{1}{\theta} + D_j^*)-1} \exp\left(-Z_j \left[ -\frac{1}{\alpha} + \sum_{i=1}^{s_j} \sum_{t=t_{ji}}^{T_{ji}} \exp(\mathbf{X}'_{jit} \boldsymbol{\beta}) \right] \right)}{\theta^{\frac{1}{\theta}} \Gamma\left(\frac{1}{\theta}\right)} \right] \\ &\quad \times \prod_{i=1}^{s_j} \prod_{t=t_{ji}}^{T_{ji}} [\exp(\mathbf{X}'_{jit} \boldsymbol{\beta})]^{D_{jit}} \\ &\propto \left[ \frac{Z_j^{(\frac{1}{\theta} + D_j^*)-1} \exp\left(-Z_j \left[ -\frac{1}{\theta} + \sum_{i=1}^{s_j} \sum_{t=t_{ji}}^{T_{ji}} \exp(\mathbf{X}'_{jit} \boldsymbol{\beta}) \right] \right)}{\theta^{\frac{1}{\theta}} \Gamma\left(\frac{1}{\theta}\right)} \right] \quad (7)\end{aligned}$$

Let

$$\begin{aligned}
A_j &= \left( \frac{1}{\theta} + D_j^* \right) \\
C_j &= \left( \frac{1}{\theta} + \sum_{i=1}^{s_j} \sum_{t=t_{ji}}^{T_{ji}} \exp(\mathbf{X}'_{jit} \boldsymbol{\beta}) \right) \\
&= \left( \frac{1}{\theta} + \sum_{i=1}^{s_j} \Lambda_{ji} \right).
\end{aligned}$$

It is clear that the posterior distribution of  $Z_j$  is  $\Gamma(A_j, 1/C_j)$  with

$$\begin{aligned}
E(Z_j|\Omega) &= A_j/C_j \\
E[\log(Z_j)|\Omega] &= \psi(A_j) - \log(C_j),
\end{aligned}$$

where  $\psi(\cdot)$  is a digamma function. Thus, given  $\Omega = (\mathbf{D}, \mathbf{X}, \theta, \boldsymbol{\beta})$  and the parameters of posterior distributions of industrial frailties  $Z_j, j = 1, \dots, S$ ,  $A_j$  and  $C_j$ , the expected joint log likelihood is

$$\begin{aligned}
&E[\log \mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \boldsymbol{\beta})|\Omega, A_j, C_j, j = 1, \dots, S] \\
&= \sum_{j=1}^S \left[ \left( \frac{1}{\theta} - 1 + D_j^* \right) E[\log(Z_j|\Omega)] - \frac{E(Z_j|\Omega)}{\theta} \right] - S \left[ \log \Gamma \left( \frac{1}{\theta} \right) + \left( \frac{1}{\theta} \right) \log(\theta) \right] \\
&\quad + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \mathbf{X}'_{jit} \boldsymbol{\beta} - E(Z_j|\Omega) \Lambda_{ji} \right) \\
&= \sum_{j=1}^S \left[ \left( \frac{1}{\theta} - 1 + D_j^* \right) [\psi(A_j) - \log(C_j)] - \frac{A_j/C_j}{\theta} \right] - S \left[ \log \Gamma \left( \frac{1}{\theta} \right) + \left( \frac{1}{\theta} \right) \log(\theta) \right] \\
&\quad + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \mathbf{X}'_{jit} \boldsymbol{\beta} - \frac{A_j}{C_j} \Lambda_{ji} \right) \tag{8}
\end{aligned}$$

Since the industrial frailties are removed by taking expectation, the MLE is obtainable by maximizing (8) .

## 2.3 MCEM Algorithm

Since the expected joint log likelihood in (8) is evaluated under  $\Omega$ ,  $A_j$ , and  $C_j$  are known, the following EM algorithm is necessary. According to the initial value settings of the Duffie et al. (2009) and Chava, et al. (2011), the initial values of  $\theta$  and  $\beta$  are set as  $(\theta^{(0)}, \beta^{(0)}) = (1, \tilde{\beta}_{noF}^{MLE})$ , where  $\tilde{\beta}_{noF}^{MLE}$  is the MLE of the default intensity model without frailty. Then, the initial values of  $A_j$  and  $C_j$  are

$$\begin{aligned} A_j^{(0)} &= \left( \frac{1}{\theta^{(0)}} + D_j^* \right) \\ C_j^{(0)} &= \left( \frac{1}{\theta^{(0)}} + \sum_{i=1}^{s_j} \sum_{t=t_{ji}}^{T_{ji}} \exp(\mathbf{X}'_{jit} \tilde{\beta}_{noF}^{MLE}) \right). \end{aligned}$$

Substitute  $(\theta^{(0)}, \beta^{(0)}) = (1, \tilde{\beta}_{noF}^{MLE})$ ,  $A_j^{(0)}$ , and  $C_j^{(0)}$  into (8), the expected joint log likelihood is evaluated as

$$\begin{aligned} &E[\log \mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \beta) | \Omega, A_j^{(0)}, C_j^{(0)}, j = 1, \dots, S] \\ &= \sum_{j=1}^S \left[ \left( \frac{1}{\theta} - 1 + D_j^* \right) [\psi(A_j^{(0)}) - \log(C_j^{(0)})] - \frac{A_j^{(0)}/C_j^{(0)}}{\theta} \right] \\ &\quad - S \left[ \log \Gamma \left( \frac{1}{\theta} \right) + \left( \frac{1}{\theta} \right) \log(\theta) \right] \\ &\quad + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \mathbf{X}'_{jit} \beta - \frac{A_j^{(0)}}{C_j^{(0)}} \exp(\mathbf{X}'_{jit} \beta) \right). \end{aligned} \quad (9)$$

and then the first-step MLE of  $\theta$  and  $\beta$  are obtained as

$$\{\tilde{\theta}^{(1)}, \tilde{\beta}^{(1)}\} = \arg \max E[\log \mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \beta) | \Omega, A_j^{(0)}, C_j^{(0)}, j = 1, \dots, S].$$

Given  $\tilde{\theta}^{(1)}$  and  $\tilde{\beta}^{(1)}$ , the  $A_j$  and  $C_j$  are re-evaluated as

$$\begin{aligned} A_j^{(1)} &= \left( \frac{1}{\tilde{\theta}^{(1)}} + D_j^* \right) \\ C_j^{(1)} &= \left( \frac{1}{\tilde{\theta}^{(1)}} + \sum_{i=1}^{s_j} \sum_{t=t_{ji}}^{T_{ji}} \exp(\mathbf{X}'_{jit} \tilde{\beta}^{(1)}) \right), \end{aligned}$$



and then the expected joint log likelihood is re-evaluated as

$$\begin{aligned}
& E[\log \mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \boldsymbol{\beta}) | \Omega, A_j^{(1)}, C_j^{(1)}, j = 1, \dots, S] \\
&= \sum_{j=1}^S \left[ \left( \frac{1}{\theta} - 1 + D_j^* \right) [\psi(A_j^{(1)}) - \log(C_j^{(1)})] - \frac{A_j^{(1)}/C_j^{(1)}}{\theta} \right] \\
&\quad - S \left[ \log \Gamma \left( \frac{1}{\theta} \right) + \left( \frac{1}{\theta} \right) \log(\theta) \right] \\
&\quad + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \mathbf{X}'_{jit} \boldsymbol{\beta} - \frac{A_j^{(1)}}{C_j^{(1)}} \exp(\mathbf{X}'_{jit} \boldsymbol{\beta}) \right). \tag{10}
\end{aligned}$$

Then, the second-step MLE of  $\theta$  and  $\boldsymbol{\beta}$  are obtained as

$$\{\tilde{\theta}^{(2)}, \tilde{\boldsymbol{\beta}}^{(2)}\} = \arg \max E[\log \mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Z}; \theta, \boldsymbol{\beta}) | \Omega, A_j^{(1)}, C_j^{(1)}, j = 1, \dots, S].$$

Repeat above algorithm  $N$  steps till the MLE estimations of  $\tilde{\theta}^{(N)}$  and  $\tilde{\boldsymbol{\beta}}^{(N)}$  converge. Then,  $\tilde{\theta}^{(N)}$  and  $\tilde{\boldsymbol{\beta}}^{(N)}$  are the final MLE estimations for the default intensity model with industrial frailty.

## 2.4 Default Intensity Model with Dual Frailties

Lin and Chen (2015) shows that the time frailty of Duffie et al. (2009) can only capture the default correlation along time  $t$  and the industrial frailty of Chava et al. (2011) can only capture the default clustering among firms  $i$ . Consequently, the conditional independence will not hold for the default intensity model with either time frailty or industrial frailty. Therefore, it is necessary to consider a default intensity model with both time and industrial frailties simultaneously. This is called the “dual-frailty default intensity model.” The default intensity in the dual-frailty default intensity model is specified as

$$\lambda_{jit} = Z_j \exp(\mathbf{X}'_{jit} \boldsymbol{\beta} + \eta Y_t), j = 1, \dots, S, i = 1, 2, \dots, m, t = t_i, t_i + 1, \dots, T_i,$$

where the time frailty is also assumed to follow an OU process:  $dY_t = -\kappa Y_t dt + dB_t$ , and the industrial frailty is also assumed to be  $Z_j \sim i.i.d.\Gamma(1/\theta, \theta)$ . The parameters under estimate include  $\theta, \beta, \eta$ , and  $\kappa$ . By assuming the independence between the time and industrial frailties, the joint likelihood function is

$$\begin{aligned}\mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}; \theta, \beta, \eta, \kappa) &= \mathcal{L}(Y; \kappa) \mathcal{L}(\mathbf{Z}; \theta) \prod_{j=1}^S \prod_{i=1}^{s_j} \mathcal{L}(D_{ji} | \mathbf{X}, \mathbf{Y}, \mathbf{Z}; \beta, \eta) \\ &= \mathcal{L}(Y; \kappa) \mathcal{L}(\mathbf{Z}; \theta) \prod_{j=1}^S \prod_{i=1}^{s_j} \left( e^{-\sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit}} \prod_{t=t_{ji}}^{T_{ji}} (\lambda_{jit})^{D_{jit}} \right),\end{aligned}$$

and then the joint log likelihood function is

$$\begin{aligned}\log[\mathcal{L}(\mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}; \theta, \beta, \eta, \kappa)] &= \log[\mathcal{L}(Y; \kappa)] + \log[\mathcal{L}(\mathbf{Z}; \theta)] + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \log(\lambda_{jit}) - \sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit} \right).\end{aligned}$$

Since there are two latent variables, time frailty and industrial frailty, they can be removed by taking double expectation on the joint log likelihood function. The double expected joint log likelihood function is

$$\begin{aligned}E \left\{ E \left[ \log[\mathcal{L}(Y; \kappa)] + \log[\mathcal{L}(\mathbf{Z}; \theta)] + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \log(\lambda_{jit}) - \sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit} \right) \right] \right\} \\ = E[\log \mathcal{L}(Y; \kappa)] + E \left\{ E \left[ \log[\mathcal{L}(\mathbf{Z}; \theta)] + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \log(\lambda_{jit}) - \sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit} \right) \right] \right\}.\end{aligned}$$

In above equation, the first term is the same as the one in (3) and the second term is also the result in (8). Therefore, double expected joint log likelihood function can be rewritten as

$$E \left\{ E \left[ \log[\mathcal{L}(Y; \kappa)] + \log[\mathcal{L}(\mathbf{Z}; \theta)] + \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \log(\lambda_{jit}) - \sum_{t=t_{ji}}^{T_{ji}} \lambda_{jit} \right) \right] \right\}$$

$$\begin{aligned}
&= E \left\{ \log \left[ \phi \left( y_1 | \mu = 0, \sigma^2 = \frac{1}{2\kappa} \right) \prod_{t=2}^T \phi \left( y_t | \mu = e^{-\kappa} y_{t-1}, \sigma^2 = \frac{1 - e^{-\kappa}}{2\kappa} \right) \right] \right\} \\
&+ E \left\{ \sum_{j=1}^S \left[ \left( \frac{1}{\theta} - 1 + D_j^* \right) [\psi(A_j) - \log(C_j)] - \frac{A_j/C_j}{\theta} \right] \right\} \\
&- S \left[ \log \Gamma \left( \frac{1}{\theta} \right) + \left( \frac{1}{\theta} \right) \log(\theta) \right] \\
&+ E \left\{ \sum_{j=1}^S \sum_{i=1}^{s_j} \left( \sum_{t=t_{ji}}^{T_{ji}} D_{jit} \mathbf{X}'_{jit} \boldsymbol{\beta} - \frac{A_j}{C_j} \exp(\mathbf{X}'_{jit} \boldsymbol{\beta}) \right) \right\}. \tag{11}
\end{aligned}$$

Using the MCEM algorithm which is similar to the one in the default intensity model with time frailty, the MLE estimators for parameters  $\kappa, \eta, \boldsymbol{\beta}$  are obtainable.

## 2.5 Fisher's Dispersion Test for Conditional Independence

Das et al. (2007) applies the Fisher's dispersion test to check the conditional independence of defaults. Denote  $U(t) = \int_0^t \sum_{i=1}^n \lambda_{is} I_{[\pi_k > s]} ds$ , where  $\pi_k$  denotes the survival time,  $I_{[\pi_k > t]}$  is the indicator function for status of default, and  $\lambda_{is}$  is the default intensity of firm  $i$  at time  $s$ . Note that  $U(t)$  represents the summation of default intensities of all survival firms from the beginning to time  $t$ . It is obvious that  $c = U(t_{k+1}) - U(t_k)$  represents the expected number of defaults within the time period  $[t_k, t_{k+1}]$ . In addition, define  $W_k = \sum_{i=1}^n I_{[t_k < \tau_i < t_{k+1}]}$  to represent the observed number of default firms with time interval  $[t_k, t_{k+1}]$ . Fixing the bin size  $\hat{c}$ , a simple test for the null hypothesis of  $W_1, \dots, W_{T_{\hat{c}}}$  being independent Poisson distributed variables with mean parameter  $c$  is the Fisher's dispersion test, where  $T_{\hat{c}}$  is the number of time intervals with length  $\hat{c}$ . Under the null of conditional independence,

$$W = \sum_{k=1}^{T_{\hat{c}}} \frac{(W_k - \hat{c})^2}{\hat{c}}$$

is distributed as a  $\chi^2$  random variable with  $T_{\hat{c}} - 1$  degrees of freedom, where

$$\begin{aligned}\hat{c} &= \hat{U}(t_{k+1}) - \hat{U}(t_k) \\ &= \sum_{t=1}^{t_{k+1}} \sum_{i=1}^n \hat{\lambda}_{is} I_{[\pi_k > s]} - \sum_{t=1}^{t_k} \sum_{i=1}^n \hat{\lambda}_{is} I_{[\pi_k > s]}.\end{aligned}$$

Note that the mentioned Fisher's dispersion test is designed to test the conditional independence along time units.

Alternatively, to test the conditional independence among individual firms  $i$ , the test statistic is considered as

$$W^* = \sum_{j=1}^S \frac{(W_j^* - \hat{c}_j^*)^2}{\hat{c}_j^*},$$

where  $W_j^* = \sum_{i=1}^{s_j} \sum_{t=1}^T D_{jit}$  denotes the number of defaults in the  $j$ th industry within the sample period and  $\hat{c}_j^* = \sum_{i=1}^T \sum_{s=1}^{s_j} \hat{\lambda}_{is}$  denotes the expected number of defaults implied by the considered model of the  $j$ th industry. The test statistic  $W^*$ , under the null of conditional independence, is distributed as a  $\chi^2$  random variable with  $S - 1$  degrees of freedom.

### 3 Empirical Studies

The sample observations are the public listed firms on Taiwan's stock market from January of 1995 to January of 2015. All data are collected from the Taiwan Economic Journal (TEJ). The explanatory variables considered in this paper are the same as in Duffie, et al. (2007) which include the firm characteristic variables, the annual returns of stock (denoted as YRT.STOCK) and distance to default (denoted as DTD), and the macroeconomic variables, the yield rate of three-month Treasury Bills and the annual return of S&P 500 index. In this paper, we replace the yield rate of three-month Treasury Bills with three-month interest rate of certificate deposit of The First Bank

in Taiwan (denoted as RF) and the annual return of S&P 500 index with the annual return of Taiwan Stock Weighted Index (denoted as YRT.INDEX). Both of the annual returns are computed as the trailing 1-year return. Combining the parameter setting of the KMV model and the solution of nonlinear equations of Ronn and Verma (1986), the distances to default are calculated in this paper. To remove the extreme values of calculated distances to default, 5 % of winsorizations from top and from bottom are considered as in Duan (2000) and Duffie et al. (2007).

There are 16 types of default events listed on TEJ and only types C, D, E, H, F, G, I, N, S, and Z are considered as default events which are the same as the ones of Huang et al. (2012). Since monthly data are used, the default month is set as the month of the day of default occurred. According to Chava et al. (2011) and Huang et al. (2012), to remove the forward bias, the previous month of explanatory variables are taken as the independent variables and the default dummy of the current month as the dependent variable. Besides, there are 81 industry classifications (remove the industries related to finance) considered in this paper. The summary statistics of explanatory variables are shown in Table 1.

Table 1. Summary Statistics for Explanatory Variables

	Mean	Median	Std	Max	Min
RF	0.018	0.011	0.015	0.068	0.004
YRT.INDEX	0.058	0.086	0.240	0.799	-0.499
DTD	5.458	4.937	3.322	22.100	-3.600
YRT.STOCK	0.150	0.021	0.680	26.440	-0.987

Note that “Std” stands for sample standard deviation, “Max” and “Min” for the

sample maximum and minimum, respectively.

Table 2. Estimation Results of Default Intensity Models

	No Frailty	Industrial Frailty	Time Frailty	Dual-Frailty
Intercept	−6.819*** (0.225)	−6.811*** (0.224)	−6.860*** (0.222)	−6.846*** (0.222)
RF	17.181*** (4.307)	15.533*** (4.317)	12.233*** (4.455)	10.803** (4.501)
YRT.INDEX	0.571* (0.292)	0.548* (0.293)	0.288 (0.311)	0.274 (0.311)
DTD	−0.298*** (0.037)	−0.289*** (0.037)	−0.292*** (0.037)	−0.283*** (0.037)
YRT.STOCK	−2.952*** (0.318)	−2.973*** (0.318)	−3.144*** (0.320)	−3.156*** (0.321)
$\theta$		0.040*** (0.006)		0.026*** (0.004)
$\eta$			0.612*** (0.084)	0.624*** (0.086)
$\kappa$			0.613*** (0.129)	0.629*** (0.131)
log-Likelihood	-1294.960	-1291.774	-1268.831	-1266.921

Note that the column titles, “No Frailty”, “Industrial Frailty”, “Time Frailty”, and “Dual Frailty” stand for for default intensity models without frailty, with industrial frailty, with time frailty, and with dual frailties, respectively.

From the estimation results shown in Table 2, the marginal effect of risk-free interest rate is always significantly positive, which indicates that the

higher level of interest rate the more risk of firm default. This is because the risk-free interest rate represents the capital cost of firms so that the default risk is high when the capital cost is high. The variable YRT.INDEX is only significantly positive for models without frailties and the model with industrial frailty. However, it is insignificant for models with time frailty and dual frailties. This is because the annual return of aggregate stock is the proxy of the opportunity cost of capital. The marginal effect of YRT.INDEX is similar to the ones of RF. As to the marginal effect of distance to default, DTD, it is significantly negative in all models. Since the distance to default is a measure of default risk, it is positively correlated with default intensity and then the marginal effect is positive rationally. In addition, the marginal effect of annual stock return of individual firm is significantly positive for all models. This result indicates the higher of individual firm's stock return the higher default risk this firm commits. This maybe because the stock return reflects the market value of individual firm and then the higher stock return implies better performance of the firm and then the default risk is relative low.

The estimated coefficient of the industrial frailty,  $\hat{\theta}$ , in the model with pure industrial frailty is 0.040 which is significant at 1 % level. And the  $\hat{\theta}$  for the model with dual frailties is 0.026 which is also significant at 1 % level. As for the estimated coefficients,  $\hat{\eta}$ , they are 0.612 and 0.624 for the models with pure time frailty and dual frailties, respectively, and are both significant at 1 % level. For the comparisons of model fitting, the fitted log joint likelihood value is getting high as the time frailty or industrial frailty or dual frailties are included in the model. The model with dual frailties has the smallest log joint likelihood value among the considered models. These results imply that no matter the time frailty or the time frailty or the dual

frailties improve the model fitting of the default intensity model.

Since the conditional independence of all sample observations depends on the conditional independence along time ( $t = 1, \dots, T$ ) and among individuals ( $i = 1, \dots, T_i$ ), the Fisher's dispersion test is conducted for sample splits along time and for sample among individual firms. For the tests for sample splitting along time, the results are shown in Table 3. The considered bin sizes include  $c = 2, 4, 6, 8, 10$  and their corresponding number of sub-samples are 100 (df = 99), 50 (df=49), 34 (df=33), 25 (df = 24), and 20 (df = 19), respectively. It is clear that the test statistics are all significant at 1 % for all bin sizes considered. This implies that the conditional independence does not hold for the frailty intensity model without frailty. For the model with an industrial frailty, all test statistics are also significant which indicates that the conditional independence does not hold for the default intensity model with an industrial frailty. As for the Duffie et al. (2009)'s model with a time frailty, all statistics are insignificant for all considered bin sizes. It can also find that all test statistics are insignificant for all bin sizes. These two results imply that the time dependence of defaults can be captured by incorporating with a time frailty. It is worth noting that the industrial frailty itself can not capture the time dependence of defaults.

Table 3. Fisher's Dispersion Tests along Time

bin	df	No Frailty		Industrial Frailty		Time Frailty		Dual-Frailty	
		W	P-value	W	P-value	W	P-value	W	P-value
2	99	173.29	0.00	171.63	0.00	106.24	0.29	101.01	0.42
4	49	82.20	0.00	78.43	0.00	43.07	0.71	38.46	0.86
6	33	53.45	0.01	49.50	0.03	21.09	0.94	17.39	0.98
8	24	44.28	0.00	42.59	0.01	17.24	0.83	19.44	0.72
10	19	66.70	0.00	65.73	0.00	13.20	0.82	15.13	0.71



Note that “bin” indicates the bin size, “ $W$ ” denotes the Fisher’s dispersion test statistic, and “P-value” is the one-tail  $p$ -value of the test statistic  $W$ .

For checking the conditional independence among the individual firms, all firms are classified into 81 groups by industrial codes specified in TEJ. Table 4 shows the testing results. The test statistic,  $W^*$ , is significant at 10 % which indicates that the conditional independence does not hold for the model without any frailty. The test statistic is also significant at 10 % for the model with time frailty. This result indicates that the default dependence can not be captured by the time frailty. However, for the model with industrial frailty or with dual frailties, the test statistics are insignificant so that the conditional independence in the default intensity model is not rejected. Finally, the test statistic is still insignificant for the default intensity model with dual frailties.

Table 4. Fisher’s Dispersion Tests among Industries

df	No Frailty		Industrial Frailty		Time Frailty		Dual-Frailty	
	$W^*$	P-value	$W^*$	P-value	$W^*$	P-value	$W^*$	P-value
80	98.407	0.080	85.375	0.320	97.633	0.088	88.899	0.232

In summary, the default dependence is confirmed from the dependence along time and among individual firms. Incorporating a time frailty in the default intensity model can only capture the time dependence but not the dependence among firms. On the other hand, incorporating an industrial frailty in the default intensity model can only capture the dependence among individual firms but not the dependence along time. However, the dependence along time and among individual firms can be captured by the dual frailties in the default intensity model. Therefore, the default intensity model with dual frailties is suggestive.

## 4 Conclusions

The default intensity model has been widely used to measure the default risk of firms. The conditional independence is a crucial condition to have the model estimation appropriately. Even though, a time frailty is introduced by Duffie et al. (2009) to capture the default dependence along time and an industrial frailty is introduced by Chava et al. (2011) to capture the default dependence among firms, Lin and Chen (2015) shows that the conditional independence does not hold for the model either only with a time frailty or only with an industrial frailty. Therefore, a dual-frailty model with both time and industrial frailties is suggested in this paper.

Assuming the time frailty is independent to the industrial frailty, the joint log likelihood is derived first in this paper. And then, the MCEM procedures are also derived to do the model estimation. Using the monthly data of Taiwan's public listed firms from January of 1995 to January of 2015, the dual-frailty model is applied. From the estimation results, not only the model fitting is improved but also the conditional independence is not rejected by the Fisher's dispersion test. Therefore, the dual-frailty default intensity model is suggestive.

## References

- Altman, Edward I. (1968), “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *The Journal of Finance*, **23**, 589–609.
- Besag, J.E. (1974), “Spatial Interaction and the Statistical Analysis of Lattice Systems,” *Journal of the Royal Statistical Society, Series B*, **36**, 192–326.
- Campbell, J. Y., J. Hilscher, and J. Szilagyi (2008), ”In Search of Distress Risk,” *The Journal of Finance*, **62**, 2899–2939.
- Chava, S., C. Stefanescu, and S. Turnbull (2011), “Modeling the Loss Distribution,” *Management Science*, **57**, 1267–1287.
- Cox, David R. (1955), “Some Statistical Methods Connected with Series of Events,” *Journal of the Royal Statistical Society*, **17**, 129–164.
- Cox, David R. (1972), “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society*, **34**, 187–220.
- Das, S., D. Duffie, N. Kapadia, and L. Saita (2007), “Common Failings: How Corporate Defaults Are Correlated,” *The Journal of Finance*, **62**, 93–117.
- Duan, J.C. (2010), “Clustered Defaults,” Working paper, Risk Management Institute, Business School, National University of Singapore.
- Duffie, D., L. Saita, and K. Wang (2007), “Multi-Period Corporate Default Prediction with Stochastic Covariates,” *Journal of Financial Economics*, **83**, 635–665.

- Duffie, D., A. Eckner, G. Horel, and L. Saita (2009), “Frailty Correlated Default,” *The Journal of Finance*, **64**, 2089–2123.
- Hammersley, J. M., and P. Clifford (1970), “Markov Fields on Finite Graphs and Lattices,” unpublished.
- Huang, R.-C. and T.-C. Jan (2011), “Frailty Correlated Default Model: An Application of Taiwan’s Public Listed Firms, ” *Journal of Chinese Statistic*, **49**, 98–122.
- Huang, R.-C., C.-S. Wu, J.-L. Lin, and J.-S. Sio (2012), “An Empirical Study of the Determinants of Financial Distress for Taiwan Corporations,” , **13**, 55–77.
- Lefebvre, Mario (2005), Applied Stochastic Processes, Springer.
- Lin, G.-Y. and M.-Y. Chen (2015), “Re-examination on Default Clustering for Taiwan Listed Firms: Industrial and Time Frailties Perspective,” Working paper, Department of Finance, National Chung Hsing University, Taichung, Taiwan.
- Ronn, E.I. and A.K. Verma (1986), “Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model,” *The Journal of Finance*, **41**, 871–896.
- Shumway, T. (2001), ”Forecasting Bankruptcy More Accurately: A Simple Hazard Model,” *The Journal of Business*, **74**, 101–124.