Clustered Defaults

Jin-Chuan Duan*

(November 12, 2009)

Abstract

Defaults in a credit portfolio of many obligors or in an economy populated with firms tend to occur in waves. This may simply reflect their sharing of common risk factors and/or manifest their systemic linkages via credit chains. One popular approach to characterizing defaults in a large pool of obligors is the Poisson intensity model coupled with stochastic covariates. A constraining feature of such models is that defaults of different obligors are independent events after conditioning on the covariates, which makes them ill-suited for modeling clustered defaults. Although individual default intensities under such models can be high and correlated via the stochastic covariates, joint default rates will always be zero, because the joint default probabilities are in the order of the length of time squared or higher. In this paper, we develop a hierarchical intensity model with three layers of shocks – common, group-specific and individual. When a common (or group-specific) shock occurs, all obligors (or group members) face individual default probabilities, determining whether they actually default. The joint default rates under this hierarchical structure can be high, and thus the model better captures clustered defaults. This hierarchical intensity model can be estimated using the maximum likelihood principle. Its predicted default frequency plot is used to complement the typical cumulative accuracy plot (CAP) in default prediction. We implement the new model on the US corporate default/bankruptcy data and find it superior to the standard intensity model.

Keywords: Default probability, default correlation, hazard rate, maximum likelihood, Poisson process, CAP, common shock, distance to default, Kullback-Leibler distance.

JEL classification code: C51, G13.

---

*Duan is with Risk Management Institute, National University of Singapore and the Department of Finance in the NUS Business School. E-mail: bizdjc@nus.edu.sg. The author would like to acknowledge the excellent research assistance from Tao Wang.

1 Introduction

Understanding the determinants of defaults is critical to many business and policy decisions. Credit analysis beyond single names is at the heart of credit portfolio management, and also has important regulatory policy implications. From a policy perspective, the severe credit crunch in the early phase of the recent financial crisis has particularly driven home the message that system-wide corporate defaults in a scale unprecedented is not only possible but also quite likely. Adding to this is a widely acknowledged role of credit rating agencies in fueling this financial crisis. Credit rating models used by the key rating agencies have been seriously questioned. As part of the remedial solutions, better credit analytical tools for modeling multiple obligors together are needed.

Defaults in a credit portfolio of many obligors or in an economy populated with firms tend to occur in waves. This may simply reflect their sharing of common risk factors and/or manifest their systemic linkages via credit chains. Two broad categories of modeling approaches in dealing with a credit portfolio have emerged in the literature – top-down and bottom-up. The top-down approach directly models the aggregate behavior at the portfolio level, and is intended for answering questions only concerning the overall portfolio. Examples are Arnsdorf and Halperin (2007), Cont and Minca (2007), Giesecke and Kim (2007), and Longstaff and Rajan (2007). In contrast, the bottom-up approach models individual names together by specifying their joint behavior. Such a model offers a wealth of information, but may deliver unsatisfactory performance at the aggregate level due to the built-in constraints from modeling individual obligors. Examples abound; for example, Li (2000), Shumway (2001), Andersen, *et al* (2003), Duffie, *et al* (2007), Duffie, et al (2008), and Peng and Kou (2009).

One popular approach to characterizing defaults in a large pool of obligors is the Poisson intensity model coupled with stochastic covariates, or the Cox process for short. Shumway (2001) and Duffie, *et al* (2007) are such examples. Azizpour and Giesecke (2008) reported 24 rail firm defaults in a single day on June 21, 1970, and used that to motivate their use of a self-exciting model (defaults generating more defaults) to deal with an abnormally large number of clustered defaults. Peng and Kou (2009) constructed a bottom-up model by observing that making cumulative intensity process jump, instead of jumps in the intensity process, can create bursty defaults. Duffie, *et al* (2008) built "frailty" into the Poisson intensity model by introducing a latent variable so as to increase default clustering.

A constraining feature of the Poisson intensity models in the bottom-up application is that defaults of different obligors are independent events after conditioning on the covariates, which makes them ill-suited for modeling clustered defaults. Das, *et al* (2007) showed by a battery of tests that the standard intensity model such as Duffie, *et al* (2007) simply does not generate enough default clustering as in the observed data. Although individual default intensities under the standard intensity models can be high and correlated via the stochastic covariates, joint default rates will always be zero, because the joint default probabilities are in the order of the length of time squared or higher. This conclusion applies to all Poisson intensity models except that of Peng and Kou (2009). The Peng and Kou (2009) approach with jump in the cumulative intensity process amounts to making the local intensity a Dirac delta function, and thus the joint default intensity does not vanish locally.

In this paper, we develop a hierarchical intensity model with three layers of shocks – common, group-specific and individual. When a common (or group-specific) shock occurs, all obligors (or group members) face individual default probabilities, determining whether they actually default. The joint default rates under this hierarchical structure can be high, and thus the model better captures clustered defaults. We develop algorithm which can compute the time-varying predicted default distribution for the credit portfolio for the standard and hierarchical intensity models. Such predicted default distributions have many applications; for example, one can compute the expected number of defaults for the next period. If coupled with the prescribed dynamics of the stochastic covariates, these predicted default distributions can be time aggregated via Monte Carlo simulations so as to cover multiple periods ahead.

This hierarchical intensity model can be estimated using the maximum likelihood principle. We implement the model on a US corporate data set with monthly frequency over the period of January 1991 to December 2008. The analysis shows that common shock is a statistically significant component. Adding a layer to create a hierarchical structure can indeed better the performance of the intensity model.

2 A hierarchical intensity model for clustered defaults

2.1 The model

Consider a credit portfolio consisting of many obligors (firms, debt issues, or individuals). For obligor (i, j), which is the j-th member of the i-the group where $i = 1, \cdots, K$ and $j = 1, \cdots, n_i$, we assume that its default is governed by the following process: for $t \geq 0$,

$$dM_{ijt} = \chi_{ijt} dN_{ct} + \zeta_{ijt} dN_{it} + dN_{ijt}. \tag{1}$$

where $N_{c0} = N_{i0} = N_{ij0} = 0$, and χ_{ijt} (or ζ_{ijt}) is a Bernoulli random variable taking value of 1 with a probability of p_{ijt} (or q_{ijt}) and 0 with a probability of $1 - p_{ijt}$ (or $1 - q_{ijt}$). χ_{ijt} and ζ_{ijt} are independent of each other and also independent across different obligors. The Poisson process N_{ct} is a common process shared by all obligors in a credit portfolio which is governed by the intensity λ_{ct}. The Poisson process specific to a group is N_{it} which is shared by all its members and subject to the intensity λ_{it}. The Poisson process unique to obligor j is the i-th group is N_{ijt}. All different Poisson processes are independent of each other.

The way to understand the above setup is as follows. N_{ct} captures the top hierarchy, say, the global event like the 2008-09 financial crisis which has severe impact beyond national boundaries. When there is no global event, obligors may still be subjected to a national event that affects many firms and individuals in that country. This is reflected in N_{it}, the middle hierarchy. When there is no common credit event globally or nationally, an individual entity can still default which is captured by N_{ijt}, the bottom hierarchy. Similarly, the three-layer hierarchical setup is equally applicable to modelling defaults due to national, industry-wide, and individual factors. Needless to elaborate, the three-layer model can be extended to more layers or reduced to just two or one layer. In the case of one layer, the hierarchial intensity model becomes exactly the standard intensity model for defaults such as that of Duffie, *et al* (2007).

When a common jump occurs, (i, j)-th obligor may or may not default depending on the value of χ_{ijt}. When its value equals one, we say that obligor (i, j) defaults. For a group-specific credit event, individual entities may react differently. The obligor (i, j) will default if $\zeta_{ijt} = 1$. As to the individual jump, an obligor defaults if its own jump occurs.

Mathematically, individual jumps will not occur concurrently even if their intensities are highly correlated. This is because the probability of a concurrent default of k obligors will have a rate equal to the length of the time interval raised to the power of $k - 1$, which becomes negligible when $k > 1$. Casting aside the limiting argument and fixing the time interval at some fixed length, the clustered default probability can in principle be raised to any desired level by increasing individual default intensities. Matching clustered default this way will, however, come at the expense of overstating individual default probabilities. This is in effect the modeling dilemma facing the default intensity models in the literature.

In our hierarchical intensity model, however, clustered defaults can occur via two channels without disturbing individual default probabilities. The two channels are: the common jump and the group-specific jump. Intuitively, defaults under the first channel will be more widespread as compared to the second one. When a common or group-specific jump happens, each obligor is facing a probability of default. If we assume for simplicity that the default probability under a common credit event is same for all obligors, say p, then the number of defaulted obligors can be described by a binomial distribution. Under such an assumption, n obligors out of the survived population of size n^* ($n^* \leq \sum_{i=1}^{K} n_i$) default at the same time has the probability of $\begin{pmatrix} n^* \\ n \end{pmatrix} p^n(1-p)^{n^*-n}$. The distribution for the number of defaults in the credit portfolio, ranging from 0 to n^*, can be easily computed with this binomial distribution. A similar calculation applies to a group-specific credit event.

Following Duffie, *et al* (2007), an obligor may leave the population due to factors other than default. For example, a firm can be de-listed from an securities exchange due to a merger, or an individual credit card holder decides to terminate the use of a particular credit card. A Poisson process L_{ijt} with intensity δ_{ijt} and $L_{ij0} = 0$ is used to model the exit for reasons other than default. This Poisson process is assumed to be independent of all other Poisson processes described earlier. Although we can also impose a hierarchical structure on exits not due to default, we opt for simplicity by focussing on the more important issue of clustered defaults.

Explicitly considering non-default exits, as opposed to simply ignoring them, is important more for the reason of characterizing the default behavior, because it must occur before a non-default exit. Interestingly, censoring does not affect the estimation of the parameters in the default intensity functions when other forms of exits are assumed to be independent of the default process. This feature comes from the fact that the overall likelihood function can be decomposed into unrelated components, and was utilized in Duffie, *et al* (2007).

We let the Poisson intensities be functions of some common state variables X_t, group-specific state variables Y_{it} and obligor-specific factors Z_{ijt}. Although it is natural to think that the common jump intensity is influenced by common state variables, it is conceivable that some group-specific or even obligor-specific factors can affect the common jump intensity. In that case, the group-specific or obligor-specific state variable is regarded as a common state variable. Needless to say, the individual default probability under a common jump may be affected by the common state variables, group-specific state variables and obligor-specific factors. Thus, we have

$$\lambda_{ct} = F(X_{t_-}) \tag{2}$$
$$\lambda_{it} = G(X_{t_-}, Y_{it_-}), \text{ for } i = 1, \cdots, K \tag{3}$$
$$\lambda_{ijt} = H(X_{t_-}, Y_{it_-}, Z_{ijt_-}), \text{ for } i = 1, \cdots, K \text{ and } j = 1, \cdots, n_i \tag{4}$$
$$p_{ijt} = P(X_{t_-}, Y_{it_-}, Z_{ijt_-}), \text{ for } i = 1, \cdots, K \text{ and } j = 1, \cdots, n_i \tag{5}$$
$$q_{ijt} = Q(X_{t_-}, Y_{it_-}, Z_{ijt_-}), \text{ for } i = 1, \cdots, K \text{ and } j = 1, \cdots, n_i \tag{6}$$
$$\delta_{ijt} = R(X_{t_-}, Y_{it_-}, Z_{ijt_-}), \text{ for } i = 1, \cdots, K \text{ and } j = 1, \cdots, n_i \tag{7}$$

where t_- denote the left limit, F, G, H and R must be non-negative functions, and P and Q must be bounded between 0 and 1. Both of which can be easily accomplished with the standard modelling techniques. The hierarchical intensity model thus far comprises a family of doubly stochastic Poisson processes or Cox processes.

If all state variable processes have continuous sample paths, it makes no difference in theory as to using t_- or t. In practice, however, one only have discretely sampled data, and t_- means using the data available at time $t - \Delta t$.

By the additivity of independent Poisson processes, the default component of the above model viewed individually can be reduced to

$$dM_{ijt} \stackrel{d}{=} \chi^*_{ijt} dN^*_{ijt} \tag{8}$$

where $\stackrel{d}{=}$ stands for distributional equivalence; N^*_{ijt} is a Poisson process with the intensity equal to $\lambda_{ct} + \lambda_{it} + \lambda_{ijt}$; and χ^*_{ijt} is a Bernoulli random variable taking value of 1 with a probability of p^*_{ijt} and 0 with a probability of $1 - p^*_{ijt}$. Note that

$$p^*_{ijt} = \frac{\lambda_{ct}}{\lambda_{ct} + \lambda_{it} + \lambda_{ijt}} p_{ijt} + \frac{\lambda_{it}}{\lambda_{ct} + \lambda_{it} + \lambda_{ijt}} q_{ijt} + \frac{\lambda_{ijt}}{\lambda_{ct} + \lambda_{it} + \lambda_{ijt}}. \tag{9}$$

It is clear that $\chi^*_{ijt} dN^*_{ijt}$ is also equivalent in distribution to a Poisson process with intensity of $p^*_{ijt}(\lambda_{ct} + \lambda_{it} + \lambda_{ijt})$ with respect to the default time. Even though they are not equivalent beyond the default time, it is irrelevant as far as modelling default is concerned. Thus, if we look at an obligor individually, the hierarchical intensity model is equivalent to the Duffie, *et al* (2007) model. But for two or more obligors considered together, the two models are not equivalent, i.e., $(M_{ijt}, M_{klt}) \stackrel{d}{\neq} (\int_0^t \chi^*_{ijs} dN^*_{ijs}, \int_0^t \chi^*_{kls} dN^*_{kls})$. The distinguishing feature of the hierarchical intensity model is therefore its ability to better capture clustered defaults.

2.2 Predicted default frequency on the the natural time scale

Perhaps, we might expect the common jump intensity to be low vis-a-vis the group-specific jump, and the group-specific jump intensity is in turn lower than individual jump intensity. But it is also plausible that common shocks actually occur more frequently, but upon occurrence, individual obligors face time varying default probabilities p_{ijt} which sometimes causes many concurrent defaults whereas other times only generates few or no concurrent defaults. In the case of a group-specific event, the concurrent defaults will be clustered in a group and the size of the default cluster will naturally depend on the magnitude of q_{ijt} in that group.

One way to appreciate the difference between the hierarchical and standard intensity models is to compare the distributions for the number of defaults when the hierarchical intensity model (HIM) is to allow for common shocks but disable the group and individual shocks. Let U be the number of defaults out of the obligor pool over the period of $[t, t + \Delta t]$.

The restricted hierarchical intensity model (rHIM) has the following distribution:

$$\begin{aligned}
Prob^{rHIM}(U=0) &= e^{-\lambda_{ct}\Delta t} + (1-e^{-\lambda_{ct}\Delta t})\prod_{i=1}^{K}\prod_{j=1}^{n_i}(1-p_{ijt}) \\
Prob^{rHIM}(U=1) &= (1-e^{-\lambda_{ct}\Delta t})\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(p_{ijt}\prod_{m=1}^{K}\prod_{l=1}^{n_i}(1-p_{mlt})^{1_{\{(m,l)\neq(i,j)\}}}\right) \\
&\vdots
\end{aligned}$$

The first part of $Prob^{rHIM}(U=0)$ is the probability of no common shock, and hence no default occurs. The second part is the probability of the event that the common shock occurs but still no obligor defaults. This probability is in sharp contrast to the one under the standard intensity model of Duffie, *et al* (2007) (DSW):

$$\begin{aligned}
Prob^{DSW}(U=0) &= \prod_{i=1}^{K}\prod_{j=1}^{n_i} e^{-\lambda_{ijt}\Delta t} \\
Prob^{DSW}(U=1) &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left((1-e^{-\lambda_{ijt}\Delta t})\prod_{m=1}^{K}\prod_{l=1}^{n_i} e^{-\lambda_{mlt}\Delta t 1_{\{(m,l)\neq(i,j)\}}}\right) \\
&\vdots
\end{aligned}$$

It is evident from the above expressions, the probability of no default can be computed rather easily by the analytical expression for either model, but for one, two more defaults, the analytical expressions will become increasingly complex, and they will not useful for computation. For the general hierarchical intensity model, adding to the complexity is the existence of group-specific and individual shocks. Therefore, the default distributions will need to be computed numerically. The exact numerical method based on the convolution principle is described in Appendix, which is a very efficient algorithm for coming up with the theoretical default distribution for either the standard or hierarchial intensity model.

The above predicted default distribution for the obligor pool is time-varying. By averaging the theoretical default distributions over time, however, we can devise a useful theoretical signature plot for the time series sample. It can then be compared to the the observed frequency computed on the natural time scale.

It is worth noting that our predicted default distribution plot is fundamentally different from the rescaled-time default distribution plot used in conjunction with the Fisher dispersion test devised by Das, *et al* (2007). Our signature plot is based on the original time scale, which is arguably more natural. Perhaps more importantly, the rescaled-time approach is not applicable to the hierarchical intensity model due to its lack of independence (conditional on stochastic covariates) across obligors.

Use the parameter estimates obtained later in the empirical study of the US corporate data, we compare one version of the hierarchical intensity model with the standard intensity model, i.e, Duffie, *et al* (2007). In Figure 1, the bars represent the observed frequencies corresponding to different numbers of defaults over the sample period (defaults per month). The two curves correspond to the averaged predicted default distributions over time under two models. Although two predicted distributions have some difference, both seem to predict the observed frequency well. The two models differ in the way that the hierarchical intensity model distributes weights more towards two ends.

If there is a variable informative of the common shock's arrival, then one can hope to reveal the difference between the the standard and hierarchical intensity models. Figures 2a and 2b are presented for this purpose. We use the average distance-to-default for the financial firms in the sample to divide the time series sample into three equal-size groups. The bottom third corresponds to the group with the lowest average distance-to-default. We compute the observed frequency of this subsample and compare it to the averaged predicted default distribution with the average taken over the subsample. The results for the bottom third sample are presented in Figure 2a. It is evident that the predicted default distribution under the hierarchical model shifts weights towards larger numbers of defaults. This is consistent with the fact that this is the subsample facing smaller market-wide distance-to-default for which we expect to see clustered defaults. The result for the top third subsample is opposite of that for the bottom third, and works out as expected. Figure 2b suggests that the hierarchical model leads to smaller numbers of defaults than does the standard intensity model when the market-wide distance-to-default is bigger.

2.3 Default correlation and double default

Default correlation and double default are two informative ways of understand the model. They are conceptually useful in examining a model's suitability for analyzing credit portfolios. The well-known default intensity model, for example, Duffie, *et al* (2007), yields too low a default correlation and double default probability vis-a-vis the empirical observations. In contrast, our hierarchical intensity model generates a much higher values in relative term as shown below.

Let $\tau_{ij} = \inf(t; M_{ijt} \geq 1)$, i.e., the random default time for the j-th firm in the i-th group. Similarly, let $\tau^*_{ij} = \inf(t; \int_0^t \chi^*_{ijs} dN^*_{ijs} \geq 1)$ and assume that N^*_{ijs} are independent Poisson processes for different obligors to mimic the Duffie, *et al* (2007) model. We also need to define the non-default exit time, $\phi_{ij} = \inf(t; L_{ijt} \geq 1)$, to meaningfully describe the joint default behavior.

A key element to understanding intensity models is the kill rate. First, let $a_t(ij, kl) = 1-(1-p_{ijt})(1-p_{klt})$ be the probability of two obligors fail to jointly survive with respect to the common credit event. In other words, it is the probability that one of the two obligors fails or both fail facing a common credit event. Similarly, $b_t(ij, kl) = 1 - (1 - q_{ijt})(1 - q_{klt})$ is for the group-specific credit event. Define

$$\alpha_t(ij, kl) \equiv a_t(ij, kl)\lambda_{ct} + 1_{\{i=k\}} b_t(ij, kl)\lambda_{it} + 1_{\{i\neq k\}}(q_{ijt}\lambda_{it} + q_{klt}\lambda_{kt}) + \lambda_{ijt} + \lambda_{klt} \quad (10)$$

and

$$\alpha^*_t(ij, kl) \equiv (p_{ijt} + p_{klt})\lambda_{ct} + q_{ijt}\lambda_{it} + q_{klt}\lambda_{kt} + \lambda_{ijt} + \lambda_{klt} \quad (11)$$

It is clear that $\alpha^*_t(ij, kl) \geq \alpha_t(ij, kl)$.

Consider two obligors in the same group – (i, j) and (i, l). We now compute the kill rate for determining joint survival over the interval $[t, t+\Delta t]$ for the hierarchical intensity model and the Duffie, *et al* (2007) model, respectively. Under the hierarchical intensity model, the kill rate for the joint survival probability is

$$\lim_{\Delta t\to 0} \frac{1 - (1 - a_t(ij, il)\lambda_{ct}\Delta t)(1 - b_t(ij, il)\lambda_{it}\Delta t)(1 - \lambda_{ijt}\Delta t)(1 - \lambda_{ilt}\Delta t)}{\Delta t} = \alpha_t(ij, il). \quad (12)$$

Note that the numerator of the kill rate is the probability that two obligors fail to jointly survive the interval after considering the common, group-specific and individual credit events together. The kill rate can in turn be used to derive the joint survival probability as follows:

$$\begin{aligned} E_0\left(1_{\{\tau_{ij}>t\}}1_{\{\tau_{ik}>t\}}\right) &= E_0\left\{E_0\left(1_{\{\tau_{ij}>t\}}1_{\{\tau_{ik}>t\}}\middle| X_s, Y_{is}, Z_{ijs}, Z_{iks}; s\in[0,t]\right)\right\} \\ &= E_0\left(e^{-\int_0^t a_s(ij,ik)ds}\right). \end{aligned} \quad (13)$$

If two obligors – (i, j) and (k, l) – are from different groups (i.e., $i \neq k$), a similar result can be derived:

$$\lim_{\Delta t\to 0} \frac{1 - (1 - a_t(ij, kl)\lambda_{ct}\Delta t)[1 - (q_{ijt}\lambda_{it} + \lambda_{ijt})\Delta t][1 - (q_{klt}\lambda_{kt} + \lambda_{klt})\Delta t]}{\Delta t} = \alpha_t(ij, kl). \quad (14)$$

Under the Duffie, *et al* (2007) model and using the mimicking structure mentioned above, the kill rate for any two obligors is always

$$\lim_{\Delta t\to 0} \frac{1 - [1 - (p_{ijt}\lambda_{ct} + q_{ijt}\lambda_{it} + \lambda_{ijt})\Delta t][1 - (p_{klt}\lambda_{ct} + q_{klt}\lambda_{it} + \lambda_{klt})\Delta t]}{\Delta t} = \alpha^*_t(ij, kl). \quad (15)$$

Double default probability and default correlation from time 0 to t for two obligors, after factoring in the censoring effect, can be computed. Their relationships to the counterparts under the Duffie, *et al* (2007) model are given in the following proposition.

Proposition 1. Censored double default probability and default correlation

Double default probability:

$$E_0\left(1_{\{\tau_{ij}\le t\wedge\phi_{ij}\}}1_{\{\tau_{kl}\le t\wedge\phi_{kl}\}}\right)\ge E_0\left(1_{\{\tau^*_{ij}\le t\wedge\phi_{ij}\}}1_{\{\tau^*_{kl}\le t\wedge\phi_{kl}\}}\right) \tag{16}$$

Default correlation:

$$Corr_0\left(1_{\{\tau_{ij}\le t\wedge\phi_{ij}\}},1_{\{\tau_{kl}\le t\wedge\phi_{kl}\}}\right)\ge Corr_0\left(1_{\{\tau^*_{ij}\le t\wedge\phi_{ij}\}},1_{\{\tau^*_{kl}\le t\wedge\phi_{kl}\}}\right) \tag{17}$$

Proof: see Appendix

The censored double default probability and default correlation are what one cares about because after an obligor exits for other reasons, default is no longer a relevant concept. By the above results, the censored double default probability or default correlation between two obligors under our hierarchical intensity structure is always higher than that under the Duffie, *et al* (2007) model. Needless to say, the directional relationship still holds true when one does not consider the effect of censoring. Moreover, it can be shown that the censored double default probability or default correlation between two obligors in the same group will be, under our hierarchical intensity model, higher than that for the two obligors in different groups but otherwise comparable.

Double default probability and default correlation as functions of time horizon can be analytically derived under either the Duffie, *et al* (2007) model or the hierarchial intensity model. The exact expressions under the constant parameter assumption are provided in Appendix. If parameters are time-varying, simulations can be used.

Another way of understanding the hierarchical intensity model vis-a-vis the Duffie, *et al* (2007) model is through a double-survival hazard rate analysis. We define a hazard rate that can be used to compute the double-survival probability in a usual way of linking the hazard rate to the survival probability. Due to other exit factors, the survival here means that neither default nor other types of exit has occurred.

The corresponding pair of hazard rates is

$$\begin{aligned} h_t(u;ij,kl) &= -\frac{\partial E_t\left(1_{\{\tau_{ij}\wedge\phi_{ij}>u\}}1_{\{\tau_{kl}\wedge\phi_{kl}>u\}}\right)/\partial u}{E_t\left(1_{\{\tau_{ij}\wedge\phi_{ij}>u\}}1_{\{\tau_{kl}\wedge\phi_{kl}>u\}}\right)} \\ &= \frac{E_t\left((\alpha_u(ij,kl)+\delta_{iju}+\delta_{klu})e^{-\int_t^u(\alpha_s(ij,kl)+\delta_{ijs}+\delta_{kls})ds}\right)}{E_t\left(e^{-\int_t^u(\alpha_s(ij,kl)+\delta_{ijs}+\delta_{kls})ds}\right)} \end{aligned} \tag{18}$$

Its counterpart under the Duffie, *et al* (2007) model is

$$
\begin{aligned}
h_t^*(u;ij,ik) &= -\frac{\partial E_t\left(1_{\{\tau_{ij}^*\wedge\phi_{ij}>u\}}1_{\{\tau_{ik}^*\wedge\phi_{ik}>u\}}\right)/\partial u}{E_t\left(1_{\{\tau_{ij}^*\wedge\phi_{ij}>u\}}1_{\{\tau_{ik}^*\wedge\phi_{ik}>u\}}\right)} \\
&= \frac{E_t\left((\alpha_u^*(ij,kl)+\delta_{iju}+\delta_{klu})e^{-\int_t^u(\alpha_s^*(ij,kl)+\delta_{ijs}+\delta_{kls})ds}\right)}{E_t\left(e^{-\int_t^u(\alpha_s^*(ij,kl)+\delta_{ijs}+\delta_{kls})ds}\right)} \qquad (19)
\end{aligned}
$$

Recall that $\alpha_s^*(ij,kl) \geq \alpha_s(ij,kl)$. If these rates are deterministic, then $h_t^*(u;ij,kl) \geq h_t(u;ij,kl)$, implies that the Duffie, *et al* (2007) model will yield a smaller double-survival probability than the hierarchical intensity model. When the rates are stochastic, a local analysis is possible. Let u approaches t to give rise to $h_t^*(t;ij,kl) \geq h_t(t;ij,kl)$, which implies that double survival locally is more likely under the hierarchical intensity model.

Note that $\alpha_t(ij,kl)+\delta_{ijt}+\delta_{klt}$ is the kill rate discussed earlier except that we have adjusted for exiting due to non-default reasons. The comparison of the hazard rate and the kill rate can be likened to the the fixed-term forward rate versus the instantaneous forward rate in the term structure literature. When the kill rate is constant, it is the same as the hazard rate. In our case where rates are stochastic, the kill rate is more helpful in assessing double survival. Although we are unable to ascertain the directional relationship between two hazard rates, we are able to determine the relationship between the double survival probabilities using the kill rates. Since $\alpha_t^*(ij,kl) \geq \alpha_t(ij,kl)$ almost surely, the double survival probability, censored or not, is always higher under the hierarchical intensity model that that under the Duffie, *et al* (2007) model. Obviously, this conclusion applies to the case where two obligors are from different groups.

It is worth noting that the complement of double survival is not double default. A defining characteristic of the Duffie, *et al* (2007) model or other standard default intensity models is that its concurrent double-default rate always equals zero regardless of how high the default intensities of the two processes are. It is evident the concurrent double default occurs with a positive rate under the hierarchical intensity model via the common or a group-specific shock. Interestingly, both double default and double survival are more likely under the hierarchical intensity model.

3 Estimation procedure

Let θ denote all parameters governing the X_t, Y_{it}, and Z_{ijt} for $i = 1, \cdots, K$ and $j = 1, \cdots, n_i$. The parameters governing F, G, H, R, P and Q functions are denoted by φ. The data set related to X_t, Y_{it}, and Z_{ijt} from time 1 to time T is denoted by D_T. Let I_t be a matrix with rows representing different groups and the column dimension equals the maximum number of obligors in groups. This matrix corresponds the status of all obligors. Prior to default or other forms of exit for an obligor, its corresponding entry in I_t is assigned a value of 0. If exiting by default at time t, the assigned value is switched to 1 and will remain fixed thereafter. If exiting due to other reasons, the assigned value is 2 and remains fixed from then on. In order to reflect the entry times of different obligors into the sample, we use V, a matrix matching the dimension of I_t, to capture these entry times.

The log-likelihood function can be decomposed into two parts:

$$\mathcal{L}(\theta, \varphi; D_T, I_T, V) = \mathcal{L}(\varphi; D_T, I_T, V) + \mathcal{L}(\theta; D_T) \tag{20}$$

Note that $\mathcal{L}(\varphi; D_T, I_T, V)$ captures the default likelihood conditional on the state variables D_T, because the processes governing defaults and other forms of exit are assumed under the model to be directed by the values of X_t, Y_{it}, and Z_{ijt}, but not the parameters governing their dynamics. The second term $\mathcal{L}(\theta; D_T)$ is simply the log-likelihood function associated with the state variables D_T whose dynamics under the model are not affected by obligor defaults.

In contrast to the Duffie, *et al* (2007) approach, we must express the default likelihood function conditional on state variables cross-sectionally at one time point, and then aggregate them over time. This is needed because of the hierarchical default structure. Specifically,

$$\mathcal{L}(\varphi; D_T, I_T, V) = \sum_{t=2}^{T} \ln\left(A_t(\varphi; D_t, I_t, V)\right) \tag{21}$$

where

$$\begin{aligned} & A_t(\varphi; D_t, I_t, V) \\ = \; & e^{-\lambda_{c(t-1)}\Delta t} \prod_{i=1}^{K} \left(e^{-\lambda_{i(t-1)}\Delta t} \prod_{j=1}^{n_i} B_{ijt} C_{ijt}^{(1)} + (1 - e^{-\lambda_{i(t-1)}\Delta t}) \prod_{j=1}^{n_i} B_{ijt} C_{ijt}^{(2)} \right) \\ & + (1 - e^{-\lambda_{c(t-1)}\Delta t}) \prod_{i=1}^{K} \left(e^{-\lambda_{i(t-1)}\Delta t} \prod_{j=1}^{n_i} B_{ijt} C_{ijt}^{(3)} + (1 - e^{-\lambda_{i(t-1)}\Delta t}) \prod_{j=1}^{n_i} B_{ijt} C_{ijt}^{(4)} \right) \end{aligned} \tag{22}$$

$$
\begin{aligned}
B_{ijt} &= 1_{\{V(i,j)>t-1\}} + 1_{\{V(i,j)\le t-1\}}[1_{\{I_{t-1}(i,j)\neq 0\}} + 1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)\neq 2\}}e^{-\delta_{ij(t-1)}\Delta t} \\
&\quad +1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)=2\}}(1-e^{-\delta_{ij(t-1)}\Delta t})] \\
C^{(1)}_{ijt} &= 1_{\{V(i,j)>t-1\}} + 1_{\{V(i,j)\le t-1\}}[1_{\{I_{t-1}(i,j)\neq 0\}} + 1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)\neq 1\}}e^{-\lambda_{ij(t-1)}\Delta t} \\
&\quad +1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)=1\}}(1-e^{-\lambda_{ij(t-1)}\Delta t})] \\
C^{(2)}_{ijt} &= 1_{\{V(i,j)>t-1\}} + 1_{\{V(i,j)\le t-1\}}\{1_{\{I_{t-1}(i,j)\neq 0\}} + 1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)\neq 1\}}(1-q_{ij(t-1)})e^{-\lambda_{ij(t-1)}\Delta t} \\
&\quad +1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)=1\}}[q_{ij(t-1)} + (1-e^{-\lambda_{ij(t-1)}\Delta t}) - q_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t})]\} \\
C^{(3)}_{ijt} &= 1_{\{V(i,j)>t-1\}} + 1_{\{V(i,j)\le t-1\}}\{1_{\{I_{t-1}(i,j)\neq 0\}} + 1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)\neq 1\}}(1-p_{ij(t-1)})e^{-\lambda_{ij(t-1)}\Delta t} \\
&\quad +1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)=1\}}[p_{ij(t-1)} + (1-e^{-\lambda_{ij(t-1)}\Delta t}) - p_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t})]\} \\
C^{(4)}_{ijt} &= 1_{\{V(i,j)>t-1\}} + 1_{\{V(i,j)\le t-1\}}\{1_{\{I_{t-1}(i,j)\neq 0\}} \\
&\quad +1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)\neq 1\}}(1-p_{ij(t-1)})(1-q_{ij(t-1)})e^{-\lambda_{ij(t-1)}\Delta t} \\
&\quad +1_{\{I_{t-1}(i,j)=0\}}1_{\{I_t(i,j)=1\}}[p_{ij(t-1)} + q_{ij(t-1)} + (1-e^{-\lambda_{ij(t-1)}\Delta t}) - p_{ij(t-1)}q_{ij(t-1)} \\
&\quad -p_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t}) - q_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t}) + p_{ij(t-1)}q_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t})]\}
\end{aligned}
$$

Note that Δt is the length of one period; for example, monthly frequency corresponds to $\Delta t = 1/12$. $V(i,j)$ is used to control for some obligors that entered the sample later than others. If the application is to track a portfolio over time without adding any new obligors in the process, $V(i,j)$ can be ignored because $1_{\{V(i,j)\le t-1\}} = 1$ and $1_{\{V(i,j)>t-1\}} = 0$.

A_t comprises two terms with the first one dealing with the event that the common shock did not occur over the time period $[t-1,t]$, which has the probability of $e^{-\lambda_{c(t-1)}\Delta t}$. The second term is for the event that the common shock did occur with a probability of $1-e^{-\lambda_{c(t-1)}\Delta t}$.

Conditional on the event of non-occurrence of the common shock, group-specific shocks can either occur or not. The first term inside the first term of A_t is for the non-occurrence of that group-specific shock whereas the second term is for the occurrence of that group-specific shock. Their respective probabilities follow the same principle as for the common event. Similarly, the two terms inside the second term of A_t deal with two possible scenarios for the group-specific shocks.

Under each of four possible combinations of common and group-specific events, we are able to figure out the appropriate probability of the joint default-exit pattern of all obligors in the sample. In the case of $B_{ijt}C^{(1)}_{ijt}$, neither the common nor group-specific shock occurred, the probability of no default or exit over $[t-1,t]$ for the (i,j)-th obligor is governed by its individual default and other exit intensities which equals $e^{-(\lambda_{ij(t-1)}+\delta_{ij(t-1)})\Delta t}$. Likewise, the probability for default or other exit follows from the specification of our intensity model.

$B_{ijt}C^{(2)}_{ijt}$ is the term specifically for the combination of no common shock but with a group-specific shock. The probability for an obligor not to default or exit for other reasons naturally becomes $(1-q_{ij(t-1)})e^{-(\lambda_{ij(t-1)}+\delta_{ij(t-1)})\Delta t}$ where the first item is the probability of no default even when a group-specific shock has already occurred. The other two items simply reflect the non-occurrence of the obligor-specific default or exit for other reasons. If a particular obligor defaulted, its probability must be the sum of its default probability due to the group-specific shock, i.e., $q_{ij(t-1)}$, and its own shock, i.e., $1-e^{-\lambda_{ij(t-1)\Delta t}}$, minus the probability of joint occurrence to avoid double counting. Conditional on the event that the common shock did not occur, a group-specific shock can either occur or not. For a particular default-exit pattern, the appropriate probability will thus be the sum of the two terms inside the first term of A_t.

The idea behind the term $B_{ijt}C_{ijt}^{(3)}$ is similar to that for $B_{ijt}C_{ijt}^{(2)}$. For the term $B_{ijt}C_{ijt}^{(4)}$, we consider the situation in which both the common and some group-specific shocks have occurred. Thus, observing a particular obligor which neither defaulted nor exited for other reasons over $[t-1,t]$, the probability must equal $(1-p_{ij(t-1)})(1-q_{ij(t-1)})e^{-(\lambda_{ij(t-1)}+\delta_{ij(t-1)})\Delta t}$. For a defaulted obligor, its probability will be the sum of three possible causes – common shock, group-specific shock or individual shock. Since all three shocks could have all occurred over $[t-1,t]$, we must adjust for double and triple counting, and thus have the result of $p_{ij(t-1)}+q_{ij(t-1)}+(1-e^{-\lambda_{ij(t-1)}\Delta t})-p_{ij(t-1)}q_{ij(t-1)}-p_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t})-q_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t})+p_{ij(t-1)}q_{ij(t-1)}(1-e^{-\lambda_{ij(t-1)}\Delta t})$. The remaining term in $B_{ijt}C_{ijt}^{(4)}$ is for the probability of exiting for other reasons, its expression is straightforward.

It is fairly easy to see the above result contains that of Duffie, *et al* (2007) as a special case. When the common and group-specific intensities are set to zero, A_t becomes $\prod_{i=1}^{K}\prod_{j=1}^{n_i}B_{ijt}C_{ijt}^{(1)}$. In the case of Duffie, *et al* (2007), one first multiply over time for each obligor and then over obligors. In our expression, we first multiply over obligors and then over time. Note that $(1-e^{-\lambda_{ij(t-1)}\Delta t})\cong\lambda_{ij(t-1)}\Delta t$ and $(1-e^{-\delta_{ij(t-1)}\Delta t})\cong\delta_{ij(t-1)}\Delta t$ when Δt is small. Using $(1-e^{-\lambda_{ij(t-1)}\Delta t})$ is more accurate because the default intensity model are likely to be applied on data that are of monthly, quarterly or yearly frequency. In this more accurate form, the only assumption is that the covariates do not change over the time period $[t-1,t]$.

The intensity function for default and that for other forms of exit are expected to be governed by different parameters. When there are no parametric restrictions linking two sets of parameters together, the following decomposed likelihood function can be very useful in numerical optimization:

$$
\begin{aligned}
& A_t(\varphi; D_t, I_t, V) \\
= \; & \left(\prod_{i=1}^{K}\prod_{j=1}^{n_i} B_{ijt}\right)\left\{e^{-\lambda_{c(t-1)}\Delta t}\prod_{i=1}^{K}\left(e^{-\lambda_{i(t-1)}\Delta t}\prod_{j=1}^{n_i} C_{ijt}^{(1)} + (1-e^{-\lambda_{i(t-1)}\Delta t})\prod_{j=1}^{n_i} C_{ijt}^{(2)}\right)\right. \\
& \left. +(1-e^{-\lambda_{c(t-1)}\Delta t})\prod_{i=1}^{K}\left(e^{-\lambda_{i(t-1)}\Delta t}\prod_{j=1}^{n_i} C_{ijt}^{(3)} + (1-e^{-\lambda_{i(t-1)}\Delta t})\prod_{j=1}^{n_i} C_{ijt}^{(4)}\right)\right\} \qquad (23)
\end{aligned}
$$

If we divide the parameter set φ into φ_B and φ_C, the term $\prod_{i=1}^{K}\prod_{j=1}^{n_i} B_{ijt}$ only contains φ_B, and the remaining terms are functions of φ_C. This decomposition will reduce the effective dimension in numerically maximizing the log-likelihood function, because φ_B and φ_C can be optimized separately.

The likelihood associated with the state variables, i.e, $\mathcal{L}(\theta; D_T)$ depends on the dynamic models adopted for the state variables. Duffie, *et al* (2007) employed the vector autogression model for the state variables. As discussed earlier, the state variable dynamics do not affect the default process estimation. Unless one is interested in default prediction beyond one period ahead, it is unnecessary to specify the state variable dynamics. Since the objective of this paper is to introduce a new approach to default modeling, we will focus on the critical difference in the default structures and avoid the complication caused by the choice of state variable dynamics.

4 Empirical analysis

4.1 Data

The data set used to compare the hierarchical and standard intensity models is a sample of US firms over the period of 1991-2008. The data frequency is monthly with the accounting data are from the Compustat quarterly and annual database. The reported figures are lagged for three months to reflect the fact that the accounting figures are typically released a couple months after the period covered. The stock market data (stock prices, shares outstanding, and market index returns) are from the CRSP monthly file. For the default/bankruptcy data, we take from the CRSP file the de-listing information and couple them with the default data obtained from the Bloomberg CACS function. Following Shumway (2001), the firms that filed for any type of bankruptcy within 5 years of de-listing are considered bankrupt. There are altogether 872 bankruptcies and defaults in the sample. Firms may exit the sample due to reasons other than default or bankruptcy, and they are lumped together as other exits. The firms with less than one year's data over the sample period are removed. The interest rates are taken from the US Federal Reserve. While Duffie, *et al* (2007) restricted to Moody's "Industrial" category, we include all firms in the data base. After accounting for missing values, there are 14,401 firms with 1,317,217 firm-month observations. Among them, 2,844 companies are financial (SIC code between 6000 and 6999) with 249,382 firm-month observations.

We follow Duffie, *et al* (2007) to use the four covariates: trailing one-year S&P 500 index return, three-month Treasury bill rate, firm's trailing one-year return, and firm's distance-to-default in accordance with Merton's model. Merton's model is typically implemented with a KMV assumption on the debt maturity and size. As described in Crosbie and Bohn (2002), the KMV assumption sets the firm's debt maturity to one year and size to the sum of the short-term debt and 50% of the long-term debt. In the literature, this KMV assumption is typically adopted only for non-financial firms because a financial company such as AIG has significantly higher liabilities that are classified as neither short-term nor long-term debt. Since we have included financial firms in our sample, we need to devise a way to handle other liabilities. Specifically, we assign a firm specific fraction of other liabilities to be added to the KMV liabilities. We then apply the maximum likelihood estimation method developed by Duan (1994,2000) to estimate this unknown parameter along with the asset return's unknown mean and standard deviation. We allow such a liability adjustment factor for both financial and non-financial firms.

In a recent article, Wang (2009) showed in the context of the Duffie, *et al* (2007) model that using the cross-sectional average of firm's distance-to-default can significantly improve the model's performance. When the average is low, individual firms are more likely to default. Since distance-to-default can be understood as an inverse of a firm's standardized leverage, Wang's finding suggests that a higher economy-wide leverage leads to higher individual defaults above and beyond the impact of their own leverages. Because our use of the maximum likelihood estimation technique gives rise to a debt value adjustment for financial firms, we can zero in on three market-wide averages of distance-to-default – financial, non-financial and overall. We will show later that the average distance-to-default of the financial firms plays a significant role in determining the arrival of common shock.

4.2 Empirical findings

The focus of our empirical analysis is on the default dynamics. We do not attempt to prescribe a time series specification for the covariates, for which Duffie, *et al* (2007) have already done. As shown earlier, the likelihood function is decomposable in a way that the default intensity structure can be estimated by treating all covariates as exogenous variables. Furthermore, the likelihood function can be decomposed in a way that the default dynamics can be estimated without knowing the exact nature of exits that occur for reasons other than defaults. These two facts when taken together allow us to just focus on the default dynamics.

All default intensity functions and individual default probability functions upon experiencing the common or group-specific shock are assumed to be exponential-linear in covariates. In Duffie, *et al* (2007), all intensity functions are also assumed to be exponential-linear in covariates.

The results for the standard intensity model using the same set of covariates as in Duffie, *et al* (2007) are reported under DSW in Table 1. Our results confirms their finding that these variables are all highly significant and their signs are also in agreement. When the common shock is introduced, we let its intensity function to depend on the average distance-to-default which is represented in three different ways – all firms, financial firms only, and non-financial firms only. The individual default probability function corresponding to the common shock occurs, i.e., p function, is specified as a function of individual distance-to-default. The results reported in Table 1 under HIM show that the parameters define the common shock are highly significant. The likelihood ratio test cannot be straightforwardly applied, however, because there are two unidentified nuisance parameters in p function under the null hypothesis of no common shock. The difference in the likelihood values between HIM and DSW is so large that any adjustment to the critical value will still be significant.

The results for HIM show that the average distance-to-default of the financial firms works just as well as the average computed using all firms, and its sign turns out as expected. But the same cannot be said about the average obtained from the nonfinancial firms. This suggests that a higher average leverage level in the financial sector is strongly indicative of the overall economy's stress level. It increases the chance of a common shock, which in turn causes clustered defaults. The variable used in p function is the firm's own distant-to-default. This variable is highly significant, meaning that individual firm's response to a common shock depends on its own leverage. The higher the leverage (a lower distance-to-default), the more likely the firm defaults upon experiencing a common shock.

The default signature plots discussed earlier are based on the parameters reported in Table 1. For the hierarchical intensity model, the plot is based on the case using the financial firms' average distance-to-default. The structural difference between two modeling approach is evident in Figures 2a and 2b, when periods are grouped according to the financial firms' average distance-to-default.

In the study of rating/default prediction models, the cumulative accuracy plot (GAP), also known as the power curve, is often employed. GAP is only concerned with rankings and totally ignores the degrees of riskiness. It lines up the obligors ranging from the most risky to least risky. Set a percentage and take a group of most risky obligors corresponding to this chosen percentage. Then, identify defaulted obligors among them, and compute the percentage represented by these defaulted ones in relation to the overall sample of defaulted obligors. GAP is a plot that relates the percentage among defaulted obligors to the percentage among the ranked obligors. GAP is a plot of the two percentages. For a large sample, the GAP for a perfect risk ranking model will quickly rises to one and levels at one. A completely uninformative risk ranking model will have the GAP plot being the line with a slope equal to one. Figure 3 presents the GAP plots corresponding to the standard and hierarchical intensity models. It is clear that the two models have highly close risk ranking performance even though their predicted default distributions differ. Therefore, GAP is uninformative in distinguishing these two models.

The two models are actually quite different if we compare their time-varying predicted default distributions. The Kullback-Leibler (KL) distance is a popular way of measuring the distance between two distribution functions at time t. We compute the KL distance using the following formula:

$$KL = \sum_{i=0}^{\infty} \ln \frac{p_t^{HIM}(i)}{p_t^{DSW}(i)} p_t^{HIM}(i). \tag{24}$$

The KL distance is non-negative and its minimum value is zero which is attainable when two distributions are identical. However, the KL distance is not a typical distance measure because it is asymmetrical. We provide a time series plot of the KL distance in Figure 4. To get a sense on how the magnitude of the KL distance translates, we note that the KL distance of a normal distribution with mean 0 and standard deviation σ to the standard normal distribution equals $\ln(\sigma) + \frac{1-\sigma^2}{2\sigma^2}$. When $\sigma = 0.75$, the KL distance is 10%. But for $\sigma = 1.25$, the KL distance becomes 4.3%. The plot indicates that the two predicted default distributions are sometimes close, but at other times they can be quite different. The difference peaks in the beginning of 2001. The KL distance is plotted alongside the average financial firm's distance-to-default. Their relationship appears to be nonlinear, reflecting the fact that we have previously learned from Figures 2a and 2b; that is, two predicted default distributions (averaged over periods with similar characteristics) differ more for either high and low values of the average distance-to-default.

5 Conclusion

In this article, a hierarchical intensity model is proposed to deal with clustered defaults. This model vis-a-vis the standard intensity model exhibits a feature that clustered defaults can be generated through a common or group-specific shocks. In this exploratory study, we have only implemented a version with a common shock and the covariate that drives the arrival of a common shock is the average distance-to-default for all financial firms. Other variables may also be informative which in turn widens the difference between the hierarchical and standard intensity models. Further exploration on potential explanatory variables may yield good results.

This new model offers an interesting theoretical feature.The multiple-default probability is proportional to the length of the measuring interval. In contrast, the standard intensity model will have such probability proportional to the time length to the power of the number of concurrent defaults. This feature should be of particular interest when one estimates the model with one data frequency, say monthly, but want to consider joint defaults over a week.

References

[1] Andersen, L., J. Sidenius and S. Basu, 2003, All Your Hedges in One Basket, *RISK* 16, November, 62-72.

[2] Arnsdorf, M. and I. Halperin, 2007, BSLP: Markovian Bivariate Spread-loss Model for Portfolio Credit Derivatives, unpublished manuscript, Quantitative Research J.P. Morgan.

[3] Azizpour, S. and K. Giesecke, 2008, Self-exciting Corporate Defaults: Contagion vs. Frailty, unpublished manuscript, Stanford University.

[4] Cont, R. and A. Minca, 2007, Reconstructing Portfolio Default Rates from CDO Tranche Spreads, unpublished manuscript, Columbia University.

[5] Crosbie, P., and J. Bohn, 2002, Modeling Default Risk, technical report, KMV LLC.

[6] Das, S.R., D. Duffie, N. Kapadia and L. Saita, 2007, Common Failings: How Corporate Defaults Are Correlated, *Journal of Finance* 62, 93-117.

[7] Duan, J.C., 1994, Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance* 4, 155-167.

[8] Duan, J.C., 2000, Correction: "Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance* 10, 461-462.

[9] Duffie, D., A. Eckner, G. Horel and L. Saita, 2008, Frailty Correlated Default, forthcoming *Journal of Finance.*

[10] Duffie, D., L. Saita and K. Wang, 2007, Multi-Period Corporate Default Prediction With Stochastic Covariatesricing Model, *Journal of Financial Economics* 83, 635-665.

[11] Giesecke, K. and B. Kim, 2007, Estimating Tranche Spreads by Loss Process Simulation, In Henderson, S.G., Biller, B., Hsieh, M.-H., Shortle, J., Tew, J.D., Barton, R.R. (Eds), Proceedings of the 2007 Winter Simulation Conference, IEEE Press.

[12] Li, D., 2000, On Default Correlation: a Copula Function Approach, *Journal of Fixed Income* 9, 43-54.

[13] Longstaff, F. and A. Rajan, 2007, An Empirical Analysis of Collateralized Debt Obligations, unpublished manuscript, University of California, Los Angeles.

[14] Peng, X. and S. Kou, 2009, Default Clustering and Valuation of Collateralized Debt Oligations, unpublished manuscript, Columbia University.

[15] Shumway, T., 2001, Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *Journal of Business* 74, 101-124.

[16] Wang, T., 2009, Determinants of Corporate Bankruptcy Using an Intensity Model, unpublished manuscript, National University of Singapore.

6 Appendix

6.1 Predicted default distributions

A. **The standard intensity model**

1. At time $t-\Delta t$, define the distribution for the cumulative number of defaults, after considering up to and including the (i,j)-th obligor, by $p_t^{ij}(k)$ for $k = 0, 1, \cdots$. Obviously, $p_t^{11}(0) = e^{-\lambda_{11(t-\Delta t)}\Delta t}$, $p_t^{11}(1) = 1 - e^{-\lambda_{11(t-\Delta t)}\Delta t}$, and $p_t^{11}(k) = 0$ for $k = 2, \cdots$.

2. Perform convolution of $p_t^{ij}(k)$ with the next obligor in the pool, i.e., the $(i, j+1)$-th obligor. Its default distribution is $q_t^{i(j+1)}(0) = e^{-\lambda_{i(j+1)(t-\Delta t)}\Delta t}$ and $q_t^{i(j+1)}(1) = 1 - e^{-\lambda_{i(j+1)(t-\Delta t)}\Delta t}$. Note that instead of tracking the cumulative number of defaults that have negligible probabilities, one can truncate the cumulative default distribution at some k^* beyond which the probability is, say, lower than 10^{-8}. The truncated default distribution needs to be normalized so as to to sum up to 1. The convolution calculation is illustrated in the following table:

$k=$	0	1	2	$\cdots$	k^*
	$p_t^{ij}(0)$	$p_t^{ij}(1)$	$p_t^{ij}(2)$	$\cdots$	$p_t^{ij}(k^*)$
	$q_t^{i(j+1)}(0)$	$q_t^{i(j+1)}(1)$	0	$\cdots$	0
Row 1	$q_t^{i(j+1)}(0)p_t^{ij}(0)$	$q_t^{i(j+1)}(0)p_t^{ij}(1)$	$q_t^{i(j+1)}(0)p_t^{ij}(2)$	$\cdots$	$q_t^{i(j+1)}(0)p_t^{ij}(k^*)$
Row 2	0	$q_t^{i(j+1)}(1)p_t^{ij}(0)$	$q_t^{i(j+1)}(1)p_t^{ij}(1)$	$\cdots$	$q_t^{i(j+1)}(1)p_t^{ij}(k^*-1)$

   The new convoluted cumulative default distribution, i.e., $p_t^{i(j+1)}(k)$, is the sum across Rows 1 and 2 in the above table.

3. Repeat the above convolution calculation for all obligors remaining in the pool at time $t - \Delta t$.

4. Denote the cumulative default distribution for the entire pool at time t by $\hat{p}_t(k)$. Average the adjusted default distributions over time to obtain $\hat{P}(k) = \frac{1}{N}\sum_{j=1}^{N}\hat{p}_{j\Delta t}(k)$ for $k = 0, 1, 2, \cdots$ where N is the number of periods of length Δt. One can also restrict the average to a subset of time periods defined by the values of some covariate.

B. **The hierarchical intensity model**

1. At time $t - \Delta t$, assume that the common shock has occurred and perform the convolution calculations similar to that described in the preceding subsection, but use the default probability $p_{ij(t-\Delta t)}$ in the hierarchical intensity model. Repeat for all obligors remaining in the pool at time $t - \Delta t$, and denote the conditional cumulative default distribution by $a_t(0), a_t(1), a_t(2), \cdots$. Then, factoring in the fact that this is a conditional distribution and zero default can also occur when

there is no common shock, the cumulative default distribution due to the common shock thus becomes

$$\begin{aligned}\hat{p}_{ct}(0) &= e^{-\lambda_{c(t-\Delta t)}\Delta t} + (1 - e^{-\lambda_{c(t-\Delta t)}\Delta t})a_t(0)\\ \hat{p}_{ct}(k) &= (1 - e^{-\lambda_{c(t-\Delta t)}\Delta t})a_t(k) \quad \text{for } k = 1, 2, \cdots .\end{aligned}$$

2. Assume that the group-specific shock has occurred and perform similar convolution calculations using the default probability $q_{ij(t-\Delta t)}$. Repeat for all obligors remaining in the group at time $t - \Delta t$, and denote the conditional cumulative default distribution by $b_{it}(0), b_{it}(1), b_{it}(2), \cdots$. Then, factoring in the fact that this is a conditional distribution and zero default can also occur when there is no group-specific shock, the default distribution due to the group-specific shock becomes

$$\begin{aligned}\hat{p}_{it}(0) &= e^{-\lambda_{i(t-\Delta t)}\Delta t} + (1 - e^{-\lambda_{i(t-\Delta t)}\Delta t})b_{it}(0)\\ \hat{p}_{it}(k) &= (1 - e^{-\lambda_{i(t-\Delta t)}\Delta t})b_{it}(k) \quad \text{for } k = 1, 2, \cdots .\end{aligned}$$

3. Perform the convolution calculation of $\hat{p}_{ct}(k)$ with $\hat{p}_{1t}(k)$ to yield the default distribution for the sum of the common and group 1-specific shocks. Note that $\hat{p}_{1t}(k)$ may take non-zero value for $k \geq 2$, this convolution calculation needs more than two rows. Suppose $\hat{p}_{ct}(k)$ with $\hat{p}_{1t}(k)$ are truncated at k^* and k^{**}, respectively, beyond which the probabilities are less than 10^{-8}. We need to construct a table similar to that in the preceding subsection, but with k^{**} rows. After group 1, continue the convolution to group 2 with the default distribution of $\hat{p}_{2t}(k)$. Repeat convolutions until all groups are exhausted.
4. Perform convolution of the default distribution for the cumulative sum of common and group-specific shocks with the individual default distributions. This can be done one obligor at a time until all are exhausted. The individual default probabilities are in the form of $1 - e^{-\lambda_{ij(t-\Delta t)}\Delta t}$ which only needs two rows in the convolution calculation. Denote the final default distribution, due to three types of shocks, by $\hat{p}_t(0), \hat{p}_t(1), \hat{p}_t(2), \cdots$.
5. Average the adjusted default distributions over time to obtain $\hat{P}(k) = \frac{1}{N}\sum_{j=1}^{N}\hat{p}_{j\Delta t}(k)$ for $k = 0, 1, 2, \cdots$ where N is the number of periods of length Δt. Again, one can restrict the average to a subset of time periods defined by the values of some covariate.

6.2 Proof of Proposition 1

First, consider

$$\begin{aligned}& E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right)\\ =\; & 1 + E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right) - E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}\right) - E_0\left(1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right) \qquad (25)\end{aligned}$$

The second term in the above expression can be further developed into

$$\begin{aligned}
& E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right) \\
= & E_0\left(1_{\{\phi_{ij}\leq\phi_{kl}\}}1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right) + E_0\left(1_{\{\phi_{ij}>\phi_{kl}\}}1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right) \\
= & E_0\left[1_{\{\phi_{ij}\leq\phi_{kl}\}}E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\middle|\,\phi_{ij},\phi_{kl}\right)\right] \\
& +E_0\left[1_{\{\phi_{ij}>\phi_{kl}\}}E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\middle|\,\phi_{ij},\phi_{kl}\right)\right] \\
= & E_0\left[1_{\{\phi_{ij}\leq\phi_{kl}\}}E_0\left(e^{-\int_0^{t\wedge\phi_{ij}}\alpha_s(ij,kl)ds-\int_{t\wedge\phi_{ij}}^{t\wedge\phi_{kl}}(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})ds}\right)\right] \\
& +E_0\left[1_{\{\phi_{ij}>\phi_{kl}\}}E_0\left(e^{-\int_0^{t\wedge\phi_{kl}}\alpha_s(ij,kl)ds-\int_{t\wedge\phi_{kl}}^{t\wedge\phi_{ij}}(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij})ds}\right)\right] \\
\geq & E_0\left[1_{\{\phi_{ij}\leq\phi_{kl}\}}E_0\left(e^{-\int_0^{t\wedge\phi_{ij}}\alpha^*_s(ij,kl)ds-\int_{t\wedge\phi_{ij}}^{t\wedge\phi_{kl}}(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})ds}\right)\right] \\
& +E_0\left[1_{\{\phi_{ij}>\phi_{kl}\}}E_0\left(e^{-\int_0^{t\wedge\phi_{kl}}\alpha^*_s(ij,kl)ds-\int_{t\wedge\phi_{kl}}^{t\wedge\phi_{ij}}(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij})ds}\right)\right] \\
= & E_0\left(1_{\{\tau^*_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau^*_{kl}>t\wedge\phi_{kl}\}}\right) \qquad (26)
\end{aligned}$$

The inequality is due to the fact that $\alpha^*_s(ij,kl)\geq\alpha_s(ij,kl)$.

Therefore,

$$E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right)\geq E_0\left(1_{\{\tau^*_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau^*_{kl}>t\wedge\phi_{kl}\}}\right)$$

because all terms involving only τ_{ij} (or τ_{kl}) in the right-hand side of equation (25) can be substituted with τ^*_{ij} (or τ^*_{kl}), a result due to our construction where an obligor individually behaves the same way under two modelling approaches.

Recall the definition for default correlation:

$$\begin{aligned}
& Corr_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}},1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right) \\
= & \frac{E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right)-E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}\right)E_0\left(1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right)}{\sqrt{\left[E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}\right)-\left(E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}\right)\right)^2\right]\left[E_0\left(1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right)-\left(E_0\left(1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right)\right)^2\right]}}
\end{aligned}$$

Apart from $E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}1_{\{\tau_{kl}\leq t\wedge\phi_{kl}\}}\right)$, all other terms are the same when τ_{ij} (or τ_{kl}) is replaced with τ^*_{ij} (or τ^*_{kl}). Thus, the directional relation follows.

6.3 Formulas for the components in the censored double default probability and default correlation under the constant parameter assumption

Impose constant parameters and compute the components needed in equation (25). Note that $E_0\left(1_{\{\tau_{ij}\leq t\wedge\phi_{ij}\}}\right) = 1 - E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}\right)$. First,

$$\begin{aligned}
& E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}\right) \\
= \; & E_0\left(e^{-(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij})(t\wedge\phi_{ij})}\right) \\
= \; & \frac{\delta_{ij}}{p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}+\delta_{ij}} + \frac{p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}}{p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}+\delta_{ij}} e^{-(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}+\delta_{ij})t}
\end{aligned}$$

Next and by a similar argument,

$$\begin{aligned}
& E_0\left(1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right) \\
= \; & \frac{\delta_{kl}}{p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}+\delta_{kl}} + \frac{p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}}{p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}+\delta_{kl}} e^{-(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}+\delta_{kl})t}
\end{aligned}$$

Thus, the remaining term to be computed is $E_0\left(1_{\{\tau_{ij}>t\wedge\phi_{ij}\}}1_{\{\tau_{kl}>t\wedge\phi_{kl}\}}\right)$, which in turn comprises the two terms as shown in equation (26).

$$\begin{aligned}
& E_0\left[1_{\{\phi_{ij}\leq\phi_{kl}\}} E_0\left(e^{-\int_0^{t\wedge\phi_{ij}}\alpha(ij,kl)ds-\int_{t\wedge\phi_{ij}}^{t\wedge\phi_{kl}}(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})ds}\right)\right] \\
= \; & E_0\left(1_{\{\phi_{ij}\leq\phi_{kl}\}} e^{-\alpha(ij,kl)(t\wedge\phi_{ij})-(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})(t\wedge\phi_{kl}-t\wedge\phi_{ij})}\right) \\
= \; & \int_0^t\int_0^{s_2}\delta_{ij}\delta_{kl}e^{-[(\alpha(ij,kl)+\delta_{ij}-p_{kl}\lambda_c-q_{kl}\lambda_k-\lambda_{kl})s_1+(\delta_{kl}+p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})s_2]}ds_1ds_2 \\
& + \int_t^\infty\int_0^t\delta_{ij}\delta_{kl}e^{-[(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})t+(\alpha(ij,kl)+\delta_{ij}-p_{kl}\lambda_c-q_{kl}\lambda_k-\lambda_{kl})s_1+\delta_{kl}s_2]}ds_1ds_2 \\
& + \int_t^\infty\int_t^{s_2}\delta_{ij}\delta_{kl}e^{-[\alpha(ij,kl)t+\delta_{ij}s_1+\delta_{kl}s_2]}ds_1ds_2 \\
= \; & \frac{\delta_{ij}\delta_{kl}}{(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}+\delta_{kl})} \\
& + \frac{\delta_{ij}(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})}{(\alpha(ij,kl)+\delta_{ij}-p_{kl}\lambda_c-q_{kl}\lambda_k-\lambda_{kl})(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}+\delta_{kl})}e^{-(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl}+\delta_{kl})t} \\
& + \frac{\delta_{ij}\left[\alpha(ij,kl)(\alpha(ij,kl)+\delta_{ij})-(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})(p_{kl}\lambda_c+q_{kl}\lambda_k+\lambda_{kl})\right]}{(\delta_{ij}+\delta_{kl})(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})(\alpha(ij,kl)+\delta_{ij}-p_{kl}\lambda_c-q_{kl}\lambda_k-\lambda_{kl})}e^{-(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})t}
\end{aligned}$$

By a similar argument, we have

$$
\begin{aligned}
& E_0\left[1_{\{\phi_{ij}>\phi_{kl}\}} E_0\left(e^{-\int_0^{t\wedge\phi_{kl}} \alpha(ij,kl)ds-\int_{t\wedge\phi_{kl}}^{t\wedge\phi_{ij}}(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij})ds}\right)\right] \\
= \; & \frac{\delta_{ij}\delta_{kl}}{(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}+\delta_{ij})} \\
& + \frac{\delta_{kl}(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij})}{(\alpha(ij,kl)+\delta_{kl}-p_{ij}\lambda_c-q_{ij}\lambda_i-\lambda_{ij})(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}+\delta_{ij})} e^{-(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij}+\delta_{ij})t} \\
& + \frac{\delta_{kl}\left[\alpha(ij,kl)(\alpha(ij,kl)+\delta_{kl})-(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})(p_{ij}\lambda_c+q_{ij}\lambda_i+\lambda_{ij})\right]}{(\delta_{ij}+\delta_{kl})(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})(\alpha(ij,kl)+\delta_{kl}-p_{ij}\lambda_c-q_{ij}\lambda_i-\lambda_{ij})} e^{-(\alpha(ij,kl)+\delta_{ij}+\delta_{kl})t}
\end{aligned}
$$

Similarly, we can compute $E_0\left(1_{\{\tau^*_{ij}\le t\wedge\phi_{ij}\}}1_{\{\tau^*_{kl}\le t\wedge\phi_{kl}\}}\right)$ by noting that the only term will be affected by moving from τ_{ij} and τ_{kl} to τ^*_{ij} and τ^*_{kl} is $\alpha(ij,kl)$. We can simply replace it with $\alpha^*(ij,kl)$. Note that the default correlation only requires the terms that have been derived.

Table 1

The maximum likelihood estimation results for the Duffie, *et al* (2007) model (DSW) and the hierarchical intensity model (HIM). HIM is implemented with using three different measures of average distance-to-default. The sample covers the period of January 1991 to December 2008, and consists of 14,401 firms with 1,317,217 firm-month observations. There are 2,844 financial firms with 249,382 firm-month observations. The total number of defaults/bankruptcies is 872.

		DSW	**HIM**		
Common Shock Intensity Function	Intercept		17.7945** (8.6799)	5.4012*** (1.6272)	16.6699** (8.0790)
	Average DTD all		-3.8191* (2.0597)		
	Average DTD financial			-0.6188** (0.2766)	
	Average DTD nonfinancial				-3.9382* (2.1505)
***p* Function**	Intercept		-6.3245*** (0.2049)	-7.6097*** (0.1988)	-6.2811*** (0.2006)
	DTD		0.949*** (0.2371)	-0.4466*** (0.0473)	0.9382*** (0.2352)
Firm-specific Shock Intensity Function	Intercept	-4.9942*** (0.0794)	-0.0381 (0.0236)	-6.3434*** (0.2072)	-0.0393* (0.0235)
	Trailing 1-Year SP500 Return	0.7727*** (0.2038)	-0.8087*** (0.0390)	0.9324*** (0.2380)	-0.8099*** (0.0386)
	3-Month Treasury Rate	-0.0505** (0.0199)	-4.542*** (0.2432)	-0.0375 (0.0235)	-4.4967*** (0.2389)
	DTD	-0.7815*** (0.0254)	-7.9697*** (0.1676)	-0.8074*** (0.0388)	-8.0156*** (0.1780)
	Trailing 1-Year Return	-3.0017*** (0.0824)	-0.4389*** (0.0452)	-4.5659*** (0.2515)	-0.4318*** (0.0461)
Log-likelihood		-5359.9	-5239.9	-5240.0	-5242.1

Note: * denotes significance at 10%, ** denotes significance at 5%, and *** denotes significance at 1%.

Figure 1

The observed default frequency over the entire sample period (January 1991 to December 2008) is used to check the predicted frequencies based on the Duffie, *et al* (2007) model (Firm specific) and the hierarchical intensity model implemented with the common and firm-specific shocks.

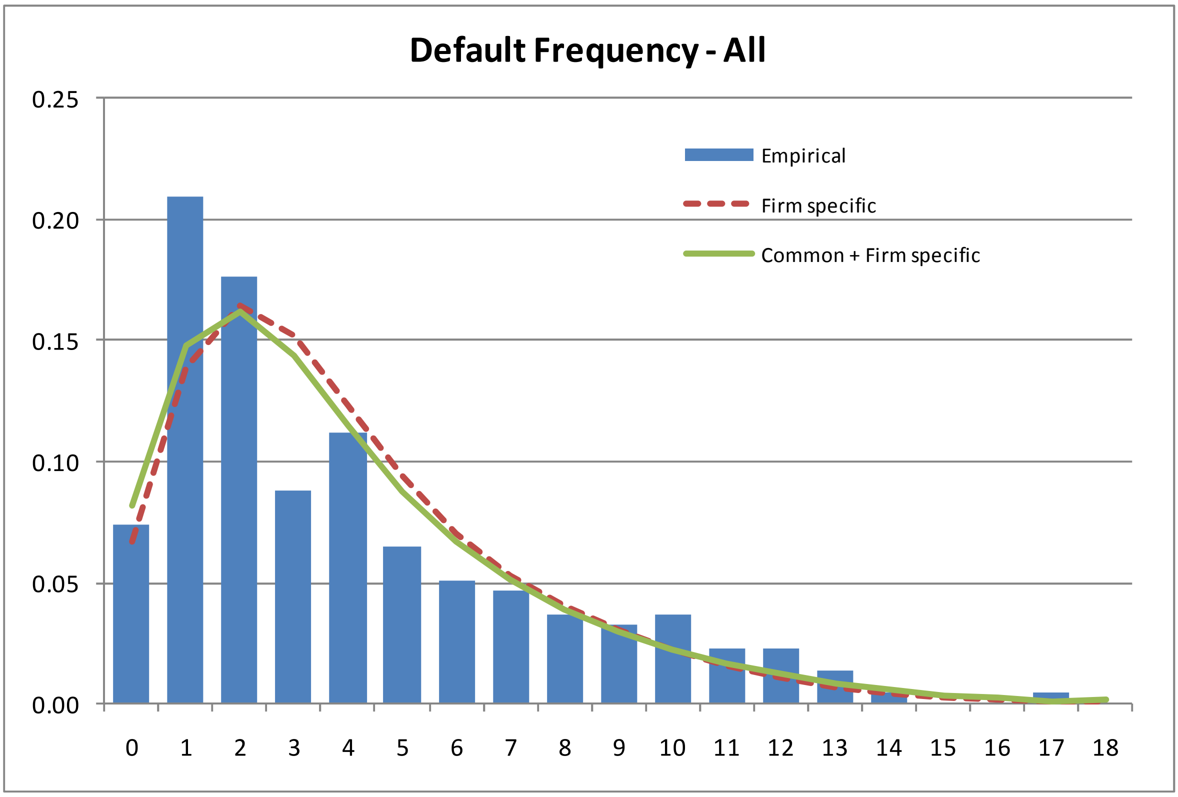

Figure 2a

The graph corresponds to the bottom third of the sample after dividing the whole sample (January 1991 to December 2008) into three groups using the average distance-to-default for the financial firms. This is used to check the predicted frequencies based on the Duffie, *et al* (2007) model (Firm specific) and the hierarchical intensity model implemented with the common and firm-specific shocks and the average distance-to-default for financial firms as reported in Table 1.

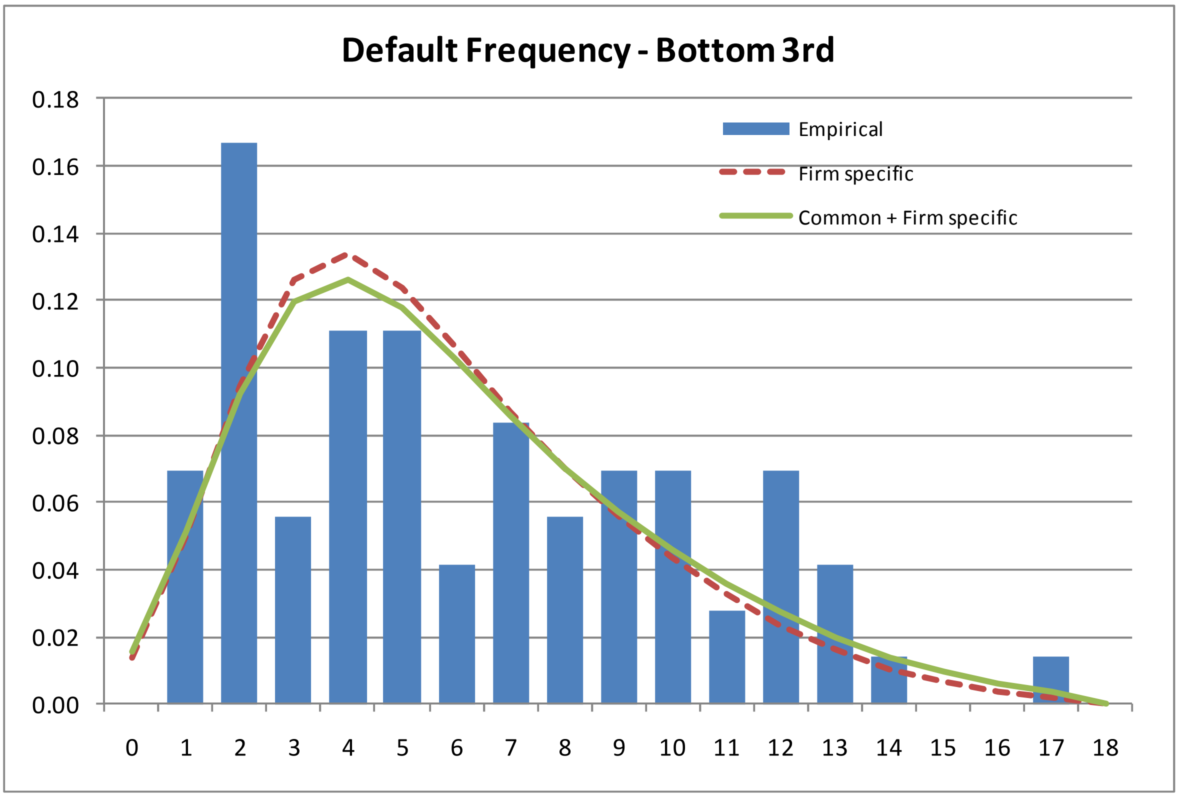

Figure 2b

The graph corresponds to the top third of the sample after dividing the whole sample (January 1991 to December 2008) into three groups using the average distance-to-default for the financial firms. This is used to check the predicted frequencies based on the Duffie, *et al* (2007) model (Firm specific) and the hierarchical intensity model implemented with the common and firm-specific shocks and the average distance-to-default for financial firms as reported in Table 1.

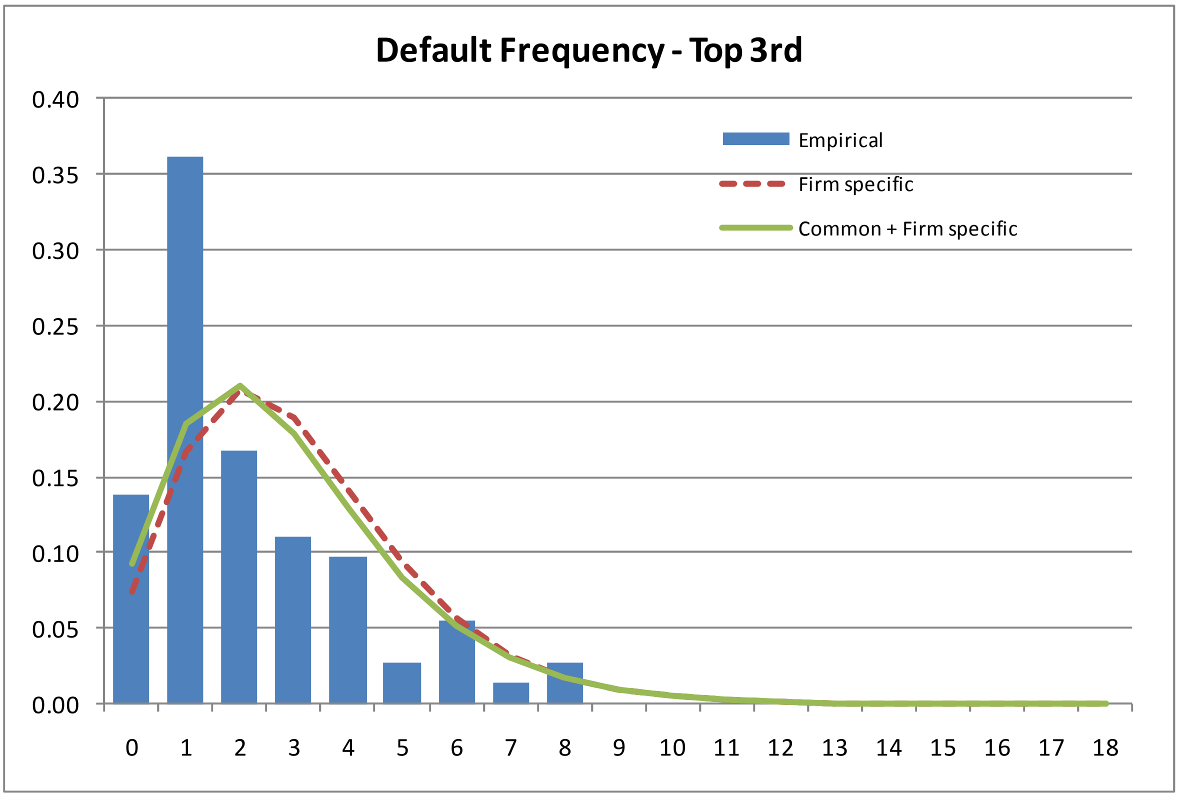

Figure 3

The cumulative accuracy plots (power curves) depict the accuracy of default predictions based solely on rank orders. The entire sample period (January 1991 to December 2008) is used. The two plots correspond to the Duffie, *et al* (2007) model and the hierarchical intensity model implemented with the common and firm-specific shocks and the average distance-to-default for financial firms as reported in Table 1.

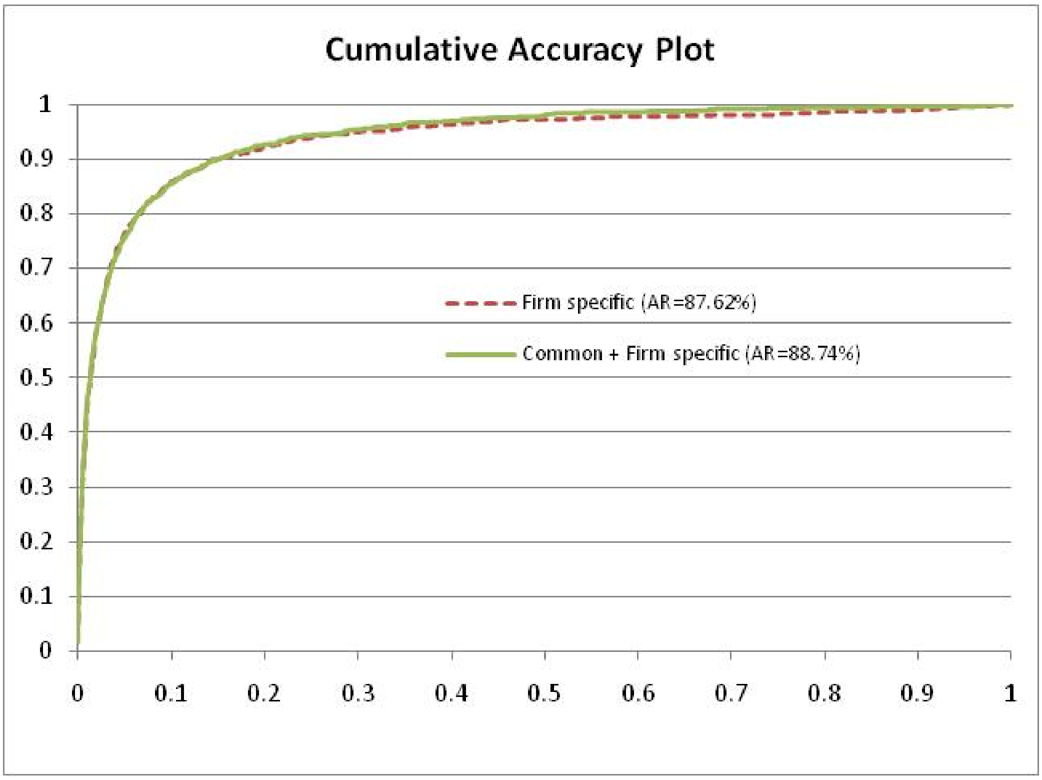

Figure 4

The Kullback-Leibler distance and the average distance-to-default for the financial firms are plotted over the sample period (January 1991 to December 2008). The Kullback-Leibler distance is between the predicted frequency distributions based on two models: the Duffie, *et al* (2007) model and the hierarchical intensity model implemented with the common and firm-specific shocks and the average distance-to-default for financial firms as reported in Table 1. It is computed using the hierarchical intensity model as the base distribution.