Online Appendix to Exploring the Sources of Default Clustering

S. Azizpour, K. Giesecke, G. Schwenkler*

Journal of Financial Economics

*Azizpour is at Apollo Global Management, Giesecke is at the Department of Management Science & Engineering, Stanford University, Schwenkler is at the Department of Finance, Boston University Questrom School of Business. Schwenkler is corresponding author: Phone (617) 358-6266, email: gas@bu.edu, web: http://www.gustavo-schwenkler.com.

Appendix A. Maximum likelihood estimation

Due to the presence of the frailty factor Z, the problem of estimating the parameter $\theta = (a, b, \kappa, \sigma, k, z)$ of the reduced-form model (1) is not standard unless $z = \sigma = 0$. We implement a variant of the filtered likelihood estimation method developed by Giesecke and Schwenkler (2018) for point process models with incomplete data. This appendix provides some details.

A.1. The likelihood function

Let D_t denote the data available at time t. These include the default data $\{T_n, U_n : n \leq N_t\}$ and the (interpolated) covariate data $\{X_s : s \leq t\}$. Let $[0, \tau]$ be the sample period. The likelihood function $\mathcal{L}_\tau(\theta)$ is the Radon-Nikodym density of the law of the data D_τ relative to its true distribution; see Section 3.1 of Giesecke and Schwenkler (2018) for a more precise statement. A maximum likelihood estimator (MLE) is a maximizer of the likelihood function. To compute the likelihood, let

$$M_\tau = \exp\left(\sum_{n\geq 1} \log(\lambda_{T_n-}) 1_{\{T_n \leq \tau\}} - \int_0^\tau (\lambda_s - 1)\mathrm{d}s\right), \tag{1}$$

and assume that $\mathbb{E}[1/M_\tau] = 1$. We define an equivalent measure $\mathbb{P}^*$ on the σ-field $\mathcal{F}_\tau$ with Radon-Nikodym density

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} = \frac{1}{M_\tau}.$$

Proposition 3.1 of Giesecke and Schwenkler (2018) states that

$$\mathcal{L}_\tau(\theta) \propto \mathbb{E}^*\left[M_\tau \mid D_\tau\right]. \tag{2}$$

The likelihood function (2) depends on the parameter θ because M_τ is a path functional of the intensity $\lambda = e^{a\cdot(1,X)} + Y + Z$. If $z = \sigma = 0$, the data set D_τ contains all relevant information to evaluate M_τ so the likelihood function simplifies to $\mathcal{L}_\tau(\theta) \propto M_\tau$. If $z, \sigma > 0$, then the conditional expectation (2) is not trivial because the data D_τ do not include any observations of the frailty Z. The conditional expectation (2) is taken with respect to the conditional $\mathbb{P}^*$-law of the frailty given the data D_τ. Girsanov's and Lévy's theorems imply that the Brownian motion W driving the frailty Z remains a Brownian motion under $\mathbb{P}^*$. As a result, Z follows the CIR process (2) under $\mathbb{P}^*$. Furthermore, Girsanov's and Watanabe's theorems imply that N is a standard Poisson process under $\mathbb{P}^*$. Therefore, N is independent of the frailty Z under $\mathbb{P}^*$. Also, Z is independent of X by assumption. Thus, the conditional

$\mathbb{P}^*$-law of Z given D_τ is governed by (k, z, σ).

A.2. Approximate likelihood function

There is no closed-form expression for the likelihood (2) if $\sigma > 0$ and $z > 0$. Giesecke and Schwenkler (2018) propose to approximate the likelihood using a rectangular quadrature method. We will use a slightly modified approximation based on a trapezoidal quadrature method that exploits the fact that our frailty Z follows a CIR process. This approximation does not interpolate the path of the frailty. It is more accurate and less susceptible to the implementation choice.

Since Z follows a CIR process under $\mathbb{P}^*$, a key result of Broadie and Kaya (2006) implies that for times $0 \leq t_1 < t_2 \leq \tau$ and points $z_1, z_2 \in \mathbb{R}_+$,

$$\begin{aligned}\phi(z_1, t_1; z_2, t_2) &= \mathbb{E}^*\left[\exp\left(-\int_{t_1}^{t_2} \lambda_u \mathrm{d}u\right) \,\middle|\, D_\tau, Z_{t_1} = z_1, Z_{t_2} = z_2\right] \\ &= \Phi(z_1, t_1; z_2, t_2) \exp\left(-\int_{t_1}^{t_2} \left(e^{a\cdot(1,X_u)} + Y_u\right) \mathrm{d}u\right), \qquad (3)\end{aligned}$$

where

$$\Phi(z_1, t_1; z_2, t_2) = \frac{I_q\left(\sqrt{z_1 z_2}\frac{4\zeta e^{-0.5\zeta\Delta}}{1-e^{-\zeta\Delta}}\right)}{I_q\left(\sqrt{z_1 z_2}\frac{4k e^{-0.5k\Delta}}{1-e^{-k\Delta}}\right)} \frac{\zeta e^{-0.5(\zeta-k)\Delta}(1-e^{-k\Delta})}{k(1-e^{-\zeta\Delta})} e^{(z_1+z_2)\left(\frac{k(1+e^{-k\Delta})}{1-e^{-k\Delta}} - \frac{\zeta(1+e^{-\zeta\Delta})}{1-e^{-\zeta\Delta}}\right)}$$

for $\Delta = t_2 - t_1$, $\zeta = \sqrt{k^2 + 2c^2}$, and I_q the modified Bessel function of the first kind of order $q = \frac{2kz}{c^2} - 1$. An application of the law of iterated expectations leads to a reformulation of the likelihood in terms of a product of terms as in (3).

Proposition A.1. *For $t \leq \tau$ and a function u on $\mathbb{R}_+$ such that $u(\lambda_t)$ is integrable,*

$$\mathbb{E}^*\left[u(\lambda_t) M_t \mid D_t\right] = e^t \mathbb{E}^*\left[u(\lambda_t)\phi(Z_{T_{N_\tau}}, T_{N_\tau}; Z_\tau, \tau) \prod_{n=1}^{N_\tau} \lambda_{T_n -}\phi(Z_{T_{n-1}}, T_{n-1}; Z_{T_n}, T_n) \,\middle|\, D_t\right].$$

Proof. Letting $\Pi_t = u(\lambda_t)\exp\left(\sum_{n=1}^{N_t} \log(\lambda_{T_n -})\right)$, we calculate

$$\begin{aligned}&\mathbb{E}^*\left[u(\lambda_t) M_t \mid D_t\right] \exp(-t) \\ &= \mathbb{E}^*\left[\Pi_t \exp\left(-\int_0^{T_{N_t}} \lambda_u \mathrm{d}u\right) \mathbb{E}^*\left[\exp\left(-\int_{T_{N_t}}^t \lambda_u \mathrm{d}u\right) \,\middle|\, D_t, (Z_u)_{u \leq T_{N_t}}, Z_t\right] \,\middle|\, D_t\right] \\ &= \mathbb{E}^*\left[\Pi_t \exp\left(-\int_0^{T_{N_t}} \lambda_u \mathrm{d}u\right) \phi(Z_{T_{N_t}}, T_{N_t}; Z_t, t) \,\middle|\, D_t\right]. \qquad (4)\end{aligned}$$

Line (4) uses the fact that, by Girsanov's theorem, the change of measure induced by M_τ does not affect the law of Z. Consequently, Z is a Markov process under $\mathbb{P}^*$. We iterate to complete the proof. $\square$

Proposition A.1 allows us to approximate the likelihood (2) using a trapezoidal quadrature rule that does not interpolate the path of the frailty Z. Algorithm A.2 summarizes the numerical scheme for terms of the form $\mathbb{E}^*\left[u(\lambda_\tau)M_\tau \mid D_\tau\right]$; the likelihood is given by $u \equiv 1$. Convergence of the scheme as the discretization becomes finer follows along the lines of Theorem 4.1 of Giesecke and Schwenkler (2018). Note that the transition law of Z is non-central chi-squared since Z follows a CIR process.

Algorithm A.2. *Fix a state-space discretization $\{z^1, \ldots, z^m\}$ for the frailty Z. Let $\mathcal{A}^i$ be a neighborhood of z^i. Define the terms*

$$F^n(k,l) = e^{T_n - T_{n-1}} \left(e^{a\cdot(1,X_{T_n-})} + Y_{T_n-} + cz^k\right) \phi(z^l, T_{n-1}; z^k, T_n),$$
$$p^n(k,l) = \mathbb{P}^*\left[Z_{T_n} \in \mathcal{A}^k \,\middle|\, D_\tau, Z_{T_{n-1}} = z^l\right],$$

for $1 \leq k,l \leq m$ and $1 \leq n \leq N_\tau$. In addition, set $T_{N_\tau+1} = \tau$ and

$$F^{N_\tau+1}(k,l) = e^{\tau - T_{N_\tau}}\, u\left(e^{a\cdot(1,X_\tau)} + Y_\tau + cz^k\right) \phi(z^l, T_{N_\tau}; z^k, \tau)$$
$$p^{N_\tau+1}(k,l) = \mathbb{P}^*\left[Z_\tau \in \mathcal{A}^k \,\middle|\, D_\tau, Z_{T_{N_\tau}} = z^l\right],$$

For a given $\theta \in \Theta$, do:

(i) Initialization: Let $\xi = (\xi(1), \ldots, \xi(m))$ be an m-dimensional vector such that $\xi(i) = 1$ if $y \in \mathcal{A}^i$ and otherwise $\xi(i) = 0$.

(ii) Iteration: For $n = 1, \ldots, N_\tau + 1$ do

- *(a) Define the $m \times m$-matrix $\hat{Q}_j$ with elements $F^n(k,l)p^n(k,l)$*
- *(b) Update the vector ξ to $\xi \leftarrow \hat{Q}_j \cdot \xi$*

(iii) Termination: Compute an approximation of $\mathbb{E}^*\left[u(\lambda_\tau)M_\tau \mid D_\tau\right]$ as $\sum_{i=1}^m \xi(i)$.

A.3. Asymptotic properties

Assume that the true parameter θ_0 lies in the interior of the parameter space Θ. If $z = \sigma = 0$, we can compute an MLE $\hat{\theta}_\tau$ analytically and Theorem 3.2 of Giesecke and Schwenkler (2018) implies that the MLE is consistent.[1] Under regularity and identifiability conditions,

[1] Condition (A3) of Giesecke and Schwenkler (2018) is irrelevant in our setting because the parameters driving the explanatory factor X can be estimated separately from the default parameter θ.

Theorem 3.3 of Giesecke and Schwenkler (2018) states that $\hat{\theta}_\tau$ will be asymptotically normal so that $\sqrt{\tau}(\hat{\theta}_\tau - \theta_0)$ converges in distribution to a multivariate normal distribution with mean 0 and variance-covariance matrix

$$\Sigma = \left(- \lim_{\tau \to \infty} \frac{1}{\tau} \nabla^2 \log \mathcal{L}_\tau(\theta_0) \right)^{-1}.$$

We use the following finite-horizon approximation of Σ:

$$\hat{\Sigma}_\tau = \left(- \frac{1}{\tau} \nabla^2 \log \mathcal{L}_\tau(\theta_0) \right)^{-1}.$$

If $\sigma > 0$ and $z > 0$, the likelihood is approximated using Algorithm A.2. Define $\mathcal{L}_\tau^A(\theta)$ as the approximate likelihood computed by Algorithm A.2. An approximate MLE $\hat{\theta}_\tau^A$ maximizes the approximate likelihood function:

$$\hat{\theta}_\tau^A \in \arg\max_{\theta \in \Theta} \mathcal{L}_\tau^A(\theta).$$

Proposition 4.3 of Giesecke and Schwenkler (2018) implies that, under mild technical conditions, the approximate MLE $\hat{\theta}_\tau^A$ will be consistent and asymptotically normal as the sample period grows and the discretization becomes finer. A finite-horizon approximation of the asymptotic variance-covariance matrix is given by

$$\hat{\Sigma}_\tau^A = \left(- \frac{1}{\tau} \nabla^2 \log \mathcal{L}_\tau^A(\hat{\theta}_\tau^A) \right)^{-1}.$$

A.4. Implementation

The numerical algorithm and approximate likelihood maximization are implemented in R and run on an eight-core 32GB Sun blade system. We use the Nelder-Mead optimization method to locate the maximizers of the likelihood function. To address the issue of local optima, we run a number of optimizations with random initial values. In order to deal with very small or very large values of the intensity and the parameters, we scale the time unit to one week.

Appendix B. Posterior mean of intensity

Since the data do not include observations of the frailty Z, the intensity λ_t cannot be measured unless $z = \sigma = 0$. We consider the posterior mean of λ_t, denoted by h_t. This appendix explains the computation of h_t.

Let $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ be the right-continuous and complete filtration generated by the observable data, i.e., $\mathcal{G}_t = \sigma(D_t) \subset \mathcal{F}_t$. The posterior mean h_t of λ_t is the optional projection of λ onto $\mathbb{G}$.[2] It satisfies

$$h_t = \mathbb{E}\left[\lambda_t \mid \mathcal{G}_t\right] \tag{5}$$

almost surely, and represents the default rate given the observable data available at t (more precisely, it is the intensity relative to $\mathbb{G}$). We exploit the change of measure defined by (1) to efficiently compute h_t using two calls of Algorithm A.2, one with $u(\lambda) \equiv 1$ and one with $u(\lambda) = \lambda$. This is based on the following result.

Proposition B.1. *Suppose* $\mathbb{E}[1/M_\tau] = 1$ *for* M_τ *defined in* (1). *For* $t \leq \tau$, *the posterior mean* h_t *satisfies almost surely*

$$h_t = \frac{\mathbb{E}^*\left[\lambda_t M_t \mid D_t\right]}{\mathbb{E}^*\left[M_t \mid D_t\right]}. \tag{6}$$

Proof. Theorem T3 of Section VI of Brémaud (1980) states that, given $\mathbb{E}[1/M_\tau] = 1$, the exponential martingale M induces an equivalent probability measure $\mathbb{P}^*$ on the σ-field $\mathcal{F}_\tau$ with Radon-Nikodym density $\frac{d\mathbb{P}^*}{d\mathbb{P}} = 1/M_\tau$. Moreover, the process $(1/M_t)_{t\leq\tau}$ is a martingale with respect to $\mathbb{F}$ by Theorem II.T8 of Brémaud (1980). Since $M_t > 0$ for all $t \leq \tau$, it follows that $\mathbb{P}$ is also absolutely continuous with respect to $\mathbb{P}^*$ on $\mathcal{F}_t$ with Radon-Nikdoym density

$$\left.\frac{d\mathbb{P}}{d\mathbb{P}^*}\right|_{\mathcal{F}_t} = M_t$$

for $t \leq \tau$. Optional projection indicates that $\mathbb{P}$ is also absolutely continuous with respect to $\mathbb{P}^*$ on the σ-algebra $\mathcal{G}_t \subseteq \mathcal{F}_t$ for all $t \leq \tau$ with Radon-Nikodym density $\mathbb{E}^*\left[M_t \,|\, \mathcal{G}_t\right]$. Now, λ is positive and $\mathbb{F}$-adapted. This implies that

$$h_t = \mathbb{E}\left[\lambda_t \,|\, \mathcal{G}_t\right] = \frac{\mathbb{E}^*\left[\lambda_t M_t \mid \mathcal{G}_t\right]}{\mathbb{E}^*\left[M_t \mid \mathcal{G}_t\right]}.$$

In order to obtain (6), note that $\mathcal{G}_t = \sigma(D_t)$. □

Appendix C. Out-of-sample simulation

This appendix describes the simulation procedure we use to forecast the conditional distribution of the total number of defaults out-of-sample.

[2]See Protter (2004) for details on the optional projection.

Inspired by Das, Duffie, Kapadia, and Saita (2007), Duffie, Saita, and Wang (2007) and Duffie, Eckner, Horel, and Saita (2009), we employ a daily Gaussian vector auto-regression of order 41 for the vector containing the rolling S&P 500 yearly return and volatility, the yield of the 3-month Treasury Bill, and the spread between the yields of the 10-year and the 1-year Treasury bonds. We employ a monthly ARMA(1,1) model for the growth rate of the industrial production growth, a quarterly MA(1) model for the GDP growth rate, and a weekly AR(1) model for the spread between the yields of BAA and AAA rated corporate bonds. We choose the optimal autoregression and moving-average orders according to the Akaike information criterion.

Using data D_t available at time t, we estimate the models of the explanatory variables and the models of default timing in Table 3. We then compute the model-implied conditional distribution of the number of defaults during $(t, t+1]$. The distribution is calculated by Monte Carlo simulation of default times using 100,000 trials. The simulation is based on the (daily) discretization of the posterior mean h_t during $(t, t+1]$ as described in Giesecke, Teng, and Wei (2012), and the simulation of the explanatory variables during that interval. We interpolate the simulated paths of the less frequently observed explanatory variables to obtain a series of daily values for all variables.

The feasibility of event simulation using h_t along with our ability to compute h_t using Proposition B.1 eliminates the need to generate paths of the frailty during $(t, t+1]$, which would be required if the simulation were based on the intensity λ. This increases the efficiency of event prediction by an order of magnitude relative to the prediction based on λ that is standard in the frailty literature.

References

Brémaud, P., 1980. Point Processes and Queues – Martingale Dynamics. Springer-Verlag, New York.

Broadie, M., Kaya, O., 2006. Exact simulation of stochastic volatility and other affine jump diffusion processes. Operations Research 54, 217–231.

Das, S., Duffie, D., Kapadia, N., Saita, L., 2007. Common failings: How corporate defaults are correlated. Journal of Finance 62, 93–117.

Duffie, D., Eckner, A., Horel, G., Saita, L., 2009. Frailty correlated default. Journal of Finance 64, 2089–2123.

Duffie, D., Saita, L., Wang, K., 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics 83, 635–665.

Giesecke, K., Schwenkler, G., 2018. Filtered likelihood for point processes. Journal of Econometrics Forthcoming.

Giesecke, K., Teng, G., Wei, Y., 2012. Numerical solution of jump-diffusion SDEs, working Paper, Stanford University.

Protter, P., 2004. Stochastic Integration and Differential Equations. Springer-Verlag, New York.