Chapter 5
Estimating Distance-to-Default with a Sector-Specific Liability Adjustment via Sequential Monte Carlo

J.-C. Duan and W.-T. Wang

Abstract Distance-to-Default (DTD), a widely adopted corporate default predictor, arises from the classical structural credit risk model of Merton (1974). The modern way of estimating DTD applies the model on an observed time series of equity values along with the default point definition made popular by the commercial KMV model. It is meant to be a default trigger level one year from the evaluation time, and is assumed to be the short-term debt plus 50% of the long-term debt. This default point assumption, however, leaves out other corporate liabilities, which can be substantial and particularly so for financial firms. Duan et al. (2012) rectified it by adding other liabilities after applying an unknown but estimable haircut. Typical DTD estimation uses a one-year long daily time series. With at most four quarterly balance sheets, the estimated haircut is bound to be highly unstable. Post-estimation averaging of the haircuts being applied to a sector of firms is thus sensible for practical applications. Instead of relying on post-estimation averaging, we assume a common haircut for all firms in a sector and devise a novel density-tempered expanding-data sequential Monte Carlo method to jointly estimate this common and other firm-specific parameters. Joint estimation is challenging due to a large number of parameters, but the benefits are manifold, for example, rigorous statistical inference on the common parameter becomes possible and estimates for asset correlations are a by-product. Four industry groups of US firms in 2009 and 2014 are used to demonstrate this estimation method. Our results suggest that this haircut is materially important, and

---

J.-C. Duan (✉)
Business School, Risk Management Institute, and Department of Economics,
National University of Singapore, Singapore, Singapore
e-mail: bizdjc@nus.edu.sg

W.-T. Wang
Risk Management Institute, National University of Singapore,
Singapore, Singapore
e-mail: rmiww@nus.edu.sg

W.K. Härdle et al. (eds.), *Applied Quantitative Finance*, Statistics and Computing,
DOI 10.1007/978-3-662-54486-0_5

varies over time and across industries; for example, the estimates are 78.97% in 2009 and 66.4% in 2014 for 40 randomly selected insurance firms, and 0.76% for all 31 engineering and construction and 83.92% for 40 randomly selected banks in 2014.

5.1 Introduction

Corporate credit risk is a common concern for all financial institutions due to their natural exposures to firms through lending activities. From the perspective of banks, the Basel Capital Accord and its compliance adds further importance to modeling credit risks. The investment community cares about corporate credit risk too due to potential losses to their portfolios. Policy makers/regulators also pay a great deal of attention to corporate credit risk because of the destabilizing effect on the economy/markets when massive corporate defaults occur. Since the seminal credit risk model of Merton (1974), viewing corporate capital structure as an option-like arrangement has gained a wide acceptance in assessing corporate default probabilities. Typically, fundamental information from the balance sheet and equity prices from the stock market are utilized in estimating the model. A particularly important risk measure out of Merton's model is distance-to-default (DTD), whose practical usage has been made popular by the commercial KMV model.

DTD is a widely adopted corporate default predictor. Its empirical estimate is typically obtained by using an observed time series of equity values along with some capital structure attributes. For practical applications, a typically complex capital structure must be simplified. This is usually done through the default point definition made popular by the KMV model. The default point is meant to be the default trigger level one year from the evaluation time, and the KMV default point, according to (Crosbie and Bohn, 2003), equals short-term debt plus 50% of the long-term debt. This default point definition, however, leaves out a firm's other liabilities, which can be substantial and particularly so for financial firms. Duan et al. (2012) proposed to add to the default point all remaining liabilities subject to a haircut, and estimated this haircut by applying the transformed-data maximum likelihood method of Duan (1994, 2000). In typical applications involving one-year long daily time series, only four quarterly balance sheets are available, which offer limited information in identifying the haircut. Thus, averaging the estimates for firms in the same corporate sector and then applying the same haircut to all firms in a two-stage estimation seems to be a sensible and practical solution. The two-stage approach has in fact been adopted by the Credit Research Initiative's live corporate default prediction system at the Risk Management Institute, National University of Singapore.

We propose a density-tempered expanding-data sequential Monte Carlo (SMC) method to estimate the haircut without relying on ad hoc averaging. This haircut is estimated jointly along with all other parameters for individual firms in the same corporate sector. This estimation task is technically challenging because of its high dimensionality (easily over one hundred parameters). Our method progressively adds a block of firms to the sample, and each time the likelihood function due to the additional data is density-tempered in a way that a somewhat arbitrary initial SMC sample of the parameters for these additional firms can be brought through a sequence of steps (reweighting, resampling and support boosting) to eventually arrive at a sample of parameters representing the distribution implied by the target likelihood.

Our method combines the two recently emerged SMC techniques: (1) density-tempered SMC by Del Moral et al. (2006) and Duan and Fulop (2015), and (2) expanding-data SMC by Chopin et al. (2013) and Fulop and Li (2013). Our method is not a simple combination of the two SMC techniques, however. Expanding data in our context is to increase the number of firms as opposed to increasing the number of observations on the same set of firms, and thus it is accompanied by an increase in the number of parameters. The second key difference is our frequentist interpretation of the estimation problem as in Chernozhukov and Hong (2003), and for which we in effect assume an improper prior, meaning that all parameters are treated equally likely before seeing the data. On the methodological front, our innovation is to do away with the need for a prior distribution in the sequential technique, which is accomplished by introducing a somewhat arbitrary but sensible initialization sampler with an analytical density function; for example, multivariate normal or truncated normal when some parameters are subject to domain restrictions. The density associated with this initialization sampler is then absorbed into the importance weight.

Joint estimation with this density-tempered expanding data SMC method is demonstrated with four sectors of US firms (insurance, banks, airlines, and engineering and construction) in 2009 and 2014, respectively. Our results suggest that this haircut is materially important, for example, the estimate is 52.19% for all 37 Engineering and Constructions in 2009 and 83.92% for 40 randomly selected banks in 2014. Joint estimation also yields estimates materially different from those obtained with the two-stage estimation method; for example, 92% for banks in 2009 under the former versus 72.61% under the latter, and the difference is way outside the 95% confidence interval obtained with the SMC method.

In addition to its methodological rigor, joint estimation has another advantage of generating asset correlations among members of a corporate sector. For example, banks and insurers show a significantly heightened level of asset correlations in 2009 as compared to 2014, which is consistent with 2009 being in the midst of a global financial crisis. For the airlines and engineering and construction sectors, a similar pattern exists but the magnitude of the difference in asset correlations are far less pronounced.

The DTD estimates generated by the two-stage method are sometimes comparable to those by our joint estimation method, for example, the engineering and construction industry in both years. For banks, however, the DTDs from the two methods are quite different. The magnitude aside, the correlations (Kendall or Pearson) between the estimates of the DTDs from the two methods exceed 80% except for banks which exhibit substantial but lower correlations as compared to other sectors.

5.2 DTD Subject to a Sector-Specific Liability Adjustment

Typical DTD estimation using a time series of equity values is performed on a firm-by-firm basis. If a corporate sector is to share a common parameter, estimation will require one to stack together all equity time series in that sector in order to reflect asset correlations among firms. To address asset correlations, we modify the Merton (1974) model by incorporating a latent common risk factor for the sector. This modification will, however, retain the Merton model's original results on a firm-by-firm basis.

5.2.1 *The Structural Credit Risk Model with a Common Liability Adjustment*

Let $V_{i,t}$ be the unobserved asset value of firm i at time t. Per usual, it follows a geometric Brownian motion, but we assume a common factor to allow for asset correlations:

$$\frac{dV_{i,t}}{V_{i,t}} = \mu_i dt + \beta_i dB_t^c + \nu_i dB_{i,t} \tag{5.1}$$

where B_t^c and $B_{i,t}$ are two independent standard Brownian motions, β_i is the firm specific coefficient used to capture how firm i responds to the common risk factor, B_t^c, and ν_i is a volatility coefficient to reflect the idiosyncratic risk of firm i. The total variance naturally becomes $\sigma_i^2 = \beta_i^2 + \nu_i^2$. Let F_{it} denote the default point at time T below which firm i will default, and $F_{i,t}$ is known at time t. The Merton (1974) model gives rise to the following equity value of firm i:

$$E_{i,t} = V_{i,t}\Psi(d_{i,t}) - F_{i,t}e^{-r(T-t)}\Psi(d_{i,t} - \sigma_i\sqrt{T-t}) \tag{5.2}$$

where $\Psi(.)$ is the standard normal cumulative distribution function and

$$d_{i,t} = \frac{\ln\left(\frac{V_{i,t}}{F_{i,t}}\right) + \left(r + \frac{\sigma_i^2}{2}\right)(T-t)}{\sigma_i\sqrt{T-t}}. \tag{5.3}$$

The time-t probability of default equals $\Psi(-DTD_{i,t})$, where

$$DTD_{i,t} = \frac{\ln\left(\frac{V_{i,t}}{F_{i,t}}\right) + \left(\mu_i - \frac{\sigma_i^2}{2}\right)(T-t)}{\sigma_i\sqrt{T-t}}. \tag{5.4}$$

The above DTD formula is, however, rarely used in practice because parameter μ is well known to be subject to huge sampling errors when daily time series is used in estimation. A modified DTD formula avoiding μ is typically used in practice; for example, Crosbie and Bohn (2003) and Duan and Wang (2012). This modified formula has also been adopted by the live corporate default prediction system of the Credit Research Initiative at the Risk Management Institute, National University of Singapore (NUS-RMI 2015). Specifically, this modified formula, denoted by DTD* is:

$$DTD_t^* := \frac{\ln\left(\frac{V_{i,t}}{F_{i,t}}\right)}{\sigma_i \sqrt{T-t}} \tag{5.5}$$

Following Duan et al. (2012), the default point is assumed to be sector-specific; that is, $F_{i,t} = SD_{i,t} + 0.5LD_{i,t} + \delta OL_{i,t}$ where the the short term debt ($SD_{i,t}$) is taken as total, the long term debt ($LD_{i,t}$) is halved, and other liabilities ($OL_{i,t}$) is subject to a unknown haircut common to all firms in the industry sector. This default point formula reduces to the KMV model's default point definition when $\delta = 0$. The ideal behind the KMV default point is a recognition that the debts of a firm typically cover a wide range of maturities, and a simple way of adapting the reality to the single-maturity set-up of the Merton model is to apply a 50% haircut to the longer-term debts. As noted in Duan et al. (2012) and further elaborated in Duan and Wang (2012), financial firms tend to have an extremely large amount of other liabilities vis-a-vis short-term and long-term debts (e.g., deposits for banks and policy obligations for insurers can amount to about 80% of their total liabilities). Thus, leaving other liabilities out of the default point will significantly distort the DTD estimate. However, the appropriate haircut is unknown and has to be estimated.

Estimating a firm-specific δ is not a sensible approach, because corporate balance sheets are available at best quarterly. The typical application of using one-year time series of daily equity values only offers three change points in liabilities, leading to a highly noisy estimate of δ. Common δ for a corporate sector is obviously a sensible compromise, but the joint estimation becomes too numerically challenging. Thus, Duan et al. (2012) employed a two-stage approach, which first estimates δ along with other model parameters for each firm in a sector, then averages all δ estimates in the sector, and finally fixing at the average δ, re-estimates other parameters for each firm in the sector. As mentioned earlier, this two-stage approach has also been adopted for the live corporate default prediction system maintained by the CRI team at the Risk Management Institute, National University of Singapore. We show later that joint estimation with all firms in a sector, instead of the two-stage approach, is actually feasible by adapting the modern density-tempered SMC technique to this specific estimation problem.

5.2.2 *The Transformed-Data Likelihood*

Duan (1994, 2000) proposed the transformed-data maximum likelihood estimation method for estimating parameters using derivative contract while the asset values are not directly observable. We apply the method to our joint estimation problem. Let $\hat{V}_{i,t}(\sigma_i, \delta)$ denote the implied asset value computed at (σ_i, δ) using the observed equity value, $E_{i,t}$, where the inverse exists and is unique, because Eq. (5.2) is monotonically increasing in $V_{i,t}$. By the process in Eq. (5.1), the N-firm one-period joint distribution at time $t-1$ is of multivariate normality with mean vector $\boldsymbol{\mu}_{1:N}$ and covariance matrix $\boldsymbol{\Sigma}_{1:N,1:N}$:

$$\begin{bmatrix} \ln(\frac{\hat{V}_{1,t}(\sigma_1,\delta)}{\hat{V}_{1,t-1}(\sigma_1,\delta)}) \\ \ln(\frac{\hat{V}_{2,t}(\sigma_2,\delta)}{\hat{V}_{2,t-1}(\sigma_2,\delta)}) \\ \vdots \\ \ln(\frac{\hat{V}_{N,t}(\sigma_N,\delta)}{\hat{V}_{N,t-1}(\sigma_N,\delta)}) \end{bmatrix} \sim \Psi(\boldsymbol{\mu}_{1:N}, \boldsymbol{\Sigma}_{1:N,1:N}). \tag{5.6}$$

where

$$\boldsymbol{\mu}_{1:N} = \begin{bmatrix} \mu_1 - \frac{1}{2}\sigma_1^2 \\ \mu_2 - \frac{1}{2}\sigma_2^2 \\ \vdots \\ \mu_N - \frac{1}{2}\sigma_N^2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{1:N,1:N} = \begin{bmatrix} \beta_1^2 + \nu_1^2 & \beta_1\beta_2 & \cdots & \beta_1\beta_N \\ \beta_2\beta_1 & \beta_2^2 + \nu_2^2 & \cdots & \beta_2\beta_N \\ \vdots & & \ddots & \vdots \\ \beta_N\beta_1 & \beta_N\beta_{12} & \cdots & \beta_N^2 + \nu_N^2 \end{bmatrix}$$

Also evident from the above, changing the sign of $\{\beta_i, i = 1, 2 \cdots, N\}$ all at once will not change the density function of the above system. For identification, therefore, one can impose a positive sign on any one of them, say, β_1, as long as it is not equal to zero.

As argued in Duan et al. (2012), a firm's asset value may change dramatically due to major investment and financing activities. Hence, the asset value implied from the observed equity value is better standardized using the corresponding book value of assets. This adjustment is to remove the scale effect so as to better capture the dynamics for the assets in place instead of reacting to jumps caused by capital structure changes. Let $\hat{\boldsymbol{W}}_{t,1:N} = [\ln(\hat{V}_{1,t}(\sigma_1, \delta)/A_{1,t}), \ln(\hat{V}_{2,t}(\sigma_2, \delta)/A_{2,t}), \cdots, \ln(\hat{V}_{N,t}(\sigma_N, \delta)/A_{N,t})]'$, where $A_{i,t}$ is book asset value of firm i at time t. The transformed-data log-likelihood function can be derived by taking into account the Jacobian of the transformation from equity value to asset value. We introduce $\boldsymbol{\theta}_{i:j} = \{(\mu_k, \beta_k, \nu_k), k = i, \cdots, j\}$ to stand for the set of the firm-specific parameters from Firm i to j inclusive. Note again that σ_k is a deduced parameter where $\sigma_k^2 = \beta_k^2 + \nu_k^2$.

For a time series sample of equity values on N firms over $t = 1, 2, \cdots, T$, denoted by $\boldsymbol{E}_{1:N}$, the log-likelihood function is

$$\begin{aligned} &\ln \mathcal{L}(\delta, \boldsymbol{\theta}_{1:N}; \boldsymbol{E}_{1:N}) \\ &= -\frac{N(T-1)}{2}\ln(2\pi) - \frac{T-1}{2}\ln\left(\det(\boldsymbol{\Sigma}_{1:N})\right) \\ &\quad -\frac{1}{2}\sum_{t=2}^{T}\left(\Delta\hat{\boldsymbol{W}}_{t,1:N} - \boldsymbol{\mu}_{1:N}\right)' \boldsymbol{\Sigma^{-1}}_{1:N,1:N}\left(\Delta\hat{\boldsymbol{W}}_{t,1:N} - \boldsymbol{\mu}_{1:N}\right) \end{aligned}$$

$$-\sum_{t=2}^{T}\mathbf{1}'_N\hat{\boldsymbol{W}}_{t,1:N}-\sum_{t=2}^{T}\sum_{i=1}^{N}\ln\Psi\left(d_{i,t}(\hat{V}_{i,t}(\sigma_i,\delta),F_{i,t}(\delta),\sigma_i)\right) \tag{5.7}$$

In the above, we have made explicit some elements of the $d_{i,t}$ function defined in Eq. (5.3) so that it is understood as a function of those model parameters. Note that $\Delta\hat{\boldsymbol{W}}_{t,1:N}=\hat{\boldsymbol{W}}_{t,1:N}-\hat{\boldsymbol{W}}_{t-1,1:N}$ and $\mathbf{1}_N$ is an N-dimensional column vector of 1.

Note that directly inverting $\boldsymbol{\Sigma}_{1:N,1:N}$ would create a heavy computational burden when n is relatively large. Under our model specification, this matrix is easily invertible with the Sherman–Morrison formula. Specifically, $\boldsymbol{\Sigma}_{1:N,1:N}$ can be decomposed into the sum of a diagonal matrix $A=diag(\nu_1^2,\ldots,\nu_N^2)$ and the outer product of a column vector $\upsilon=[\beta_1,\ldots,\beta_N]'$ with itself. If $1+\upsilon'A^{-1}\upsilon\neq 0$, then

$$\boldsymbol{\Sigma}_{1:N,1:N}^{-1}=\left(A+\upsilon\upsilon'\right)^{-1}=A^{-1}-\frac{A^{-1}\upsilon\upsilon'A^{-1}}{1+\upsilon'A^{-1}\upsilon}. \tag{5.8}$$

Missing data invariably occur in the real-life data sample. In our case, missing data can occur due to some required items in the balance sheet are occasionally absent or stock prices are not available for some firms at some time points. The likelihood function in Eq. (5.7) can be modified to allow for missing data. Specifically, one adjusts the number of firms, i.e., N, according to data availability at time t; for example, there are s firms with missing data at time t. Once the remaining $N-s$ implied asset values are computed according to Eq. (5.2), the implied asset returns of these firms again follows a multivariate normal distribution with an $(N-s)$ subvector of $\boldsymbol{\mu}_{1:N}$ and an $(N-s)$ sub-matrix of $\boldsymbol{\Sigma}_{1:N,1:N}$. Since missing data may occur differently over time, the adjustment to the likelihood function in Eq. (5.7) will have to be time-dependent. To make the computer code run efficiently, it will be useful to first sequence those firms without missing data and follow by those with missing data. Particularly, firms with similar missing data patterns are better grouped together so that the likelihood function of multiple firms can be evaluated in a larger time block.

5.3 Parameter Estimation by the Density-Tempered Expanding-Data Sequential Monte Carlo

The number of parameters in the likelihood function can be quite large; for example, there were 327 banks in the US in December 2009 giving rise to 982 parameters. Even for the relatively small airlines industry, there were 12 firms in December 2014 totaling 37 parameters to be jointly estimated. The density-tempered expanding-data SMC seems to be the only practical way for estimating such large systems.

Our density-tempered expanding-data SMC method combines the two recently emerged SMC techniques: (1) density-tempered SMC by Del Moral et al. (2006) and Duan and Fulop (2015), and (2) expanding-data SMC by Chopin et al. (2013) and Fulop and Li (2013). The common thread in these methods is to find a bridge linking the prior to the posterior distribution in the Bayesian context of parameter estimation. In the case of density-tempering, the likelihood is raised to a power between 0 (corresponding to the prior) and 1 (corresponding to the posterior) so that by applying a simple self-adapted control, one can sure-footedly migrate from a set of parameter particles representing the prior to the final set of particles for the posterior. The expanding-data SMC in the language of Duan and Fulop (2015), on the other hand, creates a bridge by gradually adding data so that the sequence of intermediate posteriors, represented by different sets of parameter particles and corresponding to various intermediate likelihoods, eventually goes to the final posterior distribution. As argued and demonstrated in Duan and Fulop (2015), density-tempering is a far more stable SMC scheme than the expanding-data approach. In our case, expanding data gradually is because handling a large number of firms all at once is not necessary and in fact not ideal in the earlier stage of estimation due to the extra computational load involved. By sequentially expanding the data set, one in effect only approximately density-temper the incremental likelihood to ensure proper distribution migrations along the way.

Our method is not a simple combination of the two SMC techniques. Expanding data in our context is to increase the number of firms as opposed to increasing the number of observations on the same set of firms, and thus it is accompanied by an increase in the number of parameters. The second key difference is our frequentest interpretation of the estimation problem, and for which we in effect assume an improper prior, meaning that all parameters are treated equally likely before seeing the data. Our methodological innovation is to do away with the prior distribution, and is done by introducing a somewhat arbitrary but sensible initial sampler with an analytical density function; for example, multivariate normal or truncated normal when some parameters are subject to domain restrictions. The corresponding methodological change needed is to replace the likelihood function, used in density-tempering or expanding-data, with the ratio of the likelihood over the initialization density.

We first define the log-likelihood function for the new firms conditional on the firms already being added ($N_s < N_q$); that is,

$$\begin{aligned}
&\ln \mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_q}; \hat{\boldsymbol{W}}_{t,N_s+1:N_q}, t = 1, \cdots, T \mid \boldsymbol{E}_{1:N_s}) \\
&= -\frac{(N_q - N_s)(T-1)}{2} \ln(2\pi) - \frac{T-1}{2} \ln\left(\det(\boldsymbol{\Sigma}_{N_s+1:N_q|N_s})\right) \\
&\quad -\frac{1}{2} \sum_{t=2}^{T} \left(\Delta \hat{\boldsymbol{W}}_{t,N_s+1:N_q} - \boldsymbol{\mu}_{t,N_s+1:N_q|N_s}\right)' \boldsymbol{\Sigma}^{-1}_{N_s+1:N_q|N_s} \\
&\qquad \left(\Delta \hat{\boldsymbol{W}}_{t,N_s+1:N_q} - \boldsymbol{\mu}_{t,N_s+1:N_q|N_s}\right) \\
&\quad - \sum_{t=2}^{T} \mathbf{1}'_{N_q-N_s} \hat{\boldsymbol{W}}_{t,N_s+1:N_q} - \sum_{t=2}^{T} \sum_{i=N_s+1}^{N_q} \ln \Psi\left(d_{i,t}(\hat{V}_{i,t}(\sigma_i, \delta), F_{i,t}(\delta), \sigma_i)\right) \qquad (5.9)
\end{aligned}$$

where

$$\boldsymbol{\mu}_{t,N_s+1:N_q|N_s} = \boldsymbol{\mu}_{t,N_s+1:N_q} + \boldsymbol{\Sigma}_{N_s+1:N_q,1:N_s}\boldsymbol{\Sigma}^{-1}_{1:N_s,1:N_s}\left(\Delta\hat{\boldsymbol{W}}_{t,1:N_s} - \boldsymbol{\mu}_{t,1:N_s}\right) \quad (5.10)$$

$$\boldsymbol{\Sigma}_{N_s+1:N_q|N_s} = \boldsymbol{\Sigma}_{N_s+1:N_q,N_s+1:N_q} - \boldsymbol{\Sigma}_{N_s+1:N_q,1:N_s}\boldsymbol{\Sigma}^{-1}_{1:N_s,1:N_s}\boldsymbol{\Sigma}'_{N_s+1:N_q,1:N_s} \quad (5.11)$$

The above two items are respectively the covariance matrix and mean vector for the $(N_q - N_s)$-dimensional asset returns corresponding to the new firm block, conditional on the asset returns of the existing N_s firms.

Our model's parameters can be divided into to two groups – common (i.e., δ) and firm-specific (i.e., $\boldsymbol{\theta}_{1:N}$). We are interested in the recursive exploration of the sequence of intermediate distributions with the recursion associated with data expansion and density-tempering. The initialization sampler's density for the firm-specific parameters from Firm i to j is denoted by $I_0(\boldsymbol{\theta}_{i:j})$, whereas the one for the common parameter is denoted by $I_0(\delta)$. For the first block of N_1 firms, its initialization sampler is independent of $I_0(\delta)$ so that the joint sampling density is $I_0(\delta)I_0(\boldsymbol{\theta}_{1:N_1})$. The intermediate distribution (up to a proportional constant) while tempering the likelihood with γ is defined as

$$f_{N_1,\gamma}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1}) \propto \left(\frac{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})}{I_0(\delta)I_0(\boldsymbol{\theta}_{1:N_1})}\right)^{\gamma} \times I_0(\delta)I_0(\boldsymbol{\theta}_{1:N_1}). \quad (5.12)$$

The term raised to power γ on the right-hand side of (5.12) is nothing but the importance weight in a sampling sense. Different SMC schemes depart in how the importance weight is controlled so as to obtain a quality sample to represent the final target distribution. Evidently, when $\gamma = 0$, $f_{N_1,0}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})$ is the initialization density. When $\gamma = 1$, $f_{N_1,1}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1}) = \mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})$, which is the likelihood function, up to a proportional constant, for the data sample up to N_1 firms.

When a new block of firms is added (taking from N_s to N_{s+1} firms), it will be more efficient to take advantage of the knowledge about the common parameter already implied by the first N_s firms and the firm-specific parameters of these N_s firms conditional on the common parameter. In real applications, the common parameter, if it were implied solely by the newly added firms, might be quite different from the common parameter suggested by the first N_s firms. When new firms are added, an ideal re-initialization sampler for the common parameter and the firm-specific parameters of the first N_s firms will be a mixture distribution combining the updated distribution revealed by these firms and the original initialization distribution. Specifically, we use the mixture distribution: $I_s^{(m)}(\delta, \boldsymbol{\theta}_{1:N_s}) = [\lambda I_s(\delta) + (1-\lambda)I_0(\delta)]I_s(\boldsymbol{\theta}_{1:N_s} \mid \delta)$ where $I_s(\delta)$ and $I_s(\boldsymbol{\theta}_{1:N_s} \mid \delta)$ denotes the distribution of the common parameter and the firm-specific parameters conditional on the common parameter derived from the SMC sample of the first N_s firms. A natural way of sampling with the conditional distribution, $I_s(\boldsymbol{\theta}_{1:N_s} \mid \delta)$, is to run regressions of $\boldsymbol{\theta}_{1:N_s}$ on δ using the SMC sample already obtained.

The tempered distribution (up to a proportional constant) when reaching N_s firms (for $s \geq 2$) is defined as

$$
\begin{aligned}
&f_{N_s,\gamma}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s}) \\
&\propto \left(\frac{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_{s-1}}; \boldsymbol{E}_{1:N_{s-1}}) \mathcal{L}(\delta, \boldsymbol{\theta}_{N_{s-1}+1:N_s}; \hat{\boldsymbol{W}}_{t,N_{s-1}+1:N_s}, t = 1, \cdots, T \mid \boldsymbol{E}_{1:N_{s-1}})}{I_{s-1}^{(m)}(\delta, \boldsymbol{\theta}_{1:N_{s-1}}) I_0(\boldsymbol{\theta}_{N_{s-1}+1:N_s})} \right)^{\gamma} \\
&\times I_{s-1}^{(m)}(\delta, \boldsymbol{\theta}_{1:N_{s-1}}) I_0(\boldsymbol{\theta}_{N_{s-1}+1:N_s}).
\end{aligned}
\tag{5.13}
$$

The initialization sampler for the firm-specific parameters associated with the newly added firms naturally uses the original initialization sampler.

The terms raised to power γ on the right-hand side of (5.13) is again the importance weight in a sampling sense, controlling sample migration from an initial distribution to the target distribution. If one can obtain a simulated sample of parameter values properly representing $\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})$, this Bayesian posterior with an improper prior, i.e., the likelihood function, shall converge to the asymptotic distribution. Hence, their sample means become the parameter estimates, and the confidence intervals can be straightforwardly obtained. Alternatively, one can use the result of Chernozhukov and Hong (2003) to justify the use of the SMC sample means and covariances in inference because the information equality holds when the correctly specified likelihood function is the target.

Advancing the tempered density will experience two cases. For the initial set of firms (i.e., N_1), moving γ to 1 can be accomplished by applying the following incremental important weight:

$$
\frac{f_{N_1,\gamma^{(2)}}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})}{f_{N_1,\gamma^{(1)}}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})} \propto \left(\frac{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})}{I_0(\delta) I_0(\boldsymbol{\theta}_{1:N_1})} \right)^{\gamma^{(2)} - \gamma^{(1)}}
\tag{5.14}
$$

Advancing from N_{s-1} to N_s firms ($s \geq 2$) can be executed with the following incremental importance weight:

$$
\begin{aligned}
&\frac{f_{N_s,\gamma^{(2)}}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})}{f_{N_s,\gamma^{(1)}}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})} \propto \\
&\left(\frac{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_{s-1}}; \boldsymbol{E}_{1:N_{s-1}}) \mathcal{L}(\delta, \boldsymbol{\theta}_{N_{s-1}+1:N_s}; \hat{\boldsymbol{W}}_{t,N_{s-1}+1:N_s}, t = 1, \cdots, T \mid \boldsymbol{E}_{1:N_{s-1}})}{I_{s-1}^{(m)}(\delta, \boldsymbol{\theta}_{1:N_{s-1}}) I_0(\boldsymbol{\theta}_{N_{s-1}+1:N_s})} \right)^{\gamma^{(2)} - \gamma^{(1)}}
\end{aligned}
\tag{5.15}
$$

While maintaining a minimum effective sample size by a self-adaptive control on γ, one must resample the parameters to even the important weights, and then follow up with several Metroplis-Hastings (MH) moves to boost the empirical support that has been reduced due to resampling. At any stage of (N_s, γ), the MH move targets $f_{N_s,\gamma}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})$ and replaces, if accepted, a subset of $(\delta, \boldsymbol{\theta}_{1:N_s})$. In fact, we need to run block MH moves, because proposing a good-quality parameter vector of a very high dimension without dividing them into blocks would be difficult. We first replace the common parameter, δ, and then proceed to replace firm-specific parameters sequentially according to how blocks of firms are added. Suppose that we have 23 firms and 5 firms are added at a time. The MH moves will comprise first proposing δ for replacement, then 15 parameters associated with the first 5 firms, then another 15 parameters for next 5 firms, and finally the last block of 9 parameters for 3 firms.

We compute the realized acceptance rates for the common parameter and each block of the firm-specific parameters after completing the MH move for the M parameter particles. The MH move will be repeated for the common parameter and blocks of firm-specific parameters, but a particular element (i.e., the common parameter or a block of firm-specific parameters) will be skipped when its cumulative realized acceptance rate has reached a target level, say, 100%. This is to ensure that the empirical support has been properly boosted but without running excessive MH moves.

A suitable proposal sampler for the common parameter or any block of firm-specific parameters is fairly easy to come by, and is typically of high quality. This is because a sample of size, say, M representing $f_{N_s,\gamma}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})$ is already available. The proposal density for the common parameter, δ, is defined as a linear regression model with normally distributed errors on a subset of m parameters, denoted by $\{\theta_1, \theta_2, \cdots, \theta_m\}$, randomly selected from the firm-specific parameters, $\boldsymbol{\theta}_{1:N_s}$; that is, δ^* is sampled based on the following regression model estimated to the parameter sample of size M:

$$\delta = a_0 + \sum_{j=1}^{m} a_j \theta_j + \epsilon, \quad \text{where} \quad \epsilon \sim N(0, \omega). \tag{5.16}$$

Naturally, a sampled δ should be discarded if it is outside of the [0, 1] interval.

For the firm-specific parameters, the proposal sampler is based on a set of regression models. Consider replacing the firm-specific parameters of a block of firms from $N_a + 1$ to N_b when the estimation has already been advanced to N_s firms. For each k between $N_a + 1$ and N_b, we use δ as regressor and estimate the following set of regressions:

$$\begin{aligned} \mu_k &= b_{k,0} + b_{k,1}\delta + \epsilon_{1,k} \\ \beta_k &= c_{k,0} + c_{k,1}\delta + \epsilon_{2,k} \\ \nu_k &= d_{k,0} + d_{k,1}\delta + \epsilon_{3,k} \end{aligned} \tag{5.17}$$

where $\epsilon_{1,k}$, $\epsilon_{2,k}$ and $\epsilon_{3,k}$ are normally distributed with mean zeros and their covariance matrix is computed from the regression residuals. Over different k's, $(\epsilon_{1,k}, \epsilon_{2,k}, \epsilon_{3,k})$ are treated as independent. In short, the proposal sampler takes the three firm-specific parameters as correlated for a firm but independent across different firms in a replacement block.

The regression parameters in effect define the proposal sampler, and these regression parameters are a function of the parameter sample of size M. So, we will use

$\mathcal{M}_{\delta,\boldsymbol{\theta}_{1:N_s}}$ to stand for these sufficient statistics. The proposed new parameters are denoted by $(\delta^*, \boldsymbol{\theta}^*_{1:N_s})$. Since we only propose a subset each time, $(\delta^*, \boldsymbol{\theta}^*_{1:N_s})$ is same as $(\delta, \boldsymbol{\theta}_{1:N_s})$ except for a particular subset being proposed for replacement.

$$\alpha_{N_s,\gamma}\{(\delta, \boldsymbol{\theta}_{1:N_1}) \Rightarrow (\delta^*, \boldsymbol{\theta}^*_{1:N_s})\}$$
$$= \min\left(1, \frac{f_{N_s,\gamma}(\delta^*, \boldsymbol{\theta}^*_{1:N_s}; \boldsymbol{E}_{1:N_s})}{f_{N_s,\gamma}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})} \frac{h(\delta, \boldsymbol{\theta}_{1:N_s} \,|\mathcal{M}_{\delta,\boldsymbol{\theta}_{1:N_s}})}{h(\delta^*, \boldsymbol{\theta}^*_{1:N_s} \,|\mathcal{M}_{\delta,\boldsymbol{\theta}_{1:N_s}})}\right) \tag{5.18}$$

By the standard argument, the target intermediate distribution in (5.13) is the stationary solution to the Markov kernel defined by the above acceptance probability. Note that we are using independent proposal, because $\mathcal{M}_{\delta,\boldsymbol{\theta}_{1:N_s}}$ reflects the whole sample of M parameter values as opposed to an individual element, $(\delta, \boldsymbol{\theta}_{1:N_s})$.

Operationally speaking, the MH acceptance probability falls into one of two cases, and each can be simplified differently.

Case (1): $s = 1$ when the operation is still on the first block of firms (i.e., $1 : N_1$)

The first ratio in (5.18) can be expressed as

$$\frac{f_{N_1,\gamma}(\delta^*, \boldsymbol{\theta}^*_{1:N_1}; \boldsymbol{E}_{1:N_1})}{f_{N_1,\gamma_1}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})} = \left(\frac{\mathcal{L}(\delta^*, \boldsymbol{\theta}^*_{1:N_1}; \boldsymbol{E}_{1:N_1})}{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_1}; \boldsymbol{E}_{1:N_1})}\right)^{\gamma} \left(\frac{I_0(\delta^*) I_0(\boldsymbol{\theta}^*_{1:N_1})}{I_0(\delta) I_0(\boldsymbol{\theta}_{1:N_1})}\right)^{1-\gamma} \tag{5.19}$$

Case (2): $s \geq 2$ when one adds another block of k firms $(N_s = N_{s-1} + k)$

The first ratio in (5.18) can be expressed as

$$\frac{f_{N_s,\gamma}(\delta^*, \boldsymbol{\theta}^*_{1:N_s}; \boldsymbol{E}_{1:N_s})}{f_{N_s,\gamma}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})}$$
$$= \left(\frac{\mathcal{L}(\delta^*, \boldsymbol{\theta}^*_{1:N_s}; \boldsymbol{E}_{1:N_s})}{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})}\right)^{\gamma} \left(\frac{I^{(m)}_{s-1}(\delta^*, \boldsymbol{\theta}^*_{1:N_{s-1}}) I_0(\boldsymbol{\theta}^*_{N_{s-1}+1:N_s})}{I^{(m)}_{s-1}(\delta, \boldsymbol{\theta}_{1:N_{s-1}}) I_0(\boldsymbol{\theta}_{N_{s-1}+1:N_s})}\right)^{1-\gamma} \tag{5.20}$$

Some of the above ratios may be further simplified to speed up calculation by utilizing the fact that $\boldsymbol{\theta}^*_{1:N_s}$ typically shares the same value with $\boldsymbol{\theta}_{1:N_s}$ over some initial segment of variable length. Assume that the firm-specific parameters to be replaced corresponds to the block of firms from $N_a + 1$ to N_b. Note that $\delta^* = \delta, \boldsymbol{\theta}^*_{1:N_a-1} = \boldsymbol{\theta}_{1:N_a-1}$, and $\boldsymbol{\theta}^*_{N_b+1:N_s} = \boldsymbol{\theta}_{N_b+1:N_s}$. Hence,

$$\frac{\mathcal{L}(\delta^*, \boldsymbol{\theta}^*_{1:N_s}; \boldsymbol{E}_{1:N_s})}{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_s}; \boldsymbol{E}_{1:N_s})} = \frac{\mathcal{L}(\delta^*, \boldsymbol{\theta}^*_{1:N_s}; \hat{\boldsymbol{W}}^*_{t,N_a+1:N_s}, t = 1, \cdots, T \mid \boldsymbol{E}_{1:N_a})}{\mathcal{L}(\delta, \boldsymbol{\theta}_{1:N_s}; \hat{\boldsymbol{W}}_{t,N_a+1:N_s}, t = 1, \cdots, T \mid \boldsymbol{E}_{1:N_a})}.$$

Finally, the second ratio in (5.18) in connection with the proposal density can naturally be simplified because sampling is only for the firm-specific parameters pertaining to a specific block of firms and the densities for the parameters outside the block are never invoked.

To summarize, the whole density-tempered expanding-data SMC algorithm along with our specific implementation parameters goes as follows:

- **Step 1: Initialization**
  Sample $(\delta, i = 1, 2, \cdots, M)$ according to the initialization density, $I_0(\delta)$, which is taken as a normal distribution with mean 0.5 and standard deviation 0.3, truncated to [0,1]. Similarly, sample $(\boldsymbol{\theta}_{1:N_1}^{(i)}, i = 1, 2, \cdots, M)$ for the first N_1 firms based on $I_0(\boldsymbol{\theta}_{1:N_1})$. We set the initial block size to 5 firms, i.e., $N_1 = 5$. $I_0(\boldsymbol{\theta}_{1:N_1})$ is a product of normal densities, and they are taken as *i.i.d.* across firms and over the three firm-specific parameters of a firm. For μ_i, the mean and standard deviation are set to 0.2 and 0.2, respectively. In the case of β_i, the mean and standard deviation are 0.15 and 0.05. β_1 is restricted to be positive because of the identification issue discussed earlier in Sect. 2.2, and its sampling is carried out with a truncated normal distribution. Finally for $\ln \nu_i$, the mean and standard deviation are set to $\ln(0.1)$ and 0.05, respectively. The initial sample is of course equally weighted, i.e., $1/M$, and M is set to 1,024.
- **Step 2: Reweighting and resampling**
  Set $\gamma^{(0)} = 0$. Start from $j = 0$ and compute the tempered incremental importance weight:

$$w_{\gamma,\gamma^{(j)}}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_1}^{(i)}) = \left(\frac{\mathcal{L}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_1}^{(i)}; \boldsymbol{E}_{1:N_1})}{I_0(\delta^{(i)}) I_0(\boldsymbol{\theta}_{1:N_1}^{(i)})} \right)^{\gamma - \gamma^{(j)}}$$

  and find γ^* such that the Effective Sample Size (ESS) is no less than B where B is set to $M/2 = 512$. This can be done with a simple grid search to find γ^* to meet the condition, which need not be exact. Note that ESS $= \frac{\left(\sum_{i=1}^{M} w_{\gamma,\gamma^{(j)}}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_1}^{(i)})\right)^2}{\sum_{i=1}^{M} w_{\gamma,\gamma^{(j)}}^2(\delta^{(i)}, \boldsymbol{\theta}_{1:N_1}^{(i)})}$.
  Resample with the incremental weights to obtain an equally weighted sample of size M.
- **Step 3: Support boosting**
  If ESS $\geq 0.9M$, this support boosting step will be skipped. Otherwise, apply the Metropolis-Hastings (MH) move to remove duplicates so as to boost the empirical support (i.e., increase the ESS). Block MH moves are run per the earlier discussion. First, δ is replaced, and then firm-specific parameters $\boldsymbol{\theta}_{1:N_1}$ are replaced in blocks with k firms at a time, and k is set to 5. Compute the realized acceptance rates (over M) for the common parameter and different blocks of firm-specific parameters. The MH move will be repeated for the common parameter and blocks of firm-specific parameters, but a particular element (i.e., the common parameter or a block of firm-specific parameters) will be skipped when its cumulative realized acceptance rate has reached a target level of 100%.
- **Step 4: Advance γ to 1**
  Set $\gamma^{(j+1)} = \gamma^*$. With the support-boosted sample in place, one computes the tempered incremental important weight and finds γ^* again as in Step 2. Reweight,

resample, and follow with support boosting according to the acceptance probability in (5.18). Repeat the operations until reaching $\gamma = 1$.

- **Step 5: Add more firms**
  Add more firms to take from N_{s-1} to N_s, where $N_s = N_{s-1} + k$ and k is set to 5 unless less than 5 firms are left. Perform re-initialization by sampling δ using $I_{s-1}^{(m)}(\delta) = \lambda I_{s-1}(\delta) + (1 - \lambda) I_0(\delta)$ and $\boldsymbol{\theta}_{1:N_{s-1}}$ from $I_{s-1}(\boldsymbol{\theta}_{1:N_{s-1}} \mid \delta)$, where λ is set at 0.8 and $I_{s-1}(\delta)$ is similar to the truncated normal sampler used in the initialization, i.e., $I_0(\delta)$, except for using the sample mean and variance of δ in the SMC sample up to N_{s-1} firms. Sampling $\boldsymbol{\theta}_{1:N_{s-1}}$ conditional on $\delta^{(i)}$ relies on the following three-dimensional multivariate regression:

$$\boldsymbol{\theta}_j = \boldsymbol{\eta}_{j,0} + \boldsymbol{\eta}_{j,1}\delta^{(i)} + \boldsymbol{\epsilon}_j, \qquad \text{where } \epsilon \sim N(\mathbf{0}, \boldsymbol{\Lambda}_j) \text{ and } j = 1, 2, \cdots, N_{s-1}.$$

  Independence across firms is assumed for this sampler, which means $I_{s-1}(\boldsymbol{\theta}_{1:N_{s-1}} \mid \delta)$ is a product of N_{s-1} three-dimensional multivariate normal densities. Again, β_1 must be restricted to be positive for the identification purpose. Thus, $\boldsymbol{\theta}_{1:1}$ is treated differently where its three elements are sampled only using their sample means and covariances obtained from the previous stage so as to avoid the complication arising from the point-specific truncation probability.
  Finally, sample the additional parameters, $(\boldsymbol{\theta}_{N_{s-1}+1:N_s}, i = 1, 2, \cdots, M)$, using the initialization sampler $I_0(\boldsymbol{\theta}_{N_{s-1}+1:N_s})$, which are normally distributed independent across firms and over different parameters for a firm. Append it to $\boldsymbol{\theta}_{1:N_{s-1}}$ to become $\boldsymbol{\theta}_{1:N_s}$. Set $\gamma^{(0)} = 0$. Start from $j = 0$, and compute the incremental important weight as in Eq. (5.15):

$$v_{\gamma,\gamma^{(j)}}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_s}^{(i)}) = \left(\frac{\mathcal{L}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_{s-1}}^{(i)}; \boldsymbol{E}_{1:N_{s-1}}) \mathcal{L}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_s}^{(i)}; \hat{\boldsymbol{W}}_{t,N_{s-1}+1:N_s}, t = 1, \cdots, T \mid \boldsymbol{E}_{1:N_{s-1}})}{I_{s-1}^{(m)}(\delta^{(i)}, \boldsymbol{\theta}_{1:N_{s-1}}^{(i)}) I_0(\boldsymbol{\theta}_{N_{s-1}+1:N_s}^{(i)})} \right)^{\gamma - \gamma^{(j)}}$$

  Find γ^* such that the ESS is no less than B, and follow with reweighting, resampling and support boosting again. Repeat until reaching $\gamma = 1$.
- **Step 6: Repeat adding more firms**
  Repeat Step 5 to take N_s to N_{s+1} until finally reaching N firms.

5.4 Empirical Implementation

5.4.1 Data

We obtain the data from the RMI-CRI database (National University of Singapore, Risk Management Institute, CRI database. Available at: http://rmicri.org [Accessed August 2015]). The data include (1) the daily market capitalization based on closing share price and number of shares outstanding on a subset of US firms in four sectors,

(2) the 3-month US Treasury interest rate series, and (3) the book values of assets and liabilities (short-term, long-term and the remainder) from quarterly balance sheets for these US firms. Share prices and interest rates are available daily, but balance sheets are released quarterly. For a given day, the relevant items are taken from the most recently available quarterly balance sheet. The firms are classified into 76 industry groups by Bloomberg Industry Classification System (BICS). To demonstrate our estimation method for the common liability adjustment factor (i.e., δ), we select four industry groups: Insurance (BICS 10008-20055), Banks (BICS 10008-20051), Airlines (BICS 10004-20018), and Engineering and Construction (BICS 10011-20082) and focus on two years: 2009 and 2014. Our sample size is 250 daily observations for each firm up to the end of the year. According to the δ estimates produced by the RMI-CRI system in its first stage of the two-stage estimation, these four industry sectors show a range of δ's that helps in gaining a better understanding of our proposed method.

Table 5.1 presents the capital structures of these four industry sectors in 2009 and 2014. The firms considered must have consecutive data for at least 22 days in a year. The smallest number of firm is 12 for the airlines industry in 2014 whereas the largest sector is banks with 327 firms in 2009. Evidently from this table, other liabilities being left out of the KMV default point formula can be quite substantial, measured as a fraction of total liabilities. This is particularly so for financial firms such as insurers and banks with other liabilities being around 80% of the total liabilities. If the haircut, i.e., δ, is not negligible, DTD of financial firms will be seriously distorted.

As Table 5.1 shows, there are many banks and insurers in their respective sectors. In the following estimation, we randomly select 40 firms common to 2009 and 2014, and do so for each of these two sectors. In these cases, we in effect jointly

Table 5.1 Capital structure of four industry sectors of US firms

	Airlines		Engineering & Construction		Banks		Insurance	
	2009	2014	2009	2014	2009	2014	2009	2014
# of firms	18	12	37	31	327	312	132	120
Average value								
Market capitalization	1979.77	13428.39	1247.59	1950.87	3311.06	5826.30	4538.22	9657.12
Short-term debt (SD)	2317.60	4608.96	629.86	783.56	8657.02	9056.90	2433.34	3074.60
Long-term debt (LD)	3299.49	3863.44	152.52	502.10	6854.90	5607.53	2711.79	2110.10
Other liabilities (OL)	2673.93	3767.06	136.28	173.28	21408.08	29162.43	24998.16	34612.81
Total liabilities (TL)	8667.21	15495.66	1597.43	2300.46	40789.69	49114.01	35191.69	47934.07
Total assets (TA)	8291.01	12239.46	918.66	1458.94	36920.00	43826.85	30143.29	39797.51
OL/TL	24.10%	23.34%	11.32%	11.72%	83.84%	89.06%	79.16%	78.80%

estimate 121 parameters (1 common plus 40 sets of 3 firm-specific parameters for each firm). Going all the way to jointly estimate using, say, 327 banks in 2009 (close to 1,000 parameters) would be methodologically feasible, but would require a GPU parallel computing implementation to complete the estimation task within a reasonable amount of time.

5.4.2 *Results*

Table 5.2 presents the results of comparing the estimated haircuts from the density-tempered expanding-data SMC method with those from the two-stage approach. The number of firms refers to the firms used in the joint estimation, not the total number of firms in that sector; for example, banks and insurers are capped at 40. The data missing rate is computed as the ratio of the number of missing day-firm observations over the maximum number of day-firms in a particular year. Missing data causes some algorithmic complications. One missing equity value, for example, results in two consecutive missing returns. Missing returns can be easily handled when a single firm is involved. Jointly estimating all firms in a sector as in this paper requires making adjustments to the conditioning set along the time dimension in order to evaluate the conditional likelihood function in Eq. (5.9). To improve computational efficiency, one needs to arrange firms with similar missing data patterns into the same group, and then leaves groups with more missing data to later processing in the sequential optimization scheme.

For the two-stage estimation, the average δ of a sector is computed over the firms in a sector (or 40 firms in the banking or insurance sector) with the haircut values generated by RMI-CRI in its first stage of the two-stage estimation. Also reported and labelled as "Used by CRI" are the haircut actually employed by the RMI-CRI live system, which are averages over a very broad division into financial and non-

Table 5.2 The haircut parameter, δ, for four industry sectors in 2009 and 2014

	Airlines		Engineering & Construction		Banks		Insurance	
	2009	2014	2009	2014	2009	2014	2009	2014
# of firms used in estimation	18	12	37	31	40	40	40	40
Missing data rate	3.64%	1.93%	1.48%	1.26%	7.53%	6.69%	0.96%	0.18%
Two-stage estimation								
Average over firms	0.3493	0.3666	0.5990	0.5009	0.7261	0.6667	0.6262	0.3136
Used by RMI-CRI	0.5671	0.3537	0.5671	0.3537	0.6898	0.5417	0.6898	0.5417
Joint estimation by SMC								
Mean	0.1693	0.0074	0.5219	0.0076	0.9200	0.8392	0.7897	0.6640
$Q_{2.5}$	0.0826	0.0002	0.2847	0.0002	0.8532	0.8170	0.7479	0.6257
$Q_{97.5}$	0.2527	0.0251	0.7450	0.0277	0.9856	0.8627	0.8356	0.6985

financial sectors as opposed to more specific sub-sectors used in this study. The joint estimation results reported in the same table provide the point estimates for the haircut for different sectors in 2009 and 2014. Also presented in the table are upper and lower values of the 95% confidence interval. These confidence intervals suggest that only engineering and construction sector has their estimated haircuts in 2009 from the two-stage method to be statistically indistinguishable from their corresponding haircuts obtained under the joint estimation method.

Table 5.3 is used to highlight the difference in the firm-specific parameters. For the two-stage estimation method, there are only two parameters (μ and σ), and their sector average values in 2009 and 2014 are reported. In contrast, the joint estimation method yields β and ν estimates in addition to μ. Note that β and ν can be combined

Table 5.3 Firm-specific parameters for four industry sectors in 2009 and 2014

	Airlines		Engineering & Construction		Banks		Insurance	
	2009	2014	2009	2014	2009	2014	2009	2014
Two-stage estimation								
μ	0.0972	0.1527	0.1271	−0.0597	−0.0528	0.0084	−0.0273	−0.0226
σ	0.2722	0.2295	0.4631	0.2893	0.1258	0.0618	0.1963	0.1188
Joint estimation by SMC								
μ								
Mean	0.2548	0.1610	0.3405	0.0223	0.6243	0.0081	0.0795	0.0091
Median	0.2364	0.2384	0.1903	0.0368	0.2449	0.0117	0.0355	0.0101
Min	−0.0056	−0.3828	−0.2527	−0.3455	−0.1439	−0.0948	−0.0776	−0.1247
Max	0.5785	0.5093	2.8807	0.5819	3.0980	0.1101	0.7667	0.2155
β								
Mean	0.3719	0.1584	0.1989	0.1248	0.7237	0.0256	0.1624	0.0574
Median	0.3644	0.1710	0.2156	0.1343	0.4694	0.0271	0.1352	0.0455
Min	0.0164	0.0677	−0.1017	0.0227	−0.2294	−0.0048	0.0657	0.0123
Max	0.6770	0.2732	0.3928	0.2272	2.6869	0.0900	0.5246	0.1648
ν								
Mean	0.2480	0.1781	0.4041	0.2567	0.2361	0.0392	0.1508	0.0959
Median	0.2095	0.1392	0.3221	0.2100	0.1560	0.0315	0.0860	0.0650
Min	0.1181	0.1066	0.1305	0.0957	0.0523	0.0199	0.0404	0.0114
Max	0.5752	0.3404	1.3683	0.9328	1.0132	0.1143	0.4923	0.6636
Asset volatility: $\sigma = \sqrt{\beta^2 + \nu^2}$								
Mean	0.4717	0.2431	0.4809	0.2952	0.7821	0.0492	0.2308	0.1142
Median	0.4738	0.2246	0.3889	0.2639	0.4971	0.0445	0.1735	0.0803
Min	0.2314	0.1362	0.2158	0.1150	0.0679	0.0267	0.1007	0.0167
Max	0.7087	0.3988	1.3721	0.9362	2.8716	0.1455	0.7130	0.6838
Asset correlation								
Mean	0.6096	0.4310	0.2627	0.2375	0.7055	0.2476	0.5828	0.3557
Median	0.6870	0.4066	0.2613	0.2273	0.8217	0.2111	0.5986	0.3465
Min	0.0186	0.1815	−0.0641	0.0032	−0.8382	−0.0895	0.0517	0.0737
Max	0.9261	0.6897	0.7452	0.6323	0.9782	0.7290	0.8768	0.7298

to produce σ estimate and also asset correlations. For some sectors, two methods yield distinctively different σ estimates; for example, airlines and banks in 2009. In general, the σ estimates by the joint estimation method are higher than those obtained by the two-stage method. The summary statistics on asset correlations suggest that asset were much more correlated in 2009 as compared to 2014. This is in agreement with the common perception of increased correlations during the 2008–2009 global financial crisis period.

Table 5.4 summarizes the DTDs generated by two estimation methods for the four sectors in 2009 and 2014. The DTD estimates generated by the two-stage method are in some cases comparable to those by the joint estimation method; for example, the engineering and construction industry in both years. For banks, however, the DTD estimates from the two methods are quite different. Generally speaking, the two-stage method yields higher DTD estimates for all sectors in 2009, when markets were more volatile then. A higher DTD implies a higher solvency, and thus the two-stage method leads to a conclusion that firms were safer than they actually were. The magnitude aside, Kendall's τ or Pearson correlation of the two set of DTD estimates exceed 80% except for banks. The correlations for banks are much lower in magnitude but still substantial. Take together, we can conclude that the DTDs from two estimation methods are materially different. When used as a default predictor in a reduced-form model, different estimation methods likely yield different prediction performances. It is reasonable to conjecture that the joint estimation will generate a better default predictor, either judging intuitively from its characteristics over the financial crisis period or simply based on its methodological rigor.

Table 5.4 DTD comparison for four industry sectors in 2009 and 2014

	Airlines		Engineering & Construction		Banks		Insurance	
	2009	2014	2009	2014	2009	2014	2009	2014
Two-stage estimation (RMI-CRI values)								
Mean	1.5749	4.6966	2.7687	4.4324	0.8708	4.1698	2.2356	5.9887
Median	1.4447	4.2140	3.0445	3.9109	0.8303	4.1269	2.4128	5.5646
Min	−0.7147	2.8421	−0.0122	0.5205	−1.2278	0.9697	−0.4286	2.4827
Max	4.0182	7.4182	6.0303	12.8785	3.0753	7.8429	6.9532	11.3443
Joint estimation by SMC								
Mean	0.8738	4.7830	2.6743	4.1897	−0.3748	3.8005	1.7170	5.5910
Median	0.6886	4.4713	2.8580	3.6562	−0.5219	3.8501	1.9077	5.0447
Min	−0.7657	2.8093	−0.0433	0.4990	−1.4501	0.8166	−0.6529	2.3012
Max	3.2526	7.6093	5.8595	13.1316	2.0759	6.5182	6.3411	10.1488
Correlation of the two methods								
Kendall	0.8382	0.9091	0.9670	0.9901	0.6410	0.8063	0.9190	0.9568
Pearson	0.9635	0.9944	0.9992	0.9988	0.7841	0.9681	0.9889	0.9972

References

Chernozhukov, V., & Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, *115*, 293–346.

Chopin, N., Jacob, P. E., & Papaspiliopoulos, O. (2013). SMC^2: a sequential Monte Carlo algorithm with particle Markov Chain Monte Carlo updates. *Journal of the Royal Statistical Society: Series B*, *75*, 397–426.

Crosbie, P., Bohn, J. 2003. Modeling Default Risk, Moodys KMV technical document.

Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, *68*(3), 411–436.

Duan, J.-C. (2000). Correction: maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, *10*, 461–462.

Duan, J.-C. (1994). Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, *4*(2), 155–167.

Duan, J.-C., & Fulop, A. (2015). Density-tempered marginalized sequential Monte Carlo samplers. *Journal of Business and Economic Statistics*, *33*(2), 192–202.

Duan, J.-C., Sun, J., & Wang, T. (2012). Multiperiod corporate default prediction ? *A Forward Intensity Approach, Journal of Econometrics*, *170*(1), 191–209.

Duan, J.-C., & Wang, T. (2012). Measuring distance-to-default for financial and non-financial firms. *Global Credit Review*, *2*, 95–108.

Fulop, A., & Li, J. (2013). Efficient learning via simulation: a marginalized resample-move approach. *Journal of Econometrics*, *176*, 146–161.

Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, *29*, 449–470.

Staff, N. U. S.-R. M. I. (2015). NUS-RMI credit research initiative technical report version: 2015 update 1. *Global Credit Review*, *5*, 113–203.