Publication Date: 01 Apr 2013
Effective Date: 01 Apr 2013
Addendum 7 to the CRI Technical Report, (Version: 2012, Update 2)

This addendum updates the Technical Report (Version: 2012, Update 2) for the extension of forecast horizon that has been implemented as of the probabilities of default (PD) released on 1 April 2013. Previously, the Credit Research Initiative (CRI) system produced default predictions to a horizon of two years. With this update, horizons of up to five years can now be computed.

This extension to a five year horizon is done by constraining the term-structure of the parameter estimates to be Nelson-Siegel (1987; hereafter NS) functions of the forward-starting time. The term-structures are further constrained so that the effect of risk factors on the forward intensity goes to zero as the horizon increases. This allows tractable and parsimonious extrapolations for horizons beyond five years.

The parameter estimation for the NS functions is based on a new numerical method (a pseudo-Bayesian sequential Monte Carlo technique) developed in a working paper by Duan and Fulop (2013). The remainder of this addendum details the new parameter estimation. Section 1 describes the parameterization of the parameters by NS functions, Section 2 gives an overview of the sequential Monte Carlo method that is used to estimate the NS functions, Section 3 describes how the parameters can be re-estimated given new data or updates of old data, and Section 4 details the calculation of the confidence intervals for the parameters estimates.

1 Smoothed parameters

Duan, Sun and Wang (2012; hereafter DSW) formulate the forward intensity model in which the forward default intensity for a firm is a function of a number of covariates. The forward default intensity for different forward starting periods are computed using different sets of parameters.

In DSW, the sets of parameters are estimated separately for each forward starting time. Parameters at different forward starting times that are associated with each covariate can be approximated by a function of the forward starting time using NS type term structure functions. DSW show that this approximation by NS functions does not negatively affect prediction performance. The RMI implementation follows Duan and Fulop (2013) to impose the functional restriction during the estimation as opposed to DSW's method of fitting the curve after parameter estimates have been obtained. This is done for two reasons.

First, it will significantly reduce the number of parameters. For example, using 12 covariates for forward default intensities up to 60 months would necessitate the joint estimation of $13\times60 = 780$ parameters. Here, 13 comes from adding an intercept to the intensity function with 12 covariates. If the coefficients corresponding to each covariate are represented by the NS function of 4 parameters, there will be at most $13 \times 4 = 52$ parameters. In fact, there will be fewer parameters as some of the NS parameters will be restricted to zero.

Second, the NS function will allow for extrapolation. For example, the 13 NS functions estimated with predictions up to 60 months can be used for prediction, say, over 72 months.

The NS function with four free parameters is:

$$h(\tau;\varrho_0,\varrho_1,\varrho_2,d) = \varrho_0 + \varrho_1\frac{1-\exp(-\tau/d)}{\tau/d} + \varrho_2\left[\frac{1-\exp(-\tau/d)}{\tau/d} - \exp(-\tau/d)\right],$$

where τ is the forward starting time (measured in years). In the RMI implementation, τ ranges from 0 to 5 years. For all covariates, the restriction $d > 0$ is imposed so that the functions converge to a value for large τ. This formulation will be used for forward intensities for both defaults and other types of exit.

For the coefficients of all stochastic covariates, the long-run level ϱ_0 is restricted to zero, because the current value of a stochastic covariate should be uninformative of default or other exits when the forward starting time goes to infinity. In other words, the coefficient of such a stochastic covariate should approach zero when τ goes to infinity.

The intercept of the forward intensity function is of course non-stochastic. Thus, ϱ_0 can have non-zero values for the intercept. With these restrictions on the NS parameters, for the example of 12 covariates there will be a total of $12 \times 3 + 1 \times 4 = 40$ parameters.

In the RMI implementation, the NS function is further constrained to be non-positive for certain covariates: DTD level and trend, Liquidity level and trend, and Profitability level and trend. Imposing these constraints is consistent with the previous implementation as documented in Technical Report, Version: 2012, Update 2.

2 Parameter estimation by Sequential Monte Carlo

Reliably estimating a system involving 40 parameters presents a numerical challenge. Moreover, the number of parameters can be greater than 40 if there are more than 12 covariates. The

RMI implmentation follows Duan and Fulop (2013) who use the sequential Monte Carlo (SMC) pseudo-Bayesian method for estimation and self-normalized statistics for inference.

Due to decomposability, the analysis can be performed separately on the default and other exit forward intensities. The data in the RMI implementation is monthly, and the sample likelihood used in estimation relies on default predictions running from 1 month to 60 months with an increment of 1 month. Naturally, default prediction is subject to data availability. Towards the end of the sample, the prediction horizon naturally decreases and stops at one-month predictions.

The following exposition closely follows the appendix in Duan and Fulop (2013). It is important to note that the RMI implementation uses the DSW model which does not contain any latent frailty or partial conditioning variable, and hence is technically much simpler in parameter estimation. For example, there is no nonlinear filtering problem.

Let θ be a set of parameters, $\pi(\theta)$ the prior, $\gamma_t(\theta)$ the pseudo posterior at time t after one makes the τ-period prediction, and $L_{j,\min(T-j\Delta t,\tau)}(\theta)$ the pseudo likelihood function at step j. The sample period is $[0,T]$ and the data in the RMI implementation is monthly, i.e., $\Delta t = 1/12$. Consider the following pseudo-posterior distribution:

$$\gamma_t(\theta) \propto \prod_{j=0}^{t} L_{j,\min(T-j\Delta t,\tau)}(\theta)\pi(\theta), \text{ for } t = 1, \dots, \frac{T}{\Delta t} - 1.$$

One can apply the sequential batch-resampling routine of Chopin (2002) together with tempering steps as in Del Moral, *et al* (2006) to advance the system. For each t, this procedure yields a weighted sample of N particles, $(\theta^{(i,t)}, w^{(i,t)})$ for $i = 1, \dots, N$, whose empirical distribution function will converge to $\gamma_t(\theta)$ as N increases. In what follows, the superscript i denotes the particle index, $i = 1, \dots, N$. Note that in the RMI implementation, $N = 1000$.

***Initialization*:** Draw an initial random sample from the prior: $\left(\theta^{(i,0)} \sim \pi(\theta),\ w^{(i,0)} = \frac{1}{N}\right)$. Here, the only role of the prior $\pi(\theta)$, is to provide the initial particle cloud from which the algorithm can start. Of course, the support of $\pi(\theta)$ must contain the true parameter value θ_0. In the RMI implementation, normal/truncated normal priors are used. Truncation applies in order to impose the restriction $d > 0$. The means for the priors are the MLE parameter estimates obtained without the NS constraints, and the standard deviations are set at 5.

***Recursions and defining the tempering sequence*:** Assume there is a particle cloud $\left(\theta^{(i,t)}, w^{(i,t)}\right)$ whose empirical distribution represents $\gamma_t(\theta)$. Then, a cloud representing $\gamma_{t+1}(\theta)$ will be reached by

combining importance sampling and Markov Chain Monte Carlo (MCMC) steps. Sometimes moving directly from $\gamma_t(\theta)$ to $\gamma_{t+1}(\theta)$ is too ambitious as the two distributions are too far from each other. This will be reflected in highly variable importance weights if one resorts to direct importance sampling. Hence, following Duan and Fulop (2013) which in turn followed Del Moral, *et al* (2006), a tempered bridge is built between the two densities and the particles are evolved through the resulting sequence of densities. In particular, assume that at $t+1$, there are P_{t+1} intermediate densities:

$$\overline{\gamma}_{t+1,p}(\theta) \propto \gamma_t(\theta) L_{t+1,\min(T-(t+1)\Delta t,\tau)}^{\xi_p}(\theta), \text{ for } p = 1, \dots, P_{t+1}.$$

This construction defines an appropriate bridge: $\xi_0 = 0$ so that $\overline{\gamma}_{t+1,0}(\theta) = \gamma_t(\theta)$, and $\xi_{P_{t+1}} = 1$ so that $\overline{\gamma}_{t+1,P_{t+1}}(\theta) = \gamma_{t+1}(\theta)$. For p between 0 and P_{t+1}, ξ^p is chosen from a grid of points to evenly distribute the weights, as decribed below. A particle cloud representing $\overline{\gamma}_{t+1,0}(\theta)$ can be initialized as:

$$\left(\overline{\theta}^{(i,t+1,0)}, \overline{w}^{(i,t+1,0)}\right) = \left(\theta^{(i,t)}, w^{(i,t)}\right).$$

Then, for $p = 1, \dots, P_{t+1}$ the sequence proceeds as follows:

- *Reweighting Step:* In order to arrive at a representation of $\overline{\gamma}_{t+1,p}(\theta)$, the particles representing $\overline{\gamma}_{t+1,p-1}(\theta)$ and the importance sampling principle can be used. This leads to:

$$\begin{aligned}
\overline{\theta}^{(i,t+1,p)} &= \overline{\theta}^{(i,t+1,p-1)} \\
\overline{w}^{(i,t+1,p)} &= \overline{w}^{(i,t+1,p-1)} \frac{\overline{\gamma}_{t+1,p}\left(\overline{\theta}^{(i,t+1,p)}\right)}{\overline{\gamma}_{t+1,p-1}\left(\overline{\theta}^{(i,t+1,p)}\right)} = \overline{w}^{(i,t+1,p-1)} L_{t+1,\min(T-(t+1)\Delta t,\tau)}^{\xi_p - \xi_{p-1}}\left(\overline{\theta}^{(i,t+1,p)}\right)
\end{aligned}$$

  To avoid particle impoverishment in sequential importance sampling where most of the weight is concentrated in a small number of particles, resample-move step is run, which is triggered whenever a measure of particle diversity, the efficient sample size ESS, defined as:

$$ESS = \frac{\left(\sum_{i=1}^{N} \overline{w}^{(i,t+1,p)}\right)^2}{\sum_{i=1}^{N}\left(\overline{w}^{(i,t+1,p)}\right)^2},$$

  falls below some preset value B. Here, resampling directs the particle cloud towards more likely areas of the sampling space, while the move step enriches particle diversity.

  In the RMI implementation, B is set to 50%. Thus, if $ESS < 50\%$, the following resampling and move steps are performed.

- *Resampling Step:* The particles are resampled proportional to their weights. If $I^{(i,t+1,p)} \in (1, \ldots, N)$ are particle indices sampled proportional to $\overline{w}^{(i,t+1,p)}$, the equally weighted particles are obtained as

$$\begin{aligned} \overline{\theta}^{(i,t+1,p)} &= \overline{\theta}^{(I^{(i,t+1,p)},t+1,p)} \\ \overline{w}^{(i,t+1,p)} &= \frac{1}{N} \end{aligned}$$

- *Move Step:* Each particle is passed through a Markov Kernel $K_{t+1,p}\left(\overline{\theta}^{(i,t+1,p)}, \cdot\right)$ that leaves $\overline{\gamma}_{t+1,p}(\theta)$ invariant, typically a Metropolis-Hastings kernel:

  1. Propose $\theta^{*(i)} \sim Q_{t+1,p}\left(\cdot \,\middle|\, \overline{\theta}^{(i,t+1,p)}\right)$.
  2. Compute the acceptance weight α, where:

$$\alpha = \min\left(1, \frac{\overline{\gamma}_{t+1,p}\left(\theta^{*(i)}\right) Q_{t+1,p}\left(\overline{\theta}^{(i,t+1,p)} \,\middle|\, \theta^{*(i)}\right)}{\overline{\gamma}_{t+1,p}\left(\overline{\theta}^{(i,t+1,p)}\right) Q_{t+1,p}\left(\theta^{*(i)} \,\middle|\, \overline{\theta}^{(i,t+1,p)}\right)}\right).$$

  3. With probability α, set $\overline{\theta}^{(i,t+1,p)} = \theta^{*(i)}$, otherwise keep the old particle.

  This step will enrich the support of the particle cloud while conserving its distribution. If the particle set is a poor representation of the target distribution, the resampling step can help adjust the location of the support. Crucially, given the importance sampling setup, the proposal distribution $Q_{t+1,p}\left(\cdot \,\middle|\, \overline{\theta}^{(i,t+1,p)}\right)$ can be adapted using the existing particle cloud.

  In the RMI implementation, block independent normal distribution proposals are fitted to the particle cloud before the move. Three (or four) Nelson-Siegel parameters corresponding to each covariate form one block. To ensure that d remains positive, any block with a non-positive value for d is discarded and the particle is resampled. Note that the likelihood ratio in the Metropolis-Hastings algorithm is not affected by this because the truncated normal creates a common adjustment term in both numerator and denominator.

  As mentioned previously, the coefficients for some covariates are required to be non-positive over all forward starting times. This is achieved by checking whether the NS curve at a particular set of three (or four) parameters meets the condition. If not, the parameter set will be discarded.

  To improve the support of the particle cloud further, one can execute multiple such Metropolis-Hastings steps each time. In the RMI implementation, additional moves are performed

only after ξ_p reaches 1. Each move uses the means implied by the particle set but all standard deviations are increased by a factor of 30%. The number of moves is set to 20 for the first time point and exponentially declines to 3 mid-way to the sample period and stays at 3 for the remainder.

When $p = P_{t+1}$ is reached, a representation of $\gamma_{t+1}(\theta)$ is:

$$\left(\theta^{(i,t+1)}, w^{(i,t+1)}\right) = \left(\overline{\theta}^{(i,t+1,P_{t+1})}, \overline{w}^{(i,t+1,P_{t+1})}\right).$$

Following Duan and Fulop (2013), the tempering sequence ξ_p is automatically set to ensure that the efficient sample size stays close to 50%. This is done by a grid search, where the ESS is evaluated at a grid of candidate ξ_p and the one that produces the closest ESS to 50% is chosen.

3 Periodic updating

In reality, portfolio credit risk models need to be updated periodically as new data arrive and/or old data are revised. With one new month of data, this means that the final date T is increased to $T + \Delta t$. A particular strength of Duan and Fulop (2013) methodology is that the estimation routine does not need to be re-initialized from the prior as the pseudo-posterior using data up to T will provide a much better proposal distribution. Let the pseudo-posterior at T be denoted by:

$$\gamma_T^{(T)}(\theta) \propto \prod_{j=0}^{T/\Delta t-1} L_{j,\min(T-j\Delta t,\tau)}^{(T)}(\theta)\pi(\theta),$$

and the pseudo-posterior at $T + \Delta t$ by:

$$\gamma_{T+\Delta t}^{(T+\Delta t)}(\theta) \propto \prod_{j=0}^{T/\Delta t} L_{j,\min(T-(j-1)\Delta t,\tau)}^{(T+\Delta t)}(\theta)\pi(\theta).$$

The superscript is introduced to differentiate the pseudo likelihoods at T and $T + \Delta t$. Due to data revisions, for example, it may be the case that $L_{j,k}^{(T+\Delta t)}(\theta) \neq L_{j,k}^{(T)}(\theta)$.

Assume that from the previous run up to T there is a weighted set of particles $(\theta^{(i,T/\Delta t-1)}, w^{(i,T/\Delta t-1)})$ representing the pseudo-posterior $\gamma_T^{(T)}(\theta)$. Next, set $\theta^{(i,T/\Delta t)} = \theta^{(i,T/\Delta t-1)}$ and reweight by

$$w^{(i,T/\Delta t)} = w^{(i,T/\Delta t-1)} \times \frac{\gamma_{T+\Delta t}^{(T+\Delta t)}\left(\theta^{(i,T/\Delta t)}\right)}{\gamma_T^{(T)}\left(\theta^{(i,T/\Delta t)}\right)}$$

Since the denominator is already available from the previous run, one only needs to compute the numerator using the new and revised data set. Then, the weighted set $(\theta^{(i,T/\Delta t)}, w^{(i,T/\Delta t)})$ represents

the new pseudo-posterior $\gamma_{T+\Delta t}^{(T+\Delta t)}(\theta)$. If the weights are too uneven, intermediate tempered densities can be constructed and resample-move steps can be executed.

The initial parameter estimation is carried out for all calibration groups using the data up to the month end of January 2013. Relevant quantities (parameter estimates, the 1000 parameter particles and corresponding weights and sample likelihoods) are saved for periodic updating for all future months.

4 Statistical Inference

The full sample size has $T/\Delta t + 1$ time series data points and predictions at time 0 all the way to the last prediction point of $T/\Delta t - 1$ will be made. Denote the pseudo-posterior mean of the parameter of the whole sample by $\hat{\theta}_{T/\Delta t-1}$:

$$\hat{\theta}_{T/\Delta t-1} \quad = \quad \frac{1}{\sum_{i=1}^{N} w^{(i,T/\Delta t-1)}} \sum_{i=1}^{N} w^{(i,T/\Delta t-1)} \theta^{(i,T/\Delta t-1)}$$

Note that $\gamma_t(\theta)$ is not a true posterior, and it cannot directly provide valid Bayesian inference. But following Duan and Fulop (2013) – which is in turn based on Shao's (2010) self-normalized statistic – inference can be performed using the t-like statistic. To test, for example, the hypothesis of the i^{th} element of θ_0, denoted by $\theta_0^{(i)}$, equal to a, one has:

$$t^* = \frac{\sqrt{T/\Delta t}\left(\hat{\theta}_{T/\Delta t-1}^{(i)} - a\right)}{\sqrt{\hat{\delta}_{i,T}}} \rightarrow^d \frac{W(1)}{\left[\int_0^1 (W(r) - rW(1))^2 dr\right]^{1/2}},$$

where $W(r)$ is a Wiener process, $\hat{\delta}_{i,T}$ is the i^{th} diagonal element of $\hat{C}_T$, and

$$\hat{C}_T = \frac{1}{(T/\Delta t)^2} \sum_{l=0}^{T/\Delta t-1} l^2 (\hat{\theta}_l - \hat{\theta}_{T/\Delta t-1})(\hat{\theta}_l - \hat{\theta}_{T/\Delta t-1})'.$$

The right-hand-side random variable for t^* does not have a known distribution, but can be easily simulated. Kiefer, *et al* (2000) reported that the 95% quantile is 5.374 and the 97.5% quantile is 6.811. These values can also be used to set up confidence intervals. Statistics are easily updated along with the periodic updating.

References

[1] Chopin, N., (2002), A Sequential Particle Filter Method for Static Models, *Biometrika* 89, 539–551.

[2] Del Moral, P., Doucet, A. and Jasra, A., (2006), Sequential Monte Carlo Samplers, *Journal of Royal Statistical Society B* 68, 411–436.

[3] Duan, J.C. and Fulop, A., (2013), Multiperiod Corporate Default Prediction with the Partially–Conditioned Forward Intensity, National University of Singapore working paper.

[4] Duan, J.C., Sun, J. and Wang, T., (2012), Multiperiod Corporate Default Prediction – A Forward Intensity Approach, *Journal of Econometrics* 170, 191–209.

[5] Kiefer, N.M., Vogelsang, T.J. and Bunzel, H., (2000), Simple Robust Testing of Regression Hypotheses. *Econometrica* 68, 695–714.

[6] Nelson, C.R. and Siegel, A.F., (1987), Parsimonious modeling of yield curves. *Journal of Business* 60, 473–489.

[7] Shao, X., (2010), A Self-Normalized Approach to Confidence Interval Construction in Time Series. *Journal of the Royal Statistical Society: Series B* 72, 343–366.