Internet Appendix to "Frailty Correlated Default"*

This document contains results and other material that is supplementary to Duffie, Eckner, Horel, and Saita (2009). Section I extends the basic model to allow for unobserved cross-sectional heterogeneity of default risk. Section II allows for a nonlinear dependence of default intensity on distance to default, through a simple non-parametric specification. Section III provides some information on out-of-sample predictive accuracy. Finally, Section IV provides the specification and parameter estimates of the time-series model for covariates.

I. Unobserved Heterogeneity

The Monte Carlo EM algorithm described in Appendix A and the Gibbs sampler described in Appendix B of Duffie, Eckner, Horel, and Saita (2009) are extended to treat unobserved heterogeneity as follows.

The extension of the Monte Carlo EM algorithm is:

0. Initialize $Z_i^{(0)} = 1$ for $1 \leq i \leq m$ and initialize $\theta^{(0)} = (\widehat{\beta}, 0.05, 0)$, where $\widehat{\beta}$ is the maximum likelihood estimator of β in the model without frailty.

1. (Monte-Carlo E-step.) Given the current parameter estimate $\theta^{(k)}$, draw samples $\left(Y^{(j)}, Z^{(j)}\right)$ for $j = 1, \ldots, n$ from the joint posterior distribution $p_{Y,Z}(\,\cdot \mid W, D, \theta^{(k)})$ of the frailty sample path $Y = \{Y_t : 0 \leq t \leq T\}$ and the vector $Z = (Z_i : 1 \leq i \leq m)$ of unobserved heterogeneity variables.

*Citation format: Duffie, Darrell, Andreas Eckner, Guillaume Horel, and Leandro Saita, 2009, Internet Appendix to "Frailty Correlated Default," *Journal of Finance* Vol. 64, 2089-2123, `http://www.afajof.org/IA/2009.asp`.

This can be done, for example, by using the Gibbs sampler described below. The expected complete-data log-likelihood is now given by

$$
\begin{aligned}
Q\left(\theta, \theta^{(k)}\right) &= E_{\theta^{(k)}}\left(\log \mathcal{L}\left(\theta \mid W, Y, Z, D\right)\right) \\
&= \int \log \mathcal{L}\left(\theta \mid W, y, z, D\right) p_{Y,Z}\left(y, z \mid W, D, \theta^{(k)}\right) \, dy \, dz. \quad \text{(IA.1)}
\end{aligned}
$$

Using the sample paths generated by the Gibbs sampler, $(IA.1)$ can be approximated by

$$
\widehat{Q}\left(\theta, \theta^{(k)}\right) = \frac{1}{n} \sum_{j=1}^{n} \log \mathcal{L}\left(\theta \mid W, Y^{(j)}, Z^{(j)}, D\right). \quad \text{(IA.2)}
$$

2. (M-step.) Maximize $\widehat{Q}(\theta, \theta^{(k)})$ with respect to the parameter vector θ, using the Newton-Raphson algorithm. Set the new parameter estimate $\theta^{(k+1)}$ equal to this maximizing value.

3. Replace k with $k+1$, and return to Step 2, repeating the MC E-step and the M-step until reasonable numerical convergence.

The Gibbs sampler for drawing from the joint posterior distribution of $\{Y_t : 0 \leq t \leq T\}$ and $\{Z_i : 1 \leq i \leq m\}$ works as follows:

0. Initialize $Y_t = 0$ for $t = 0, \ldots, T$. Initialize $Z_i = 1$ for $i = 1, \ldots, m$.

1. For $t = 1, \ldots, T$ draw a new value of Y_t from its conditional distribution given Y_{t-1}, Y_{t+1} and the current values for Z_i. This can be done using a straightforward modification of the Metropolis-Hastings algorithm described in Appendix B by treating $\log Z_i$ as an additional covariate with corresponding coefficient equal to one.

2. For $i = 1, \ldots, m$, draw the unobserved heterogeneity variables $Z_1, \ldots, Z_m$ from their conditional distributions given the current path of Y. See below.

3. Store the sample path $\{Y_t : 0 \leq t \leq T\}$ and the variables $\{Z_i : 1 \leq i \leq m\}$. Return to Step 1 and repeat until the desired number of scenarios has been drawn, discarding the first several hundred as a burn-in sample.

It remains to show how to draw the heterogeneity variables $Z_1, \ldots, Z_m$ from their conditional posterior distribution. First, we note that

$$p\left(Z \mid W, Y, D, \theta\right) = \prod_{i=1}^{m} p\left(Z_i \mid W_i, Y, D_i, \theta\right),$$

by conditional independence of the unobserved heterogeneity variables Z_i. In order to draw Z from its conditional distribution, it therefore suffices to show how to draw the Z_i's from their conditional distributions. Recall that we have chosen the heterogeneity variables Z_i to be gamma distributed with mean one and standard deviation 0.5. A short calculation shows that in this case the density parameters a and b are both four. Applying Bayes' rule,

$$\begin{aligned} p\left(Z_i \mid W, Y, D, \theta\right) &\propto p_\Gamma\left(Z_i; 4, 4\right) \mathcal{L}\left(\theta \mid W_i, Y, Z_i, D_i\right) \\ &\propto Z_i^3 e^{-4Z_i} e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} \left[D_{it} \lambda_{it} \Delta t + (1 - D_{it})\right], \end{aligned} \tag{IA.3}$$

where $p_\Gamma\left(\,\cdot\,; a, b\right)$ is the density function of a Gamma distribution with parameters a and b. Plugging (7) of Duffie, Eckner, Horel, and Saita (2009) into $(IA.3)$ gives

$$\begin{aligned} p\left(Z_i \mid W, Y, D, \theta\right) &\propto Z_i^3 e^{-4Z_i} \exp\left(-\sum_{t=t_i}^{T_i} \tilde{\lambda}_{it} e^{\gamma Y_t} Z_i\right) \prod_{t=t_i}^{T_i} \left[D_{it} \lambda_{it} \Delta_t + (1 - D_{it})\right] \\ &= Z_i^3 e^{-4Z_i} \exp\left(-A_i Z_i\right) \cdot \left\{ \begin{array}{cl} B_i Z_i & \text{if company } i \text{ did default} \\ 1 & \text{if company } i \text{ did not default} \end{array} \right\}, \end{aligned} \tag{IA.4}$$

for company specific constants A_i and B_i. The factors in $(IA.4)$ can be combined to give

$$p\left(Z_i \mid W_i, Y, D_i, \theta\right) = p_\Gamma\left(Z_i; 4 + D_{i,T_i}, 4 + A_i\right). \tag{IA.5}$$

This is again a Gamma distribution, but with different parameters, and it is therefore easy to draw samples of Z_i from its conditional distribution.

Table IA.I shows the MLE of the covariate parameter vector β and the frailty parameters η and κ together with estimated standard errors. We see that, while including unobserved heterogeneity decreases the coefficient η of dependence (sometimes called volatility) of the default intensity on the OU frailty process Y from 0.125 to 0.112, our general conclusions regarding the economic significance of the covariates and the importance of including a time-varying frailty variable remain. Moreover, Figure IA.I shows that the posterior distribution of the frailty qualitatively remains essentially the same.

Table IA.I
MLE Frailty with Unobserved Heterogeneity

Maximum likelihood estimates of the intensity parameters in the model with frailty and unobserved heterogeneity. Asymptotic standard errors are computed using the Hessian matrix of the likelihood function at $\theta = \widehat{\theta}$.

	Coefficient	Std. error	t-statistic
constant	−0.895	0.134	−6.7
distance to default	−1.662	0.047	−35.0
trailing stock return	−0.427	0.074	−5.8
3-month T-bill rate	−0.241	0.027	−9.0
trailing S&P 500 return	1.507	0.309	4.9
latent factor volatility	0.112	0.022	5.0
latent factor mean reversion	0.061	0.017	3.5

II. Nonlinearity Check

Duffie, Eckner, Horel, and Saita (2009) assume a linear dependence of the log-intensity on the covariates. This assumption might be overly restrictive, especially in the case of the distance to default (DTD), which explains most of the variation of default intensities across companies and across time. It is indeed possible that, if the response of the true log-intensity to DTD is faster than linear, then the latent variable in our current formulation would be higher when DTDs go well below normal and vice versa.

To check the robustness of our findings with respect to the linearity assumptions, we therefore re-estimate the model using a non-parametric model for the contribution of distance to default, replacing $DTD(t)$ with $-\log U(t)$ in (1) of Duffie, Eckner, Horel, and Saita (2009), where $U(t) = f(DTD(t))$ and $f(x)$ is the non-parametric kernel-smoothed fit of one-year frequency of default in our sample at distance to default of x. Figure IA.II shows the historical occurrence of different levels of distance to default for 402,434 firm-months, while Figure IA.III shows the estimated relationship between the current level of DTD and the annualized default intensity. For values of $DTD \leq 9$, a Gaussian kernel smoother with bandwidth equal to one was used to obtain the intensity estimate, whereas due to lack of data the tail of the distribution was approximated by a log-linear relationship, smoothly extending the graph in Figure IA.II.

Using this extension, we re-estimate the model parameters as before. Table IA.II shows the estimated covariate parameter vector $\widehat{\beta}$ and frailty parameters

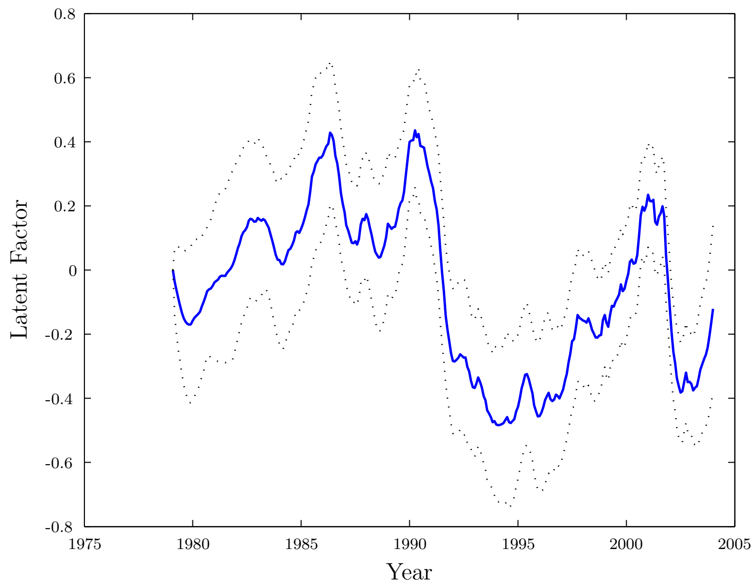

Figure IA.I. Conditional posterior mean $\{E(\eta Y_t \,|\, \mathcal{F}_T) : 0 \leq t \leq T\}$ with one-standard deviation bands, for the scaled Ornstein-Uhlenbeck frailty variable ηY_t in the model that also incorporates unobserved heterogeneity.

$\widehat{\eta}$ and $\widehat{\kappa}$ together with "asymptotic" estimates of standard errors.

Comparing Table II in Duffie, Eckner, Horel, and Saita (2009) and Table IA.II in this Internet Appendix, we see that none of the coefficients linking a firm's covariates to its default intensity has changed noteworthily. In particular, the coefficient now linking the default intensity and $-\log U(t)$ is virtually the same as the coefficient for DTD in the original model. Note, however, that the intercept has changed from -1.20 to 2.28. This difference is due to the fact that $-\log U(t) \approx DTD - 3.5$. Indeed, for the intercept at $DTD = 0$ in Figure IA.III we have $10^{-1.5} \approx 0.032 \approx \exp(-1.20 - 2.28)$. In addition, the posterior path of the latent Ornstein-Uhlenbeck frailty variable looks as before (not shown here).

III. Out-of-Sample Accuracy Ratios

This section provides out-of-sample accuracy ratios for our model and some variants. Given a future time horizon and a particular default prediction model, the "power curve" for out-of-sample default prediction is the function f that

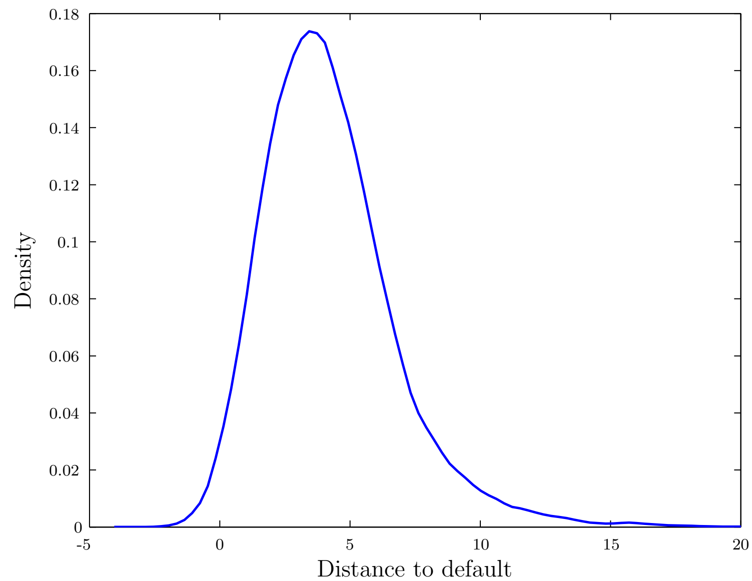

Figure IA.II. DTD population density. Population density estimate of distance to default for 402,434 firm-months between January 1979 and March 2004. The estimate was obtained by applying a Gaussian kernel smoother (bandwidth equal to 0.2) to the empirical distribution.

maps any x in $[0, 1]$ to the fraction $f(x)$ of firms that default before the time horizon that were initially ranked by the model in the lowest fraction x of the population. For example, for the model without frailty, on average over 1993 to 2004, the highest quintile of firms ranked by estimated default probability at the beginning of a year accounted for 92% of firms defaulting within one year. Power curves for the model without frailty are provided in Duffie, Saita, and Wang (2007).

The "accuracy ratio" of a model with power curve f is defined as

$$2\int_0^1 (f(x) - r(x))\, dx,$$

where $x \mapsto r(x) = x$, the identity, is the expected power curve of a completely uninformative model, one that sorts firms randomly. So, a random-sort model has an expected accuracy ratio of zero. A "crystal ball" perfect-sort model has an accuracy ratio of one minus the total ex-post default rate. The accuracy

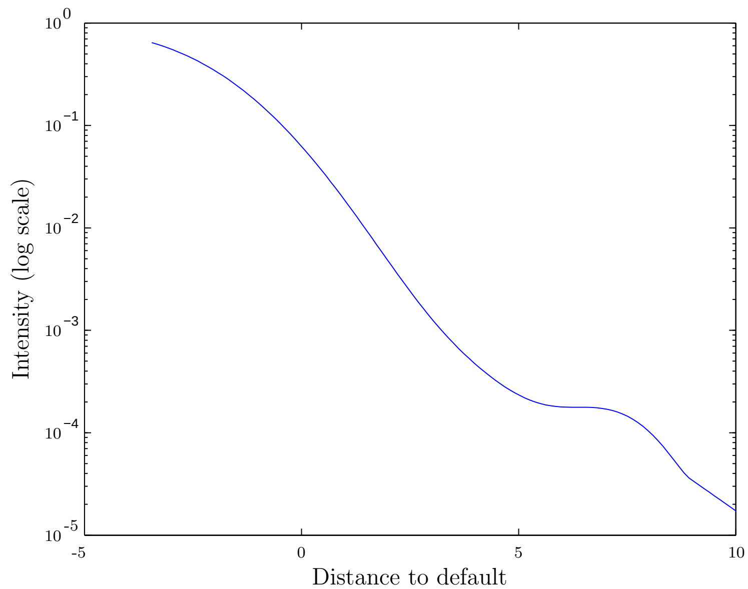

Figure IA.III. Default frequency vs. DTD. Non-parametric estimate of the dependence of annual default frequency on the current level of distance to default (DTD). For values of distance to default less than 9, a Gaussian kernel smoother with bandwidth of 1 was used to obtain the intensity estimate. For DTD larger than 9, a log-linear relationship was assumed.

ratio is a benchmark for comparing the default prediction accuracy of different models.

Duffie, Saita, and Wang (2007), who do not allow for frailty, find accuracy ratios that are an improvement on those of any other model in the available literature. A comparison of the accuracy ratios found in Duffie, Saita, and Wang (2007) with those for the frailty model shown in Figure IA.IV shows that accuracy ratios are essentially unaffected by allowing for frailty. This may be due to the fact that, because of the dominant role of the distance-to-default covariate, firms generally tend to be ranked roughly in order of their distances to default, which of course do not depend on the intensity model. Accuracy ratios, however, measure ordinal (ranking) quality, and do not fully capture the out-of-sample ability of a model to estimate the magnitudes of default probabilities. Our results, however, suggest that the frailty model that we have proposed does not improve the out-of-sample accuracy of the magnitudes of firm-level

Table IA.II
MLE Frailty Nonlinearity Check

Maximum likelihood estimates of the intensity parameters θ in the model with frailty, replacing distance to default with $-\log(f(DTD))$, where DTD is distance to default and $f(\,\cdot\,)$ is the non-parametric kernel estimated mapping from DTD to annual default frequency, illustrated in Figure IA.III. The frailty volatility is the coefficient η of dependence of the default intensity on the standard Ornstein-Uhlenbeck frailty process Y. Estimated asymptotic standard errors were computed using the Hessian matrix of the expected complete data log-likelihood at $\theta = \widehat{\theta}$.

	Coefficient	Std. Error	t-statistic
Constant	2.279	0.194	11.8
$-\log(f(DTD))$	-1.198	0.042	-28.6
Trailing stock return	-0.618	0.075	-8.3
3-month T-bill rate	-0.238	0.030	-8.1
Trailing S&P 500 return	1.577	0.312	5.1
Latent factor volatility	0.128	0.020	6.3
Latent factor mean reversion	0.043	0.009	4.8

estimates of default probabilities over the model without frailty.

Figure IA.V shows accuracy ratios for the variant of our model that replaces the unobserved frailty variable Y with the one-year trailing average default rate. The accuracy ratios are comparable to those of the model with frailty.

IV. Summary of Covariate Time-Series Model

We summarize here the particular parameterization of the time-series model for the covariates that we adopt from Duffie, Saita, and Wang (2007). Because of the high-dimensional state vector, which includes the macroeconomic covariates as well as the distance to default and size of each of almost 3000 firms, we opt for a Gaussian first-order vector autoregressive time series model with the following simple structure.

The 3-month and 10-year Treasury rates, r_{1t} and r_{2t}, respectively, are modeled by taking $r_t = (r_{1t}, r_{2t})'$ to satisfy

$$r_{t+1} = r_t + k_r(\theta_r - r_t) + C_r \epsilon_{t+1}\,,$$

where $\epsilon_1, \epsilon_2, \ldots$ are independent standard-normal vectors, C_r is a 2×2 lower-triangular matrix, and the time step is one month. Provided C_r is of full rank,

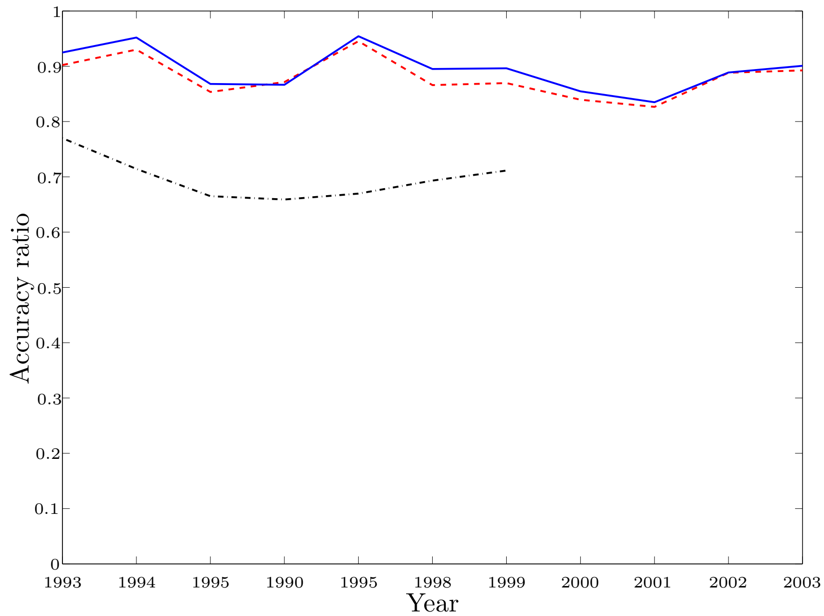

Figure IA.IV. Out-of-sample accuracy ratios. Out-of-sample accuracy ratios (ARs), based on models estimated with data up to December 1992. The solid line provides one-year-ahead ARs based on the model without frailty. The dashed line shows one-year-ahead ARs for the model with frailty. The dash-dot line shows five-year-ahead ARs for the model with frailty.

this is a simple arbitrage-free two-factor affine term-structure model. Maximum likelihood parameter estimates and standard errors are reported in Duffie, Saita, and Wang (2007).

For the distance to default D_{it} and log-assets V_{it} of firm i, and the trailing one-year S&P500 return, S_t, we assume that

$$\begin{bmatrix} D_{i,t+1} \\ V_{i,t+1} \end{bmatrix} = \begin{bmatrix} D_{it} \\ V_{it} \end{bmatrix} + \begin{bmatrix} k_D & 0 \\ 0 & k_V \end{bmatrix} \left(\begin{bmatrix} \theta_{iD} \\ \theta_{iV} \end{bmatrix} - \begin{bmatrix} D_{it} \\ V_{it} \end{bmatrix} \right) + \\ + \begin{bmatrix} b \cdot (\theta_r - r_t) \\ 0 \end{bmatrix} + \begin{bmatrix} \sigma_D & 0 \\ 0 & \sigma_V \end{bmatrix} \eta_{i,t+1} \tag{IA.6}$$

$$S_{t+1} = S_t + k_S(\theta_S - S_t) + \xi_{t+1}, \tag{IA.7}$$

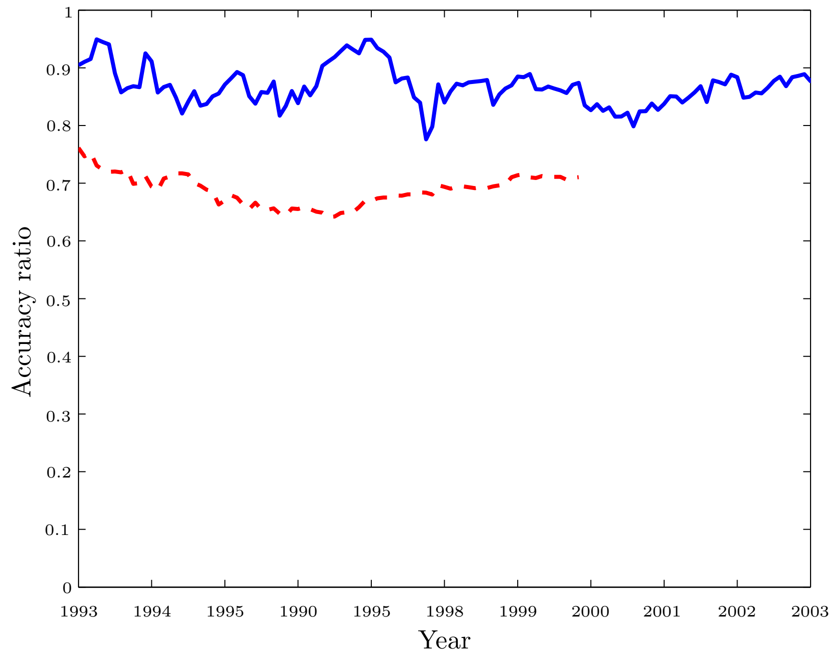

Figure IA.V. Alternative out-of-sample accuracy ratios. Out-of-sample one-year (solid line) and five-year (dashed line) accuracy ratios (ARs), based on the DSW model enhanced with the trailing one-year default rate as an additional covariate.

where

$$\begin{aligned}\eta_{it} &= Az_{it} + Bw_t \\ \xi_t &= \alpha_S u_t + \gamma_S w_t\end{aligned} \tag{IA.8}$$

for $\{z_{1t}, z_{2t}, \ldots, z_{nt}, w_t : t \geq 1\}$ that are *iid* two-dimensional standard normal, all independent of $\{u_1, u_2, \ldots\}$, which are independent standard normals. The 2×2 matrices A and B have $A_{12} = B_{12} = 0$, and are normalized so that the diagonal elements of $AA' + BB'$ are one. For estimation, some such standardization is necessary because the joint distribution of η_{it} (over all i) is determined by the six (non-unit) entries in $AA' + BB'$ and BB'. Our standardization makes A and B equal to the Cholesky decompositions of AA' and BB', respectively. For simplicity, although this is unrealistic, we assume that ϵ is independent of (η, ξ). The maximum likelihood parameter estimates, with standard errors, are provided in Duffie, Saita, and Wang (2007) and are relatively unsurprising.

REFERENCES

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

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