

Specification Analysis of Reduced-Form Credit Risk Models*

Antje Berndt

PRELIMINARY

First Version: January 30, 2006

Current Version: April 4, 2007

Abstract

This paper employs non-parametric specification tests developed in Hong and Li (2005) to evaluate several one-factor reduced-form credit risk models for actual default intensities. Using estimates for actual default probabilities provided by Moody's KMV from 1994 to 2005 for 106 U.S. firms in seven industry groups, we strongly reject popular univariate affine model specifications. As a good compromise between goodness-of-fit and model simplicity we propose to assume that the logarithm of the actual default intensity follows an Ornstein-Uhlenbeck process, also known as the Black-Karasinski (BK) model. For the BK model specification, we find that there is substantial mean-reversion in actual log-default intensities, with an average half-time of roughly 18 months. Our results also show that the level of pairwise correlation in log-default intensities differs across industries. It is higher among oil and gas companies, and lower for healthcare firms.

JEL Classifications: C14, C22, C23, C24, C52

Keywords: non-parametric specification tests, credit risk models

*I thank Moody's KMV for providing access to the EDF data. The author is with the Tepper School of Business, Carnegie Mellon University. Email: aberndt@andrew.cmu.edu.

1 Introduction

This paper estimates a time-series model for U.S. corporate default intensities using one-year default probabilities as estimated by Moody’s KMV EDF rates. Our data consists of 12 years of monthly EDF rates for 106 firms from five industry groups: broadcasting and entertainment, cars, healthcare, oil and gas, and retail. We employ the non-parametric specification tests developed in Hong and Li (2005) to evaluate several one-factor reduced-form credit risk models for actual default intensities. Our findings strongly reject popular univariate affine model specifications such as the Ornstein-Uhlenbeck model, the CIR model and the CIR model with jumps. As a good compromise between goodness-of-fit and model simplicity we propose to assume that the log default intensity follows an Ornstein-Uhlenbeck process, also known as the Black-Karasinski (BK) model.

Because of a substantial small-sample bias in the firm-specific maximum-likelihood estimates of the mean-reversion coefficients, and to account for the co-movement in default risk across firms, we then impose a joint distribution of EDF rates across firms in the same industry. We employ the EM algorithm together with Gibbs sampling to account for missing and censored data in our sector-by-sector estimation strategy. Using the BK model specification, we find that there is substantial mean-reversion in actual log-default intensities, with an average half-time of roughly 18 months. Our results also show that the level of pairwise correlation is different for different industry groups. It is higher among oil and gas companies, and lower for healthcare firms.

The remainder of this paper is structured as follows. Section 2 describes our data source for conditional default probabilities. In Section 3, we introduce four parametric models for default intensities, and Section 4 describes the strategy and results for the reduced-form time-series model estimation. In Section 5, we perform nonparametric tests of the different model specifications.

2 Data

We use the one-year Expected Default Frequency (EDF) data provided by Moody’s KMV as our measure of actual default probabilities. We will discuss this measure only briefly, referring the reader to Berndt, Douglas, Duffie, Ferguson, and Schranz (2005) for a more detailed description. The concept of the EDF measure is based on structural credit risk framework of Black and Scholes (1973) and Merton (1974). In these models, the equity of a firm is viewed as a call option on the firm’s assets, with the strike price equal to the firm’s liabilities. The “distance-to-default” (DD), defined as the number of standard deviations of asset growth by which its assets exceed a measure of book liabilities, is a sufficient statistic of the likelihood of default. In the

current implementation of the EDF model, to the best of our knowledge, the liability measure is equal to the firms short-term book liabilities plus one half of its long-term book liabilities. Estimates of current assets and the current standard deviation of asset growth (volatility) are calibrated from historical observations of the firms equity-market capitalization and of the liability measure. For a detailed discussion, see, for example, Appendix A in Duffie, Saita, and Wang (2005).

Crosbie and Bohn (2001) and Kealhofer (2003) provide more details on the KMV model and the fitting procedures for distance to default and EDF. Unlike the Merton model, where the likelihood of default is the inverse of the normal cumulative distribution function of DD, Moody's KMV EDF measure uses a non-parametric mapping from DD to EDF that is based on a rich history of actual defaults. Therefore, the EDF measure is somewhat less sensitive to model mis-specification. The accuracy of the EDF measure as a predictor of default, and its superior performance compared to rating-based default prediction, is documented in Bohn, Arora, and Korbalev (2005). Duffie, Saita, and Wang (2005) construct a more elaborate default prediction model, using distance to default as well as other covariates. Their model achieves accuracy that is only slightly higher than that of the EDF, suggesting that EDF is a useful proxy for the physical probability of default. Furthermore, the Moodys KMV EDF measure is extensively used in the financial services industry. As noted in Berndt, Douglas, Duffie, Ferguson, and Schranz (2005), 40 of the worlds 50 largest financial institutions are subscribers.

We obtain monthly one-year EDF values from Moody's KMV for the time period July 1993 through March 2004. For the majority of firms in our sample, we observe all 12 years of data. As indicated by Kurbat and Korbalev (2002), Moody's KMV caps its one-year EDF estimate at 20%. Since this truncation, if untreated, would bias our estimator, we explicitly account for this censoring with the associated conditional likelihood, as explained below. Moody's KMV also truncates the EDF below at 2 basis points. Moreover, there is a significant amount of integer-based granularity in EDF data below approximately 10 basis points. We therefore remove from the sample any firm whose sample average EDF is below 10 basis points. There were occasional missing data points. These gaps were also treated exactly, assuming the event of censoring is independent of the underlying missing observation. Table 1 lists the firms for which we have EDF data, showing the number of monthly observations for each as well as the number of EDF observations that were truncated from above at 20% or truncated from below at 0.02%. We also report the firm's average one-year EDFs. Figure 2 displays the time series of the median EDF rates for the healthcare, oil-and-gas, and broadcasting-and-entertainment sectors.

Table 1: Description of EDF data. Source: Moody's KMV.

ticker	sector	uncensrd.	$\geq 20\%$	$\leq 0.02\%$	total	mean(EDF)	ticker	sector	uncensrd.	$\geq 20\%$	$\leq 0.02\%$	total	mean(EDF)
233647Q	B&E	55	0	18	73	10	IPG	B&E	119	0	1	120	95
ABC	H	113	0	0	113	165	JNJ	H	26	0	118	144	2
ABT	H	82	0	62	144	4	KMG	O&G	123	0	21	144	21
ADELQ	B&E	97	16	0	113	763	KMI	O&G	144	0	2	144	28
AGN	H	126	0	18	144	11	KMP	O&G	139	0	3	142	20
AHC	O&G	137	0	7	144	14	KRI	B&E	70	0	50	120	4
AMGN	H	36	0	108	144	4	L	B&E	81	0	21	102	24
APA	O&G	144	0	0	144	30	LH	H	120	0	0	120	215
APC	O&G	144	0	0	144	21	LLY	H	101	0	43	144	5
BAX	H	144	0	0	144	18	MCCC	B&E	56	0	0	56	584
BC	B&E	120	0	0	120	45	MDT	H	39	0	105	144	3
BEV	H	142	2	0	144	309	MGLH	H	103	19	0	122	718
BHI	O&G	144	0	0	144	27	MMM	H	42	0	78	120	3
BJS	O&G	144	0	0	144	48	MRO	H	50	0	70	120	4
BLC	B&E	115	0	5	120	21	MRK	O&G	144	0	0	144	29
BMY	H	54	0	90	144	8	NBR	O&G	144	0	0	144	40
BR	O&G	129	0	15	144	11	NEV	O&G	137	0	0	137	205
BSX	H	123	0	21	144	22	NOI	O&G	97	0	0	97	105
CAH	H	144	0	0	144	19	OCR	H	131	0	13	144	74
CAM	O&G	113	0	0	113	55	OEL	O&G	124	0	0	124	94
CCU	B&E	142	0	2	144	24	OMC	B&E	120	0	0	120	31
CHIR	H	144	0	0	144	26	OXY	O&G	132	0	12	144	22
CHK	O&G	130	13	0	143	520	PDE	O&G	144	0	0	144	192
CHTR	B&E	50	11	0	61	924	PFE	H	36	0	84	120	2
CMCSA	B&E	144	0	0	144	53	PHA	H	77	0	23	100	10
CNG	U	72	0	13	85	5	PKD	O&G	144	0	0	144	300
COC	O&G	45	0	0	45	25	PRM	B&E	107	3	0	110	340
COP	O&G	133	0	11	144	8	PXD	O&G	144	0	0	144	124
COX	B&E	116	0	0	116	22	RCL	B&E	144	0	0	141	80
CVX	O&G	39	0	105	144	3	RIG	O&G	136	0	4	140	45
CYH	H	97	0	0	97	124	SBI	B&E	113	0	0	113	270
DCX	C	68	0	6	74	41	SGP	H	61	0	59	120	11
DGX	H	93	0	0	93	103	SLB	O&G	85	0	35	120	15
DIS	B&E	103	0	41	144	10	SUN	O&G	120	0	0	120	29
DO	O&G	97	0	10	107	27	THC	H	144	0	0	144	65
DVN	O&G	135	0	9	144	23	TLM	O&G	126	0	18	144	21
DYN	U	121	13	0	134	450	TRI	H	67	0	0	67	180
EEP	O&G	114	0	6	120	15	TSG	B&E	94	0	3	97	47
ENRNQ	O&G	105	1	1	107	44	TSO	O&G	135	0	0	135	168
EP	O&G	143	1	0	144	244	TWX (AOL)	B&E	144	0	0	144	61
EPD	O&G	76	0	0	76	24	UCL	O&G	117	0	3	120	15
F	C	143	0	1	144	26	UHS	H	120	0	0	120	30
FST	O&G	142	0	0	142	282	UNH	H	113	0	7	120	19
GDT	H	115	0	2	117	17	VIA	B&E	139	0	5	144	17
GENZ	H	144	0	0	144	22	VLO	O&G	144	0	0	144	40
GLM	O&G	120	0	0	120	51	VPI	O&G	144	0	0	144	128
GM	C	144	0	0	144	33	WFT	O&G	143	0	1	144	87
HAL	O&G	144	0	0	144	52	WLP	H	137	0	7	144	26
HCA	H	140	0	4	144	37	WMB	U	135	8	1	144	236
HCR	H	120	0	0	120	41	WYE	H	86	0	58	144	10
HMA	H	95	0	25	120	21	XOM	O&G	0	0	120	144	2
HRC	H	131	4	0	135	233	XTO	O&G	120	0	0	120	122
HUM	H	142	0	0	142	86	YBTVA	B&E	120	0	0	120	398
ICCI	T	65	0	0	65	326	total		11,894	91	1,545	13,530	

* H: Healthcare; O&G: Oil and Gas; B&E: Broadcasting and Entertainment; C: Cars; T: Telecommunications; U: Utilities. ** In basis points.

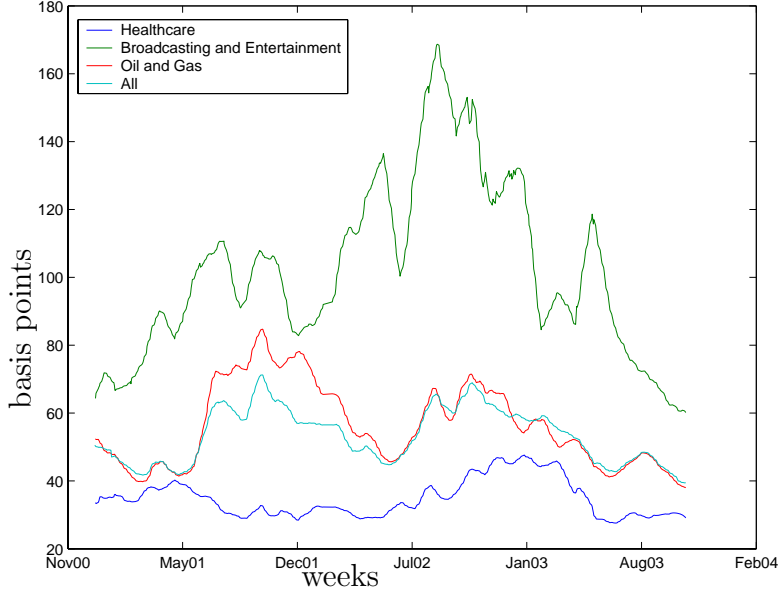


Figure 1: Median one-year EDF rates by sector. Source: Moody’s KMV.

3 Parametric Model for Default Intensity

The default intensity of an obligor is the instantaneous mean arrival rate of default, conditional on all current information. To be slightly more precise, we suppose that default for a given firm occurs at the first event time of a (non-explosive) counting process N with intensity process λ , relative to a given probability space (Ω, \mathcal{F}, P) and information filtration $\{\mathcal{F}_t : t \geq 0\}$ satisfying the usual conditions. In this case, so long as the obligor survives, we say that its default intensity at time t is λ_t . Under mild technical conditions, this means that, conditional on survival to time t and all information available at time t , the probability of default between times t and $t + h$ is approximately $\lambda_t h$ for small h . We also adopt the relatively standard simplifying doubly-stochastic, or Cox-process, assumption, under which the conditional probability at time t , for a currently surviving obligor, that the obligor survives to some later time T , is

$$p(t, T) = E \left(e^{-\int_t^T \lambda(s) ds} \mid \mathcal{F}_t \right). \quad (1)$$

We study four one-factor models for the default intensity that are a special case of the system of stochastic dynamic equations

$$d\lambda_t = [\alpha_0 + \alpha_1 \lambda_t + \alpha_2 \lambda_t \log \lambda_t] dt + [\beta_0 + \beta_1 \lambda_t^\nu] dB_t + \gamma_J \lambda_t \Delta J_t. \quad (2)$$

Here, $\alpha_0, \alpha_1, \alpha_2, \beta_0$, and β_1 are constants, $\nu \in \{0.5, 1\}$ and $\gamma_J \in \{0, 1\}$. (B, B^v) is a two-dimensional standard Brownian motion and ΔJ is a pure

Table 2: Model specifications for actual default intensities.

Id	Model	α_0	α_1	α_2	β_0	β_1	β_2	ν	γ_J
OU	Ornstein-Uhlenbeck	✓	✓		✓				
CIR	Cox-Ingersoll-Ross	✓	✓			✓		0.5	
CIRJ	Cox-Ingersoll-Ross w jumps	✓	✓					0.5	1
BK	Black-Karasinski		✓	✓		✓		1	

jump process, whose jump sizes are independent and whose jump times are those of an independent Poisson process with mean jump arrival rate l . Table 2 gives the details for each of the six model specifications that are subject of our study by showing the values of the exponent ν and the indicator γ_J and by indicating with “✓” those coefficients that appear in nonzero form.

Each of the specifications in Table 2 models the default intensity as a mean-reverting stochastic and, except for the OU model, non-negative process. The first three specifications OU, CIR and CIRJ belong to the class of affine processes with jump diffusions, and closed-form solutions for the survival probabilities in (1) are available. (See Duffie and Kan (1996) and Duffie and Gârleanu (2001) for details.) The BK model is used, for example, by Berndt, Douglas, Duffie, Ferguson, and Schranz (2005) to describe the time-series behavior of actual default arrival intensities. For this model, given the log-autoregressive form of the default intensity in (A.5) in the appendix, there is no closed-form solution available for the one-year EDF, $1 - p(t, t + 1)$ from (1). We therefore rely on numerical lattice-based calculations of $p(t, t + 1)$ and employ the two-stage Hull and White (1994) procedure for constructing trinomial trees.

Duffee (2002) observes that excess returns on corporate bonds are *(i)* small, on average, and that they *(ii)* exhibit a substantial predictable variation. We now examine whether we find similar evidence for actual default intensities. Since we do not observe instantaneous default intensities directly, we will rely on the one-year EDF observations as close proxies for this exploratory analysis and compute the ratio of the sample average of one-year EDFs over the sample standard deviation of the EDFs for each firm in our sample. Table 3 shows cross-sectional summary statistics for these ratios, and Figure 2 plots a histogram of the ratios across all firms. We find evidence that the ratio $\frac{E_t \lambda_{t+h}}{\text{Var}_t \lambda_{t+h}}$ can take on values both above and below one. The latter occurs, using EDFs as a proxy of λ , for roughly one-third of the firms in our sample. Comparing the two one-factor non-negative pure-diffusion models CIR and BK, an

Table 3: In-sample statistics for ratios of a firm’s average one-year EDFs to their standard deviation

Sector	mean	std. dev.	1 st quartile	median	3 rd quartile	firms
Healthcare	1.221	0.480	0.885	1.151	1.389	36
Oil and Gas	1.174	0.440	0.975	1.092	1.346	42
B & E	1.153	0.305	0.968	1.081	1.348	21
Cars	1.421	0.384	1.128	1.513	1.691	3
Others	1.280	0.816	0.583	1.287	1.977	4
All	1.197	0.441	0.929	1.111	1.387	106

attractive feature of the BK model is that

$$\begin{aligned}
 \frac{E(\lambda_{t+h}|\lambda_t)}{\sqrt{Var(\lambda_{t+h}|\lambda_t)}} &= \frac{e^{\theta(1-k)}\lambda_t^k e^{1/2s^2}}{e^{\theta(1-k)}\lambda_t^k \sqrt{e^{2\frac{\sigma^2}{2\kappa}(1-k^2)} - e^{\frac{\sigma^2}{2\kappa}(1-k^2)}}} \\
 &= \frac{1}{\sqrt{e^{\frac{\sigma^2}{2\kappa}(1-k^2)} - 1}} \\
 &\rightarrow \frac{1}{\sqrt{e^{\frac{\sigma^2}{2\kappa}} - 1}} \quad \text{as } h \rightarrow \infty
 \end{aligned}$$

for all time steps h , whereas for the CIR model we have

$$\begin{aligned}
 \frac{E(\lambda_{t+h}|\lambda_t)}{\sqrt{Var(\lambda_{t+h}|\lambda_t)}} &= \frac{\theta(1-k) + k\lambda_t}{\sqrt{\theta\frac{\sigma^2}{2\kappa}(1-k)^2 + \lambda_t\frac{\sigma^2}{\kappa}k(1-k)}} \\
 &\rightarrow \frac{\sqrt{2\theta\kappa}}{\sigma} \quad \text{as } h \rightarrow \infty \text{ or as } \lambda_t \rightarrow 0 \\
 &\geq 1 \quad \text{if the Feller condition is satisfied.}
 \end{aligned}$$

The BK model specification, therefore, offers the flexibility of the conditional standard deviation of λ to exceed the conditional mean for $\sigma^2 > 2 \log(2) \kappa$, whereas in the CIR model that is not possible in the long-run.

4 Estimation Strategy

For our analysis, we will ignore misspecification of the EDF model itself and assume that $1 - p(t, t + 1)$ is indeed the current one-year EDF. From the Moody’s KMV data, we then observe $p(t, t + 1)$ at successive dates $t, t + h, t + 2h, \dots$, where h is one month. From these observations, we will estimate

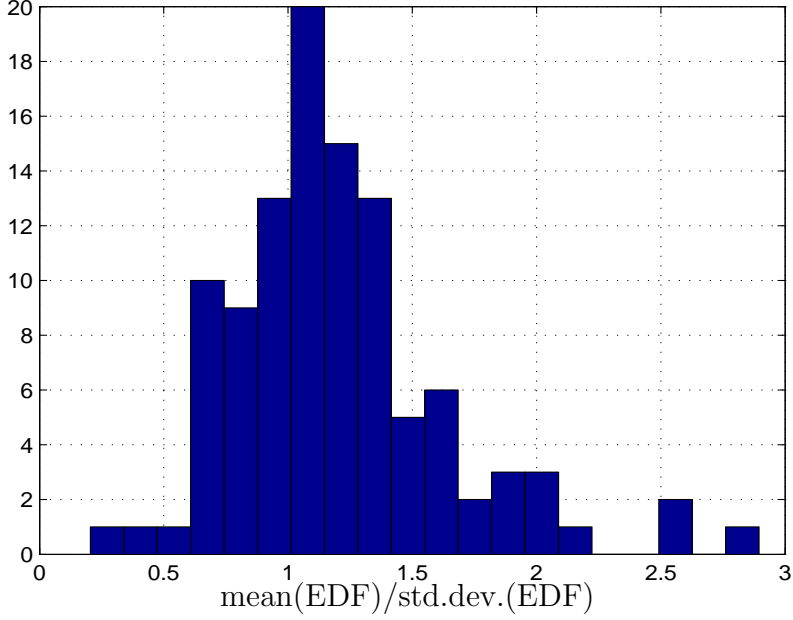


Figure 2: Distribution of the ratio of a firm's average EDF over their standard deviation.

a time-series model of the underlying intensity process λ , for each firm, under the four different model specifications in Table 2. In total, we analyzed 106 firms from seven industry groups.

The data for firm i is the one-year EDF level $Y_{j^i}^i$ at month j^i , for a subset $\{t_{j_0^i}, \dots, t_{j_{N_i}^i}\}$ of $N+1$ month-end times t_0, t_1, \dots, t_N . Our maximum likelihood estimator (MLE) $\hat{\Theta}$ of the parameter vector Θ treats the effects of missing and of truncated EDFs. For each date t_j , let \mathcal{O}_j denote the subset of firms $\{1, \dots, I\}$ for which we observe an uncensored EDF rate at that time, and let \mathcal{C}_j and \mathcal{M}_j denote the subset of firms for which the EDF data at time t_j is truncated and missing, respectively. Then we can define $Y_j^{\mathcal{O}} = \{Y_j^i; i \in \mathcal{O}_j\}$, $Y_j^{\mathcal{C}} = \{Y_j^i; i \in \mathcal{C}_j\}$, and $Y_j^{\mathcal{M}} = \{Y_j^i; i \in \mathcal{M}_j\}$ as the collection of uncensored, truncated and missing EDF observations at time t_j , respectively. Finally, $Y_j = Y_j^{\mathcal{O}} \cup Y_j^{\mathcal{C}} \cup Y_j^{\mathcal{M}}$ collects all EDF data at time t_j .

The complete data likelihood of $Y = \{Y_j : j = 1, \dots, N\}$ evaluated at outcomes $y = \{y_j : j = 1 : N\}$, using the usual abuse of notation for measures, is defined by

$$\mathcal{L}(Y; \Theta) dy = \prod_{j=0}^{N-1} P(Y_{j+1} \in dy_{j+1}; Y_j = y_j, \Theta)$$

where $P(\cdot; Y_n = y_n, \Theta)$ denotes the distribution of $\{Y_{n+1}, Y_{n+2}, \dots\}$ associated with initial condition y_n for Y_n , and associated with parameter vector Θ . A maximum likelihood estimator (MLE) $\hat{\Theta}$ for Θ solves

$$\sup_{\Theta} l(Y; \Theta), \quad (3)$$

where $l(Y; \Theta) = \log(\mathcal{L}(Y; \Theta))$.

Let λ^i denote the default intensity process for firm i , and let

$$X^i = \begin{cases} \lambda^i & \text{if } \lambda^i \text{ is OU, CIR, or CIRJ,} \\ \log \lambda^i & \text{if } \lambda^i \text{ is BK.} \end{cases}$$

denote the vector of state variables for firm i . If Θ is the true parameter vector, then $Y_t^i = G(X_t^i; \Theta)$ for some deterministic function $G(\cdot; \Theta)$ dictated by the modeled EDF

$$G(X_t^i; \Theta) = 1 - E_{\Theta} \left(e^{-\int_t^{t+1} \lambda_s(X_s^i) ds} \mid X_t^i \right), \quad (4)$$

where E_{Θ} denotes expectation associated with the parameter vector Θ . Let $X_j^{\Theta, i} = G^{-1}(Y_j^i; \Theta)$ denote the vector of state variables for firm i at time t_j that would be implied by a non-censored EDF observation Y_j^i , assuming the true parameter vector is Θ , and set $X_j^{\Theta} = (X_j^{\Theta, 1}, \dots, X_j^{\Theta, I})$. Letting $DG(\cdot; \Theta)$ denote the Jacobian of $G(\cdot; \Theta)$ with respect to its first argument, and using standard change-of-measure arguments, we can rewrite the likelihood and log-likelihood as

$$\begin{aligned} \mathcal{L}(Y; \Theta) &= \prod_{j=0}^{N-1} P(X_{j+1}^{\Theta}; X_j^{\Theta}, \Theta) \frac{1}{|\det DG(X_{j+1}^{\Theta}; \Theta)|} \\ l(Y; \Theta) &= \sum_{j=0}^{N-1} \log(P(X_{j+1}^{\Theta}; X_j^{\Theta}, \Theta)) - \log(|\det DG(X_{j+1}^{\Theta}; \Theta)|) \end{aligned} \quad (5)$$

In our case, $|\det DG(X_{j+1}^{\Theta}; \Theta)|$ simplifies to $\prod_i |DG(X_{j+1}^{\Theta, i}; \Theta)|$.

4.1 Firm-by-Firm Parameter Estimation

The MLE $\hat{\Theta}$ of the parameter vector Θ is first obtained considering each firm separately. Our methodology for maximum-likelihood estimation of the parameters of the default intensity treats the effects of missing EDF data as well as censoring of EDFs by truncation from above. For each firm i , let \mathcal{Q}^i , \mathcal{C}^i , and \mathcal{M}^i denote the set of months for which the values Y_j^i are observed without censoring, with censoring from above, and are missing (at random), respectively. In particular, $Y^i = \{Y_{j_0^i}^i, \dots, Y_{j_{N^i}^i}^i\} = \{Y_j^i; j \in \mathcal{O}^i \cup \mathcal{C}^i\}$ is the collection of all EDF observations for firm i .

Suppose, to pick an example of a censoring outcome from which the general case can easily be deduced, that, for months k through $\bar{k} > k + 1$ inclusive, the EDFs are truncated at $\zeta = 20\%$. That means that the censored and observed value Y_j^i is 20%, implying that the actual EDF was equal to or larger than 20%. Let us also assume that the EDF data between months $l + 1$ and \bar{l} , inclusive, are missing, but that we have EDF observations without censoring for all other months. That is, $\mathcal{O}^i = \{0, \dots, k, \bar{k} + 1, \dots, l, \bar{l} + 1, \dots, N\}$, $\mathcal{C}^i = \{j : k + 1 \leq j \leq \bar{k}\}$, and $\mathcal{M}^i = \{j : l + 1 \leq j \leq \bar{l}\}$. Then, the likelihood of the EDF observations Y^i evaluated at outcomes $y = \{y_j : j \in \mathcal{O}^i\}$, using the usual abuse of notation for measures, is defined by

$$\begin{aligned}
\mathcal{L}(Y^i; \Theta) dy &= \prod_{j=0}^{k-1} P(Y_{j+1}^i \in dy_{j+1}; Y_j^i = y_j, \Theta) \\
&\times P(Y_{k+1}^i \geq \zeta, \dots, Y_{\bar{k}}^i \geq \zeta; Y_k^i = y_k, Y_{\bar{k}+1}^i = y_{\bar{k}+1}, \Theta) \\
&\times P(Y_{\bar{k}+1}^i \in dy_{\bar{k}+1}; Y_{\bar{k}}^i = y_{\bar{k}}, \Theta) \\
&\times \prod_{j=\bar{k}+1}^{l-1} P(Y_{j+1}^i \in dy_{j+1}; Y_j^i = y_j, \Theta) \\
&\times P(Y_{l+1}^i \in dy_{l+1}; Y_l^i = y_l, \Theta) \\
&\times \prod_{j=\bar{l}+1}^{N-1} P(Y_{j+1}^i \in dy_{j+1}; Y_j^i = y_j, \Theta),
\end{aligned}$$

Using standard change-of-measure arguments, we can rewrite the likelihood as

$$\begin{aligned}
\mathcal{L}(Y^i; \Theta) dy &= \prod_{j=0}^{k-1} P(X_{j+1}^{\Theta,i}; X_j^{\Theta,i}, \Theta) \frac{1}{|\det DG(X_{j+1}^{\Theta,i}; \Theta)|} \\
&\times P(Y_{k+1}^i \geq \zeta, \dots, Y_{\bar{k}}^i \geq \zeta; Y_k^i = y_k, Y_{\bar{k}+1}^i = y_{\bar{k}+1}, \Theta) \\
&\times P(X_{\bar{k}+1}^{\Theta,i}; X_{\bar{k}}^{\Theta,i}, \Theta) \frac{1}{|\det DG(X_{\bar{k}+1}^{\Theta,i}; \Theta)|} \\
&\times \prod_{j=\bar{k}+1}^{l-1} P(X_{j+1}^{\Theta,i}; X_j^{\Theta,i}, \Theta) \frac{1}{|\det DG(X_{j+1}^{\Theta,i}; \Theta)|} \\
&\times P(X_{l+1}^{\Theta,i}; X_l^{\Theta,i}, \Theta) \frac{1}{|\det DG(X_{l+1}^{\Theta,i}; \Theta)|} \\
&\times \prod_{j=\bar{l}+1}^{N-1} P(X_{j+1}^{\Theta,i}; X_j^{\Theta,i}, \Theta) \frac{1}{|\det DG(X_{j+1}^{\Theta,i}; \Theta)|}. \tag{6}
\end{aligned}$$

The second term on the right-hand side of (6) is equal to

$$q(Y^i; \Theta) = P(X_{k+1}^{\Theta,i} \geq G^{-1}(\zeta; \Theta), \dots, X_{\bar{k}}^{\Theta,i} \geq G^{-1}(\zeta; \Theta); \\ X_k^{\Theta,i} = G^{-1}(y_k; \Theta), X_{\bar{k}+1}^{\Theta,i} = G^{-1}(y_{\bar{k}+1}; \Theta), \Theta).$$

In Appendix A we describe, for each of the model specifications in Table 2, how to compute $q(Y; \Theta)$ by Monte Carlo integration.

A MLE $\hat{\Theta}^i$ for Θ of firm i solves

$$\sup_{\Theta} \mathcal{L}(Y^i; \Theta). \quad (7)$$

The firm-by-firm parameter estimates are summarized in Table 4.

4.2 Sector-by-Sector Parameter Estimation

A Monte-Carlo analysis revealed substantial small-sample bias in the MLE estimators, especially for mean reversion. We therefore impose that all firms within one industry have the same level of mean reversion κ and volatility σ , while allowing for a firm-specific level parameter θ . The Brownian motions driving the default intensities have a constant pairwise correlation across all firms in the sector. For example, for the BK model, we generalize (2) by assuming that X_t^i of firm i satisfies the Ornstein-Uhlenbeck equation

$$dX_t^i = \kappa (\theta^i - X_t^i) dt + \sigma \left(\sqrt{\rho} dB_t^c + \sqrt{1 - \rho} dB_t^i \right), \quad (8)$$

where B^c and B^i are independent standard Brownian motions, independent of $\{B^j\}_{j \neq i}$, and the constant pairwise within-sector correlation coefficient ρ is an additional parameter to be estimated.

We then employ the Expectation-Maximization (EM) algorithm to find a maximum likelihood estimator $\hat{\Theta}$. The EM algorithm starts with an initial guess $\Theta^{(0)}$ and iterates the following two steps:

- **E-step:** Compute

$$Q(\Theta | \Theta^{(m)}) = E[l(Y; \Theta) | Y_o, \Theta^{(m)}] \quad (9)$$

- **M-step:** Find $\Theta^{(m+1)}$ that maximizes $l(\Theta | \Theta^{(m)})$.

It is well known that this iteration always increases the likelihood value (see, for example, Dempster, Lair, and Rubin (1977)). We stop the iteration if the change in the parameters falls below ϵ , for ϵ small.

Table 4: Summary statistics for fitted parameters.

Parameter	OU	CIR	CIRJ	BK
α_0				
mean	23.23	18.32	18.83	
std dev	55.91	44.23	47.15	
median	1.99	6.08	6.16	
1st quartile	0.65	4.01	4.18	
3rd quartile	15.17	14.01	14.45	
α_1				
				$\alpha_1 - \frac{1}{2\beta_1^2}$
mean	0.50	0.09	0.17	1.59
std dev	0.43	0.17	0.29	1.47
median	0.48	0.03	0.01	0.98
1st quartile	0.07	0.00	0.00	0.63
3rd quartile	0.75	0.12	0.22	1.89
α_2				
mean				0.53
std dev				0.41
median				0.38
1st quartile				0.27
3rd quartile				1.53
β_0				
mean	253.14			
std dev	369.74			
median	66.79			
1st quartile	27.31			
3rd quartile	354.00			
β_1				
mean		4.86	4.64	1.35
std dev		3.63	3.62	0.40
median		3.48	3.46	1.31
1st quartile		2.83	2.86	1.12
3rd quartile		5.28	4.74	1.53
γ_J				
mean			4.22	
std dev			35.75	
median			0.03	
1st quartile			0.01	
3rd quartile			0.08	

In our analysis we will approximate the expectation in (9) by its MC estimate. Using a standard change of variable argument, the expectation in (9) can be written as

$$\int l(Y(X); \Theta) f_X(X^c, X^M | X^o, \Theta^{(m)}) d(x^c, x^M).$$

We use the Systematic-Scan Gibbs Sampler (see, for example, Liu (2003)) to impute censored and missing data points. Initially, we set all components of Y^c equal to 20% and initialize all missing data points in Y^M via linear interpolation. Define $Y^{cM} = (Y^c, Y^M)$ and align $Y_{cm} = (Y_{cm,1}, \dots, Y_{cm,N})$, and similarly for the associated X . At the $(g + 1)$ -st iteration of the Gibbs Sampler:

- Draw, for $j = 2, \dots, N$, $X_j^{cM,(g+1)}$ from the conditional distribution

$$f_X(X_j^{cM,(g+1)} | X_1^{cM,(g+1)}, \dots, X_{j-1}^{cM,(g+1)}, X_{j+1}^{cM,(g+1)}, \dots, X_N^{cM,(g+1)}; X^o; \Theta)$$

The proof of the following lemma is provided in Appendix B.

Lemma 1. *Let $\mathcal{S} = \{1, \dots, S\}$ denote some subset of firms $\{1, \dots, I\}$ and let \mathcal{S}^c be its complement. Let's fix some time point t_j between t_0 and t_N . Then, the conditional distribution of $X_{\mathcal{S}^c,j}$ given X_{j-1} , $X_{\mathcal{S},j}$, and X_{j+1} is normal with mean $\mu = w\mu_1 + (1 - w)\mu_2$ and variance-covariance matrix $\Sigma = wA_{22}^{-1}$, where $w = (1 + e^{-2\kappa h})^{-1}$, A_{21} and A_{22} are the lower-left $(I - S) \times S$ and lower-right $(I - S) \times (I - S)$ submatrix of Σ_ϵ^{-1} , respectively, and*

$$\begin{aligned} \mu_1 &= \theta_{\mathcal{S}^c} + e^{-\kappa h}(X_{\mathcal{S}^c,j-1} - \theta_{\mathcal{S}^c}) - (A_{22})^{-1}A_{21}\epsilon_{\mathcal{S},j}, \\ \mu_2 &= \theta_{\mathcal{S}^c} + e^{\kappa h}(X_{\mathcal{S}^c,j+1} - \theta_{\mathcal{S}^c}) + e^{\kappa h}(A_{22})^{-1}A_{21}\epsilon_{\mathcal{S},j+1}. \end{aligned}$$

Sector-by-sector estimates for the BK model are shown in Table 5, with asymptotic standard error estimates in parentheses. Note that the ML estimates for the correlation parameter ρ are quite different for different industry groups. It is higher among oil and gas companies, and lower for healthcare firms. This is confirmed when computing the average pairwise correlation of the firms' innovations to log EDFs by sector, which are reported in Table 6.

Note that according to Figure 2 the healthcare sector has, at least on average, the firms with the highest credit-quality, whereas oil and gas companies are more often of medium credit quality. Table 7 shows the average pairwise correlation of innovations to log EDFs for firms in different median-EDF brackets. For our sample period, pairwise correlation seem to be lower among firms with very low default risk and also among firms with a very high probability of default. They are higher among firms of median credit quality. As shown in Appendix C, however, the pairwise correlation between λ^i and λ^j in the BK model does not depend on θ^i , θ^j or the level of λ .

Table 5: Sector EDF-implied default intensity parameters for the BK model.

	mean($\hat{\theta}$)	$\hat{\kappa}$	$\hat{\sigma}$	$\hat{\rho}$	no. firms
Oil and Gas	3.219	0.393 (0.051)	1.212 (0.036)	0.257 (0.021)	32
Healthcare	3.276	0.538 (0.064)	1.399 (0.049)	0.109 (0.017)	16
B & E	3.855	0.549 (0.090)	1.350 (0.069)	0.229 (0.028)	15

Table 6: Average pairwise correlation of the firms' innovations to log EDFs by sector.

	Healthcare	Oil and Gas	B & E	Cars
Healthcare	12.60%	9.68%	11.08%	7.66%
Oil and Gas	9.68%	25.38%	11.82%	10.64%
B & E	11.08%	11.82%	19.89%	16.31%
Cars	7.66%	10.64%	16.32%	25.66%

5 Non-parametric Specification Test

We now describe a nonparametric specification test for the model specifications in Table 2. The test is based on the nonparametric specification test of Hong and Li (2005). We extend their method to include jump-diffusions. Adapted to our problem statement, the EDF process $Y_t = G(X_t; \Theta)$ will first be transformed to $Z_t = \log(1 - Y_t) = g(X_t; \Theta)$, where $g(x; \Theta) = \log(1 - G(x; \Theta))$. We will then treat Z_t as our observed continuous-time jump-diffusion process that follows the SDE:

$$dZ_t = \mu(Z_t; \Theta) dt + \sigma(Z_t; \theta) dW_t + \Delta J_t(Z_t)$$

where, for the six model specifications described in Table 2,

$$\begin{aligned} \mu(Z; \Theta) &= \frac{\partial g}{\partial X} \Big|_{X=X_t^\Theta} \mu(X) + \frac{1}{2} \frac{\partial g^2}{\partial^2 X} \Big|_{X=X_t^\Theta} \sigma(X)^2, \\ \sigma(Z; \Theta) &= \frac{\partial g}{\partial X} \Big|_{X=X_t^\Theta} \sigma(X) \\ \Delta J_t(Z_t) &= \int_0^\infty g(X_{t-} + x) - g(X_{t-}) \mu_X(dx). \end{aligned}$$

Table 7: Average pairwise correlation of the firms' innovations to log EDFs by credit quality.

median edf (bps)	0-10	10-50	50-100	100-500	500-1000	> 1000
0-10	6.56	8.61	7.36	7.22	6.05	1.47
10-50	8.61	17.89	13.82	13.35	10.48	0.79
50-100	7.36	13.82	16.41	13.46	10.01	0.52
100-500	7.22	13.35	13.46	14.39	10.24	0.27
500-1000	6.05	10.48	10.01	10.24	8.46	0
> 1000	1.47	0.79	0.52	0.27	0	-

In order to apply the specification test we have to transform the sample (Z_0, \dots, Z_N) using the dynamic probability integral transform

$$\begin{aligned} \xi_j(\Theta) &= \int_{-\infty}^{Z_j} P(Z_j = z | Z_{j-1}, \Theta) dz \\ &= \int_{-\infty}^{Z_j} \frac{P(X_j^\Theta = g^{-1}(z; \Theta) | X_{j-1}^\Theta, \Theta)}{\left| \frac{\partial g(x; \Theta)}{\partial x} \right|_{x=X_j^\Theta}} dz. \end{aligned} \quad (10)$$

$(Z_0(\Theta), \dots, Z_N(\Theta))$ will be available in closed-form, up to the parameter vector Θ , for the affine jump diffusions. In particular, $g(x; \Theta) = A(1; \Theta) + B(1; \Theta)x$ is linear in x , so the partial derivative $\frac{\partial g(x; \Theta)}{\partial x}$ does not depend on x , and hence

$$\begin{aligned} \xi_j(\Theta) &= \frac{1}{|B(1; \Theta)|} \int_{-\infty}^{Z_j} P(X_j^\Theta = g^{-1}(z; \Theta) | X_{j-1}^\Theta, \Theta) dz \\ &= \frac{1}{|B(1; \Theta)|} P(X_j \leq g^{-1}(Z_j; \Theta) | X_{j-1}^\Theta, \Theta). \end{aligned} \quad (11)$$

For all other model specifications we will need to employ numerical techniques to evaluate $\frac{\partial G(x; \Theta)}{\partial x} \Big|_{x=X_t^\Theta}$, and also numerical quadrature methods to compute the integral.

If the model is correctly specified, the series $\{\xi_j\}_{j=0}^N$ is i.i.d. $U[0, 1]$. The test statistic suggested by Hong and Li (2005) is

$$\hat{Q}(j) = \frac{1}{V_0^{1/2}} \left[(N-j)h\hat{M}(j) - hA_h^0 \right], \quad (12)$$

where $h = h(n)$ is a bandwidth such that $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$, and

$$\begin{aligned} \hat{M}(j) &= \int_0^1 \int_0^1 (\hat{\eta}_j(z_1, z_2) - 1)^2 dz_1 dz_2 \\ \hat{\eta}_j(z_1, z_2) &= \frac{1}{N-j} \sum_{\tau=j+1}^N K_h(z_1, \hat{\xi}_\tau) K_h(z_2, \hat{\xi}_{\tau-j}) \\ K_h(x, y) &= \begin{cases} \frac{1}{h} \frac{k\left(\frac{x-y}{h}\right)}{\int_{-x/h}^1 k(u) du} & \text{if } x \in [0, h) \\ \frac{1}{h} k\left(\frac{x-y}{h}\right) & \text{if } x \in [h, 1-h] \\ \frac{1}{h} \frac{k\left(\frac{x-y}{h}\right)}{\int_{-1}^{(1-x)/h} k(u) du} & \text{if } x \in (1-h, 1] \end{cases} \\ k(u) &= \frac{15}{16} (1-u^2)^2 1_{\{|u| \leq 1\}}. \end{aligned}$$

The non-stochastic centering and scaling factors are

$$\begin{aligned} A_h^0 &= \left[\left(\frac{1}{h} - 2 \right) \int_{-1}^1 k^2(u) du + 2 \int_0^1 \int_{-1}^b k_b^2(u) k^2(u) du db \right]^2 - 1, \\ V_0 &= 2 \left[\int_{-1}^1 \left(\int_{-1}^1 k(u+v) k(v) dv \right)^2 du \right]^2, \\ k_b(\cdot) &= \frac{k(\cdot)}{\int_{-1}^b k(v) dv}. \end{aligned}$$

As suggested in Hong and Li (2005), we use $h = n^{-1/6} \text{std}(\{\hat{\xi}\})$. Under the correct model specification, Hong and Li (2005) show that

$$\hat{Q}(j) \rightarrow_d N(0, 1), \quad (13)$$

$$\text{cov}(\hat{Q}(i), \hat{Q}(j)) \rightarrow_p 0 \quad \text{for } i \neq j \quad (14)$$

Under model misspecification, on the other hand, we have

$$\hat{Q}(j) \rightarrow_p \infty.$$

Hence, we compare the test statistic $\hat{Q}(j)$ with the upper-tailed $N(0, 1)$ critical value C_α at the level α and, if $\hat{Q}(j) > C_\alpha$, reject the null hypothesis of correct model specification at level α .

Figure 3 plots the histogram of generalized residuals, across all firms, and Figure 4 displays the $Q(j)$ test statistics, $j = 1, \dots, 20$, for the OU, CIR and BK model specifications. Finally, Table 8 shows the rejection rates based on $\hat{Q}(1)$ statistics for 106 firms in our sample.

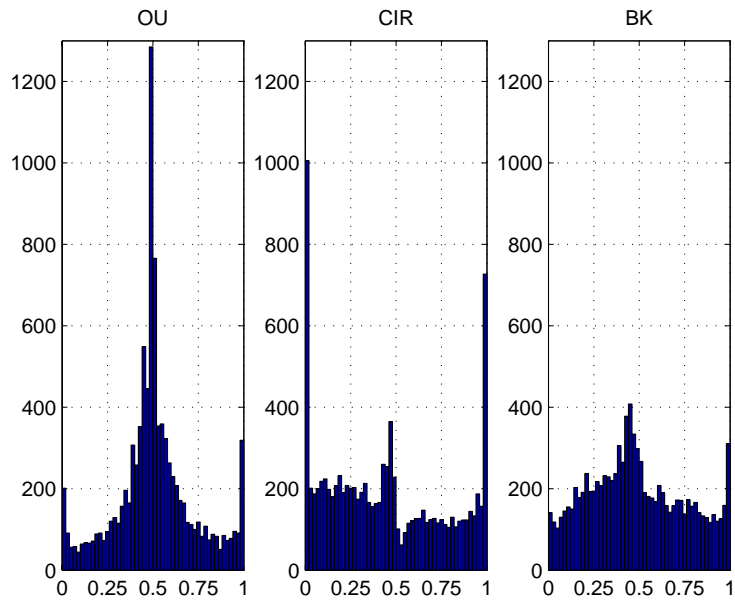


Figure 3: Histogram of generalized residuals, across all firms.

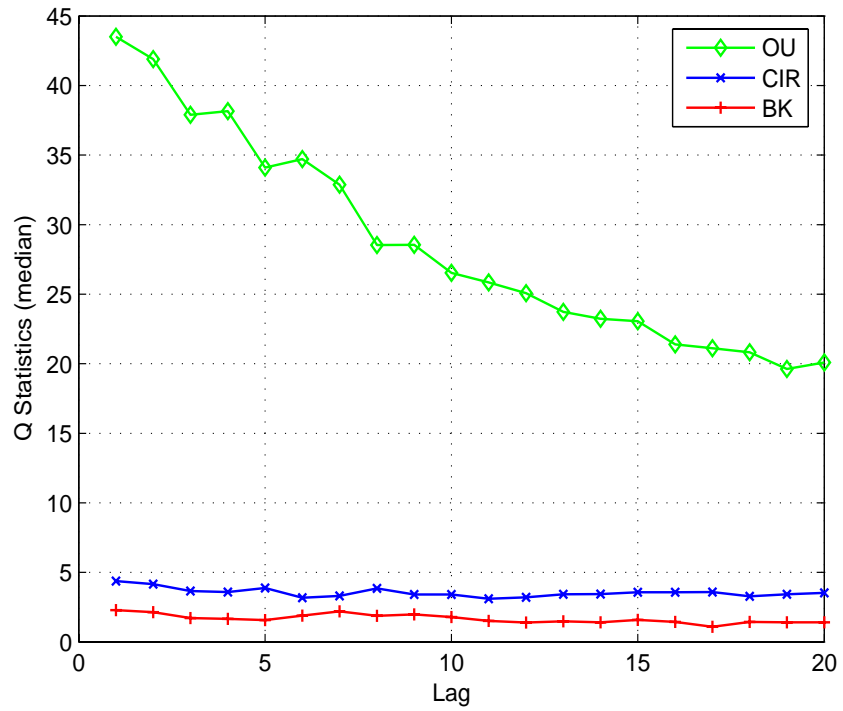


Figure 4: $Q(j)$ test for default intensity model specifications.

Table 8: Rejection rates based on $\hat{Q}(1)$ statistics.

Significance Level	OU	CIR	BK
1%	0.988	0.724	0.408
5%	0.988	0.803	0.539
10%	1.000	0.855	0.553

A Discussion of Model Specifications

In this appendix, we study the model specifications in Table 2 of actual default intensities with regard to (i) their functional form of $G(X_t^i; \Theta)$ in (4), (ii) the transition densities $P(\cdot; X_j^i, \Theta)$, and (iii) simulating missing and censored data from $P(\cdot; X_k^i, X_{k+1}^i, \Theta)$.

A.1 OU Model

In the Ornstein-Uhlenbeck model specification, the state variable X equals λ and follows the stochastic process

$$dX_t = \kappa(\theta - X_t) dt + \sigma dB_t, \quad (\text{A.1})$$

For (A.1), G is available in closed form

$$G(x; \Theta) = 1 - e^{A(1; \Theta) + B(1; \Theta)x},$$

where $k = e^{-\kappa\Delta}$ and

$$\begin{aligned} A(\Delta; \Theta) &= -\theta \left(\Delta + \frac{1-k}{\kappa} \right) + \frac{1}{2} \frac{\sigma^2}{\kappa^2} \left(\Delta + 2\frac{1-k}{\kappa} - \frac{1-k^2}{2\kappa} \right), \\ B(\Delta; \Theta) &= -\frac{1-k}{\kappa}. \end{aligned}$$

The conditional transition probability $P(X_{t+\Delta}; X_t, \Theta)$ is normal with conditional mean $(1-k)\theta + kX_t$ and conditional variance $\frac{\sigma^2}{2\kappa}(1-k^2)$.

We observe that for any time t between times s and u , the conditional distribution of X_t given X_s and X_u is a normal distribution with mean $M(t|s, u)$ and variance $V(t|s, u)$ given by

$$\begin{aligned} M(t|s, u) &= \frac{1 - e^{-2\kappa(u-t)}}{1 - e^{-2\kappa(u-s)}} M(t|s) + \frac{e^{-2\kappa(u-t)} - e^{-2\kappa(u-s)}}{1 - e^{-2\kappa(u-s)}} M(t|u), \\ V(t|s, u) &= \frac{V(t|s)V(u|t)}{V(u|s)}, \end{aligned}$$

where, for times t before u , we let

$$\begin{aligned} M(u|t) &= \theta + e^{-\kappa(u-t)}(X(t) - \theta) \\ V(u|t) &= \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(u-t)}) \\ M(t|u) &= e^{\kappa(u-t)}(X(u) - \theta(1 - e^{-\kappa(u-t)})) \end{aligned}$$

denote the conditional expectation and variance, respectively, of X_u given X_t , and the conditional expectation of X_t given X_u . As a consequence, letting

$Z_k = X(t_k)$, we can easily simulate from the joint conditional distribution of $(Z_{k+1}, \dots, Z_{\bar{k}})$ given Z_k and $Z_{\bar{k}+1}$ which is given by

$$P(Z_{k+1}, \dots, Z_{\bar{k}} | Z_k, Z_{\bar{k}+1}) = P(Z_{k+1} | Z_k, Z_{\bar{k}+1}) \prod_{j=1}^{\bar{k}-(k+1)} P(Z_{k+j+1} | Z_{k+j}, Z_{\bar{k}+1}).$$

We are now in a position to estimate the quantity in (7) by generating some “large” integer number J of independent sample paths $\{(Z_{k+1}^j, \dots, Z_{\bar{k}}^j); 1 \leq j \leq J\}$ from the joint conditional distribution of $(Z_{k+1}, \dots, Z_{\bar{k}})$ given Z_k and $Z_{\bar{k}+1}$, and by computing the fraction of those paths for which $Z_i^j \geq g^{-1}(\zeta)$ for all i in $\{k+1, \dots, \bar{k}\}$.

A.2 CIR Model

In the Cox-Ingersoll-Ross model specification, the state variable X equals λ and follows the stochastic process

$$dX_t = \kappa(\theta - X_t) dt + \sigma\sqrt{X_t} dB_t, \quad (\text{A.2})$$

For (A.1), G is available in closed form

$$G(x; \Theta) = 1 - e^{A(1; \Theta) + B(1; \Theta)x},$$

where (see, for example, Duffie and Gârleanu (2001))

$$B(\Delta; \Theta) = \frac{1 - e^{b_1 s}}{c_1 + d_1 e^{b_1 \Delta}}$$

$$A(\Delta; \Theta) = \frac{m(-c_1 - d_1)}{b_1 c_1 d_1} \log \frac{c_1 + d_1 e^{b_1 s}}{c_1 + d_1} + \frac{m}{c_1} s,$$

where, with $n = -\kappa$, $p = \sigma^2$, and $m = \kappa\theta$, we have

$$c_1 = \frac{-n + \sqrt{n^2 - 2pq}}{2q}$$

$$d_1 = \frac{n + \sqrt{n^2 - 2pq}}{2q}$$

$$b_1 = -\frac{d_1(n + 2qc_1) - (nc_1 + p)}{c_1 - d_1}.$$

A.3 CIRJ Model

The default intensity follows an (CIRJ) process with jumps:

$$dX_t = \kappa(\theta - X_t) dt + \sigma\sqrt{X_t} dB_t + \Delta J_t, \quad (\text{A.3})$$

where B is a standard Brownian motion and J_t is a pure jump process, whose jump sizes are independent and exponentially distributed with mean μ and whose jump times are those of an independent Poisson process with mean jump arrival rate l . The long-run mean is given by $\bar{m} = \theta + l\mu/\kappa$. The s -year conditional survival probability is given by

$$E_t \left[e^{q \int_t^{t+s} \lambda_u du} \right] = 1 - e^{\alpha(s) + \beta(s)\lambda_t},$$

for $q = -10^{-4}$ and where the coefficients $\alpha(s)$ and $\beta(s)$ are given by

$$\begin{aligned} \beta(s) &= \frac{1 - e^{b_1 s}}{c_1 + d_1 e^{b_1 s}} \\ \alpha(s) &= \frac{m(-c_1 - d_1)}{b_1 c_1 d_1} \log \frac{c_1 + d_1 e^{b_1 s}}{c_1 + d_1} + \frac{m}{c_1} s + \\ &\quad \frac{l(a_2 c_2 - d_2)}{b_2 c_2 d_2} \log \frac{c_2 + d_2 e^{b_2 s}}{c_2 + d_2} + \left(\frac{l}{c_2} - l \right) s, \end{aligned}$$

where, with $n = -\kappa$, $p = \sigma^2$, and $m = \kappa\theta$, we have

$$\begin{aligned} c_1 &= \frac{-n + \sqrt{n^2 - 2pq}}{2q} \\ d_1 &= \frac{n + \sqrt{n^2 - 2pq}}{2q} \\ b_1 &= -\frac{d_1(n + 2qc_1) - (nc_1 + p)}{c_1 - d_1} \\ a_2 &= \frac{d_1}{c_1} \\ b_2 &= b_1 \\ c_2 &= 1 - \frac{\mu}{c_1} \\ d_2 &= \frac{d_1 + \mu}{c_1}. \end{aligned}$$

A.4 BK Model

The default intensity follows a Black-Karasinski (**BK**) process:

$$d \log \lambda_t = \kappa(\theta - \log \lambda_t) dt + \sigma dB_t. \quad (\text{A.4})$$

After some preliminary diagnostic analysis of the EDF data set, we opted to specify a model under which the logarithm $X_t^i = \log \lambda_t^i$ of the default intensity of firm i satisfies the Ornstein-Uhlenbeck equation

$$dX_t^i = \kappa(\theta^i - X_t^i) dt + \sigma \left(\sqrt{\rho} dB^c + \sqrt{1 - \rho} dB_t^i \right), \quad (\text{A.5})$$

where $B = (B^c, B^1, \dots, B^I)'$ is a $I + 1$ -dimensional standard Brownian motion, and θ^i , κ , σ , and ρ are constants to be estimated. In particular, we have imposed a joint distribution of EDFs across firms through imposing joint normality of the Brownian motions driving each firm's EDFs, with a constant cross-firm correlation structure. The behavior for $\lambda = e^X$ is sometimes called a Black-Karasinski model.¹

From (A.5), for any time t and time step h (which is $1/12$ in our application), the discretely sampled log-intensity process $X = (X^1, \dots, X^I)'$ satisfies

$$X_{t+h} = b_0 + b_1 X_t + \epsilon_{t+h}, \quad (\text{A.6})$$

where $b_1 = e^{-\kappa h}$, $b_0 = (1 - b_1)\theta$, $\theta = (\theta^1, \dots, \theta^I)$, and $\epsilon_{t+h}, \epsilon_{t+2h}, \dots$ are *iid* normal with mean zero and variance-covariance matrix $\Sigma_\epsilon = \sigma^2(1 - e^{-2\kappa h})/(2\kappa)\Gamma$, where Γ is a $I \times I$ matrix with 1's on the diagonal and ρ everywhere else. In particular we have

$$p(t, t + \Delta t) = g(\lambda_t; \Delta t) \quad (\text{A.7})$$

This leaves us with a vector $\Theta = (\{\theta^i\}, \kappa, \sigma, \rho)$ of unknown parameters to estimate from the available monthly EDF observations of a given firm. In general, given the log-autoregressive form of the default intensity in (A.5), there is no closed-form solution available for the one-year EDF, $1 - p(t, t + 1)$ from (1). We therefore rely on numerical lattice-based calculations of $p(t, t + 1)$. Our current parameter estimates are for the two-stage procedure for constructing trinomial trees proposed by Hull and White (1994).

A.5 Simulating missing and censored data

We suppress Θ in what follows in order to simplify notation. We observe that for any time t between times s and u , the conditional distribution of $X(t)$ given $X(s)$ and $X(u)$ is a normal distribution with mean $M(t | s, u)$ and variance $V(t | s, u)$ given by

$$\begin{aligned} M(t | s, u) &= \frac{1 - e^{-2\kappa(u-t)}}{1 - e^{-2\kappa(u-s)}} M(t | s) + \frac{e^{-2\kappa(u-t)} - e^{-2\kappa(u-s)}}{1 - e^{-2\kappa(u-s)}} M(t | u), \\ V(t | s, u) &= \frac{V(t | s)V(u | t)}{V(u | s)}, \end{aligned}$$

where, for times t before u , we let

$$\begin{aligned} M(u | t) &= \theta + e^{-\kappa(u-t)}(X(t) - \theta) \\ V(u | t) &= \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(u-t)}) \\ M(t | u) &= e^{\kappa(u-t)}(X(u) - \theta(1 - e^{-\kappa(u-t)})) \end{aligned}$$

¹See Black and Karasinski (1991).

denote the conditional expectation and variance, respectively, of $X(u)$ given $X(t)$, and the conditional expectation of $X(t)$ given $X(u)$. As a consequence, letting $Z_k = X(t_k)$, we can easily simulate from the joint conditional distribution of $(Z_{k+1}, \dots, Z_{\bar{k}})$ given Z_k and $Z_{\bar{k}+1}$ which is given by

$$P(Z_{k+1}, \dots, Z_{\bar{k}} | Z_k, Z_{\bar{k}+1}) = P(Z_{k+1} | Z_k, Z_{\bar{k}+1}) \prod_{j=1}^{\bar{k}-(k+1)} P(Z_{k+j+1} | Z_{k+j}, Z_{\bar{k}+1}).$$

We are now in a position to estimate the quantity in (7) by generating some “large” integer number J of independent sample paths $\{(Z_{k+1}^j, \dots, Z_{\bar{k}}^j); 1 \leq j \leq J\}$ from the joint conditional distribution of $(Z_{k+1}, \dots, Z_{\bar{k}})$ given Z_k and $Z_{\bar{k}+1}$, and by computing the fraction of those paths for which $Z_i^j \geq g^{-1}(\zeta)$ for all i in $\{k+1, \dots, \bar{k}\}$.

B Proofs

We will prove the following extension to Lemma 1.

Lemma 2. *Let's fix some time point t_j between t_0 and t_N , and let $\mathcal{J}_1 = \{1, \dots, J_1\}$ denote the subset of firms $\{1, \dots, I\}$ for which we observe the EDF rate at time t_{j-1} and that did not exit our sample at time t_j . Also, let $\mathcal{J}_2 = \{1, \dots, J_2\}$ denote the set of firms in \mathcal{J}_1 that did not exit our sample at time t_{j+1} . Let $\mathcal{S}_1 = \{1, \dots, S_1\}$ ($\mathcal{S}_2 = \{1, \dots, S_2\}$) denote the subset of firms in \mathcal{J}_1 (\mathcal{J}_2) for which we have an EDF observation at time t_j , and let \mathcal{S}_1^c (\mathcal{S}_2^c) be its complement. Then, the conditional distribution of $X_{\mathcal{S}_1^c, j}$ given $X_{\mathcal{J}_1, j-1}$, $X_{\mathcal{S}_1, j}$, and $X_{\mathcal{J}_2, j+1}$ is normal with mean $\mu = \Sigma(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$ and variance-covariance matrix $\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$, where $\Sigma_1 = (A_{22})^{-1}$, $\Sigma_2 = e^{2\kappa h}(B_{22})^{-1}$, and*

$$\begin{aligned} \mu_1 &= \theta_{\mathcal{S}_1^c} + e^{-\kappa h}(X_{\mathcal{S}_1^c, j-1} - \theta_{\mathcal{S}_1^c}) - (A_{22})^{-1}A_{21}\epsilon_{\mathcal{S}_1, j}, \\ \mu_2 &= \theta_{\mathcal{S}_2^c} + e^{\kappa h}(X_{\mathcal{S}_2^c, j+1} - \theta_{\mathcal{S}_2^c}) + e^{\kappa h}(B_{22})^{-1}B_{21}\epsilon_{\mathcal{S}_2, j+1}. \end{aligned}$$

Here, A_{21} and A_{22} are the lower-left $(J_1 - S_1) \times S_1$ and lower-right $(J_1 - S_1) \times (J_1 - S_1)$ submatrix of $\Sigma_{\epsilon, \mathcal{J}_1}^{-1}$, respectively. Similarly, B_{21} and B_{22} are the lower-left $(J_2 - S_2) \times S_2$ and lower-right $(J_2 - S_2) \times (J_2 - S_2)$ submatrix of $\Sigma_{\epsilon, \mathcal{J}_2}^{-1}$, respectively.

Proof. We have

$$f(X_{\mathcal{S}_1^c, j} | X_{\mathcal{J}_1, j-1}, X_{\mathcal{S}_1, j}, X_{\mathcal{J}_2, j+1}) \propto f(X_{\mathcal{S}_1^c, j} | X_{\mathcal{J}_1, j-1}, X_{\mathcal{S}_1, j}) f(X_{\mathcal{J}_2, j+1} | X_{\mathcal{S}_2, j}, X_{\mathcal{S}_2^c, j}). \quad (\text{B.1})$$

For the first term on the right-hand side of (B.1) we have

$$\begin{aligned} X_{S_1^c, j} | X_{\mathcal{J}_1, j-1}, X_{S_1, j} &\sim \theta_{S_1^c} + e^{-\kappa h} (X_{S_1^c, j-1} - \theta_{S_1^c}) + \epsilon_{S_1^c, j} | \epsilon_{S_1, j} \\ &\sim MN(\mu_1, \Sigma_1). \end{aligned} \quad (\text{B.2})$$

Working towards the second term on the right-hand side of (B.1) we know

$$\begin{aligned} X_{\mathcal{J}_2, j+1} | X_{S_2, j}, X_{S_2^c, j} &\sim \theta_{\mathcal{J}_2} + e^{-\kappa h} (X_{\mathcal{J}_2, j} - \theta_{\mathcal{J}_2}) + \epsilon_{\mathcal{J}_2, j+1} \\ &\sim \theta_{\mathcal{J}_2} + e^{-\kappa h} (X_{\mathcal{J}_2, j} - \theta_{\mathcal{J}_2}) + MN(0, \Sigma_{\epsilon, \mathcal{J}_2}). \end{aligned}$$

Hence,

$$\log(f(X_{\mathcal{J}_2, j+1} | X_{S_2, j}, X_{S_2^c, j})) \propto -\frac{1}{2}(\epsilon'_{S_2, j+1}, \epsilon'_{S_2^c, j+1}) \Sigma_{\epsilon, \mathcal{J}_2}^{-1} (\epsilon'_{S_2, j+1}, \epsilon'_{S_2^c, j+1})'$$

In particular, the right-hand side of this equation is

$$\begin{aligned} &-\frac{1}{2}(\epsilon'_{S_2, j+1}, (X_{S_2^c, j+1} - (\theta_{S_2^c} + e^{-\kappa h} (X_{S_2^c, j} - \theta_{S_2^c}))))' \Sigma_{\epsilon, \mathcal{J}_2}^{-1} \\ &\quad (\epsilon'_{S_2, j+1}, (X_{S_2^c, j+1} - (\theta_{S_2^c} + e^{-\kappa h} (X_{S_2^c, j} - \theta_{S_2^c}))))', \end{aligned}$$

which equals, up to a constant,

$$\begin{aligned} &-\frac{1}{2} \left(e^{-2\kappa h} X'_{S_2^c, j} B_{22} X_{S_2^c, j} - 2e^{-\kappa h} X'_{S_2^c, j} B_{21} \epsilon_{S_2, j+1} \right) \\ &-\frac{1}{2} \left(-2e^{-\kappa h} X'_{S_2^c, j} B_{22} (X_{S_2^c, j+1} - \theta_{S_2^c} (1 - e^{-\kappa h})) \right) \end{aligned}$$

Consequently,

$$\begin{aligned} &f(X_{\mathcal{J}_2, j+1} | X_{S_2, j}, X_{S_2^c, j}) \\ &\propto \text{MNpdf}(X_{S_2^c, j}; \theta_{S_2^c} + e^{\kappa h} (X_{S_2^c, j+1} - \theta_{S_2^c}) + e^{\kappa h} B_{22}^{-1} B_{21} \epsilon_{S_2, j+1}, e^{2\kappa h} B_{22}^{-1}) \\ &= \text{MNpdf}(X_{S_2^c, j}; \mu_2, \Sigma_2) \end{aligned} \quad (\text{B.3})$$

From Equations (B.2) and (B.3), we conclude that

$$X_{S_1^c, j} | X_{\mathcal{J}_1, j-1}, X_{S_1, j}, X_{\mathcal{J}_2, j+1} \sim \text{MN}(\mu, \Sigma).$$

□

C Implied Correlation Structure for BK Model

For the BK model in (8) we have

$$\begin{aligned}
\rho^{i,j} &= CORR_t(\lambda_{t+h}^i, \lambda_{t+h}^j) \\
&= \frac{COV_t(\lambda_{t+h}^i, \lambda_{t+h}^j)}{\sqrt{Var_t(\lambda_{t+h}^i)}\sqrt{Var_t(\lambda_{t+h}^j)}} \\
&= \frac{COV_t(\lambda_{t+h}^i, \lambda_{t+h}^j)}{\sqrt{Var_t(\lambda_{t+h}^i)}\sqrt{Var_t(\lambda_{t+h}^j)}} \\
&= \frac{E_t \left[(\lambda_{t+h}^i - e^{m_t^i(h)+1/2v_t^i(h)})(\lambda_{t+h}^j - e^{m_t^j(h)+1/2v_t^j(h)}) \right]}{\sqrt{e^{2m_t^i(h)+v_t^i(h)}(e^{v_t^i(h)} - 1)}\sqrt{e^{2m_t^j(h)+v_t^j(h)}(e^{v_t^j(h)} - 1)}} \\
&= \frac{E_t \left[\lambda_{t+h}^i \lambda_{t+h}^j \right] + e^{m_t^i(h)+1/2v_t^i(h)} e^{m_t^j(h)+1/2v_t^j(h)}}{\sqrt{e^{2m_t^i(h)+v_t^i(h)}(e^{v_t^i(h)} - 1)}\sqrt{e^{2m_t^j(h)+v_t^j(h)}(e^{v_t^j(h)} - 1)}} \\
&= \frac{e^{-1/2(v_t^i(h)+v_t^j(h))} E_t \left[e^{u_{t+h}^i} e^{u_{t+h}^j} \right] - 1}{\sqrt{e^{v_t^i(h)} - 1}\sqrt{e^{v_t^j(h)} - 1}} \\
&= \frac{e^{-1/2(v_t^i(h)+v_t^j(h))} e^{1/2Var(u_{t+h}^i+u_{t+h}^j)} - 1}{\sqrt{e^{v_t^i(h)} - 1}\sqrt{e^{v_t^j(h)} - 1}} \\
&= \frac{e^{-1/2(v_t^i(h)+v_t^j(h))} e^{1/2Var(u_{t+h}^i+u_{t+h}^j)} - 1}{\sqrt{e^{v_t^i(h)} - 1}\sqrt{e^{v_t^j(h)} - 1}} \\
&= \frac{e^{-1/2(v_t^i(h)+v_t^j(h))} e^{1/2(v_t^i(h)+v_t^j(h)+2\rho_{u^i,u^j} \sqrt{v_t^i(h)}\sqrt{v_t^j(h)})} - 1}{\sqrt{e^{v_t^i(h)} - 1}\sqrt{e^{v_t^j(h)} - 1}} \\
&= \frac{e^{\rho_{u^i,u^j} \sqrt{v_t^i(h)}\sqrt{v_t^j(h)}} - 1}{\sqrt{e^{v_t^i(h)} - 1}\sqrt{e^{v_t^j(h)} - 1}},
\end{aligned}$$

where

$$\begin{aligned}
m_t^i(h) &= \theta^i + e^{-\kappa^i h}(\log(\lambda_t^i) - \theta^i) \\
v_t^i(h) &= \sigma^{i2} \frac{1 - e^{-2\kappa^i h}}{2\kappa}.
\end{aligned}$$

In particular, the pairwise correlation between λ^i and λ^j does not depend on θ^i , θ^j or the level of λ_t^i or λ_t^j .

References

- Berndt, A., R. Douglas, D. Duffie, M. Ferguson, and D. Schranz (2005). Measuring Default Risk Premia from Default Swap Rates and EDFs. Working Paper, Stanford University.
- Black, F. and P. Karasinski (1991). Bond and Option Pricing when Short Rates are Log-Normal. *Financial Analysts Journal*, 52–59.
- Black, F. and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81**, 637-654.
- Bohn, J., N. Arora, and I. Korbalev (2005). Power and Level Validation of the EDF CREDIT Measure in the U.S. Market. Working Paper, Moody's KMV.
- Crosbie, P. and J. Bohn (2001). Modeling Default Risk. Working Paper, KMV.
- Dempster, A. P., N. M. Lair, and D. B. Rubin (1977). Maximum Likelihood from Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1-38.
- Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* **57**, 405-43.
- Duffie, D. and N. Gârleanu (2001). Risk and Valuation of Collateralized Debt Obligations. *Financial Analysts Journal* **57**, January–February, 41–62.
- Duffie, D., L. Saita, and K. Wang (2005). Multiperiod Corporate Default Probabilities with Stochastic Covariates. forthcoming, *Journal of Financial Economics*.
- Duffie, D. J. and R. Kan (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance* **6**, 379-406.
- Hong, Y. and H. Li (2005). Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates. *RFS* **18**, 37-84.
- Hull, J. and A. White (1994). Numerical Procedures for Implementing Term Structure Models I: Single Factor Models. *Journal of Derivatives* **2**, 7-16.

- Kealhofer, S. (2003). Quantifying Credit Risk I: Default Prediction. *Financial Analysts Journal*, January–February, 30–44.
- Kurbat, M. and I. Korbalev (2002). Methodology for Testing the Level of the EDF Credit Measure. Working Paper, Technical Report 020729, Moody's KMV.
- Liu, J. S. (2003). *Monte Carlo Strategies in Scientific Computing*. New York, NY: Springer Series in Statistics.
- Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance* **29**, 449-70.