On-line Appendix for

Anchoring Credit Default Swap Spreads to Firm Fundamentals

Jennie Bai[1] and Liuren Wu[2]

A. Variations in structural model implementations

Merton (1974) assumes that the total asset value (A) of a company follows a geometric Brownian motion with instantaneous return volatility σ_A, the company has a zero-coupon debt with principal D and time-to-maturity T, and the firm's equity (E) is a European call option on the firm's asset value with maturity equal to the debt maturity and strike equal to the debt principal. The company defaults if its asset value is less than the debt principal at the debt maturity. These assumptions lead to the following two equations that link the firm's equity value E and equity return volatility σ_E to its asset value A and asset return volatility σ_A,

$$E = A \cdot N(d + \sigma_A \sqrt{T}) - D \cdot N(d), \tag{1}$$

$$\sigma_E = N(d + \sigma_A \sqrt{T}) \sigma_A A / E. \tag{2}$$

Equation (1) is the European call option valuation formula that treats the equity as a European call option on the company's asset value with strike equal to the debt principle D and expiration equal to the debt maturity date T. Equation (2) is derived from equation (1) and provides a link between the equity return volatility (σ_E) and the asset return volatility (σ_A). In the two equations, $N(\cdot)$ denotes the cumulative normal density and d is a standardized measure of *distance to default*,

$$d = \frac{\ln(A/D) + (r - \frac{1}{2}\sigma_A^2)T}{\sigma_A \sqrt{T}}, \tag{3}$$

with r denoting the instantaneous riskfree rate.

[1]Georgetown University, McDonough School of Business, 37th and O Streets, Washington, DC 20057; tel: +1-202-687-5695; email: jennie.bai@georgetown.edu.

[2]Baruch College, Zicklin School of Business, One Bernard Baruch Way, Box B10-225, New York, NY 10010; tel: +1-646-312-3509; fax: +1-646-312-3451; email: liuren.wu@baruch.cuny.edu.

The Merton (1974) model is highly stylized in its assumptions on both the asset value dynamics and the capital structure. For asset value dynamics, well-documented discontinuous price movements and stochastic volatilities are ignored. For the capital structure, most companies have more than just a zero-coupon bond. Despite its stylized nature, the model captures two major determinants of credit risk — financial leverage and business risk — and combines them into a standardized distance-to-default measure, which normalizes the financial leverage (log asset to debt ratio, $\ln(A/D)$) by the asset return volatility over the maturity of the debt ($\sigma_A \sqrt{T}$). It is well known that given the same financial leverage, firms with riskier business operations can have a higher chance of default. The distance-to-default measure normalizes the financial leverage by the business risk so that it reflects the number of standard deviations that the asset value is away from the debt principal.[3] The standardized measure becomes comparable across firms that have different levels of business risk.

To compute a firm's distance to default, we take the company's market capitalization as its equity value E, the company's total debt as a proxy for the principal of the zero-coupon bond D, and the one-year realized stock return volatility as an estimator for stock return volatility σ_E. We further assume zero interest rates ($r = 0$) and fix the debt maturity at $T = 10$ years for all firms. Since our focus is on the cross-sectional difference across firms, choosing any particular interest rate level for r or simply setting it to zero generates negligible impacts on the cross-sectional performance. We solve for the firm's asset value A and asset return volatility σ_A from the two equations in (1) and (2) via an iterative procedure.

The literature has implemented the Merton (1974) model in several variations. This section discusses the rationale behind our particular choices and investigates the effects of these variations on the cross-sectional-explanatory power of the resulting MCDS and WCDS.

[3] Under the assumed dynamics, the log asset value $\ln A_T$ has a normal distribution with a risk-neutral mean of $\mu = \ln A_t + (r - \frac{1}{2}\sigma_A^2)T$ and a variance of $V = \sigma_A^2 T$. Hence, the negative of the distance to default $-d = (\ln D - \mu)/\sqrt{V}$ can be formally interpreted as the number of standard deviations by which the log debt principal exceeds the mean of the terminal log asset value.

A.1. Maturity choices

By regarding equity as an option on the asset, the Merton model uses the option maturity T to control the relative contribution of asset volatility to the equity value and hence default probability. We choose a relatively long option maturity to give more weight to the asset volatility in the determination of the default probability. By contrast, the KMV implementation as documented in Crosbie and Bohn (2003) sets the maturity T to one year, matching their default probability forecasting horizon. In Panel A of Table 1, we show how the maturity choice affects the cross-sectional explanatory power of the model on the five-year CDS spread. As in the main text of the paper, each date, we use a random half of the universe to calibrate the model and generate CDS valuations on the whole universe. The left side in the panel reports the time-series average of the cross-sectional R-squared estimates on the in-sample half of the universe for both MCDS and WCDS valuation. The right side reports the corresponding average R-squared estimates on the out-of-sample half of the universe.

[Table 1 about here.]

The results in panel A of Table 1 show that choosing a short maturity such as one year in the Merton model can generate significantly lower cross-sectional explanatory power on the five-year CDS spreads. The performance improves and reaches a plateau once the chosen maturity is five years or longer.

A.2. Debt principal proxy

KMV proposes to approximate the debt principal by the firm's current liabilities plus one half of its long-term debt, but we find that using total debt generates slightly better performance than the KMV choice for explaining the cross-sectional variation of the five-year CDS. Nevertheless, we accommodate alternative leverage measures in our WCDS construction. In panel B of Table 1, we compare our implementation ("Total Debt") with the implementation of using current liability plus half of long-term liability ("CL+LL/2"). In the latter case, we use the ratio of total debt to market capitalization to replace the ratio of current liability plus half of long-term liability to the market capitalization as the alternative leverage measure. The results show that using current liability plus half of long-term liability generates lower cross-sectional explanatory power for MCDS; however, by accommodating the other measure as an alternative leverage measure in the WCDS construction, the two choices generate similar performances for WCDS.

A.3. Solving for asset value and asset return volatility

To solve for the firm's asset value and asset return volatility, several studies, e.g., Crosbie and Bohn (2003), Vassalou and Xing (2004), Duffie, Saita, and Wang (2007), and Bharath and Shumway (2008), propose a more computationally intensive iterative approach that involves the time series history: Starting with an initial guess on asset return volatility σ_A, they solve for the history of the firm's asset value based on the firm's equity value history and equation (1). Then, they estimate the asset return volatility σ_A from log changes on the solved asset value series. Duan (1994) and Ericsson and Reneby (2005) propose to embed this iterative procedure into a maximum likelihood framework.

Computing σ_A from the asset value time series can be problematic when changes in the asset value are induced by financing or investment decisions instead of operating activities. In such cases, returns on assets, upon which the asset return volatility σ_A should be measured, can be quite different from log changes in the asset value series. Duan, Sun, and Wang (2012) propose to scale the asset value by the corresponding book value as an approximate correction for this issue. Directly solving the two equations (1) and (2) completely circumvents this scaling issue.

In panel C of Table 1, we analyze the effect of different asset return volatility calculation methods. We label our method as the two-equations-two-unknowns (TETU) approach, and we consider two alternative approaches based on asset value history (AVH). In AVH-I, we take the suggestion from Duan, Sun, and Wang (2012) and scale the extracted asset value by the corresponding book value of total asset. We compute log returns on this scaled asset value and estimate the sample standard deviation of the return series over the past year as the asset return volatility. The cross-sectional explanatory power of MCDS computed from this approach averages around 57% and 55% for the two sub-samples, respectively, markedly lower than the average performance of the MCDS computed from our TETU approach at 65% and 64%.

Scaling the asset value by its book value removes the impact of financing activities such as mergers and issuance or retirement of stocks or debt, but it also has the un-intended effect of mitigating the operating activities because the book value of equity can change due to both financing activities (e.g., issuing or retiring stock) and operating activities (retained earnings). The latter should not be excluded from the asset value change as it is an integral part of the return on asset. As a result, the sample standard deviation computed from the scaled asset value time series may not accurately reflect the true asset return volatility.

In AVH-II, we do not use the asset value history directly, but rather rely on the stock return history and adjust the stock return by the asset-stock sensitivity. Specifically, at each date, we first compute the daily stock total return based on the adjusted stock price series (adjusted for splits and dividend payments), and then convert this stock return into asset return based on the hedge ratio implied by the Merton model:

$$R^A_{t+1} = \left[R^E_{t+1}\right] \left[\frac{E_t}{A_t} \cdot \frac{1}{N(d_t + \sigma_A \sqrt{T})}\right], \tag{4}$$

where R^E_{t+1} in the first bracket denotes the daily stock return, and the second bracket contains the hedge ratio from the Merton (1974) model. We compute the asset return volatility σ_A from this asset return (R^A_{t+1}) time series. The return relation in (4) is analogous to the volatility relation in (2) used in our TETU approach. Had the hedging ratio (the second bracket in equation (4)) been a constant, this AVH-II approach would have generated the same result as the TETU approach. The fact that the hedging ratio varies over time can lead to some differences. The last row in panel C of Table 1 shows that with the AVH-II approach, the the cross-sectional explanatory power for MCDS becomes 65% for the in-sample half, and 63% for the out-of-sample half, very close to our TETU approach. Given the similar performance, we choose the TETU approach because it is much less computationally intensive and it fits our cross-sectional focus better.

A.4. Equity as a barrier option

A commonly proposed alternative to the Merton (1974) model is to assume as in Leland (1994) and Leland and Toft (1996) that the firm can default any time before the debt maturity when the firm's asset value falls below a certain threshold. In this case, equity becomes a call option on the firm's asset value with a knock-out barrier. If one sets the barrier to the debt principal value and assumes zero rates, the model implies that the equity value is equal to the intrinsic value of the knock-out barrier option, $E = A - D$, and the market value of debt is equal to its principal value. Equity return volatility no longer plays a role in determining the firm's asset value. The "naive" Merton alternative proposed in Bharath and Shumway (2008) can be justified under this barrier option assumption. In panel D of Table 1, we compare the Merton model performance with this alternative on the simple implementation of the Leland model. This simple alternative performs better than the Merton model implementation with $T = 1$, but worse than our Merton model implementation with $T = 10$.

B. Local polynomial regressions and bandwidth choice

In constructing MCDS and WCDS, we have used local polynomial regressions to capture the potentially nonlinear relations between two variables x and y,

$$y = f(x) + e, \tag{5}$$

where $f(\cdot)$ denotes the local polynomial form and e denotes the regression error.

To illustrate the specifics of the regression, suppose that we have N observations of the pair $(y_i, x_i)_{i=1}^N$ and we intend to generate an estimate for $\hat{y}_k$ at $x = x_k$. The estimate $\hat{y}_k$ is obtained from a weighted least square regression,

$$\hat{y}_k = X_k(X^\top W_k X)^{-1} X^\top W_k y, \tag{6}$$

where y denotes the $N \times 1$ vector of the dependent variable observations and X is a matrix made of polynomials of x. Our MCDS construction involves a local quadratic regression, in which case $X = [1, x, x^2]$ is an $N \times 3$ matrix. Our WCDS construction involves many local linear regressions, in which case $X = [1, x]$ is an $N \times 2$ matrix. To construct the weighting matrix W_k, we use a Gaussian kernel, with W_k being a diagonal matrix with the ith diagonal element given by,

$$W_k^i = \exp\left(-\frac{(x_i - x_k)^2}{2h^2}\right), \tag{7}$$

where h is the bandwidth that controls the relative weighting of different observations. Intuitively, observations that are closer to x_k obtain higher weights. The weight declines as the distance $|x_i - x_k|$ increases, with the declining speed controlled by the bandwidth h. A higher bandwidth assigns more uniform weights across observations and thus generates more smoothing. In the limit of $h \to \infty$, W_k becomes an identity matrix and we are essentially running just one global regression.

Under normal distribution assumptions on x and the Gaussian kernel, textbooks, e.g., Simonoff (1996), often propose a default optimal bandwidth choice that balances between smoothing and fitting,

$$\hat{h} = (4/3)^{(1/5)} \sigma_x / N^{(1/5)}, \tag{8}$$

where σ_x denotes the standard deviation of x.

To show how the bandwidth choice affects the cross-sectional performance both in sample and out of sample, we repeat the exercise under different bandwidth choices while setting the maturity to ten years. Table 2 shows the effects of the bandwidth choice on the cross-sectional explanatory powers of MCDS and WCDS. When we set the bandwidth to half of the default choice, the in-sample fitting for both MCDS and WCDS becomes better, but the out-of-sample becomes worse, showing signs of out-of-sample instability. The default bandwidth choice generates both good in-sample and out-of-sample performances. As we set the bandwidth to twice the default bandwidth or higher, the in-sample and out-of-sample performance becomes very much similar to each other. In our main text analysis, we set the bandwidth to twice the default choice.

[Table 2 about here.]

A related choice for the local polynomial regression is the order of the polynomial. If the relations are reasonably flat, one can choose an order of zero and essentially perform local averages around each target region. However, when the relation has a steep slope, a local average regression can generate biases at the boundaries when the averaging can only use data from one side. In this case, a local linear regression can help reduce the bias at the boundaries. We use the local linear regression to link $\ln(\text{CDS}/\text{MCDS})$ to each additional firm characteristic. When we map $\ln(\text{RCDS})$ to the market observation $\ln(\text{CDS})$, we choose a local quadratic regression because the observed relations between the two often show a convex shape. The second-order polynomial helps capture this observed curvature well even with a large bandwidth.

References

Bharath, S. T., and T. Shumway. “Forecasting Default with the Merton Distance to Default Model.” *Review of Financial Studies*, 21 (2008), 1339–1369.

Crosbie, P., and J. Bohn. “Modeling Default Risk.” Working paper, Moodys KMV (2003).

Duan, J.-C. “Maximum Likelihood Estimation Using Price Data of the Derivative Contract.” *Mathematical Finance*, 4 (1994), 155–167.

Duan, J.-C.; J. Sun; and T. Wang. “Multiperiod Corporate Default Prediction — A Forward Intensity Approach.” *Journal of Econometrics*, 170 (2012), 191–209.

Duffie, D.; L. Saita; and K. Wang. “Multi-period Corporate Default Prediction with Stochastic Covariates.” *Journal of Financial Economics*, 83 (2007), 635–665.

Ericsson, J., and J. Reneby. “Estimating Structural Bond Pricing Models.” *Journal of Business*, 78 (2005), 707–735.

Leland, H. E. “Risky Debt, Bond Covenants and Optimal Capital Structure.” *Journal of Finance*, 49 (1994), 1213–1252.

Leland, H. E., and K. B. Toft. “Optimal Capital Structure, Endogenous Bankruptcy and the Term Structure of Credit Spreads.” *Journal of Finance*, 51 (1996), 987–1019.

Merton, R. C. “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates.” *Journal of Finance*, 29 (1974), 449–470.

Simonoff, J. S. *Smoothing Methods in Statistics*. Springer-Verlag, New York (1996).

Vassalou, M., and Y. Xing. “Default Risk in Equity Returns.” *Journal of Finance*, 59 (2004), 831–868.

TABLE 1

Variations in structural model implementations and cross-sectional explanatory power

Entries report the time-series averages of the R-squared estimates from cross-sectional regressions of log market CDS quotes on the logarithm of MCDS and WCDS, respectively. Each day, we take a random half of the universe for calibration and generate predictions on the whole universe. The left side reports the average R-squared on the in-sample half of the universe whereas the right side reports the average R-squared on the out-of-sample half of the universe. Panel A shows the effect of maturity choice in the Merton model implementation. Panel B shows the effect of debt principal choice either as total debt or current liability (plus half of long-term liability (CL+LL/2). Panel C shows the effect of different asset return volatility calculation methods. Our default implementation solves the asset return volatility together with the asset value from two equations implied by the Merton model. We label it as the two-equations-two-unknowns (TETU) approach. We also consider two alternative approaches, where the asset return volatility is estimated as the sample annualized standard deviation of daily log returns on the extracted asset value over the past year. We label them as asset-value-history (AVH) approaches I and II. Finally, panel D compares the CDS computed from the Merton model to that from a simplified version of the Leland model, where equity is a barrier option on the asset and is valued as the difference between the asset value and the debt principal.

Choices	In-sample R^2		Out-of-sample R^2	
	MCDS	WCDS	MCDS	WCDS
Panel A. Maturity choice				
1	0.40	0.71	0.37	0.66
3	0.58	0.75	0.56	0.71
5	0.62	0.76	0.61	0.73
10	0.65	0.77	0.64	0.74
15	0.66	0.78	0.65	0.75
20	0.66	0.78	0.65	0.75
Panel B. Debt principal choice				
Total Debt	0.65	0.77	0.64	0.74
CL+LL/2	0.62	0.77	0.62	0.75
Panel C. Asset return volatility estimation method choice				
TETU	0.65	0.77	0.64	0.74
AVH-I	0.57	0.74	0.55	0.70
AVH-II	0.65	0.77	0.63	0.73
Panel D. Different model assumptions				
Merton	0.65	0.77	0.64	0.74
Leland	0.63	0.75	0.61	0.70

TABLE 2

Bandwidth choice in local polynomial regressions and the cross-sectional explanatory power

Entries report the time-series averages of the R-squared estimates from cross-sectional regressions of log market CDS quotes on the logarithm of MCDS and WCDS, respectively. Each day, we take a random half of universe for calibration and generate predictions on the whole universe. The left side reports the average R-squared on the in-sample half of the universe whereas the right side reports the average R-squared on the out-of-sample half of the universe. The performance shows the effects of bandwidth choice for the local quadratic and local linear regressions in the MCDS and WCDS construction, with $\hat{h}$ denoting the textbook default choice.

Bandwidth	In-sample R^2		Out-of-sample R^2	
Choice	MCDS	WCDS	MCDS	WCDS
$.5\hat{h}$	0.67	0.81	0.63	0.71
$1\hat{h}$	0.66	0.79	0.64	0.74
$2\hat{h}$	0.65	0.77	0.64	0.74
$3\hat{h}$	0.65	0.76	0.64	0.74