Online Appendix:

The Cross-Section of Recovery Rates and Default Probabilities Implied by Credit Default Swap Spreads

Redouane Elkamhi	Kris Jacobs	Xuhui Pan
University of Toronto	University of Houston	Tulane University

June 25, 2014

This online appendix proceeds as follows. Section I discusses the economic determinants of the firm-specific latent factor. Section II discusses potential identification problems when estimating recovery rates, and also presents a Monte Carlo analysis that addresses potential identification problems. Section III provides additional information on data used in the regressions that document the determinants of recovery rates. Section IV presents a figure with additional information on model errors. Section V presents a figure that illustrates the importance of reliable estimates of recovery rates. Finally, Section VI documents the cross-sectional relationship between recovery rates and default probabilities.

I Economic Determinants of the Risky Term Structure

For models with latent variables, a central question is what the latent variables capture. We answer this question by regressing the estimated firm-specific latent factor on economic and financial variables that are known to explain credit spreads. Table 1 reports the results of this exercise. To motivate the use of the regressors, we also present regressions of the CDS spreads on the same variables. Finally, we also regress the errors on the same variables, to provide additional insight into the error structure, and to investigate if any economic variable is highly correlated with the model errors.

Table 1 presents estimation results for two regressions. Identifying economic and financial variables that explain the spreads and latent factors is not our main objective, and we limit ourselves to variables that have been used in existing studies. The first regression, reported in the first three columns, exclusively uses the variables suggested by the Merton (1974) model and other structural models of credit risk: interest rates, leverage, and volatility. Interest rates are captured by the ten-year Treasury yield. Volatility is measured using a simple exponentially weighted moving average of squared returns. Following the literature on the determinants of CDS spreads, leverage is measured as long term debt, Compustat variable DLTT, divided by the sum of long term debt and market value of equity, which are the product of stock price (Compustat variable PRCCD) and shares outstanding (Compustat variable CSHOC). The second regression includes a number of other explanatory variables suggested by the credit risk literature:[1] the slope of the term structure, which is defined as the spread between ten-year and two-year Treasury yield, the VIX volatility index, the S&P500 return, liquidity, which we measure using the Pastor-Stambaugh (2003) equity market liquidity variable. Treasury yields are from the Federal Reserve Board, and the Pastor-Stambaugh liquidity is obtained from Stambaugh's website.

Columns 1 and 4 report on regressions for the CDS spread. Columns 2 and 5 report on the latent variable, and columns 3 and 6 on the errors. We run the regressions on a firm-by-firm basis and report the median of the point estimates and standard errors. Standard errors are computed using a Newey-West correction with the integer value of $4 * (T/100)^{2/9}$ lags. Because there is only one latent factor per firm, we only run one regression per firm in the case of columns 2 and 5. For the CDS premia and the squared errors, we run three regressions per firm, one for each maturity, and we report the medians for the distribution of all estimated coefficients.

The results in column 1 indicate that interest rates, leverage, and volatility are all sta-

[1] See for example Collin-Dufresne, Goldstein, and Martin (2001), Bakshi, Madan, and Zhang (2006a), Ericsson, Jacobs, and Oviedo (2009), Tang and Yan (2010), and Zhang, Zhou, and Zhu (2009) for studies that use these variables to explain credit spreads.

tistically significant determinants of the CDS spreads, although volatility is only significant at the 10% level. The median R^2 is 43.27%. These results are consistent with existing findings, see for instance Ericsson, Jacobs, and Oviedo (2009). The median point estimates are also roughly consistent with estimates from existing papers. Column 4 indicates that from the other economic variables, only the VIX is statistically significant. Despite this, the increase in R^2 is substantial. Columns 2 and 5 indicate that, not surprisingly, the latent factor attempts to capture the movements in these variables; from a statistical perspective the emphasis hereby is on the time-variation in leverage. The economic and financial variables explain a large proportion of the variation in the latent factor.

The regression analysis for the errors in columns 3 and 6 demonstrates that the errors are not statistically significantly related to the economic and financial variables, suggesting that the latent factor does an adequate job of capturing the time variation in these variables. We use squared errors in the regression because the loss function used in estimation consists of squared errors. However, the R^2 indicates that some of the variation in the errors can be explained by the economic and financial variables. These results suggest that some gains in model fit may be possible by including a second latent factor, but that the resulting gain is not likely to be spectacular. Since our objective is to estimate recovery rates rather than to optimize model fit, we do not further pursue this.

II Monte Carlo Analysis

While it has occasionally been argued that recovery rates cannot be identified from credit-risky securities such as corporate bonds or credit default swaps, recent studies have argued conclusively that under the assumption of recovery of face value, identification is possible. See for instance Pan and Singleton (2008) and Schneider, Sogner, and Veza (2010) for discussions. Pan and Singleton (2008) argue that the use of multiple tenors in estimation helps identification, and for this reason we simultaneously use three tenors in estimation.

The issue of econometric identification is a complex one, and the same terminology is often used to refer to two somewhat different issues. The first type of identification problem is one where the theoretical model is ill-suited for econometric estimation, in the sense that no matter how many data are brought to bear on the question, certain parameters cannot be identified. In the credit risk literature, this is the case under the recovery of market value assumption, in which case the recovery rate cannot be identified. See for instance Duffie and Singleton (1999) for a discussion. This problem is akin to trying to identify two unknowns from one equation. In the case of the recovery of market value, this occurs because a higher recovery rate can be compensated by a lower default probability, while yielding the same price for the risky security.

The second type of econometric identification problem is more subtle, and much more prevalent. It concerns a situation where mathematically the problem is tractable, and in principle the parameters can be identified, but the problem and the available data are such that it is difficult to precisely identify and estimate the parameters of the problem. When pricing credit risky securities, the recovery rate and the default probability are both critically important, and this will be reflected in the statistical loss function used in estimation. The loss function, which can be a log-likelihood, or a sum of squares, may not be very informative in certain directions. Therefore, estimation of both default probabilities and recovery rates jointly from credit risky securities will always be fraught with this type of identification problem. The severity of the identification problem depends on many issues, such as the theoretical model, the available data, and the estimation method. This does not mean that the parameters cannot be estimated. It does mean, however, that care must be taken in the interpretation of the results, and that one has to carefully check for evidence of model misspecification and identification problems.

We perform a Monte Carlo analysis to assess the robustness of our estimation methodology and to detect potential identification problems of the second type. We simulate time series of three years worth of daily CDS spreads by Monte Carlo for a typical firm in our sample, for all three tenors. We subsequently perturb the parameters of the data generating process by adding a random noise, drawing from a normal distribution with zero mean and standard deviation equal to two standard deviation of the empirical distribution of parameters. We use the resulting parameter values as starting values for the numerical search that fits the simulated CDS spreads. We repeat this experiment one hundred times.

Table 2 shows the resulting parameter distribution. It is very tightly distributed around the parameters of the data generating process. The averages of the estimated recovery rates are very close to the true recovery rate, and the same applies to the parameters governing default probabilities. None of the t-statistics for the differences between the estimated parameters and the true parameters are greater than 0.5, suggesting differences are not statistically significant. This Monte Carlo experiment confirms that our econometric methodology is able to reliably estimate model parameters, and recovery rates in particular. We conclude that recovery rates and default probabilities are adequately identified in our econometric implementation.

III The Determinants of Recovery Rates: Data Sources

This section provides additional detail on data used in the regressions that document the determinants of recovery rates.

Panel A of Table 3 presents descriptive statistics for the explanatory variables, and Panel B presents a correlation matrix for the explanatory variables. The firm-specific variables are obtained from Compustat. Following Acharya, Bharath, and Srinivasan (2007), leverage is defined as long term debt, Compustat variable DLTT, divided by total assets, Compustat variable AT. Firm size is defined as the natural logarithm of total assets. Tangibility is defined as the ratio of property, plant and equipment, Compustat variable PPEGT, to total assets.

The industry variables are also computed from Compustat. We use two industry distress variables. The first is a dummy variable that takes the value one if the median stock return of all the firms in the 3-digit NAICS industry code is less than -20% in a fiscal year, where returns are calculated using the Compustat variable PRCC. The second industry distress variable is a dummy variable that takes the value one if the median sales growth in a 3-digit NAICS code industry is negative for any of the last two years.[2] Sales is measured using the Compustat SALE variable. We also define the industry's Q-ratio as the median of the ratio of the market value to book value of all the firms in the 3-digit NAICS code, measured using the MKVALT and AT variables; we proxy industry illiquidity by the inverse of the quick ratio for the industry, computed as current assets minus inventory divided by current liabilities, for which we use the variables ACT, INVT, and LCT. We define median industry

[2] We also investigated the other distress measures proposed by Acharya, Bharath, and Srinivasan (2007), but for our sample, these did not result in realistic proportions of distressed firms in the sample.

leverage as the industry median ratio of long term debt to total assets, computed using all firms in the 3-digit NAICS code. Industry specificity is defined as the median ratio of machinery and equipment to total assets of all firms in the 3-digit NAICS code, using the variables FATE and AT. Finally, the number of peer firms in the industry is defined using the 3-digit NAICS code.

IV Model Errors

Figure 1 presents the ratio of the RMSE for the model with estimated recovery and the model with 40% recovery, for all three tenors. Note that even though the model with estimated recovery rate nests the model with 40% recovery rate, for some firms the ratio is larger than one for one of the maturities because we use all three maturities jointly in estimation. The most important conclusion from Figure 1 is that for many firms, the RMSE for the model with estimated recovery rate is only a fraction of the RMSE for the model with 40% recovery rate.

V The Impact of Default Probabilities and Recovery Rates on CDS Premia

To motivate the importance of reliable estimates of the recovery rate, consider Figure 2. It depicts how CDS premia for a typical A-rated company (top) and a typical B-rated company (bottom) are affected by survival probabilities and recovery rates. The benchmark pricing, based on our estimation results, yields a CDS premium of 25.81 basispoints for the A-rated company and 416.94 for the B-rated company. The grey line indicates how deviations from the benchmark values of survival probabilities, measured in percentages on the X-axis, affect CDS premiums, and the black line does the same for recovery rates. Both figures indicate that an incorrect estimate of the recovery rate affects the CDS premium even more than an incorrect estimate of the survival probability, especially for large deviations. Note that the resulting changes in CDS premiums are large compared to the RMSEs in our empirical exercise, which are 0.873 basis points on average for a typical A-rated company and 40.025 basispoints for a typical B-rated company.

VI The Cross-Sectional Relation Between Recovery Rates and Default Probabilities

Figure 3 presents estimated recoveries and average five-year default probabilities for all 152 firms. The correlation between average default probabilities and recoveries is positive, at 38.79%. Correlations for one-year and three-year default probabilities (not reported) are also positive, but somewhat lower. Note that Figure 3 represents the cross-sectional dependence between the recovery rate, which is assumed constant over time, and the average of the default probability, which is time varying. Different results may obtain if one estimates both the recovery rate and the default probability as time-varying.

References

Acharya, V.; S. Bharath; and A. Srinivasan. "Does Industry-Wide Distress Affect Defaulted Firms? Evidence from Creditor Recoveries." *Journal of Financial Economics,* 85 (2007), 787–821.

Bakshi, G.; D. Madan; and F. Zhang. "Investigating the Role of Systematic and Firm-Specific Factors in Default Risk: Lessons from Empirically Evaluating Credit Risk Models." *Journal of Business,* 79 (2006a), 1955-1987.

Collin-Dufresne, P.; R. Goldstein; and S. Martin. "The Determinants of Credit Spreads." *Journal of Finance,* 56 (2001), 2177-2207.

Duffie, D., and K. Singleton. "Modeling Term Structures of Defaultable Bonds." *Review of Financial Studies,* 12 (1999), 687-720.

Ericsson, J.; K. Jacobs; and R. Oviedo. "The Determinants of Credit Default Swap Premia." *Journal of Financial and Quantitative Analysis,* 44 (2009), 109-132.

Longstaff, F.; S. Mithal; and E. Neis. "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market." *Journal of Finance* 60 (2005), 2213-2253.

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

Pan, J., and K. Singleton. "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads." *Journal of Finance*, 63 (2008), 2345-2384.

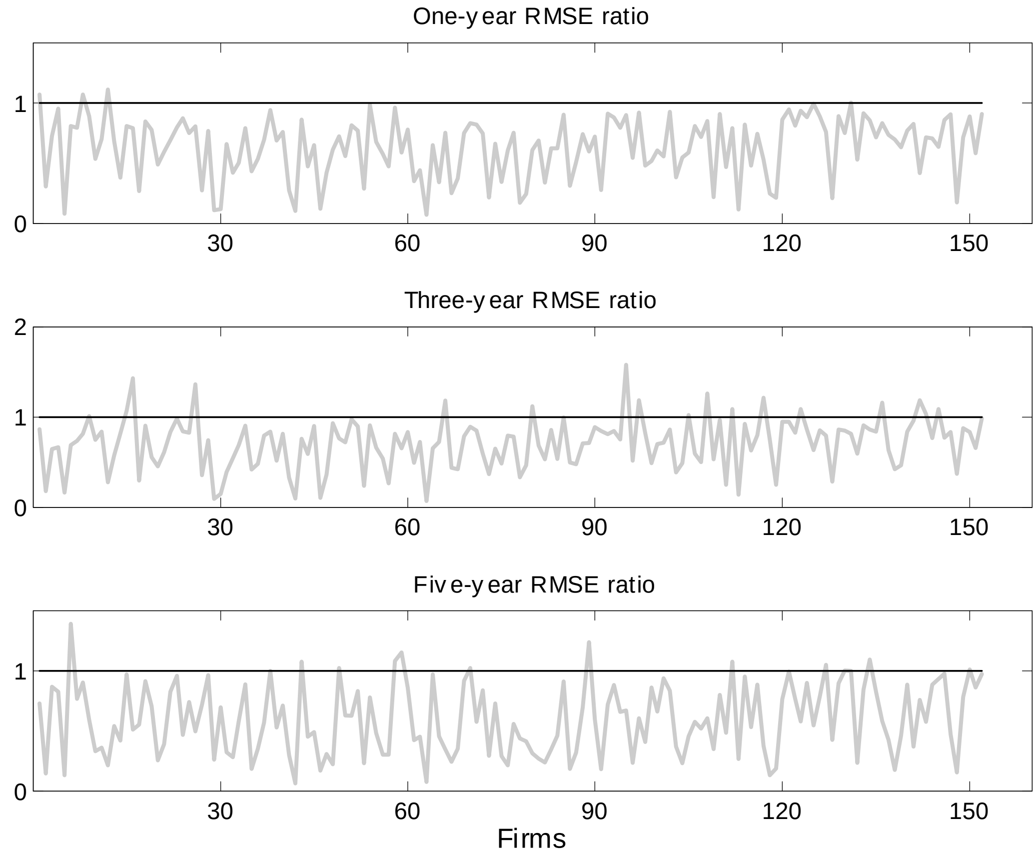
Pastor, L., and R. Stambaugh. "Liquidity Risk And Expected Stock Returns." *Journal of Political Economy*, 111 (2003), 642-685.

Schneider, P.; L. Sogner; and T. Veza. "The Economic Role of Jumps and Recovery Rates in the Market for Corporate Default Risk." *Journal of Financial and Quantitative Analysis*, 45 (2010), 1517-1547.

Tang, D., and H. Yan. "Market Conditions, Default Risk, and Credit Spreads." *Journal of Banking and Finance,* 34 (2010), 743–753.

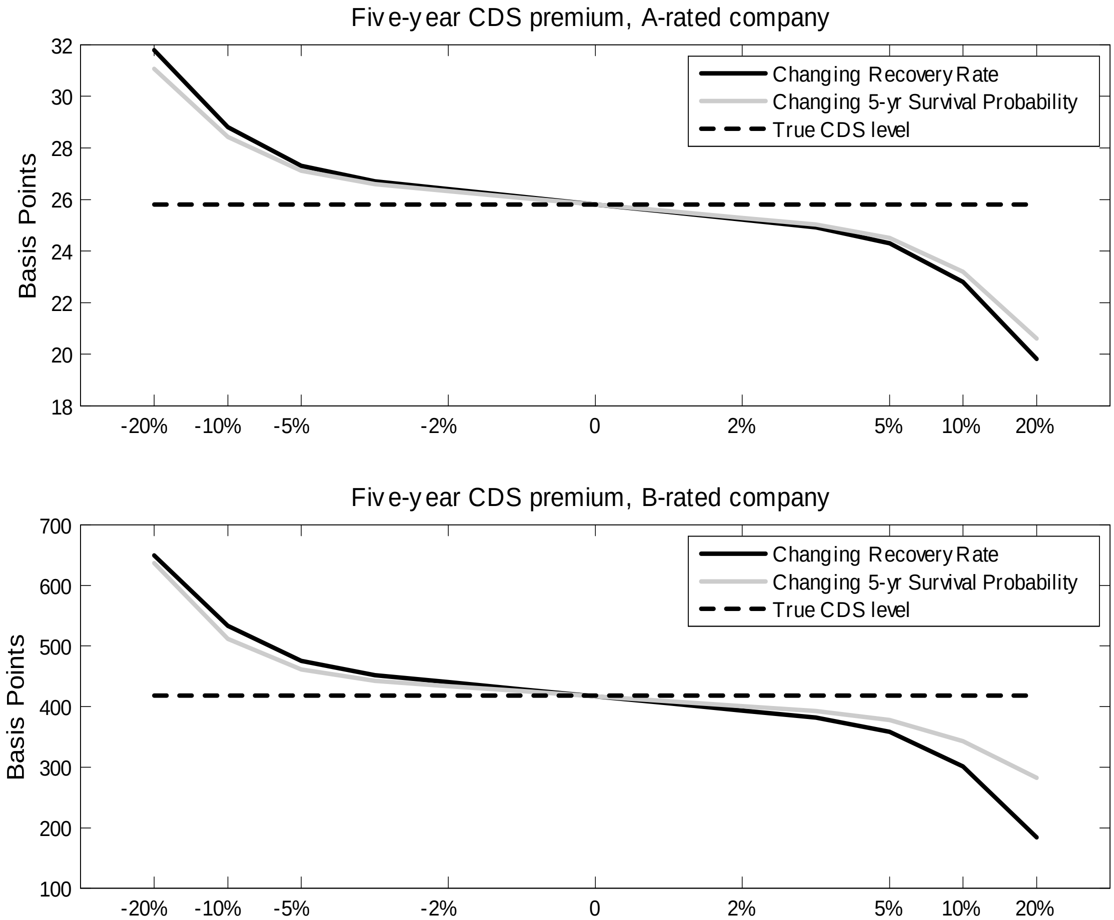
Zhang, B.; H. Zhou; and H. Zhu. "Explaining Credit Default Swap Spreads with Equity Volatility and Jump Risks of Individual Firms." *Review of Financial Studies,* 22 (2009), 5099-5131.

Figure 1: RMSE for Estimated Recovery Divided by RMSE for 40% Recovery.



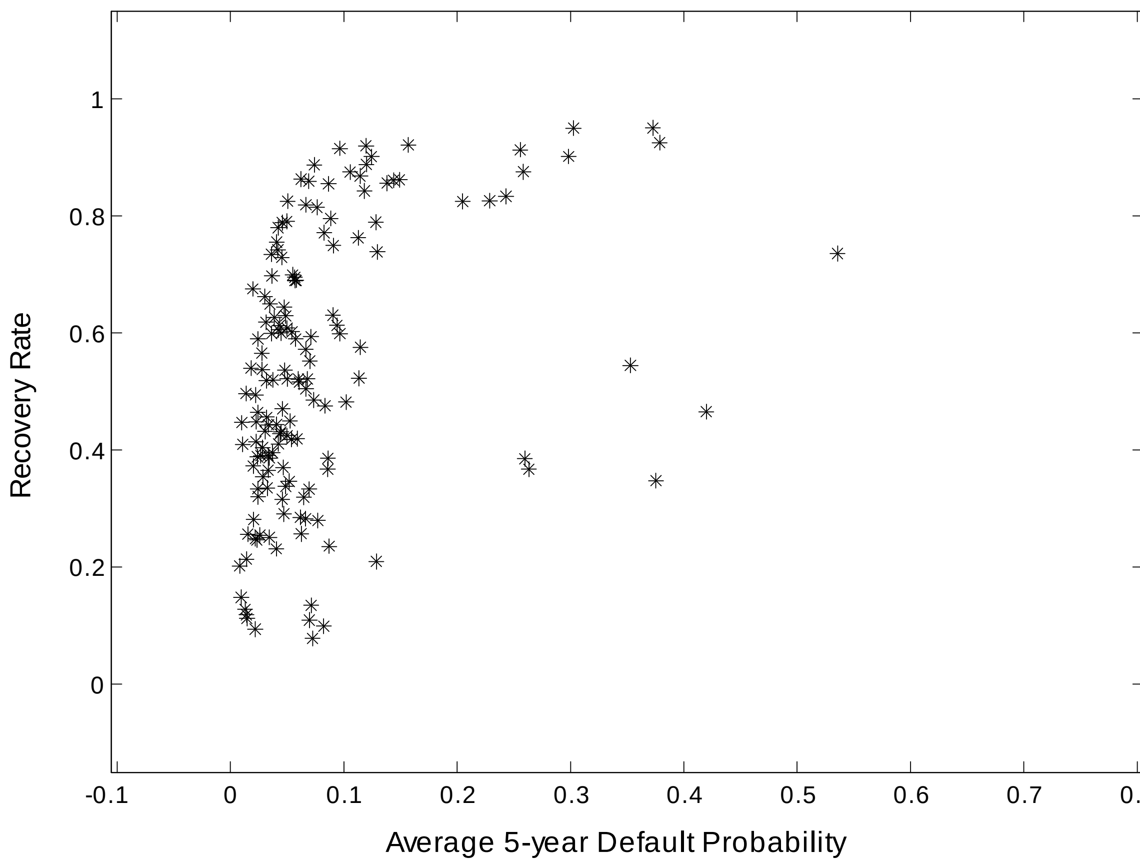
Notes to Figure: The top panel shows the ratio of the RMSE with estimated recovery and the RMSE with 40% recovery for one-year CDS spreads, for all 152 firms. The middle and bottom panels show the same ratio for three-year and five-year CDS spreads respectively.

Figure 2: The Impact of Default Probabilities and Recovery Rates on CDS Premia.



Notes to Figure: We show how CDS premia for a typical A-rated company (top) and a typical B-rated company (bottom) are affected by survival probabilities and recovery rates. The benchmark pricing, based on our estimation results, results in a CDS premium of 25.81 basispoints for the A-rated company and 416.94 for the B-rated company. The dark grey line indicates how deviations from the benchmark values of survival probabilities, measured in percentages on the X-axis, affect CDS premiums, and the light grey line does the same for recovery rates.

Figure 3: Cross-Sectional Relation Between Average Default Probabilities and Recovery Rates.



Notes to Figure: We show the cross-sectional relation between average estimated five-year default probabilities and estimated recovery rates, using estimates for all 152 firms.

Table 1: Economic Determinants of CDS Spreads and the Latent Factor

	CDS	Latent Factor	Squared Residuals	CDS	Latent Factor	Squared Residuals
Constant	0.0019**	-0.1227	4.78E-08	0.0019**	-0.1661	1.92E-08
	(2.35)	(-0.20)	(0.86)	(2.00)	(-0.18)	(0.41)
10-Yr Treasury	-0.0004***	-0.1882	-1.51E-08	-0.0005***	-0.2503*	-9.87E-09
	(-2.94)	(-1.33)	(-1.16)	(-2.83)	(-1.93)	(-0.85)
Leverage	0.0053***	4.9304***	8.38E-08	0.0048***	3.6869***	-1.60E-09
	(3.42)	(3.34)	(0.95)	(2.60)	(3.08)	(-0.02)
Volatility	0.0108*	5.9968	7.91E-07	0.0068	4.7600	3.96E-07
	(1.93)	(1.48)	(1.09)	(1.39)	(1.53)	(0.66)
Slope				-0.0001	-0.1877*	1.49E-08
				(-0.48)	(-1.95)	(1.08)
VIX				4.14E-05*	0.0300*	2.06E-09
				(1.90)	(1.68)	(1.25)
Liquidity				0.0006	0.4064	3.51E-09
				(0.77)	(0.54)	(0.13)
S&P500 Return				0.0007	0.0718	8.50E-08
				(0.13)	(0.02)	(0.19)
R^2	43.27%	48.03%	4.75%	56.63%	65.18%	8.30%

Notes: This table reports the median of coefficients and Newey-West t-statistics across all firms and maturities for CDS spreads and squared errors, as well as for the firm-specific latent factor. We run separate regressions for the one, three, and five-year maturity for all firms, and then get the sample median of all coefficients and Newey-West t-statistics.

Table 2: Estimated Parameter Distribution from Monte Carlo Simulations

	α	β_1	β_2	β_3	β_4	μ_j	Φ_j	Σ_j	u_1	u_2	u_3	Recovery
						Data Generating Process						
	0.00095	-0.00011	0.00003	-0.00075	-0.00241	-0.00248	0.99984	0.00050	1.07E-08	1.94E-08	1.08E-08	0.53742
Percentile						**Distribution of Estimated Parameters**						
Average	0.00097	-0.00011	0.00003	-0.00076	-0.00235	-0.00253	0.99983	0.00051	1.10E-08	2.00E-08	1.11E-08	0.53793
2.5%	0.00085	-0.00012	0.00003	-0.00085	-0.00261	-0.00278	0.99978	0.00039	8.39E-09	1.63E-08	8.80E-09	0.46295
25%	0.00093	-0.00012	0.00003	-0.00079	-0.00244	-0.00263	0.99982	0.00046	1.01E-08	1.85E-08	1.00E-08	0.51760
50%	0.00097	-0.00011	0.00003	-0.00076	-0.00235	-0.00253	0.99984	0.00050	1.10E-08	2.00E-08	1.11E-08	0.53779
75%	0.00100	-0.00011	0.00003	-0.00074	-0.00225	-0.00244	0.99985	0.00056	1.19E-08	2.14E-08	1.21E-08	0.56464
97.5%	0.00110	-0.00010	0.00003	-0.00068	-0.00215	-0.00227	0.99988	0.00066	1.36E-08	2.41E-08	1.35E-08	0.61422
Std Err	0.00006	0.00001	0.00000	0.00004	0.00013	0.00013	0.00002	0.00007	1.41E-09	1.99E-09	1.30E-09	0.03555

Notes: We report the parameters of the data generating process and the distribution of estimated parameters from the Monte Carlo simulations. (α, β_1, β_2,β_3,β_4) capture the dynamics of the hazard rate process; (μ_j, Φ_j, Σ_j) capture the dynamics of the firm-specific latent factor; and (u_1,u_2,u_3) are the standard deviations of measurement errors of the 1-year, 3-year, and 5-year CDS spreads.

Table 3: Summary Statistics and Correlation Matrix for Firm and Industry Variables

Panel A: Summary Statistics

Variable	#Obs	Average	Std. Dev.	Min	Max
Leverage	576	0.23	0.13	0.01	0.71
Firm Size	576	10.09	1.27	7.39	13.87
Tangibility	546	0.52	0.36	0	1.66
Ind. Distress 1	576	0.06	0.24	0	1
Ind. Distress 2	576	0.05	0.22	0	1
Ind. Q	576	1.01	0.54	0.12	2.49
Ind. Leverage	433	0.16	0.11	0	0.50
Ind. Specificity	424	0.16	0.18	0	0.70
Ind. Illiquidity	543	1.25	0.49	0.34	2.46
Ind. Peer Firms	576	284.81	302.16	7	1023

Panel B: Correlation Matrix

	Leverage	Firm Size	Tangibility	Ind. Distress1	Ind. Distress2	Ind. Q	Ind. Leverage	Ind. Specificity	Ind. Illiquidity	Ind. Peer Firms
Leverage	1									
Firm Size	-0.1084	1								
Tangibility	0.2400	-0.3262	1							
Ind. Distress 1	0.2157	-0.0543	-0.0612	1						
Ind. Distress 2	0.0913	-0.0428	-0.1077	0.0873	1					
Ind. Q	-0.1076	-0.2213	0.1721	-0.2234	-0.1865	1				
Ind. Leverage	0.3617	-0.0923	0.2793	0.0637	0.1203	-0.4072	1			
Ind. Specificity	0.2245	-0.2921	0.5267	0.0438	0.0291	0.0810	0.3255	1		
Ind. Illiquidity	-0.1951	0.0783	-0.3750	-0.0493	0.0850	0.3838	-0.3690	-0.4721	1	
Ind. Peer Firms	-0.0861	0.2643	-0.1968	-0.0043	0.0905	0.1090	-0.3992	-0.4016	0.6242	1

Notes: Panel A reports summary statistics of firm and industry variables used in the regressions in Table 4 in the paper. Panel B reports the correlation matrix for these variables.