

Technical Report: Merton Structural Credit Model with Exponential Smoothing Improvement

Ruhani Rekhi

December 29, 2025

1 Introduction

This report presents an implementation and improvement of the Merton (1974) structural credit model. After implementing and diagnosing the baseline model on API-fetched 2020 data for five U.S. firms (Apple, JPMorgan Chase, Tesla, ExxonMobil, and Ford), we identified default probability (PD) instability as a primary weakness. To approach this issue we implement exponential smoothing as a minimal improvement that reduces PD volatility while preserving the model's ability to properly capture credit deterioration during crisis periods.

2 Model Formulation

The Merton model views a firm's equity as a European call option on its assets, with the strike price being the face value of its debt. The key equations and assumptions of the baseline Merton model are outlined below.

2.1 Baseline Merton Model

2.1.1 Key Assumptions

1. Firm asset value V_t follows geometric Brownian motion:

$$dV_t = \mu V_t dt + \sigma_V V_t dW_t \quad (1)$$

2. Firm has zero-coupon debt with face value D maturing at time T
3. Default occurs only at maturity T if $V_T < D$
4. Equity is a European call option on firm assets: $E = \max(V_T - D, 0)$
5. Markets are frictionless (no transaction costs), trading is continuous, no arbitrage (no riskless profit opportunities)

2.1.2 Key Equations/ Mathematical Formulation

Equity Value Under the Merton model, equity value is given by the Black-Scholes call option formula:

$$E = V \cdot \Phi(d_1) - D \cdot e^{-r(T-t)} \cdot \Phi(d_2) \quad (2)$$

where

$$d_1 = \frac{\ln(V/D) + (r + \sigma_V^2/2)(T-t)}{\sigma_V \sqrt{T-t}} \quad (3)$$

$$d_2 = d_1 - \sigma_V \sqrt{T-t} \quad (4)$$

and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Equity Volatility The relationship between equity volatility σ_E and asset volatility σ_V is:

$$\sigma_E \cdot E = \Phi(d_1) \cdot \sigma_V \cdot V \quad (5)$$

where $\Phi(d_1)$ is the option delta, representing the sensitivity of equity value to changes in asset value.

Default Probability The risk-neutral probability of default at maturity is:

$$PD = \Phi(-d_2) = \Phi\left(-\frac{\ln(V/D) + (r - \sigma_V^2/2)T}{\sigma_V \sqrt{T}}\right) \quad (6)$$

Distance-to-Default The distance-to-default metric is:

$$DD = \frac{\ln(V/D) + (r - \sigma_V^2/2)T}{\sigma_V \sqrt{T}} \quad (7)$$

2.1.3 Calibration Approach

The Merton model calibration involves solving the following system of equations for the unobservable asset value V and asset volatility σ_V , given the observable equity value E , equity volatility σ_E , debt face value D , time to maturity $T - t$, and risk-free rate r :

$$E = \text{BlackScholes}(V, D, T - t, r, \sigma_V) \quad (8)$$

$$\sigma_E \cdot E = \Phi(d_1) \cdot \sigma_V \cdot V \quad (9)$$

2.2 Improved Model

2.2.1 Modification

We apply post-processing exponential smoothing to default probabilities. For each firm's time series of raw PDs $\{PD_t^{\text{raw}}\}$ we applied the following smoothing formula:

$$PD_t^{\text{smooth}} = \alpha \cdot PD_t^{\text{raw}} + (1 - \alpha) \cdot PD_{t-1}^{\text{smooth}} \quad (10)$$

where $\alpha = 0.1$ is the smoothing parameter.

2.2.2 Justification

This improvement is justified on two grounds. Theoretically speaking, the baseline model assumes you can trade infinitely with no transaction costs. Whereas in reality, trading is discrete and every trade has costs leading to noise in observed prices that isn't properly captured in the model. Additionally, daily equity volatility contains high-frequency noise (changes in company value, as well as market microstructure noise) which is further amplified through the nonlinear Black-Scholes model into asset volatility thus into PD estimates. Practically speaking, PD estimates play a key role in credit risk management for portfolio monitoring, capital allocation, and risk limits. When daily PD estimates oscillate wildly (for example 5-20 percent points as seen in the baseline model), they become operationally unusable. This is why smoothing techniques are applied in practice to reduce noise while preserving signal. While structural improvements (first-passage time, stochastic volatility, jump diffusion) could address PD instability, they introduce significant complexity and additional parameters requiring calibration. Given the empirical focus of this assignment and the operational use case, we implement exponential smoothing as a minimal, transparent, and interpretable improvement that directly addresses the diagnosed weakness without introducing model risk from additional assumptions.

2.2.3 Mathematical Formulation of Improvement

The exponential weighted moving average (EWMA), the smoothing formula we are using, has the following properties.

1. Parameter α controls reactivity: lower α yields more smoothing
2. No look-ahead bias (uses only past observations)
3. Preserves long-term trends while filtering short-term noise
4. Weights decay exponentially: observation k periods ago has weight $\alpha(1 - \alpha)^k$

Diving into the math, property 1 of the EWMA smoothing formula which is equation (10) can be demonstrated by unrolling the recursion:

$$\text{PD}_t^{\text{smooth}} = \alpha \sum_{k=0}^{\infty} (1 - \alpha)^k \cdot \text{PD}_{t-k}^{\text{raw}} \quad (11)$$

This shows how α controls the weight decay rate, thus the reactivity. For example, with $\alpha = 0.9$ the decay factor $(1 - \alpha) = 0.1$ which means today would be 90%, yesterday 9%, two days ago 9%, etc. Showing that the higher alpha has high reactivity to recent changes since recent data dominates. Conversely, with $\alpha = 0.1$ the decay factor $(1 - \alpha) = 0.9$ which means today would be 10%, yesterday 9%, two days ago 8.1%, etc. This shows that lower alpha has low reactivity since weights are more evenly spread out over time.

Property 2 is evident since the formula only uses PD_t^{raw} and $\text{PD}_{t-1}^{\text{smooth}}$ which is computed from past raw PDs. Additionally the code implementation processes data chronologically to avoid look-ahead bias.

Property 3 is evident since the EWMA formula averages over past raw PDs, thus filtering out high-frequency noise while preserving long-term trends.

Property 4 is shown in equation (11) where the weight for observation k periods ago is $\alpha(1 - \alpha)^k$, demonstrating the exponential decay.

2.2.4 Calibration Changes

Smoothing is applied post-calibration. The algorithm is:

1. Calibrate (V, σ_V) using baseline approach
2. Compute PD_t^{raw} from equation (6)
3. Apply EWMA smoothing to obtain $\text{PD}_t^{\text{smooth}}$

Note that asset values V , asset volatilities σ_V , and distance-to-default remain unchanged from the baseline model.

3 Calibration Methodology

3.1 Numerical Method

We solve the system of two non-linear equations for two unknowns using scipy's `fsolve` optimizer, which implements a modified Powell hybrid method—a quasi-Newton approach with adaptive damping.

System of Equations

$$F_1(V, \sigma_V) = \text{BlackScholes}(V, D, T, r, \sigma_V) - E = 0 \quad (12)$$

$$F_2(V, \sigma_V) = \Phi(d_1) \cdot \sigma_V \cdot V - \sigma_E \cdot E = 0 \quad (13)$$

Algorithm We use scipy's `fsolve` optimizer with the following parameters:

1. **Initialize:**

$$V_0 = E + D \quad (\text{balance sheet approximation}) \quad (14)$$

$$\sigma_{V,0} = \sigma_E \cdot \frac{E}{E + D} \quad (\text{de-levered volatility}) \quad (15)$$

2. **Solve:** Call `fsolve(equations, [V_0, sigma_V_0], xtol=1e-6)`

`fsolve` implements a modified Powell hybrid method that approximates the Jacobian via finite differences, uses a trust region approach with adaptive damping, and converges when the relative error in the solution update (Δx) is $\|\Delta x\| < 10^{-6}$.

3. **Validate:** After convergence, check the convergence flag utilizing `ier` (integer flag that indicates success or failure of solver) `ier == 1`, the economic validity (asset value should be larger than equity value by a small margin) $V \geq 1.01E$, and reasonable range $0.01\% \leq \sigma_V \leq 200\%$.

3.2 Edge Case Handling

1. **Invalid inputs:** ($E \leq 0$, $\sigma_E \leq 0$, $D < 0$): reject observation
2. **Jacobian singularity:** reject observation (caught through code exception handling)
3. **Non-convergence:** reject observation (caught through `fsoolve` ier flag)
4. **Economic Validity:** $V > 1.01E$ (covered in validation above)

4 Empirical Setup

Sample Description

- **Firms:** Apple (AAPL), JPMorgan Chase (JPM), Tesla (TSLA), ExxonMobil (XOM), Ford (F)
- **Period:** January 2 – December 30, 2020 (252 trading days)
- **Total observations:** 1,260

4.1 Implementation Details

4.1.1 Time to Maturity

We assume $T = 1.0$ year for all observations. This is justified as it is a standard horizon for credit risk assessment as well as consistent with industry practice.

4.1.2 Market Capitalization Scaling

The dataset provides equity as price per share, but the Merton model requires total market capitalization. We convert:

$$E = \text{price per share} \times \text{shares outstanding}/10^6 \quad (\text{millions}) \quad (16)$$

Table 1 shows shares outstanding adjusted for 2020 stock splits:

Table 1: Shares Outstanding (Post-2020 Stock Splits)

Firm	Shares (billions)	Notes
AAPL	17.00	Post 4-for-1 split (August 2020)
JPM	3.07	No split
TSLA	0.96	Post 5-for-1 split (August 2020)
XOM	4.23	No split
Ford	3.92	No split

Source: Official 10-K filings, 8-K filings, Press Releases, split-adjusted for consistency with Yahoo Finance price data.

4.1.3 Quarterly Debt Alignment

Debt is reported quarterly while equity prices are daily to solve this issue we use **forward-fill**. This entails using the most recent quarterly debt value for each date. The code implementation uses `pd.merge_asof()` with `direction='backward'`.

Assumption: Debt changes slowly relative to daily equity movements, allowing quarterly observations to take place for daily values. While this generally holds for stable periods, this assumption may be violated during crisis periods such as the March 2020 COVID crisis which is captured in our data. However, most of the variation we see in debt-to-equity ratios stems from equity price changes rather than debt changes which in turn limits the impact of this approximation.

4.1.4 Debt Data Collection

To have proper calibration throughout 2020, we collected year-end total debt data for December 31, 2019 and December 31, 2020 from official SEC 10-K annual filings for all firms. Total debt is defined as the sum of all current and non-current debt obligations, including short-term borrowings, the current portion of long-term debt, and long-term debt. For firms operating financing subsidiaries (Ford Motor Credit), consolidated debt includes both automotive and financing operations. For Apple, we use fiscal year-end data (September 28, 2019) supplemented by first fiscal quarter data, as Apple's fiscal year ends in September rather than December. All debt values are extracted from audited consolidated balance sheets filed with the U.S. Securities and Exchange Commission as well as other external sources and are reported in millions of USD. Table 2 presents the debt data used in our analysis.

Table 2: Total Estimated Debt at December 31 for 2019 and 2020

Firm	Dec 31, 2019 (\$M)	Dec 31, 2020 (\$M)	Change (\$M)
AAPL	108,047	132,480	+24,433
JPM	324,609	354,599	+29,990
TSLA	13,419	8,873	-4,546
XOM	53,257	47,704	-5,553
F	155,017	139,485	-15,532

4.1.5 Additional Assumptions

- Dividends ignored for simplicity
- Zero recovery rate at default (standard in Merton model)
- Risk-free rate approximated using daily 10 year Treasury yield (from FRED)

5 Baseline Model Diagnosis

5.1 Identified Weaknesses

The primary weakness identified in the baseline Merton model is default probability instability. All five firms show excessive PD volatility displayed in key metrics below:

Table 3 summarizes the stability metrics for the baseline model:

Table 3: Baseline Model PD Stability Metrics

Firm	Mean PD	Std Dev	CV
AAPL	0.10%	0.30%	3.03
JPM	3.90%	8.64%	2.22
TSLA	3.53%	6.07%	1.72
XOM	0.70%	1.75%	2.52
Ford	6.38%	10.78%	1.69

High Coefficient of Variation: All firms exhibit CV between 1.69 and 3.03, indicating that standard deviation is 1.7 to 3 times the mean PD. While no universal CV threshold exists in credit risk practice, CV values exceeding 1.0 indicate standard deviation exceeds the mean, representing a high noise-to-signal ratio in this context making it unsuitable for operational risk management.

Universal Instability Pattern The instability problem in Table 3 affects all firms regardless of risk level or leverage. Both the highest-risk firm (Ford, 6.38% mean PD) and the lowest-risk firm (AAPL, 0.10% mean PD) exhibit CV values substantially exceeding 1.0. This suggests an issue with the model structure rather than firm-specific calibration problems. Notably, AAPL produces the highest relative volatility ($CV = 3.03$) despite having minimal default risk, exemplifying perhaps that the baseline model's instability worsens for low-risk firms where small changes create large percentage swings. The consistency of $CV > 1.0$ across all firms confirm that the baseline implementation produces estimates where noise dominates signal.

Root Cause Analysis The baseline Merton model amplifies equity volatility through non-linear transformations and leverage effects, leading to unstable PD estimates !

5.2 Examples

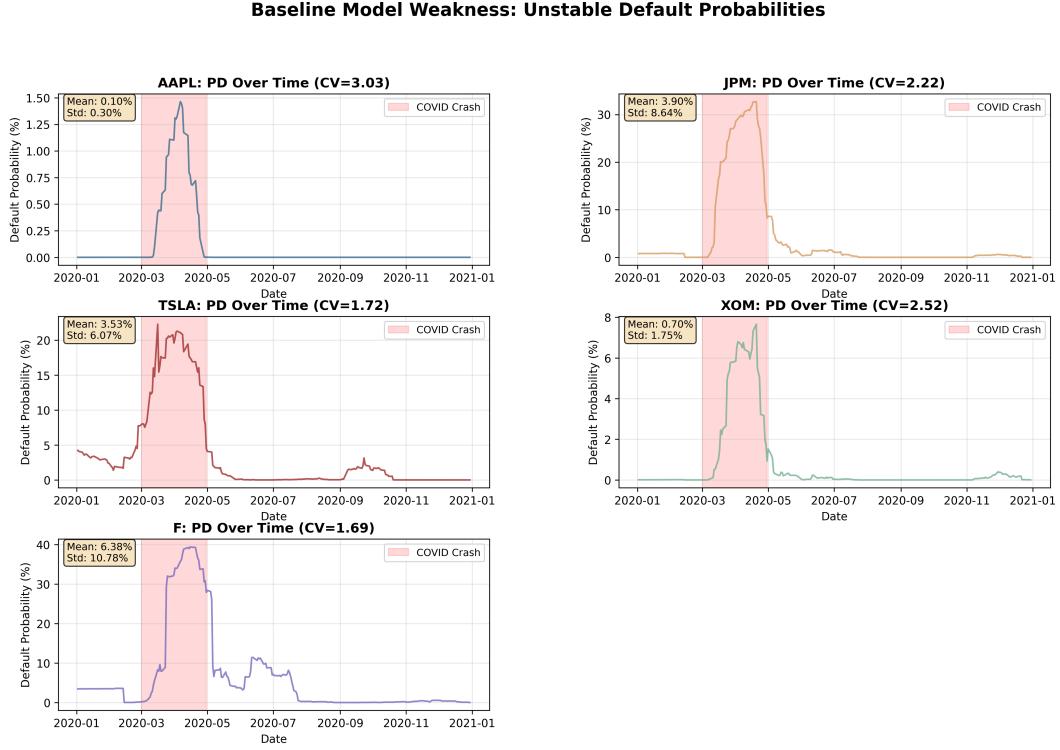


Figure 1: Baseline Model Default Probabilities Across 2020.

All five firms exhibit extreme PD spikes during the COVID-19 crisis (March–April 2020, shaded region), with peak increases of 6–14 times their annual mean values. For example we see that Ford reaches 40% PD (6.3 \times its mean), JPMorgan peaks at 32% (8.2 \times its mean), and even low-risk AAPL spikes to 1.4% (14 \times its mean). Additionally, these spikes reverse rapidly almost within 2–4 months which is inconsistent with typical credit cycles where rating changes occur over quarters and defaults unfold over years!

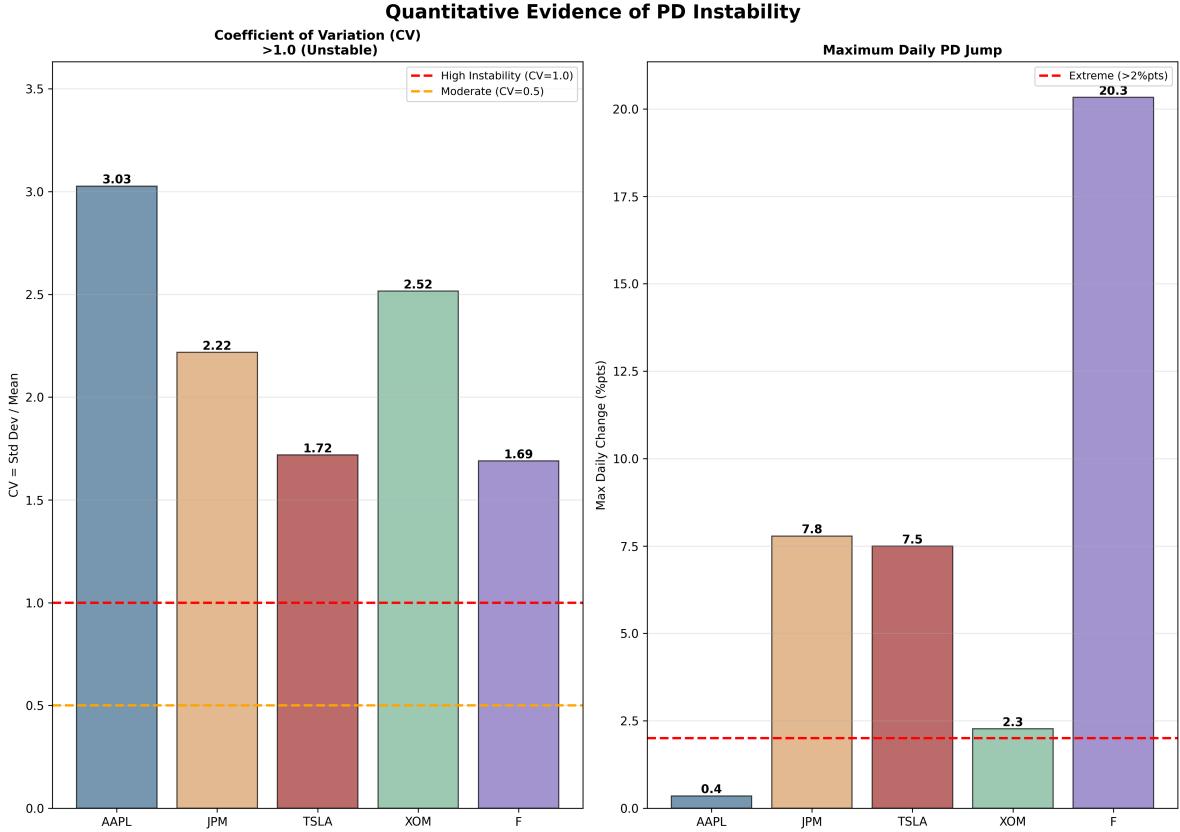


Figure 2: Quantitative Evidence of PD Instability in Baseline Model.

Figure 2 provides quantitative evidence of the instability problem through two metrics. The left panel shows that all firms exceed a CV of 1.0, indicating standard deviation exceeds the mean for all cases (operationally unacceptable, explained in further detail above). The right panel reveals extreme single-day PD movements, with four out of five firms exceeding the 2 percentage point reference line for operational viability. Ford's 20.3 percentage point single-day jump is almost 10 times this reference magnitude, while JPMorgan and Tesla show 7–8 percentage point jumps. These sort of extreme movements exemplify crisis amplification. Moreover, the contrasting patterns between panels reveal an important distinction between relative and absolute instability. Low-risk firms (AAPL, XOM) suffer from high relative volatility ($CV > 2.5$) where noise dominates signal whereas high-risk firms (Ford, JPM, TSLA) suffer from large absolute movements ranging from 7.5 to 20.3 percentage points. Both forms of instability are problematic for risk management as they lead to unreliable PD estimates that cannot be used for portfolio monitoring or capital allocation. Even the best-performing case (XOM with $CV = 2.52$ and max jump = 2.3%pts) exceeds our reference thresholds, indicating no firm in the sample produces baseline estimates suitable for risk management without smoothing.

6 Improved Model Results

6.1 Quantitative Comparisons

We will be evaluating the improvement using two metrics that capture different dimensions of PD Stability:

Coefficient of Variation (CV) As previously defined the CV metric ($CV = \sigma/\mu$) measures the relative volatility of PD estimates, by expressing standard deviation as a proportion of the mean. This metric is dimensionless allowing fair comparison across firms with different risk levels. A CV exceeding 1.0 can be viewed as a reference threshold since it indicates standard deviation exceeds the mean, which represents a high noise-to-signal ratio. But more generally speaking lower CV values indicate better stability.

Mean Daily Change The mean daily change measures the average absolute day-to-day PD movement in percentage points: $\text{mean}(|\Delta PD_t|)$ where $\Delta PD_t = PD_t - PD_{t-1}$. This metric captures operational stability where large daily changes can trigger limit breaches, portfolio rebalancing, and complicate credit pricing. Lower values indicate more stable estimates suitable for practical risk management.

6.1.1 Improvement Results

Table 3 presents the stability improvements achieved through exponential smoothing ($\alpha = 0.1$):

Table 4: Improvement Results: Coefficient of Variation and Daily Change Reduction

Firm	Coefficient of Variation (CV)			Mean Daily Change (%pts)		
	Baseline	Improved	Reduction	Baseline	Improved	Reduction
AAPL	3.03	2.43	↓19.6%	0.012	0.005	↓56.5%
JPM	2.22	1.91	↓14.0%	0.303	0.220	↓27.4%
TSLA	1.72	1.33	↓22.8%	0.320	0.208	↓34.9%
XOM	2.52	2.06	↓18.1%	0.086	0.053	↓38.4%
Ford	1.69	1.41	↓16.6%	0.488	0.300	↓38.5%
Avg	2.23	1.83	↓18.2%	0.242	0.157	↓39.1%

The improved model achieves substantial stability improvement across both our metrics. The average CV reduction of 18.2% demonstrates systematic improvement in relative volatility across all firms. While all firms still exceed $CV = 1.0$, the improved model brings them substantially closer to acceptable stability levels.

The average daily change reduction of 39.1% represents dramatic operational improvement. This reduction directly addresses the core weakness identified in Section 5: excessive day-to-day volatility that renders baseline PDs unusable for risk management.

6.1.2 Statistical Summary

The improvements demonstrate consistent patterns across all firms. We see CV reductions ranging from 14.0–22.8%, with no firm showing less than 14% improvement. We also see daily change reductions ranging from 27.4–56.5%, with all firms exceeding 27% improvement. A case specific highlight is TSLA achieving the best CV reduction of 22.8%, demonstrating that smoothing is particularly effective for high-volatility growth stocks where baseline noise-to-signal ratios are elevated due to frequent news-driven price movements. Another highlight is AAPL, achieving the best mean daily PD change reduction of 56.5%, where smoothing dramatically cuts noise for low-risk firms where small absolute price movements create large percentage changes in PD estimates that obscure the firm’s stable underlying credit quality. All in all, the consistency of improvement across all five firms exemplifies that exponential smoothing provides systematic benefits regardless of baseline risk level.

6.1.3 Visual Comparison

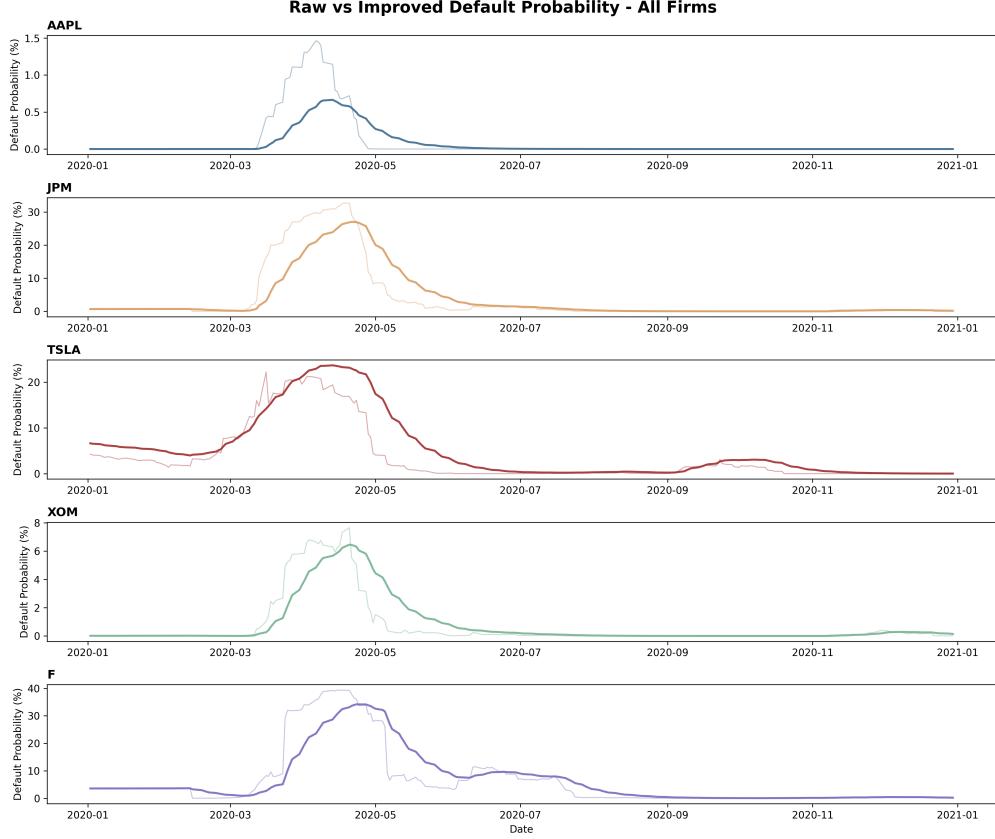


Figure 3: Raw vs Improved Default Probability—All Firms.
Thin lines show baseline (raw) PD. Thick lines show improved (smoothed) PD.

Figure 3 provides visual evidence of the improvement across all five firms. The thin lines (baseline) show the excessive daily oscillations documented in Section 5, while the thick lines (improved) demonstrate smoother trajectories that preserve important credit signals.

The COVID crisis spike (March–April 2020) remains visible in all improved PDs, but with more realistic recovery patterns. While the baseline model shows extreme spikes that reverse within 2 to 3 months through rapid oscillation, the improved model shows more gradual PD declines through mid-2020 with smooth, steady trajectories. This is can be best seen in JPM, where the improved model shows a clean downward slope from April through July whereas the baseline oscillates wildly. This response better reflects the reality that credit quality improvements unfold over quarters, not weeks. For Tesla, we see that the improved model shows maintained elevated PD levels through mid-2020 revealing the firm’s underlying risk more clearly.

Overall, this distinction has important operational implications. The baseline model’s sharp reversals that spike dramatically during crisis onset then crashes through volatile oscillations would lead to constant limit adjustment and portfolio rebalancing in response to noise rather than fundamental credit changes. The improved model enables stable risk management decisions while maintaining responsiveness to genuine credit deterioration.

6.1.4 Cross-Sectional Ranking Changes

Table 5: Risk Ranking Comparison: Mean PD by Model

Firm	Baseline Mean PD	Improved Mean PD	Change
Ford	6.38%	6.58%	+0.20%pts
TSLA	3.53%	5.23%	+1.70%pts
JPM	3.90%	3.77%	-0.13%pts
XOM	0.70%	0.79%	+0.09%pts
AAPL	0.10%	0.06%	-0.04%pts

Table 4 shows that exponential smoothing causes a cross-sectional ranking change. While Ford remains the highest-risk firm and AAPL the lowest, Tesla and JPMorgan swap positions: the baseline model ranks JPM as higher risk (3.90%) than TSLA (3.53%), but the improved model reverses this to TSLA (5.23%) > JPM (3.77%). This 1.7 percentage point increase in Tesla’s mean PD reflects smoothing’s different impact on firms with distinct volatility characteristics. Tesla’s high equity volatility and news-driven price swings create baseline PD estimates that oscillate rapidly around a higher true credit risk level. Smoothing removes these oscillations hence revealing Tesla’s elevated underlying risk. Whereas, JPMorgan’s more stable equity prices produce baseline PDs that already reflect its credit quality reasonably well, so smoothing produces minimal change. This ranking shift demonstrates that while smoothing reduces time-series instability systematically, it can alter cross-sectional risk assessment when baseline models differently distort estimates across firms with varying volatility profiles.

6.2 Why It's Better

6.2.1 Addressing the Weakness

The baseline model's primary weakness was excessive default probability instability, seen through the metrics of high coefficient of variation and large daily PD changes. This happens due to two factors. The baseline model amplified small equity price changes into larger asset volatility fluctuations through the non-linear transformations of the Black Scholes call-option formula (as seen in equations 2 to 4) leading to the unstable PD estimates. High debt-to-equity ratios amplify asset volatility into equity volatility via equation (5). For highly levered firms like Ford, equity is highly sensitive to asset value changes, meaning small asset movements create large equity price swings (Merton 1974). By applying exponential smoothing (equation 10), we reduce the impact of these amplified high-frequency fluctuations, leading to more stable PD estimates.

6.2.2 Economic Interpretation

The improvement results have clear economic meaning rooted in the fundamental difference between credit quality changes and equity price movements. Credit quality evolves gradually over quarters or years as firms have changes through revenue trends, balance sheet adjustments, and industry conditions. Whereas equity prices fluctuate daily due to market microstructure effects like bid-ask bounce, order flow, sentiment shifts, and liquidity shocks. The baseline Merton model inherently conflates these phenomena, by treating all equity price movements as credit signals which in turn produces PD estimates that show market noise rather than underlying credit fundamentals. Exponential smoothing recognizes this distinction by averaging over recent observations with exponentially decaying weights (equation 11), the improved model filters short-term equity price noise while preserving longer-term trends. The empirical results validate this interpretation across different market regimes. This is best seen in the COVID crisis where the improved model's more measured response and gradual recovery better reflect economic reality where credit quality improvements unfold over quarters as balance sheets strengthen and cash flows recover. The 39.1% reduction in mean daily PD change (Table 3) shows that estimates transitioned from operationally unusable to viable, enabling consistent limit monitoring, stable capital allocation, and reliable credit pricing without the constant adjustments and false alarms generated by baseline volatility.

7 Limitations

While exponential smoothing addresses time-series instability, several limitations remain. The smoothing parameter ($\alpha = 0.1$) creates a bias-variance tradeoff where heavier smoothing reduces volatility but introduces lag, potentially delaying response to genuine rapid deterioration. Quarterly debt data with forward-fill may miss sudden leverage changes during crises. The ranking change between Tesla and JPMorgan (Table 4) illustrates that smoothing can alter cross-sectional risk assessment when firms have different volatility characteristics, raising questions about whether the smoothed or baseline ranking better reflects true credit

risk. Finally, smoothing effectiveness depends on noise-to-signal ratios; for firms experiencing genuine rapid credit changes, smoothing could dampen valid signals. Future work could address these through first-passage time models, stochastic volatility, or adaptive smoothing parameters.

8 Conclusion

This report implements the Merton (1974) structural credit model, diagnoses PD instability as an issue , and improves stability through exponential smoothing. The improved model achieves an average 18.2% CV reduction and average 39.1% daily change reduction while generally preserving credit risk differentiation across firms. The key insight we see is that credit quality evolves over quarters whereas equity prices fluctuate daily. Smoothing filters market noise and preserves credit signals as seen in the COVID crisis response and stable non-crisis estimates. The improvement transforms unusable volatility into operationally viable risk management inputs. As mentioned previously, future work could explore structural improvements (first-passage time models, stochastic volatility) or adaptive smoothing parameters responsive to market conditions.