Estimation of Probability of Default (PD)

Introduction to Credit Risk and Applications in Python

Abhishek Sharma **Alireza Miriashtiani**

July 2025

University of Freiburg

Agenda

1. **Part 1 – Conceptual Framework & Model Families**
    - Introduction to Credit Risk & PD
    - Core Approaches to PD Estimation
    - Model Comparison & Transition
2. **Part 2 – Macroeconomic PD Modeling & Case Study**
    - Macro PD Model Setup
    - Stress Testing Methodology
    - Empirical Application on US Firms
    - Conclusions & Extensions

Introduction to Credit Risk

- **Credit risk** refers to the potential that a borrower or counterparty fails to meet their contractual debt obligations.
- It is a major source of financial risk for banks, bond investors, and any institution with exposure to loans or credit instruments.
- Managing credit risk is essential for maintaining solvency, liquidity, and profitability.
- Regulatory frameworks like Basel II and III require capital to be held based on credit risk exposure.

Defining Probability of Default (PD)

- **PD** quantifies the likelihood that a borrower will default on their obligations within a specific time horizon (usually 1 year).
- It is a fundamental input into:
  - Regulatory capital requirements
  - Internal credit scoring and risk assessment
  - Loan pricing and provisioning
  - Portfolio optimization and stress testing
- In practice, PD is usually estimated using one of several model types.

Three Core Approaches to PD Estimation

1. **Fundamental & Classical Statistical Models**
   Maps accounting and macro variables to PD via transparent logistic / probit regressions.
2. **Structural / Option-Theoretic Models**
   Treats equity as a call on assets; default occurs when asset value falls below debt (Merton / KMV).
3. **Machine-Learning & Non-linear Models**
   Learns complex patterns with RF, boosting, NNs; tackles class imbalance with tools such as SMOTE.

Each approach has distinct data requirements, strengths, and limitations.

Fundamental & Classical Statistical Models – Overview

- Combines firm-level **accounting ratios** and **macroeconomic indicators** to assess credit risk.
- **Accounting models**: Altman Z-score, Ohlson O-score, Zmijewski – quick discriminant scores from balance-sheet and income statement data.
- **Macroeconomic models**: logistic / probit frameworks linking PD to GDP growth, unemployment, interest rates; penalised and rare-event variants handle sparse defaults.
- **When to use**: limited market data (private firms, stress testing) but high interpretability demands (IRB, audit); trade-off is linear boundaries and medium predictive power.
- Transparent, backward-looking, and a foundation for regulatory stress testing (e.g. Basel III).

Altman Z-Score (1968)

- Combines 5 financial ratios into a single score:
  - Working Capital / Total Assets
  - Retained Earnings / Total Assets
  - EBIT / Total Assets
  - Market Value of Equity / Book Value of Total Liabilities
  - Sales / Total Assets
- Formula:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5$$

- Interpretation:
  - $Z < 1.81$: High default risk
  - $Z > 3.00$: Low default risk

Classical Statistical Prediction – Logit / Probit Overview

- Estimates the probability of default from binary outcomes (default / no-default).
- Typical predictors: financial ratios, firm size, market indicators, macroeconomic data.
- Core techniques: **logistic** and **probit** regressions – interpretable coefficients, extend ratio models.
- Widely used in industry credit scoring due to simplicity and transparency.

Logistic vs Probit Models

Logit: $P(D=1) = \dfrac{1}{1+e^{-(\alpha+\beta' X)}}$ Probit: $P(D=1) = \Phi(\alpha + \beta' X)$

	Logit	**Probit**
Link function	Logistic sigmoid	Normal CDF
Coeff. meaning	Odds ratio	Z-value shift
Typical use	Credit scoring, marketing	Econometrics, limited-dep. vars

Ohlson (1980) and Zmijewski (1984)

- Logistic regression frameworks using accounting variables:
  - **Ohlson**: Uses 9 variables (e.g., size, leverage, liquidity, performance).
  - **Zmijewski**: Uses 3 main variables:
    - Net Income / Total Assets
    - Total Liabilities / Total Assets
    - Current Assets / Current Liabilities
- These models predict a binary default outcome.
- Advantages: statistically rigorous, interpretable coefficients.

Empirical Comparison: PIT vs. TTC Estimation Approaches

Sources: Poghosyan et al. (2018), Jacobs Jr. (2022)

- **Point-in-Time (PIT):**
  - Uses current firm- and macro-level variables
  - Highly responsive to short-term shocks
- **Through-the-Cycle (TTC):**
  - Uses historical averages to smooth business cycles
  - Captures structural default risk, not cyclical noise
- **Hybrid Models:**
  - Combine PIT sensitivity with TTC stability
  - Aim to improve forecast accuracy across scenarios

Why This Matters + Data & Methodology

Model Selection Depends On:

- **Forecasting:** PIT models preferred
- **Capital planning:** TTC models align with regulatory needs
- **Balanced insights:** Hybrid models offer best of both

Poghosyan et al. (2018):

- Panel of 14,000+ firms in 18 countries (1995–2014)
- Variables: profitability, leverage, liquidity, GDP, inflation
- PIT: uses current variables; TTC: long-run moving averages

Jacobs Jr. (2022):

- 24 years of U.S. corporate default data
- Logistic regression separating cyclical vs. structural factors
- Backtested against historical default rates

Key Findings and Contributions

- **PIT Models:**
  - Detect early warning signs
  - Ideal for stress testing and crisis response
- **TTC Models:**
  - Reduce false positives during downturns
  - Align with long-term regulatory frameworks
- **Hybrid Models (Best Performing):**
  - Fuse accounting, market, and macro signals
  - Show superior predictive accuracy
  - Provide practical estimation framework
- **Contribution:**
  - Validates hybrid modeling at scale
  - Supports structured integration of accounting and macroeconomic data

Fundamental & Classical Statistical Models – Pros & Cons

Strengths

- **Transparent**: coefficients and decision rules easy to interpret and audit.
- **Data-light**: works with readily available accounting statements; no market prices needed.
- **Regulator-friendly**: aligns with IRB and Basel model governance expectations.
- **Fast deployment**: simple to implement, calibrate and back-test.

Weaknesses

- **Linear decision boundary**: may miss complex, non-linear patterns.
- **Backward-looking**: relies on historical ratios that can lag current conditions.
- **Collinearity risk**: many financial ratios are highly correlated.
- **Medium predictive power**: generally lower AUROC than modern ML on rich datasets.

Option-Theoretic Approach - Overview

- Introduced by **Merton (1974)** as a structural model of credit risk.
- Treats a firm's equity as a **call option** on its assets.
- Default occurs if the value of the firm's assets falls below its debt at maturity.
- Requires market data on:
  - Asset volatility
  - Firm value

Merton Model Mechanics

- **Inputs:** Market value of equity E_0, book value of debt D, asset volatility σ_V, risk-free rate r, time horizon T.
- **Model intuition:** Default occurs if firm value $V_T < D$ at time T. Equity is viewed as a call option on firm assets with strike D.
- **Distance to Default (DD):**

$$DD = \frac{\ln(V_0/D) + \left(r - \frac{1}{2}\sigma_V^2\right) T}{\sigma_V \sqrt{T}}$$

- **Probability of Default:**

$$PD = \Phi(-DD) = \Phi\left(-\frac{\ln(V_0/D) + \left(r - \frac{1}{2}\sigma_V^2\right) T}{\sigma_V \sqrt{T}}\right)$$

- **Interpretation:**
  - Higher DD $\rightarrow$ lower PD (firm is safer)
  - Forward-looking, market-sensitive credit risk metric

KMV Model and Expected Default Frequency (EDF)

- The KMV model is a commercial adaptation of the Merton model developed by Moody's.
- It estimates a firm's **Expected Default Frequency (EDF)** using:
  - Market value of assets
  - Asset volatility
  - Default point (typically short-term debt + half of long-term debt)
- Maps the **Distance to Default (DD)** to historical default probabilities.
- Produces a forward-looking, market-implied PD.

KMV Model – Market-Implied Default Risk

- **Commercial adaptation** of Merton's model developed by Moody's (CreditMonitor).
- **Inputs:**
  - Estimated market value of assets V_0
  - Asset volatility σ_V
  - **Default point (DP)** = short-term debt + 0.5 × long-term debt
- **Distance to Default:**

$$DD = \frac{\ln(V_0/DP) + \left(\mu - \frac{1}{2}\sigma_V^2\right) T}{\sigma_V \sqrt{T}}$$

- **Expected Default Frequency (EDF):**

$$PD = \text{EDF} = \text{historical frequency associated with } DD$$

- **Features:**
  - Forward-looking and market-sensitive
  - Replaces closed-form normal CDF with empirical DD-to-PD mapping from historical default data

Option-Based / Structural Models – Pros & Cons

Strengths

- **Market-sensitive & forward-looking**: PD adjusts as equity prices move.
- **Theoretically grounded**: derives from Merton (1974) capital–structure option view.
- **Real-time monitoring**: distance-to-default can be updated intraday.
- **Consistent framework**: links PD with expected loss, option pricing, and CDS spreads.

Weaknesses

- Requires liquid equity market data (price, volatility) and timely balance-sheet inputs.
- Assumes log-normal asset returns and constant volatility—may mis-specify fat tails.
- Simplifies capital structure; ignores covenants, drawdowns, hybrids.
- Limited applicability to private firms or illiquid markets.

Machine-Learning & Non-linear Models

- **Common algorithms**: Random Forest, Gradient Boosting (XGBoost, LightGBM), Support Vector Machines, Neural Networks.
- **Class-imbalance remedies**: SMOTE, ROSE, class-weights.
- Leverages high-dimensional accounting, macro, market, or text data for superior predictive accuracy.
- **Next** $\rightarrow$ *Explainable AI* to open the black box.

Explainable AI (XAI) for PD Models

SHAP

- Consistent additive feature attribution
- Global and local explanations

LIME

- Local surrogate models
- Model-agnostic, quick what-if insights

Feature Importance

- Gini / permutation scores
- Fast sanity check for regulators

ECB TRIM and Basel Model-Risk guidance expect clear, auditable explanations for complex models.

Rare-Event Problem & Solutions

- Corporate defaults are typically $<2\%$ of observations severe class imbalance.
- Imbalance can bias coefficients and inflate Type-II error.
- Mitigation options:
  - **SMOTE**: generates synthetic default cases.
  - **Penalised logistic**: ridge / Firth variants reduce small-sample bias.
  - **Ensembles**: bagging / boosting improve robustness.

Penalised Logistic + SMOTE – Empirical Example

- Case study on Turkish commercial credit portfolio.
- Combined **SMOTE** oversampling with **ridge-penalised logistic** regression.
- Result: materially higher AUROC and better calibration vs traditional logit.
- Illustrates practical viability of hybrid statistical fixes in emerging-market data.

Machine-Learning & Non-linear Models – Pros & Cons

Strengths

- **High predictive accuracy**: discovers complex, non-linear relationships.
- **Automatic feature handling**: tree ensembles and deep nets cope with large variable sets.
- **Flexible data ingestion**: integrates ratios, market data, text, ESG signals.
- **Adaptive**: models can be retrained frequently as new data arrive.

Weaknesses

- Often perceived as a "**black box**"; requires XAI for audit.
- **Data-hungry**: needs large, labelled datasets to avoid overfitting.
- **Higher computational cost**: tuning and retraining can be resource-intensive.
- **Regulatory scrutiny**: Basel/ECB demand robust validation and explainability.

Summary Comparison

Aspect	Fundam. & Stat.	Option-Based	ML & Non-linear
Main data	Ratios, macro vars	Equity price, σ, debt	Mixed: ratios, text, market
Explainability	High (coeffs, logic)	Medium (model-based)	Low–Med. (XAI tools)
Accuracy (AU-ROC)	Medium	Medium	High
Regulatory fit	Basel/IRB-friendly	Partial fit	Requires XAI, validation
Cost to deploy	Low (simple, fast)	Moderate (data needs)	High (data + tuning)

Selecting the Right PD Model

- Model choice depends on:
  - Type of firm (listed vs. private)
  - Data availability
  - Need for interpretability vs. predictive power
- In practice, hybrid models are often used to balance strengths.

Future Trends in PD Estimation

- **Machine Learning Integration:** Increasing use of advanced algorithms for better predictive accuracy and pattern recognition.
- **Alternative Data Sources:** Incorporating non-traditional data (e.g., transaction data, social media, ESG scores) to enrich models.
- **Macroeconomic Sensitivity:** Greater emphasis on forward-looking, macro-adjusted PDs (e.g., PIT vs. TTC models).
- **Explainable AI (XAI):** Balancing model complexity with interpretability for regulatory compliance and trust.
- **Real-Time Risk Monitoring:** Leveraging big data and cloud computing for dynamic, high-frequency PD updates.

Part 2: Macroeconomic Default Modelling

What We've Covered So Far

- Firm-level PD estimation using **Fundamental & Classical**, **Option-Based** and **Machine Learning** techniques
- Comparative insights from PIT, TTC, and hybrid models
- Practical implications of model choice for PD accuracy and stability

What Comes Next

- How macroeconomic variables shape default dynamics
- Case study: Stress-testing US corporates under macro scenarios
- Estimating PD values under both base and stressed scenarios to see the impact

Let's explore how macro environments interact with credit risk and observe economy-wide PD forecasts.

Introduction

- Estimating probabilities of default is the first step in assessing the credit exposure and potential losses faced by financial institutions.
- Probabilities of default act as the basic inputs when evaluating systemic risk and stress testing financial systems in measuring the likelihood of defaults, especially under Basel II's capital adequacy framework.
- There are variety of models that attempt to measure the probability of default, e.g., macroeconomic-variable-based models.
- These models show that default rates in the financial, corporate, and household sectors increase during recessions.
- This observation has led to the implementation of econometric models that attempt to explain default indicators, such as probabilities of default or default rates, using macroeconomic variables.

What are Macro Economic Based Models for PD?

- Models that explicitly link the probability of default (PD) to macroeconomic variables such as GDP growth, unemployment, inflation, and interest rates.
- They use statistical or econometric methods to quantify how changes in the macroeconomy influence credit risk within firms, sectors, or entire portfolios.
- Useful for transforming point-in-time PDs into forward-looking and scenario-based estimates, aligning credit risk measurement with economic cycles.
- They enable financial institutions to evaluate the sensitivity of credit risk to adverse economic conditions and to incorporate systematic risk factors into their risk management frameworks.

Why Use Macro-Based PD Models?

- Default rates are not static and tend to rise during economic downturns and fall during expansions, driven by systematic macroeconomic conditions.
- These models enable risk managers to quantify this relationship, providing a more realistic estimation of credit risk across different phases of the economic cycle.
- They are essential for stress testing and scenario analysis, allowing the estimation of PDs under adverse macroeconomic conditions for regulatory capital planning.
- Forward-looking PD estimation using macro-based models aligns with IFRS 9 requirements for expected credit loss (ECL) calculations, improving provisioning accuracy.
- They enhance portfolio risk management by identifying sectors or borrower segments that are more vulnerable to macroeconomic shocks, allowing targeted risk mitigation strategies.

Advantages and Disadvantages of Macroeconomic Models

Chan-Lau (2006) lists three **advantages** of macroeconomic models:

- This type of model is very suitable for designing stress scenarios.
- Easily available and helpful for comparison purposes.
- The default rate used to estimate the model is observed historically, so that one can avoid making assumptions.

Disadvantages:

- They require a data time span longer than one business cycle to accurately capture the impact of the business cycle on default probabilities.
- These models are vulnerable to the Lucas critique, as their parameters and functional forms may not remain stable over time.
- Economic data often have substantial reporting lags, complicating the estimation and forecasting of macroeconomic models with current information.

Detailed Types of Models

- **Logistic Regression:** Models the log-odds of PD as a linear function of macroeconomic variables, handling the bounded nature of PD between 0 and 1.
- **Linear Regression:** Can be applied to transformed PDs, simpler but may predict values outside [0,1]. Useful for exploratory analysis.
- **Latent Factor Models:** Models systemic risk using unobservable factors influenced by macro variables, capturing correlation among defaults within sectors.
- **Vector Autoregression (VAR):** Captures the dynamic interdependence between macroeconomic variables and PDs, useful for scenario-based forecasting and stress testing.
- **Dynamic Factor Models:** Reduces dimensionality by extracting latent factors from macroeconomic variables, which are then used to model PDs with time dynamics.
- **Panel Data Models:** Utilise firm or sector-level panel data over time, allowing for the estimation of both time and entity effects, providing granular insights on macroeconomic drivers of PDs.

Categories of Macroeconomic Models

Macroeconomic models can be categorized as:

- **Exogenous Models:** Economic variables are independent of financial distress. The relationship between economic variables and default rates is modeled as:

  $$pd_t = g(x_1, x_2, \ldots, x_n) + \epsilon$$

  Drawback: Unrealistic assumption of a constant relationship between macroeconomic variables and default rates.

- **Endogenous Models:** Economic variables to be influenced by financial distress. The Vector Autoregressive (VAR) methodology is commonly used, expressed as:

  $$Z_{t+1} = \alpha_t + \sum_{j=1}^{p} \beta_j Z_{t+1-j} + \epsilon_{t+1}$$

  Inference in VAR models depends on the choice of lags: too many lags reduce degrees of freedom, while too few may omit important dependencies.

Advantages of Using a Logit Model in PD Estimation

- **Interpretability:** Maps macro predictors to PD clearly; coefficients have intuitive meanings.
- **Valid Probability Range:** Ensures PD estimates remain within [0,1].
- **Non-linear Effects:** Captures diminishing/increasing marginal effects realistically.
- **Statistical Robustness:** MLE-based, consistent, efficient estimation.
- **Panel and Dynamic Ready:** Supports AR(1) persistence, fixed effects, and latent factors.
- **Stress Testing Compatible:** Enables scenario-based PD shifts and Monte Carlo simulation.
- **Industry Standard:** Widely used in credit risk and Basel frameworks.

Summary: Logit models provide a robust, interpretable, and structurally appropriate framework for PD estimation and macro stress testing.

Data and Definitions

- **Default Rate:** The default rate is the ratio of the number of firms in default to the total number of firms in a given quarter.
- **Economy-wide default rate** ($pd_{t,0}$):

$$pd_{t,0} = \frac{\text{Number of defaults in all sectors}}{\text{Average number of firms in all sectors}}$$

- **Sector-specific default rate** ($pd_{t,i}$):

$$pd_{t,i} = \frac{\text{Number of defaults in sector } i}{\text{Average number of firms in sector } i}$$

Data and Definitions

- **Macroeconomic Variables:** The selected macroeconomic variables are chosen to assess risks and for stress test scenarios. These may not always have the highest explanatory power but are important for scenario analysis:
  - **Gross Domestic Product (GDP):** Reflects total demand in the economy. Lower GDP growth can reduce sales for firms, increasing the likelihood of default.
  - **Interest Rate:** Higher interest rates increase the cost of debt for firms, raising default risk.
  - **Exchange Rate:** Affects international operations. Higher exchange rates lower costs for importers but can increase prices for exporters, impacting their competitiveness.
  - **Stock Market Return and Volatility:** Per Merton's theory, stock market returns are negatively related to default probability, while volatility is positively related.
  - **Oil Price:** Higher oil prices increase operational costs for firms, potentially increasing default probabilities.

The Model (a)

- We use an **AR(1) logit linear model** to estimate the probability of default (PD) dynamically over time.
- Let the log-odds of the default probability at time t be:

$$(PD_t) = \log\left(\frac{PD_t}{1 - PD_t}\right)$$

- The model is specified as:

$$(PD_t) = \beta_0 + \beta_1(PD_{t-1}) + \beta_2 \cdot GDP_{t-1} + \beta_3 \cdot IR_{t-1} + \beta_4 \cdot HY_{t-1} + \epsilon_t$$

- Here:
  - β_0: intercept term.
  - β_1: captures AR(1) persistence in PD.
  - $\beta_2, \beta_3, \beta_4$: coefficients on macroeconomic drivers (GDP growth, interest rates, high-yield spread).
  - ϵ_t: residual error capturing unexplained risk.
- This structure enables capturing both the **persistence in defaults** and the **systematic influence of macroeconomic factors** on PDs.

The Model (b)

- After estimating the model parameters (β) via OLS, we use the model to:
  - Forecast future PDs recursively (dynamic 1-step ahead forecasts).
  - Simulate PD paths under baseline and stress scenarios.
- **Conversion back to PD:** After simulating or predicting logit values, we convert them back to probabilities:

$$PD_t = \frac{\exp((PD_t))}{1 + \exp((PD_t))}$$

- **Key Features of the Model:**
  - Captures **persistence** in PDs via the AR(1) term.
  - Reflects **systematic risk** via macroeconomic variables.
  - Allows for **scenario-based stress testing** using simulated shocks.

Dynamic Effects of Shocks in the Macro-PD Model

Short-Term Effect:

- A small, one-period macroeconomic shock ΔX leads to:

$$\Delta(PD_t) \approx \beta^\top \Delta X$$

  indicating an immediate, proportional impact of the shock on the log-odds of default.

Long-Term Effect:

- Due to AR(1) persistence (β_1), shock effect accumulates over time.
- The long-term cumulative impact on the logit PD becomes:

$$\Delta(PD) \approx \frac{\beta^\top \Delta X}{1 - \beta_1}$$

  showing that the closer β_1 is to 1, the slower the adjustment and the larger the long-term amplification.

Key Insight: The AR(1) coefficient (β_1) controls the speed of adjustment, determining how quickly PDs respond to macro shocks in the system.

Stress Testing: Theoretical Framework (a)

Objective:

- To evaluate how credit risk (PD) responds under adverse but plausible macroeconomic scenarios.
- Commonly used in regulatory frameworks (e.g., Basel) for assessing capital adequacy and risk resilience.

Stress Testing Process:

- Establish a **baseline scenario** with expected macroeconomic conditions.
- Define a **stress scenario** introducing adverse macroeconomic shocks (e.g., a decline in GDP, increase in interest rates).
- Apply the scenario to a credit risk model to simulate the impact on default probabilities.

Core Principle:

- Stress testing helps institutions quantify vulnerabilities in their portfolios by translating macroeconomic shocks into PD shifts.

Stress Testing: Theoretical Framework (b)

Model Specification for Stress Testing:

- Credit risk models typically use an autoregressive structure with macroeconomic drivers:

$$(PD_t) = \beta_0 + \beta_1(PD_{t-1}) + \beta_2 X_{t-1} + \epsilon_t$$

  where:
  - (PD_t): Log-odds of default at time t.
  - (PD_{t-1}): Persistence via AR(1).
  - X_{t-1}: Vector of macroeconomic variables (GDP growth, interest rates, etc.).
  - ϵ_t: Random disturbances capturing unexplained risk.
- Under a stress scenario, adverse shocks are applied to X_{t-1} to evaluate their impact on PD.

Key Concept:

- By simulating macro shocks within this framework, we observe how PDs dynamically adjust under stressed conditions.

Stress Testing: Theoretical Framework (c)

Simulation-Based Stress Testing:

- **Monte Carlo Simulation** is commonly used to model the distribution of PD outcomes under baseline and stress scenarios.
- The process involves:
  - Simulating multiple paths of macroeconomic variables with and without shocks.
  - Applying these paths to the credit risk model to generate PD projections.
  - Converting logit values back to PD:

$$PD_t = \frac{\exp((PD_t))}{1 + \exp((PD_t))}$$

  - Aggregating results to analyze the distribution of mean PDs (e.g., 1-year ahead) under each scenario.
- This allows comparison of metrics (mean, percentiles) between baseline and stress scenarios to quantify credit risk sensitivity.

Purpose: To assess the potential impact of macroeconomic stress on portfolio credit risk in a systematic, quantifiable manner.

Properties of the Data Used

- **Time Span:** Quarterly data from **2015-Q1 to 2024-Q4**. (US Data)
- **Data Structure:**
  - Sector-level panel capturing credit risk indicators.
  - Observations prepared as logit-transformed default probabilities for numerical stability and boundedness.
  - Includes lagged PD values to track changes over time.
- **Macroeconomic Variables Included:**
  - GDP growth (quarter-over-quarter percentage changes)
  - 3-month interest rates
  - High-yield credit spreads
- **Data Preparation:**
  - Missing values from lag construction removed.
  - All variables aligned quarterly for consistent analysis.
  - Dataset cleaned for modeling and time series analysis workflows.

AR(1) Logit PD Model: Estimated Parameters

- Estimated via manual OLS on logit-transformed PD.
- Model specification:

$$(PD_t) = c + \phi(PD_{t-1}) + \beta_1 GDP_{t-1} + \beta_2 IR_{t-1} + \beta_3 HY_{t-1} + \epsilon_t$$

- Number of observations: **n = 39**

Variable	**Coef**	**StdErr**	**t**	**P**
Constant	-5.2010	1.2611	-4.12	0.000
logit_pd_lag1	+0.3802	0.1510	2.52	0.017
gdp_qoq_pct	+0.0019	0.0240	0.08	0.939
t3m_rate	+0.0284	0.0247	1.15	0.259
hy_oas	+0.0826	0.0471	1.76	0.088

Table: Manual OLS Estimates

Signs mildly consistent with expectations: GDP (-), interest (+), HY (+).

Interpretation of OLS Estimation Results

- **Constant** is negative and significant ($p < 0.01$): low baseline PD level.
- **Lagged logit(PD)** is positive and significant ($p = 0.017$):
  - Confirms persistence in default probabilities across quarters.
- **GDP growth** is positive but highly insignificant ($p = 0.939$):
  - Coefficient near zero; no meaningful impact on PDs.
  - Sign contradicts expectations (expected negative).
- **3-month rate** is positive but insignificant ($p = 0.259$):
  - Suggests potential link between funding costs and PDs, but evidence is weak.
- **HY-OAS** is positive and borderline significant ($p = 0.088$):
  - Indicates PDs rise with credit market stress (wider spreads).
- **Conclusion**: Persistence dominates; macro variables show expected signs (except GDP), but lack strong explanatory power in this sample.

Persistence and Diagnostics

Persistence (AR(1)):

- Estimated:

$$\hat{\phi} = 0.3802$$

  indicating moderate persistence in PDs with relatively fast adjustment to macro shocks.

Diagnostics:

- **Tail Risk:** 99.5% Gaussian = 0.6812, Student-t (df=6) = 0.9804, Fat-tail multiplier = 1.44 (✓)
- **Out-of-Sample MAPE:** 17.2% (2023-Q1 to 2024-Q4) (✓)
- Indicates reasonable predictive performance with AR(1) structure enhancing stability.

Monte Carlo Stress Testing

- We use a **Student-t residual shock** based simulation engine:

$$\mathrm{logit}(PD_t) = c + \phi \cdot \mathrm{logit}(PD_{t-1}) + \beta^\top X_{t-1} + \epsilon_t$$

  with $\epsilon_t \sim t(df = 6)$, scaled to model residual variance.
- Simulation Parameters:
  - **200,000 paths**
  - **8 quarters (2 years) horizon**
  - **Shocks applied:** GDP –1% T3m -0.3% HY-OAS +0.5 % for 2 quarters
- Baseline scenario and mild GDP shock scenario are compared to evaluate default rate sensitivity.
- PD paths are converted from logit scale:

$$PD_t = \frac{\exp(\mathrm{logit}(PD_t))}{1 + \exp(\mathrm{logit}(PD_t))}$$

  and aggregated to compute 1-year-ahead mean PD distributions.

Quantitative Results of Stress Tests

Scenario Comparison:

- **Baseline Scenario:** Latest macro conditions without additional shocks.
- **Stress Scenario:** GDP –1 %, T3m –0.30 %, HY-OAS +0.50 % for two quarters.

Key Quantitative Results (1-Year Ahead):

	Baseline	Mild Shock	Change
Median PD (%)	0.045	0.046	+0.001
95th Percentile PD (%)	0.066	0.068	+0.002
99th Percentile PD (%)	0.082	0.084	+0.002

Interpreting the Coefficients & Stress-Test Outcome

Macro elasticities

GDP growth: $\beta_1 = +0.0020$ ($p = 0.94$): *statistically zero* – PD is insensitive to a 1-pp QoQ GDP swing.

3-month rate: $\beta_2 = +0.0284$ ($p = 0.26$): higher policy rates raise PD only marginally; the effect is economically tiny.

HY-OAS: $\beta_3 = +0.0826$ ($p = 0.09$): largest macro coefficient, yet still small impact on PD.

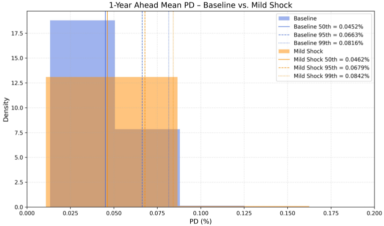
Resulting 1-year PD shift under the mild shock

- Median: **0.045 %** $\rightarrow$ **0.046 %** (+0.001 pp)
- 99^{th} percentile: **0.082 %** $\rightarrow$ **0.084 %** (+0.002 pp)
- Histogram shows a uniform, almost imperceptible right-shift.

Take-away: Lagged PD dominates short-run dynamics; macro coefficients are too small to matter unless the scenario is far more severe (multi-year GDP contraction or multi-percentage-point spread blow-out). The current mild recession-style shock shows small impact on PD which is consistent to the previous findings.

Interpreting the Stress-Test Chart

- The chart compares the **1-year-ahead mean PD** under:
  - **Baseline (blue)** no macro shock.
  - **Mild shock (orange)** GDP –1% rate –0.30 %, HY-OAS +0.50 % for two quarters.
- **Vertical lines:** 50th (solid), 95th (dashed) and 99th (dotted) percentiles.
- **Observation:**
  - PDs cluster around 0.045
  - Shock shifts the entire distribution slightly rightward by roughly 0.002 pp at the tail.
- Consistent with *Simons & Rolwes (2009)*: modest, short-lived GDP shocks translate into limited near-term PD movement.



Takeaway: Under a mild recession-style shock, incremental credit-risk changes are economically negligible; harsher scenarios are needed to reveal material risk shifts.

Limitations and Next Steps

Limitations:

- Lack of sectoral-wise analysis.
- No correlation modeling across macro shocks.
- Mild scenario may understate severe crisis impact.

Next Steps:

- Multi-scenario and multi-variable stress testing.
- Variance-based sensitivity decomposition.