TIME_SERIES_FORECASTING

August 29, 2025

Forecasting Financial Risk Metrics: From Holt-Winters to Gradient Boosting

This notebook demonstrates a comparison of Holt-Winters exponential smoothing and XGBoost for forecasting charge-off and recovery amounts in consumer lending. The project covers complete model development lifecycle including EDA, hyperparameter tuning, cross-validation, and performance evaluation, with practical applications for credit risk management.

1. Exploratory data analysis including time series decomposition and stationarity testing
2. Individual time series visualization and comparative analysis
3. Generalized trend analysis using HP Filter model
4. Holt-Winters model selection and hyperparameter tuning to train, test and forecast 12 months of predictions
5. Recovery Ratio Analysis and Business Implications

1.0 Exploratory Data Analysis and Time Series Characterization

Overview This analysis utilizes a monthly time series dataset spanning 78 periods from January 2019 through June 2025, containing charge-off balances (CO_BAL), recovery balances (REC_BAL), and recovery ratios for an unsecured personal loan portfolio. The dataset captures 6.5 years of financial performance data with complete monthly observations.

Key Dataset Characteristics

- Charge-off balances average $705,805 with high volatility (std: $325,979), ranging from $229,726 to $1.43 million, indicating significant fluctuations in portfolio losses over the observation period
- Recovery balances average $200,030 with lower relative volatility (std: $78,735), ranging from $81,281 to $533,886, demonstrating more stable collection performance
- Recovery ratios average 33.8% with substantial variation (std: 0.17), ranging from 7.8% to 82.2%, reflecting the cyclical nature of collection effectiveness and varying market conditions
- Data completeness is 100% with no missing values across all variables and consistent monthly frequency

The substantial range and standard deviation across the data variables indicate the presence of trend, seasonality, and volatility patterns typical of financial time series data. As a result, fundamental analysis of the time series data should be performed to ensure accurate forecasting.

1.1 Stationary Testing The statistical tests reveal that charge-offs and recoveries behave very differently over time. Charge-offs failed the stationarity test (p-value: 0.989), indicating they follow trends and can continue growing or declining without reverting to a typical level. This means charge-offs don't have a stable long-term average they return to - they can drift upward or downward and stay at those new levels.

Recovery balances passed the stationarity test (p-value: 0.000), showing they consistently return to their historical average of around $200,030. When recoveries spike above or drop below this typical range, they tend to move back toward the center rather than establishing new permanent levels.

This difference means the two series require different forecasting approaches. Charge-offs need models that can handle ongoing trends and growth patterns, while recoveries can use simpler methods since they naturally fluctuate around their long-term average. Understanding this distinction helps explain why charge-offs may be harder to predict than recoveries.

1.2 Seasonality & Decomposition Analysis The seasonal decomposition reveals distinct patterns in both charge-offs and recoveries. For charge-offs, the trend component shows a clear U-shaped pattern with a decline from 2019 through mid-2021 (COVID), followed by steady growth from 2022 onward (Fed Funds Rate Increase), reaching approximately $1.1 million by 2025. The seasonal component demonstrates consistent annual cyclicality with regular peaks and troughs throughout each year, while the residual component shows significant volatility, particularly during 2020-2021 and 2023, indicating periods of unusual activity beyond normal seasonal patterns.

Recovery data exhibits different characteristics with a more complex trend pattern that shows initial decline through 2021, followed by recovery and growth from 2022-2024, then stabilization around $240,000 by 2025. The seasonal component displays regular annual patterns similar to charge-offs but with smaller amplitude variations. The residuals show notable spikes, particularly in 2020 and 2023, suggesting external factors significantly impacted recovery performance during these periods.

The decomposition confirms that both series contain meaningful seasonal and trend components that can be modeled, with charge-offs showing stronger trending behavior (supporting the non-stationarity finding) while recoveries demonstrate more bounded variation around evolving trend levels (consistent with eventual mean reversion). The substantial residual components in both series indicate the presence of irregular shocks that pure trend and seasonal models may struggle to capture.

1.4 Correlation Analysis & Outlier Detection The scatter plot reveals a weak positive correlation (0.187) between charge-offs and recoveries, indicating that higher charge-off periods are associated with slightly higher recovery amounts, though the relationship is not strong. The temporal clustering provides more insight than the correlation itself, showing distinct behavioral patterns across different economic periods.

The lower-left cluster dominated by 2020-2021 observations (red and orange points) reflects the pandemic period when regulatory forbearance and collection moratoriums significantly suppressed both charge-offs and recoveries. During this time, charge-offs remained artificially low due to extended delinquency periods (120-day extensions to 180+ days) while recovery efforts were constrained by regulatory restrictions and reduced collection activity.

In contrast, the upper-right portion shows increased density of 2024-2025 observations (dark blue points), indicating a return to normal collection practices combined with deteriorating credit conditions. This rightward shift suggests charge-offs have resumed normal patterns while moving to higher absolute levels, likely driven by macroeconomic pressures including elevated interest rates and inflationary impacts on borrower payment capacity.

The single recovery outlier at approximately $534K represents an exceptional collection event that falls well outside normal operational parameters. The overall pattern demonstrates how external economic and regulatory factors can create distinct regime shifts in the charge-off/recovery relationship, with implications for model stability across different market conditions.

The annual correlations show highly unstable relationships between charge-offs and recoveries, ranging from strongly negative (-0.613 in 2024) to moderately positive (0.425 in 2021). The 2021 peak correlation likely reflects synchronized pandemic recovery effects, while 2024's strong negative correlation suggests charge-offs increased substantially without corresponding recovery improvements.

2.0 Individual Series Analysis and Visualization

This section presents individual time series visualizations and comparative analysis to examine the temporal patterns, volatility, and relationships between charge-offs and recoveries across the 78-month observation period. The visualizations reveal distinct behavioral differences between the two series and highlight regime shifts corresponding to major economic events from 2019-2025.

2.1 Historical Charge-Off Performance Gross charge-offs declined significantly from 2019 through mid-2021, reaching a low of approximately $230K during the pandemic forbearance period. From 2022 onward, charge-offs have increased dramatically, rising from these historic lows to over $1.4M by 2025, representing more than a 500% increase from the trough levels.

2.2 Historical Recoveries Performance Recovery amounts show high volatility throughout the period, fluctuating between approximately $80K and $534K without a clear directional trend. The data shows "mean-reverting" behavior that hovers near $200K, with spikes in early 2019, mid-2020, and a significant peak exceeding $530K in 2023, followed by stabilization in the $200K-$330K range through 2025.

2.3 Historical Charge-Off and Recoveries Combined Performance The combined visualization shows the dramatic impact of COVID-era regulatory interventions on both charge-offs and recoveries from mid-2020 through mid-2022. The shaded period highlights when forbearance measures and softened colelction efforts artificially suppressed normal credit loss patterns, creating a distinct trough that contrasts sharply with pre-pandemic and recovery-phase behaviors.

COVID Period (Shaded Area):

- Charge-offs dropped to historic lows around $230K-$400K
- Recoveries remained relatively stable but constrained in the $100K-$250K range
- Both series show reduced volatility during regulatory intervention

Post-COVID Divergence:

- Charge-offs surge dramatically from 2022 onward, reaching $1.4M+ by 2025
- Recoveries return to pre-pandemic patterns, fluctuating around $200K-$300K with occasional spikes

2.4 Ratio Visualization and Trend The recovery ratio analysis reveals three distinct operational regimes with dramatically different collection effectiveness patterns. During the pre-pandemic period, ratios remained relatively stable around 30%, reflecting normal collection operations. The COVID moratorium period shows initial volatility followed by improving ratios that peaked near 80% as charge-offs were artificially suppressed while some recovery activities continued. The post-2022 period demonstrates a sharp deterioration in collection effectiveness, with ratios declining to 15-20% as charge-offs surged while recovery amounts failed to scale proportionally.

Key Insights:

- Pre-COVID Baseline (2019-2020): Recovery ratios maintained steady 30% average with moderate volatility, indicating normal collection operations
- COVID Peak Distortion (2020-2022): Ratios artificially inflated to 50-80% due to regulatory forbearance suppressing charge-offs while maintaining some collection activity
- Post-Pandemic Deterioration (2022-2025): Sharp decline in collection effectiveness to 15-20% as charge-offs resumed aggressive growth without proportional recovery scaling
- Regime Instability: The dramatic shifts across periods indicate that any modeling assumption of stable recovery relationships will likely fail across different market conditions
- Collection Capacity Constraints: The declining trend post-2022 suggests collection operations may be overwhelmed by the surge in charge-off volumes

3.0 HP Filter Trend Analysis and Visualization

The HP Filter provides is one approach to decompose time series data into long-term trend and short-term cyclical components – this can help identify enable the identification of underlying structural patterns. This analysis applies HP filtering to both charge-off and recovery data to isolate the general direction of the data movements outside from temporary fluctuations outside. The decomposition helps distinguish between temporary regime shifts caused by external events (such as COVID-19) and genuine structural changes in the underlying data.

3.1 Charge-Off Trend Decomposition The charge-off trend was generated using a higher smoothing parameter lambda = 1000, rather than a lambda = 129600 (standard for monthly data). This allows the trend to be more responsive to the actual directional changes in your data, particularly capturing the U-shaped recovery pattern from the 2021 trough. The trend and 95% confidence interval bands appear to appropriately capture the range of charge-offs with only 2 data points falling outside the 95% confidence bands out of 78 total observations. This results in a 2.56% breach rate, which is well within the expected 5% for 95% confidence intervals.

3.2 Recovery Trend Decomposition The recovery trend was generated using a lower smoothing parameter lambda = 129600 which is typical for monthly data. The trend and 95% confidence interval bands appear to appropriately capture the range of charge-offs with only 2 data points falling outside the 95% confidence bands out of 78 total observations. This results in a 2.56% breach rate, which is well within the expected 5% for 95% confidence intervals.

3.3 Combined Charge-offs and Recoveries The overlapping confidence bands during the COVID period reveal how regulatory interventions fundamentally disrupted normal credit patterns. When charge-offs were artificially suppressed through forbearance or forgiveness measures, recoveries appear to have continued at typical levels - the usual gap between these metrics disappeared temporarily overriding the natural market relationship.

The dramatic separation of confidence bands after 2022 demonstrates the return to normal operational patterns, but with charge-offs resuming at permanently elevated levels. While recovery bands remained relatively stable throughout the entire period, charge-off bands expanded significantly as the series moved from suppressed pandemic levels to new highs. This pattern confirms that the COVID period represented a temporary disruption rather than a structural change in the underlying relationship between charge-offs and recoveries.

4.0 Forecast Balances with Holt-Winters Exponential Smoothing

This section applies Holt-Winters exponential smoothing to generate 12-month forecasts for charge-off and recovery balances. The approach includes model fitting, validation using train/test splits, and forecast generation with confidence intervals. Due to the different statistical properties identified in previous sections, charge-offs and recoveries require distinct modeling strategies to optimize forecast accuracy.

4.1a Charge-Off Model Selection and Hyperparameter Tuning The baseline Holt-Winters model demonstrates reasonable performance with an R-squared of 0.676, indicating the model explains approximately 68% of the variance in charge-off data. The 21.61% MAPE on training data provides a benchmark for in-sample accuracy, while the RMSE of $149,938 quantifies the typical forecast error magnitude. This initial model uses default parameters that may not be optimal for the specific characteristics of charge-off data. To improve forecast accuracy, a systematic hyperparameter tuning process will test different combinations of trend types (additive vs multiplicative), seasonal patterns, and damping parameters to identify the configuration that minimizes forecast error while maintaining model stability.

The hyperparameter tuning initially suggested that multiplicative components outperform additive approaches based on training metrics, with multiplicative trend and seasonal components achieving the best in-sample performance (R^2: 0.67, MAPE: 21.35%). However, out-of-sample validation revealed critical generalization issues with the multiplicative model, producing a negative R^2 (-1.032) and systematic over-forecasting despite reasonable MAPE performance. When tested on holdout data, the additive trend and seasonal configuration demonstrated superior forecasting accuracy with an 11.65% MAPE and positive R^2 (0.408), representing a 45% improvement in forecast error. This validates the importance of out-of-sample evaluation in model selection, as training performance alone can be misleading for volatile financial time series. The additive model (trend='add', seasonal='add', damped=False) will be used for final charge-off forecasting due to its superior generalization performance and minimal forecast bias.

4.1b In-Sample Model Performance The model demonstrates strong in-sample performance with perfect mean alignment of $619,944 for both actual and fitted values. This confirms the predictions are unbiased. The slightly lower standard deviation in fitted values ($249,862 vs $265,320) indicates the model appropriately smooths extreme volatility while preserving the overall variance structure.

The plot shows the fitted values tracking actual data closely throughout most periods -even effectively capturing the U-shaped pattern from the COVID trough through recent recovery. While some deviations occur during high-volatility periods where the model (The model appears to lag behind rapid directional changes), the overall alignment between actual and fitted values suggests the multiplicative configuration has learned underlying patterns rather than memorizing noise.

4.1c Out-Of-Sample Model Performance The model's forecasts remain within confidence bounds, indicating it effectively captures uncertainty. The additive model smooths volatility while preserving trend, and test predictions stay within statistical limits, which is important for avoiding outlier-driven planning errors. These results support the model choice: with 11.65% MAPE and positive R^2 of 0.408, the additive trend + seasonal configuration delivers a reasonably accurate and reliable forecast.

4.1d Future Period Forecasting The future forecast shows charge-offs continuing their upward trajectory from current levels around \$1.4M toward \$1.7M by mid-2026, with seasonal patterns creating temporary peaks and valleys. The confidence intervals appropriately widen over the forecast horizon, ranging from roughly \$1.4M to \$1.9M by the end of the period, acknowledging inherent uncertainty while maintaining realistic bounds based on the model's validated error structure. The entire yellow shaded area represents plausible scenarios, with the central forecast serving as the most likely outcome rather than a definitive prediction, requiring risk management strategies that account for the full range of potential charge-off levels.

The orange linear trend line clearly demonstrates a persistent upward trajectory in charge-offs. Regardless of estimated short-term or seasonal fluctuations, risk management should plan for continued growth in loss levels over the forecast horizon.

4.2a Recoveries Model Selection and Hyperparameter Tuning This baseline recovery model shows weak performance with an R-squared of only 0.308, indicating the model explains less than one-third of the variance in recovery data. The 25.26% MAPE is substantially higher than the charge-off model's performance. This poor fit likely identifies challenges for modeling stationary recovery data using exponential smoothing methods designed for trending series.

The hyperparameter tuning has identified the baseline model of multiplicative components to achieve the best performance. As aresult, this model will be used going forward.

4.2b In-Sample Model Performance The recovery model shows perfect mean alignment (\$193,619 for both actual and fitted values) but significant over-smoothing, with fitted standard deviation (\$44,550) representing only 57% of the actual standard deviation (\$78,697). The plot appears to confirm this issue - while the green dotted fitted values track the general level, they fail to capture the substantial month-to-month fluctuations that characterize recovery data. As a result, this over-smoothing explains the poor R-squared (0.308) despite unbiased mean prediction.

4.2c Out-Of-Sample Model Performance The test results show significant variation and forecasting failure. The 64.09% MAPE is nearly three times worse than training performance (25.26%), while the R^2 of -4.665 indicates forecasts perform nearly five times worse than simply using the mean. The visual confirms the issue - the purple test predictions show extreme volatility, including forecasts dropping to near zero, which is operationally impossible for recovery operations. Finally, the 0.88x ratio shows systematic under-forecasting, but this is misleading given the wild swings in predictions.

As a result, the Holt-Winters model is fundamentally inappropriate for recovery data, as evidenced by the dramatic deterioration from training to test performance. The model is essentially attempting to impose trend and seasonal structure on inherently volatile, stationary data. So, alternative approaches should be approached.

To perform this test, we will start with an ARIMA model, then a SARIMA and finally, extrapolate off the HP Filter trend data an estimate forecast and confidence interval band.

No need to plot the data. We can already see its operating in a very narrow band. Lets try a Sarimax.

Band is still too narrow. The data may be too volatile for forecasting. The best approach might be the simplest approach - leverage our hp filter trend to forecast out.

4.7a Forecasting on full dataset