EN - Public Report - Módulo 3

Sprint 2 Summary

Date: August 17–29, 2025

Sprint Objective

To adjust and implement a new simulation pipeline to estimate, on a monthly horizon, **(i)** the probability of recession at each future step and **(ii)** the probability of **at least one** recession occurring within that horizon.

The approach combines a Probit model for the binary event with a VAR(1) model for predictor dynamics and uses Monte Carlo Simulation to propagate uncertainties.

Methodological Approach (Overview)

- **Probit Regression** — models $P(Recession_t = 1 | X_t) = \Phi(\beta_0 + X_t \beta)$. Coefficients estimated by MLE via Newton–Raphson.
- **VAR(1)** — models the temporal evolution of predictors: $X_t = c + A X_{t-1} + \varepsilon_t$. Parameters estimated by OLS.
- **Monte Carlo** — samples Probit coefficients (to capture uncertainty in β) + simulates multiple future paths of predictors via VAR(1), computes recession probabilities at each path, and aggregates statistics (means, quantiles, probability of any recession).

Implementation (Main Components and Functions)

- **Utility Functions**
  - `std_norm_cdf`: Standard normal CDF (uses `scipy.stats.norm.cdf` when available; fallback via `erf`).
  - `_probit_mle`: Probit MLE with Newton–Raphson — computes log-likelihood, gradient, and Hessian; adds small regularizations and fallback via `pinv` when necessary.
  - `fit_var1`: Estimates c, A, and Σ (innovation covariance) via OLS with minimum sample check.
  - `simulate_var1`: Generates future X paths using Cholesky decomposition with eigenvalue fallback to ensure numerical robustness.
- **Main Function:** `monte_carlo_recession`
  - Validates input (target column and predictors).
  - Standardizes predictors (z-scores) for numerical stability.
  - Fits Probit and VAR(1).
  - Samples β from an asymptotic Normal (`vcov`) and simulates `n_sim` predictor paths.
  - For each path, computes $P_t = \Phi(\beta_i' Z_t)$ and estimates whether at least one recession occurred (simulating Bernoulli per period).
  - Aggregates results: mean, percentiles (5% and 95%), probability of any recession, β statistics (mean/std), and Average Marginal Effects (AME) at the sample mean.

- Implements numerical safeguards: jitter for near-singular matrices, probability bounds, and matrix inversion alternatives.

- **Minimal Example** under `if __name__ == "__main__":` generates synthetic data, runs the simulation (horizon = 6, n_sim = 2000), and prints sample results.

Results Delivered in This Sprint

- Complete and modular pipeline for probabilistic recession simulation integrating Probit + VAR(1) + Monte Carlo.
- Robust implementation with numerical treatments (jitter, Cholesky/eigendecomposition fallback, pinv).
- Consolidated output (`MonteCarloOutput`) containing:
  - `prob_mean`: mean recession probability per horizon step.
  - `prob_q05` / `prob_q95`: 5% / 95% intervals per step.
  - `prob_any_recession`: estimated probability of at least one recession within the horizon.
  - `beta_mean` / `beta_std`: summary of β samples.
  - `ame_at_mean`: Average Marginal Effects of predictors.
  - `details`: metadata (Probit convergence, iterations, VAR1 parameters, feature means/stds, number of simulations, etc.).
- Functional test with synthetic data demonstrating full execution and expected outputs.

Technical Notes and Limitations

- **"Any recession" computation:** the current code evaluates the occurrence of at least one recession per path by simulating Bernoulli draws each period using `rng.uniform() < p_t`.

  A deterministic alternative would be `1 - ∏(1 - p_t)` (conditional expectation). The two have distinct interpretations — the first simulates binary realizations per period, the second computes the aggregated probability conditional on p_t.

- **VAR uncertainty:** currently, parameter uncertainty in VAR(1) is not explicitly sampled (only β uncertainty via the Probit `vcov` is captured). Capturing dynamic uncertainty would require bootstrap or Bayesian sampling of VAR parameters.
- **Sample size and stability:** VAR(1) requires a minimum series length (check already implemented). In small samples, estimates of Σ and A may be noisy.
- **Probit assumptions:** observations are independent conditional on X_t; if there is autocorrelation in the Probit error or unmodeled dependencies, estimates may be biased.
- **Performance:** large `n_sim` increases computational cost; the function is vectorized when possible, but simulations can be parallelized for speed gains.

Sprint 3 Summary

Date: September 1–12, 2025

Sprint Objective

The third sprint focused on **comparing the Monte Carlo Simulation model with other probabilistic methods**, evaluating performance, robustness, and ability to capture economic recessions.

Additionally, **new modeling and variable transformation approaches** were explored to improve accuracy and stability.

Methodology

Five models with distinct statistical and temporal characteristics were tested:

Model	Type	Temporal Evolution	Sampling	Accuracy	Precision	Recall
Multiple Regression	Linear Statistical	No	No	N/A	N/A	N/A
Logistic Bootstrap	Nonlinear Statistical	Yes	No	1.0	0.0	0.0
Markov Chain	Temporal Probabilistic	No	Yes	0.556	0.0	0.0
Threshold VAR (TVAR)	Temporal Statistical	No	Yes	1.0	1.0	1.0
Monte Carlo	Probabilistic Simulation	Yes	Yes	0.875	1.0	0.875

Each method was evaluated using performance metrics (accuracy, precision, recall, F1-score, Brier Score, and Log Loss), allowing detailed comparative analysis of recession prediction capacity.

Main Results and Interpretation

Multiple Regression

- **R^2 (0.845):** Explains 84.5% of data variance; strong performance for linear relationships.
- **Limitation:** Does not capture temporal dynamics or regime changes — inadequate for complex economic forecasting.

Logistic Bootstrap

- **Accuracy (1.0)** and **Precision/Recall (0.0):** Predicts only the dominant class (no recession).
- **Limitation:** Suffers from **class imbalance**, failing to identify actual recessions.

Markov Chain

- **Accuracy (0.556):** Near-random performance.
- **Brier Score (0.253), Log Loss (1.113):** Indicate poor probabilistic calibration.
- **Limitation:** State transition assumption based solely on previous period is too simplistic for complex economic data.

Threshold VAR (TVAR)

- **Accuracy, Precision, Recall, F1 (1.0):** Perfect performance, capturing regime shifts.
- **Brier Score (0.354):** Good calibration but sensitive to chosen threshold and lag length.
- **Limitation:** May overfit small samples.

Monte Carlo (current model)

- **Accuracy (0.875), Precision (1.0), Recall (0.875), F1 (0.933):** Excellent balance between identification and precision.
- **Brier Score (0.173):** Well-calibrated probabilities.
- **Log Loss (0.538):** Moderate, consistent confidence in predictions.
- **Limitation:** Computationally intensive and sensitive to input data quality.

Comparative Summary

- **Monte Carlo:** best balance between precision and recall; provides rich, interpretable probabilistic forecasts.
- **Threshold VAR:** top binary classification performance but parameter-dependent.
- **Markov Chain:** simple temporal structure but low predictive power.
- **Logistic Bootstrap:** fails to handle class imbalance.
- **Multiple Regression:** useful for static explanation, not for dynamic forecasting.

Sprint 3 Conclusions

- The **Monte Carlo model** remained robust, showing **strong probabilistic metrics and stability** even with new variables and transformations.
- **Markov Chain** and **Logistic Bootstrap** exhibited limitations in capturing actual recessions, reinforcing the advantage of models incorporating temporal dependence and multivariate uncertainty.
- **Threshold VAR** achieved top performance but requires further validation to avoid overfitting.
- Variable normalization and preprocessing improved **model stability and probabilistic performance**.

Sprint 3 consolidated the empirical comparison of probabilistic methods, confirming that the Monte Carlo Simulation model offers the best balance between performance and interpretability.

The study reinforces its **advantage in forecasting recessions**, particularly by integrating temporal dependence and parameter uncertainty.

Although Threshold VAR achieved maximum classical performance metrics, Monte Carlo maintains greater flexibility and robustness, justifying its continuation and refinement in the next module.

Sprint 4 Summary

Date: September 14–26, 2025

Sprint Objective

The fourth sprint aimed to **evaluate the resilience and generalization of the Monte Carlo Simulation model** through three main approaches:

1. **Cross-validation** — measuring discriminative power and forecast calibration.
2. **Stress testing** — testing model response under extreme economic scenarios (severe, mild, and boom conditions).
3. **Rolling validation (temporal continuous validation)** — assessing stability across different historical windows, simulating series forecasts.

The final goal was to **detect possible overfitting** and **measure predictive consistency** under adverse situations.

Applied Methodology

Three types of tests were conducted based on the `monte_carlo_recession` model code developed in previous sprints:

🔷 Cross-Validation (TimeSeriesSplit)

- **Split:** 5 temporal folds.
- **Metrics:**
  - **AUC (Area Under the Curve):** measures ability to distinguish recessions (positive class) from non-recessions.
  - **Brier Score:** measures probability calibration (lower is better).

🔷 Stress Testing

- Simulation of three extreme macroeconomic scenarios:
  - *Mild recession:* moderate GDP and CLI drop.
  - *Severe recession:* sharp GDP, CLI, and USSLIND fall.
  - *Boom:* strong economic growth and indicator improvement.
- For each scenario, new Monte Carlo simulations with 2000 iterations were performed.

🔷 Rolling Validation

- **Structure:** initial window of 120 observations, with sequential horizon-by-horizon predictions.
- **Metrics:** AUC and Brier Score per window to assess temporal stability.

Results Obtained

Cross-Validation

Metric	Mean Value	Interpretation
AUC	*nan*	The model failed to produce useful discrimination values (class separability issue).
Brier Score	**0.423**	Indicates moderate calibration — probabilities are somewhat distant from actual outcomes.

Performance was limited. The absence of AUC shows difficulty in distinguishing recessions from non-recessions, while the Brier Score indicates predictions are somewhat calibrated but still imprecise.

Stress Testing

Scenario	Estimated Recession Probability
Mild recession	1.000
Severe recession	1.000
Boom	1.000

The model reacted deterministically, assigning 100% recession probability in all scenarios — even during expansion periods.

This extreme response may indicate:

- **Overfitting** to historical conditions.
- **Lack of sensitivity** to input variations across economic regimes.

Rolling Validation

Metric	Mean Value	Observations
Mean AUC	*nan*	The model lacked discriminative capacity across temporal windows.
Mean Brier	**0.434**	Slightly worse calibration than in cross-validation.
Brier_all	0.0 → 1.0	High variability, indicating temporal instability.

In some windows, the model produced nearly perfect forecasts; in others, it failed completely.

This demonstrates lack of temporal consistency — possibly due to dependence on specific training periods or highly time-correlated variables.

Sprint 4 Conclusions

- **Cross-validation:** limited performance and absence of AUC indicate that the model **does not discriminate economic states well**.
- **Stress testing:** deterministic behavior (100%) suggests **overfitting** and **low adaptability to new regimes**.
- **Rolling validation:** unstable performance over time, with large Brier Score variability.

The Monte Carlo Simulation model shows a solid probabilistic structure but still lacks satisfactory generalization.

It tends to react in extreme and poorly calibrated ways when exposed to new or stressed economic conditions.

Tests revealed generalization and stability issues, with deterministic stress responses and low temporal discrimination in cross and rolling validations.

Despite these limitations, the process was crucial for identifying structural weaknesses and guiding methodological refinements for the final project stage, which will aim to increase robustness, calibration, and sensitivity under distinct economic scenarios.

Sprint 5 Summary

Date: September 29 – October 9, 2025

Sprint Objective

The main objective of Sprint 5 was to **consolidate the results obtained in previous sprints and formally integrate them into the Final Course Project (TCC)**.

This phase focused on the **systematization of the technical and analytical content** developed in earlier modules into structured sections of the final academic document, presenting:

1. The **practical structure of models integrated with Monte Carlo Simulation**;
2. The **comparison between Logistic and Probit regressions** and their economic implications;

3. The **counterproof analysis of models** against the Monte Carlo Simulation;
4. And the **detailed presentation of obtained results**, including metrics, validations, and theoretical interpretations.

The general goal of this sprint was to **transform the project's empirical and experimental results into documented scientific content** with academic rigor and methodological coherence.

Development and Content Produced

During Sprint 5, four new sections were written and added to the TCC, based on the technical work developed throughout Sprints 2, 3, and 4.

Sprint 5 Conclusion

Sprint 5 marked the **final integration of technical and theoretical content into the TCC document**, highlighting:

- Structuring of sections **4.0 to 5.0**, incorporating methodology, comparisons, and results;
- Adaptation of technical reports (Sprints 2–4) to academic language;
- Organization of metrics, tables, and analyses into scientific format;
- Consolidation of conclusions and recommendations for future work.

These advances established the proposed model as a **comprehensive probabilistic system for recession forecasting**, validated through multiple approaches and with critical analysis of its limitations and strengths.

Sprint 5 marked the **transition from the experimental phase to the scientific documentation phase**.

The Monte Carlo Simulation model, initially developed and tested in independent modules, was formally described, justified, and compared in performance terms.

As a result, the TCC now incorporates not only quantitative outcomes but also the **theoretical and methodological foundation necessary** to consolidate the research as an original contribution to the study of **recession unpredictability and economic cycle randomness**.

Conference Submissions:

https://github.com/Inteli-College/2025-1A-T01-G18-INTERNO/tree/main/Modulo%203/Sprint%205/Submissao_Congressos

- **CNMAC-2025 —** National Congress on Applied and Computational Mathematics
- **ICSE SEIP —** 47th International Conference on Software Engineering
- **ICSE 2026 NIER Track —** 48th International Conference on Software Engineering