



ACTUARIAL SOCIETY
OF SOUTH AFRICA

Advancing Actuarial Excellence

Predicting the Pulse of the JSE

Can Machine Learning Edge Out the Historical Average Equity Risk Premium?

Nabila Patel, Furqaan Khan, and Eden Gross

School of Statistics and Actuarial Science, University of the Witwatersrand
ASSA Retirement Matters Seminar 2026



01

The Core Question

What the equity risk premium is, and why it matters

02

Literature Review

The historical average benchmark versus the promise of machine learning

03

The South African Market

Structural realities that set the JSE apart from developed markets

04

Data & Methodology

Predictors, model classes, and the out-of-sample evaluation framework

05

Results

Benchmark resilience, regularisation gains, and the failure of complex ML

06

Conclusion

Contributions, limitations and avenues for future research

What Is the Equity Risk Premium?



The extra return investors demand for holding risky equities instead of a safe asset

$$\text{ERP} = \text{return on equities} - \text{risk-free rate}$$

Why It Matters for Actuaries



Asset Allocation

Strategic and tactical allocation decisions for retirement funds



Discount Rates

The rate used to value long-term liabilities and growth assets



Return Assumptions

Long-term return assumptions built into fund valuations



PART ONE

THE CORE QUESTION

Can machine learning beat the historical average when forecasting the South African equity risk premium?



1. Benchmark

How strong is the historical average as a benchmark for forecasting the South African ERP?



2. Machine Learning

Can machine learning outperform the historical average?



3. Evaluation

Does the forecasting window matter?

- Expanding
 - Rolling
-



PART TWO

LITERATURE REVIEW

Setting the scene



Goyal & Welch (2008)

- Evaluated several major ERP predictors proposed in the literature
- Tested using out-of-sample forecasting
- Nearly 100 years of US data



THE RESULT

Most predictors failed to outperform the historical average.



The Pushback



Campbell & Thompson (2008)

THE CONSTRAINT

- Sensible restrictions
- Excess returns restricted to be non-negative



THE INSIGHT

- Small forecasting gains may still have economic value



Can Machine Learning Do Better?



Gu, Kelly & Xiu (2020)

WORKS WELL FOR

Individual stock returns



FALLS SHORT FOR

The aggregate premium



The ML Promise, and Its Limit



Xu & Liu (2024)

- Updates Goyal-Welch for the ML era
- LASSO, Ridge, Elastic Net and ensembles

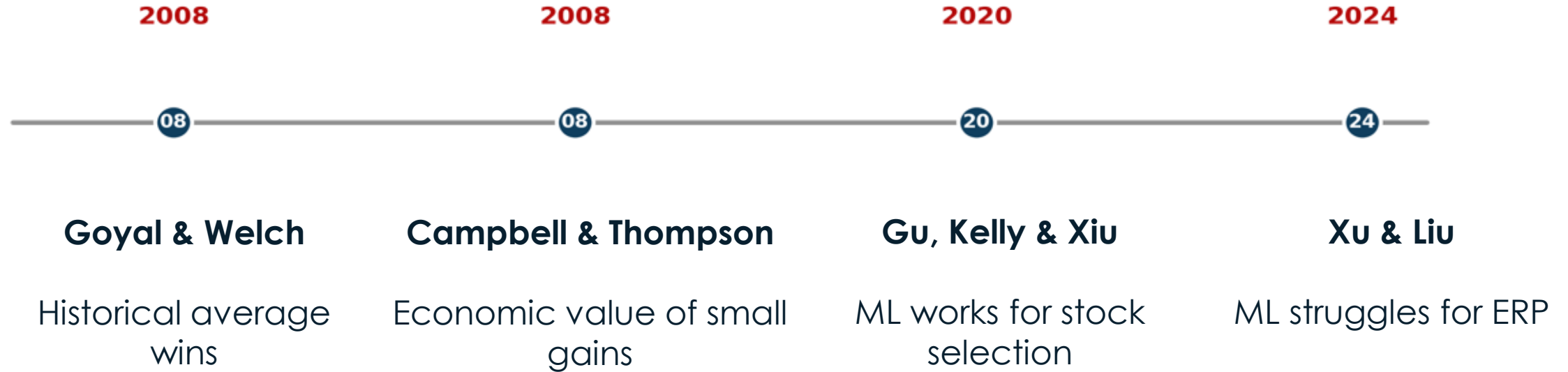


Modern machine learning largely fails to close the gap.

Historical average is extremely difficult to beat.



The Gap in the Literature



**Almost all evidence comes from developed markets.
What happens in an emerging market like South Africa?**



1 The Setting

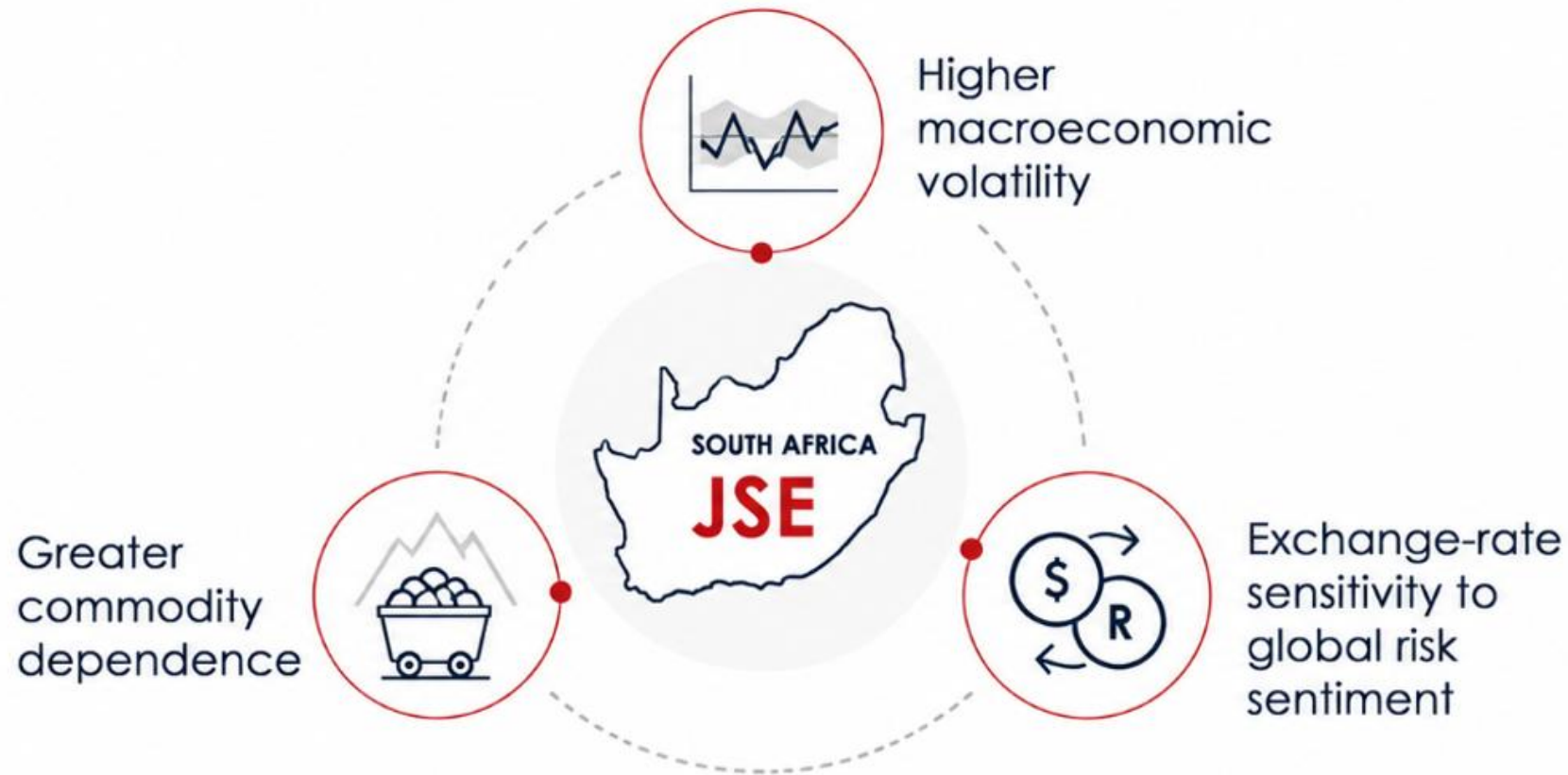
2 The Method

PART THREE

THE SOUTH AFRICAN MARKET

Two deliberate departures from the existing literature

Departure One: A Different Market



Could these differences create additional predictability?

Departure Two: A Different Evaluation Framework

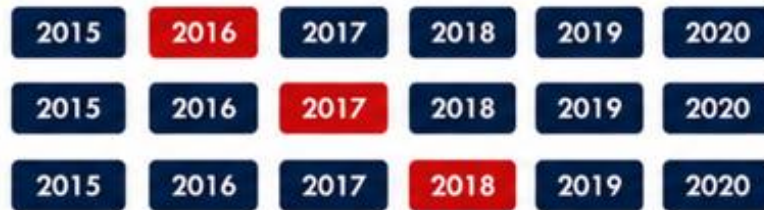


Why Standard Cross-Validation Is Unsuitable

Standard Machine Learning



Randomly split into folds



■ Training set ■ Testing set

✓ Works well for independent observations

Financial Time Series



Chronological order matters



⚠ Look-ahead Bias: The model sees the future when making the forecast

Markets move in one direction: forward

Departure Two: Out-of-Sample Evaluation



Expanding Window

(Growing)



The training window expands using all available past data

Rolling Window

(Fixed Size)



The training window rolls forward using a fixed number of past periods



Every forecast uses only information available at that point in time.



No future information



Real-world forecasting



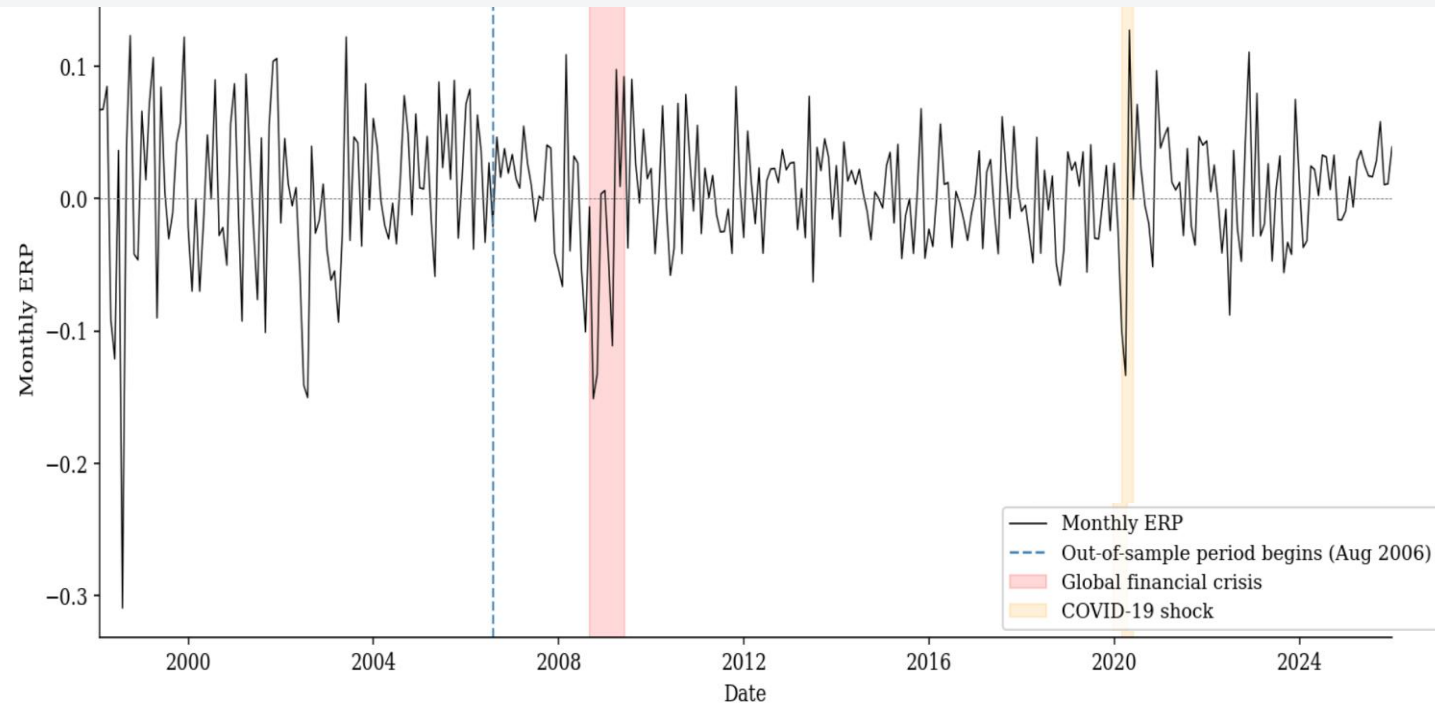
Fair comparison across models



PART FOUR

DATA & METHODOLOGY

Data & Predictor Variables



Jan 1998 – Dec 2025

Full sample period (332 monthly observations)

Aug 2006 – Nov 2025

Out-of-sample evaluation window (232 forecasts)

100 months

Rolling / initial training window length

Figure 1: South African Equity Risk Premium, January 1998 to December 2025

Valuation Ratios

DY, EY, book-to-market

Macroeconomic

T-bill rate, bond yield, term spread

Market-Based

Momentum, volatility, liquidity

Commodity & FX

Gold price (USD), exchange rate

Almost Two Decades of Out-of-Sample Testing



Monthly JSE data, from the Bloomberg and IRESS expert database. Data were imputed where necessary.

332

monthly observations

January 1998 – December
2025

232

out-of-sample forecasts

August 2006 – November
2025

6

model families tested

against two benchmark
specifications

2

evaluation windows

expanding and rolling
design

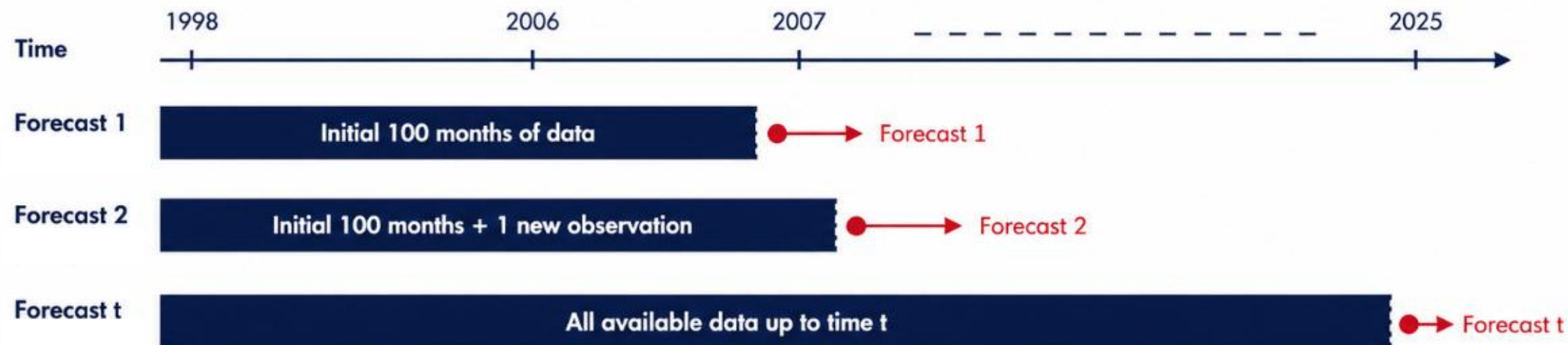


The out-of-sample window spans the global financial crisis, a full commodity cycle, and the COVID-19 period — turbulence that turns out to matter a great deal for these results.



Expanding Historical Average

Starts with 100 months and grows over time



Methodology: Balancing Flexibility with Constraint



Alternative Benchmarks	Historical Average (12-, 36-, 100-month), Exponential Smoothing ($\alpha = 0.1, 0.3, 0.5$)	Equal or geometrically declining historical weights
Regularised Linear	Ridge (L2), LASSO (L1), Elastic Net (L1 + L2)	Coefficient shrinkage and/or variable selection to penalise complexity
Non-Linear Ensemble	Random Forest, XGBoost	Aggregating decision trees, sequentially or independently, to capture non-linear trends
Deep Learning	Long Short-Term Memory (LSTM)	Gated neural memory structures capturing complex, long-term temporal dependencies

A Walk-Forward Loop, Repeated 232 Times



— — — — — Repeat one month at a time, 232 times, walking forward through the full out-of-sample period

Expanding Window

Training sample grows every month; the window only ever gets longer.

Rolling Window

Fixed-length window slides forward, dropping the oldest month as a new one is added.



Out-of-Sample R^2

$$R_{OOS}^2 = 1 - \frac{SSE \text{ of the model}}{SSE \text{ of the benchmark}}$$

A lower SSE
implies better
performance

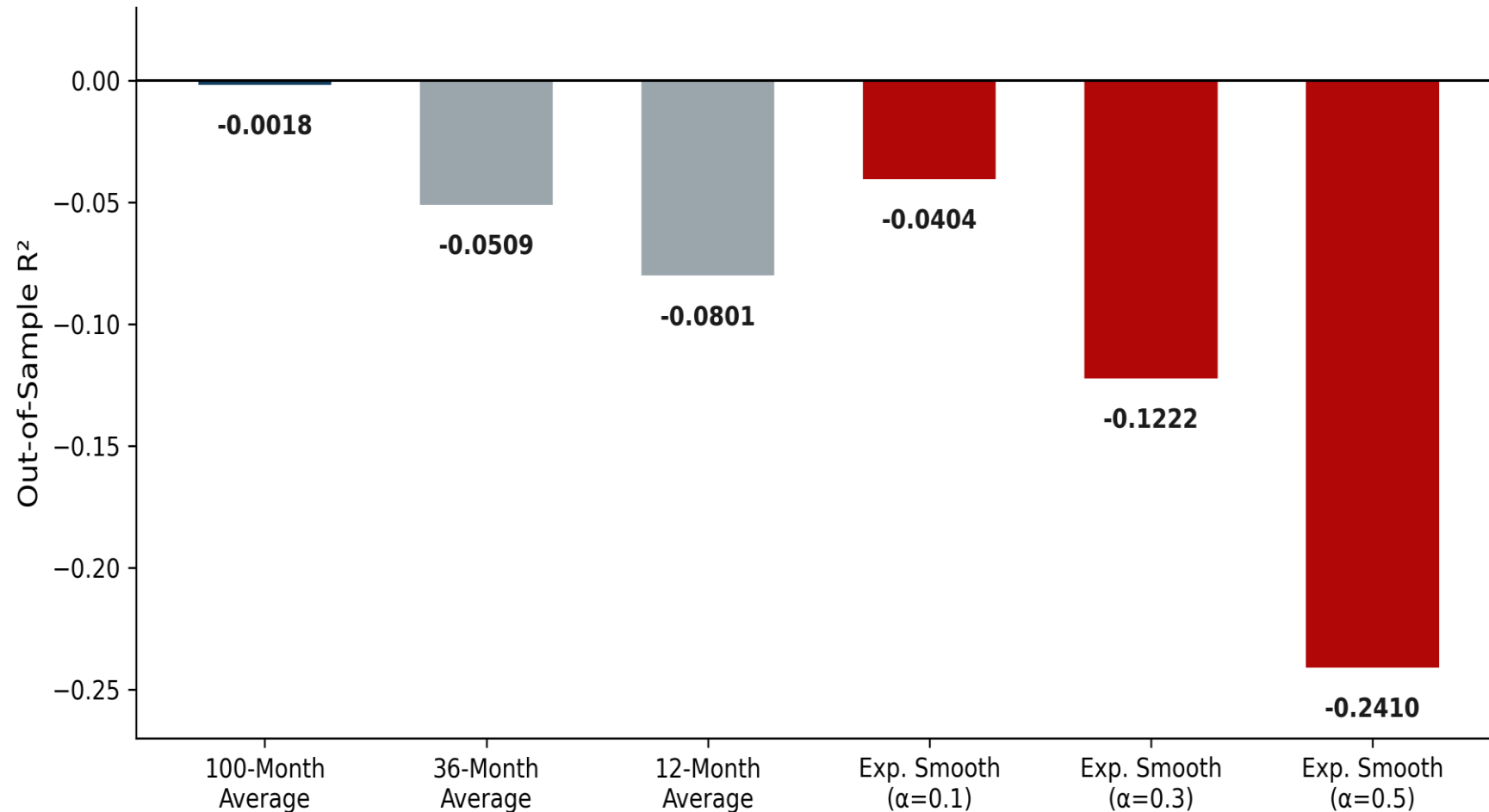
$R_{OOS}^2 > 0$	$R_{OOS}^2 = 0$	$R_{OOS}^2 < 0$
Outperforms benchmark	Matches benchmark	Underperforms benchmark
		



PART FIVE

RESULTS

Results I: The Resilience of the Benchmark



Every alternative failed

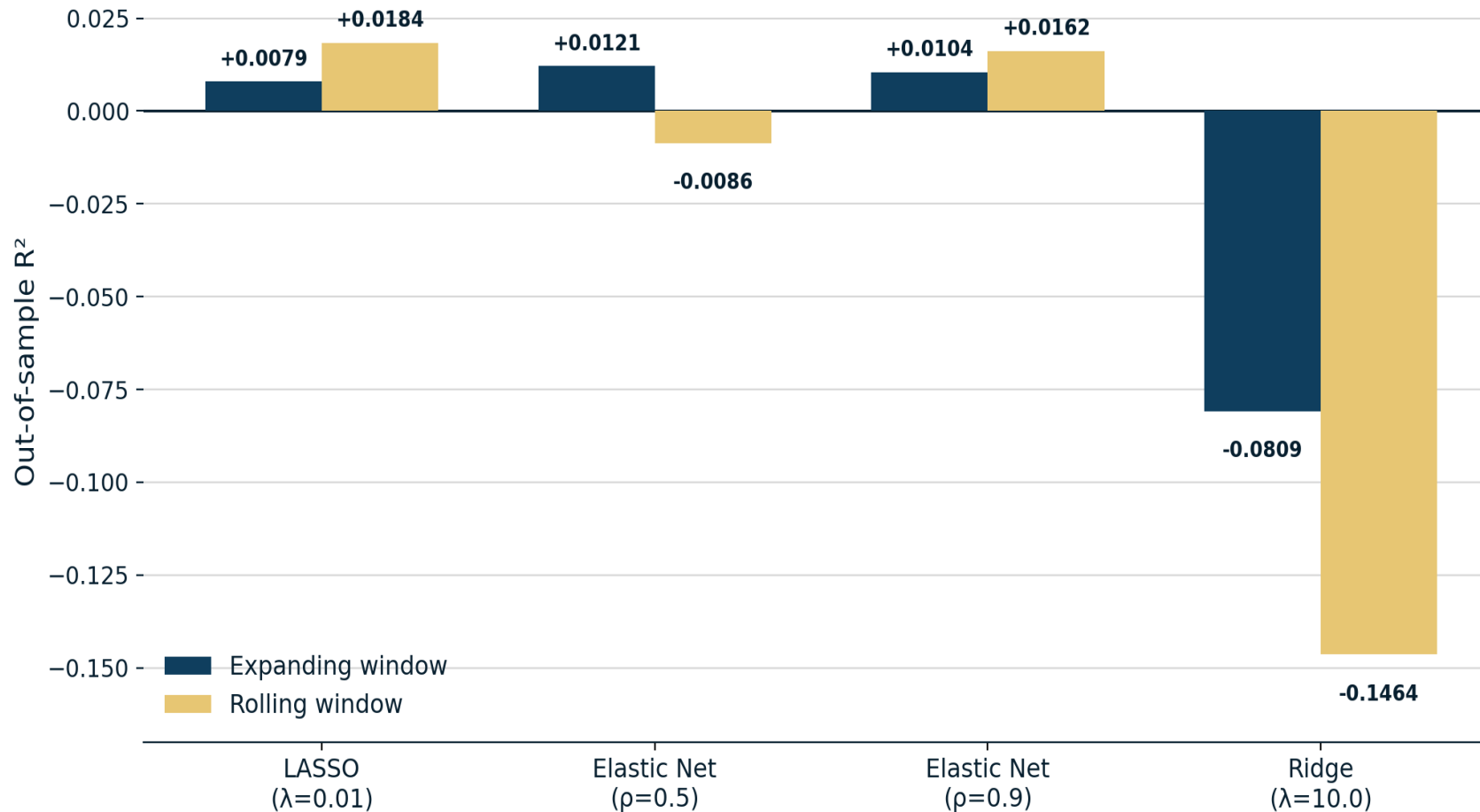
Shorter rolling averages and exponential smoothing all underperform the simple historical average.

$$R_{oos}^2 < 0$$

Underperforms benchmark

Figure 2: Out-of-Sample Forecast Performance (R^2): Benchmark models

Results II: The Triumph of Regularisation



LASSO

Best-performing model

Elastic Net

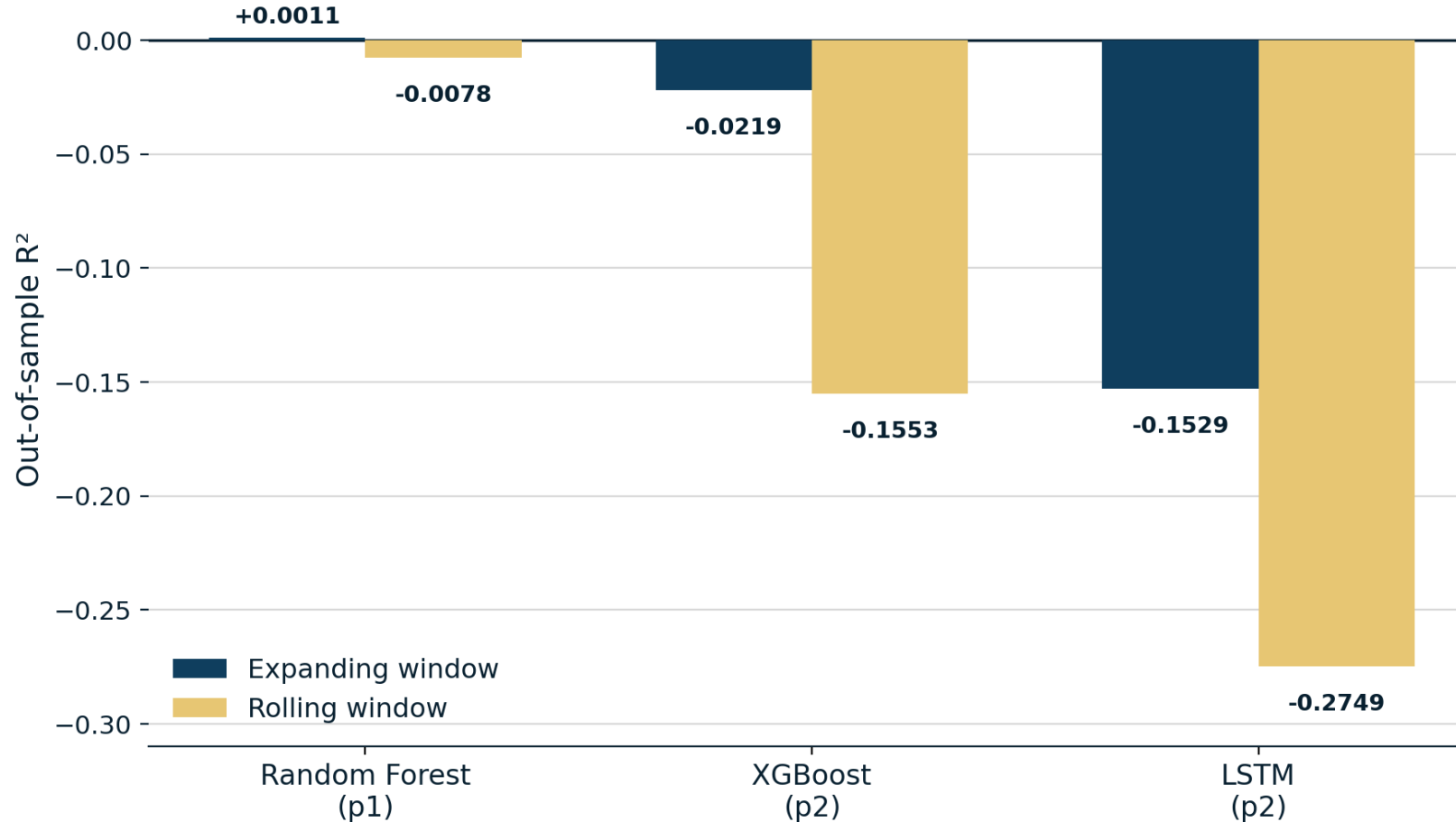
Consistent gains

Ridge

Completely failed to outperform the benchmark

Figure 3: Out-of-Sample Forecast Performance (R^2): Regularised Models

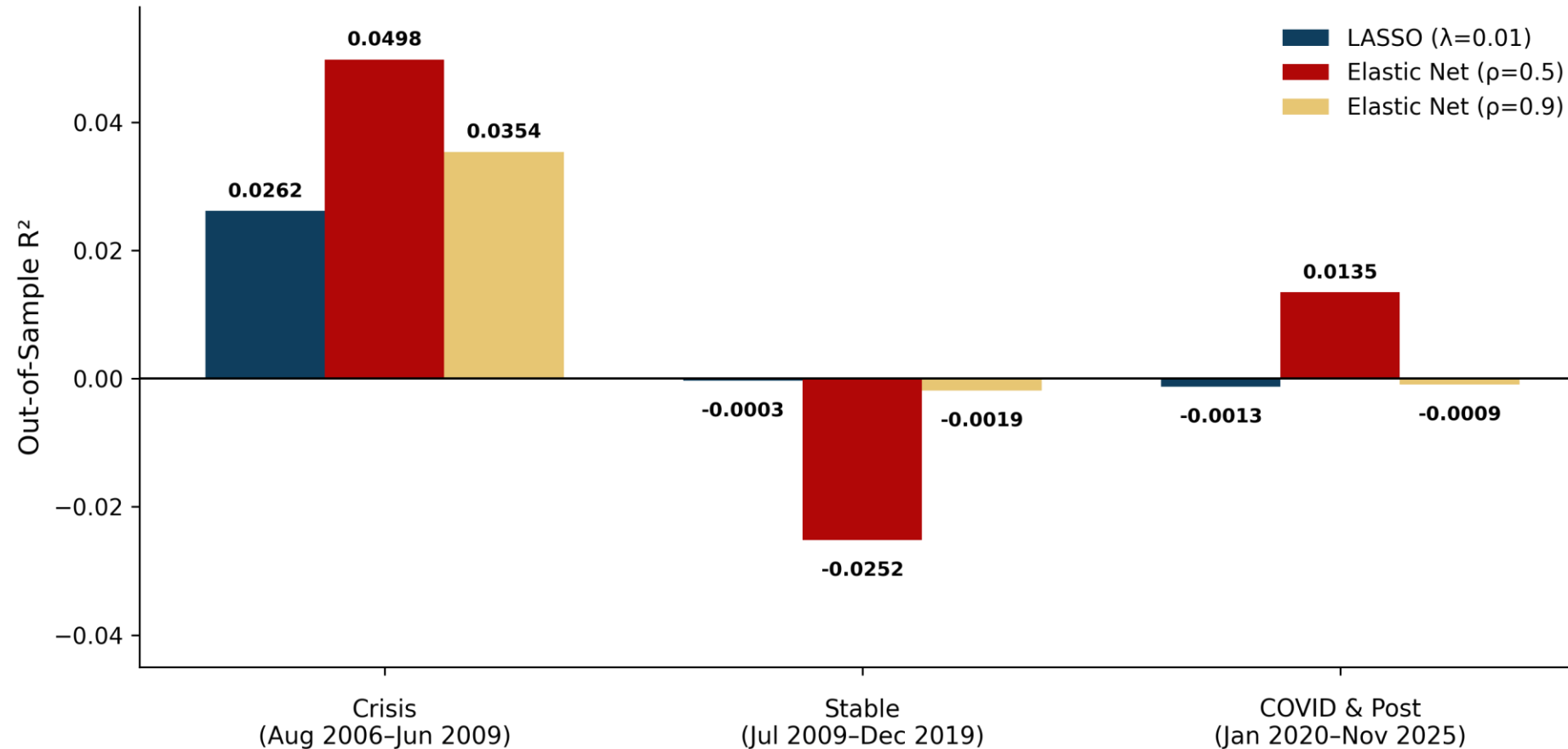
Results III: The Fall of the Machines



Random Forest, XGBoost and LSTM all failed to beat the benchmark under both window specifications. LSTM performed worst of all.

Figure 4: Out-of-Sample Forecast Performance (R^2): ML Models

Regime Dependency: The Crisis Edge



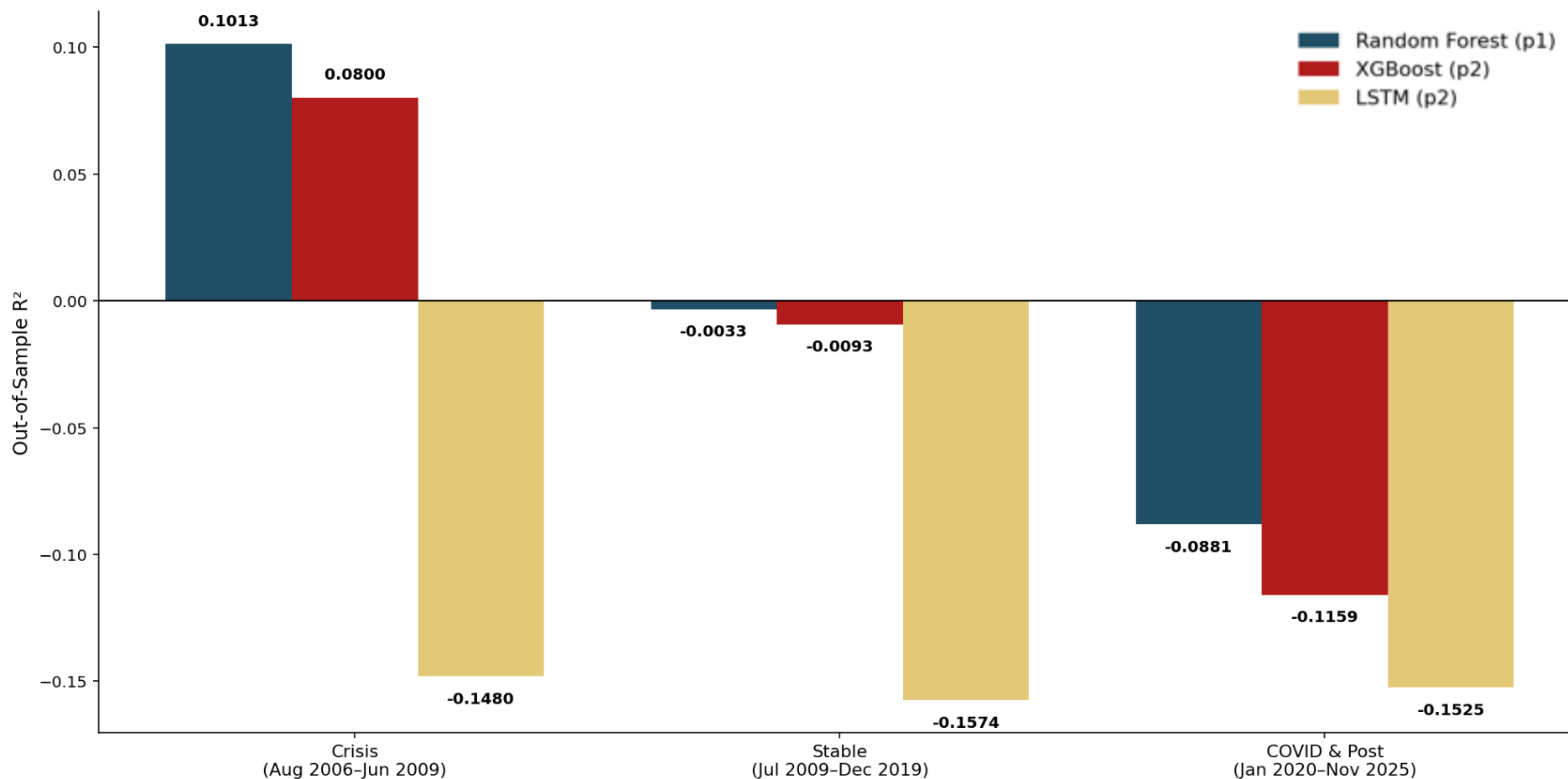
Predictability is concentrated in volatile periods

$$R_{OOS}^2 < 0$$

Underperforms benchmark

Figure 5: Regime Dependent Out-of-Sample Forecast Performance (R^2): Regularised Models

Regime Dependency: The Crisis Edge



Predictability is concentrated during August 2006 to June 2009

$$R_{OOS}^2 < 0$$

Underperforms benchmark

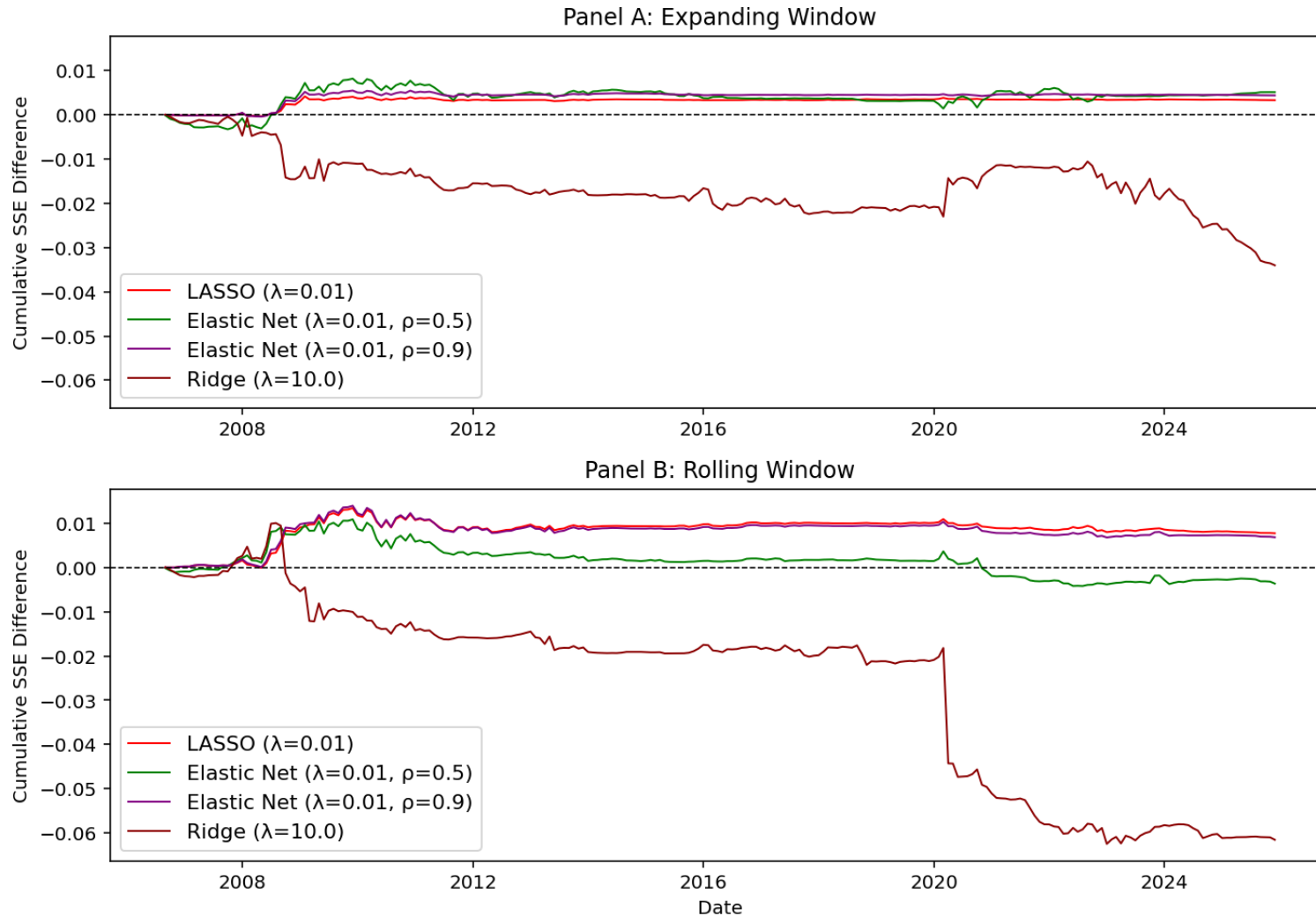
Figure 5: Regime Dependent Out-of-Sample Forecast Performance (R^2): ML Models



PART SIX

CONCLUSION

Understanding Model Performance



LASSO

Best-performing model

Elastic Net

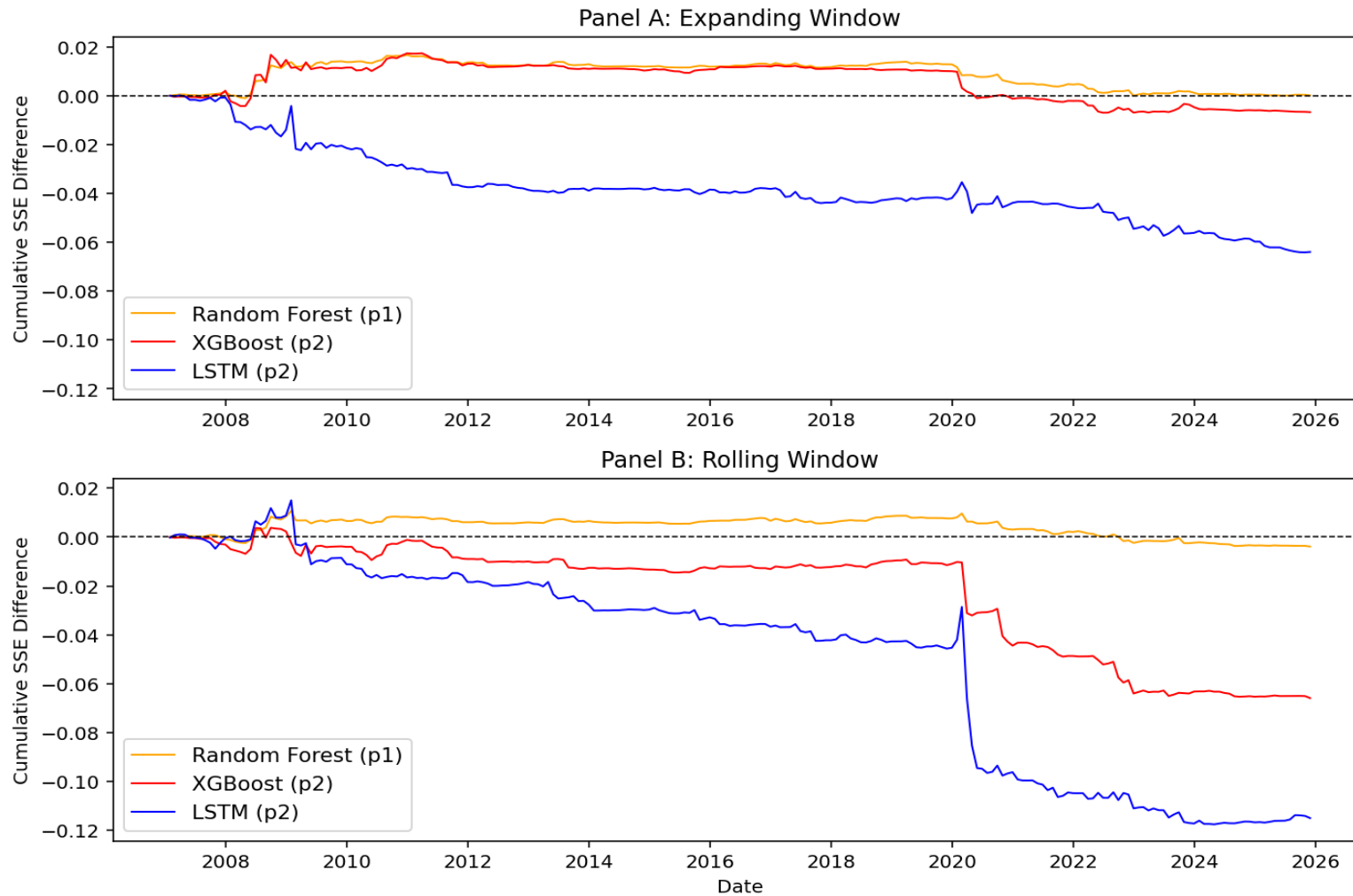
Consistent gains

Ridge

Completely failed to outperform the benchmark

Figure 6: Cumulative forecast improvement over the benchmark: Regularised Models

Understanding Model Performance



Random Forest, XGBoost and LSTM all failed to beat the benchmark under both window specifications. LSTM performed worst of all.

Figure 7: Cumulative forecast improvement over the benchmark: ML Models

Conclusion



1	The historical average held up.	No alternative benchmark was able to achieve a lower error metric.
2	Only appropriately regularised models added value.	LASSO (moderate penalisation) and Elastic Net (L1-dominated) delivered a small edge over the historical average
3	Sophistication alone did not help.	Ridge Regression, Random Forest, XGBoost and the LSTM all failed to beat the benchmark
4	Predictive relationships are time-varying.	Some models performed under rolling and some under expanding windows, testament to a bias variance trade-off
5	Emerging-market predictability exists, even if slightly	Extracting it requires careful control of model complexity and tuning



Limitations

- Modest sample size
- Set of predictors may not be appropriate
- Implicit independence assumptions
- Does not talk to the applicability to emerging markets as a whole
- Results are purely theoretical

Future Research

- Test whether these results hold across other emerging markets
- Expand (or contract) the predictor set
- Move from statistical significance to an explicit, costed measure of economic value, incorporating realistic transaction costs



Questions & Discussion

Thank you

Nabila Patel • 2677272@students.wits.ac.za

Furqaan Khan • 2694917@students.wits.ac.za

Eden Gross • eden.gross@wits.ac.za



ACTUARIAL SOCIETY
OF SOUTH AFRICA

Advancing Actuarial Excellence

Educate • Develop • Uphold