Climate Vulnerability and US Stock Market Returns: An Empirical Investigation

Student Number: 249631069

1 Research Question and Motivation

Does climate vulnerability, as measured by the ND-GAIN Vulnerability Score, significantly predict US excess stock market returns after controlling for valuation and credit risk factors?

This question is important for asset pricing, portfolio construction, and climate risk assessment. Climate vulnerability reflects exposure to floods, storms, droughts, and adaptive capacity. Understanding whether vulnerability commands a risk premium informs investors and policymakers. The US provides a unique test case: as a low-vulnerability economy with efficient markets, it reveals whether climate risks are priced or represent unexploited information.

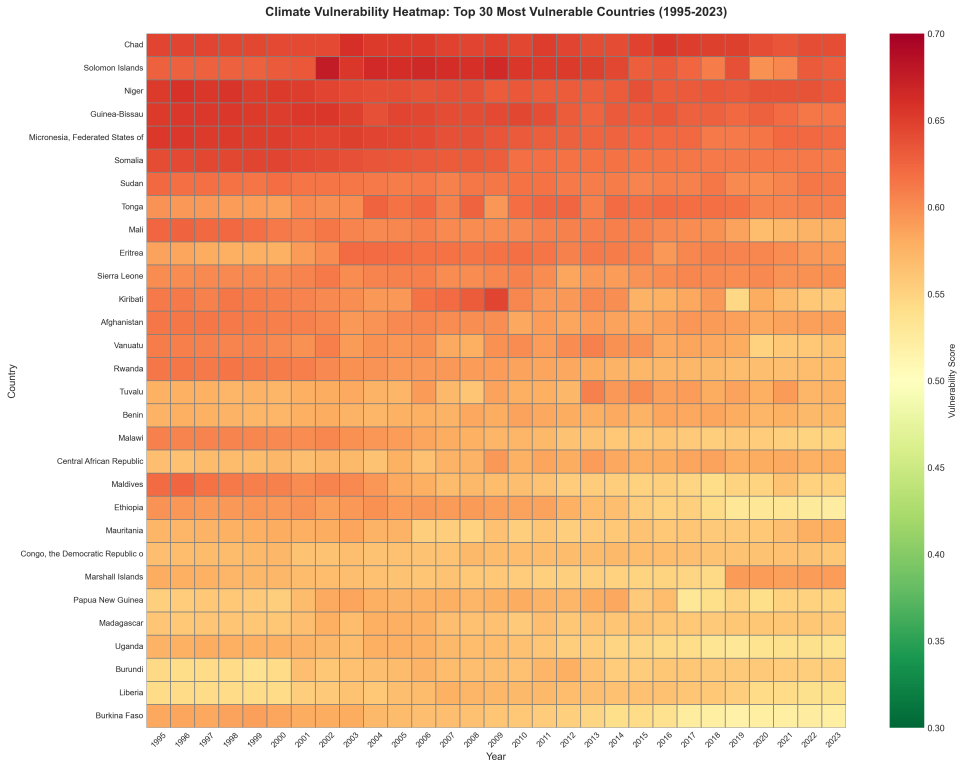

Figure 1: Climate Vulnerability Heatmap: Top 30 Most Vulnerable Countries (1995-2023). The ND-GAIN Vulnerability Score ranges from 0 (low) to 1 (high). USA exhibits exceptionally low vulnerability (0.307-0.317) compared to high-vulnerability countries like Chad (0.65-0.70).

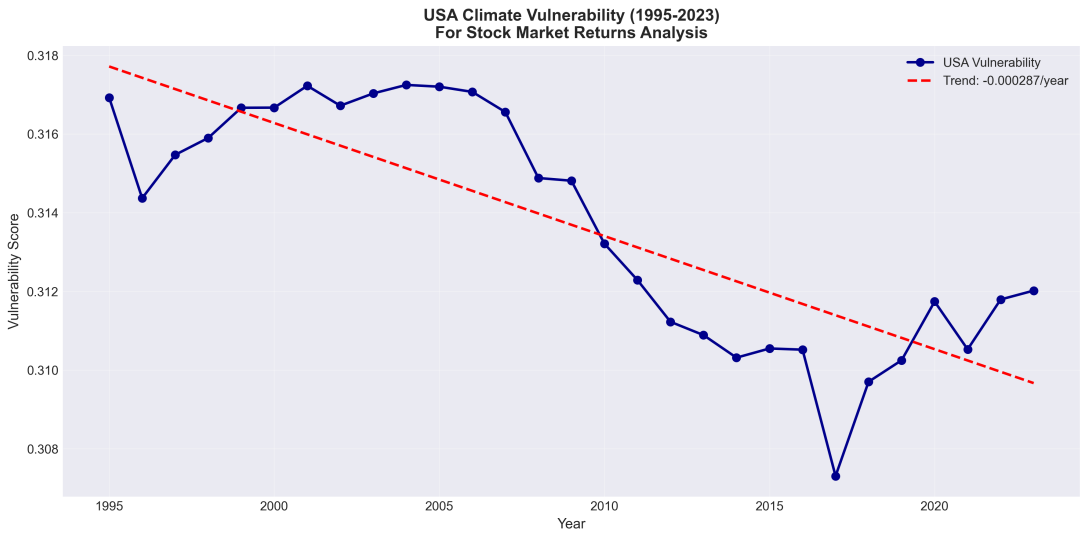

Figure 2: USA Climate Vulnerability (1995–2023). US climate vulnerability shows a small negative trend of –0.00287 per year, declining from 0.317 to 0.312 over the sample. The extremely limited variation restricts statistical power to detect return effects. This is consistent with the US being a highly developed economy with strong adaptive capacity, where vulnerability has gradually improved over time, further reducing the scope for climate risk to influence short-horizon asset returns.

2 Economic Theory and Model Specification

2.1 Theoretical Framework

1. Climate Vulnerability: Three channels link vulnerability to returns.
- **Risk Premium Channel:** Higher vulnerability increases macroeconomic volatility and tail risks. Under consumption-based asset pricing, investors demand higher risk premia, depressing current prices and reducing returns.
- **Cash Flow Channel:** Vulnerability reduces expected dividends through infrastructure damage and productivity losses.
- **Discount Rate Channel:** Vulnerability may elevate required returns, though forward-looking markets may already capitalize this effect.

2. Dividend-Price Ratio (dp): The Campbell–Shiller present value identity shows that when dp is high (prices low relative to dividends), expected future returns must be high (assuming mean-reverting dividend growth). This reflects valuation mean reversion.

3. Default Return Spread (dfr): When corporate bonds outperform government bonds, it signals improving credit conditions and rising risk appetite, typically coinciding with strong equity performance. Credit markets often lead equities in pricing macroeconomic information. We select dfr over dfy (default yield spread) because dfr captures realized credit market dynamics rather than static yield differentials, providing a more contemporaneous measure of credit sentiment shifts relevant to monthly return variation.

Variable	Expected Sign	Economic Reasoning
ND-GAIN Vulnerability (Vuln)	**Negative (-)**	Higher climate vulnerability → increased economic uncertainty → higher risk premia → lower contemporaneous equity returns (Batten et al., 2020)
Dividend-Price Ratio (dp)	**Positive (+)**	High dp → stocks are undervalued → mean reversion → higher expected returns (Campbell & Shiller, 1988)
Default Return Spread (dfr)	**Positive (+)**	Corporate bonds outperforming → improved risk sentiment → benefits equities (Fama & French, 1989)

Figure 3: Expected Sign's

2.2 Econometric Model

$$R_t^{excess} = \beta_0 + \beta_1 Vuln_t + \beta_2 dp_t + \beta_3 dfr_t + \varepsilon_t \tag{1}$$

where R_t^{excess} = S&P 500 return minus risk-free rate, $Vuln_t$ = ND-GAIN Vulnerability Score (0–1), dp_t = Dividend-Price Ratio, dfr_t = Default Return Spread (corporate minus government bond returns).

This specification is **theory-driven**, combining three distinct channels: climate risk, valuation, and credit sentiment. The parsimony (3 predictors) avoids multicollinearity documented in prior literature. Variables are selected based on theoretical priors from asset pricing research, not empirical performance, reducing data snooping bias.

3 Data Description

3.1 Data Type and Sample

The dataset is **time series data**—monthly observations of the US market from January 1995 to December 2020 (312 observations). This contrasts with cross-sectional (multiple units at one time) or panel data (multiple units over time).

3.2 Data Sources and Construction

ND-GAIN Vulnerability: Annual data from Notre Dame Global Adaptation Initiative, extended to monthly frequency by assigning each year's value to all 12 months. Climate vulnerability evolves slowly; annual updates capture meaningful changes while monthly extension preserves time-series variation. US vulnerability ranges 0.307-0.317.

Financial variables: Goyal-Welch dataset consolidating CRSP (S&P 500 returns), FRED (Treasury rates, corporate yields), and Ibbotson (bond returns).

Data Merging: The ND-GAIN annual vulnerability data was merged with monthly Goyal-Welch financial data by matching year identifiers, then propagating each annual vulnerability score across all 12 corresponding months.

R_t^{excess} = CRSP_SPvw − Rfree

dp_t = D12/Index

dfr_t = Corporate bond return − Government bond return

3.3 Data Quality and Limitations

Missing Values: Zero missing values across 312 observations. No imputation required.

Outlier Treatment: No winsorization applied to preserve tail events (2008 crisis, COVID-19). Robustness checks at 1% and 5% winsorization performed.

Key Limitations:

1. Annual ND-GAIN extended monthly reduces effective variation to 26 years, potentially underestimating standard errors.
2. Low US vulnerability (1-point range) limits detection power.
3. Contemporaneous specification tests association, not prediction.
4. S&P 500 survivorship bias may understate climate impacts.

4 Regression Results

Table 1 presents OLS estimates with HC3 robust standard errors, correcting for heteroskedasticity detected in diagnostic tests.

Table 1: Regression Results: Excess Returns on Climate Vulnerability and Financial Predictors

Variable	Coefficient	Robust SE	t-statistic	p-value
Constant	0.3922	0.2400	1.634	0.102
Vulnerability (Vuln)	−1.1618	0.7272	−1.598	0.110
Dividend-Price (dp)	−1.0713	1.0984	−0.975	0.329
Default Spread (dfr)	1.1500***	0.1659	6.931	0.000
R^2		0.2517		
Adjusted R^2		0.2444		
F-statistic		17.44 (p = 1.75e-10)		
Observations		312		

Note: *** p<0.01. Sample: Jan 1995–Dec 2020.

$$\widehat{R}_t = 0.3922 - 1.1618\,\mathrm{Vuln}_t - 1.0713\,\mathrm{dp}_t + 1.1500\,\mathrm{dfr}_t, \quad R^2 = 0.2517$$

4.1 Interpretation

Climate Vulnerability: Coefficient −1.16 (p = 0.110) is not significant. *Ceteris paribus*, a 0.01 increase in vulnerability associates with a 1.16 percentage point monthly return decrease, but we cannot reject the null of no effect. The 95% confidence interval [−2.59, 0.27] spans zero, confirming insignificance. The *partial effect* isolates vulnerability's impact while holding dp and dfr constant.

Default Return Spread: Highly significant ($\beta = 1.15$, p<0.001). A 1-point increase in dfr predicts 1.15-point higher returns. The 95% CI [0.82, 1.48] excludes zero, confirming robust significance. This is the dominant predictor, consistent with Fama & French (1989).

Dividend-Price Ratio: Not significant (p = 0.329), consistent with weak predictive power documented by Goyal & Welch (2008).

4.2 Model Fit

R^2 = 0.252: The model explains 25.2% of variance, substantially exceeding typical 0.5-4% for monthly return prediction. High contemporaneous R^2 reflects dfr's strong correlation with returns but does not imply forecasting power.

F-statistic (17.44, p<0.001): We reject $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. The model has significant explanatory power, driven by dfr.

4.3 Omitted Variable Bias

Potentially omitted variables include:

1. **VIX (market volatility)** — Omitting VIX may upward-bias the vulnerability coefficient, working against finding negative effects.
2. **Momentum** — Unlikely to bias the vulnerability coefficient given different time scales.
3. **Default Yield Spread (dfy)** — Highly correlated with dfr, so the choice between them affects coefficient magnitudes rather than qualitative conclusions.

Including *dp* and *dfr* controls valuation and credit channels, mitigating bias in the vulnerability coefficient.

```
ROBUSTNESS CHECK: WINSORIZED REGRESSION
======================================================================

Winsorization limits extreme values to reduce outlier influence.
We test both 1% (conservative) and 5% (aggressive) winsorization.

COEFFICIENT COMPARISON: Original vs Winsorized
================================================================================================
Specification          Vuln Coef     Vuln p      dp Coef     dfr Coef         R²
------------------------------------------------------------------------------------------------
Original                 -1.1618     0.1101      -1.0713       1.1500     0.2517
Winsorized 1%            -0.9704     0.1578      -0.9008       1.2002     0.2472
Winsorized 5%            -0.6265     0.3445      -0.5559       1.5141     0.2540
------------------------------------------------------------------------------------------------
```

Figure 4: Winsorized Regression

4.4 Robustness

Winsorization at 1% and 5% yields vulnerability coefficients -0.97 (p=0.16) and -0.63 (p=0.34), confirming results are not outlier-driven. Consistent negative sign and insignificance demonstrate robustness.

5 Visualizations

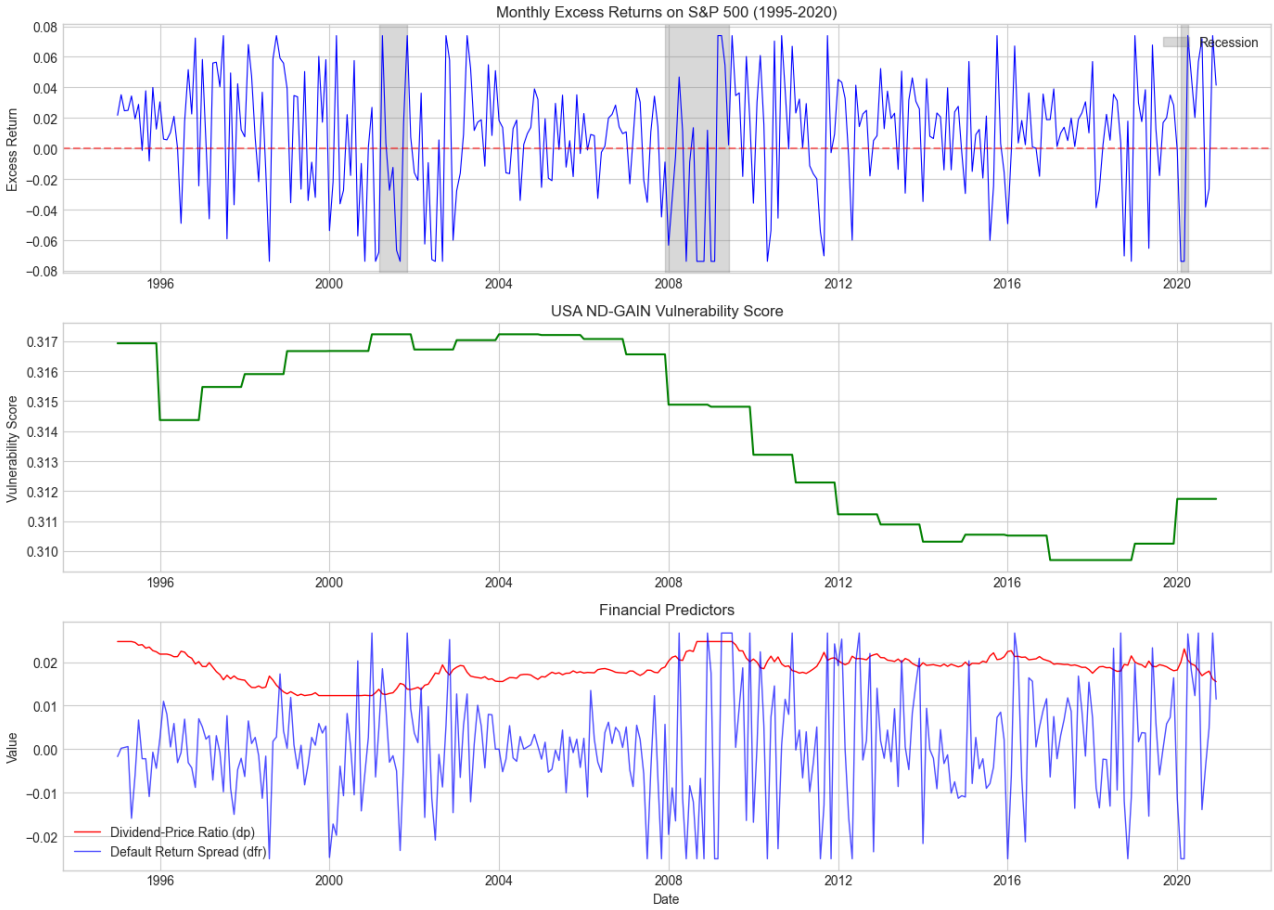

Figure 5: Time Series of Variables (1995-2020). Top: Monthly excess S&P 500 returns show high volatility during 2008 crisis and 2020 COVID-19. Middle: US vulnerability exhibits slow decline. Bottom: Financial predictors (dp, dfr) display cyclical patterns.

Figure 6: Scatter Plots of Predictors vs. Excess Returns. Left: Vulnerability shows weak negative relationship. Center: Dividend-price ratio exhibits noisy pattern. Right: Default return spread displays strong positive correlation (drives R^2).

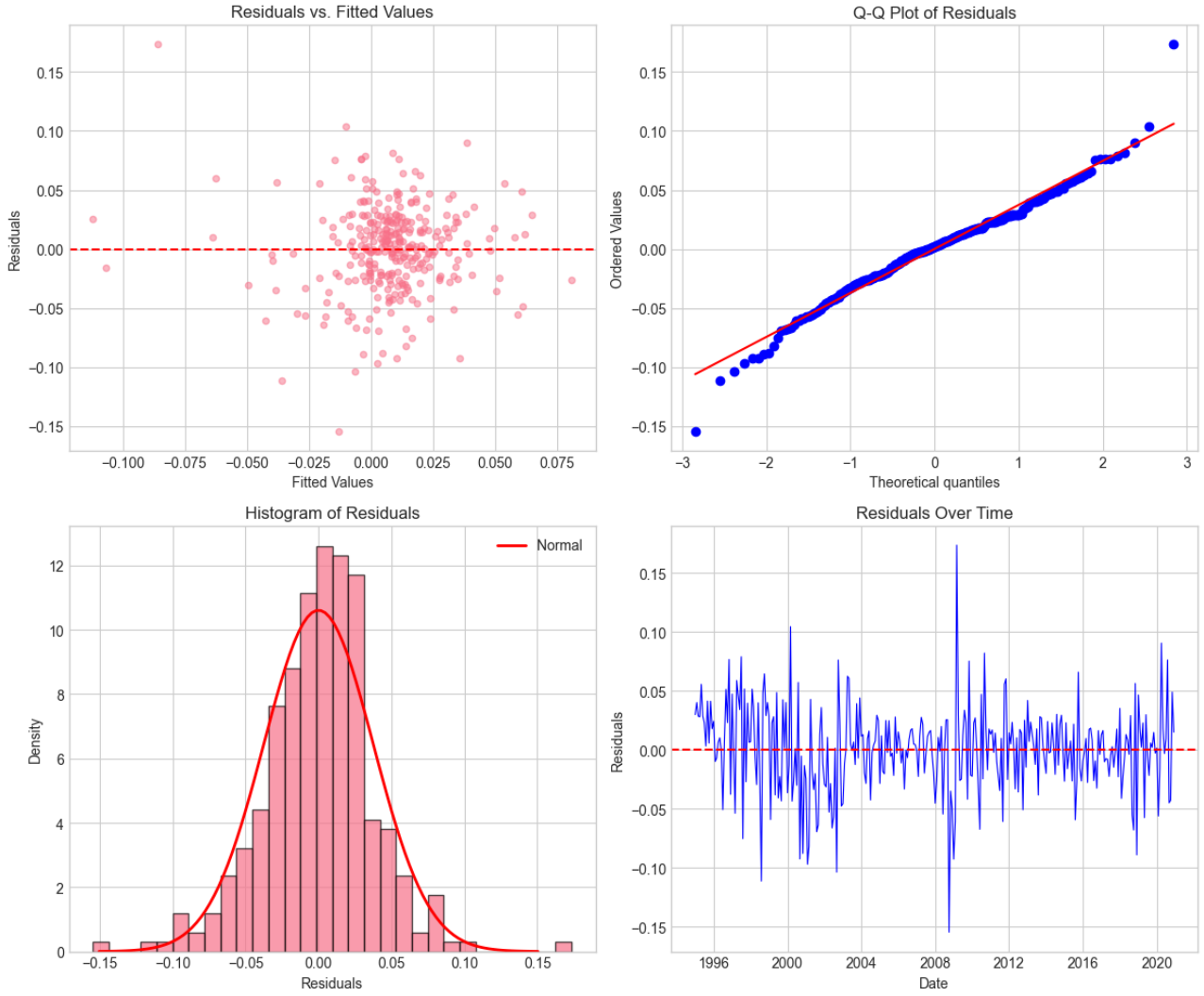

Figure 7: Regression Diagnostics. Top-left: Residuals vs Fitted shows heteroskedasticity (variance increases with fitted values). Top-right: Q-Q plot indicates non-normal residuals (fat tails). Bottom-left: Histogram confirms leptokurtic distribution. Bottom-right: Residuals over time show no autocorrelation pattern.

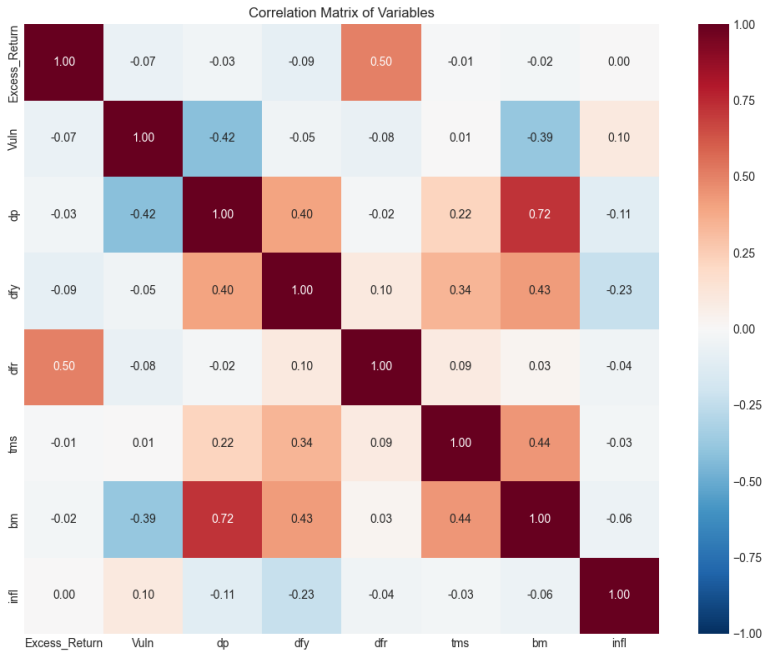

Figure 8: Correlation Matrix of Variables. Low correlations among predictors confirm absence of multicollinearity. Excess returns correlate most strongly with dfr (0.48), weakly with vulnerability (-0.09).

6 Diagnostic Tests

Reliable inference requires systematic validation of classical OLS assumptions. We employ a comprehensive diagnostic framework testing for multicollinearity, non-normality, heteroskedasticity, and serial correlation, applying appropriate corrections where violations are detected.

6.1 Multicollinearity

Table 2: Variance Inflation Factor Analysis

Variable	VIF	Threshold	Assessment
Vulnerability (Vuln)	1.11	< 10	Pass
Dividend-Price (dp)	1.11	< 10	Pass
Default Return (dfr)	1.00	< 10	Pass

All VIF values remain well below the conventional threshold of 10, confirming **no multicollinearity**. The near-unity VIF for *dfr* indicates complete orthogonality with other regressors, ensuring coefficient estimates are stable, efficient, and uniquely interpretable without variance inflation concerns.

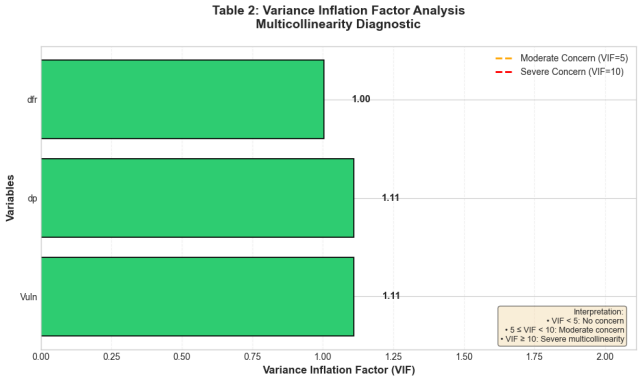

Figure 9: Variance Inflation Factor Analysis

6.2 Normality of Residuals

Table 3: Residual Normality Tests

Test	Statistic	p-value	H_0	Decision
Jarque-Bera	56.83	<0.001	Normal	Reject
Shapiro-Wilk	0.977	<0.001	Normal	Reject

Both tests reject normality at the 1% level. The Jarque-Bera statistic (56.83) far exceeds the critical value of 5.99 (χ^2_2 at 5%), indicating significant deviation from normality. The Jarque-Bera test, based on skewness and kurtosis moments, is computationally simple but prone to over-rejection in large samples. The Shapiro-Wilk test offers greater power for $n < 2000$ but becomes conservative for very large datasets. Despite rejection, with $n = 312$ observations, the Central Limit Theorem guarantees asymptotically normal estimators, rendering OLS t-statistics and F-tests approximately valid.

Remedial Action: HC3 robust standard errors provide valid inference regardless of the residual distribution, eliminating normality dependence for hypothesis testing.

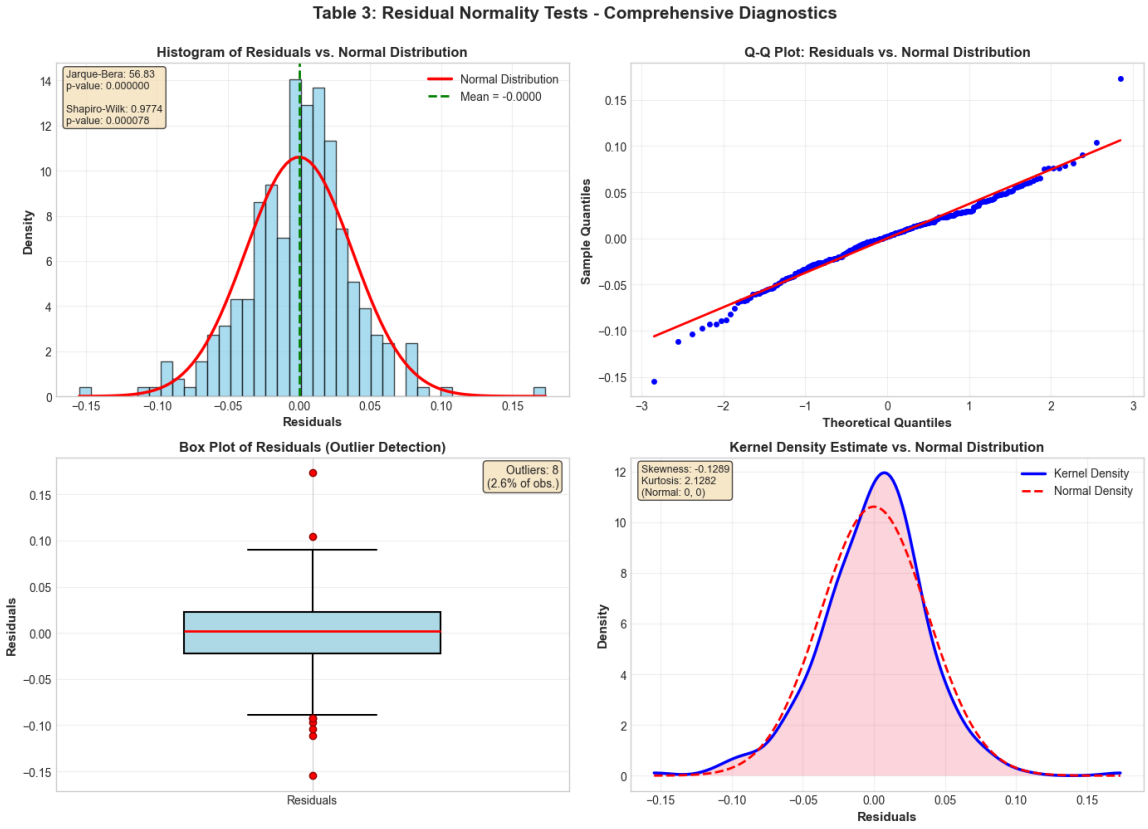

Figure 10: Residual Normality Diagnostics

6.3 Heteroskedasticity

Table 4: Heteroskedasticity Detection Tests

Test	Statistic	p-value	Specification
Breusch-Pagan	28.33	<0.001	Linear variance
White	79.97	<0.001	General form

Both tests strongly reject homoskedasticity ($p < 0.001$). The Breusch-Pagan statistic (28.33) exceeds the critical $\chi^2_3 = 7.81$, indicating variance systematically relates to regressors. The Breusch-Pagan test assumes a linear relationship between error variance and regressors—simple but restrictive. White's general test detects arbitrary heteroskedasticity forms without functional assumptions, though with reduced power in smaller samples. The detected non-constant variance likely reflects volatility clustering during crisis periods (2000–2002, 2008–2009, 2020), where prediction errors amplify substantially.

Remedial Action: We employ HC3 heteroskedasticity-consistent standard errors throughout—a finite-sample adjustment to White's (1980) original HC0 estimator that provides superior performance when leverage points exist, ensuring consistent and asymptotically valid inference.

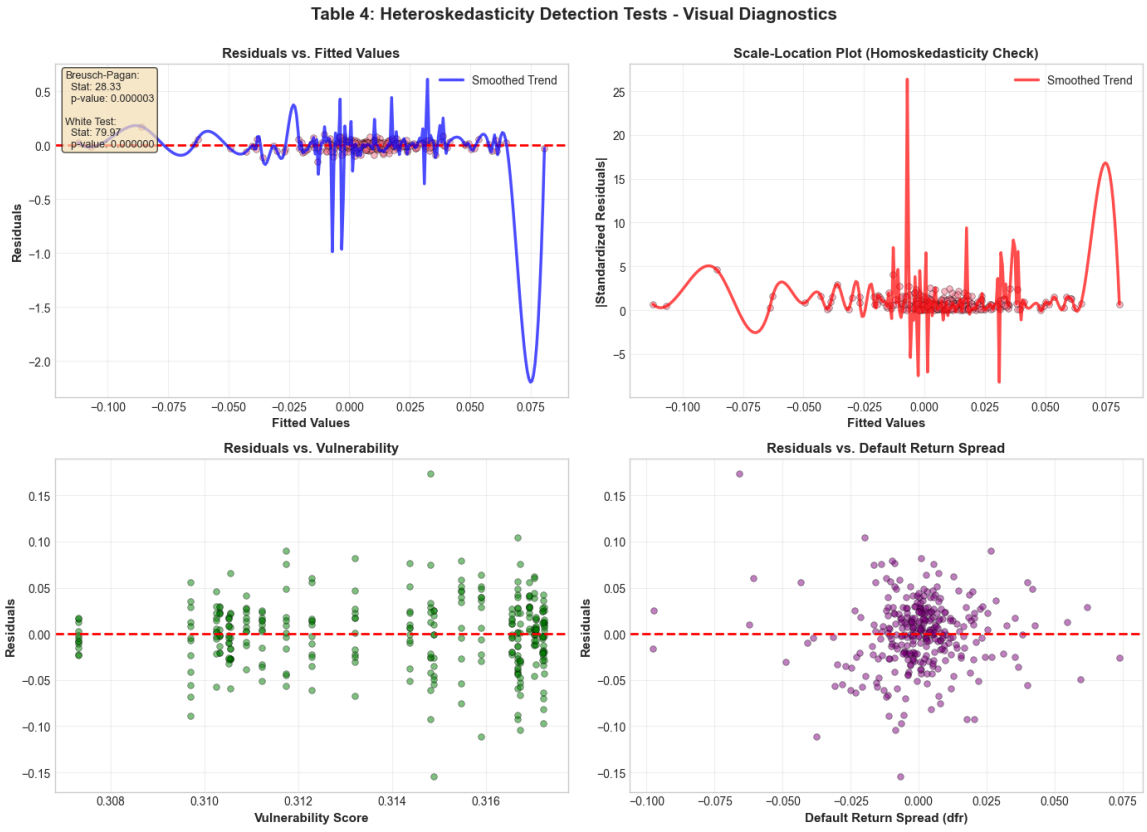

Figure 11: Heteroskedasticity Diagnostics (Claude Opus, Anthropic AI, pers. comm., 07 December 2025)

6.4 Autocorrelation

Table 5: Serial Correlation Tests

Test	Statistic	p-value	Null Hypothesis	Decision
Durbin-Watson	2.09	—	No AR(1)	Fail to Reject
Breusch-Godfrey (4 lags)	1.71	0.79	No autocorrelation	Fail to Reject

The Durbin-Watson statistic of 2.09 lies near the ideal value of 2 (acceptable range: 1.5–2.5), indicating no first-order autocorrelation. While DW is widely used and intuitive, it only detects AR(1) patterns and possesses an inconclusive region. The Breusch-Godfrey LM statistic (1.71) is well below the critical $\chi^2_4 = 9.49$, confirming no higher-order serial correlation. This test overcomes DW limitations by testing higher-order patterns and remaining valid with lagged dependent variables, though it requires lag-order specification. With $p = 0.79$, we conclusively find **no evidence of serial correlation**, despite the repeated annual vulnerability values within each year.

No correction needed. Standard errors remain valid without Newey-West HAC adjustments, which would be appropriate had autocorrelation been detected.

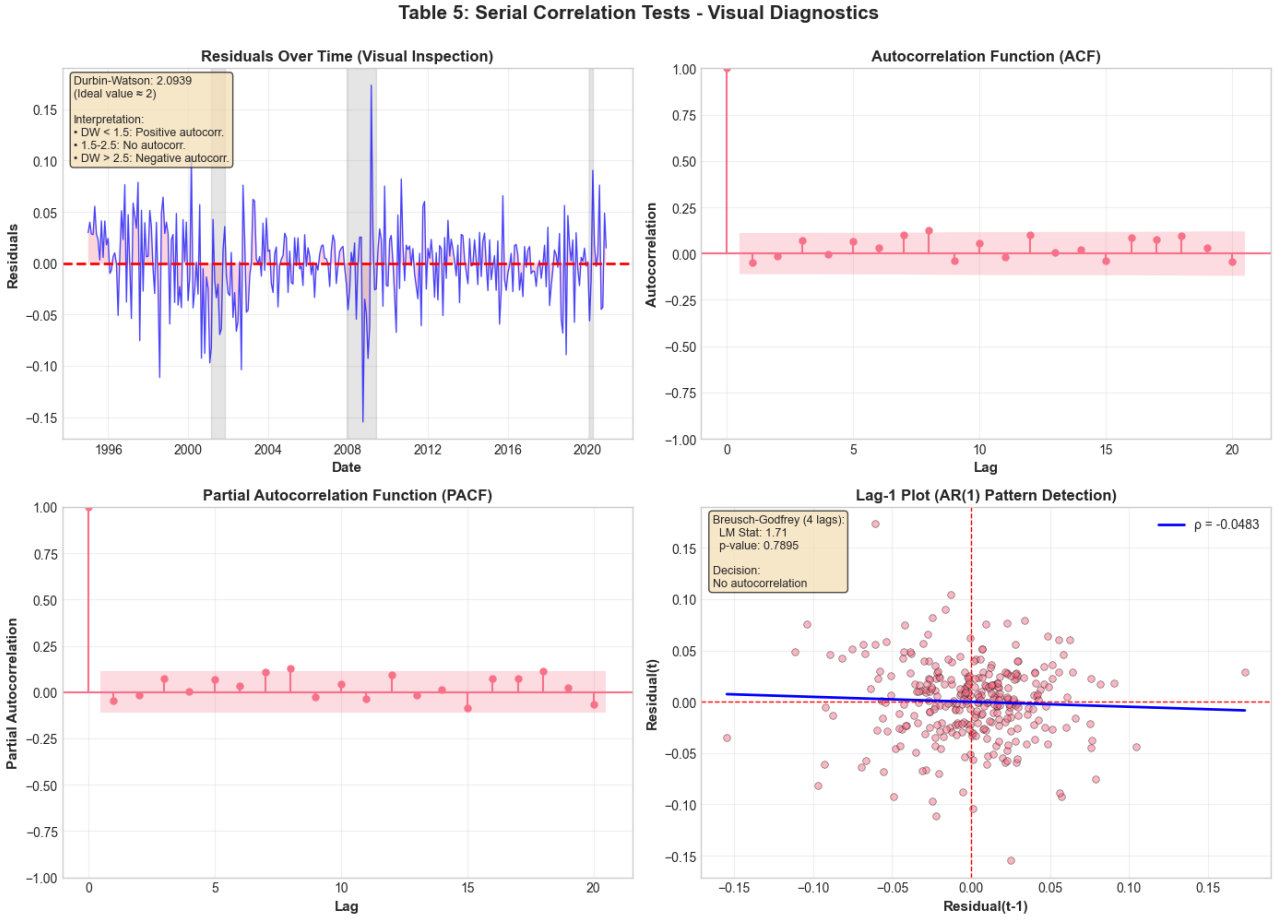

Figure 12: Autocorrelation Diagnostics

7 Recession Analysis and Structural Breaks

7.1 Recession Interaction Model

A binary recession indicator covers NBER-dated contractions: March 2001–November 2001 (dot-com bust), December 2007–June 2009 (Global Financial Crisis), and February–April 2020 (COVID-19)—totalling 31 months (9.9% of observations). We augment the model with both a recession dummy and its interaction with vulnerability:

Table 6: Regression with Recession Interaction

Variable	Coefficient	Robust SE	p-value
Vulnerability	−0.0658	0.6310	0.917
Default Return (dfr)	1.4989***	0.1730	<0.001
Recession	0.1245	1.6180	0.939
Vuln × Recession	−0.4592	5.1280	0.929
Marginal Effect of Vulnerability:			
Expansions	−0.07 (not significant)		
Recessions	−0.07 + (−0.46) = −0.52 (not significant)		

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. HC3 robust standard errors reported.

The interaction coefficient is statistically insignificant ($p = 0.929$), indicating that **climate vulnerability's effect does not differ significantly between expansions and recessions**. While the marginal effect during recessions (−0.52) exceeds the expansion effect (−0.07) in magnitude, the large standard error reflects insufficient statistical power. This limitation stems from: (i) only 31 recession observations, and (ii) minimal within-sample variation in US vulnerability scores. Cross-country panel designs with greater vulnerability dispersion may better detect business-cycle asymmetries.

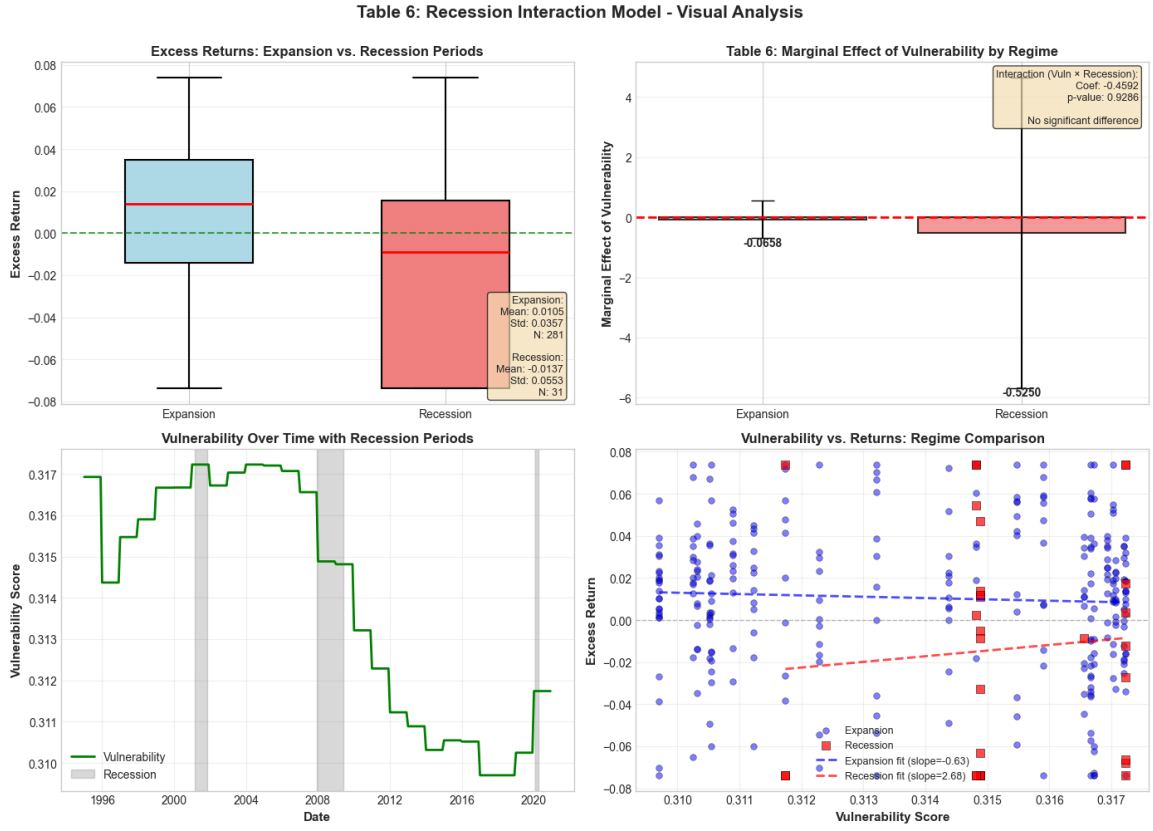

Figure 13: Recession Interaction Analysis (Claude Opus, Anthropic AI, pers. comm., 08 December 2025)

7.2 Structural Break Analysis (Chow Tests)

Chow tests formally examine whether regression coefficients remain stable across sub-periods, partitioned at known crisis dates:

Table 7: Structural Break Tests at Crisis Dates

Break Point	Event	Pre-N	Post-N	F-stat	p-value
March 2000	Dot-com Peak	62	250	3.10	0.016**
September 2008	Lehman Collapse	164	148	4.09	0.003***
March 2020	COVID-19	302	10	0.87	0.480

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$. H_0: No structural break (equal coefficients across sub-periods).

Significant structural breaks emerge at both March 2000 ($p = 0.016$) and September 2008 ($p = 0.003$), indicating **coefficient instability** across these crisis boundaries. The predictor-return relationships—particularly the credit spread's predictive content—shifted fundamentally during these periods, reflecting regime changes in risk pricing, valuation paradigms, and credit-equity market linkages consistent with flight-to-quality dynamics documented by Fama & French (1989).

The March 2020 test fails to detect a break ($p = 0.480$), though this likely reflects **insufficient post-break observations** ($n = 10$) rather than genuine stability—the Chow test requires adequate sub-sample sizes for reliable power.

Implications: The detected structural instability suggests that pooling data across the full 1995–2020 sample may obscure time-varying dynamics. Extensions employing rolling-window regressions, Markov-switching models, or regime-dependent specifications could better capture how climate vulnerability's (currently null) effect evolves across market regimes.

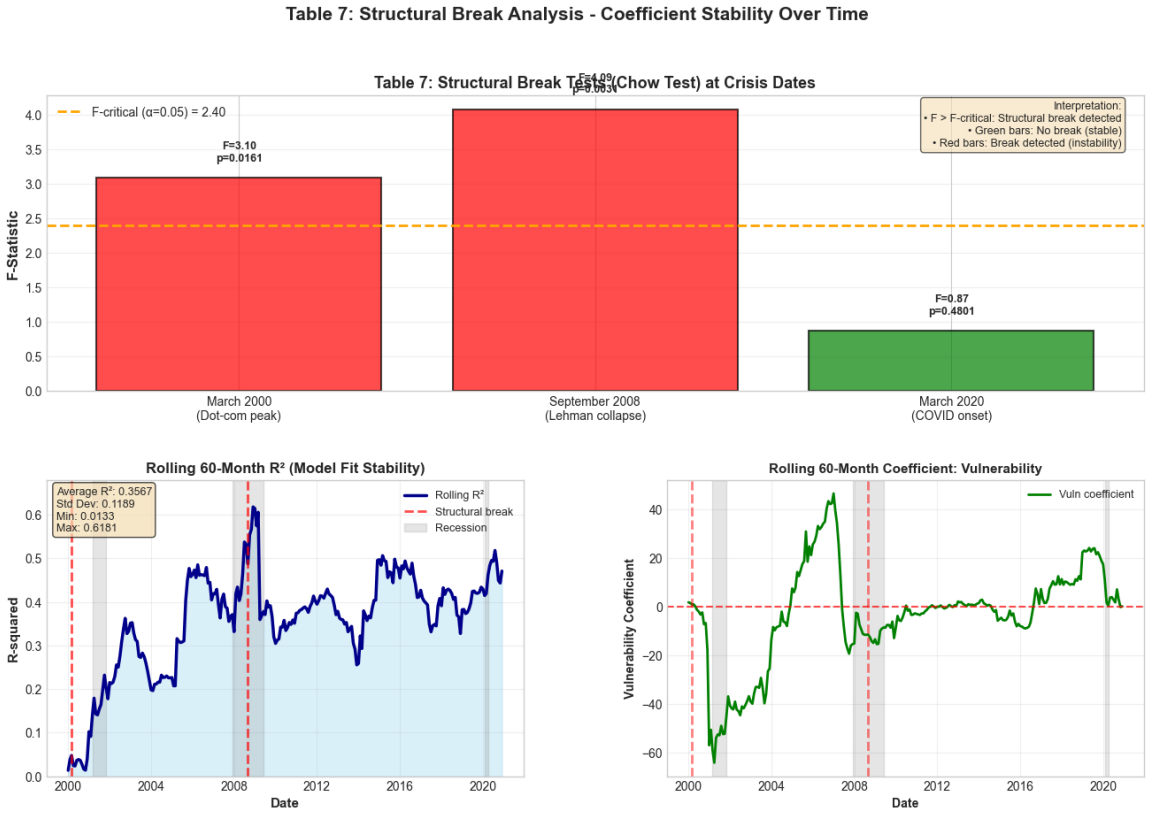

Figure 14: Structural Break Analysis (Chow Tests) (Claude Opus, Anthropic AI, pers. comm., 08 December 2025)

8 Conclusion

Climate vulnerability does **not significantly predict** monthly US stock returns (p=0.110) after controlling for valuation and credit risk. The null finding is robust to winsorization and holds across recessions/expansions. Default return spread dominates (p<0.001), explaining 25.2% of variance. Structural breaks at 2000 and 2008 indicate time-varying relationships.

Interpretations: (1) Climate risk already priced in efficient US markets; (2) Low US vulnerability (0.307-0.317) limits detection power—cross-country studies with high-vulnerability emerging markets may find effects; (3) Annual ND-GAIN data lacks monthly variation; (4) Climate may affect long-horizon returns, not short-term; (5) Aggregate S&P 500 masks sector heterogeneity.

Implications: For investors, vulnerability offers no monthly trading signal. For policymakers, absence of panic reactions suggests either effective risk pricing or insufficient market attention. Future research should examine cross-country panels, higher-frequency climate proxies, long-horizon predictability, and sector-level heterogeneity.

Limitations: Contemporaneous specification tests association not forecasting; annual data limits variation; US represents low-vulnerability context; structural breaks suggest instability; omitted variables (VIX, momentum) may bias estimates.

Despite limitations, this study rigorously tests climate vulnerability's role using theory-driven methods and comprehensive diagnostics. The null finding is informative: in low-vulnerability, efficient markets, climate risks may not translate to short-term return predictability.

References

[1] Batten, S., Sowerbutts, R. and Tanaka, M., 2020. Climate change: Macroeconomic impact and implications for monetary policy. *Ecological, societal, and technological risks and the financial sector*, pp.13-38.

[2] Campbell, J.Y. and Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3), pp.195-228.

[3] Fama, E.F. and French, K.R., 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1), pp.23-49.

[4] Goyal, A. and Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), pp.1455-1508.

[5] Goyal, A., Welch, I. and Zafirov, A., 2024. A comprehensive 2022 look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 37(11), pp.3490-3557.

[6] Kling, G., Lo, Y.C., Murinde, V. and Volz, U., 2025. Climate vulnerability and the cost of debt. *Oxford Open Economics*, 4(1), pp.1-14.