Insurance Solvency Analysis of EU countries with Script R and Python

Introduction

In this project, two scripts were developed, in R and Python respectively, to analyse the relationship between the solvency conditions of EU countries and various economic and financial variables.

The aim is to explore potential correlations between these indicators and to visualise the behaviour of these solvency measures in different European countries.

In this work, the integrated use of R and Python allowed a double validation of the results and the optimisation of each step of the analysis, combining the statistical power of R with the flexibility and automation of Python. This dual approach ensured more reliable results.

Variables considered

The variables considered were extracted in quarterly format from the :

" https://data.ecb.europa.eu/" data from insurance companies, considering aggregated life and non-life data.

" https://data.worldbank.org/" macroeconomic data

The variables considered are as follows.

Variable Assets:

A10 (Currency and deposits, representing liquid assets)

A20 (Loans, credit granted to others)

A30 (Shares and units in investment funds, showing holdings in shares and mutual funds)

A40 (Insurance technical reserves, which include provisions for future insurance claims).

A50 (Other receivables, representing outstanding payments due)

A60 (Investment fund shares/units held specifically by insurance companies)

A70 (Debt securities, representing bonds and similar financial instruments)

A80 (Other activities, a miscellaneous category)

Variable Liabilities:

L20 (Loans taken out by the insurance undertaking, with indication of foreign debt)

L30 (debt securities issued, if the insurance undertaking has issued bonds)

L40 (insurance technical reserves (liabilities), covering provisions for future insurance claims)

L50 (Defined benefit pension fund reserves, defined benefit plan obligations)

L51 (Defined Contribution Pension Fund Reserves, for defined contribution pension liabilities)

L70 (Other liabilities, a category for various other obligations)

Economic Variables:

GDP (GDP)

HICP (Harmonised Index of Consumer Prices)

Total indicator

T00 (Total assets and liabilities, providing an overview of the entire balance sheet of the company)

Script "Project.R/Project.ipynb".

The first phase of the analysis involved the preparation and cleaning of the dataset using the `Project` script. Initially, data were imported from CSV files, which were organised and renamed to ensure consistency and ease of use in the subsequent analysis. A renaming function was applied to standardise column names and ensure that each dataset maintained a consistent structure, facilitating the analysis and modelling process.

In addition, the script focuses mainly on comparing two solvency measures, **T00** and **L40**, providing a detailed assessment of the financial stability of the countries considered, with a specific focus on Spain and Italy.

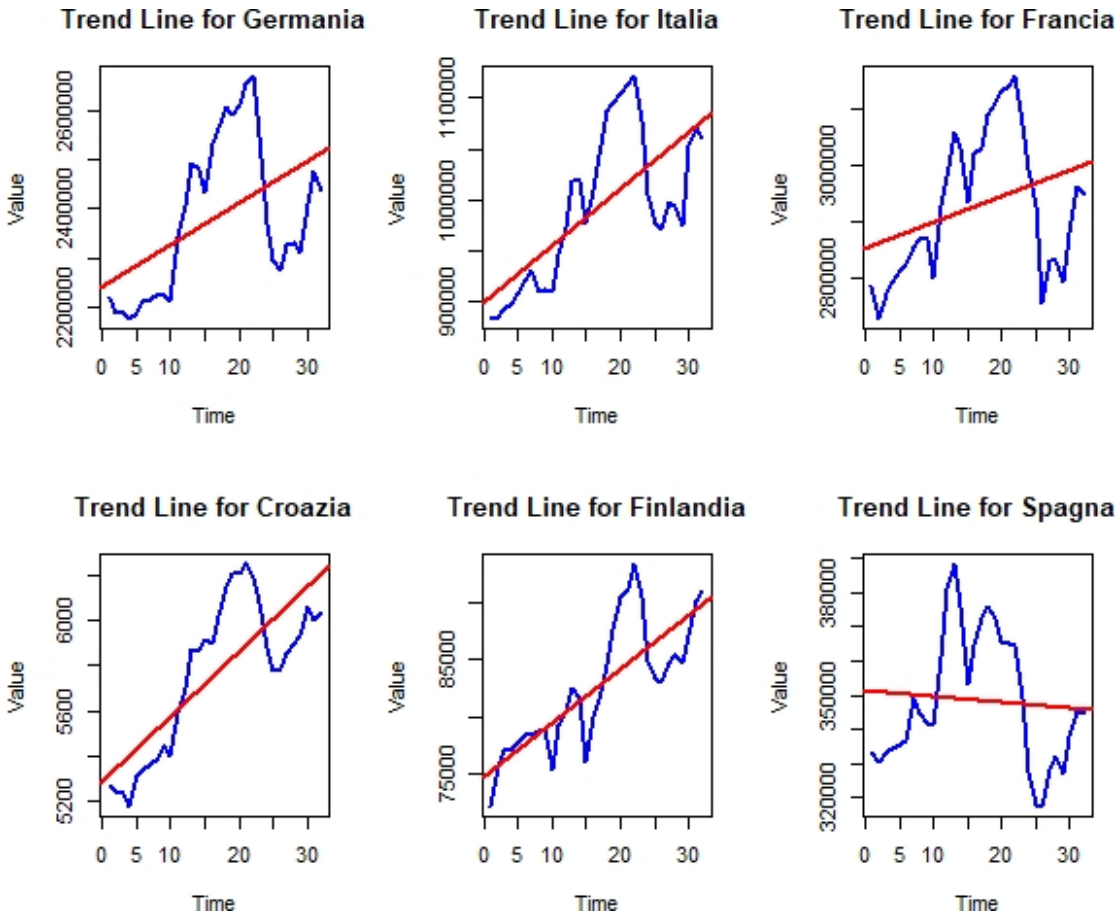
The script also includes statistical tests to verify the distribution of the data, such as **the Shapiro test** and some **customised Gamma** tests, which are useful for evaluating the type of model to be implemented.

Most countries did not present normal distributions, and this was visible both graphically (through the use of histograms) and from the test results.

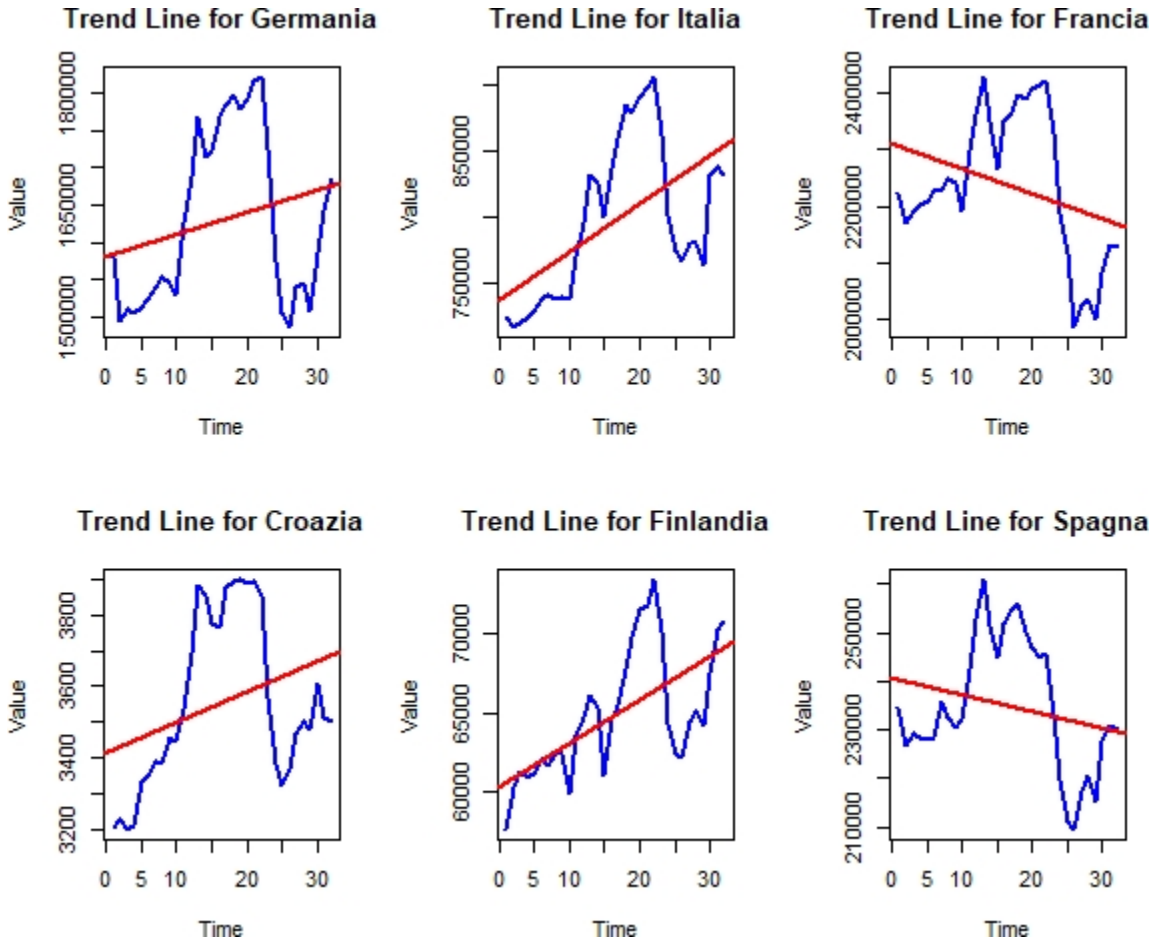
Customised tests were also carried out to check for the presence of other distributions, in the hope of capturing as much information as possible. However, even where the various tests yielded positive results regarding the presence of gamma distributions, there was no evidence in the graphical representations, nor was there any real advantage in using models better suited to deal with this type of distribution of the dependent variable (such as GLM models).

This preparation and validation phase thus resulted in a standardised dataset ready for advanced analysis, with a focus on comparison and statistical validity of the T00 and L40 solvency measures.

Graphs of T00 with trend line



Graphs of L40 with trend line



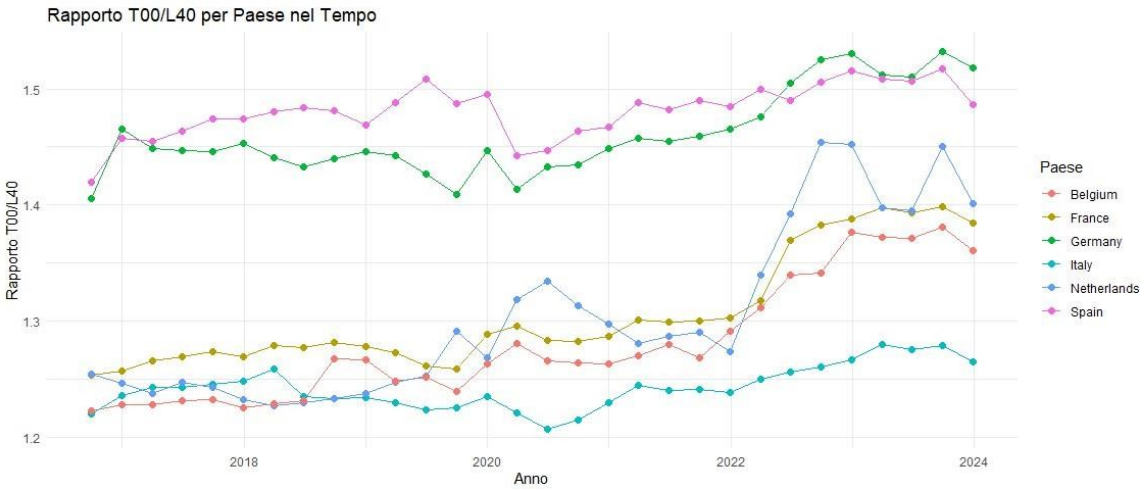
Concluding remarks part one

In the first instance, there was the intention to study only the variable T00 as a dependent variable, which was, however, too closely linked to macroeconomic changes in GDP and inflation, and thus not a suitable measure to describe insurance solvency.

It was therefore concluded to relate the "**Total-assets-liabilities**" (**T00**) with the "**Insurance technical reserves reported by insurance corporations**" (**L40**), resulting in a solvency measure weighted on the amount of capital set aside to cover potential future claims, thus being much more suitable to indicate insurance solvency.

Its values are close to 1, where larger values indicate greater solvency while values close to or even smaller, indicate a greater risk of default.

Below is a graph of the index values over time for some relevant EU countries.



Script "Templates.R/Models.ipynb".

Models' processes economic and solvency data for several European countries, including Italy, France, Germany, Spain and others. For each country, the script performs the following operations:

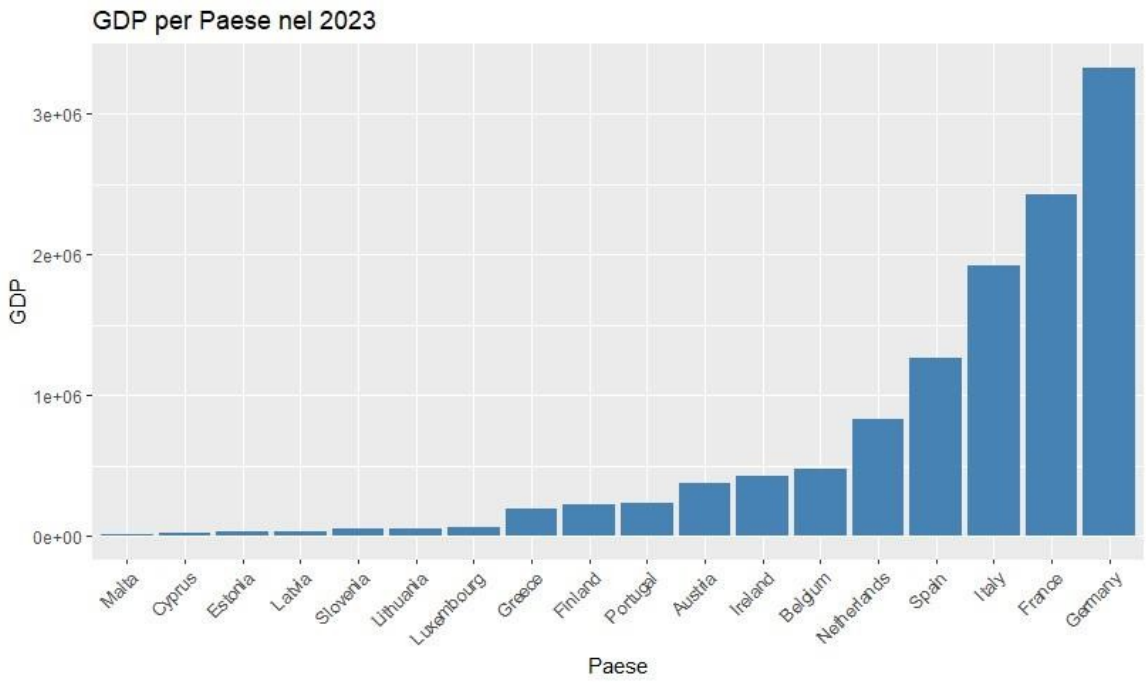
- Select the first 32 observations (**time interval September 2016-June 2024**) for all variables considered.
- It constructs temporary data frames for each country and produces visualisations via histograms to assess the distribution of these variables.
- It conducts regression analyses to explore the relationship between solvency variables and financial and economic indicators such as GDP and HICP.

An initial approach was to consider a general model that took into account data from all countries simultaneously. This approach, however, proved inconclusive, as the inherent characteristics of each country made any general model and interpretation unfeasible and uninterpretable.

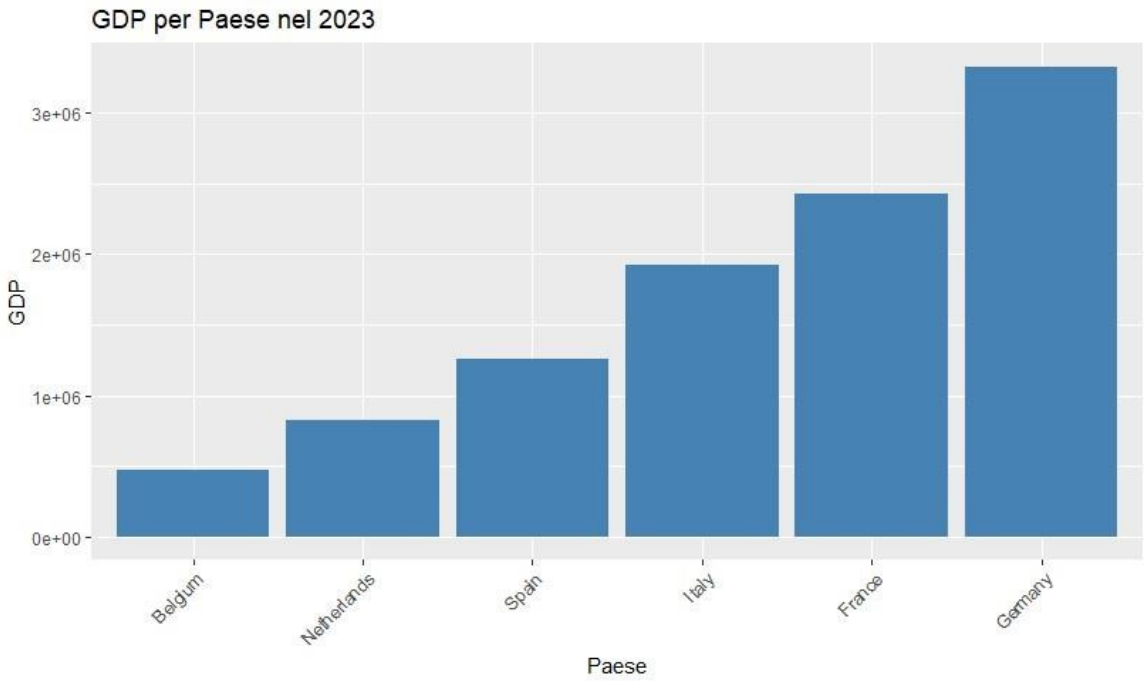
Thus, the focus shifted to the individual analysis of countries over time.

Having confirmed the change of focus, it was necessary to skim off the countries that could actually be analysed from those that were unsuitable. There were two selection criteria:

search for countries with **higher** and more stable **GDP** over time, and the **normality of** the data of the index variable T00/L40.



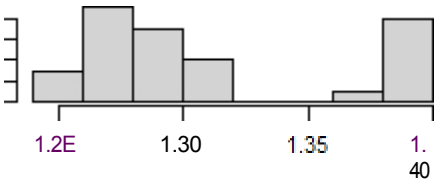
The first selection criterion, that using the GDP variable, ensures that the analysis concerns significant and less volatile economies, thus with greater comparability and economic relevance. This criterion led to an initial skimming of countries, focusing on Germany, France, Italy, Spain, the Netherlands and Belgium.



The second selection criterion, concerning the Normality of the distributions for the variable T00/L40, guarantees a strong robustness of the results and a more effective interpretability of the conclusions, as well as a further reduction of countries. In fact, the only countries to present normal distributions, according to Shapiro's test (a test particularly suitable for distributions with few values, in our case 32), were Spain and Italy.

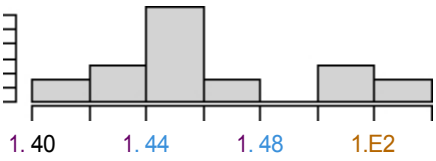
Below is a histogram representation of T00/L40 of the European countries, filtered by the first selection criterion.

France



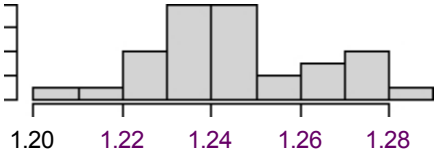
TO 0[[pa ese]]/L40[[pa ese]]

Germany



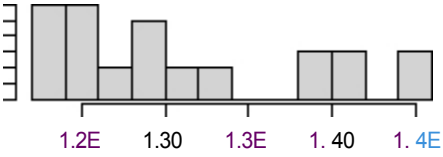
TO 0[[pa ese]]/L40[[pa ese]]

ltaly



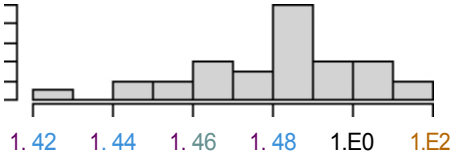
TO 0[[pa ese]]/L40[[pa ese]]

Netherlands



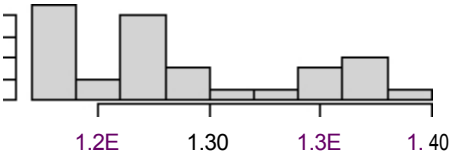
TO 0[[pa ese]]/L40[[pa ese]]

Spain



TO 0[[pa ese]]/L40[[pa ese]]

Belgium



TO 0[[pa ese]]/L40[[pa ese]]

Models Spain

Of the models used to analyse Spain's insurance solvency, the best model obtained from the R^2 perspective turns out to be a model obtained by the **stepwise** method

Below is **Summary** , **VIF** and **tests of** the **basic assumptions** of the selected models**:**

```
> summary(model_stepwise_spain)

Call:
lm(formula = T00/L40 ~ L20 + L50 + A30 + A40 + A80, data =
data_all_country$country ==
    "Spain", ])

Residuals:
       Min          1Q      Median          3Q         Max
-0.0144661 -0.0034754 0.0004712 0.0040067 0.0138663

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.813e-01  6.510e-02  15.073  2.30e-14  ***
L20  5 . 972e-06 3.950e-06 1 .512  O .143
L50  5 , 428e-06  4,994e-07  10,869  3,64e-11  *** A30  -
1,492e-06 1,741e-07 -8,572 4,73e-09 *** A40  6 , 724e-
06  1,336e-06  5 ,034  3,07e-05  ***  A80  3 , 825e-05
7,456e-06 5 ,130 2,39e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' 1

Residual standard error: 0.007427 on 26 degrees of freedom
Multiple R-squared:.906,        Adjusted R-squared:0.8879
F-statistic: 50.1 on 5 and 26 DF, p-value: 1.586e-12

> car::vif(model_stepwise_spain)
     L20      L50      A30      A40      A80
2.657611 1.658491 4.949411 2.981859 3.347959
```

Following the removal of the variable L20 (which was found to be non-significant with a p-value of 0.143)

there is a slight reduction in explained variability, while multicollinearity is significantly reduced.

```
> summary(model_spain)

Call:
lm(formula = T00/L40 ~ +L50 + A30 + A40 + A80, date =
date_all_countries[from ta_all_countries$Country ==
    "Spain", ])

Residuals:
       Min          1Q      Median          3Q         Max
-0.0149463 -0.0038443 -0.0002728 0.0043494 0.0139679

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1,016e+00  6,229e-02  16,316  1,65e-15  ***
L50  5 , 232e-06  4,936e-07  10,599  4,02e-11  *** A30  -
1,348e-06 1,490e-07 -9,044 1,17e-09 *** A40  8 , 210e-
06  9,256e-07  8 ,870  1,74e-09  ***  A80  3 , 404e-05
7,079e-06 4 ,808 5,10e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' 1

Residual standard error: 0.007602 on 27 degrees of freedom
Multiple R-squared: 0.8977, Adjusted R-squared: 0.8825
F-statistic: 59.23 on 4 and 27 DF, p-value: 5.644e-13

> car::vif(model_spain)
     L50      A30      A40      A80
1.546909 3.463312 1.366934 2.880396
```

Residue diagnostics

```
> shapiro.test(model_stepwise_spain$residuals)

        Shapiro-Wilk normality test

data: model_stepwise_spain$residuals
W = 0.98141, p-value = 0.8395
```

```
> lmtest::bptest(model_stepwise_spain)

        studentsBreusch-Pagan test

data: model_stepwise_spain
BP = 7.5023, df = 5, p-value = 0.1859
```

```
> lmtest::dwtest(model_stepwise_spain)

        Durbin-Watson test

data: model_stepwise_spain DW
= 2.0892, p-value = 0.3028
alternative hypothesis: true autocorrelation is greater than 0
```

All basic assumptions on residuals are therefore largely verified.

Model interpretation Spain

Interpretation **of model_spain:**

L50 (Defined Benefit Pension Fund Reserves):

- **Coefficient**: 5.428×10^{-6} (p-value < 0.001)
- **Interpretation**: This positive and highly significant coefficient indicates that an increase in defined benefit pension fund reserves is associated with an increase in the T00/L40 ratio. This suggests that defined benefit pension reserves provide greater stability to insurance companies, improving their ability to cover future claims.

A30 (Shares/units of investment funds and equities)

- **Coefficient**: -1.492×10^{-6} (p-value < 0.001)
- **Interpretation**: An increase in A30 is associated with a decrease in the ratio T00/L40. This suggests that holdings in shares and investment funds carry a higher risk for companies, which could reduce their financial stability in the event of market fluctuations.

A40 (Insurance technical reserves):

- **Coefficient**: 6.724×10^{-6} (p-value < 0.001)
- **Interpretation**: An increase in technical reserves is associated with an increase in the T00/L40 ratio. This is expected, as higher technical reserves offer more protection against future risks, improving the soundness of the company.

A80 (Other Activities):

- **Coefficient**: 3.825×10^{-5} (p-value < 0.001)
- **Interpretation**: This variable also has a significant positive effect, indicating that other assets, if properly managed, can contribute to greater financial stability of insurance companies. 'Other assets' may include diversified assets, which help to balance the overall portfolio and improve liquidity.

Italy Models

Of the models used to analyse Italy's insurance solvency, the best model obtained from the point of view of R^2 is again a model obtained by the **stepwise** method

Summary and **VIF** of the models considered are given below**:**

```
> summary(model_stepwise_italy)

Call:
lm(formula = T00/L40 ~ HICP + L20 + L50 + L70 + A30 + A40 + A60,
    date = data_all_countries[data_all_countries$Country == "Italy", ])

Residuals:
       Min         1Q     Median         3Q        Max
-0.0079059 -0.0025869 -0.0006758 0.0026565 0.0079164

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.314e+00 3.653e-02 35.986 < 2e-16 ***
HICP-2       .288e-03 1.139e-03 -2.010 0.055843 .
L201          .831e-06.028e-06          .782 0.087459 .L501
             .796e-06.044e-07          .7885.76e-09 ***
L701          .466e-05.506e-06          .954 0.062473 .
A30-3        .555e-07 5.449e-08 -6.525 9.51e-07 ***
A40-4        .770e-06 1.959e-06 -2.435 0.022705 *
A60-3        .003e-07 7.013e-08 -4.282 0.000258 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' 1

Residual standard error: 0.004489 on 24 degrees of freedom
Multiple R-squared: 0.9607, Adjusted R-squared: 0.9492
F-statistic: 83.71 on 7 and 24 DF, p-value: 2.64e-15

> car::vif(model_stepwise_italy)
     HICP       L20       L50       L70       A30       A40
22.151347  9.247546 16.246950  5.857374  7.746101  8.429911
      A60
18.869400
```

Following the removal of the variables HICP, L70, A40, by means of an iterative procedure, there is again a slight reduction in the variability explained, against a strong reduction in multicollinearity.

```
> summary(model_italy)

Call:
lm(formula = T00/L40 ~ +L20 + L50 + A30 + A60, date =
date_all_countries[from ta_all_countries$Country ==
    "Italy", ])

Residuals:
       Min          1Q      Median          3Q         Max
-0.0081089 -0.0028824 -0.0009905 0.0028038 0.0087950

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.246e+00 1.876e-02 66.423 < 2e-16 ***
L208         .637e-07.030e-07          .143 0.041272 *L501
             .392e-06404e-07          .9121.71e-10 ***
A30-2       .444e-07 2.282e-08 -10.709 3.20e-11 ***
A60-1       .886e-07 4.621e-08 -4.082 0.000356 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' 1

Residual standard error: 0.004798 on 27 degrees of freedom
Multiple R-squared: 0.9494, Adjusted R-squared: 0.9419 F-
statistic: 126.8 on 4 and 27 DF, p-value: < 2.2e-16


> car::vif(model_italy)
     L20      L50      A30      A60
1.245039 6.712279 1.189836 7.173432
```

Residue diagnostics

```
> shapiro.test(model_stepwise_italy$residuals)

        Shapiro-Wilk normality test

data: model_stepwise_italy$residuals
W = 0.97836, p-value = 0.7506
```

```
> lmtest::bptest(model_stepwise_italy)

        students Breusch-Pagan test

data: model_stepwise_italy
BP = 3.897, df = 7, p-value = 0.7916
```

```
> lmtest::dwtest(model_stepwise_italy)

        Durbin-Watson test

data: model_stepwise_italy DW
= 2.0594, p-value = 0.2221
alternative hypothesis: true autocorrelation is greater than 0
```

All basic assumptions on residuals are therefore largely verified.

Interpretation models Italy

In the interpretation of the models, we focus on the analysis of the model obtained by means of the stepwise procedure since, having more variables, it is more interesting from an explanatory point of view.

HICP (Harmonised Index of Consumer Prices):

- **Coefficient**: -0.002288 (p-value = 0.055843)
- **Interpretation**: An increase in inflation is associated with a decrease in the T00/L40 ratio. This implies that a rising inflation environment may erode the real value of assets relative to technical reserves, reducing the perceived stability of insurance companies. Although close to the 5% significance level, its impact requires attention.

L20 (Contracted Loans):

- **Coefficient**: 1.831×10^{-6} (p-value = 0.087459)
- **Interpretation**: The positive coefficient indicates that increased borrowing has a positive effect on the T00/L40 ratio, but the p-value suggests that it is not statistically significant at the 5% level. This implies that the effect of loans is not determined with sufficient certainty, and may not be a relevant factor in the present context.

L50 (Defined Benefit Pension Fund Reserves):

- **Coefficient**: 1.796×10^{-6} (p-value < 0.001)
- **Interpretation**: This positive and highly significant coefficient indicates that an increase in defined benefit pension reserves is associated with an improvement in the T00/L40 ratio. Pension reserves provide greater stability, helping companies to manage future liabilities more effectively.

L70 (Other liabilities):

- **Coefficient**: 1.466×10^{-5} (p-value = 0.062473)
- **Interpretation**: The coefficient suggests that an increase in the other liabilities is associated with an increase in the T00/L40 ratio. Although the p-value is close to the significance limit, it could imply that effective management of the other liabilities may contribute to greater stability.

A30 (Shares/units of investment funds and equities)

- **Coefficient**: -3.555×10^{-7} (p-value < 0.001)
- **Interpretation**: An increase in this variable is associated with a decrease in the T00/L40 ratio. This suggests that companies with greater exposure to the stock market may be more vulnerable to market risks, compromising their financial stability.

A40 (Insurance technical reserves):

- **Coefficient**: -4.770×10^{-6} (p-value = 0.022705)
- **Interpretation**: An increase in technical reserves is associated with a decrease in the T00/L40 ratio. This is counterintuitive, as one would expect an increase in reserves to improve stability. However, it may reflect the fact that higher technical reserves are needed to meet higher liabilities, so the company may be in a higher risk phase.

A60 (Shares/units of investment funds held by insurance companies)

- **Coefficient**: -3.003×10^{-7} (p-value < 0.001)
- Interpretation: Similar to A30, higher exposure in stocks/shares o f investment funds is associated with a decline in the T00/L40 ratio. This suggests that investments in risky assets may undermine companies' ability to cope with future claims, highlighting the need for prudent investment management.

Conclusion

Based on the linear regression models analysed for insurance companies in Italy and Spain, we can draw some significant conclusions regarding the economic and financial variables that influence the stability and performance of the insurance sector in both countries. Here are the main considerations:

Financial Stability of Insurance Companies

- **Technical and Pension Reserves**: The financial stability of insurance companies is strongly related to insurance technical reserves (A40) and defined benefit pension fund reserves (L50). An increase in these reserves is associated with an improvement in the T00/L40 ratio, indicating that companies need to maintain adequate levels of reserves to cover future liabilities. The prudent management of these reserves is crucial for policyholder confidence and the sustainability of the industry.

Inflation and Market Risk

- **Impact of the** HICP: In Italy, the Harmonised Index of Consumer Prices (HICP) shows a negative effect on the T00/L40 ratio, suggesting that rising inflation may jeopardise the capital stability of insurance companies. Companies should be prepared to implement strategies to mitigate the impact of inflation on their balance sheets.

- **Investment Exposure**: Both models reveal that a higher exposure in equities and investment funds (A30 and A60) is associated with a decrease in the T00/L40 ratio. This indicates that insurance companies in both countries need to adopt prudent risk management to prevent market fluctuations from undermining their financial stability.

Management of Liabilities

- **Other Liabilities**: The presence of other liabilities (L70) has a positive influence on financial stability in Italy, although the results are less clear in Spain. This suggests that companies may need to actively manage their liabilities to maintain a sound financial balance and a positive reputation.

Diversity in the Models of Italy and Spain

- **Differences in Coefficients**: Although there are some similarities in the significant variables, differences in the coefficients and p-values between the two models indicate that economic and financial dynamics may vary significantly between countries.

Strategic Implications for the Sector

- **Planning and Risk Management**: Insurance companies need to take a strategic approach to resource planning, considering economic trends and market trends. Diversification of investments and sound reserve management can help mitigate the risks associated with inflation and market volatility.
- Constant Monitoring: Constant monitoring of economic variables and company performance is essential. Companies must be ready to adapt to changes in the economic environment to maintain their competitiveness and stability.

In summary, the analysis of the regression models for Italy and Spain highlights the complexity of the relationships between economic/financial variables and the stability of insurance companies. Companies need to pay attention to reserve management, the impact of inflation and their investment strategies to ensure sound financial performance in a changing economic environment.