

Louvain School of Management
and Instituto Superior Técnico

Natural Gas Price Impact on the Euro Area Sovereign Bonds

The feasibility of a simplistic approach

Author's:
Inês Patrícia Dias Cabral

Master's thesis with the view of getting the degrees:
Master in Industrial Engineering and Management, professional focus
Financial Management +
Master in Business Engineering, professional focus Financial
Engineering

Supervisor's:
Prof. Leonardo Iania and Prof. Cláudia Nunes Philippart

Academic Year 2022-2023



Natural Gas price impact on the Euro area Sovereign Bonds

The feasibility of a simplistic approach

Inês Patrícia Dias Cabral

*Thesis to obtain the Master of Science
Double Degree in Business Engineering and Industrial Engineering and
Management*

Instituto Superior Técnico & Louvain School of Management

July 2023

Supervisors:

Prof. Leonardo Iania
Prof. Cláudia Nunes Philippart

Examination Committee:

Chairperson: Prof. Tânia Rute Xavier de Matos Pinto-Varela
Members of the Committee: Prof. Isabel M. Rodrigues

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to both my supervisors. To Professor Leonardo Lania for introducing me to the theme of this thesis, for increasing my interest in the world of Financial Econometrics by providing me several opportunities to dive deeper into the current tendencies of research in this field, and for the support. To Professor Cláudia Nunes Philippart for the constant availability, support, and guidance from the first day I started developing this dissertation and additionally for aiding me in the major selection previously to being my supervisor. Their expertise and insightful feedback significantly shaped the direction of this thesis.

Furthermore, I would like to extend my acknowledgement to Instituto Superior Técnico for the rigorous and demanding education it gave me, starting from my Bachelor's degree, and for providing me the opportunity to shape my path, that I gradually discovered over the course of these six years. I would also like to express my gratitude to the Université Catholique de Louvain and Louvain School of Management for their contribution to improving my understanding of Financial Econometrics. This last year has provided me with a new and enriching perspective, further enhancing my academic journey. Additionally, I would like to express my gratitude to the several Professors that have shaped my path, namely the ones that gave valuable input to the development of this dissertation. After all, I would like to express my utmost gratitude to both of these esteemed institutions, and the involved individuals, for providing me with the priceless chance to pursue this Double Master Degree. This experience has undeniably enriched the end of my academic journey.

On a more personal level, I would like to thank my parents and my brother for the patience and support for me to find and follow my path throughout my life. Their guidance to make me always want to achieve my best and to never shut doors has made me reach this final step. You are my best friends, and without you this wouldn't be possible. Furthermore, I would like to thank my mother for dropping her busy life and helping me in the challenging final stages of this thesis.

I would like to thank my boyfriend for being my rock, for listening to my worries, insecurities, and even the theoretical reasoning behind this thesis. Thank you for always lifting me up and pushing me to do my best. Thank you for your unconditional support to every decision I have made and for always being there, even at miles away.

Thank you to the rest of my family for being there when I needed the most, namely to all my four grad-parents for the love and support they have given me in this journey. Furthermore, thank you to my uncle, my aunt and my cousins for receiving me with open arms and for being a second home when I was far away from my own, I will never forget your care and support in the transitions I had to face in these past six years.

Finally, I want to thank my friends. To the ones from my childhood and upbringing, for shaping the happiest memories and for keep being present in my live. To the ones that joined me in the "Nobre Instituto" journey since day one, for sharing with me my fist steps in university and for backing me up and always being there in my many changes of path, you have made these years unforgettable. From these friends, a thank you to the ones who made part of the Student's Union adventure, to work with you has been one of the most fulfilling experiences I can remember, what we achieved and the values we shared will always be present in my live.

In conclusion, I would like to express my deepest appreciation to all those whose names may not be mentioned here but have played a role in this work. Your support, whether big or small, has made a difference.

Thank you.

Abstract

This master's thesis presents a comprehensive analysis of the relationship between the Title Transfer Facility (TTF) natural gas contract price and sovereign bond yields using some simplistic methods. In fact, the relationship between commodity prices and sovereign bond yields has become a topic of interest due to its potential implications for financial markets and the macroeconomic landscape.

This dissertation investigates the impact of natural gas prices on sovereign bond yields in the European Union (EU) and analyzes their evolving relationship. The introduction provides an overview of the factors influencing sovereign bond yields, and the complexity of the natural gas market in the EU, emphasizing the importance of understanding the expected behavior of bond yields in response to changes in natural gas prices.

The key research aims to fill the gap in understanding the specific relationship between natural gas prices and sovereign bond yields, providing a simple and comprehensive understanding of this contemporary and current topic. The selected countries—Germany, Italy, Spain, Hungary, and France—offer insights into the varying degrees of dependence in natural gas, specially of Russian origin, diversification needs, and the influence of energy mixes, providing a comprehensive analysis of natural gas dynamics in Europe.

The analysis is divided into four sections: Ordinary Least Squares (OLS) regression, Least Absolute Selection and Shrinkage Operator (LASSO) regression, Shapley Values, and a periodical analysis. The results obtained from these methods provide an initial evidence of the existing relationship between the TTF variable and bond yields.

The OLS regression results reveal a significant association between the TTF variable and changes in yields, particularly in the longer term. The LASSO regression further validates these findings, demonstrating the importance of the TTF variable in explaining bond yield variations. The Shapley Values analysis quantitatively assesses the variable contributions, consistently highlighting the impact of the TTF variable on yields, albeit with some variations across countries and time periods. The periodical analysis reveals shifts in the average contribution of the TTF variable during different periods, further emphasizing its relevance.

However, the validation and discussion of the applied models highlight challenges associated with the underlying assumptions and limitations of the statistical methods. Violations of the OLS assump-

tions, such as autocorrelation, heteroskedasticity, and non-normality in the residuals, compromise the accuracy and reliability of the estimated coefficients. The presence of mis-specification and possible omitted variable bias raises concerns about the interpretation and robustness of the results.

In fact the reliance on machine learning techniques, including LASSO and Shapley Values, presents both advantages and limitations. Biased variable selection and the assumption of linearity in these methods can affect the interpretation of the results. Additionally, the non-linear relationship between the explanatory and dependent variables, as well as the reactive and memory-based nature of the data, pose challenges to model specification.

In conclusion, while the statistical methods applied provide some insights into the relationship between the TTF variable and bond yields, careful consideration of the underlying assumptions, model validation, and alternative regression techniques is essential. The limitations of machine learning techniques emphasize the need for rigorous statistical review and validation to ensure accurate interpretation of the results. The complex data characteristics require meticulous model specification and further investigation of alternative methodologies.

Keywords

Natural Gas; Sovereign Bond Yields; Ordinary Least Squares (OLS) regression; Least Absolute Selection and Shrinkage Operator (LASSO) regression; Shapley Values; model mis-specification;

Resumo

A presente dissertação de mestrado apresenta uma análise abrangente da relação entre o preço do gás natural segundo o contrato Holandês do tipo Title Transfer Facility (TTF) e as obrigações soberanas de países da União Europeia, utilizando vários métodos simplificados. A relação entre os preços de matérias-primas e as obrigações soberanas tornou-se um tema de interesse devido às suas potenciais implicações nos mercados financeiros e na paisagem macroeconómica. A introdução fornece uma visão geral dos fatores que influenciam as obrigações soberanas e a complexidade do mercado do gás natural na União Europeia, enfatizando a importância de compreender o comportamento esperado do objeto financeiro de interesse em resposta às mudanças nos preços do gás natural.

A pesquisa tem como objetivo preencher a lacuna na compreensão da relação específica dos preços do gás natural, fornecendo uma compreensão simples e abrangente deste tópico contemporâneo e atual. Os países selecionados - Alemanha, Itália, Espanha, Hungria e França - oferecem um leque diversificado no sentido de representarem os diferentes graus de dependência em relação ao gás natural, especialmente de origem russa, e as suas gerais necessidades de diversificação, proporcionando uma análise abrangente da dinâmica do gás natural na Europa.

A análise é dividida em quatro seções: a primeira sobre a regressão Ordinary Least Squares (OLS), regressão Least Absolute Selection and Shrinkage Operator (LASSO), Valores de Shapley e uma análise periódica. Os resultados obtidos por meio destes métodos fornecem evidências iniciais da relação existente entre a variável TTF (tipo de contrato de gás natural) e os preços das obrigações soberanas.

Os resultados da regressão OLS revelam uma associação significativa entre a variável TTF e os preços das obrigações soberanas, especialmente a longo prazo. A regressão LASSO confirma estes resultados, demonstrando a importância da variável TTF na explicação das variações os preços das obrigações. A análise dos Valores de Shapley avalia mais quantitativamente as contribuições das variáveis, destacando consistentemente o impacto da variável TTF nos preços, embora com algumas variações entre os países e períodos considerados. A análise periódica revela mudanças na contribuição média da variável TTF durante diferentes períodos, enfatizando ainda mais sua relevância.

No entanto, a validação e discussão dos modelos aplicados destacam os desafios associados às suposições subjacentes e limitações dos métodos estatísticos. Violações das suposições da regressão

OLS, como autocorrelação, heteroscedasticidade e não-normalidade nos resíduos, comprometem a precisão e robustez dos coeficientes estimados. A presença de má especificação do modelo inicial e a possibilidade da existência de variáveis omitidas levanta preocupações sobre a interpretação previamente realizada dos resultados.

De facto, a dependência de técnicas de "machine-learning", incluindo LASSO e Valores de Shapley, apresentam tanto vantagens como limitações. A seleção viesada de variáveis e a suposição de linearidade nestes métodos podem afetar a interpretação dos resultados. Além disso, a relação não linear entre as variáveis explicativas e dependentes, bem como a natureza reativa e baseada em memória dos dados, representam desafios para a especificação do modelo.

Em conclusão, embora os métodos estatísticos aplicados forneçam algumas informações sobre a relação entre a o preço do gás natural e os preços das obrigações soberanas, é essencial uma cuidadosa consideração das suposições subjacentes, validação do modelo e técnicas alternativas de regressão. As limitações das técnicas de "machine-learning" enfatizam a necessidade de uma revisão estatística rigorosa e validação para garantir uma interpretação precisa dos resultados. As características complexas dos dados exigem uma especificação cuidadosa do modelo e investigação adicional de metodologias alternativas.

Palavras Chave

Gás Natural; Obrigações Soberanas; Regressão Ordinary Least Squares (OLS); Regressão Least Absolute Selection and Shrinkage Operator (LASSO); Valores de Shapley; Má Especificação do Modelo.

Contents

1	Introduction	1
1.1	Sovereign Bond Yields	3
1.2	Natural Gas	6
1.2.1	European Union Natural Gas Industry	6
1.2.1.A	European Union Natural Gas Trade	7
1.2.1.B	European Union Natural Gas Demand	8
1.2.1.C	European Union Natural Gas Production	8
1.2.1.D	European Union Natural Gas Supply	9
1.2.2	Russian Natural Gas Crisis	11
1.2.2.A	Previous Relationship	11
1.2.2.B	Current Crisis	12
1.2.3	Countries Characteristics	13
1.2.3.A	Germany	13
1.2.3.B	Italy	13
1.2.3.C	Spain	14
1.2.3.D	France	14
1.2.3.E	Hungary	15
1.3	Event Analysis	15
1.4	Literature Review	17
1.5	Positioning and Research Question	20
1.6	Organization of the Document	21
2	Data Analysis	23
2.1	Data description	25
2.1.1	Statistical Description	27
2.2	Stationarity tests and Transformations	28
2.2.1	Explanatory Variables	29
2.2.2	Dependent Variables	32

2.3	Correlation	35
3	Methodology	37
3.1	Ordinary Least Squares Regression	39
3.2	Least Absolute Selection and Shrinkage Operator	40
3.3	Shapley Values	42
3.4	Periodical Analysis	43
4	Results Discussion	45
4.1	Ordinary Least Squares Regression	47
4.1.1	OLS Commodity Comparison	48
4.2	LASSO	49
4.2.1	LASSO Commodity Comparison	51
4.3	Shapley Values	53
4.3.1	Shapley Values Commodity Comparison	54
4.4	Periodical Analysis	55
4.4.1	Least Absolute Selection and Shrinkage Operator	55
4.4.2	Shapley Values	57
4.5	The Overall Results	58
5	Results Validation	61
5.1	Ordinary Least Square Regression	63
5.2	Cross-validation	65
5.3	Other Regression Tryouts	65
5.4	LASSO	66
5.5	Shapley Values	68
5.6	Data Characteristics	69
5.7	The Machine Learning Tendency	69
6	Conclusion	71
6.1	Conclusions	73
6.2	Future Work	75
	Bibliography	77
A	Supplementary Plots and Graphics	81
A.1	Event Analysis	81
A.2	Raw Explanatory Data	82

A.3	Results Full Sample	83
A.3.1	OLS Results	83
A.3.2	LASSO Results	85
A.3.3	Shapley values Results	89
A.4	Results Period Analysis	96
A.4.1	LASSO Results	96
A.4.2	Shapley Values Results	100
A.5	Residual Plots	107
A.6	Full Residual Tests Statistics	114
A.7	Tryouts	114
A.7.1	Interaction	115
A.7.2	Manually removed Outliers	116
A.7.3	Robust	117
A.7.4	Lag 1 day	118
A.7.5	Lag 1 month	119
A.7.6	Lag 1 year	120

List of Figures

1.1	Sovereign Bond Yields evolution from January 2018 to June 2023	4
1.2	Natural Gas Prices (TTF Dutch contract) from January 2018 to June 2023	7
1.3	Dependence on Russian Gas; Sources: [1]	10
1.4	Sources: IMF staff based on ENTSOG (System Development Map, 2021, and Transparency platform) [1]	11
1.5	Changes in Germany Bond Yields (10 year in red and 1 year in orange)	16
2.1	Graphical representation Title Transfer Facility (TTF)	30
2.2	Graphical representation VIX	31
2.3	Graphical representation POL	31
2.4	Graphical representation F2S	31
2.5	Graphical representation OIL	31
2.6	Graphical representation Change in Germany Yields	33
2.7	Graphical representation Change in Italy Yields	33
2.8	Graphical representation Change in Spain Yields	33
2.9	Graphical representation Change in France Yields	33
2.10	Graphical representation Change in Hungary Yields	34
2.11	Graphical representation Spread in Germany Yields	34
2.12	Graphical representation Spread in Italy Yields	34
2.13	Graphical representation Spread in Spain Yields	34
2.14	Graphical representation Spread in France Yields	34
2.15	Graphical representation Spread in Hungary Yields	34
2.16	Correlation matrix of the variables	35
4.1	Coefficients for each LASSO λ value for BGe10	50
4.2	Coefficients for each LASSO λ value for SpBGe10	50
4.3	Shapley Values for the change in yields for Italy 1y bond	53
4.4	Force Plot of the Shapley Values for the change in yields for Italy 1y bond	53
4.5	TTF Coefficients for sub-sample period for BGe10	56

4.6	TTF period Shapley Values for Blt1	57
4.7	Force Plot of the TTF period Shapley Values for Blt1	57
A.1	Event Analysis Italy	81
A.2	Event Analysis Spain	81
A.3	Event Analysis France	82
A.4	Event Analysis Hungary	82
A.5	Raw VIX	82
A.6	Raw POL	82
A.7	Raw F2S	83
A.8	Raw OIL	83
A.9	Lasso coefficients for BGe1	85
A.10	Lasso coefficients for BGe10	85
A.11	Lasso coefficients for Blt1	86
A.12	Lasso coefficients for Blt10	86
A.13	Lasso coefficients for BSp1	86
A.14	Lasso coefficients for BSp10	86
A.15	Lasso coefficients for BFr1	86
A.16	Lasso coefficients for BFr10	86
A.17	Lasso coefficients for BHn1	87
A.18	Lasso coefficients for BHn10	87
A.19	Lasso coefficients for SpBGe1	87
A.20	Lasso coefficients for SpBGe10	87
A.21	Lasso coefficients for SpBlt1	87
A.22	Lasso coefficients for SpBlt10	87
A.23	Lasso coefficients for SpBSp1	88
A.24	Lasso coefficients for SpBSp10	88
A.25	Lasso coefficients for SpBFr1	88
A.26	Lasso coefficients for SpBFr10	88
A.27	Lasso coefficients for SpBHn1	88
A.28	Lasso coefficients for SpBHn10	88
A.29	Relative Shapley Values for BGe1	89
A.30	Force plot for Shapley values BGe1	89
A.31	Relative Shapley Values for BGe10	89
A.32	Force plot for Shapley values BGe10	89
A.33	Relative Shapley Values for Blt1	90

A.34	Force plot for Shapley values Blt1	90
A.35	Relative Shapley Values for Blt10	90
A.36	Force plot for Shapley values Blt10	90
A.37	Relative Shapley Values for BSp1	90
A.38	Force plot for Shapley values BSp1	90
A.39	Relative Shapley Values for BSp10	91
A.40	Force plot for Shapley values BSp10	91
A.41	Relative Shapley Values for BFr1	91
A.42	Force plot for Shapley values BFr1	91
A.43	Relative Shapley Values for BFr10	91
A.44	Force plot for Shapley values BFr10	91
A.45	Relative Shapley Values for BHn1	92
A.46	Force plot for Shapley values BHn1	92
A.47	Relative Shapley Values for BHn10	92
A.48	Force plot for Shapley values BHn10	92
A.49	Relative Shapley Values for SpBGe1	92
A.50	Force plot for Shapley values SpBGe1	92
A.51	Relative Shapley Values for SpBGe10	93
A.52	Force plot for Shapley values SpBGe10	93
A.53	Relative Shapley Values for SpBlt1	93
A.54	Force plot for Shapley values SpBlt1	93
A.55	Relative Shapley Values for SpBlt10	93
A.56	Force plot for Shapley values SpBlt10	93
A.57	Relative Shapley Values for SpBSp1	94
A.58	Force plot for Shapley values SpBSp1	94
A.59	Relative Shapley Values for SpBSp10	94
A.60	Force plot for Shapley values SpBSp10	94
A.61	Relative Shapley Values for SpBFr1	94
A.62	Force plot for Shapley values SpBFr1	94
A.63	Relative Shapley Values for SpBFr10	95
A.64	Force plot for Shapley values SpBFr10	95
A.65	Relative Shapley Values for SpBHn1	95
A.66	Force plot for Shapley values SpBHn1	95
A.67	Relative Shapley Values for SpBHn10	95
A.68	Force plot for Shapley values SpBHn10	95

A.69	Lasso coefficients for Period analysis BGe1	96
A.70	Lasso coefficients for Period analysis BGe10	96
A.71	Lasso coefficients for Period analysis Blt1	96
A.72	Lasso coefficients for Period analysis Blt10	96
A.73	Lasso coefficients for Period analysis BSp1	97
A.74	Lasso coefficients for Period analysis BSp10	97
A.75	Lasso coefficients for Period analysis BFr1	97
A.76	Lasso coefficients for Period analysis BFr10	97
A.77	Lasso coefficients for Period analysis BHn1	97
A.78	Lasso coefficients for Period analysis BHn10	97
A.79	Lasso coefficients for Period analysis SpBGe1	98
A.80	Lasso coefficients for Period analysis SpBGe10	98
A.81	Lasso coefficients for Period analysis SpBlt1	98
A.82	Lasso coefficients for Period analysis SpBlt10	98
A.83	Lasso coefficients for Period analysis SpBSp1	98
A.84	Lasso coefficients for Period analysis SpBSp10	98
A.85	Lasso coefficients for Period analysis SpBFr1	99
A.86	Lasso coefficients for Period analysis SpBFr10	99
A.87	Lasso coefficients for Period analysis SpBHn1	99
A.88	Lasso coefficients for Period analysis SpBHn10	99
A.89	Relative Shapley Values for Period Analysis BGe1	100
A.90	Force plot for Shapley Values for Period Analysis BGe1	100
A.91	Relative Shapley Values for Period Analysis BGe10	100
A.92	Force plot for Shapley Values for Period Analysis BGe10	100
A.93	Relative Shapley Values for Period Analysis Blt1	101
A.94	Force plot for Shapley Values for Period Analysis Blt1	101
A.95	Relative Shapley Values for Period Analysis Blt10	101
A.96	Force plot for Shapley Values for Period Analysis Blt10	101
A.97	Relative Shapley Values for Period Analysis BSp1	102
A.98	Force plot for Shapley Values for Period Analysis BSp1	102
A.99	Relative Shapley Values for Period Analysis BSp10	102
A.100	Force plot for Shapley Values for Period Analysis BSp10	102
A.101	Relative Shapley Values for Period Analysis BFr1	103
A.102	Force plot for Shapley Values for Period Analysis BFr1	103
A.103	Relative Shapley Values for Period Analysis BFr10	103

A.104	Force plot for Shapley Values for Period Analysis BFr10	103
A.105	Relative Shapley Values for Period Analysis BHn1	104
A.106	Force plot for Shapley Values for Period Analysis BHn1	104
A.107	Relative Shapley Values for Period Analysis BHn10	104
A.108	Force plot for Shapley Values for Period Analysis BHn10	104
A.109	Relative Shapley Values for Period Analysis SpBGe1	105
A.110	Relative Shapley Values for Period Analysis SpBGe10	105
A.111	Relative Shapley Values for Period Analysis SpBl1	105
A.112	Relative Shapley Values for Period Analysis SpBl10	105
A.113	Relative Shapley Values for Period Analysis SpBSp1	106
A.114	Relative Shapley Values for Period Analysis SpBSp10	106
A.115	Relative Shapley Values for Period Analysis SpBFr1	106
A.116	Relative Shapley Values for Period Analysis SpBFr10	106
A.117	Relative Shapley Values for Period Analysis SpBHn1	107
A.118	Relative Shapley Values for Period Analysis SpBHn10	107
A.119	ACF residuals plot for BGe1	107
A.120	QQ plot for residuals plot for BGe1	107
A.121	Residuals vs Fitted plot for BGe1	107
A.122	ACF residuals plot for BGe10	107
A.123	QQ plot for residuals plot for BGe10	107
A.124	Residuals vs Fitted plot for BGe10	107
A.125	ACF residuals plot for BFr1	108
A.126	QQ plot for residuals plot for BFr1	108
A.127	Residuals vs Fitted plot for BFr1	108
A.128	ACF residuals plot for BFr10	108
A.129	QQ plot for residuals plot for BFr10	108
A.130	Residuals vs Fitted plot for BFr10	108
A.131	ACF residuals plot for Bl1	108
A.132	QQ plot for residuals plot for Bl1	108
A.133	Residuals vs Fitted plot for Bl1	108
A.134	ACF residuals plot for Bl10	109
A.135	QQ plot for residuals plot for Bl10	109
A.136	Residuals vs Fitted plot for Bl10	109
A.137	ACF residuals plot for BSp1	109
A.138	QQ plot for residuals plot for BSp1	109

A.139	Residuals vs Fitted plot for BSp1	109
A.140	ACF residuals plot for BSp10	109
A.141	QQ plot for residuals plot for BSp10	109
A.142	Residuals vs Fitted plot for BSp10	109
A.143	ACF residuals plot for BHn1	110
A.144	QQ plot for residuals plot for BHn1	110
A.145	Residuals vs Fitted plot for BHn1	110
A.146	ACF residuals plot for BHn10	110
A.147	QQ plot for residuals plot for BHn10	110
A.148	Residuals vs Fitted plot for BHn10	110
A.149	ACF residuals plot for SpBGe1	110
A.150	QQ plot for residuals plot for SpBGe1	110
A.151	Residuals vs Fitted plot for SpBGe1	110
A.152	ACF residuals plot for SpBGe10	111
A.153	QQ plot for residuals plot for SpBGe10	111
A.154	Residuals vs Fitted plot for SpBGe10	111
A.155	ACF residuals plot for SpBFr1	111
A.156	QQ plot for residuals plot for SpBFr1	111
A.157	Residuals vs Fitted plot for SpBFr1	111
A.158	ACF residuals plot for SpBFr10	111
A.159	QQ plot for residuals plot for SpBFr10	111
A.160	Residuals vs Fitted plot for SpBFr10	111
A.161	ACF residuals plot for SpBl1	112
A.162	QQ plot for residuals plot for SpBl1	112
A.163	Residuals vs Fitted plot for SpBl1	112
A.164	ACF residuals plot for SpBl10	112
A.165	QQ plot for residuals plot for SpBl10	112
A.166	Residuals vs Fitted plot for SpBl10	112
A.167	ACF residuals plot for SpBSp1	112
A.168	QQ plot for residuals plot for SpBSp1	112
A.169	Residuals vs Fitted plot for SpBSp1	112
A.170	ACF residuals plot for SpBSp10	113
A.171	QQ plot for residuals plot for SpBSp10	113
A.172	Residuals vs Fitted plot for SpBSp10	113
A.173	ACF residuals plot for SpBHn1	113

A.174	QQ plot for residuals plot for SpBHn1	113
A.175	Residuals vs Fitted plot for SpBHn1	113
A.176	ACF residuals plot for SpBHn10	113
A.177	QQ plot for residuals plot for SpBHn10	113
A.178	Residuals vs Fitted plot for SpBHn10	113

List of Tables

2.1	Retrieved Variables Description - Part 1	26
2.2	Retrieved Variables Description - Part 2	27
2.3	Statistics Table	28
2.4	Explanatory Variables Stationary Tests Results Table	30
2.5	Transformed Explanatory Variables Stationary Tests Results Table	31
2.6	Dependent Variables Stationary Tests	32
4.1	TTF coefficients and p-value for the different regressions	47
4.2	OIL coefficient and p-value for the different regressions	49
4.3	Summary table LASSO Results	50
4.4	Summary table OIL vs TTF LASSO Results	52
4.5	Summary table Shapley Results	54
4.6	Summary table Commodity Comparison Shapley Results	55
4.7	Summary period analysis table LASSO Results	56
4.8	Summary table Shapley period analysis Results	58
5.1	Residual tests results	64
5.2	Summary table LASSO Results	68
A.1	Final table with Coefficient and P-Value for VIX	83
A.2	Final table with Coefficient and P-Value for F2S	84
A.3	Final table with Coefficient and P-Value for the desired variable	84
A.4	Regression's F-statistic, p-value, and R-squared	85
A.5	Residual tests results	114
A.6	Final table with Coefficient and P-Value for TTF Interpolation Tryout	115
A.7	Residual tests results Interpolation Tryout	115
A.8	Final table with Coefficient and P-Value for TTF in Manually removed Outliers Tryout	116
A.9	Residual tests results Manually removed Outliers Tryout	116
A.10	Final table with Coefficient and P-Value for TTF in Robust Regression Tryout	117
A.11	Residual tests results Robust Regression Tryout	117

A.12 Final table with Coefficient and P-Value for TTF Lag 1 Day Tryout	118
A.13 Residual tests results Lag 1 Day Tryout	118
A.14 Final table with Coefficient and P-Value for TTF Lag 1 Month Tryout	119
A.15 Residual tests results Lag 1 Month Tryout	119
A.16 Final table with Coefficient and P-Value for TTF Lag 1 Year Tryout	120
A.17 Residual tests results Lag 1 Year Tryout	120

Acronyms

ADF	Augmented Dickey Fuller
bcm	billion cubic meter
DWD	Daily Working Days
ECB	European Central Bank
EONIA	Euro Overnight Index Average
ESI	European Sentiment Indicator
ESTR	Euro Short-Term Rate
EU	European Union
FSRU	Floating Storage Regasification Units
KfW	Kreditanstalt für Wiederaufbau
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LASSO	Least Absolute Selection and Shrinkage Operator
LNG	Liquefied Natural Gas
OIS	Overnight Indexed Swaps
OLS	Ordinary Least Squares
PP	Phillips-Perron
TTF	Title Transfer Facility

1

Introduction

Contents

1.1	Sovereign Bond Yields	3
1.2	Natural Gas	6
1.3	Event Analysis	15
1.4	Literature Review	17
1.5	Positioning and Research Question	20
1.6	Organization of the Document	21

The relationship between commodity prices and sovereign bond yields has become a topic of interest due to its potential implications for financial markets and the macroeconomic landscape. Understanding the expected behavior of bond yields in response to changes in natural gas prices can provide valuable insights for policymakers, investors, and market participants. This introduction draws upon various sources to shed light on the factors influencing sovereign bond yields and their connection to the current natural gas market behaviour.

1.1 Sovereign Bond Yields

The study of sovereign bond yields holds considerable importance owing to the crucial function played by this particular bond market in providing financial insights and the broad range of factors that impact these yields. Sovereign bonds are frequently regarded as secure investments, contributing to portfolios with a high level of diversification and serving as a safeguard during times of crisis, commonly referred to as the "flight to quality" phenomenon [2].

Although sovereign debt is generally considered to carry lower risk compared to private-sector debt due to the government's ability to raise funds through taxation [3], borrowing costs for governments are influenced by macroeconomic and financial conditions the country in question faces. This leads to differences in the bond yields, even within the same monetary union, as it was the case in the European sovereign debt crisis, which highlighted divergent macroeconomic performances and solvency risks among Eurozone countries, leading to different pricing in the bond market [4].

Therefore, analyzing the behavior of sovereign bond markets can also provide insights into the overall health of a country's economy and its perceived creditworthiness. This information can be useful to the main market players in making informed decisions about investments, risk management strategies and even monetary and fiscal policies. Furthermore, understanding the determinants of yield differentials and analyzing the effects of global events on sovereign bond markets, such as the COVID-19 pandemic and the recent invasion of Ukraine, can provide very valuable insights [2].

Looking at the information contained within the sovereign bond yields, the ones with longer-term horizons hold multifaceted significance. Firstly, they serve as a criterion for the membership in the Economic and Monetary Union (EMU), as stated in [5], as an EMU candidate's long-term sovereign bond yield should not exceed two percentage points above the average yield rate of the three best-performing EMU member states. Secondly, the private sector frequently uses government bonds as collateral, benchmarks for pricing other fixed-income securities due to the already mentioned characteristic that

they are presumed to be risk-free financial investments. Finally, the significance of the yield rates reflects market perceptions of a country's fundamentals, debt-servicing capacity, and global risk factors, as mentioned before.

In recent decades, there has been a downward trend in long-term government bond yields in developed economies, and in some cases, yields have even turned negative [2], as it can be observed in Figure 1.1. While macroeconomic fundamentals and fiscal accounts are important factors in determining government borrowing costs [6], the relationship between sovereign bond yields and these factors can deviate in the short term, especially during periods of financial stress [6]. Therefore, it is crucial to differentiate between the long-run and short-run determinants of borrowing costs.

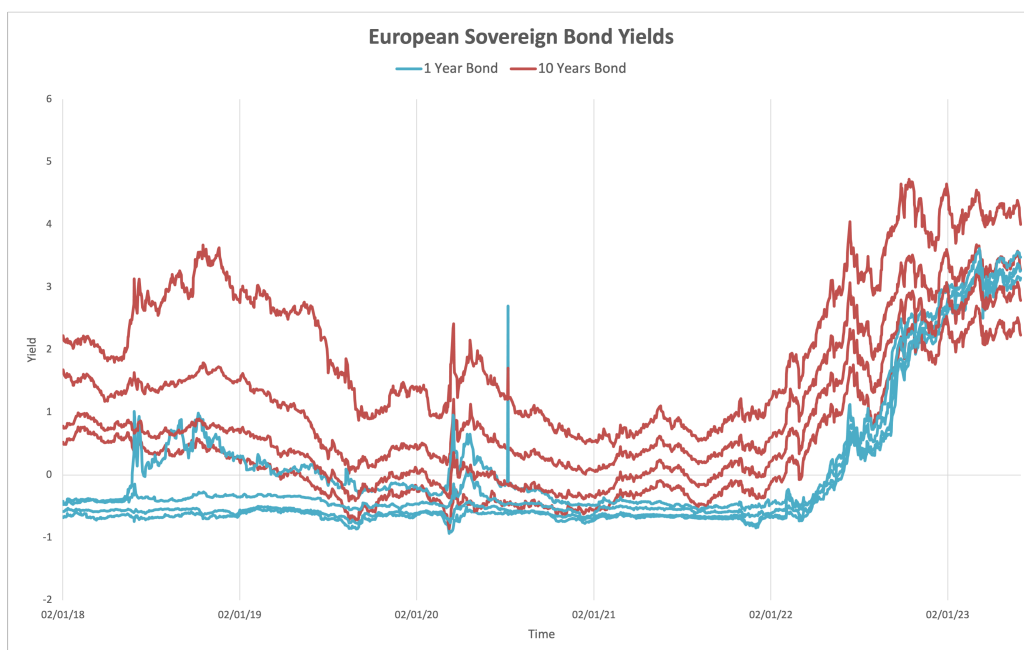


Figure 1.1: Sovereign Bond Yields evolution from January 2018 to June 2023

Bond markets have evolved into an essential and necessary source of funding [3]. Consequently, it is crucial to have an in-depth understanding of the variables affecting sovereign bond rates, given their growing relevance. Sovereign bonds have gained popularity and significance in recent years, especially in Emerging Markets, particularly following the Global Financial Crisis of 2008–09 and the subsequent euro-zone sovereign debt crisis [5]. During these periods, European sovereign bond spreads have increased relatively to corresponding German bonds, which serve as a benchmark. This underscores the importance of studying sovereign bonds during times of political and economic turmoil, such as the one we currently face.

Adding to the complex dynamics of the European bond market, Warin et al. (2023) [5] highlights the challenges faced by the European economy at the end of 2022. The risks arising from the multi-layered crisis, including the ongoing pandemic, accelerating food and energy prices, and a security crisis in Europe, were actively factored into the daily yield rates of sovereign bonds issued by European Union (EU) governments, as we can observe in Figure 1.1.

Before analysing the graphic above, it is significant to consider that the behavior of the yield curve, which can take various shapes over time, provides valuable insights into the dynamics of bond yields [7]. The yield curve can exhibit increasing and concave patterns, with yield dynamics showing persistence and spread dynamics displaying less persistence. Volatility tends to be higher in the short end of the yield curve, while long rates exhibit greater persistence [7]. Even though there are several theoretical characteristics regarding the behaviour of the yield curve, this aspect will be the most relevant for the dissertation. Further information regarding this curve can be found in Kosowski et al.(2015) [8].

In Figure 1.1, the blue colour represents the 1 year bond yields and the red represents the 10 year bond yields. Looking at the graphic we can observe initially a period of more dispersion of the bond yields of the different countries, specially in the 10 year yield. Afterwards, this dispersion shortens but there is an observable peak in all the bond yields correspondent to the beginning of the Covid-19 pandemic. This behaviour becomes more steady, until the escalation of the bond yields in overall, corresponding to the spike in inflation and the beginning of the Ukraine invasion by Russia. As we know, this period also witnessed significant interest rate hikes by central banks across the EU, reflecting a more strict approach of financial markets after reassessing the macroeconomic fundamentals EU member crisis state. It is also observable that the value of the 1 year bond yields even surpass some 10 year bond yields, signalling the inversion of the upward shape of the yield curve in some countries.

Based on this simplified analysis, we may conclude that variables such as international risk factors, credit risk, and liquidity risk spread may be the key drivers to influence changes in government bond rates. These various factors play an important role in explaining yield differentials and highlight the need to consider other sources of risk factors when assessing bond markets [5], such as the price of natural gas.

To better comprehend the influence of natural gas prices on sovereign bond rates, it is necessary to investigate the link between European nations and their use of natural gas, as well as certain crucial historical events in Europe related to this commodity. Examining Europe's energy dependency on natural gas, as well as its connections with gas-producing nations and geopolitical dynamics, can give significant insights into the variables that drive bond market movements. We may acquire a better grasp of how

changes in natural gas prices may impact sovereign bond rates in European nations by taking into account the past history and the specific characteristics of the European natural gas market.

1.2 Natural Gas

Natural gas is a popular energy source that has been gaining importance in recent years. Its widespread use in many sectors makes it a versatile fuel that can meet the energy demands of diverse industries. In 2008, according to the International Energy Agency energy outlook, natural gas was expected to cover 25% of the world's energy needs by 2030, with its consumption set to be double by that time [9].

The natural gas industry has three primary segments, namely upstream, midstream, and downstream. The upstream segment is responsible for the exploration and production of natural gas. The midstream segment is in charge of transporting natural gas to local distribution grids, large-scale industrial users, and power plants. Gas is transported across continents through high-pressure transmission pipelines, and overseas, it is transported in tankers as Liquefied Natural Gas (LNG) [10]. Finally, the downstream segment delivers gas to small domestic and business consumers.

One of the most significant advantages of natural gas is its lower carbon content than coal and oil, making it a favored fuel from an environmental perspective [11]. Natural gas emits 45% less carbon than coal [12] and 25% less carbon than oil [11]. For this main reason the EU taxonomy has recognized natural gas infrastructure investments as sustainable, if they aligned with the EU's decarbonization targets [12].

Natural gas is primarily used for energy generation and heating, but it is also an essential input into production processes in some industries, such as the chemical industry. The chemical industry accounts for the largest portion of industrial demand, with the food industry and the metalwork industry following closely behind [12]. The reliance on natural gas for process heat and the potential for a gas shortage in the future pose significant challenges for the industry sector.

1.2.1 European Union Natural Gas Industry

The Natural Gas industry in the EU has undergone significant changes over the past two decades, with the introduction of competition in the market for natural gas being one of the most notable developments [10]. Despite this, the EU remains a net importer of fossil fuels, including natural gas, and is subject to pressure from third countries in ensuring energy supply security [13].

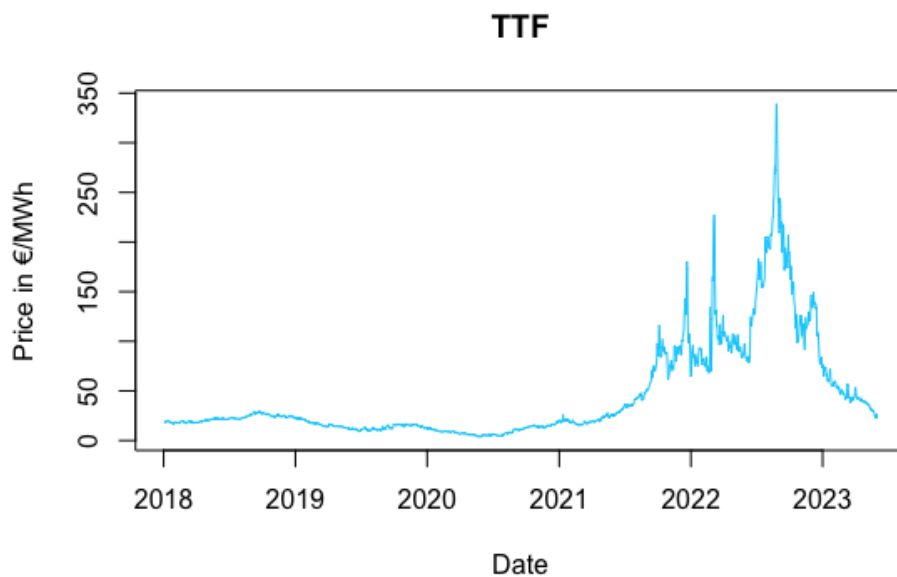


Figure 1.2: Natural Gas Prices (TTF Dutch contract) from January 2018 to June 2023

This issue was brought to the forefront in 2022, when the EU experienced a significant price crisis in the natural gas market, where the price of this commodity surged to historic highs, almost 20 times higher than the average European natural gas price in 2020 [14]. This can be observed in Figure 1.2, and was mainly driven by the supply and demand dynamics of the market, with the second half of 2022 characterized by record levels of European gas prices [12]. The EU's heavy reliance on natural gas imports, combined with the volatility of global energy markets, has highlighted the need for the EU to continue diversifying its energy sources and investing in renewable energy technologies.

To gain a deeper understanding of the current issue surrounding natural gas in the EU, it is crucial to examine the broader characteristics of natural gas trade within the EU, including the developments in demand, production, and supply of this fossil fuel.

1.2.1.A European Union Natural Gas Trade

The gas industry in the EU has undergone significant evolution in terms of how trade has been conducted over time. In the past, public coordination took the shape of direct state intervention, in the form of joint-ventures between gas producers and national/local authorities, between whom the international gas transmission and wholesale trade was managed. This resulted in municipal gas companies to be responsible for managing local distribution networks and retail trade. Traditionally, the European gas market operated through long-term contracts that included take-or-pay and destination clauses, and

where gas prices were closely tied to the prices of oil products [10].

Since the turn of the century, however, the EU has sought to create a more competitive gas market by introducing competition between gas producers and suppliers in the upstream segment and in the wholesale and retail segments. Therefore, as a result, the short-term transactions became favored over traditional long-term contracts, and liquid spot markets have emerged in places where the different sources of supply would connect with the demand of Natural Gas. To make this happen, competing traders have been given access to essential facilities in transport, distribution, and storage to reach their customers [10].

Despite these changes, the EU's gas trade has been subject to significant economic and political shifts in the international gas market, which have forced the EU to adapt its policies to ensure that its gas trade remains secure and sustainable. This was intended to be achieved by the Energy Union policy of 2015, which included components such as energy security (reliable and adequate energy supply at reasonable prices [13]), a fully integrated European energy market, energy efficiency, decarbonization, and research&innovation. The EU's continued adaptation to changing economic and political conditions in the global gas market will likely continue to shape the future evolution of the EU's gas industry [10].

1.2.1.B European Union Natural Gas Demand

As previously mentioned, the natural gas is a significant contributor to various sectors such as electricity production, heating, industry, and transport [9]. At an European level, natural gas accounts for 25% of the primary energy supply. In fact previously, during the 90s and early 2000s, natural gas production in the EU was steady until it reached its peak in 1996, and then started declining in 2004, as it will be explored in the subsection below. Nevertheless, natural gas demand continued to increase, and an example of this phenomenon was the increase of around 1% of the number of registered EU natural gas customers during the period 2006-2007 [9].

The COVID-19 pandemic had a notable impact on the EU's gas industry demand. The subsequent depletion of European underground stored gas stocks, coupled with soaring demand following the lifting of COVID-19-related restrictions, resulted in significant changes in the market [14]. Factors such as reduced renewable energy supplies and the need for natural gas-fired power plants to compensate for intermittent renewable power further contributed to the increased demand for natural gas in Europe [12]. These developments have shaped the recent dynamics of the EU's gas industry demand, reflecting the evolving energy landscape and the challenges faced during the pandemic.

1.2.1.C European Union Natural Gas Production

Another relevant variable to the shift of this commodity's dynamic is the production of conventional gas in Europe, as it has been on a downward trajectory, with an expected decline of 110 billion cubic meter

(bcm) per year (corresponding to 40% of the 2013 production) between 2013 and 2030 [15]. This decline is primarily influenced by the performance of the three main producers in Europe, namely Norway, the Netherlands, and the UK. On the other hand, the prospects for unconventional gas production, including shale gas, tight gas, and coal bed methane, were not promising before 2020, as by 2030 it is projected to contribute less than 20 bcm to the overall production [15]. This decline in domestic production has led to an increased reliance on imported natural gas, as we will see further. However, there is a more optimistic outlook for renewable biogas, which could potentially increase from 14 bcm in 2012 to 50 bcm by 2030 [15]. Nonetheless, challenges related to subsidies introduce uncertainty to these larger figures [15].

Consequently, Europe is expected to only be able to replace approximately half of the decline in conventional gas production with unconventional/renewable sources, and much of this alternative production was projected to not be available until the latter half of the 2020s [15].

1.2.1.D European Union Natural Gas Supply

All in all we can understand from the previous information that the EU heavily depends on imports for its natural gas supply, with 45% of its natural gas being imported [11]. Since 2006 this import dependency was expected to increase due to limited reserves within the EU as mentioned above [11]. In 2001, Russia, Norway, and Algeria accounted for 80% of all gas imports to the EU, with Russia's Gazprom being the largest supplier at 30.7% [11]. Indeed, competition between a variety of suppliers has evolved mainly in northwest Europe, while other regions are still supplied by one single supplier, which is Russia [10]. In 2011, the EU's total net supplies were covered by various sources, with 33% from indigenous production (the Netherlands and UK), 19% from Norway, 24% from Russia, 9% from Algeria, 8% from Qatar, and the remaining 7% from countries such as Nigeria, Trinidad and Tobago, Egypt, Libya, and many others [9]. In fact, in recent years Russia remained the largest supplier of natural gas to the EU, with more than 40% of total imports coming from this country in 2020 [1]. Also in the same year (2020) the EU's import dependency for natural gas from foreign countries was around 84% [1]. It is important to note that the dependence on Russia for overall energy natural gas varies significantly across EU countries [1], as we can observe in Figure 1.3.

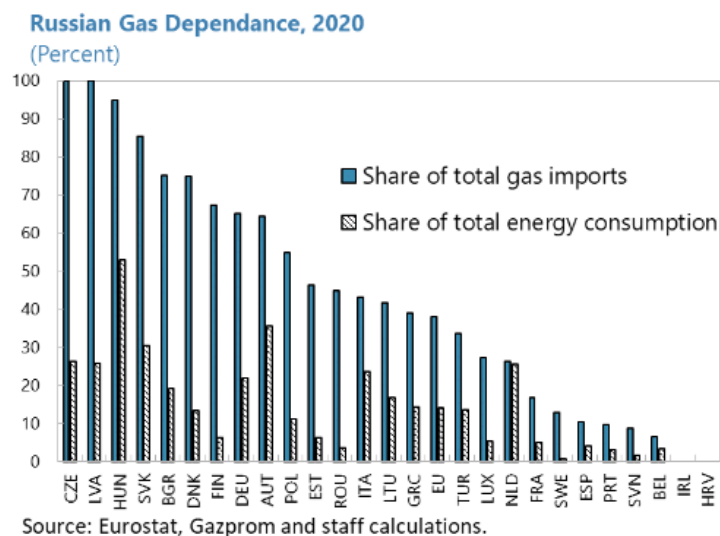


Figure 1.3: Dependence on Russian Gas; Sources: [1]

The gas industry supply distribution in Europe is influenced by various factors and sources. Pipeline gas flows from different regions (including Russia, Norway, the United Kingdom, Northern Africa, and the Caspian region) contribute to Europe’s gas supply [1]. Russian pipelines play a significant role in delivering gas to Europe through routes like Germany, Poland, Ukraine, and Turkey [1]. Norwegian gas enters Europe through countries such as Germany, the Netherlands, Belgium, the UK, and Denmark [1]. Gas from Northern Africa and Azerbaijan also enters Europe through countries like Spain, Italy, Turkey, and Greece [1]. The distribution of gas supply in Europe involves a mix of non-Russian pipelines, Russian pipelines, and LNG import terminals [1]. Currently, non-Russian pipelines account for 30% of total gas import capacity, while Russian pipelines make up 42%, and LNG import terminals contribute another 28 % [1]. This is present in the Figure 1.4.

The EU gas industry has experienced supply disruptions, with different factors contributing to these events. Prior to 2006, disruptions were mainly caused by accidents, terrorism incidents, and monopolistic practices [9]. However, recent disruptions have been influenced by non-European geopolitical conflicts, which have had significant impacts on the European economy [9], as these disruptions have led to spikes in energy prices and raised concerns about the security and reliability of gas supply [1]. In October 2021, it was reported that storage levels, particularly at sites owned or controlled by Gazprom, were low, raising suspicions of attempts to manipulate European energy supplies [14]. This topic will be further explored in the next sub-section.

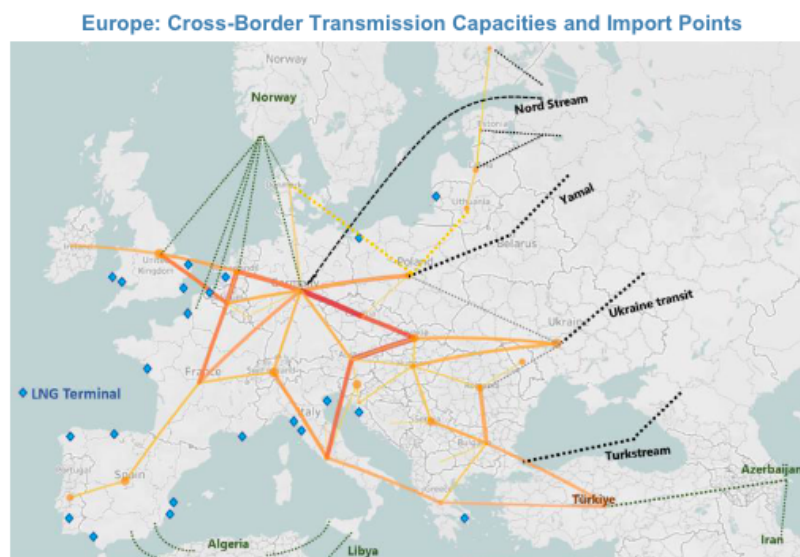


Figure 1.4: Sources: IMF staff based on ENTSOG (System Development Map, 2021, and Transparency platform) [1]

In Figure 1.4 the lines that are thicker and darker (from yellow to orange to red) have higher transmission capabilities. The pipeline directions and flows inside the EU are not displayed in the figure. Pipelines for imports into Europe are depicted by dotted lines. Indicated by dashed yellow lines are pipelines that are anticipated to start operating in the next year.

1.2.2 Russian Natural Gas Crisis

1.2.2.A Previous Relationship

Overall, the previous relationship between the EU and Russia in gas supply has been marked by fluctuations in export volumes, contractual obligations, geopolitical conflicts, and the need for the EU to enhance energy security through various strategic initiatives.

In the 1990s, Russian gas exports to Europe consistently exceeded 100 bcm per year, reaching over 160 bcm in the mid-2000s before declining and recovering to pre-recession levels in 2013 [15]. The majority of these exports were governed by long-term contracts with binding clauses, such as take-or-pay agreements, as seen previously [15].

Disruptions in the European gas supply from 2006 onward were mainly a result of non-European geopolitical conflicts [9]. Notably, instances like Russia's use of natural gas as a political tool in its relationship with Ukraine, led to interruptions in gas flow to Europe in 2006 and 2009 [9]. However, in response to these challenges, the EU intensified its focus on energy security after the 2014 Russian-Ukrainian crisis [13].

1.2.2.B Current Crisis

The current crisis in EU-Russia gas supply involves factors such as reduced availability of LNG, insufficient storage build-up, potential gas shortages, and the influence of geopolitical events.

One contributing factor was the increased demand and transit of LNG to North-East Asia, which has resulted in a reduced availability of LNG for Europe [1]. Insufficient storage build-up by Gazprom and its subsidiary Astora during the summer of 2021 has raised concerns about a potential gas shortage in Europe and led to significant price increases [1]. The failure to ensure adequate storage of gas is noteworthy, the under utilization of pipeline transits and the avoidance of storage buildup may have been influenced by Gazprom's objectives, such as pressuring German authorities regarding the Nord Stream II pipeline and encouraging importers to return to long-term contracts [12].

Another factor was the absence of storage filling during 2021, which was caused by reduced imports from Russia via Ukrainian and Yamal transits, leading to a decrease in deliveries beyond long-term contract obligations [12]. European buyers initially responded by reducing imports of Russian gas and increasing purchases of LNG. However, the rising prices of LNG reversed this trend, resulting in increased usage of the Ukrainian transit after the invasion of Ukraine in February 2022, resulting in record sales of Russian natural gas during this period [12].

As expected, a prolonged full shut-off of Russian gas could lead to gas shortages and substantial GDP losses in the most Russian natural gas dependent countries [15]. However, storage facilities in the EU have seen rapid accumulation since April 2022, reaching 58 % of capacity by the end of June, which contributes to the overall resilience of Europe's gas supply [15].

The future challenges of EU gas supply involve contractual obligations, evolving gas demand, and the need for reliable supply. Currently there are still some European companies that have contractual obligations to import minimum quantities of Russian gas, which limit the scope for reducing deliveries. Failure to fulfill these obligations or terminating contracts prematurely could lead to significant financial obligations through international legal claims [15]. Despite efforts to diversify energy sources, modeling of European gas demand and global gas and LNG supply indicates that Europe will still require a substantial amount of Russian gas in the coming years. Estimates suggest a minimum of 100 bcm and potentially over 200 bcm of Russian gas will be needed in the period leading up to 2030, independent of long-term contracts [15].

To address these challenges, the EU will need to focus on ensuring the reliability and security of gas supply. This includes maintaining infrastructure, such as pipelines and storage facilities, to guarantee uninterrupted flow. Additionally, efforts to diversify the energy mix, promote energy efficiency, and explore renewable energy sources can help reduce dependence on Russian gas in the long run [15].

1.2.3 Countries Characteristics

The selected countries— Germany, Italy, Spain, Hungary, and France— provide a good sample of different natural gas relationships in Europe, as together these countries offer insights into the varying degrees of dependence, the need for diversification, and the influence of energy mixes in shaping natural gas dynamics in Europe.

1.2.3.A Germany

Germany holds a significant position in the natural gas market, characterized by its high dependence on imports from Russia and the challenges it faces in diversifying its sources of supply.

This country heavily relies on Russian gas, with an import share of 55% in 2020, and approximately 46 bcm in 2021 [12]. This dependency on Russian gas exposes Germany to potential disruptions in supply, as seen in instances of reduced deliveries from Russia [1].

One of the factors contributing to Germany's reliance on Russian gas is the historical abundance of pipelines gas from Russia, which has limited the development of domestic LNG import capabilities [12]. Consequently, Germany has explored options to mitigate supply shortages by considering the chartering of Floating Storage Regasification Units (FSRU) to facilitate short-term domestic LNG imports [1].

However, Germany faces challenges in its gas transmission infrastructure, including bottlenecks within the country and limitations in north-south transmission [12]. These limits have an influence not only on Germany's capacity to transport gas from alternative sources through its regional distribution networks but also on the sharing of gas imports within the nation and on imports into Central and Eastern Europe [12].

1.2.3.B Italy

Italy holds a significant position in the natural gas market, and similarly to Germany, is characterized by its reliance on imports, particularly from Russia, and the challenges it faces in diversifying its gas supply. Its energy mix heavily relies on natural gas, which accounts for 40% of the energy mix, with 94% of the gas being imported, and 40% of those imports coming from Russia [12].

Italy, along with Germany, has experienced a significant reduction or cessation of Russian gas deliveries as of early July 2022 [1]. The transmission constraints within the country and some regional distribution systems limit the ability to transport gas from alternative sources, leaving Italy vulnerable [1].

This commodity contributes 59% to electricity production in Italy [12], making it particularly susceptible to the effects of a gas supply disruption [1]. While Italy has greater potential to secure alternative gas supplies compared to some other countries, the ability to replace Russian gas remains limited, with authorities stating that they can only replace around two-fifths of the gas in 2022 [1]. This shortfall, cou-

pled with reduced inventories, would result in significant challenges in meeting Italy's gas consumption needs [1].

In terms of storage facilities, Italy is noted as a country where Gazprom does not own storage facilities [14]. However, data from Gas Infrastructure Europe indicate that gas storage levels in Italy have reached near-normal levels, suggesting a certain level of preparedness [14].

1.2.3.C Spain

Spain's relationship with natural gas is characterized by its role as the EU's largest LNG hub, accounting for over 35 % of its import capacity. LNG imports play a significant role in Spain's gas supply, with gas coming from Northern Africa, Azerbaijan, and other sources through import terminals. However, not all of the stated LNG import capacity is usable, as technical constraints and limited connections to central Europe affect its effectiveness [1].

Being one of the countries not highly reliant on Russian gas and without infrastructure bottlenecks, Spain has a more diversified gas supply compared to countries like Hungary [1]. While Spain has a substantial import capacity for LNG, it can only export a limited portion of it, specifically 10 %, to France. This suggests that Spain primarily serves as an import hub rather than a significant exporter of natural gas [1].

In terms of gas reserves as a share of annual consumption, Spain falls into the category of countries with low reserves but well-connected to international LNG markets, along with the UK, Belgium, and Portugal. Consequently, any inventory build-up or draw-down in Spain would have minimal impact on the rest of Europe [1].

1.2.3.D France

France's position and characteristics in the natural gas market are influenced by its reliance on Russian gas, limitations in gas transmission to neighboring countries, and its lower reliance on natural gas compared to other energy sources as Italy and Germany. The share of natural gas in France's energy mix is only 16%, indicating a lesser dependence on this energy source [12]. Furthermore, the import share of Russian natural gas to France in 2021 was only 17%, with Norwegian natural gas having a higher import share of 36%, followed by Algeria, the Netherlands, and Nigeria [12]. France's significant use of nuclear power, with a share of 69% in electricity production, sets it apart from Germany, which is phasing out nuclear power [12].

While France has some reliance on Russian gas, it also has direct access to LNG import capacity, and alternative pipeline supply routes, providing it with some flexibility and the ability to adjust [1]. However, the transmission of gas from France to most neighboring gas systems faces bottlenecks and

constraints, along with regulatory and technical limitations linked to the odorization of this commodity, which could limit gas sharing within France and potentially constrain gas imports into Central and Eastern Europe [1].

In terms of storage capacity, France does not have storage facilities owned by Gazprom, and data suggests that gas storage levels in France have reached near-normal levels in 2021 [14].

1.2.3.E Hungary

Hungary's relationship with natural gas is characterized by a high reliance on Russian gas, similar to its neighboring countries Czechia and Slovakia. The main alternative supply routes for these countries pass through constrained countries such as Germany, Austria, and Italy, limiting their gas inflows. Consequently, significant shortages are likely to emerge in these regions, resulting in a potential need for extremely high prices to clear the regional gas market [1].

The model suggested by the IMF in [1] suggests that losses would be more significant in countries where infrastructure constraints lead to substantial decreases in gas supply. Sectors with higher natural gas intensity, particularly in these countries, would suffer the most, as expected. The direct impact on gas-intensive industries in these countries is expected to be substantial due to the significant gas shortages anticipated. Furthermore, trade spillovers would exacerbate these effects, resulting in an estimated aggregate GDP loss of around 4 % in the baseline scenario for each country during the year of 2023, show a really significant impact [1].

1.3 Event Analysis

The objective of this section is to demonstrate the patterns exhibited by bond yields in response to significant occurrences in the natural gas sector and other macroeconomic factors. The upcoming Data section 2 will provide an explanation of the methodology employed to obtain the series under analysis (changes in bond yields derived by applying a first differences transformation to the bond yields). This data will be employed to conduct an event analysis, examining the dynamics of bond yield fluctuations observable in a graphical representations. The 1-year yields are represented in orange, while the 10-year yields are depicted in red. However, for the sake of simplicity, only one nation's graphic will be presented in the main text (Germany), with the remaining nations' graphics available in **Appendix A.1**. Nevertheless, the discussion below will be done for all the five countries under study.

The analysis focuses on a specific time frame, spanning from the start of 2022 to June 2023, which covers the significant events associated with the invasion of Ukraine by Russia. The dates that have been examined are as follows:

- **24th of February 2022** - Russia Invades of Ukraine;

- **8th of April 2022** - Release of 5th sanction package aimed to the imported Russian Fuels;
- **3rd of June 2022** - Release of the 6th sanction package aimed to ban Russian fossil fuels;
- **23rd of June 2022** - Germany warns of Lehman Brothers-Like Contagion from Russian Gas Cuts;
- **5th of July 2022** - Natural Gas price soars 700%;
- **29th of August 2022** - European Energy slumps as German Gas stores fill ahead of time;
- **1st of October 2022** - Gazprom Halts Supplies to Italy in Latest Energy battle;
- **15th of March 2023** - Fall of Credit Suisse (included due to the noticeable impact);

In order to enhance identification, the events connected with macroeconomics and politics will be denoted by the color blue, those concerning Germany will be indicated by the color green, and the events associated with natural gas will be represented by the color purple.

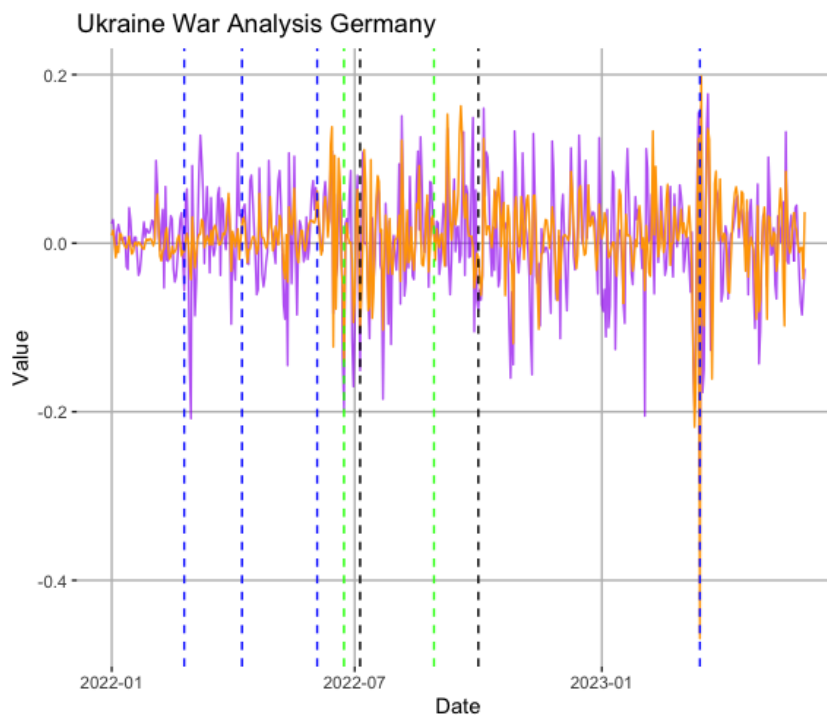


Figure 1.5: Changes in Germany Bond Yields (10 year in red and 1 year in orange)

Upon analyzing the acquired graphic representations, such as the one depicted in Figure 1.5, which showcases the variations in bond yields for Germany, it becomes evident that the onset of the Ukraine conflict had an apparent effect on long-term yields. This pattern is observed across all the countries analyzed, with the exception of Hungary.

The implementation of the fifth sanction packages targeting fossil fuel imports from Russia does not have an immediate and direct impact. However, the sixth package implementation resulted in a minor increase in short-term yield differences across all countries, observed shortly after the announcement date.

Regarding the Germany's alert about the potential spread of a financial crisis similar to the Lehman Brothers event, it is observed that all nations experienced a decrease in bond yields, both in the short and long term, which coincided precisely with this declaration.

Additionally regarding the starting point of the increase in natural gas prices, it can be determined with precision the reaction of the bond yields. It is noteworthy that, in most countries, there is a significant negative peak accompanied by a relatively less prominent positive peak. However, Hungary stands out as an exception, experiencing a minor positive peak in both yield changes.

The anticipated filling of natural gas storage in Germany's does not appear to have an apparent impact on the resulting graphics.

The cessation of Gazprom's supply to Italy has had a noticeable effect on the economic trajectory of the country, as evidenced in **Appendix A.1**. Interestingly, similar impacts have been observed in other countries, with the exception of Hungary. It is worth noting, however, that the magnitude of this impact in Hungary is not particularly noteworthy when compared to the fluctuations experienced in other countries.

Finally, the inclusion of the fall of Credit Suisse bank serves to identify a prominent decline observed across all analyzed countries, highlighting the manner in which these fluctuations are influenced by significant macroeconomic events.

1.4 Literature Review

This literature review aims to synthesize the existing research on the relationship between natural gas prices and sovereign bond yields, drawing insights from relevant studies.

The relationship between different types of variables and sovereign bond yields has been a topic of interest in the literature on financial markets and macroeconomics. Various studies have explored the impact of different macroeconomic factors, including oil prices, exchange rates, and global risk aversion, on bond yields in different countries and regions.

Focusing on commodity prices, several studies have focused on the relationship between oil prices and different financial instruments, providing valuable insights into the dynamics between these variables. For instance, Miller and Ratti (2009) [16] analyzed the long-run relationship between the world price of crude oil and the international stock market, finding a clear connection between the variables for

six OECD countries. Ravischandran (2010) [17] examined the influence of oil prices on the stock market of Gulf Cooperation Council (GCC) countries, concluding that there is a significant long-term effect of oil prices on market returns. Sariannidis et al. (2010) [18] observed a positive correlation between crude oil prices and 10-year bond returns in the U.S. stock market, using a GARCH model. Aloui et al. (2012) [19] explored the impacts of oil price shocks on stock markets in emerging countries, finding a relationship between oil prices and the stock market as a whole.

A more recent example of this type of work was the study of the relationship between oil price changes and emerging market sovereign bond spreads by Chen et al. (2021) [2]. This article finds that oil-specific demand shocks have both statistically and economically significant impacts on emerging market sovereign bond spreads. These findings suggest that financial market speculative activity associated with oil-specific demand shocks may contribute to search-for-yield behavior in emerging market bond markets, leading to reduced sovereign spreads.

In light of the recent turmoil in natural gas prices, particularly due to the war with Ukraine, it is crucial to examine the potential impact of these price fluctuations on sovereign bond yields, especially in the context of the EU. As we can see, while the literature has predominantly focused on oil prices, the specific dynamics of natural gas prices and their implications for sovereign bond yields remain relatively unexplored. However Halser et al. (2023) [12] has examined the economic impact of a curtailment of natural gas flows from Russia on Germany, given its significant import dependency. The study emphasizes the need for import diversification and demand adjustments to avoid a natural gas shortage. While the focus is primarily on the German context, the findings highlight the potential consequences of disruptions in natural gas supply and the importance of addressing import dependency to ensure energy security.

In the context of the impact of other macroeconomic factors, there has been a significant focus on the determinants of sovereign bond yield spreads. Studies by Poghosyan (2012) [6] and Blancard et al. (2019) [20] indicate that macroeconomic fundamentals, government debt levels, and institutional factors play a crucial role in explaining sovereign bond yield spreads.

In a more climate driven macroeconomic variables, another relevant study by Boitan et al. (2021) [21] analyzes the impact of climate change metrics on the cost of government borrowing, expressed as sovereign bond yields and risk premiums, in EU countries. The research reveals that climate-vulnerable countries with lower climate disaster managerial abilities pay higher risk premiums on their sovereign debt. Additionally, it suggests that Mediterranean countries face higher sovereign bond yields compared to Atlantic-region countries, indicating the influence of climate change patterns on borrowing costs within the EU. Moreover, in this same topic Cevik et al. (2022) [22] find that climate change vulnerability and re-

silience significantly affect the cost of government borrowing, highlighting the importance of considering environmental risks in assessing sovereign risk.

Also recently, in the context of the impact of macroeconomic factors on bond yields, Koroleva et al. (2020) [23] investigate the relationship between macroeconomic factors and government bond yields. The study reveals a negative association between the yield of government bonds and gold, while indicators such as inflation, oil prices, and GDP per capita show a positive association with bond yields. These findings highlight the multifaceted nature of macroeconomic determinants and their influence on bond yields. Additionally, Dewachter et al. (2014) [24] contributes to the existing literature by shedding light on the importance of economic fundamentals in explaining sovereign bond yield spreads within the Euro area.

In the fields of overall characteristics of the bond yields within the different members of the EU, Consoli et al. (2020) [25] has proposed an indicator of fragmentation in the sovereign bond market of the Eurozone, which measures divergences in borrowing costs between core and peripheral European countries. The study identifies the absence of fragmentation as periods when bond yields of the two groups share a common stochastic trend in the long run. The indicator of fragmentation is responsive to systemic stress events and negatively related to the European Central Bank (ECB) monetary policy actions, emphasizing the role of ECB interventions in addressing fragmentation concerns.

Finally looking at the most recent developments in incorporating machine learning techniques in the assessment of the impact of different factors along the yield curve, Bianchi (2021) [26] explores the predictive power of machine learning methods, specifically extreme trees and neural networks (NNs), in the context of bond return predictability. Overall, the study highlights the efficacy of machine learning methods in capturing bond return predictability and emphasizes the varying influence of different factors along the yield curve.

Following this, in the light of this recent machine learning incorporation tendency, several studies use Least Absolute Selection and Shrinkage Operator (LASSO) method to examine the impact of specific variables on bond yields. For example, Bredin et al. (2021) [27] employed the LASSO method to predict WTI oil market returns by utilizing Nelson-Siegel (NS) factors derived from the term structure of WTI oil futures. The LASSO approach allowed the researchers to select the most informative factors while considering a wide range of macroeconomic indicators and oil market-specific predictors. The researchers chose LASSO to address the challenge of selecting relevant variables from a wide range of predictors, resulting in a more comprehensive understanding of the relationship between oil market returns and bond yields. Ciner et al.(2022) [28] utilized the Lasso penalty in a quantile regression framework to investigate the determinants of cryptocurrency returns during the volatile period of COVID-

19. By employing the LASSO method, the researchers aimed to select the most influential variables among a relatively large set of predictors. However much earlier Huang et al. (2009) [29] investigated the forecasting of bond risk premia by identifying macro factors with significant predictive power. To achieve this, the researchers employed the LASSO approach to select relevant variables and improve forecasting accuracy.

Moreover, there is also a new tendency to assess the impact of a given variable using Shapley Values. Several studies have employed Shapley values to examine different aspects of financial markets. For example, Wang (2021) [30] develops a risk meter called FRM@China, which utilizes a quantile-lasso regression model to assess risk transfer between different sectors and predict systemic risk. To understand the contribution of macro features in this "black box" situation, the study expands the FRM approach by incorporating Shapley values. The findings highlight the significance of short-term interest rates and forward guidance as drivers of risk. Additionally, Consolli (2022) [25] utilizes economic news within a neural network framework to forecast the Italian 10-year interest rate spread. By incorporating Shapley values, the study identifies the most impactful features influencing the model's output. The use of Shapley values provides insights into the importance of different variables and their contributions to the forecasting performance. Finally, Cerutti (2021) [31] explores the association between macrofinancial factors and the evolution of covered interest parity (CIP) deviations in the post-financial crisis period. The study employs the concept of Shapley values to compare the relative strength of risk-related factors and monetary policy-related factors in explaining the dollar basis.

These findings provide a foundation for further research on the impact of natural gas prices on sovereign bond yields in the EU. Future studies could explore the specific channels through which natural gas price fluctuations affect bond yields, considering factors such as energy security, import dependency, and the role of natural gas in the energy mix. Additionally, examining the interaction between natural gas prices and other macroeconomic variables in the EU context would contribute to a more comprehensive understanding of the relationship between natural gas prices and sovereign bond yields.

1.5 Positioning and Research Question

The research landscape regarding the impact of natural gas prices on sovereign bond yields is relatively short, with limited studies addressing this specific relationship, mostly focusing on the volumes of natural gas and not specifically on prices. Despite the importance of both energy markets and government debt markets, there is a noticeable gap in understanding how fluctuations in natural gas prices can influence sovereign bond yields. Furthermore, while existing research has explored the relationship between various macroeconomic variables and bond yields, there is a lack of studies that have employed both

Lasso and Shapley values as complementary methods to assess the impact of variables on sovereign bond yields. By adopting this different and quite simplistic approach, the study aims to provide a more comprehensive understanding of the relationship between natural gas prices and sovereign bond yields, throughout the time span between 2018 and 2023 (characterized by a great macro-financial and political volatility), shedding light on this specific impact which is a contemporary and current topic.

Research Question : *Does the Natural Gas Price impacts Sovereign Bond Yields in the EU, and how as it been evolving in recent years?*

1.6 Organization of the Document

The present thesis is structured in the following manner: In Chapter 1, the problem is defined, accompanied by a comprehensive analysis of the current natural gas situation in the European Union (EU). Additionally, a thorough literature review of prior research in the field is presented, and the dissertation is situated by articulating the main research question.

In Chapter 2, the data will be thoroughly described and examined, highlighting its key characteristics prior to its utilization in the analysis. Furthermore, Chapter 3 will provide a comprehensive theoretical and practical description of the methodology and approaches employed.

Subsequently, Chapter 4 will present and analyze the results that have been obtained, drawing preliminary conclusions. Chapter 5 will thoroughly examine the validity of the findings mentioned previously.

Finally, Chapter 6 will provide a comprehensive summary of the discussions that have been conducted throughout the dissertation, provide some suggestions on the future work to develop and will serve as the concluding section of the work.

2

Data Analysis

Contents

2.1	Data description	25
2.2	Stationarity tests and Transformations	28
2.3	Correlation	35

This section will discuss the approach that was employed for data selection, including the characteristics of the data, the necessary transformations, and the final inputted data for subsequent analysis in this dissertation.

2.1 Data description

Initially the main idea was to follow and use the same variables as used by Dewachter et al. (2014) [24] since it uses a series of very different macro factors and variables that can be seen as "fundamental" to the spreads of bond yields. However, the frequency of this data is monthly, and the main target of the present thesis is to assess the impact of a commodity price variable in a period of high volatility, that has suffered great daily fluctuations. For this main reason I opted to work with high frequency time series data (in daily frequency). This led to the elimination of the GDP, debt and inflation variables specific to each country analysed, as well as the European Commission's European Sentiment Indicator (ESI).

Having this in mind, this dissertation will use Time Series in **daily frequency** over the period **from the beginning of January of 2018 to the beginning of June of 2023**. The countries under study, as mentioned before, are Germany, Italy, Spain, France and Hungary. The retrieved data for the explanatory variables is the following:

- The Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX), obtained from CBOE Exchange, which expresses the existing volatility of the Standard & Poor's (S&P) 500 stock market index options, and is used as a measure of global Financial volatility Financial markets;
- The European Economic Policy Uncertainty Index (POL), which measures economic uncertainty related to policies and therefore is produced by Economic Policy Uncertainty. The construction of this index, suggested by Baker et al.(2013) [32], is based on newspaper coverage of policy-related economic uncertainty and disagreements among economic forecasters regarding upcoming economic indicators for France, Germany, Italy, Spain, and the United Kingdom. This index is widely used as a broad indicator of political risk in the euro area;
- The Flight to Safety (F2S) which measures the liquidity premium that can be interpreted as a common liquidity factor across the euro area bond market, and that is calculated as a spread between the yield on the German government-guaranteed Kreditanstalt fur Wiederaufbau (KfW) bond (which is a German government-owned development bank) and the German sovereign bond. Both series were retrieved from Bloomberg and averaged across maturities, following the same methodology as described in Dewachter et al. (2014) [24];

- The Crude Oil price (OIL), represented by Brent Crude Oil Future contract price retrieved from the Commodities market through the website [33], and it is given in dollars per barrel;
- The Natural Gas price, represented by the Dutch Title Transfer Facility (TTF) Gas Future contract price as this is the key European benchmark price [1]. The values are given in Euro/MWh and were retrieved from the Commodities market through the following website [34];

The raw graphical representations of the data can be found in **Appendix A.2**

In terms of the dependent variables, the daily value for zero coupon bond yields of maturities 1 and 10-years were retrieved for each of the analysed countries. In addition, the Overnight Indexed Swaps (OIS) was acquired, which is related to the inter-bank interest rate in Europe known as the Euro Overnight Index Average (EONIA). This will be represented by the Euro Short-Term Rate (ESTR) series, which reflects the wholesale borrowing costs of Euro area banks for maturities of 1 and 10 years, as obtained from Bloomberg (EESWE1 Curncy and EESWE10 Curncy). This data was utilized as a benchmark variable to compute the spread for each country across different maturities. The objective of this study is to examine the utilization of the 1 and 10-year sovereign yield spread as an indicator of sovereign risk. **Table 2.2** presents a concise overview of the data obtained for the current study.

Table 2.1: Retrieved Variables Description - Part 1

Factor	Description	Frequency	Source
VIX	Chicago Board Options Exchange (CBOE) Market Volatility Index	Daily Working Days (DWD)	Cboe Exchange [35]
POL	European Economic Policy Uncertainty Index	Daily	Economic Policy Uncertainty [36]
F2S	Flight to Safety, a liquidity factor across the euro area bond market	DWD	Bloomberg (TckrKfW: <i>BI665</i> ; TckrBGe : <i>I0160</i>)
OIL	Brent Crude Oil Future contract price	DWD	Commodity Market (Tckr : <i>LCOU3</i>)
TTF	Dutch TTF Gas Future contract	DWD	Commodity Market (Tckr : <i>TFAc1</i>)

Table 2.2: Retrieved Variables Description - Part 2

Factor	Description	Frequency	Source
OIS	Overnight Indexed Swap or Euro short-term rate	DWD	Bloomberg (Tckr : EESWE)
BGe(1/10)	Germany Zero-coupon yield for government bonds for maturities of 1 and 10 years	DWD	Bloomberg (Tckr : I0160)
BIt(1/10)	Italy Zero-coupon yield for government bonds for maturities of 1 and 10 years	DWD	Bloomberg (Tckr : I0400)
BSp(1/10)	Spain Zero-coupon yield for government bonds for maturities of 1 and 10 years	DWD	Bloomberg (Tckr : I0610)
BFr(1/10)	France Zero-coupon yield for government bonds for maturities of 1 and 10 years	DWD	Bloomberg (Tckr : I0140)
BHn(1/10)	Hungary Zero-coupon yield for government bonds for maturities of 1 and 10 years	DWD	Bloomberg (Tckr : I0142)

Since the majority of the data was expressed in daily working days, except the POL that was fully daily, I opted to perform an interpolation on the missing values using the spline method. This method mathematically reproduces flexible shapes, allowing the identification the points where adjacent functional pieces join. This is based usually on low-order polynomials, which are chosen to fit the data between two consecutive knots [37].

2.1.1 Statistical Description

Next the main goal was to calculate descriptive statistics to characterize this interpolated data, namely the mean (*Mean*), median (*Median*), standard deviation (*SD*), minimum value (*Min*), maximum value (*Max*), skewness (*Skewness*) and kurtosis (*Kurtosis*). For these last two statistics, the sample skewness was obtained by dividing the third standardized moment by the cube of the standard deviation, which is commonly used to assess the asymmetry of a data-set. For the kurtosis I calculated the raw one, which measures the kurtosis without subtracting 3 units, which if subtracted corresponds to the deviation from a normal distribution. The overall results are presented in **Table 2.3**.

Table 2.3: Statistics Table

Variable	Mean	Median	SD	Min	Max	Skewness	Kurtosis
VIX	21.108	19.630	8.095	9.122	82.690	2.362	10.325
POL	156.528	122.720	113.322	4.050	861.100	2.197	6.070
F2S	0.254	0.219	0.111	0.104	0.651	1.034	0.372
OIL	70.653	70.118	19.422	19.330	127.980	0.167	0.309
TTF	45.073	21.810	50.930	3.510	339.195	2.073	4.610
OIS1	0.038	-0.475	1.229	-0.775	3.771	1.981	2.324
OIS10	0.556	0.199	1.022	-0.638	3.105	1.132	0.082
BGe1	-0.163	-0.643	1.111	-0.932	3.336	2.080	2.760
BGe10	0.304	-0.050	0.904	-0.859	2.697	1.154	0.195
Bl1	0.326	-0.150	1.116	-0.567	3.616	1.737	1.769
Bl10	2.105	1.881	1.203	0.455	4.740	0.450	-1.033
BSp1	0.044	-0.422	1.104	-0.684	3.351	1.994	2.397
BSp10	1.208	0.927	1.025	0.010	3.678	0.937	-0.253
BFr1	-0.097	-0.576	1.128	-0.847	3.410	2.065	2.692
BFr10	0.669	0.318	0.980	-0.415	3.201	1.175	0.210
BHn1	3.005	0.723	4.340	0.001	13.751	1.454	0.506
BHn10	3.910	2.967	2.193	1.527	9.870	1.155	-0.104

2.2 Stationarity tests and Transformations

The examination of stationarity is a crucial aspect when analyzing a Time Series. However, it is essential to be cautiousness when categorizing a Time Series as stationary, as it can fall under either the category of strictly stationary or weakly stationary process. Both concepts involve the examination of the mean and auto-covariance of the series, where the first applies to the mean of the probability distribution of the series, while the second refers to the covariance between X_t and its own lagged value.

To argue that the time series exhibits strict stationarity, it is presumed that the joint distribution of

$$(X_t, X_{t+j_1}, X_{t+j_2} + \dots + X_{t+j_n})$$

remains unchanged for any given value of

$$j_1, j_2, \dots, j_n$$

As a consequence the mean and the auto-covariance of the series remain constant over time. The primary challenge in ensuring strict stationarity of a time series lies in the precise distinction of its dis-

tribution. Due to its stochastic nature, determining the distribution of the process presents a significant challenge, if not an impossible task.

For this reason, most literature considers sufficient to only test whether both the mean and the autocovariance of the series are kept constant over time, in another words, test for covariance-stationarity, also known as weak stationarity. Based on this information, from this point onward, when the word stationarity is mentioned it refers to covariance-stationarity.

To assess the stationarity of the variables three different tests were performed: the Augmented Dickey Fuller (ADF), the Phillips-Perron (PP), and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests for unit roots.

Dickey and Fuller proposed a formal test for stationarity, emphasizing that testing for non-stationarity is equivalent to examining the presence of a unit root [38]. However, there has been criticism regarding the low power of the Dickey-Fuller test, only having a successful identification on the order of 30%. Consequently, the test is not considered appropriate, and many other tests for unit root suffer from low power as well. To address the limitations of the Dickey-Fuller test, the Augmented Dickey-Fuller (ADF) test was developed. The ADF test is an extension of the simple Dickey-Fuller test and accounts for the presence of autocorrelation in the error term. By including additional lagged terms of the dependent variables, the ADF test aims to eliminate the problem of autocorrelation. As a result, the Augmented Dickey-Fuller test is typically used instead of the simple Dickey-Fuller test [38]. Similarly the Phillips-Perron (PP) test also tries to tackle the autocorrelation issue, however by fully ignoring any serial correlation. Both these stationarity tests the null hypothesis of the existence of unit root in the time series.

On the other hand, the KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test is designed to complement the two previous unit root tests. It serves to test both the unit root hypothesis and the stationarity hypothesis, allowing us to differentiate between series that appear stationary, series that seem to have a unit root, and series for which the data or tests do not provide sufficient information to determine their stationarity or integration status [39]. One of the main advancements introduced by the KPSS test is the consideration of error autocorrelation. However, estimating the long-run variance poses a practical challenge in conducting the test. The method used to correct for autocorrelation in the KPSS test is similar to the Phillips-Perron corrections employed in unit root tests [39]. In this case the tested null hypothesis is that the data is stationary.

2.2.1 Explanatory Variables

The stationarity test results are presented in Table 2.4, containing first the value of the statistic and the correspondent p-value for each one of the tests described above.

Observing the values we can conclude that the KPSS test results indicate that neither variables are

Table 2.4: Explanatory Variables Stationary Tests Results Table

Variable	PP_Stat	PP_P_Val	ADF_Stat	ADF_P_Val	KPSS_Stat	KPSS_P_Val
VIX	-4.80	0.01	-4.38	0.01	2.57	0.01
POL	-16.75	0.01	-3.31	0.07	2.18	0.01
F2S	-2.98	0.17	-2.48	0.37	7.45	0.01
OIL	-1.96	0.59	-1.73	0.69	7.59	0.01
TTF	-2.66	0.30	-2.24	0.48	10.55	0.01

stationary, even if sometimes the PP and ADF support stationary series (for instance the case of VIX). Being aware of the characteristics of these tests, I will only consider a variable to be stationary if the three tests provide the same conclusions.

Following usual procedures, transformations of the variables were applied. In order to convert the VIX, the **First Differences** transformation was employed, as demonstrated in the study conducted by Barway et al. (2019) [40]. The chosen transformation for both the POL and F2S variables was the **Difference in natural logarithms** or **Growth Rate**. The rationale for selecting this type of transformation was based on the fact that the POL variable exhibited elevated values, given its nature as an index, and conversely, the fact that this transformation would provide to the F2S variable coefficients more interperability. In line with common practice in the literature, the growth rate was employed as the transformation for both commodity prices, namely OIL and TTF. The resulting variables from these transformations are graphically represented as follows:

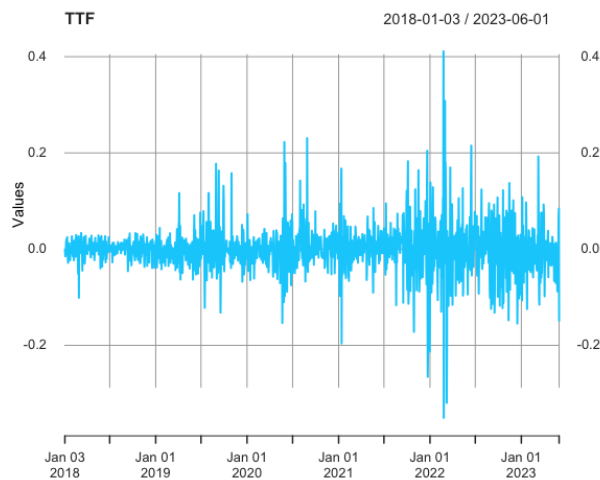


Figure 2.1: Graphical representation TTF

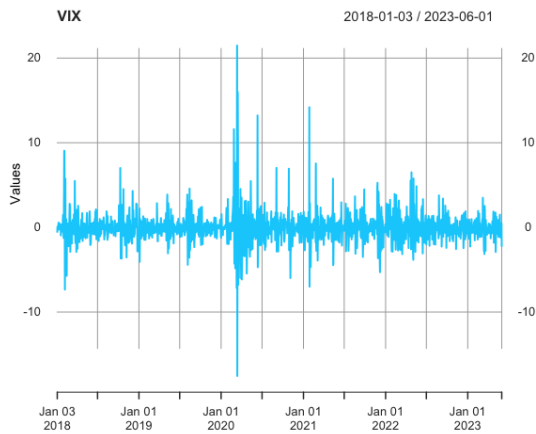


Figure 2.2: Graphical representation VIX

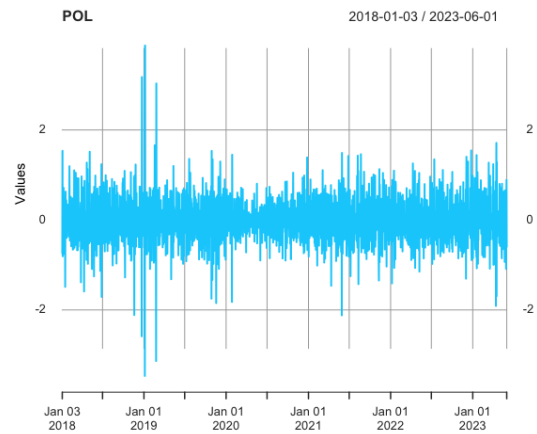


Figure 2.3: Graphical representation POL

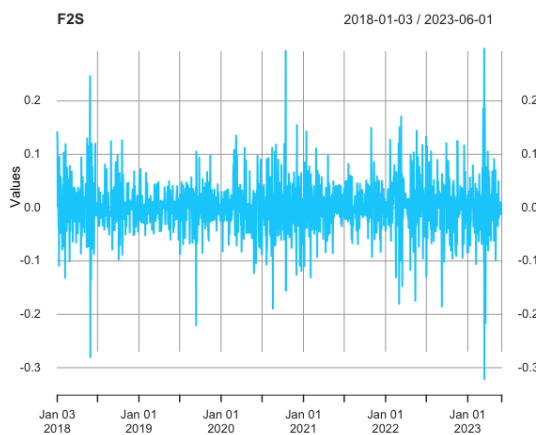


Figure 2.4: Graphical representation F2S

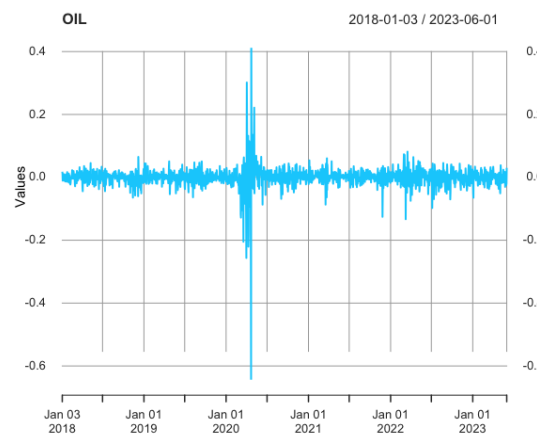


Figure 2.5: Graphical representation OIL

Reassessing stationarity, Table 2.5 displays the findings related to this of the transformed variables. Contrary to the original variables, we make the assumption of stationarity. Therefore, it can be inferred that the usual transformations employed were effective in addressing the initial problem of non-stationarity.

Table 2.5: Transformed Explanatory Variables Stationary Tests Results Table

Variable ¹	PP_Stat	PP_P_Val	ADF_Stat	ADF_P_Val	KPSS_Stat	KPSS_P_Val
VIX	-39.84	0.01	-12.32	0.01	0.03	0.1
POL	-106.54	0.01	-19.91	0.01	0.00	0.1
F2S	-39.13	0.01	-12.53	0.01	0.05	0.1
TTF	-30.50	0.01	-13.02	0.01	0.20	0.1
OIL	-37.36	0.01	-10.62	0.01	0.07	0.1

2.2.2 Dependent Variables

For the the case of the dependent variables to test, I have opted to consider two different types of variables, the first type being the Changes in Bond Yields series, which could be obtained by the simple First Differences transformation to the retrieved Bond Yields (they will assume the acronym of the B(Country)(Maturity) from this point onward). The second type is the Bond Yield Spreads, which, as mentioned before, would be a simple subtraction of the correspondent maturity OIS to the retrieved Bond Yields (they will assume the acronym of the SpB(Country)(Maturity) from this point onward).

After the construction of the dependent variables, the results of the stationarity tests were the following, represented in Table 2.6.

Table 2.6: Dependent Variables Stationary Tests

Variable	PP_Statistic	PP_P_Value	ADF_Statistic	ADF_P_Value	KPSS_Statistic	KPSS_P_Value
BGe1	-37.24	0.01	-11.40	0.01	1.39	0.01
BGe10	-35.50	0.01	-12.51	0.01	0.53	0.04
Bl1	-70.78	0.01	-13.37	0.01	0.39	0.08
Bl10	-37.88	0.01	-12.31	0.01	0.30	0.10
BSp1	-35.98	0.01	-11.63	0.01	1.48	0.01
BSp10	-34.54	0.01	-12.80	0.01	0.57	0.03
BFr1	-37.46	0.01	-11.08	0.01	1.55	0.01
BFr10	-35.12	0.01	-12.51	0.01	0.54	0.03
BHn1	-35.33	0.01	-11.56	0.01	0.97	0.01
BHn10	-32.14	0.01	-13.20	0.01	0.15	0.10
SpBGe1	-3.29	0.07	-3.06	0.13	7.86	0.01
SpBGe10	-2.73	0.27	-2.09	0.54	10.39	0.01
SpBl1	-6.92	0.01	-4.02	0.01	11.63	0.01
SpBl10	-2.45	0.39	-2.41	0.40	8.23	0.01
SpBSp1	-4.18	0.01	-3.51	0.04	11.95	0.01
SpBSp10	-3.85	0.02	-3.23	0.08	7.18	0.01
SpBFr1	-3.79	0.02	-3.46	0.05	8.80	0.01
SpBFr10	-3.99	0.01	-3.08	0.12	3.63	0.01
SpBHn1	-1.39	0.84	-1.47	0.80	16.03	0.01
SpBHn10	-2.23	0.48	-1.69	0.71	15.06	0.01

The results show that for the Changes in Yields variables (upper part of the table), the ADF and PP tests lead to accept the stationarity, while the KPSS test sets the majority of them as being non stationary. In contrary, the results for the Spreads (lower part of the table) show that while some series are stationary for the PP and ADF test, none of them is stationary according to the KPSS test.

¹To ease clarity in presentation, the variables will retain their "original" names, although it should be noted that they will henceforth refer to the transformed variables. The same approach will be utilized for the Figures.

Since there were no additional transformations to the initial sought dependent variables, their resulting graphical presentation is the following, where the line in orange represents the short-term 1-year bond equivalent and the red line the long-term correspondent:

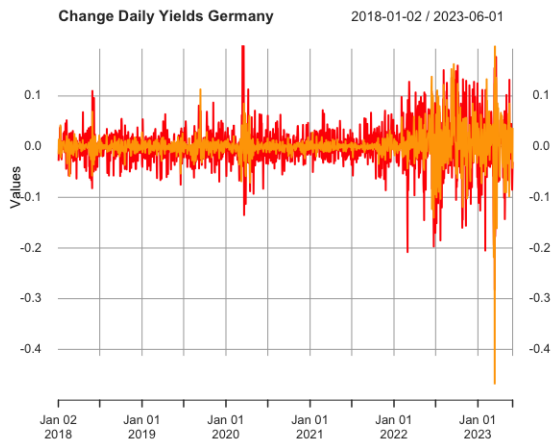


Figure 2.6: Graphical representation Change in Germany Yields

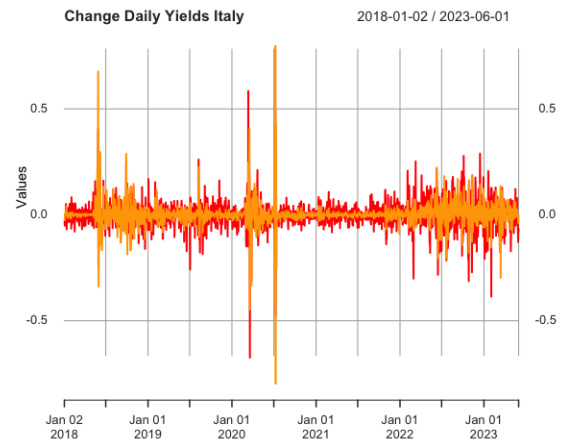


Figure 2.7: Graphical representation Change in Italy Yields

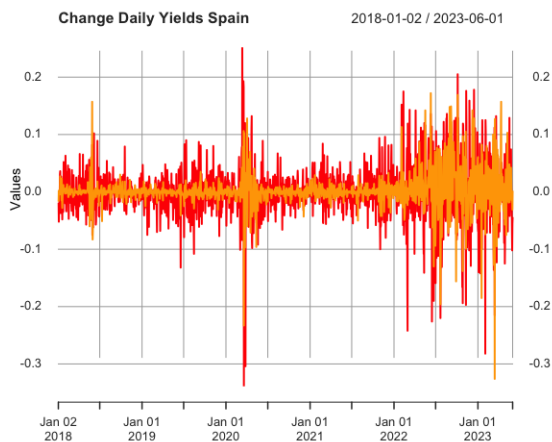


Figure 2.8: Graphical representation Change in Spain Yields

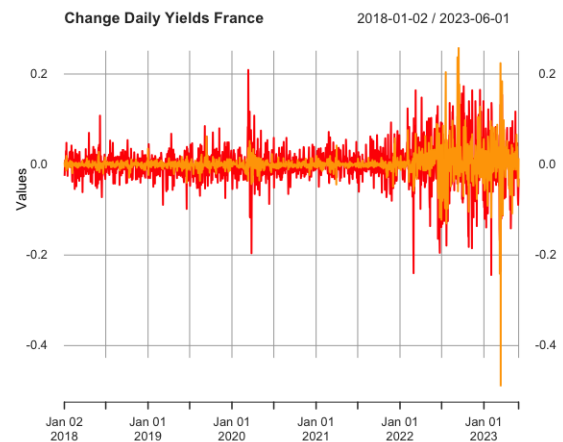


Figure 2.9: Graphical representation Change in France Yields

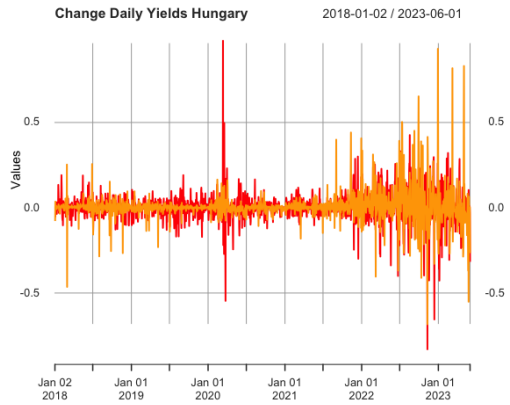


Figure 2.10: Graphical representation Change in Hungary Yields

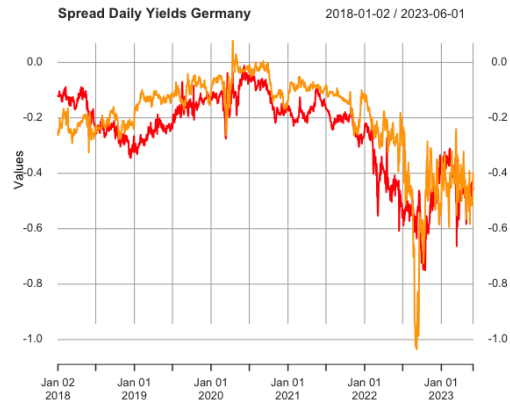


Figure 2.11: Graphical representation Spread in Germany Yields

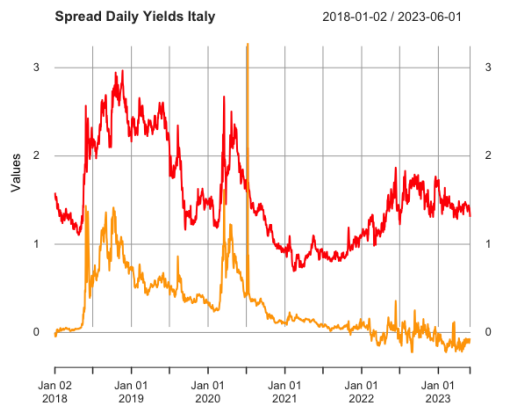


Figure 2.12: Graphical representation Spread in Italy Yields

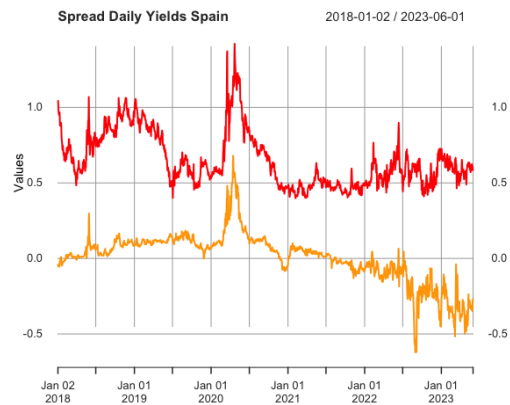


Figure 2.13: Graphical representation Spread in Spain Yields

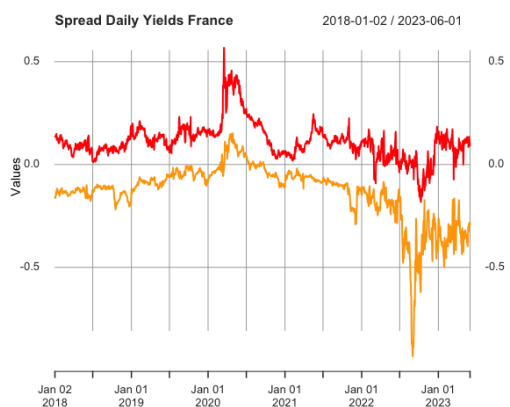


Figure 2.14: Graphical representation Spread in France Yields

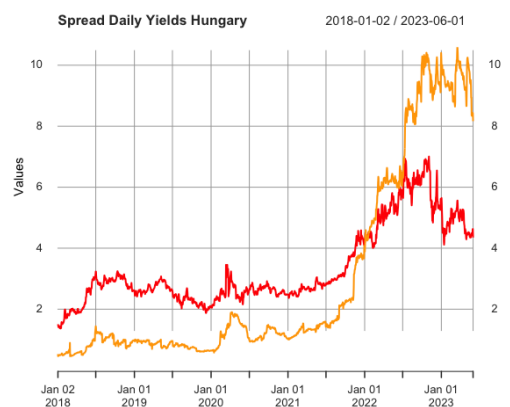


Figure 2.15: Graphical representation Spread in Hungary Yields

From looking at the graphics we can observe that in terms of Spreads, the dimension of the spreads really changes from country to country, being specially elevated in Italy and Hungary. This is also observable in the Changes in yields however the scale difference is not that significant.

2.3 Correlation

To conclude the present Chapter, there was an assessment of whether the obtained final variables are significantly correlated. The matrix that represents visually the existing correlation is the presented in Figure 2.16.

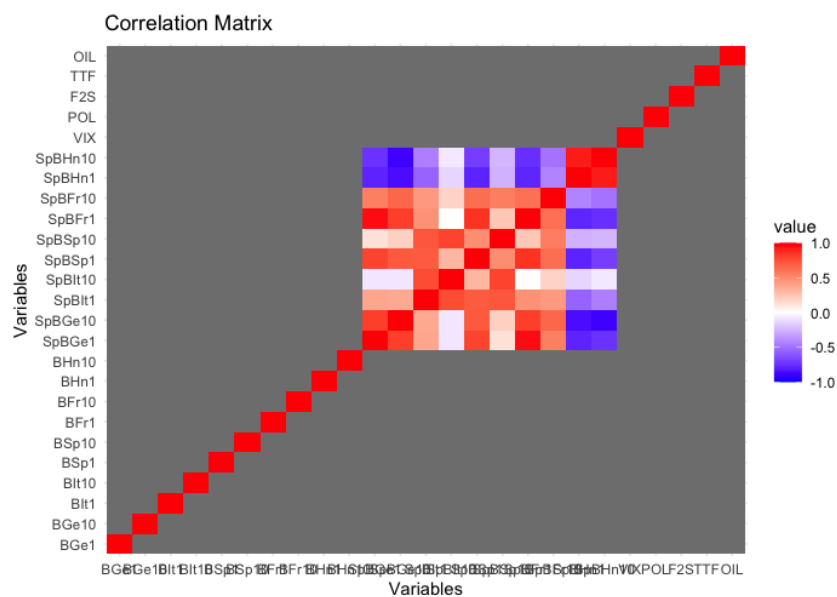


Figure 2.16: Correlation matrix of the variables

It is observable that the majority of the correlations are not statistically significant (represented in grey), being the only correlated variables the Spreads, which was expected considering the way these variables were significant. This leads to the conclusion that the calculated data does not present correlation, which will be a relevant characteristic in the following research.

3

Methodology

Contents

3.1 Ordinary Least Squares Regression	39
3.2 Least Absolute Selection and Shrinkage Operator	40
3.3 Shapley Values	42
3.4 Periodical Analysis	43

This section aims to describe the methods that will be employed to reach a conclusion to the postulated research question. The main objective was to adopt a very parsimonious and simplistic approach by adopting and adapting statistical and machine learning methods, as it will be further described.

A very relevant note is that I will initially proceed with the assumption that the necessary statistical conditions for model validation are satisfied, as I found it often to be assumed in the literature of this field. However, it is important to note that addressing and verifying these assumptions will be a priority at a later stage of the study.

3.1 Ordinary Least Squares Regression

In conducting a thesis on the impact of natural gas prices on sovereign bond yields, I will firstly apply one popular and commonly used statistical method which is the Ordinary Least Squares (OLS) regression. OLS regression is a widely used method due to its simplicity and interpretability, as it aims to measure the effects of the explanatory variables on the dependent variable, by estimating the coefficients of variables that linearly affect the dependent variable. The estimation technique is based on minimizing the residual sum of squares (RSS), which involves finding the values of the parameters that best fit the observed sample [41].

Mathematically this method is given by:

$$Y = X'\beta + u \quad (3.1)$$

where Y is a vector containing the dependent variable, X is a matrix containing the independent variables to test, β is a vector containing the determined coefficients and u is the vector containing the error term [41]. The solution of the minimization of the RSS that leads to the estimate of β , hereby denoted by $\hat{\beta}_{OLS}$, is given by:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (3.2)$$

Under certain assumptions, an advantage of OLS is that the coefficient estimates obtained from this method are unbiased. OLS estimates are linear functions of the dependent variable, and they have the smallest variance among all linear unbiased estimators within the class of linear estimators [41] [42], and therefore there is an underlying assumption that the data follows a linear relationship.

In terms of more practical implementation of this method, the OLS was obtained by training the model in ten different subsets of the full data-set, being therefore able to select the coefficients that would minimize the overall mean sample error. This is known as cross-validation and is very common in

machine learning. This can be synthesised by the following steps:

- Split the data-set into 10 approximately equal-sized subsets or folds.
- Select the first fold as the validation set and the remaining nine folds as the training set. Fit the OLS model using the training data.
- Use the OLS model to predict the target variable for the instances in the validation fold.
- Compare the predicted values to the actual values in the validation fold and calculate the performance metric of interest in this case the mean squared error.
- Repeat in total 10 times the same procedure, each time selecting a different fold as the validation set.
- Compute the average performance metric across the 10 folds to obtain an overall assessment of the model's performance.
- Choose the set of coefficients that best perform on average.

The main reason why this approach was adopted is mainly due to the characteristic of achieving a more representative estimate of the underlying relationship between the studied variables, as it tests different subsets and selects the one that performs best on the validation set, and the plus that this allows to control in a certain way over-fitting of the model, a quite usually attributed characteristic to the OLS.

3.2 Least Absolute Selection and Shrinkage Operator

The research then uses LASSO regression, a penalized regression approach introduced by Tibshirani in 1996 [29]. The LASSO method aims to construct more efficient models by removing redundant variables through the imposition of a penalty in the objective function, facilitating variable selection. It should be noted that it is more of a data mining technique than an econometric tool. The consistency of LASSO variable selection relies on certain nontrivial conditions that may fail in large-scale macroeconomic forecasting scenarios.

One advantage of using LASSO regression is its ability to handle a larger number of potential predictors while still achieving variable selection. The loss function of the LASSO regression consists of the sum of squared errors, similar to OLS regression, but with the addition of a penalty term. This penalty term is determined by a threshold parameter, estimated using cross-validation, and controls the amount of shrinking applied to the coefficients, as described by the following expression:

$$\hat{\beta}_{Lasso} = \arg \min_{\beta} RSS(\beta) + \lambda \sum_{i=1}^N |\beta_i| \quad (3.3)$$

Generally LASSO regression offers three key advantages over traditional OLS estimation. First, it addresses the issue of forecasting accuracy by reducing the variance of predicted values without sacrificing excessive bias. Second, it enhances the interpretability of the model by selecting a smaller subset of predictors that exhibit the strongest effects, providing a clearer understanding of the underlying determinants. Thirdly, LASSO regression continuously shrinks coefficients toward zero as the regularization parameter (λ) increases, automatically setting small estimated coefficients to zero and reducing model complexity [29]. Finally the dynamic feature selection capability of the LASSO is another advantageous aspect. As the estimation window expands, the LASSO selects variables that contribute successfully to forecasting within that specific window. This provides insights into variable importance over time, as the selected variables by the LASSO may change. [27].

Mostly this last advantage and characteristic of LASSO will be the focus of the incorporation of this method in the present research. In fact the main idea behind the inclusion of this method is to really assess firstly whether LASSO selects the natural gas Dutch TTF variable as relevant . This was done by performing the cross validation of the λ hyper-parameter and check the state of the TTF coefficient. This cross validation followed a similar chronology as the one described for the OLS, with the following differences:

- There was previously defined a grid of λ values to be tested.
- The steps of the OLS were repeated for each value of the λ parameter, only this time applying the OLS with the above LASSO restriction.
- The λ value that was choosen corresponds to the best performance metric, in this case the λ that minimizes the average out-of-sample variance.
- The LASSO model was retrained using the selected λ value on the entire training set to ensures that the final model benefits from the maximum amount of training data.

Then the second application of the LASSO, and in a way more insightful to the overall analysis, was to change the restriction hyper-parameter (λ) value, from a more loosen to a stricter value (whithin the values defined for the cross-validation grid), and observe the graphical evolution of the attributed TTF coefficient, compared to the other coefficients. This was achievable by storing the resulting coefficients for each tested λ and plot these vectors in a graphical representation.

3.3 Shapley Values

I then move to the application of Shapley values, which is a fairly new machine learning technique that provides insights into the marginal effects of different factors in a cooperative game or coalition [30].

The Shapley value, introduced by Shapley (1997), has desirable properties, such as efficiency, symmetry, and linearity. It enables to fairly distribute the collective payout among players by assigning each player their expected marginal contribution [43]. By applying this concept to the relationship between natural gas prices and sovereign bond yields, it will provide a means to quantify the average contribution of natural gas prices to the prediction of bond yields within different coalitions or combinations of features [30].

To better illustrate the reasoning behind this method, let's consider the following solely illustrative example. Suppose we have a prediction model for sovereign bond yields that uses three features: Natural Gas Prices (NGP), Inflation Rate (IR), and GDP Growth (GDP). These three features are the "players" in the cooperative game we are considering. In this case let's define:

- N: The set of all players. In our example, $N = \text{NGP, IR, GDP}$.
- S: A subset of players forming a coalition. In our example, S can be any combination of these three features. For instance, one possible coalition is $S = \text{NGP, IR}$, another could be $S = \text{IR, GDP}$, and so on.

Now, let's define $v(S)$ and reach the Shapley value equation:

- $v(S)$: The total expected sum of payouts that the members of coalition S can obtain by cooperation. In other words, it represents the prediction performance when using only the features present in coalition S. For example, $v(\text{NGP, IR})$ would be the prediction performance when using only Natural Gas Prices and Inflation Rate as features.
- Shapley Value equation:

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \times [v(S \cup \{i\}) - v(S)] \quad (3.4)$$

where:

- φ_i : The Shapley value for player i (in our case, NGP, IR, or GDP).
- $|S|$: The number of players in coalition S.
- $|N|$: The total number of players.
- $v(S)$: The total expected sum of payouts for coalition S (as mentioned earlier).

- The sum extends over all subsets S of N excluding the player i . It represents all the possible combinations of players where player i is added to different coalitions.

The Shapley value quantifies the average contribution of each feature (player) to the prediction of bond yields across all possible coalitions or combinations of features. It provides a fair way of distributing the collective payout (prediction performance) among the features based on their marginal contributions.

By calculating the Shapley value for each feature either for this example or for the problem of interest of this theses, the importance of each feature can be assessed in the prediction model and understand how much they contribute to the overall accuracy or performance.

However, calculating the exact Shapley value can be computationally intensive, particularly when dealing with a large number of features [44]. Additionally it is important to note that the interpretation of Shapley values should be done carefully. The Shapley value represents the average contribution of a feature value to the prediction within different coalitions, but it should not be confused with the difference in prediction when a feature is removed from the model [44].

Furthermore, the computation of Shapley values requires the availability of data related to all variables, including both dependent and explanatory variables, rather than solely relying on the prediction function, which in this particular thesis is represented by a simple linear model. In addition the inclusion of correlated features can pose challenges in terms of sampling and maintaining the symmetry axiom of the Shapley value [44]. However, in the context of this dissertation, we have already evaluated the absence of correlation, making it irrelevant in this particular example.

In practical terms, initially the Shapley values were computed for the various dependent variables in order to conduct the test. In order to determine the proportionate contributions, a Force Plot was utilized, as the values are dependent upon the scale of the contributions. In fact, Force Plots are commonly employed in network analysis and data visualization to acquire a deeper understanding of the structure and patterns present in complex systems. They facilitate the visualization of relationships between the input variables, such as the Shapley values. The utilization of both these plots and the relative Shapley values bars graphic representation provide comparable and more comprehensible overall results.

3.4 Periodical Analysis

After obtaining the initial results from the complete dataset, I proceeded to divide the data into three distinct sub-samples based on significant macroeconomic events. The first subset, referred to as the "*Before Covid*" period, spans from January 2, 2018, to January 1, 2020. The second subset, known as the "*Covid Pandemic*" period, covers the time frame from January 2, 2020, to January 1, 2022. Lastly,

the third subset, labeled the "*Ukraine War*" period, encompasses the period from January 2, 2022, to June 1, 2023.

For the three datasets, I employed the same analytical techniques as for the complete dataset, specifically the LASSO and Shapley values methods. However, this time the focus was on examining the changes in the impact of the TTF variable across three distinct time periods.

4

Results Discussion

Contents

4.1	Ordinary Least Squares Regression	47
4.2	LASSO	49
4.3	Shapley Values	53
4.4	Periodical Analysis	55
4.5	The Overall Results	58

Prior to starting this chapter, and for clarity purposes, it is important to note that the primary procedure involves performing tests on a total of twenty distinct models for each of the dependent variables, utilizing the same set of five explanatory variables (VIX, POL, F2S, OIL, and TTF). It is worth highlighting that particular emphasis will be placed on analyzing the behavior of the TTF variable, as it is directly related to the research question (once it represents the Natural gas price).

4.1 Ordinary Least Squares Regression

The first set of results was obtained for the OLS using cross validation techniques, performed as described in the above **Methodology Section 3**. In order to summarize the overall (quite numerous) results, in Table 4.1 it is included the dependent variable of the regression, the TTF coefficient and the correspondent p-value. In the table the results in bold correspond to the statistically relevant coefficients, or in other words the coefficients with a p-value smaller than 0.05. The Appendix containing all tables with the full results is present in **Appendix A.3.1**.

Dependent Variable	TTF Coefficient	TTF p-value
BGe1	-0.0277	0.0234
BGe10	-0.0625	0.0004
Bl1	-0.0688	0.1802
Bl10	-0.1024	0.0016
BSp1	-0.0122	0.3596
BSp10	-0.0803	0.0002
BFr1	-0.0284	0.0287
BFr10	-0.0748	0.0001
BHn1	0.0914	0.0251
BHn10	0.1436	0.0005
SpBGe1	0.2310	0.0036
SpBGe10	0.0967	0.1795
SpBl1	-0.0861	0.6226
SpBl10	-0.5707	0.0323
SpBSp1	0.1126	0.1740
SpBSp10	-0.1423	0.1147
SpBFr1	0.1522	0.0288
SpBFr10	-0.0331	0.4583
SpBHn1	-3.0821	0.0528
SpBHn10	-0.5951	0.3591

Table 4.1: TTF coefficients and p-value for the different regressions

Based on the analysis of the summary table provided, it can be inferred that the Natural Gas TTF contract exhibits a greater influence on the changes in yields, as indicated by the data presented in the upper half of the table. Conversely, the impact of the Natural Gas TTF contract on the Spreads of the same yields appears to be comparatively less significant, as evident from the information presented in

the lower half of the table.

Additionally, in the context of changes in yields (upper table half), it is observed that the TTF variable exhibits greater significance in longer-term cases compared to shorter-term cases. This is evident from the statistical significance of all longer-term yield changes in the results. The results associated to the Spreads (upper table half) do not exhibit a distinct trend.

However it is important to give a note on the expected sign for this coefficient and the actual sign given by the regression results. We note that the estimated value of the TTF coefficient is the estimated value for a change (in the bond itself or on the spread) in the expected value of the sovereign bond yields if the TTF contract increases by 1%. As given by the context described in the **Introduction Section 1**, even though the relationship between the dependent and TTF variables is expected to be complex (due to the various factor that influence both). The reasoning would be that if geopolitical tensions cause a spike in natural gas prices, it may lead to higher bond yields due to increased risk aversion and uncertainty among investors. Additionally if higher natural gas prices significantly impact the cost structure of these industries, it can have broader implications for economic growth and productivity. This, in turn, can influence bond yields, with higher prices potentially leading to higher yields as investors demand higher compensation for potential economic risks. However there is also a reasoning that sustains the opposite effect, if higher natural gas prices are truly driven by reduced supply, it may indicate disruptions in energy markets, potentially impacting economic activity and leading to lower bond yields.

Observing again the summarized results in Table 4.1 we can observe that the majority of the regressions estimate a negative relationship between both variables, sustaining that the final reasoning may prevail in the results. Nevertheless the relationship can vary across countries, which is observable in the first half of the table for the case of Hungary.

Finally and once again, the behavior of the Spread exhibits a higher degree of randomness, leading to unpredictable changes in the TTF coefficient sign. Consequently, it becomes challenging to comprehensively identify the underlying factors driving these results.

4.1.1 OLS Commodity Comparison

In order to examine a potential comparison between the two commodities encapsulated in the five explanatory variables (OIL and TTF), the focus will now be on the outcomes of the OLS regression for the OIL variable. The summary table provided below presents the findings for the OIL variable, which we will compare to the previously discussed results for the TTF variable.

Dependent Variable	OIL Coefficient	OIL p-value
BGe1	0.1142	0.0000
BGe10	0.1887	0.0000
Bl1	0.0091	0.9331
Bl10	-0.0473	0.4886
BSp1	-0.0153	0.5882
BSp10	0.0854	0.0623
BFr1	0.0658	0.0166
BFr10	0.1908	0.0000
BHn1	0.0557	0.5180
BHn10	0.0210	0.8083
SpBGe1	0.1439	0.3902
SpBGe10	0.0988	0.5158
SpBl1	-0.6591	0.0746
SpBl10	-0.9631	0.0872
SpBSp1	-0.1728	0.3231
SpBSp10	-0.1977	0.2991
SpBFr1	0.0277	0.8506
SpBFr10	-0.0381	0.6867
SpBHn1	-0.9097	0.7866
SpBHn10	-1.1984	0.3819

Table 4.2: OIL coefficient and p-value for the different regressions

In this particular instance, it is evident that the model does not attribute as much statistical significance to the variable OIL when compared to the outcome obtained for the TTF variable. Additionally, it is noticeable that the statistical significant cases apply exclusively to two nations, namely Germany and France. This serves as an initial indication of the commodity comparison between OIL and TTF, which will be further explored within the subsequent sections of this chapter.

Although the utilization of OLS regression provides a general indication of the potential influence of natural gas on sovereign bond yields, it will be supplemented by the following methodology, as delineated in the preceding section, thereby adding to the results that were already obtained and discussed in this section

4.2 LASSO

In accordance with the methodology described in Section 3, the results of the LASSO implementation include graphical depictions of the evolution of the five explanatory variables coefficient values (y-axis) in relation to the LASSO restriction coefficient, λ (x-axis). Below are two examples of graphics that illustrate the German ten-year changes in yields and the German ten-year yield spread. The remaining images for each of the twenty distinct models, one for each dependent variable, can be found in Appendix A.3.2.

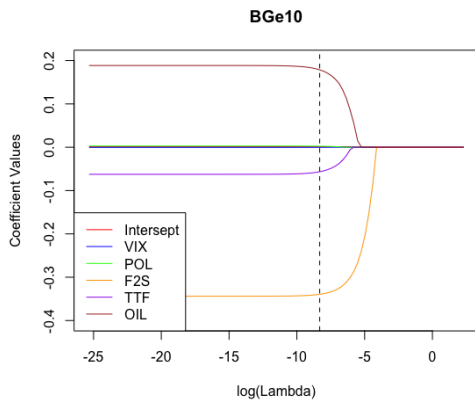


Figure 4.1: Coefficients for each LASSO λ value for BGe10

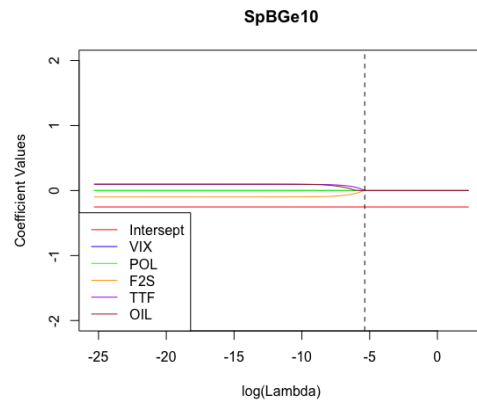


Figure 4.2: Coefficients for each LASSO λ value for SpBGe10

To summarize the results depicted in the graphical representations, a summary table was constructed. This table presents various metrics for each regression, including the dependent variable of the regression (referred to as "*Dependent Var*"), the last variable to be unselected by the LASSO method (also known as the "*Most important Variable*"), the lambda value at which the coefficient of the TTF variable becomes zero (denoted as " λ to TTF=0"), the sign of the TTF coefficient, and the sign of the TTF variable for the optimal calculated lambda (referred to as "*Signal TTF at opt λ* "). The table that is obtained is as follows:

Dependent Var	Most Important Var	λ to TTF=0	Sign TTF	Signal TTF at opt λ
BGe1	F2S	0.001	(-)	(-)
BGe10	F2S	0.003	(-)	(-)
Bl11	Intercept	0.003	(-)	0
Bl110	VIX	0.005	(-)	(-)
BSp1	F2S	0.0008	(-)	0
BSp10	F2S	0.004	(-)	(-)
BFr1	F2S	0.001	(-)	(-)
BFr10	F2S	0.003	(-)	(-)
BHn1	Intercept	0.005	(+)	0
BHn10	Intercept	0.007	(+)	(+)
SpBGe1	Intercept	0.012	(+)	(+)
SpBGe10	Intercept	0.005	(+)	0
SpBl11	Intercept	0.005	(-)	0
SpBl110	Intercept	0.038	(-)	(-)
SpBSp1	TTF	0.005	(+)	0
SpBSp10	Intercept	0.009	(-)	0
SpBFr1	Intercept	0.007	(+)	(+)
SpBFr10	Intercept	0.002	(-)	0
SpBHn1	TTF	0.152	(-)	(-)
SpBHn10	TTF	0.038	(-)	(-)

Table 4.3: Summary table LASSO Results

Upon examination of the summary table related to the LASSO results, notable observations can be made. Specifically, in the first half of the table, it is evident that the F2S variable is consistently identified as the most significant among the other explanatory variables. Furthermore, it is noteworthy that none of the results indicate the natural gas TTF variable as the most important variable in the regression. Upon examination of the third column, it becomes evident that there exists a noticeable pattern that the long-term alterations in yields require a more strict constraint, denoted by a higher value of λ , in order to exclude the natural gas TTF variable from the regression. Upon examining the sign of the TTF coefficient, it becomes evident that the sign aligns with the findings derived from the OLS analysis, as anticipated. The final column of the table indicates that there is a limited number of situations where the TTF variable is excluded at the optimal value of lambda. This observation aligns with the findings of the previous OLS assessment. Specifically, the regressions in which the coefficient for TTF was not found to be statistically significant are also the regressions in which the TTF variable is omitted by the LASSO, with the exception of the change in 1-year Hungary bond yield.

On the other hand, in the second part of the table, relating to the spreads, a notable difference arises due to the existence of certain regressions that identify the TTF as the most important variable for both the yield spreads of Hungary and the short-term yield spread of Spain. However, the significance of F2S is largely diminished, as the majority of the regressions indicate the Intercept as the most important variable. In relation to the second column, there is a notable shift in the observed patterns during the first half of the table. This shift applies to the short-term yield spreads, which demonstrate higher values of λ in order to remove the TTF variable from the regression equation, as compared to the first half of the table. The only exception to this pattern is Spain, which maintains the same behavior as described in the preceding paragraph. Ultimately, the final two columns result in identical findings in terms of the mirrored outcomes observed in the OLS regression. Nevertheless, it is worth noting that there are a greater number of instances where the TTF is not incorporated at the optimal λ . The findings of this section of the table exhibit a higher degree of randomness compared to the preceding section, as similarly indicated in the analysis of the OLS results.

4.2.1 LASSO Commodity Comparison

In order to keep the same structure as in the previous OLS section, I will proceed by conducting a direct comparison between two explanatory variables associated with commodities, namely the TTF and OIL. In order to provide a concise overview of the LASSO regression results for this comparison, the forthcoming table will present the sign of the OIL variable (referred to as "Sign OIL") and identify the variable that is excluded last by the LASSO restriction, λ (referred to as "Last commodity to zero"). The findings are as follows:

Dependent Var	Sign OIL	Last commodity to zero
BGe1	(+)	OIL
BGe10	(+)	OIL
Bl1	(+)	TTF
Bl10	(-)	TTF
BSp1	(-)	TTF
BSp10	(+)	TTF
BFr1	(+)	OIL
BFr10	(+)	OIL
BHn1	(+)	TTF
BHn10	(+)	TTF
SpBGe1	(+)	TTF
SpBGe10	(+)	TTF
SpBl1	(-)	OIL
SpBl10	(-)	TTF
SpBSp1	(-)	TTF
SpBSp10	(-)	TTF
SpBFr1	(+)	TTF
SpBFr10	(-)	TTF
SpBHn1	(-)	TTF
SpBHn10	(-)	TTF

Table 4.4: Summary table OIL vs TTF LASSO Results

Observing these results for the OIL variable, it is possible to conduct a simple comparison to the TTF variable. Overall the sign of the OIL price drifts the majority of times from regression to regression, as we observed in the OLS results. Furthermore, the initial section of the table shows that the indicated commodity variable, which is considered to be more important by the LASSO, remains equal within the country's short and long-term variables. Overall there is a clear preference for the TTF variable.

This analysis has given the input of a more in-depth understanding of the strength of the coefficients that were attributed to the initial OLS analysis, however and as mentioned in the **Methodology Section 3** this method is quite limited in terms of understanding exactly to which extent does our interest variable, the TTF, contributes to the overall change in the two yields behaviour metrics, beyond the already known and discussed limitations of the LASSO model. This leads us to the next method described in the **Methodology 3**, the Shapley Values.

This analysis has provided a deeper understanding of the strength of the coefficients derived from the initial OLS analysis. However, as stated in the **Methodology Section 3**, this method has certain limitations in accurately determining the extent to which our variable of interest, TTF, influences the overall changes in each dependent variable. This brings us to the subsequent approach described by the **Methodology 3** chapter, namely the Shapley Values.

4.3 Shapley Values

The Shapley Values calculation was executed in accordance with the methodology outlined in the Methodology Section 3. The resulting output consisted of multiple graphical bar representations, illustrating the relative Shapley values for each coefficient included in the previous OLS and LASSO models. To improve the visualization of the exact percentages corresponding to the Shapley Value of a variable within a coalition alongside the other explanatory variables, a corresponding force plot has been created. An example of these two plots is provided below, and all of the other resulting graphics can be found in **Appendix A.3.3**.

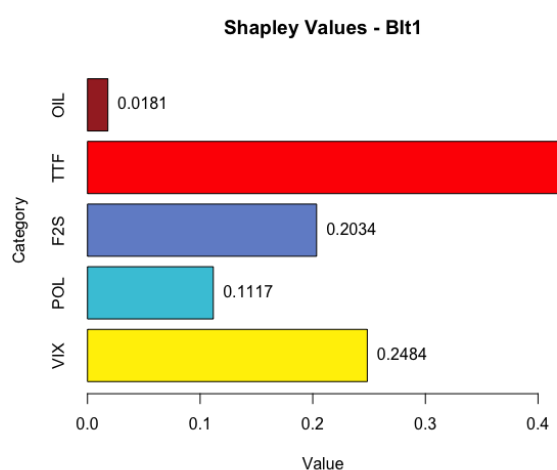


Figure 4.3: Shapley Values for the change in yields for Italy 1y bond

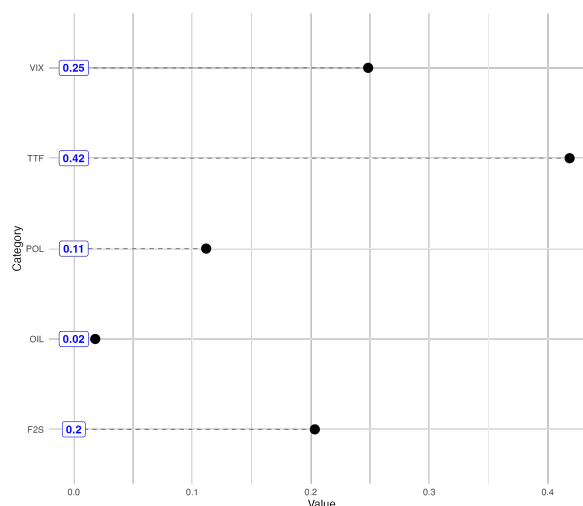


Figure 4.4: Force Plot of the Shapley Values for the change in yields for Italy 1y bond

The table presented below provides a brief summary of the key results. Firstly, it outlines the percentage of the TTF's contribution in relation to the other explanatory variables for each dependent variable. Secondly, it identifies the variable that made the greatest contribution and specifies the corresponding percentage.

Dependent Var	Force Plot	TTF contribution	Var Higher Impact	Percentage Higher Impact
BGe1		0.01	F2S	0.9
BGe10		0.03	F2S	0.86
Bl1		0.42	TTF	0.42
Bl10		0.17	VIX	0.74
BSp1		0.02	F2S	0.92
BSp10		0.20	F2S	0.61
BFr1		0.03	F2S	0.9
BFr10		0.06	F2S	0.81
BHn1		0.6	TTF	0.6
BHn10		0.21	VIX	0.74
SpBGe1		0.8	TTF	0.8
SpBGe10		0.44	TTF	0.44
SpBl1		0.11	OIL	0.82
SpBl10		0.56	TTF	0.56
SpBSp1		0.66	TTF	0.66
SpBSp10		0.66	TTF	0.66
SpBFr1		0.9	TTF	0.9
SpBFr10		0.19	F2S	0.74
SpBHn1		0.91	TTF	0.91
SpBHn10		0.51	TTF	0.51

Table 4.5: Summary table Shapley Results

Looking at the above Table results, it is noticeable that the first portion of the table indicates a tendency for the overall contribution of TTF to hold greater significance in longer-term bonds compared to shorter-term bonds, except in the case of Italy. Also in the first half of the table, it is observed that the F2S variable tends to make greater contributions, with only two exceptions where the TTF variable exhibits the highest contribution. Additionally, it is evident that a significant proportion of the highest contribution is monopolized, with the percentage exceeding 80% in approximately half of the cases. In the second part of the Table, a significant difference arises in the TTF Shapley Value, in which this variable assumes the largest percentage of contribution in nearly all times. The overall conclusions in this second half of the table remain consistent with the previous sections, with the exception that the significance of TTF tends to be reversed, becoming more important in the short term rather than the long term.

4.3.1 Shapley Values Commodity Comparison

The commodities were compared again, and the end result is displayed in the table below, which provides a similar structure as the one in the equivalent LASSO subsection:

Dependent Var	Force Plot OIL contribution	Commodity with highest contribution
BGe1	0.07	OIL
BGe10	0.08	OIL
BIt1	0.02	TTF
BIt10	0.07	TTF
BSp1	0.01	TTF
BSp10	0.04	TTF
BFr1	0.05	OIL
BFr10	0.11	OIL
BHn1	0.06	TTF
BHn10	0.01	TTF
SpBGe1	0.12	TTF
SpBGe10	0.14	TTF
SpBIt1	0.82	OIL
SpBIt10	0.04	TTF
SpBSp1	0.3	TTF
SpBSp10	0.31	TTF
SpBFr1	0.04	TTF
SpBFr10	0.06	TTF
SpBHn1	0.04	TTF
SpBHn10	0.44	TTF

Table 4.6: Summary table Commodity Comparison Shapley Results

It is interesting to find that these findings align with the conclusions presented in the previous section on LASSO, thus reinforcing the same overall conclusions. However, they offer a more precise quantification of the degree to which each variable influences the dependent variable.

4.4 Periodical Analysis

The prior sections looked at various parsimonious approaches for evaluating the influence of the TTF variable in the dependent variables for the full dataset. However, as previously discussed, the period under consideration exhibited significant volatility, leading to the interest to analyse the three sub-samples outlined in the Methodology Section. As stated in the Methodology section 3, the analysis will concentrate on the LASSO and Shapley Values. These methods have been chosen because they already generate the essential outcomes of the OLS, as demonstrated in the previous sections.

4.4.1 Least Absolute Selection and Shrinkage Operator

In the context of LASSO regression, graphical representations of the coefficients were obtained, similar to the ones presented in the previous LASSO section. However, this representation exclusively depicted the TTF coefficient for each of the defined periods. An illustration of the obtained graphics is depicted below, while the full set of acquired graphics can be found in **Appendix A.4.1**.

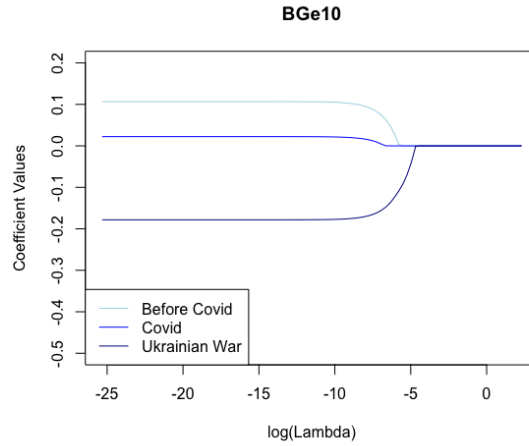


Figure 4.5: TTF Coefficients for sub-sample period for BGe10

Similarly to the methodology employed in the previous section on Results and Discussion, a summary table was constructed. This table includes the following information: the time period during which the TTF coefficient is set to zero in first place, denoted as the "Least important TTF" period; the time period during which TTF is set to zero in last place and considered the most significant, referred to as the "Most important TTF" period; and three columns indicating the sign of the TTF variable during each respective period, labeled as "Sign BC," "Sign CP," and "Sign UW." The findings are as follows:

Var	Least important TTF	Most important TTF	Sign BC	Sign CP	Sign UW
BGe1	BC	UW	(+)	(+)	(-)
BGe10	CP	UW	(+)	(+)	(-)
Bl1	CP	UW	(+)	(-)	(-)
Bl10	CP	UW	(+)	(+)	(-)
BSp1	CP	UW	(+)	(+)	(-)
BSp10	CP	UW	(+)	(+)	(-)
BFr1	BC	UW	(+)	(+)	(-)
BFr10	CP	UW	(+)	(+)	(-)
BHn1	BC	UW	(-)	(-)	(+)
BHn10	CP	UW	(+)	(+)	(+)
SpBGe1	BC	UW	(-)	(-)	(+)
SpBGe10	CP	BC	(+)	(-)	(-)
SpBl1	BC	CP	(+)	(-)	(-)
SpBl10	UW	CP	(-)	(-)	(-)
SpBSp1	BC	UW	(-)	(-)	(+)
SpBSp10	UW	CP	(-)	(-)	(-)
SpBFr1	BC	UW	(+)	(-)	(+)
SpBFr10	BC	CP	(-)	(-)	(-)
SpBHn1	BC	UW	(-)	(-)	(-)
SpBHn10	BC	?	(+)	(-)	(+)

Table 4.7: Summary period analysis table LASSO Results

On observing the above table, a notable result is the evident discrepancy between the first half and the second half of the table. While the patterns in the first half can be inferred to some extent, the second half appears to follow a more random pattern, making it difficult to derive a clear interpretation, thus leading to the decision of leaving it out of the analysis for this particular segment of the findings.

Looking at the first half of the table, we can see that, as predicted, given that it coincides with a period of rising prices, the TTF variable takes on greater importance in all cases during the Ukraine War Period. The least significant time then changes, whereas the Covid-19 period tends to be the most frequently chosen. The shift in the signal of the coefficient, which nearly always takes place in the Ukraine War period, is another significant and pertinent conclusion. This could point to a change in the way the price of natural gas is connected to the bond yield changes. Beyond this and returning to the OLS's initial findings, it is confirmed that the sign assigned to the coefficient of the TTF variable for the full sample is the same as the sign during the Ukraine War period, which may imply that when the whole sample of data is taken into account, this period has a greater influence than the others. This is not surprising because cross-validation procedures were used to get the OLS estimator, which means that since the gas has a greater impact on the change in bond yields during the Ukraine War period, this estimate will be have a greater influence as it offers a more favorable fit. This shift in dynamic is an intriguing result that, from an intuitive standpoint, makes sense considering the reasoning shown in the previous OLS section.

4.4.2 Shapley Values

The results of the Shapley Values tests generated a visual representation in the form of a bars graphic and a Force Plot, similar to the full sample case. However, these outputs only included the Shapley values for the TTF variable in the three different periods considered. The following two images presented below serve as an illustrative example of the full outcomes that are available in **Appendix A.4.2**.

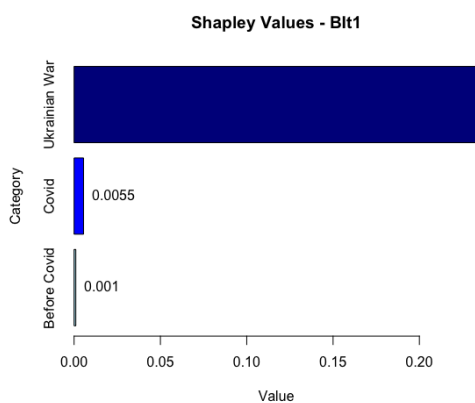


Figure 4.6: TTF period Shapley Values for Blt1

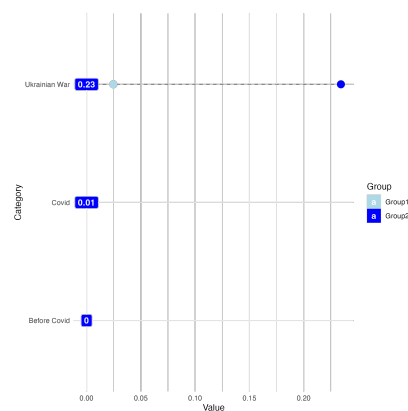


Figure 4.7: Force Plot of the TTF period Shapley Values for Blt1

Var	Before Covid TTF	Covid Pandemic TTF impact	Ukraine War TTF impact
BGe1	0	0.02	0.02
BGe10	0.05	0.01	0.14
Bl1	0	0.01	0.23
Bl10	0.01	0	0.33
BSp1	0.28	0.01	0.03
BSp10	0.64	0	0.26
BFr1	0.05	0.06	0.04
BFr10	0.1	0.02	0.22
BHn1	0.07	0.23	0.38
BHn10	0.71	0.01	0.35

Table 4.8: Summary table Shapley period analysis Results

Upon a thorough review of the results, it becomes evident that certain unanticipated results have emerged. Despite certain similarities to the outcomes derived from the previous analysis conducted using the LASSO method, the general tendency towards a greater influence of the TTF variable during the period of the Ukraine war is not consistently evident in many cases. Recalling the concept of Shapley values, which quantify an average impact on the value of prediction resulting from changes in bond yields, it becomes apparent that the evaluation of the TTF variable's influence across different scenarios is challenging. Nevertheless, there are several interesting results that can be identified as potential developments, specifically during the period covering the Covid-19 pandemic and the Ukraine War. It is evident that, with the exception of the one-year bond in France, the majority of contributions either increase or remain constant. Another observed behavior is the decline in the average contribution from the period before the Covid outbreak, with some exceptions.

Overall the results of this analysis sustain the initial LASSO results of the change of dynamic between the growth rate of the TTF Dutch natural gas contract and the changes in the Sovereign Bond Yields.

4.5 The Overall Results

The results presented and analyzed in this section have consistently demonstrated that the three methods agree on the relationship between our variable of interest (TTF) and the dependent variables. The results has demonstrated certain signs of each country's connection with natural gas (described in the Introduction Chapter 1), as well as the comparison of the effects of oil and gas prices on sovereign bond yields. In addition, there is a noticeable shift in the relationship between our dependent variables and TTF during the period characterized by higher volatility in natural gas prices (UW). However, the expected effect of the natural gas variable on bond yields is not as evident as initially hypothesized, although this effect can still be observed.

However, before drawing the main conclusions, it is essential to consider whether this analysis, which employed a parsimonious approach, is truly valuable. Moreover, it is crucial to examine whether the initial assumption regarding the fulfillment of the required statistical conditions might jeopardize these findings, and if so, to which extent. The upcoming chapter will focus on this specific matter.

5

Results Validation

Contents

5.1	Ordinary Least Square Regression	63
5.2	Cross-validation	65
5.3	Other Regression Tryouts	65
5.4	LASSO	66
5.5	Shapley Values	68
5.6	Data Characteristics	69
5.7	The Machine Learning Tendency	69

This section will mainly focus on the validation and discussion of the application of the models used in order to answer the research question of the dissertation. Thus I check the main assumptions regarding the models previously used, and I refer to the literature on this topic, where this kind of verification is rarely presented. I will first focus on the key and basic OLS assumptions.

5.1 Ordinary Least Square Regression

The OLS relies on several crucial and well known assumptions to allow accurate estimation and interpretative results:

- **Linearity:** The relationship between the dependent variables and explanatory variable is assumed to be linear.
- **Independence:** The explanatory variables used in the regression analysis are assumed to be independent of each other. This assumption ensures that the errors or residuals are not systematically related to each other (auto-correlation).
- **Homoscedasticity:** The variance of the errors is constant across all levels of the explanatory variables.
- **No endogeneity:** OLS assumes that there is no endogeneity, meaning that each variable is not influenced by the error term. This assumption is crucial to establish a causal relationship between the OLS variables.
- **No perfect multicollinearity:** OLS assumes that there is no perfect linear relationship between the variables. Perfect multicollinearity occurs when one variable can be perfectly predicted from a combination of other variables.

These assumptions can be tested by analysing the behaviour of the residuals regarding their auto-correlation, heteroskedasticity and non-normality.

The tests I employ to evaluate these properties of the residuals are the Breusch-Pagan test for heteroskedasticity, which is based on the null hypothesis that the data is normally distributed, the Ljung-Box Q-test for the auto-correlation, where the null hypothesis is that the residuals are independent (no auto-correlation), and, finally, the Shapiro-Wilk test, which is based on the null hypothesis that the data is normally distributed. The table below contains the test results for the OLS that were first obtained. Additionally, the **Appendix A.5** has various graphical representations of the residuals, including the normality quantile-quantile plot, an ACF (autocorrelation function) plot, and a fitted vs. residuals plot.

Dep_Var	Autocorr_Test p-value	Homosk_Test p-value	Norm_Test p-value
BGe1	0.0000	0.0000	0.0000
BGe10	0.0000	0.0000	0.0000
Blt1	0.0000	0.8971	0.0000
Blt10	0.0000	0.0068	0.0000
BSp1	0.0000	0.0000	0.0000
BSp10	0.0000	0.0031	0.0000
BFr1	0.0000	0.0000	0.0000
BFr10	0.0000	0.0002	0.0000
BHn1	0.0000	0.3412	0.0000
BHn10	0.0000	0.0000	0.0000
SpBGe1	0.0000	0.1456	0.0000
SpBGe10	0.0000	0.9934	0.0000
SpBlt1	0.0000	0.2754	0.0000
SpBlt10	0.0000	0.6434	0.0000
SpBSp1	0.0000	0.0274	0.0000
SpBSp10	0.0000	0.0790	0.0000
SpBFr1	0.0000	0.2697	0.0000
SpBFr10	0.0000	0.0577	0.0000
SpBHn1	0.0000	0.7541	0.0000
SpBHn10	0.0000	0.9954	0.0000

Table 5.1: Residual tests results

Only the p-values of the residual tests are presented in the table above, while the statistics values can be found in Appendix A.6. Upon examination of the table, it becomes evident that the obtained results exhibit a violation of the necessary requirements for ensuring the reliability of the findings under consideration. This violation is primarily manifested through the presence of auto-correlation in the residuals in all regressions. In relation to the phenomenon of heteroskedasticity, certain cases indicate the presence of homoskedasticity, while the outcomes of normality tests consistently indicate non-normal distributions. When the OLS model appears invalid, the primary considerations are the first two traits (auto-correlation and heteroskedasticity). For the last trait, normality, if the residuals do not conform to a normal distribution, having a large sample, the central limit theorem (CLT) can be invoked to provide justification for trusting the output OLS test statistics and confidence intervals, as stated by Wooldridge in [42].

The key concern in the results obtained from the analysis is the evident presence of auto-correlated errors, which significantly undermines the reliability of the estimated coefficients and calls into question the statistical significance of the findings. Auto-correlated residuals in the regression model lead to inefficiency in estimating the regression coefficients, resulting in sub-optimal forecasts and invalid significance tests. This issue challenges the interpretation and usefulness of the presented results [45].

Moreover, the presence of autocorrelated residuals also indicates potential mis-specification in the OLS regression equations. Such mis-specification can arise from the omission of relevant variables

or the inclusion of irrelevant ones [45]. It is common for regression equations to suffer from misspecification, which can further compromise the accuracy and validity of the estimated coefficients. Recognizing and addressing this issue is crucial for ensuring the reliability of the results.

In the context of auto-correlated residuals in literature, there is some sign of omission of these tests even if there are several earlier warnings about the dangers of auto-correlation in the residuals. Once again, if a regression equation demonstrates strong auto-correlation in the residuals, as it was observed in the above results, it indicates serious problems in interpreting the coefficients of the OLS results [45].

The question that arises is whether the utilization of the cross-validation technique to estimate the coefficients of the OLS mitigates these concerns.

5.2 Cross-validation

The use of cross-validation (CV) in estimating the coefficients of the residuals does not necessarily invalidate the concerns with auto-correlation discussed earlier. While Bergmeir et al. (2018) [46] suggests that cross-validation can adequately control overfitting in well-fitted models, it highlights the importance of checking for serial correlation in the residuals to detect cases where cross-validation procedures may yield a systematic underestimation of the error, which is the current case.

If the model is heavily mis-specified and produces residuals with non-zero serial correlation, the assumptions underlying cross-validation are violated as demonstrated in this article [46]. When misspecification occurs the CV method fails to work effectively. In such cases, the validity of the proposed CV is compromised, as the assumptions related to serial correlation and model specification are no longer met.

5.3 Other Regression Tryouts

In order to investigate the potential resolution to the inherent problem in the OLS model, an initial approach was taken to employ a robust version of the OLS. This decision was made due to the evident presence of extreme behaviors in the data, which can be attributed to the volatility observed during the period under consideration. Additionally, given the susceptibility of the basic OLS method to outliers, a modified version of the OLS was used to mitigate the impact of these outliers. One key modification involves the implementation of robust covariance matrix calculation, which effectively addresses the issue of heteroscedasticity and accounts for the presence of outliers. The utilization of this robust covariance matrix estimation technique enhances the accuracy of the standard errors associated with the

regression coefficients, thereby enabling the execution of valid hypothesis testing and the estimation of confidence intervals. The initial findings indicate a decreasing statistical significance of all the coefficients. Nevertheless, it is important to note that this finding is not considered valid due to the unresolved auto-correlation issue of the residuals. The results tables are located in **Appendix A.7.3**.

Furthermore, the removal of outliers was performed manually using Turkey's Fence Rule. This rule involves calculating the inter-quartile range (IQR) and identifying outliers as data points that fall outside the fences. In this case, the fences were determined as points beyond 1.5 times the IQR. The findings consistently demonstrate identical residual characteristics, as evidenced in **Appendix A.7.2**, thereby making it an unsuitable solution.

Moreover, in order to assess whether the issue was linked to the disparity between the behaviors of the explanatory variables and the dependent variable, various time lags were examined. These time lags included a lag of one day, a lag corresponding to one month, and a final lag of one year. Once again, the results do not demonstrate any improvement in the residuals, as can be observed in **Appendix A.7.4**, **Appendix A.7.5**, and **Appendix A.7.6**.

The final attempt involved the inclusion of new variables, specifically interaction variables. These interaction variables were computed as the product of two distinct explanatory variables, resulting in the interaction between the two. In this instance, I solely examined the interactions between all variables and the TTF variable, while also including the squared value of the TTF as an input in the model. One of the primary findings was the shift in statistical significance from the first-order to the second-order TTF. However, it is important to note that the residuals continue to deviate from the expected requirements. The outcomes of these tryouts are documented in **Appendix A.7.1**.

All of these tryouts lead to the suspicion that the issue related to the results may be linked to the assumption of linearity, which will be further examined in the subsequent discussion. The subsequent discussion will focus on the implications of an underlying unspecified model in the other methodologies applied, beyond the OLS.

5.4 LASSO

One of the primary concerns associated with the LASSO is related to the biased selection of variables that exhibit high correlation among themselves. This arises due to the constraint imposed on the OLS regression, which lacks the capability to differentiate between correlated variables and instead randomly

selects one from this variable cluster. However, considering the correlation properties of data employed in this dissertation, this issue will not pose a problem.

The question that is posed on the validity of the obtained graphical representation of the coefficients for different strengths of constraint are directly affected by the problems already tackled for the underlying model. The existing literature on this topic is limited, with most articles focusing on the issues related to bias in the estimator. It is anticipated that the performance of the estimator will be even worse than OLS due to regularization. This is because LASSO tends to estimate larger coefficients on unstable features, as it selects only a subset of predictors. Consequently, this exacerbates the problem of omitted variables that may already exist in the underlying model, which is not new in the context of LASSO [47]

In terms of the application made specifically, the main question resides in whether the mis-specification of the model and the inherently possible non included variables would change the selection of the existing variables and would change the values of the constraint at which the variables are removed from the regression model. In order to verify this, an experiment was conducted wherein the OIL price related variable was eliminated from the LASSO model. The objective was to observe the resulting lambda value that would subsequently remove the TTF variable from the regression model, and to determine if any changes occurred.

The findings from all the regression analyses indicate that the inclusion or exclusion of the OIL variable leads to changes in the value of the TTF variable, yet to a small extent or in most cases remaining the same. This demonstrates that the inclusion of additional variables that may interact with the tested system of variables can alter the obtained results. Consequently, it can be concluded that this method lacks robustness. However, it is important to note that the results presented and discussed in the previous chapter, when employing this particular method, can only be considered valid to the extent that we can assert that, within the context of a system incorporating the given explanatory variables, variable "x" holds greater significance for the underlying model compared to the other variables. This claim, however, is inevitably flawed. The outcomes of the comparison are presented in the table below, focusing just on the changes in bonds. The spreads, both with and without the OIL variable, show identical results.

Dependent Var	λ to TTF=0 without OIL	λ to TTF=0 with OIL
BGe1	0.001	0.001
BGe10	0.002	0.003
Bl1	0.003	0.003
Bl10	0.005	0.005
BSp1	0.0008	0.0008
BSp10	0.004	0.004
BFr1	0.001	0.001
BFr10	0.003	0.003
BHn1	0.005	0.005
BHn10	0.007	0.007

Table 5.2: Summary table LASSO Results

5.5 Shapley Values

The Shapley values function package used calculates the relative importance of variables in linear regression avoiding the collinearity. Since this package assumes linearity it is mainly model dependent, which will have an impact on the results, making them dependent to the circumstances of the underlying model.

In this method, the ones categorized as model-dependent aim to identify the aspects that are considered significant by the model. In this case the Shapley values function assesses the local impact on model predictions [48].

The remaining question is whether the introduction of additional variables in the calculation of Shapley Values leads to a complete modification of the results, similar to the discussion previously held for LASSO. The resulting answer remains consistent with the previous conclusion, as the interaction between the newly introduced variables and the preexisting ones can result in varying outcomes. In this particular scenario, there are new approaches for conducting a model-independent computation of the Shapley values. These approaches offer the potential to detect non-linear relationships and reveal inconsistencies in the initial results that might otherwise remain undetected [48]. Regrettably, due to constraints imposed by the software and the newness of this approach, I was unable to implement a model-independent method for calculating Shapley values. Nevertheless, this proposal will be further examined ahead.

The results obtained in this case are valid within the specified model framework of OLS. However, similar to LASSO, these results cannot be generalized to infer the contributions of the variables to the changes in the dependent variable due to the mis-specification of the underlying model.

5.6 Data Characteristics

From the above discussion it became clear that the main problem in the validation of the previous analysis resides on the underlying model, which by itself relies on the data characteristics.

When considering the characteristics of the data, it is important to note that although it exhibits favorable behavior in terms of stationarity and lack of correlation with other explanatory variables, it is still challenging to assume a linear relationship with the tested dependent variables solely based on the graphical analysis. Indeed, particularly in the context of spreads, it is nearly illogical to suggest that a variable characterized by a well-defined structure and trajectory can be linearly explained by highly fluctuating explanatory variables. The phenomenon of bond changes exhibits some improvement as the dependent variables also demonstrate a significant degree of oscillation. Nevertheless, this situation is not without drawbacks, as it may introduce challenges in aligning the observed behaviors. Despite the various explanations and potential factors contributing to the inadequate fit of the underlying model, it is evident that the model in question is ineffective, primarily due to the initial assumption of a linear relationship. Indeed, the outcomes of the correlation tests have already indicated this evident finding, which was further demonstrated by the results of the residuals tests.

Beyond this, the type of response of the variables, namely whether they were more reactive or more memory based, and the inclusion of these two types in the same regression can also be identified as a limitation, as it requires more careful model specification. Different lag structures and functional forms may be needed to adequately capture the dynamics and interactions between these variables.

Finally The presence of a potential omitted variable bias in the regression may contribute to the invalidation of the underlying model. However, in order to establish a strong argument, it is necessary to demonstrate both the correlation between a specific omitted variable and the dependent variable, as well as the correlation between this omitted variable and the other explanatory variables, which would determine the direction of the bias. Nevertheless, the trial carried out to explain the complex relationship between the variables inputted in the models using such a limited model may reinforce this suspicion.

5.7 The Machine Learning Tendency

Throughout the development of the current thesis, a comprehensive review of relevant literature was conducted, focusing on the established methodologies and techniques employed in the domains of statistics and machine learning. Particularly, the last field has gained popularity and increased employment due to its strong overall performance and ease of access to the associated tools and applications.

However there is a lack of precision in the specification of the necessary statistical requirement for model validation in the researched papers. Namely in the Shapley values applications, this method is used as a way to "*compare and summarize the relative strength of both types of factors without delving into the complications of structural modeling*" [31] not really showing the required steps for model validation, reporting only the regression standard errors and not the characteristics of the residuals in the several tests they perform. Also the Wang et al.(2022) [30] paper, at any point mentions residuals or statistical model validation, using also a very parsimonious underlying model to calculate the Shapley Values and not mentioning the possible limitations of interpreting directly the obtained results. Finally, in this method literature we can find a more careful and detailed study of the Shapley Values applications in Brandsæter et al.(2022) [43], where the limitations on the correctness of the results is tested in cluster identification, and even though this goes a bit further than the scope of this master thesis, it shows a more thoughtful approach to the application of this methodology.

Additionally in the case of the LASSO methodology selection, while Bredin et al.(2021) [27] and Huang et al.(2009) [29] both show the results of residual auto-correlation test results, this is not provided or mentioned by Ciner et al.(2022) [28] work. In the first two papers, the first does not find the auto-correlation limitation, however the second, Huang et al.(2009) [29] identifies this problem within his results and identifies it as a "*methodological issue*", viewing its presence as "*an indication of misspecification*".

The lack of clarity in the presentation of the required validation steps before applying the machine learning techniques can lead to a dangerous path, as it was proven in the present work. Even with quite reasonable results, it has been proven that these do not give the conclusions that were initially sought. Therefore these techniques even if powerful do not have the ability to substitute completely the rigor of a parallel or even previous statistical review.

6

Conclusion

Contents

6.1	Conclusions	73
6.2	Future Work	75

6.1 Conclusions

In conclusion, the analysis conducted using different methods assuming necessary assumptions, namely OLS, LASSO, and Shapley Values, has provided some insights into the relationship between the Natural Gas TTF contract and sovereign bond yields.

The OLS results indicated that the TTF contract tended to be more relevant for changes in yields than for yield spreads. There was a tendency for the TTF variable to be more significant in longer-term yield changes compared to shorter-term changes. The majority of the regressions showed a negative relationship between the TTF variable and bond yields, aligning with the expected reasoning based on economic impacts. However, the relationship varied across countries and market conditions.

The LASSO analysis further supported the findings from OLS, revealing the importance of the TTF variable in longer-term yield changes. The results showed that the TTF variable was often omitted at higher constraint levels, indicating its relevance in the longer run. For yield spreads, the pattern was not so clear, with some regressions highlighting the significance of the TTF variable. The comparison with the OIL variable consistently showed the TTF variable as more important.

The Shapley Values analysis provided additional insights into the contribution of the TTF variable to yield changes. The results indicated that the TTF variable's importance varied across periods, with the Ukraine War period often showing higher impact. However, the overall contributions were not consistently observable due to the nature of Shapley Values.

The periodical analysis focused on sub-samples of data, considering the volatile nature of the overall period. Both LASSO and Shapley Values analyses suggested changes in the TTF variable's importance and impact during different periods, with the Ukraine War period often exhibiting a significant relationship.

However, taking a step back to check the initial assumptions, the analysis of the data and application of different statistical methods, has revealed important limitations and challenges in the validation of the underlying model assumptions. The OLS regression, which formed the basis of the analysis, relies on assumptions of linearity, independence, homoscedasticity, no endogeneity, and no perfect multicollinearity. However, the residuals tests results showed violations of these assumptions.

The presence of autocorrelated errors in the OLS model raises concerns about the accuracy of coefficient estimates and the validity of hypothesis testing. This issue indicates potential mis-specification of the model and the omission of relevant variables, further compromising the reliability of the results. It

is worth noting that the literature on this topic has emphasized the dangers of auto-correlation and the need for proper validation of the model assumptions.

The use of cross-validation in estimating the coefficients of the OLS model does not solve the concerns related to autocorrelation. While cross-validation can control overfitting in well-fitted models, it relies on assumptions that may be violated in the presence of autocorrelated residuals. The mis-specification of the underlying model and the violation of assumptions undermine the effectiveness of cross-validation.

Attempts to address the limitations of the OLS model through robust regression, removal of outliers, lagged variables, and interaction terms did not yield satisfactory results. The residuals continued to exhibit characteristics that indicated mis-specification and violated assumptions.

The LASSO regression and Shapley Values results were influenced by the underlying model's limitations. LASSO, while addressing multicollinearity, is not robust to mis-specification and may lead to biased estimation. The utilization of the Shapley Values algorithm offered valuable insights into the significance of variables. However, it is important to note that this approach does not address or alleviate the inherent limitations of the underlying model. Nonetheless, both methods showcase the significance of the explanatory variables within the system of chosen variables utilized in the model.

The data characteristics, such as the possible non-linear and oscillating behavior of the variables, pose challenges to the assumption of linearity in the model. The presence of reactive and memory-based variables further complicates the model specification. These characteristics, combined with the potential omitted variable bias, contribute to the misalignment between the model and the data.

In the literature reviewed there was a tendency to not clearly showcase the validation of statistical assumptions. Many papers focused on the implementation of methods without thoroughly discussing the limitations and potential incremental biases that may arise from mis-specification and violated assumptions.

In light of these findings, it is clear that the models/techniques used in this study do not replace the rigor of statistical model validation. The limitations and challenges identified in the underlying model call into question the reliability and interpretation of the results obtained. Further research and refinement of the model specification are necessary to address these limitations and provide more robust and accurate insights into the relationship between the Natural Gas TTF contract and sovereign bond yields

of European Countries.

6.2 Future Work

Future research should aim to address the limitations and challenges identified in this study to provide more robust and accurate insights into the relationship between the Natural Gas TTF contract and sovereign bond yields.

The natural first direction to pursue would be to refine the model specification in order to better capture the non-linear and oscillating behavior of the variables, to expand the analysis to include a broader range of relevant variables, to extend the analysis to country-specific studies to account for country-specific factors that were not included in the current work and also to expand the time period to capture potential structural changes and identify long-term trends.

However before pursuing all these more broad steps, in the literature I have found two approaches that I would implement initially, as they arise from similar problems as the ones I faced in the course of this thesis, and that are quite new.

The first one would be to change the LASSO approach to be a Supervised Adaptive Group LASSO (SAGLASSO) applied to an unrestricted auto-regressive distributed lag model (ARDL) instead of correcting for the existing auto-correlation as it was performed by Huang et al. (2009) [29]. This methodology simultaneously applies shrinkage to the coefficients of both current and lagged variables, effectively identifying important predictors while considering their potential delayed impact. Additionally the non-linear restriction in the ARDL model is relaxed. All in all the results of applying this methodology could solve the initial linearity problem and shed light into the necessary lag structure that the inputted data would require to explain the bond yields dynamics.

And the last would be to consider the work developed by Fryer et al. (2021) [48] mentioned before in the Chapter 4. As in the present thesis, the majority of the game formulations found in the machine learning and statistics literature can be categorized as model-dependent. However this study introduces the concept of model-independent Shapley values. These Shapley values utilize a measure of non-linear dependence as the defining characteristic. The major advantage of these Shapley values lies in their capacity to reveal and attribute non-linear dependencies among features. One of the ways to do this is that they calculate for each feature the attributed dependence on labels (ADL) calculated without any model, the attributed dependence on prediction (ADP) which is calculated using the output of a trained

model and compare both. This comparison aims to identify disparities in the attributions between reliance on labels and reliance on predictions generated by an improperly specified model. Unlike Shapley values that rely solely on the model, this approach allows for a comparison between the absence of a model and its outputs. Greater disparities between ADL and ADP indicate that the model inadequately captures the underlying dependence structure. Additionally the paper suggest the calculation of the attributed dependence on residuals (ADR) for each feature. If the feature in question exhibits a higher level of dependence on residuals compared to the other features it suggests that this feature plays a significant role in shaping the residuals of the model, or in other words, there is evidence that the model's predictions demonstrate a stronger reliance on this particular feature than the actual labels do.

These two methods would initially be good identifiers of the underlying problems regarding the thought of model, and would guide then the more general approaches of extending the model itself to more complex ones, never the less this last approach.

Finally considering the initial objective of this dissertation, on whether the Natural Gas Price impacts Sovereign Bond Yields in the EU, and how as it been evolving in recent years, while the analysis provided insights into this relationship, the research question regarding its impact and evolution in recent years has not been fully answered. The findings suggest a significant but complex relationship, with variations across countries and market conditions. However, limitations in the model assumptions, violations of statistical requirements, and data characteristics pose challenges. To fully address the research question, further research is needed to refine the model, consider additional variables, account for non-linear dependencies, and explore different time periods. These steps will enhance the validity and provide a more conclusive understanding of the relationship between the Natural Gas Price and Sovereign Bond Yields in the EU.

Bibliography

- [1] G. D. Bella, M. Flanagan, F. Toscani, A. Pienkowski, K. Foda, M. Stuermer, and S. Maslova, "Natural gas in europe: The potential impact of disruptions to supply," *IMF Working Papers*, vol. 145, p. 1, 2022. [Online]. Available: <https://elibrary.imf.org/openurl?genre=journal&issn=1018-5941&volume=2022&issue=145>
- [2] S. S. Chen, S. Huang, and T. Y. Lin, "How do oil prices affect emerging market sovereign bond spreads?" *Journal of International Money and Finance*, vol. 128, p. 102700, 2022.
- [3] M. A. Nasir, T. N. L. Le, Y. Ghabri, and L. D. T. Huynh, "Sovereign bonds and flight to safety: Implications of the covid-19 crisis for sovereign debt markets in the g-7 and e-7 economies," *International Review of Financial Analysis*, vol. 86, p. 102548, 2023.
- [4] M. Costola and M. Iacopini, "Measuring sovereign bond fragmentation in the eurozone," *Finance Research Letters*, vol. 51, p. 103354, 2023.
- [5] T. Warin and A. Stojkov, "Sovereign bond yield differentials across europe: A structural entropy perspective," *Entropy*, vol. 25, p. 630, 2023.
- [6] T. Poghosyan, "Long-run and short-run determinants of sovereign bond yields in advanced economies," *Economic Systems*, vol. 38, p. 1000, 2014.
- [7] F. X. Diebold and C. Li, "Forecasting the term structure of government bond yields," *Journal of Econometrics*, vol. 130, p. 337, 2006.
- [8] R. L. Kosowski and S. N. Neftci, "Chapter 3 - cash flow engineering, interest rate forwards and futures," in *Principles of Financial Engineering (Third Edition)*, ser. Academic Press Advanced Finance, R. L. Kosowski and S. N. Neftci, Eds. San Diego: Academic Press, 2015, p. 53.
- [9] M. Flouri, C. Karakosta, C. Kladouchou, and J. Psarras, "How does a natural gas supply interruption affect the eu gas security? a monte carlo simulation," *Renewable and Sustainable Energy Reviews*, vol. 44, p. 785, 2015.

- [10] A. Correljé, "The european natural gas market," *Current Sustainable/Renewable Energy Reports*, vol. 3, p. 28, 2016.
- [11] R. G. Egging and S. A. Gabriel, "Examining market power in the european natural gas market," *Energy Policy*, vol. 34, p. 2762, 2006.
- [12] C. Halser and F. Paraschiv, "Pathways to overcoming natural gas dependency on russia—the german case," *Energies*, vol. 15, p. 4939, 2022.
- [13] R. Krikštolaitis, V. Bianco, L. Martišauskas, and S. Urbonienė, "Analysis of electricity and natural gas security. a case study for germany, france, italy and spain," *Energies*, vol. 15, p. 1000, 2022.
- [14] V. Milov, "European gas price crisis: Is gazprom responsible?" *European View*, vol. 21, p. 66, 2022.
- [15] R. Dickel, E. Hassanzadeh, J. Henderson, A. Honoré, L. El-Katiri, S. Pirani, H. Rogers, J. Stern, and K. Yafimava, "Reducing european dependence on russian gas : distinguishing natural gas security from geopolitics," *Oxford Institute for Energy Studies*, p. 81, 2014.
- [16] J. Miller and R. Ratti, "Crude oil and stock markets: Stability, instability, and bubbles," *Energy Economics*, vol. 31, no. 4, p. 559, 2009.
- [17] K. Ravichandran and K. A. Alkhatlan, "Impact of oil prices on gcc stock market," *Research in Applied Economics*, vol. 2, p. E4, 2010.
- [18] N. Sariannidis, G. Galyfianakis, and E. Drimbetas, "The effect of financial and macroeconomic factors on the oil market," *International Journal of Energy Economics and Policy*, vol. 5, p. 1084, 2015.
- [19] C. Aloui, K. Nguyen, and H. Njeh, "Assessing the impacts of oil price fluctuations on stock returns in emerging markets," *Economic Modelling*, vol. 29, p. 2686, 2012.
- [20] G. Capelle-Blancard, P. Crifo, M. A. Diaye, R. Oueghlissi, and B. Scholtens, "Sovereign bond yield spreads and sustainability: An empirical analysis of oecd countries," *Journal of Banking and Finance*, vol. 98, p. 156, 2019.
- [21] I. A. Boitan and K. Marchewka-Bartkowiak, "Climate change and the pricing of sovereign debt: Insights from european markets," *Research in International Business and Finance*, vol. 62, p. 102685, 2022.
- [22] S. Cevik and J. T. Jalles, "This changes everything: Climate shocks and sovereign bonds [U+204E]," *Energy Economics*, vol. 107, p. 105856, 2022.

- [23] E. Koroleva and M. Kopeykin, "Understanding of macro factors that affect yield of government bonds," *Risks*, vol. 10, p. 166, 2022.
- [24] H. Dewachter, L. Iania, M. Lyrio, and M. de Sola Perea, "A macro-financial analysis of the euro area sovereign bond market; a macro-financial analysis of the euro area sovereign bond market," 2014. [Online]. Available: <http://hdl.handle.net/10419/144471>
- [25] S. Consoli, L. T. Pezzoli, and E. Tosetti, "Neural forecasting of the italian sovereign bond market with economic news," *Journal S. Stat Series A*, vol. 185, p. s197, 2022.
- [26] D. Bianchi and M. Büchner, "Bond risk premia with machine learning," *The review of Financial Studies*, vol. 34, p. 1046, 2021.
- [27] D. Bredin, C. O'Sullivan, and S. Spencer, "Forecasting wti crude oil futures returns: Does the term structure help?" *Energy Economics*, vol. 100, p. 105350, 2021.
- [28] C. Ciner, B. Lucey, and L. Yarovaya, "Determinants of cryptocurrency returns: A lasso quantile regression approach," *Finance Research Letters*, vol. 49, p. 102990, 2022.
- [29] J.-Z. Huang and Z. Shi, "Determinants of bond risk premia," 2009. [Online]. Available: <https://www.fordham.edu/download/downloads/id/3452/huangshi-jan2011pdf.pdf>
- [30] R. . Wang, M. . Althof, and W. Härdle, "A financial risk meter for china," 2021. [Online]. Available: <http://hdl.handle.net/10419/248438>
- [31] E. M. Cerutti, M. Obstfeld, and H. Zhou, "Covered interest parity deviations: Macrofinancial determinants," *Journal of International Economics*, vol. 130, p. 103447, 2021.
- [32] S. R. Baker, N. Bloom, and S. J. Davis, "Measuring economic policy uncertainty," 2013. [Online]. Available: www.policyuncertainty.com
- [33] Investing.com, "Brent Oil Futures - Oct 23 (LCOV3)," 2023, last accessed 23 June 2023. [Online]. Available: <https://www.investing.com/commodities/brent-oil>
- [34] Investing.com , "Dutch TTF Natural Gas Futures - Aug 23 (TFAc1)," 2023, last accessed 23 June 2023. [Online]. Available: <https://www.investing.com/commodities/dutch-ttf-gas-c1-futures-historical-data>
- [35] Cboe Exchange, "Historical Data for Cboe VIX® Index and Other Volatility Indices," 2023, last accessed 22 January 2023. [Online]. Available: https://www.cboe.com/tradable_products/vix/vix_historical_data/

- [36] Economic Policy Uncertainty, "Europe Monthly Index," 2023, last accessed 22 January 2023. [Online]. Available: https://www.policyuncertainty.com/europe_monthly.html
- [37] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, "A review of spline function procedures in r," *BMC Medical Research Methodology*, vol. 19, 2019.
- [38] R. Mushtaq, "Augmented dickey fuller test," 2011. [Online]. Available: <https://ssrn.com/abstract=1911068>
- [39] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics*, vol. 54, no. 1-3, p. 159, 1992.
- [40] S. Barwary and T. Abazari, "Preprocessing data: A study on testing transformations for stationarity of financial data," 2019. [Online]. Available: <http://www.diva-portal.org/smash/get/diva2:1334843/FULLTEXT01.pdf>
- [41] J. D. Hamilton, *Time Series Analysis*. New Jersey: Princeton University Press, 1994.
- [42] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*, ser. ISE - International Student Edition. South-Western, 2009.
- [43] A. Brandsæter and I. K. Glad, "Shapley values for cluster importance: How clusters of the training data affect a prediction," *Data Mining and Knowledge Discovery*, 2022.
- [44] C. Molnar, "Interpretable machine learning - a guide for making black box models explainable," 2023. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/shapley.html>
- [45] C. W. J. Granger and P. Newbold, "Spurious regressions in econometrics," p. 111, 1974.
- [46] C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Computational Statistics and Data Analysis*, vol. 120, p. 70, 2018.
- [47] K. Kuang, R. Xiong, P. Cui, S. Athey, and B. Li, "Stable prediction with model misspecification and agnostic distribution shift." [Online]. Available: www.aaai.org
- [48] D. V. Fryer, I. Strumke, and H. Nguyen, "Model independent feature attributions: Shapley values that uncover non-linear dependencies," *PeerJ Computer Science*, vol. 7, p. e582, 2021.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/lsm