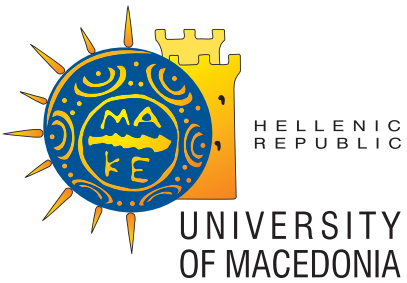

SCHOOL OF BUSINESS ADMINISTRATION

DEPARTMENT OF ACCOUNTING AND FINANCE

MASTER'S PROGRAM IN ACCOUNTING AND FINANCE

Thesis

GREEK – TURKISH ARMS RACE: AN APPROACH USING NEURAL NETWORKS

by

SPYRIDON NTELLIS

Supervisor: Prof. Achilleas Zapranis

A thesis submitted in fulfilment of the requirements for the degree of

Master of Science in Accounting and Finance

December 2021

Dedicated to my brother Thanasis
and to my parents, Nikolaos and Vasiliki
as a tribute to their never-failing love and support

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my professors at the Master's Program in Accounting and Finance for their eagerness to impart their knowledge and contribute towards the advancement of science. I also wish to acknowledge the inspiration provided by my fellow students; their stimulating questions, their remarks and the discussions we had have been an invaluable source of diverse views and interesting insights.

A special thank you belongs to my supervisor, Prof. Achilleas Zapranis, for his guidance, thoughtful comments and recommendations; for his encouragement and patience; for inspiring me to think outside the box and always proceed with diligence.

Last but by no means least, I must express my profound gratitude to my family for their unconditional support throughout my studies and through the process of researching and writing this thesis. This thesis would not have been possible without them.

ABSTRACT

Greece and Turkey constitute a prime example of perennial rivalry. Relations between the two countries have been entangled in a series of unresolved disputes, which have brought Greece and Turkey on the brink of war on several occasions. Defence economics research and, more specifically, quantitative arms race research, has concerned itself with the rivalry between the two neighbouring countries. However, the approaches adopted, based mainly on traditional econometrics, have proved inconclusive as to whether a Greek–Turkish arms race exists. Meanwhile, researchers have expressed concerns over the statistical and methodological issues involved.

This thesis revisits one of the approaches that have only scarcely been applied in related literature: the approach of neural networks. It aims to answer the question whether an arms race between Greece and Turkey exists, whilst unveiling the methodological issues involved. Three different models (A, B and C), based on neural networks, are developed for years 1963–2018, each one utilising a different set of input variables. Of them, only Model C achieves a performance considerably superior to that of the benchmark models used. An assessment of input significance through the use of SHAP values on Model C reveals that Turkey-related variables are not prime determinants of Greek defence spending. The analysis also highlights several technical issues: the ambiguity of what is termed an 'arms race', the intricacies involved in choosing input variables, data reliability issues, correlation significance testing and the impact of correlations between variables on input significance measures. These issues highlight the need for more rigorous research design and testing procedures, as well as for a careful interpretation of the results.

Keywords: Greece, Turkey, Arms Race, Neural Networks, Explainable Artificial Intelligence (XAI)

ΠΕΡΙΛΗΨΗ ΣΤΗΝ ΕΛΛΗΝΙΚΗ

Ελλάδα και Τουρκία αποτελούν ένα εξέχον παράδειγμα διαρκούς αντιπαλότητας στην Ανατολική Μεσόγειο. Κατά την διάρκεια της νεότερης ιστορίας και παρά τις όποιες περιόδους προσέγγισης και ειρηνικής συνύπαρξης, οι σχέσεις των δύο χωρών χαρακτηρίζονται συστηματικά από έντονο ανταγωνισμό, ο οποίος σε πολλές περιπτώσεις τις έχει οδηγήσει στα πρόθυρα μιας πολεμικής σύγκρουσης.

Η έντονη αντιπαλότητα μεταξύ των δύο χωρών έχει υπάρξει αντικείμενο μελέτης πολλών ερευνητών στον χώρο των οικονομικών της άμυνας και ειδικότερα στον χώρο της ποσοτικής ανάλυσης των εξοπλιστικών ανταγωνισμών. Η σχετική έρευνα βασίστηκε αρχικά στην πρωτοποριακή μελέτη του Richardson (1960*a*), ο οποίος μοντελοποίησε τους εξοπλιστικούς ανταγωνισμούς με την χρήση διαφορικών εξισώσεων. Καθώς τα εμπειρικά αποτελέσματα από την εφαρμογή του συγκεκριμένου μοντέλου δεν υπήρξαν ικανοποιητικά, η έρευνα έχει στραφεί έκτοτε κατά κύριο λόγο στην χρήση κλασικών οικονομετρικών μεθόδων. Τα αποτελέσματα των ερευνών ωστόσο παραμένουν ασαφή, με τους ερευνητές να συμπεραίνουν άλλοτε την ύπαρξη και άλλοτε την ανυπαρξία ενός εξοπλιστικού ανταγωνισμού μεταξύ Ελλάδος και Τουρκίας. Παράλληλα, η σχετική έρευνα έχει αναδείξει στατιστικά και εν γένει μεθοδολογικά ζητήματα, τα οποία επηρεάζουν την αξιοπιστία των ερευνών και πιθανώς εξηγούν την απουσία σαφούς συμπεράσματος.

Μεταξύ των σχετικών ερευνών, αξιοσημείωτη είναι η προσέγγιση των Refenes et al. (1995), οι οποίοι εξετάζουν το ζήτημα του εξοπλιστικού ανταγωνισμού με την χρήση νευρωνικών δικτύων. Τόσο οι Refenes et al. (1995), όσο και οι Andreou and Zombanakis (2000, 2011), οι οποίοι εφάρμοσαν επίσης την ίδια μέθοδο, καταλήγουν στο συμπέρασμα ότι οι κύριοι προσδιοριστικοί παράγοντες της ελληνικής αμυντικής δαπάνης σχετίζονται άμεσα με την Τουρκία και, ως εκ τούτου, υφίσταται εξοπλιστικός ανταγωνισμός μεταξύ των δύο χωρών. Σε καμμία από τις ανωτέρω μελέτες, ωστόσο, δεν γίνεται αναλυτικά λόγος για τα διάφορα τεχνικά ζητήματα που δύνανται να επηρεάσουν τα αποτελέσματα.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η επαναπροσέγγιση του ελληνοτουρκικού εξοπλιστικού ανταγωνισμού με την χρήση νευρωνικών δικτύων και η ανάδειξη των σημαντικότερων σχετικών μεθοδολογικών ζητημάτων. Για τον σκοπό αυτό, αναπτύσσονται τρία διαφορετικά μοντέλα – νευρωνικά δίκτυα για την πρόβλεψη της μεταβολής της ελληνικής αμυντικής δαπάνης την περίοδο 1963–2018. Τα μοντέλα διαφέρουν ως προς τις μεταβλητές που χρησιμοποιούνται ως προβλεπτικοί παράγοντες. Ειδικότερα,

η επιλογή των μεταβλητών στο Μοντέλο Α βασίζεται στην αρχική μελέτη του Richardson (1960*a*), στο Μοντέλο Β επιλέγονται οι ίδιες μεταβλητές με εκείνες των Refenes et al. (1995), ενώ στο Μοντέλο Γ πραγματοποιείται μια ad hoc επιλογή έξι μεταβλητών. Τρία διαφορετικά (αφελή) μοντέλα χρησιμοποιούνται ως σημείο αναφοράς για τον προσδιορισμό της ελάχιστης αποδεκτής απόδοσης των μοντέλων Α, Β και Γ.

Από τα τρία μοντέλα Α, Β και Γ, μόνο το μοντέλο Γ επιτυγχάνει απόδοση ουσιωδώς καλύτερη από τα μοντέλα αναφοράς, καθώς και από τα μοντέλα που έχουν αναπτυχθεί κατά την πρότερη έρευνα. Ειδικότερα, το μοντέλο Γ επιτυχάνει σφάλμα RMSE ίσο με 7,98%, σφάλμα MAE ίσο με 6,29% και συντελεστή προσδιορισμού R^2 ίσο με 40%.

Στο μοντέλο Γ πραγματοποιείται εν συνεχεία ανάλυση σημαντικότητας των μεταβλητών με την χρήση των τιμών SHAP. Από την ανάλυση των τιμών προκύπτει ότι οι σημαντικότεροι προβλεπτικοί παράγοντες της μεταβολής της ελληνικής αμυντικής δαπάνης είναι η μεταβολή της αμυντικής δαπάνης το προηγούμενο έτος και η μεταβολή των ελληνικών εξαγωγών ως ποσοστό του ΑΕΠ κατά το προηγούμενο έτος. Σε κάθε περίπτωση, οι μεταβλητές που σχετίζονται με την Τουρκία δεν αποτελούν βασικούς προσδιοριστικούς παράγοντες της ελληνικής αμυντικής δαπάνης και, κατά συνέπεια, φαίνεται ότι η Ελλάδα δεν μετέχει σε εξοπλιστικό ανταγωνισμό με την Τουρκία.

Παράλληλα, η ανάλυση και των τριών μοντέλων φωτίζει πληθώρα μεθοδολογικών και στατιστικών ζητημάτων. Τα ζητήματα αυτά αφορούν την μεταβλητότητα της μεταβολής της ελληνικής αμυντικής δαπάνης και τον τρόπο με τον οποίο αυτή επηρεάζει το βέλτιστο επιτεύξιμο αποτέλεσμα, την ύπαρξη συσχέτισης μεταξύ των μεταβλητών των μοντέλων και την επίδραση που αυτή έχει στις μεθόδους ανάλυσης σημαντικότητας, την επιλογή των στατιστικών επιπέδων σημαντικότητας, την μέθοδο επιλογής μεταβλητών στα νευρωνικά δίκτυα, την επάρκεια και αξιοπιστία των δεδομένων που χρησιμοποιούνται κατά την εκπαίδευση των δικτύων, καθώς και την ύπαρξη μεροληψίας προδιαγραφής. Όλα τα ανωτέρω ζητήματα πιθανώς εξηγούν την απουσία ενός σαφούς συμπεράσματος επί του ζητήματος της ύπαρξης ή μη ενός εξοπλιστικού ανταγωνισμού μεταξύ Ελλάδος και Τουρκίας.

Συνολικά, η παρούσα διπλωματική εργασία παρέχει ένα γενικό πλαίσιο μελέτης των εξοπλιστικών ανταγωνισμών με τη χρήση νευρωνικών δικτύων, φωτίζοντας σημαντικά ζητήματα που άπτονται του σχεδιασμού της έρευνας και της αξιολόγησης των εκάστοτε ευρημάτων. Μέσα από την ανάλυση των ζητημάτων αυτών, φανερώνεται η ανάγκη για επαρκή και περισσότερο αξιόπιστα δεδομένα, για εναλλακτικές μεθοδολογικές προσεγγίσεις επί του ζητήματος των εξοπλιστικών ανταγωνισμών, καθώς και για νέους αλγορίθμους που δεν επηρεάζονται από την απουσία μεγάλου όγκου δεδομένων.

Λέξεις – Κλειδιά: Ελλάδα, Τουρκία, Εξοπλιστικός ανταγωνισμός, Νευρωνικά δίκτυα, Επεξηγήσιμη τεχνητή νοημοσύνη

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In 2020, Greek–Turkish relations entered a period of severe escalation in the aftermath of the Syrian refugee crisis. This escalation was not unprecedented in the relations of the two countries. Greece and Turkey, both located in the tumultuous Eastern Mediterranean area, have historically been characterised by strained relations, despite temporary periods of rapprochement. These tense relations between the two countries as well as the level of their military expenditure over time have urged researchers to portray Greece and Turkey as a prime example of an arms race relationship.

Literature on arms races in general is extensive and provides interesting insights into the subject. Researchers have examined most rivalling countries in an attempt to uncover whether there actually is an arms racing relationship between them. The question ultimately arising from this investigation has been whether arms races could lead to the outbreak of a war. This consideration was particularly significant for the relations between USA and USSR in the era of Cold War, but it was also important for the turbulent Balkan peninsula.

Arms race research has raised another significant question as well: How does spending in the context of arms races affect the economy, growth potential and society of the countries engaged in it? The spending perspective on the issue has assumed greater importance in recent years, considering the international expansion of defence spending. Global military expenditure was estimated at about 1.922 trillion USD (constant 2018 prices) in 2019, having surged 9.83% in the last five years (Stockholm International Peace Research Institute [SIPRI], 2020). This accounts for almost 5.27 billion USD per day or, in other words, 97.57% of the United Nations *biannual* budget for 2018–2019 (United Nations [UN], 2017) on a daily basis! For Greece in particular, the subject of military spending was brought once again under the spotlight after the country's recent economic crisis. The publicity given to the issue and the discussions in which it resulted, highlighted the need for a detailed insight into the defence spending dynamics of rival countries, so that policy-makers are able to reach proper decisions regarding defence spending.

Arms races analysis dates back at least to Montesquieu (Luterbacher, 1985), but it was the first formalisation of arms races models by Richardson (1960*a*) that spawned considerable research in the field. Richardson provided researchers with a simple yet powerful model in an era of antagonism between West and East, which threatened to lead the whole planet to a nuclear holocaust. Despite its extensive use by researchers, however, empirical results of Richardson's

model have not been satisfactory. Thus, most approaches taken thereafter have mainly been based on traditional econometrics. Unfortunately, results on many rivalling dyads have been ambiguous, while researchers have voiced considerable concerns over the suitability of the statistical methods chosen and the reliability of data used in many of these approaches.

Among the country dyads analysed in the context of arms race research, Greece and Turkey constitute a prominent example owing to their turbulent relations. However, research on the Greek–Turkish arms race yielded inconclusive results, with researchers reporting either the inexistence of an arms race or the existence of a unilateral race on the part of Greece. As in the case of other arms races, there have also been concerns over the methodological approaches taken.

A noteworthy exception in Greek–Turkish arms race literature, and in arms race research in general, was that of Refenes et al. (1995), who employed artificial neural networks in order to study Greek defence expenditure. This approach was taken by only two other researchers in the following years. In all related cases, researchers concluded that variables related to Turkish military capabilities were prime determinants of Greek defence expenditure.

Taking the aforementioned into consideration and given the renewed interest in machine learning and artificial intelligence, this thesis revisits arms race considerations in the context of Greek–Turkish relations for years 1963–2018. More specifically, the Greek–Turkish defence expenditure dynamics are analysed through the use of artificial neural networks. Original research in the field is extended using state-of-the-art techniques for model estimation and input significance evaluation.

The aim of this thesis is to provide researchers with a solid framework for the study of arms races in the specific case of Greece and Turkey, shedding light on oft-overlooked —yet critical— methodological issues. More specifically, the objective of this thesis is three-fold: to establish whether there is an arms race between Greece and Turkey and provide Greek financial administrators and decision-makers with valuable insights into the interaction of Greek–Turkish armaments, as proxied by defence expenditure; to regenerate interest in the use of neural networks and machine learning techniques to examine defence spending dynamics as well as other issues where conventional statistics prevail; to highlight substantial statistical and methodological issues involved in designing, executing and evaluating research using neural networks.

Thesis Overview

The structure of this thesis is as follows: Chapter 2 concerns itself with a review of related literature. It provides a historical note on Greek–Turkish relations since World War I, as well as a comprehensive literature review on arms race modelling and a detailed technical analysis of neural networks and SHAP values. Chapter 3 analyses the methodology employed in this thesis. Results are presented in Chapter 4, followed by a discussion of related issues in Chapter 5 and conclusions in Chapter 6.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

2.1 Greece and Turkey: Overview

2.1.1 A Historical Note on Greek–Turkish Relations After 1923

Greek–Turkish relations constitute a prime example of trouble and perennial rivalry. Despite temporary periods of cordiality, relations have always been tense.

The defeat of Greek forces in the Greco–Turkish War of 1919–1922, which led to the Mandatory Population Exchange Agreement between Greece and Turkey, the Treaty of Lausanne in 1923, the renunciation of irredentism (*Megali Idea*) by Greece (Grigoriadis, 2011) and the founding of modern Turkey, marked the beginning of a period of mutual friendship between Greece and Turkey. This friendship was mainly attributed to the countries' post-war leaders, Eleftherios Venizelos and Kemal Atatürk (Demirel, 1998; Stephanopoulos, 1998). Both countries fought at the Korean War at the side of the UN forces (Moustakis, 2003) and became NATO members in February 1952. Later, Greece and Turkey both contributed to peacekeeping forces in Bosnia and Albania (Turan and Barlas, 1999). Despite cooperation in the context of NATO, membership of NATO is argued to have led to an intensification of disputes (Krebs, 1999).

On 6 September 1955, a pogrom was launched against Istanbul's Greek minority. The pogrom, which accelerated emigration of ethnic Greeks from Turkey, was estimated to have led to a decrease of ethnic Greek population in Istanbul from 65,108 to 49,081 between 1955 and 1960 (Dundar, 1999). This incident, in addition to Greek–Cypriot guerilla unrest that had already developed in Cyprus (which was inhabited by both Greek and Turkish populations) (Heraclides, 2010), effectively terminated the period of friendship between Greece and Turkey.

Greek–Cypriot armed campaign against British rule in Cyprus in the 1950s, in favour of a union with Greece (*enosis* — Greeks accounted for 82% of the population of the island in 1955) finally resulted in the establishment of an independent Republic of Cyprus, whose constitution strived to ensure representation of both Greek- and Turkish-Cypriot communities. Despite constitutional provisions, disputes arose soon after Cypriot independence. In 1964, a planned Turkish invasion of Cyprus was prevented by US diplomatic intervention (Krebs, 1999; Larrabee, 2012).

In July 1974, a coup d'état, inspired by the Greek military junta (which had seized power in 1967) with the intention of declaring *enosis*, triggered a Turkish invasion of the island and the collapse of the Greek military dictatorship (Heraclides, 2010). The invasion, in the aftermath of which Turkey controlled 36% of the island's territory, inaugurated a period of high tension in Greek–Turkish interactions (Constas, 1991).

Tensions in Cyprus have continued to exist since the Turkish invasion. On 15 November 1983, Turk–Cypriots declared an independent Turkish Republic of Northern Cyprus (TRNC), which is so far only recognised by Turkey. In 2004, the Republic of Cyprus became a member of the European Union as a representative of the whole island; in the same year, the UN–brokered Annan Plan for the re-unification of Cyprus was rejected by the majority of Greek–Cypriots despite being approved by Turkish–Cypriots. To this day, talks under the auspices of the United Nations to reach a mutually accepted compromise remain deadlocked. In light of continuing Turkish aggression towards Cyprus, a breakthrough is rather improbable.

Greek–Turkish relations deteriorated significantly in the aftermath of the Turkish invasion of Cyprus (Heraclides, 2012). Greece temporarily withdrew from NATO's military arm after the invasion (Larrabee, 2012). Its rivalry with Turkey became the main point of focus, since its communist neighbours were no longer considered to pose an immediate threat (Avramides, 1997; Ifestos and Platias, 1992) — this focus was illustrated in the defence doctrine officially declared in 1985, where Turkey is identified as a principal threat and the Warsaw Pact is only an indirect one (Platias, 1991). Significant international changes, such as the collapse of communism in the late 1980s did not seem to affect Greek–Turkish relations, given both countries' improved relations with their communist neighbours (Avramides, 1997).

The outbreak of a war was only narrowly averted in 1987 and 1996 (Athanassiou and Kollias, 2000; Matthews, 1999). The 'earthquake diplomacy', initiated after powerful earthquakes hit both countries in 1999, contributed to the inauguration of a period of cordial relations and to a shift in Greek government's adamantly negative stance regarding Turkey's accession to the EU (Larrabee, 2012). In fact, in October 2005, Greece voted for Turkey to begin entry negotiations with the EU. Although a series of co-operation agreements were signed during this period of détente, persisting disputes related to the Aegean Sea were not settled (Heraclides, 2012).

Frictions became the rule after 2016 (Heraclides, 2019). In 2016, after the failed Turkish coup attempt, eight Turkish soldiers fled to Greece, seeking political asylum. The Supreme Court of Greece denied their extradition, which was requested by Turkey. Two years later, in March 2018, Turkey detained two Greek military officers who crossed into Turkey on espionage charges and released them after almost six months of imprisonment. In 2020, the Greek–Turkish conflict

entered yet another period of escalation due to refugees and migrants being pushed to Turkish–Greek border and conflicts over maritime zones. Considering these recent advancements, as well as the fact that Turkey has evolved into an 'authoritarian' state by Turkish scholars' account (Çandar, 2019; Insel, 2019), there seems to be no reasonable expectation of a breakthrough in Greek–Turkish relations in the following years.

Overall, Greek–Turkish relations are characterised by a wide range of unresolved issues and disputes:

- The Cyprus issue
- The breadth of territorial waters in the Aegean Sea. Both Greece and Turkey currently possess six nautical miles of territorial waters. Greece claims a right to expand its waters to 12 nautical miles, in accordance with the 1982 United Nations Convention on the Law of the Sea (UNCLOS). Turkey, which has not ratified UNCLOS, has threatened Greece with war in the case it attempts to enforce the 12-mile rule unilaterally (*casus belli*, see e.g. Moustakis (2003)).
- The extent of the Exclusive Economic Zone and continental shelf rights, as well as research and rescue rights
- The extent of national airspace. Greece currently claims ten miles, of which Turkey recognises only six.
- Greek militarisation of eastern Aegean Sea islands that are close to Turkey
- Greek sovereignty over several small islets in the Aegean Sea
- Illegal immigration and refugee influx from the Turkish coast of the Aegean
- The Muslim minority of Western Thrace in Greece. The Muslim minority, predominantly of Turkish origin but comprising other ethnic groups as well, is persistently referred to by Turkey as 'Turkish', raising concerns over potential territorial claims to the region (Heraclides, 2019).

The Greek–Turkish conflict is argued to be an identity-based conflict, stemming from both countries' collective identities and national historical narratives, which are built upon a vilification of one another (see e.g. Heraclides, 2012). Mutual suspicion and fears arising from this portrayal of one another constitute a prime impediment to a reconciliation. Moreover, incapable and populist leaders as well as the reproduction of conflict through education (Heraclides, 1980; Keyder, 2005) serve to exacerbate disputes (Heraclides, 2012).

2.1.2 Growth and Defence Expenditure in Greece and Turkey

In 2018, Greece and Turkey were both reported as countries of very high human development, ranking 32nd and 59th respectively among 189 countries based on their Human Development Index (United Nations Development Programme/Human Development Report Office [UNDP/H-DRO], 2019). With a Gross Domestic Product of $23,547 per capita in 2018, Greece ranked below the European Union and OECD average (World Bank, 2020*b*), while Turkey was reported to have a per capita GDP of $15,069 in the same year. Greek population in 2018 was estimated at about 10.73 million (an increase of 28.82% since 1960); Turkish population was estimated at about 82.32 million (having grown by almost 200% since 1960).

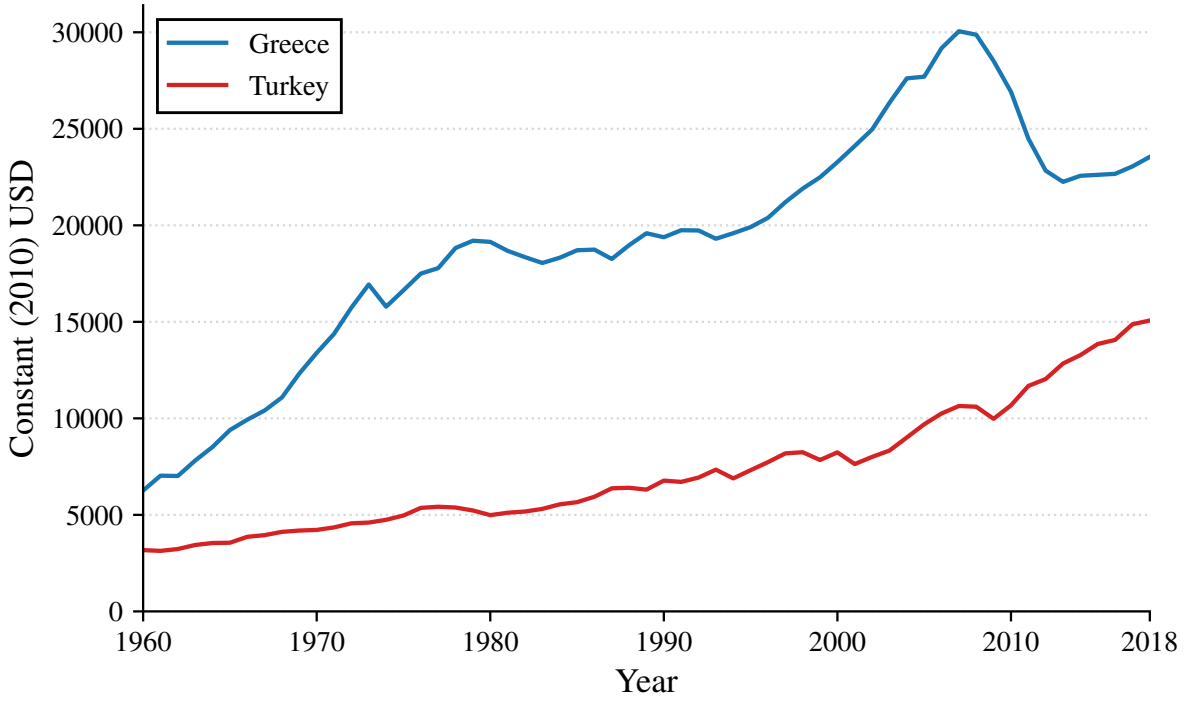

Figure 2.1: Per capita GDP for Greece and Turkey in constant 2010 USD (1960–2018). Source: World Bank (2020*b*)

The Greek economy underwent periods of accelerated growth, stagnation and crisis during years 1960–2018. Years 1960–1980 were characterised by a vast increase of the Greek Gross Domestic Product (GDP), which rose from 52.15 billion USD (constant 2010 values) to 184.59 billion (an increase of about 254%). The economy continued growing throughout the 1980s and 1990s (albeit at a lower pace), while from 1996 onwards fast-paced growth resumed. The Great Recession of 2007 and the Greek government-debt crisis led to a sharp downturn; real GDP declined from 332.06 billion USD in 2007 to 243.99 billionn USD in 2013. Years 2013–2018 were characterised by alternating periods of contraction and modest growth.

Public debt constitutes a thorn in the flesh of Greek economy. Despite increasing by only about 11% in terms of GDP during the period 1960–1980 (11.58% to 22.53%), it grew to 100.29% of GDP within the next 13 years. Public debt remained roughly unchanged during the following years, however the Great Recession led to a debt that amounted to 172% of GDP in 2011. Greece has since been struggling to finance its debt, which reached 181% of its GDP in 2018.

The Turkish economy grew significantly during the period 1960–2018, despite facing a series of crises. The country's GDP has been growing steadily since 1960, reaching 1,240 billion USD in 2018 (a 13-fold increase over the 1960 value of 87.23 billion USD). Turkey faced drastic recessions in 1980, 1994, 1999, 2001 and 2009; however, all of these recessions were characterised by a rapid resurgence of growth in the following years.

The Turkish economy has been characterised by long-standing hyperinflationary conditions. Inflation only reached a single-digit level in 2004 after 33 years of double-digit levels. Inflation levels remained mostly single-digit throughout the rest of 2000s and 2010s. However, in 2017 and 2018 inflation rose above the 10% mark, reaching 11.14% and 16.33% respectively. Per capita real GDP for both Greece and Turkey is presented in Figure 2.1.

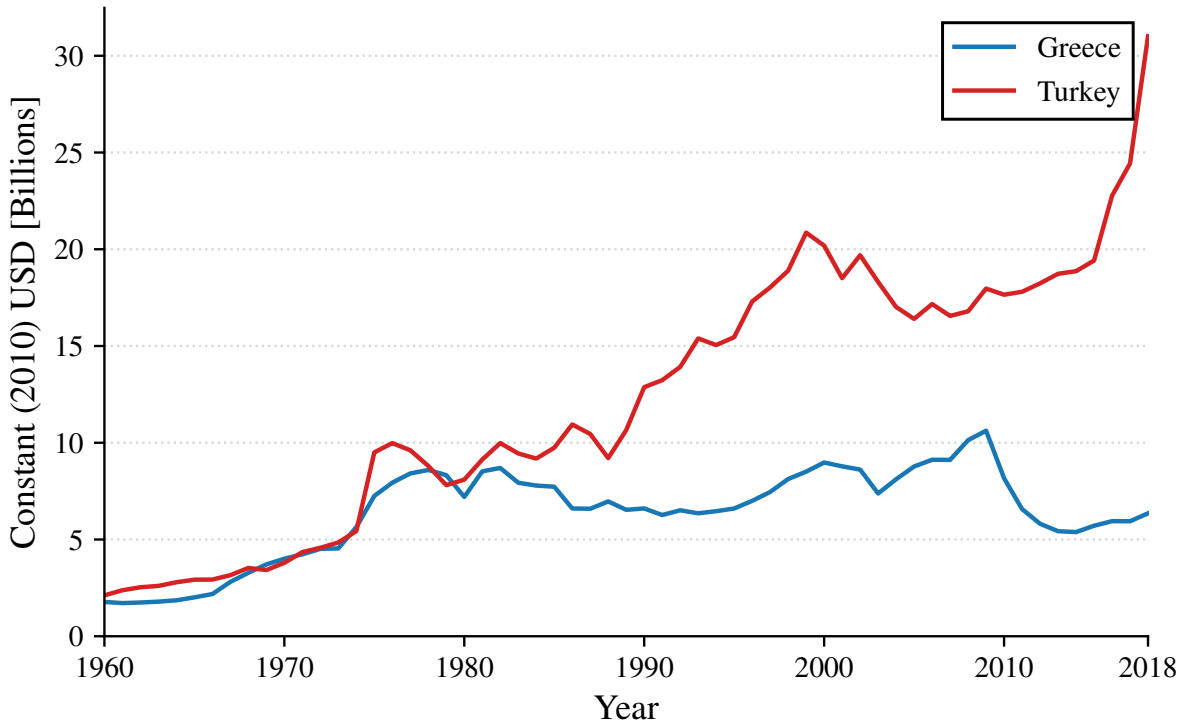

Figure 2.2: Annual military expenditure of Greece and Turkey in billions of constant 2010 USD (1960–2018). Source: Author's calculations based on Stockholm International Peace Research Institute [SIPRI] (2020)

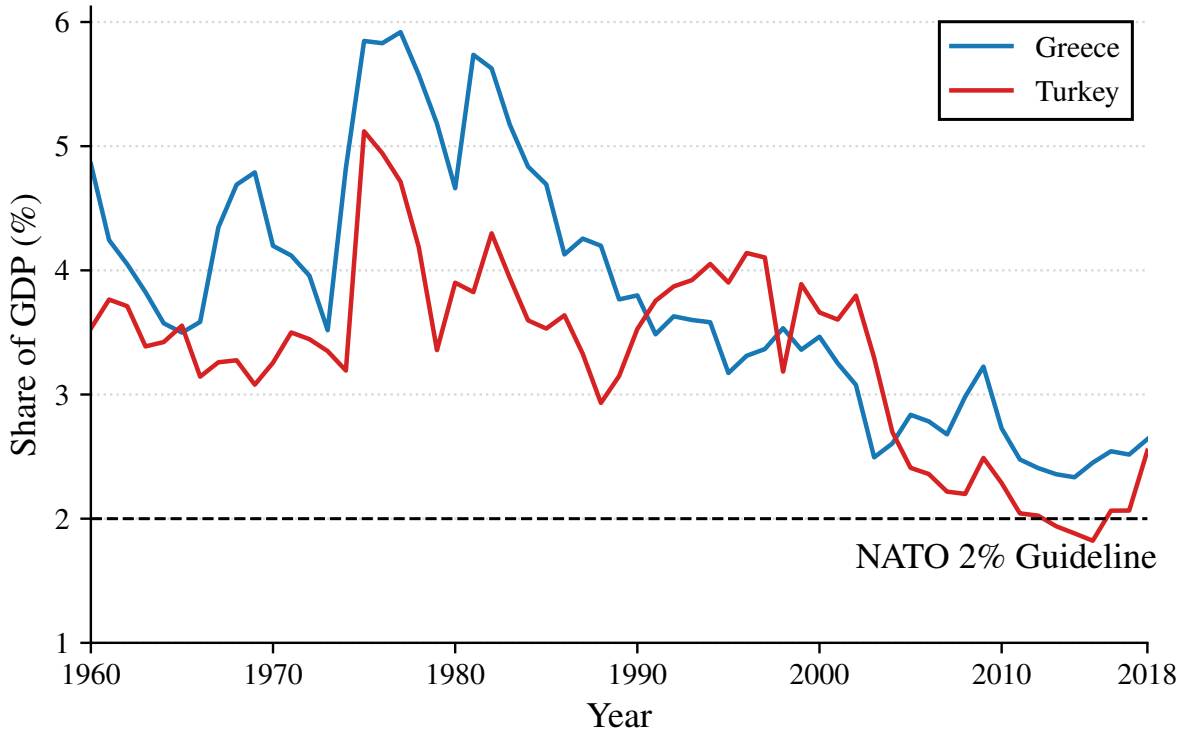

Figure 2.3: Annual military expenditure of Greece and Turkey as a share of GDP (1960–2018), including NATO 2% guideline established in 2006. The 2% guideline reflects security considerations made in 2006 and *is not* indicative of past conditions. Source: Stockholm International Peace Research Institute [SIPRI] (2020)

As far as military expenditure is concerned, Greek nominal military expenditure exhibited a vast increase during the periods 1967–1969 (70.21% in total; potentially due to military rule), 1974–1975 (59.76%; due to the Turkish invasion of Cyprus) and in 1981 (18.37%). The Greek economic crisis had a pronounced effect on military expenditure, which decreased about 45% within only 3 years (2009–2012). Greek spending was also notably reduced in 1980, 1986 and 2003. When examined as a share of GDP, Greek spending has consistently been over 2.33%, averaging at about 3.80%. In general, Greece, a member of NATO since 1952 and of the EU since 1981, has been reported to typically spend a greater ratio of its GDP on defence compared to the EU and NATO averages (Athanassiou et al., 2002; Kollias, 2018).

Turkish nominal military expenditure skyrocketed in years 1974–1975 (96.07% change in total), while notable increases were also reported in years 1970–1971 (27.1% in total), 1989–1990 (39.81% in total) and 2016–2018, after the failed coup attempt (59.73%). Significant decreases of over 10% were only reported in 1979 and 1988. Turkish spending as a share of GDP has also steadily been over 2%, with the sole exception of years 2013–2015. The average figure stands at 3.31%, close to that of Greece and among the highest averages in NATO.

Greek and Turkish expenditure graphs reveal notable similarities, at least up to the late 1980s, which may imply some degree of interdependence. From 1990 onwards, Turkish expenditure exhibits a steep upward trend. Greek expenditure exhibits a less pronounced increase as well, at least until late 2000s, when the economic crisis struck. A significant note that should be made is that the use of conscripts by both countries keeps personnel costs at an artificially lower level.

The annual expenditures of both Greece and Turkey are presented in Figure 2.2 (nominal values) and Figure 2.3 (share of GDP). Average annual change of spending is 2.7% for Greece and 5.3% for Turkey. The respective standard deviations of 9.8% and 5.3% are indicative of the high variance exhibited by both countries' expenditure. Military expenditure data also shows that the termination of Cold War did not have a distinct effect on Greek and Turkish spending.

Greece and Turkey, both NATO members since 1952, are bound to expenditure commitments. Since 2006, NATO defence ministers have committed to spending at least 2% of their GDP on defence (North Atlantic Treaty Organization [NATO], 2020). Although this is merely a guideline and not mandatory, NATO (2020) deems it a significant indicator of political will. Greece and Turkey have typically been spending more than 2% of their GDP on defence since the establishment of the guideline (with the marginal exception of years 2013–2015 for Turkey).

2.2 Arms Race Modelling

2.2.1 Introductory Remarks

The notion of arms races is well-known in research, especially in the realms of international relations and biology. Military arms races in particular are a recurrent topic of public discourse, bearing a mostly negative connotation, since they are commonly associated with conflict and war. Their existence can be traced back to ancient civilisations; it has been argued, however, that they have become a distinctive phenomenon only since the industrial revolution (Huntington, 1958).

Research on arms races includes a wide literature on many of their aspects: their causes and motives, their onset, manifestation and termination, their implications for the countries involved and for international security in general. Research on the aforementioned aspects is interdisciplinary, spanning from the fields of Political Science and Economics to that of Psychology. Nevertheless, arms races are a key issue lying at the root of strategic studies and, as such, they are closely linked to the notion of national security.

Arms races may be quantitative or qualitative. Quantitative races refer to increases or decreases in absolute armament levels, while qualitative races refer to changes in the characteristics of armaments or, in other words, to technological advances. As far as this distinction is concerned, Huntington (1958) suggests that qualitative races can be more perilous compared to quantitative ones. Gray (1971) argues that both quantitative and qualitative aspects are involved in arms races between countries.

2.2.2 Defining an 'Arms Race'

Arms races have been defined in many different ways in literature. Huntington (1958, p. 41) defines an arms race as 'a progressive, competitive peacetime increase in armaments by two states or coalitions of states resulting from conflicting purpose or mutual fears'. Steiner (1973, p. 5) perceives an arms race between two nations or sets of nations as 'repeated, competitive, and reciprocal adjustments of their war-making capacities'. Gray (1971, p. 40) defines it as 'two or more parties perceiving themselves to be in an adversary relationship, who are increasing or improving their armaments at a *rapid* rate and structuring their respective military postures with a *general* attention to the past, current, and anticipated military and political behaviour of the other parties'. As Wallace (1979) notes, most definitions are similar in that arms races are characterised by a *concurrent unusual increase* in the military expenditure of two nations, driven solely by the pressure of the military rivalry itself.

Given the definitions outlined above, determining what constitutes 'a rapid rate' or 'unusual increase' is crucial. The main approach adopted in literature (Diehl, 1983; Gibler et al., 2005; Rider et al., 2011; Sample, 1997) defines a rapid build-up as growth equal to or greater than 8% in either military expenditure or military personnel in each of three successive years. An alternative approach often cited in relevant literature is that of Horn (1987). According to this approach, an arms race exists when two criteria are met: first, the average growth of military expenditure in the period before a dispute is higher than the whole-study-period average; second, the average growth in the second half of the pre-dispute period is higher than the corresponding growth of the first half.

Despite the framework analysed above, arms race definitions have received criticism by researchers. Rattinger (1984) argues that the issue of properly defining arms races is neglected in literature and, as a result, there is no historical list of arms races based on specific standardised criteria. Instead, the absence of standardisation leads to researchers providing many different lists of cases. Rattinger also draws up a comprehensive list of points that need to be clarified in order to obtain a concise definition of arms races: first, what is a suitable proxy of armaments; second,

whether the perception of hostility and mutual directedness of military effort is an intrinsic part of the definition of an arms race; third, how the onset and termination of arms races can be precisely defined.

Finally, it is worth noting that some researchers recommend avoiding the term 'arms race' altogether due to its ambiguity (Bellany, 1975). Moreover, the politicisation of the term makes it subject to broader, less thoughtful interpretations, thus contributing to this ambiguity.

2.2.3 Quantitative Models of Arms Races

Arms race models can be subsumed within the broader context of conflict modelling. Quantitative analysis in this realm dates at least one century, when Lanchester (1916) devised his linear and square laws of attrition. Among such models, which consist of systems of differential equations, the best-known model is probably that of Richardson (1960*a*), which has inspired a multitude of econometric models. Arms races have also been modelled as a repeated two-person game, in which each participant chooses between a high and a low share of military expenditure (Smith et al., 2000). To overcome the inability of game-theory models to capture time dynamics, evolutionary games and ABMS have also been proposed as models for studying conflicts (Kadera et al., 2020).

Two types of models are prominent in early literature: action–reaction and domestic structure models. Action–reaction models consider competitive relations to be the driving force of arming, while domestic structure models deem internal factors as determinants of armaments (e.g. Senghaas, 1990). These two types of models are generally not considered mutually exclusive (Buzan, 1987); in fact, research has led to the prevalence of complex models that incorporate factors from both types of models (Batchelor et al., 2002).

Richardson's Bilateral Arms Race Model

One of the oldest and most influential models of defence expenditure is the arms race model proposed originally by Richardson (1919). The model became widely known through monographs published posthumously (Richardson, 1960*a*,*b*) as well as through the work of Boulding (1962).

Although Richardson's model is one of many arms race models, it constitutes the first formalisation of action–reaction arms race models and has been a point of reference for many years. Thus, a presentation of this model is of vital importance for a complete understanding of relevant literature.

In an attempt to examine arms races —which were thought to have significantly contributed to the outbreak of World War I— and interpret this complex issue through a simple and

elegant model, Lewis Fry Richardson developed a quantitative model using his mathematical skills as a physicist. In the case of two countries A and B, the model makes the following assumptions about the conditions that lead these two countries to either increasing or decreasing their armaments (Caspary, 1967):

- Country A increases its armaments in proportion to the level of armaments of country B due to a perception that it is militarily threatened. A similar reaction to the armaments of Country A is expected from Country B.
- The current level of armaments exerts a burden on the economy, which places a proportionate constraint upon additional spending.
- In the absence of a military threat from another country, a constant rate of arming is postulated, which is driven by ambitions, grievances, external pressure and other factors.

These ideas were mathematically represented using a system of linear differential equations as follows:

$$dx/dt = ky(t) - ax(t) + g \qquad (k > 0) \tag{2.1}$$

$$dy/dt = lx(t) - by(t) + h \qquad (l > 0) \tag{2.2}$$

where: x and y are the levels of armaments for countries A and B at time t
k and l ('defence coefficients') indicate the rate of increase in armaments as a response to a unit of armaments of the other country
a and b ('fatigue coefficients') indicate the extent of the burden incurred due to the costs involved in maintaining armaments
g and h ('grievance terms') indicate the level of armaments that remains constant, irrespective of the level of armaments of countries A and B and which may be driven by prejudices, grievances, ambitions etc.
dx/dt and dy/dt express the rate of change of x and y in the unit of time

It is apparent that the rate at which each of the countries changes its level of armaments is positively related to the fear linked with its opponent's arms level, negatively related to the burden placed on its economy by its own arms level and positively related to enduring ambitions or grievances. Richardson's arms race actors exhibit what is called *myopic behaviour*, in that they only react to present activities of their respective adversaries.

For an equilibrium to occur, there must be no change in armament levels for both opponents (i.e. $dx/dt = dy/dt = 0$); the following reaction functions are therefore yielded:

$$x = \frac{k}{a}y + \frac{g}{a} \tag{2.3}$$

$$y = \frac{l}{b}x + \frac{h}{b} \tag{2.4}$$

Defining stability and instability in the context of this model is quite straightforward. Assuming grievance terms g and h are positive, an equilibrium exists if the product of the slopes in equations (2.3) and (2.4) is less than unity, namely:

$$\frac{k}{a}\frac{l}{b} < 1 \implies kl < ab \tag{2.5}$$

This means that an equilibrium exists if the product of the 'fatigue coefficients' is greater than the product of the 'defence coefficients'. In this case, the equilibrium is stable and the point (x_0, y_0), at which it occurs, is as follows:

$$x_0 = \frac{hk + gb}{ab - kl} \qquad y_0 = \frac{gl + ha}{ab - kl} \tag{2.6}$$

In case the product of the slopes in equation (2.5) is more than unity (indicating that drivers of build-up exceed dampening factors), then an equilibrium is possible if grievance terms are negative, but this equilibrium would be unstable. In this case of instability, endless escalation as a concept could potentially describe a situation where escalation reaches a threshold in which its nature changes, i.e. a war breaks out.

Richardson (1960*a*, p. 12) carefully sets the boundaries of his research by stating that 'The equations are merely a description of what people would do if they did not stop to think.' He also makes a number of theoretical observations, which offer insightful perspectives on power distribution, cooperation, the psychological aspects of arms races and even defence burdens.

On a closing note, it should be noted here that Richardson and researchers thereafter extended their models to account for more than two rivalling countries. Since only the case of Greece and Turkey is examined in this thesis, related research for multi-country models is beyond the scope of the thesis.

Beyond Richardson: Criticism and Alternative Models

Richardson's arms race model, perhaps partly due to the simplicity in which he attempted to approach a rather complex issue, stimulated considerable research in the field. This research attempted to deal with shortcomings of the original model, introduce more sophisticated models that resemble reality more closely or apply the model to specific types of armaments.

Several interpretations of Richardson's equations have been employed. A common interpretation is transforming the differential equations into a first difference form to facilitate the use of econometric methods (Majeski and Jones, 1981):

$$x(t) - x(t-1) = ky(t-1) - ax(t-1) + g \tag{2.7}$$

$$y(t) - y(t-1) = lx(t-1) - by(t-1) + h \tag{2.8}$$

Empirical applications of Richardson's model have not been satisfactory. Studies have reported statistically insignificant coefficient estimates, unanticipated coefficient signs, unstable coefficients, poor measures of fit or significant changes when research is broken down to more periods (Majeski and Jones, 1981; Moll and Luebbert, 1980; Rattinger, 1975, 1976*b*, 1984). Results are also arguably sensitive to the exact measure of military expenditure being used and to other specification-related choices (Smith, 2020). Stoll (1982), using Monte Carlo simulations, concludes that Richardson's equations may lead to misleading inferences about arms acquisition processes. In fact, Richardson's validation process of his own model allows for alternative contradictory representations (Wagner et al., 1975).

Of course, criticism and shortcomings in Richardson's work should not serve to disregard his contribution to arms race modelling. Richardson's research on arms races is still being cited and commonly used as a point of reference. Applications of his model stretch way beyond the field of peace research; firms competition (Chalikias and Skordoulis, 2014) and autonomous vehicles (Riaz and Niazi, 2017) are notable examples of the last decade. The simplicity of Richardson's model has therefore proved to be both an advantage and a disadvantage: The model can be employed in many fields; however, drawing concrete conclusions based on empirical analysis has proved particularly challenging (Smith, 2020).

Having established that the empirical results of Richardson's model are not satisfactory, researchers have proposed a multitude a models that build upon the original model and its common interpretation illustrated above. Changes in the proposed models include, among others, the incorporation of additional variables (such as technology or economic constraints), different representations of armaments (manpower or major weapons inventories instead of using defence expenditures as per the original Richardson analysis) and lagging variables for more than one year (see e.g. Ferejohn, 1976; Hollist, 1977*b*; Intriligator, 1975; Luterbacher, 1975; Majeski and Jones, 1981; Rattinger, 1976*b*; Saaty, 1968; Schelling, 1966). Researchers stressed the simplicity of Richardson−type models and the consequent need to incorporate additional domestic, social or psychological factors (Moll and Luebbert, 1980; Rattinger, 1984).

Caspary (1967) proposes a non-linear model, incorporating a cost constraint, a desired

armaments level and scaling constants, in an attempt to introduce economic theory in Richardson's model. Defining stability in Caspary's model, which contains non-first degree terms and is therefore non-linear, is complicated. Nevertheless, as Luterbacher (1975) notes, in a non-linear arms race, stability or instability may manifest itself depending on the phase in which the race is and this is why non-linear models are more capable of describing the way in which arms races evolve.

Lambelet (1973) proposes a two-theatre armaments interaction model, differentiating between conventional and strategic forces and attempting to create indices of these forces. Later, Lambelet et al. (1979) introduce a semi-logarithmic reciprocal specification as a model that satisfactorily incorporates overall resource constraints.

Zinnes et al. (1976) propose allowing both positive and negative coefficients in Richardson's equations, while Intriligator and Brito (1984) propose a system of differential equations with the intention of modelling a missile arms race. Hill (1978) considers a system of Richardson differential equations, introducing time lags. Later, Majeski and Jones (1981) propose a linear autoregressive model similar to the one examined by Hill (1978) in order to account for the influence of past expectations. McCubbins (1983) proposes a different approach, arguing that arms races occur between weapon systems with incompatible policy goals.

Ostrom (1978) proposes a detailed 'reactive linkage' model in an attempt to uncover the dynamics behind the process of policy-making regarding US defence expenditure. However, this model comes with its own set of restrictions (McGinnis, 1991) and is tailored to a US-specific process.

Rattinger (1984) argues that the arms race question should be answered within the wider context of the determinants of military expenditure. This is the approach that has since dominated research, given the poor performance of action–reaction models.

Choosing suitable factors to model military expenditure has been a rather challenging task. Factors that qualify as military expenditure determinants may be the perceptions of threat, political factors, the existence of military industry, the race for resources between different government agencies and of course external determinants (Rattinger, 1984). A list of such factors that have been used in related literature includes, inter alia: GDP (Collier and Hoeffler, 2007; Dunne et al., 2008; Solarin, 2017; Sun and Yu, 1999), lagged spending (Dunne et al., 2003, 2008; Solarin, 2017), urbanisation (Gupta et al., 2001; Solarin, 2017), external aid (Collier and Hoeffler, 2007), trade openness (Dunne et al., 2008), participation in international war (Collier and Hoeffler, 2007; Dunne and Perlo-Freeman, 2003*b*; Dunne et al., 2008), civil war (Collier and Hoeffler, 2007; Dunne and Perlo-Freeman, 2003*b*), membership of NATO (Solarin,

2017), rival countries' spending (Seiglie, 2016), neighbours' spending (Flores, 2011), population (Dunne et al., 2008; Solarin, 2017), democracy index (Dunne et al., 2008; Solarin, 2017), trade levels (Dunne and Perlo-Freeman, 2003*b*; Seiglie, 2016), allies' spending (Flores, 2011; Seiglie, 2016), the security web (Dunne and Perlo-Freeman, 2003*a*; Dunne et al., 2008; Solarin, 2017), globalisation (Dunne et al., 2008; Solarin, 2017), length of borders and land area (Hewitt, 1992), public opinion (Eichenberg and Stoll, 2003), government form and political system (Albalate et al., 2012; Kim et al., 2013; Zuk and Thompson, 1982) and even CO_2 emissions (Mourad and Nehme, 2019).

An initial analysis of the different models used to model the demand for military expenditure can be found in Smith (1995). As in the case of earlier literature, methodology differs significantly, ranging from static and dynamic panel data analysis (Dunne and Perlo-Freeman, 2003*b*) to panel data regressions (Dunne et al., 2008), pooled data regression analysis (Collier and Hoeffler, 2007) and cross-country data (Dunne and Perlo-Freeman, 2003*a*).

Beyond classic econometric models, attention has also been drawn to highly-complex, chaotic systems as being able to capture a great deal of the complexity involved in real systems (Hill, 1992). However, simple non-linear deterministic models can arguably render predictability impossible (Saperstein, 1984).

Finally, another type of analysis employed by researchers is causality analysis. Majeski and Jones (1981), in their study of twelve dyads using causality analysis, conclude that two-way (linear) causality is not present in any of the cases, while also claiming that in some cases arms races are asymmetric, namely only one of the two nations is racing.

Statistical Considerations

Researchers have pointed out several shortcomings regarding the methodological approaches taken in arms race research. Rattinger (1976*a*) questions the validity of models proposed due to the problems of serial correlation and the large number of parameters. The issue of whether the statistical methods employed are suitable, considering that in many models there were only few annual observations but many parameters, is aptly illustrated in the old saying recalled by Luterbacher (1975, p. 213): 'with four parameters you can fit an elephant and with eight you can make him [*sic*] wiggle his [*sic*] tail'. This saying may sound exaggerated, but drawing an elephant with four parameters is actually possible, as proved by Mayer et al. (2010).

Researchers have also expressed concerns about excessive data manipulation (Luterbacher, 1975) and about the methods employed in general (Anderton, 1989; Diehl, 1983; Smith, 1989). Smith (1995) lays commonly overlooked considerations regarding Cochrane–Orcutt type trans-

formations, which are applied to correct serial correlation, while Dunne et al. (2005) highlight a series of statistical issues for unit root tests, co-integration tests and the vector autoregression framework (VAR).

Research on arms race modelling focusses for the most part on using measures of fit as the main criterion to judge how 'good' a model is. This disregards the fact that 'good measures of fit do not necessarily imply the right model', as Anderton (1989, p. 350) aptly states. To this point, Wallace and Wilson (1978), in their comparative test of models, conclude that three models are equally capable of explaining British military expenditure for the period 1870–1914.

Hollist (1977*a*), in a comparative test of eight models, illustrates that models exhibit varying explanatory power from country to country, thus concluding that 'models incorporating different mixes of independent variables are appropriate in different empirical contexts' (Hollist, 1977*a*, p. 339). Moreover, McGinnis (1991) stresses that drawing specific boundaries between the effects of explanatory variables is not always possible, as these effects may be indirectly influenced by several complex interactions. Anderton (1989) underlines the importance of examining the sensitivity of estimates to outliers.

Finally, the establishment of a model that accurately represents an arms race race relationship is also hindered by structural changes, which imply that parameters are not stable. Indeed, evidence of parameter shifts has been reported in literature (Cusack and Ward, 1981; Lucier, 1979).

Choosing Representative and Reliable Data

It is apparent that datasets play a significant role in drawing consistent conclusions. In this context, two common issues are the following: First, poor quality of available data; second, choosing datasets that correspond satisfactorily to the variables they describe (armaments in particular).

As far as poor quality is concerned, changing data sources may produce a substantial discrepancy in the results obtained (Brauer, 2007; Cusack and Ward, 1981). Lebovic (1999) reports considerable discrepancies between military expenditure data reported by SIPRI and those reported by World Military Expenditures and Arms Transfers (WMEAT). Discrepancies may also exist between revisions of the same source (Brauer, 2002, p. 90). Another interesting issue raised by Anderton (1989) is using inferior data when superior data are readily available.

Brzoska (1981) lists an array of reasons that impact the reliability of data on military expenditure. These include: the fact that institutions reporting military expenditure depend on government data beyond their control; corrections for inflation, exchange rates, etc.; modes

of data preparation; diverse reasons for producing information. In addition, data reported by countries may intentionally be inaccurate in order to mislead rivals or for domestic reasons.

If the purpose of arms race modelling is to study the behaviour of policy-makers, then identifying and using the correct source among many available sources (i.e. the source actually used by policy-makers) is pivotal and particularly challenging. Policy-makers may not always consult up-to-date or accurate sources of data to make decisions and the sources chosen may change throughout a large period of time. Of course, if policy-makers are believed to react to the real values of the variables, then the aforementioned issue is irrelevant.

In addition, it has been argued that decisions by policy-makers are not always rational or based on complete information (Jervis, 1976). Hammond (1993, p. 47) argues that it is 'subjective interpretations of the actions of others' that often underlie arming decisions. Lambelet (1986) also underlines the issue of misperceptions in arms races and war.

The question of a suitable proxy of armaments is also of great significance. If the reaction parameters in the model employed in this thesis are found to be insignificant, does this truly imply that an arms race does not exist? Could this instead be a indication that an inappropriate index was used?

Different measures of armaments have been used in literature. Among all measures, defence expenditure is the one that is predominantly used. This choice is not without issues (Bellany, 1975; Luterbacher, 1975; McCubbins, 1983; Rattinger, 1976*b*). Defence expenditure statistics include expenses purely unrelated to military capabilities, while they fail to capture qualitative aspects of armaments, such as procurement programmes and major decisions made by alliances (Rattinger, 1984), morale and courage, quality of planning, tactics, efficiency, level of equipment maintenance, quality and speed of information flow (Kollias, 1996). Dunne and Smith (2007) raise the issue that aggregate expenditure data may fail to capture cases where a country switches from employing conventional forces to developing nuclear forces or aiding terrorism to counter its adversary.

Even when the specific case of military expenditure is considered, it is unclear whether levels, logarithms, the change or share of military expenditure out of GDP or central government expenditure are more suitable measures. Although each choice comes with its own set of advantages and disadvantages, Brauer (2002, p. 88) argues that level data is more appropriate when testing for the (in)existence of an arms race, since they 'indicate actual or expected fighting capabilities of oneself vis-à-vis the putative adversary'. The choice of level data is also supported by Looney and Mehay (1990) and Gonzalez and Mehay (1990). Kollias (1996), on the other hand, argues that the most appropriate measure of military capability is the stock of military

capability, that the military expenditure finances. In this case, however, the multitude of weapon systems used by countries, fluctuations in quantity and quality, and the issue of depreciation pose an intractable problem in obtaining reliable estimates. In addition, using stock data disregards personnel training expenditure.

The use of aggregate expenditure data spawns additional complexities. If more than one non-correlated races occur simultaneously, then one may neutralise another, in which case aggregate levels would remain constant. Unfortunately, acquiring highly disaggregated data is not a perfect solution. Literature presents evidence that different components of the defence budget —at least in the case of the United States— are determined by different sets of explanatory variables (Mintz, 1988, 1989; Mintz and Hicks, 1984). This highlights another significant issue: Which sections of military expenditure should be included in a measure of armaments? For instance, should increases in military personnel salaries be included? On the one hand, such increases are not directly related to a country's capabilities. On the other hand, they may constitute a significant motive for the military personnel to increase its productivity.

Researchers have attempted to alleviate the issues arising from the use of military expenditure data by introducing strategic capability measures (Intriligator, 1975). The related studies, however, focus on countries that possess nuclear capabilities and, thus, it has been argued that the capability measures mentioned above may not be suitable for countries short of nuclear capabilities (Deger and Sen, 1983).

Overall, finding a perfect measure poses significant difficulties (Busch, 1970; Luterbacher, 1975). However, there seems to be a consensus among researchers that military expenditure is the best available measure of armaments (Brauer, 2002). Of course, the issue of data reliability remains and it should be given more attention by researchers.

Wider Theoretical Considerations

As is the case with many issues in social sciences, the very nature of the issue under examination poses restrictions to the types of quantitative analysis that can be carried out accurately and responsibly. In the early years of arms race modelling, Boulding (1962) and Rapoport (1957) question whether classical mathematics in the form of Richardson's differential equations could be adequately applied to arms races. Boulding (1962, p. 24) stresses that 'the classical apparatus of physical mechanical systems [...] has only a very limited applicability to social systems'. A similar statement is made by Rapoport (1957) about using classical analysis to model human behaviour.

A broader consideration refers to the theoretical foundation of the literature on arms

races. Wohlstetter (1974, p. 80) underlines the discrepancy between modelling group behaviour and modelling the results of this behaviour:

> The trouble with most arms race theories has been that they start by assuming an accelerating competition and then look about for some mechanism that might conceivably explain it – a simple pair of differential equations with an exponential solution (as in Richardson), worst case dynamics, explosive interservice rivalries, etc. It would be better to start, however, with the actual gross behavior of the parties in the competition.

Concerns have also been voiced about the suitability of empirical research. When examining empirical results, Buzan (1983, 1987) criticises approaches of arms races as not fitting within what is implied by the notion of 'racing', maintaining that arms racing literature does not focus on arduous, abnormal competitions between countries, but instead chooses a broader meaning. Empirical results indicating an inexistence of reaction in ostensibly evident cases of arms races have also led to scepticism about the whole notion of arms races (Moll and Luebbert, 1980; Zinnes, 1980).

Researchers who have attempted to summarise arms race literature seem to be dissatisfied by the approaches adopted and the knowledge gained through those (see e.g. Anderton, 1989; Brauer, 2002). Brauer (2002, p. 90) notes that '*post hoc* rationalization of one's findings is very easy', while Anderton (1989, p. 362) sheds light to the issue of publication bias, when urging journals to 'be willing to publish papers that report "insignificant" results'.

2.2.4 The Case of Greece and Turkey

The Greek–Turkish conflict has drawn substantial attention in defence economics literature, given the strained relations and perennial disputes between the two countries. The Turkish threat, more specifically Turkey's revisionism, is considered to be the main threat to Greek national interests (Athanassiou et al., 2002) and the decisive factor that drives increased military spending for Greece (Kollias, 2018).

Part of the relevant literature attempts to test for the presence of an arms race between Greece and Turkey using Granger causality analysis; Granger causality in this case can be considered as the statistical equivalent to Richardson's action–reaction model. All authors agree that two-way causality is what constitutes an arms race. However, results are inconclusive. Majeski (1985) reports that there is significant mutual interaction between Greece and Turkey for the period 1949–1975. So do Kollias and Paleologou (2002) and Dritsakis (2004) for periods

1950–1999 and 1960–2001 respectively. Strong evidence of an arms race is reported by Kollias and Makrydakis (1997) as well. On the other hand, Georgiou et al. (1996) find little empirical evidence that an arms race exists for the period 1960–1990 using a VAR specification. So do Paparas et al. (2016) for years 1957–2013. Brauer (2002), in his review, deems this approach as incapable of capturing complexities in Greek–Turkish spending dynamics, such as interactions, where high levels of Greek expenditure for reasons irrelevant to Turkey spawn a racing reaction from Turkey.

A handful of studies model arms races as an iterated two-person game, in which Greece and Turkey choose either a high or a low share of spending each year. Evidence for periods 1958–1997 (Smith et al., 2000) and 1958–2004 (Şahin and Özsoy, 2008) suggests that Greece and Turkey do not engage in an 'action–reaction' type of arms race and that a rather internal explanation of expenditure is more plausible.

Most studies employ econometric methods, yielding inconclusive results as well. A study by Georgiou (1990) for the period 1958–1987 does not corroborate the hypothesis of an arms race. However, the reliability of these results is questioned by Kollias (1994), who argues that the use of military expenditure as a share of GDP is not an appropriate measure to investigate the existence of an arms race.

Kollias (1991) applies Richardson's arms race model, which does not support the hypothesis of an arms race for periods 1950–1986 and 1974–1986. However, the estimations of the linear models that Kollias further employed suggest that Greek military spending is in fact affected by Turkish spending and the relative size of the arms forces.

Seiglie (1992) estimates a linear model of military expenditure using cross-sectional data for 55 countries (including Greece) for periods 1968–1971 and 1972–1976 and reaches the conclusion that opponents' military expenditure is statistically significant at the 5% level only for the period 1968–1971. Kapopoulos and Lazaretou (1993) report that, for the period 1962–1988, Turkish military spending exerts a notable effect on Greek security.

Kollias (1996), using data for the period 1960–1992, concludes that the Greek military expenditure is considerably more influenced by the Turkish one in the short and long run than it is by NATO spending. On the contrary, two decades later, Kollias et al. (2016) reports that Turkish military expenditure is not a significant determinant of Greek spending for years 1990–2014, with domestic factors exerting pronounced influence.

Avramides (1997) attempts to model Greek defence expenditure based on economic principles; he uses a Stone–Geary utility function and the Deaton–Muellbauer functional form for the period 1950–1989. He concludes that Greece responds to Turkish defence spending in

the long run as well as that it has been free-riding on its NATO allies for the period prior to 1974.

Dunne and Nikolaidou (2001) perform a linear regression on the change of Greek military expenditure for years 1960–1996 and report that the lagged value of change of Turkish spending has a positive and significant effect. Five years later, Dunne et al. (2005) consider a variety of specifications and estimation techniques, including classic Richardson-type models, a VAR specification as well as different measures of armaments, concluding that an an arms race between Greece and Turkey is rather implausible. Researchers stress, however, that results are highly sensitive to sample and specification.

Öcal (2002) employs a Smooth Transition Regression model for the period 1956–1994, which suggests the existence of asymmetric effects between Greek and Turkish expenditure, with increases in Turkish expenditure carrying more substantial effects on Greek expenditure than the opposite. Evidence of an asymmetric long-run relationship is also reported by Öcal and Yildirim (2009), using threshold co-integration analysis for the period 1956–2003.

Using an Autoregressive Distributed Lag model (ARDL) for the period 1960–1998, Kollias and Paleologou (2003) report that the Turkish expenditure exerts a positive and significant effect on Greek spending. The change of NATO spending is also reported to have a positive and significant effect. Nikolaidou (2008) employs an ARDL model as well for the period 1961–2005, including a dummy variable for Cyprus invasion in 1974. The Turkish military expenditure is found to be a statistically significant explanatory variable for Greek military expenditure.

The issue of what constitutes a proper measure of armaments and how this choice affects results is also apparent in Greek–Turkish arms race literature. A case in point is that of Georgiou et al. (1996) and Kollias and Makrydakis (1997) mentioned above, where the former uses shares of expenditure in GDP and the latter uses levels of expenditure. Avramides (1997, p. 173) also notes that 'levels of defence expenditures and their GDP shares cannot be taken as similar measures of intentions or perceptions'.

As far as statistical considerations are concerned, Smith (1998) stresses the importance of visualising data before applying statistical methods. Using this as a starting point, Brauer (2002) notes that a simple five-year difference in the base year used for military expenditures leads to great differences in the resulting time series, which spawns scepticism about the validity of the results. Brauer also criticises pre-statistical work as being too weak.

In general, Brauer (2002, p. 93) concludes that 'In a word, as in the case of Turkey, we do not know much at all about the determinants of Greek military spending either.' Brauer (2002, p. 101) stresses 'how little it [related literature] concerns itself with *political economy* and how much of the literature narrowly sticks to *pure* economics, *mathematical statistics*, and

econometrics'. Nevertheless, Andreou and Zombanakis (2003) make a bold claim that an arms race between Greece and Turkey is a reality, maintaining that indications to the contrary in relevant literature should be attributed to weaknesses in the statistical techniques applied.

Neural Networks Approach

A notable exception in earlier Greek–Turkish arms race literature is that of Refenes et al. (1995), who adopt an artificial neural networks approach to gain a better insight into Greek defence expenditure. The results they obtain through this method suggest that the ratio of Greek to Turkish armed forces (Turkish quantitative advantage) and the Turkish military spending per soldier are the most significant explanatory variables of the growth rate of Greek spending.

The approach of using neural networks is also adopted later by Andreou and Zombanakis (2000), who attempt to differentiate between financial and human resources and nevertheless report that, in all cases, variables pertaining to Turkey constitute top determinants of Greek defence expenditure. In addition, they call for attention to the human resources factor of the arms race, stressing the demographic downturn witnessed in Greece in comparison to the increasing Turkish population.

Andreou and Zombanakis (2011), in an attempt to revisit research by Andreou and Zombanakis (2000), develop different neural networks for Greece and Turkey for the period 1961–2008. An input significance analysis of the results suggests that the three main determinants of Greek spending are the Turkish spending as a share of GDP, the Turkish per capita spending and the Greek per capita spending. The corresponding factors for Turkish spending are the Turkish armed forces personnel per 1000 inhabitants, the Turkish per capita spending and the Turkish GDP rate of growth. Although the network employed for the Turkish expenditure exhibits signs of overfitting, the results indicate that racing is unidirectional on the side of Greece, which corroborates the results of Andreou and Zombanakis (2000). Andreou and Zombanakis (2011), as well as Katsaitis et al. (2019) later, reiterate the emphasis on the human resources parameter.

Unfortunately, although the neural networks approach was proposed almost two and a half decades ago, it has only scarcely been examined in related literature, despite the extensive use of traditional econometrics. This may be attributable to the 'black box' nature of neural networks as well as to the advanced knowledge required for reliable input significance analysis.

2.3 Artificial Neural Networks

2.3.1 Introductory Remarks

In recent years, there has a been a rapid growth in Artificial Intelligence (AI) and machine learning in the field of data analysis and intelligent applications development (Sarker, 2021). Big data availability, cloud computing, increased computing power and improved hardware are considered to be some of the drivers of this growth. Modern AI methods are employed in many activity sectors (Russell and Norvig, 2016), attaining high levels of performance in complex problem-solving and, thus, becoming particularly significant for the future of human society (West, 2018). This increasing significance of AI has also led to a fierce competition between governments, which aim to leverage AI in order to increase their economic power and influence (Buchanan, 2019).

Artificial neural networks (hereafter simply referred to as 'neural networks') are a machine learning method inspired by the functioning of biological neurons. Originally formalised by McCulloch and Pitts (1943), neural networks as a field of research have enjoyed great popularity in recent years due to a growing interest in data analysis and machine learning models.

Modern applications of neural networks extend over a wide range of disciplines. A —by no means exhaustive— list of applications includes, inter alia, autonomous driving (Wu et al., 2017), vessel route prediction (Zissis et al., 2015), quantum chemistry (Balabin and Lomakina, 2009), game playing (Silver et al., 2016), biological signal classification (Sengupta et al., 2016), disease diagnosis (Ganesan et al., 2010), infrastructure reliability analysis (Nabian and Meidani, 2018), coastal engineering (Dwarakish et al., 2013), drought forecasting (Mishra and Desai, 2006), face recognition (Lawrence et al., 1997), speech recognition (Abdel-Hamid et al., 2014), handwriting recognition (Graves and Schmidhuber, 2009), sign language recognition (Adithya et al., 2013), classification of Android phone malware (Nix and Zhang, 2017), credit card fraud detection (Fu et al., 2016), spam e-mail detection (Clark et al., 2003), criminal recidivism (Caulkins et al., 1996) and algorithmic trading (Sezer and Ozbayoglu, 2018).

Concerning time series analysis, neural networks offer several advantages over traditional statistical methods. Neural networks are proved to be universal approximators of functions (Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989) and their derivatives (White et al., 1992), while also being capable of approximating ordinary least squares and nonlinear least squares regression (White, 1992*b*; White and Stinchcombe, 1992) as well as nonparametric regression (White, 1992*a*). As White (1992*c*, p. 79) aptly states, 'neural networks are capable in principle

of providing good approximations to just about anything one would like.' Neural networks are also able to provide adequate estimations of linear functions (White, 1992*a*,*b*; White and Gallant, 1992; White and Stinchcombe, 1992), requiring almost no a priori assumptions about the underlying process being modelled (Refenes et al., 1995). Moreover, neural networks can form different functions in different segments of the sample space (Wasserman, 1989).

2.3.2 Fundamental Notions

Artificial Neuron

The basic units, of which neural networks are comprised, are called neurons. Neurons are modelled in a way that mimics their biological counterparts. A typical artificial neuron comprises input paths (corresponding to a biological neuron's dendrites), output paths (corresponding to a biological neuron's axons) and propensities which influence its output (as in biological neurons). The neuron combines input signals, including the effects of propensities (bias), and produces an output signal.

Signal flow inside the typical artificial neuron is illustrated in Figure 2.4. Each input path i carries a signal x_i and has a strength represented by a weight w_i. The neuron calculates the weighted sum of all input signals plus bias b. The total input z is therefore:

$$z = \sum_i x_i w_i + b = w^T x + b \tag{2.9}$$

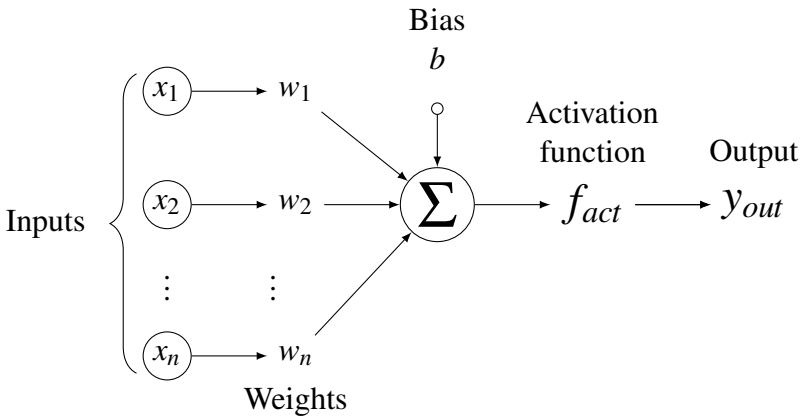

Figure 2.4: Illustration of signal flow inside an artificial neuron

The output signal (y_{out}) of a neuron is a —usually nonlinear— transformation of total input z. This transformation is performed through the application of a function $g(\cdot)$, called an activation function. Commonly used activation functions include the following:

Binary step function:

$$y_{out} = \begin{cases} 0 & \text{if } z < 0 \text{ (threshold)} \\ 1 & \text{if } z \geqslant 0 \end{cases} \tag{2.10}$$

Logistic sigmoid function:

$$y_{out} = \frac{1}{1+e^{-z}} \tag{2.11}$$

Hyperbolic tangent function (tanh):

$$y_{out} = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \tag{2.12}$$

Rectified linear unit function (ReLU):

$$y_{out} = \begin{cases} 0 & \text{if } z \leqslant 0 \\ z & \text{if } z > 0 \end{cases} \tag{2.13}$$

Artificial neurons (such as those described above) combine to form neural networks. Neural networks usually comprise an input layer, an output layer and one or more hidden layers in-between. Each of these layers comprises multiple neurons, which are connected to other neurons in adjacent layers.

Modern neural networks appear in a wide variety of topologies. These topologies may constitute accurate representations of biological neural networks (e.g. Gluck and Bower, 1988; Granger et al., 1989), but they may as well deviate considerably from biological functioning (e.g. Rumelhart and McClelland, 1986). Beyond Multilayer Perceptrons, which will be analysed later, modern forms of neural networks include:

- Other types of feed-forward networks, such as Radial Basis Function networks (RBF), Convolutional Neural Networks (CNN), Autoencoders, Probabilistic networks and Deep stacking networks
- Recurrent Neural Networks (RNN), including Long Short-Term Memory networks (LSTM), Hopfield networks and Boltzmann machines
- Modular networks, which comprise several small networks
- Other models, such as Generative Adversarial Networks (GAN), Support Vector Machines (SVM), Wavelet Neural Networks and Neural Turing Machines

Learning Process

Learning in neural networks occurs through an iterative process of readjusting the weights (and bias, if included) until the network produces the desired output y within a given tolerance; in other words, until the defined cost function is minimised. Statistically, learning can be described as follows (Zapranis and Refenes, 1999): Given a sample $S_n = (x_i, y_i), i = 1, \ldots, N$, derived from an unknown function $\sigma(x)$ in which a stochastic component ε with zero mean is added, learning consists in determining an estimator $g(x; w) = \hat{\sigma}(x)$ of $\sigma(x)$, where w represents network weights. As no assumptions are made in advance concerning the functional form of $\sigma(x)$, the neural network is a non-parametric estimator of the conditional density $E[y|x]$.

Learning methods are typically categorised as follows (Du and Swamy, 2019):

- Supervised learning, whose task is to infer a function $f : X \to Y$ from a given training dataset $\{(x_i, y_i)|i = 1, \ldots, N\}$, where $x_i \in X$ is a training example and $y_i \in Y$ is the known label of x_i. The learning process is driven by a measure of discrepancy between estimated and real outputs. Supervised learning is commonly used for optimisation, classification and signal processing.

- Unsupervised learning, whose task is to capture significant patterns from unlabelled input data, based on correlations exhibited between these data. It is commonly used for clustering, feature extraction and signal coding.

- Reinforcement learning, which concerns itself with how an artificial agent ought to take action in order to maximise the cumulative expected reward. Reinforcement learning constitutes a special case of supervised learning, in which the desired output is not exactly known. Instead, only information about the success or failure of an answer is provided. Reinforcement learning is commonly used in control theory.

Other learning methods include semi-supervised learning (in which unlabelled data are jointly used with labelled data), ordinal regression, manifold learning, transfer learning, multi-view learning, multilabel learning and multiple-instance learning.

Several caveats apply regarding the learning process and its ability to generalise: First, input data may be imprecise or insufficient to infer a mapping of inputs to outputs. Second, the network may overfit; that is, the network may produce excellent results on training data but perform poorly on unseen test data. This situation is usually the result of training with too many examples, features or epochs, as well as the result of improper data sampling and splitting methods.

Several methods may be used to control generalisation and avert overfitting. These include early stopping (Prechelt, 1998), regularisation (Du and Swamy, 2019) and dropout (Srivastava et al., 2014).

2.3.3 Multilayer Perceptrons

Multilayer Perceptrons (MLP) are a commonly used class of feed-forward neural network models. They consist of an input layer, an output layer and one or more hidden layers. These layers comprise several units, which are fully interconnected to units in the adjacent layer. Data flows from the input to the output layer only (feed-forward). The architecture of MLPs is illustrated in Figure 2.5. The network presented in the figure consists of an input layer with S_1 input units, an output layer with S_M output units and $(m-1)$ hidden layers, each consisting of S_m hidden units.

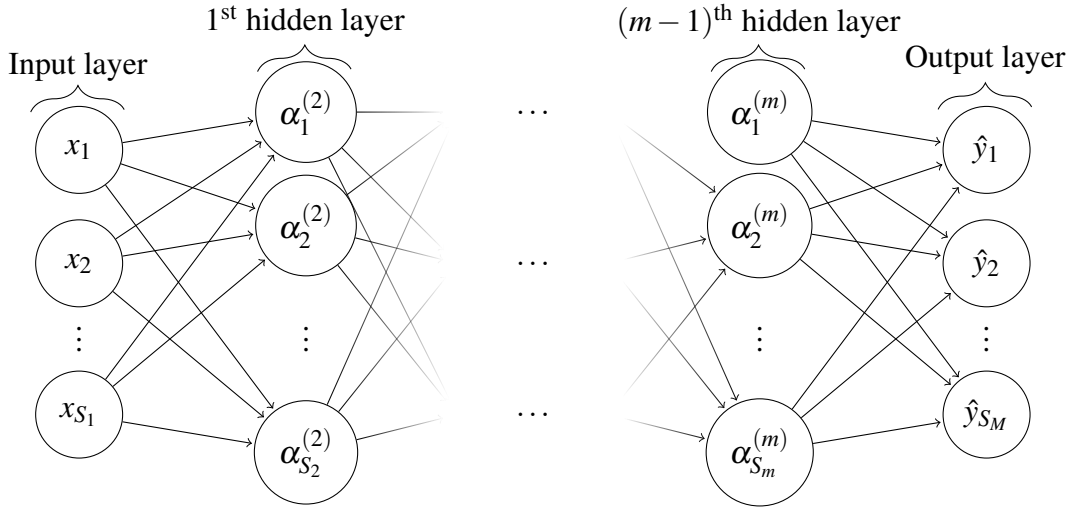

Figure 2.5: Representation of a fully interconnected $(M = m+1)$-layer neural network. Each layer consists of $S_m, m = 1,\ldots,M$ units.

Using the following notation:

- x is the input vector and $\hat{y}$ is the output vector
- S_m is the number of units in layer $m = 1,\ldots,M$
- $W^{(m-1)}$ is a matrix of order $S_{m-1} \times S_m$, which contains the values of the weights that connect units of layer $m-1$ to units of layer m $(m = 2,\ldots,M)$
- $b^{(m)} = \left(b_1^{(m)},\ldots,b_{S_m}^{(m)}\right)^T$ is the bias vector $(m = 2,\ldots,M)$
- $\alpha^{(m)}$ is the output vector of units in layer $m = 2,\ldots,M$ and $\alpha^{(1)} = x$ and $\alpha^{(M)} = \hat{y}$

- $g^{(m)}(\cdot)$ denotes the activation function, applying $g_i^{(m)}(\cdot)$ to the i^{th} component of the vector within $(m = 1, \ldots, M)$

for $m = 2, \ldots, M$ we obtain:

$$z^{(m)} = \left[W^{(m-1)}\right]^T \alpha^{(m-1)} + b^{(m)} \tag{2.14}$$

$$\alpha^{(m)} = g^{(m)}\left(z^{(m)}\right) \tag{2.15}$$

Activation functions are typically selected to be the same for all units in a layer. In addition, it is common that an activation function is chosen for the first $M-1$ layers and another function is chosen for layer M.

A considerable advantage of MLPs is that they are universal approximators. It has been mathematically proved that an MLP with a single hidden layer, in which a sigmoidal activation function is used, is capable of approximating any continuous multivariate function (Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989; Xiang et al., 2005). (Huang, 2003; Tamura and Tateishi, 1997) also examine the universal approximation capability of two-hidden-layer MLPs.

Backpropagation Learning Algorithm

Backpropagation, short for *backpropagation of errors*, is a well-known and popular learning algorithm for supervised learning tasks (Rumelhart and McClelland, 1986), despite being biologically improbable (Du and Swamy, 2019). Backpropagation employs a gradient-search method to minimise a loss function, which approximates the discrepancy between network output values and desired values (Du and Swamy, 2019).

The backpropagation algorithm can be simply described as follows: Given an input pattern, network weights are randomly initialised and a feed-forward pass is performed, producing an output pattern. This output pattern is then compared to a target pattern and the discrepancy (error) between actual and desired values is calculated for each output unit. This error is then propagated backwards and network weights are adjusted in a direction that minimises the error.

The backpropagation algorithm demands a continuous, nonlinear, monotonically increasing, differentiable activation function (Du and Swamy, 2019). In this thesis, the backpropagation algorithm is presented in an MLP, such as the one illustrated in Figure 2.5.

In regression problems —such as the problem examined in this thesis— the cost function C that is commonly used to measure the discrepancy between actual network output $\hat{y}_i$ and desired output y_i is the mean squared error (MSE):

$$C = \frac{1}{N}\sum_{i \in S} C_p = \frac{1}{N}\sum_{p \in S} \left\| \hat{y}_p - y_p \right\|^2 \tag{2.16}$$

where N is the sample size, y_p is the desired output for element p, $\hat{y}_p$ is the calculated network output for element p, S is a set comprising all training pattern pairs (x_p, y_p) and:

$$C_p = \left\|\hat{y}_p - y_p\right\|^2 = c_p^T c_p \tag{2.17}$$

$$c_p = \hat{y}_p - y_p \tag{2.18}$$

where the i^{th} element of c_p is $c_{p,i} = \hat{y}_{p,i} - y_{p,i}$.

Network weight matrix $W^{(m-1)}$ and bias vector $b^{(m)}$, $m = 2,\ldots,M$ can be merged in a matrix $W = \left[w_{ij}\right]$. Error function C is typically minimised by applying a gradient-descent optimisation algorithm; weights are adjusted according to the following formula:

$$\Delta_p W = -\eta \frac{\partial C_p}{\partial W} \tag{2.19}$$

where η is a small positive number called learning rate.

The derivative of (2.19) can be calculated by applying the chain rule:

$$\frac{\partial C_p}{\partial w_{uv}^{(m)}} = \frac{\partial C_p}{\partial z_{p,v}^{(m+1)}} \frac{\partial z_{p,v}^{(m+1)}}{\partial w_{uv}^{(m)}} \tag{2.20}$$

The second factor of (2.20) is calculated as follows:

$$\frac{\partial z_{p,v}^{(m+1)}}{\partial w_{uv}^{(m)}} = \frac{\partial}{\partial w_{uv}^{(m)}} \left(\sum_{\omega=1}^{S_m} w_{\omega v}^{(m)} \alpha_{p,\omega}^{(m)} + b_v^{(m+1)} \right) = \alpha_{p,u}^{(m)} \tag{2.21}$$

The first factor of (2.20) can be calculated by applying the chain rule:

$$\frac{\partial C_p}{\partial z_{p,v}^{(m+1)}} = \frac{\partial C_p}{\partial \alpha_{p,v}^{(m+1)}} \frac{\partial \alpha_{p,v}^{(m+1)}}{\partial z_{p,v}^{(m+1)}} = \frac{\partial C_p}{\partial \alpha_{p,v}^{(m+1)}} \dot{g}_v^{(m+1)} \left(z_{p,v}^{(m+1)} \right) \tag{2.22}$$

For units in the output layer it is:

$$\frac{\partial C_p}{\partial \alpha_{p,v}^{(m+1)}} = c_{p,v}, \quad m = M-1 \tag{2.23}$$

while for units in hidden layers it is:

$$\frac{\partial C_p}{\partial \alpha_{p,v}^{(m+1)}} = \sum_{\omega=1}^{S_{m+2}} \frac{\partial C_p}{\partial z_{p,\omega}^{(m+2)}} w_{v\omega}^{(m+1)}, \quad m = 1,\ldots,M-2 \tag{2.24}$$

Thus, if a delta function is defined by:

$$\delta_{p,v}^{(m)} = -\frac{\partial C_p}{\partial z_{p,v}^{(m)}}, \quad m = 2,\ldots,M \tag{2.25}$$

then, by substituting (2.22) and (2.23) in (2.25), for units in the output layer $(m = M-1)$ it is:

$$\delta_{p,v}^{(M)} = -c_{p,v} \dot{g}_v^{(M)} \left(z_{p,v}^{(M)} \right) \tag{2.26}$$

and by substituting (2.22) and (2.24) in (2.25), for units in the hidden layers $(m = 1, \ldots, M-2)$ it is:

$$\delta_{p,v}^{(m+1)} = \dot{g}_v^{(m+1)}\left(z_{p,v}^{(m+1)}\right)\sum_{\omega=1}^{S_{m+2}} \delta_{p,\omega}^{(m+2)} w_{v\omega}^{(m+1)} \tag{2.27}$$

Therefore, the derivative in (2.20) can be rewritten as:

$$\frac{\partial C_p}{\partial w_{uv}^{(m)}} = \frac{\partial C_p}{\partial z_{p,v}^{(m+1)}}\frac{\partial z_{p,v}^{(m+1)}}{\partial w_{uv}^{(m)}} = -\delta_{p,v}^{(m+1)}\alpha_{p,v}^{(m)} \tag{2.28}$$

Biases may be updated in one of two ways: They may be treated as special weights from values permanently set to unity. Alternatively, a gradient-descent method w.r.t. $b^{(m)}$ may be employed, by applying the procedure analysed above. Given that biases can be treated as weights, they are typically disregarded in applications (Du and Swamy, 2019).

Optimisations

There are three types of gradient descent:

- Batch gradient descent: In batch gradient descent, the gradient of the cost function is computed for the entire training dataset for one weight update to be performed. As such, convergence is slow and it may even be intractable for large datasets that do not fit in memory.
- Stochastic gradient descent: In stochastic gradient descent, only a single, randomly chosen, training example is used for gradient calculation and weight update. As weights are updated for each training sample, the cost function tends to exhibit significant fluctuations.
- Mini-batch gradient descent: In mini-batch gradient descent, the gradient of the cost function is computed for a mini-batch of n training examples. Mini-batch sizes typically range between 50 and 256, but they may vary depending on the dataset. Mini-batch gradient descent combines the advantages of batch and stochastic gradient descent, offering a stable and fast convergence.

A shortcoming of gradient descent is that the error surface is multi-dimensional and may therefore contain numerous local minima. As a result, training the network often requires experimentation with different initial weights, adjusting the learning rate, or adding a momentum term to avoid being trapped in suboptimal local optima and to achieve better convergence performance.

For instance, adding a momentum term changes the backpropagation algorithm illustated in (2.19) as follows:

$$\Delta_p W(t) = -\eta \frac{\partial C_p(t)}{\partial W(t)} + \alpha \Delta W(t-1) \tag{2.29}$$

where α is the momentum factor, typically $0 < \alpha \leqslant 1$.

In modern applications of neural networks, commonly used gradient descent optimisation algorithms include: Adagrad (Duchi et al., 2011), Adadelta (Zeiler, 2012), RMSProp (Hinton, 2014), Adam (Kingma and Ba, 2014) and AMSGrad (Reddi et al., 2019). Recently proposed optimisers include AdamW (Loshchilov and Hutter, 2019), QHAdam (Ma and Yarats, 2019) and AggMo (Lucas et al., 2019).

Algorithms based on gradient descent are first-order learning algorithms. Another broad category of learning algorithms are the second-order methods, such as quasi-Newton or variable metric algorithms (e.g. Davidon–Fletcher–Powell, Broyden–Fletcher–Golfarb–Shanno algorithms) and conjugate gradient methods (e.g. Fletcher–Reeves, Polak–Ribiere algorithms). Polak (1991), Press et al. (1992) and van der Smagt (1994) offer an extensive review of such methods.

Adam Optimiser

Adam (Kingma and Ba, 2014), short for *adaptive moment estimation*, is an adaptive learning rate optimisation algorithm; this means that it computes individual learning rates for different parameters. To adapt the learning rate for each weight of the neural network, Adam uses estimations of the first and second moments of the gradients (i.e. the mean and uncentred variance). More specifically, on each iteration, the algorithm calculates an exponential moving average of the gradient and the squared gradient as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\, g_t \tag{2.30}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\, g_t^2 \tag{2.31}$$

where: m_t and v_t are moving averages for current iteration t; they are initialised with zeros on first iteration

g_t is the gradient on current iteration t

β_1 and β_2 are proposed parameters that control the decay rates of the moving averages; their recommended default values are 0.9 and 0.999 respectively

The default values of β_1 and β_2 as well as the initialisation of moving averages with zeros result in a bias of moment estimates towards zero. The bias-corrected estimates are as

follows:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{2.32}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{2.33}$$

Given the estimates above, the Adam update rule is as follows:

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \tag{2.34}$$

where: w represents model weights

η is the step size; its recommended default value is 0.001

ε is an adequately small number to prevent division by zero; its recommended default value is $1e-8$

Adam optimiser is a commonly used algorithm in the field of deep learning and it is typically recommended by researchers (see e.g. Ruder, 2016). The authors of the original paper also present empirical evidence that Adam compares favourably to other optimisation algorithms by being memory and computationally efficient.

Despite being particularly popular and bearing many desirable properties, Adam bears some disadvantages. It may not converge to an optimal solution for some tasks, while Wilson et al. (2017) and Keskar and Socher (2017) have showed that, for some tasks, stochastic gradient descent exhibits better generalisation capabilities. Researchers have attempted to deal with the shortcomings of Adam by proposing improved optimisers, such as AdamW (Loshchilov and Hutter, 2019) and QHAdam (Ma and Yarats, 2019).

2.3.4 Input Significance Estimation

The use of neural networks has been shown to lead to better predictions compared to traditional approaches. However, neural networks are often perceived as 'black boxes', because they provide little insight into their internal decision-making process. While their predictive ability per se is useful, as machine learning models get deployed across many critical industries (such as finance, healthcare or defence, where a false positive may have far-reaching consequences), the perception of machine learning models as black boxes as well as problems with bias and susceptibility to attacks lead to a mistrust to these models.

Concerning bias (the fact that models may include an imprint of the unconscious biases of their developer), Angwin et al. (2016), for instance, have shown that predictions made by a widely used criminal risk assessment tool are racially biased. Another similar example is that of

Caliskan et al. (2017), who demonstrate that applying machine learning may replicate semantic biases.

Humans are also generally reluctant to adopt methods not directly interpretable and tractable (Zhu et al., 2018). This leads to simple models being preferred for their ease of interpretation, although they may be less accurate than more complex ones.

Another concern is that of attack methods that have recently emerged. Imperceptible alterations hidden in deep neural networks have been proved to affect many different types of networks, causing them to make targeted errors (Carlini and Wagner, 2017; Moosavi-Dezfooli et al., 2017; Papernot et al., 2016).

The aforementioned considerations about interpretability, bias and unintentional discriminatory behaviour stress the need for an insight into the way in which models make decisions. The field which concerns itself with these problems is called Explainable AI. The goals to which different applications of Explainable AI cater are the following (Barredo Arrieta et al., 2020): trustworthiness, causality, transferability, informativeness, confidence, fairness, accessibility, interactivity and privacy awareness.

Although Explainable AI was developed as an attempt to tackle the issues analysed above, in the future, it might actually be a legal obligation to explain how models function, especially if these are broadly deployed and have an effect on significant decisions. A case in point is that of the European Union, which introduced a form of 'right to explanation' in its General Data Protection Regulation (GDPR), which came into force in 2018 (Goodman and Flaxman, 2016).

Two significant remarks should however be made. First, unequivocally not trusting a model because its functioning is not completely comprehended does not constitute sound advice for researchers or practicioners. Testing on unseen data is probably a better basis for trust. Second, there is an ongoing debate in recent literature about how different models cater to what humans consider a 'good' explanation, which is the role of human intuition in model results evaluation, or whether explainability is always helpful in a task-specific setting (see e.g. Kumar et al., 2020; Passi and Jackson, 2018). Poursabzi-Sangdeh et al. (2018), for instance, make an interesting note about the evaluation of 'interpretable' models:

> Participants who were shown a clear model with a small number of features were better able to simulate the model's predictions. However, contrary to what one might expect when manipulating interpretability, we found no improvements in the degree to which participants followed the model's predictions when it was beneficial to do so. Even more surprisingly, increased transparency hampered people's ability to

> detect when the model makes a sizable mistake and correct for it, seemingly due to information overload.

Proceeding to a taxonomy of Explainable AI models, there is a distinction between models that provide transparency about how they function internally (which is not the case with neural networks) and models that can be explained using post-hoc interpretations about how they behaved in particular cases and why.

Concerning post-hoc interpretations in particular, it has been argued that a measure of input significance applicable to all different types of neural networks does not exist (Sarle, 2000). The issue is rather entangled, as Masters (1994, p. 191) notes:

> The question of which features in the training set are used by a particular feed-forward network can be excruciatingly difficult to answer. It is easier to discuss tempting methods that *do not work* than it is to find methods that do, so that will be done first.

Nevertheless, many methods for post-hoc interpretations have been proposed in literature, each of them bearing its own set of advantages and disadvantages. These methods utilise different means, such as (see Barredo Arrieta et al., 2020, for a thorough analysis):

- Text explanations
- Visual explanation techniques
- Local explanations, in which explanations to less complex solution subspaces, which are relevant for the whole model, are provided
- Explanations by example, in which representative examples of the inner workings of the model under analysis are extracted
- Explanations by simplification, in which a simplified model is built based on the model that needs to be explained. The simplified model strives to achieve an optimal resemblance to the original model and retain a comparable performance, while reducing complexity.
- Feature relevance explanation methods, which compute relevance scores for a model's input variables and compare these scores to estimate the importance of each variable.

Feature relevance explanation methods have been particularly popular in feed-forward networks analysis. An indicative list of such methods —although by no means exhaustive— is

the following (Cao et al., 2016; Gevrey et al., 2003; Goh, 1995; Montaño and Palmer, 2003; Sarle, 2000; Sung, 1998; Wang et al., 2000; Yang and Zhang, 1997; Zapranis and Refenes, 1999):

- Weighted average of input weights
- Sums of input weights
- Elaborate functions of weights, such as the sum of products of normalized weights, proposed by Garson (1991)
- Partial derivatives, whether average derivatives, average absolute derivatives, average squared derivatives or average derivative magnitude
- Average elasticity or average elasticity magnitude of output with respect to an input
- Differences in output corresponding to a given change in an input
- Change in loss function when an input is perturbed, replaced by a fixed value (e.g. its mean) or removed completely
- Change in the coefficient of determination when an input is perturbed
- The profile method, used by Lek et al. (1995, 1996), which analyses one input at a time, clamping the values of all other inputs at a fixed level
- Neural network committee–based sensitivity analysis, proposed by Cao et al. (2016), which utilises a set of neural network models instead of only a single optimal model

Partial derivative and input perturbation methods have been proved to exhibit better performance compared to other methods based on weights (Gedeon, 1997; Wang et al., 2000).

2.3.5 Shapley Additive Explanations (SHAP)

Shapley Additive Explanations (SHAP) is a novel model-agnostic, perturbation-based method, with a theoretical foundation in game theory, which provides interpretability to machine learning models. The main idea behind SHAP is calculating the contribution of each input value to the prediction made by the underlying model in a similar way to Shapley values which determine the contribution of each agent in a game.

SHAP was proposed by Lundberg and Lee (2017) in an attempt to unify several existing methods: LIME (Ribeiro et al., 2016), Shapley sampling values (Strumbelj and Kononenko,

2014), DeepLIFT (Shrikumar et al., 2017), QII (Datta et al., 2016), Layer-wise relevance propagation (Bach et al., 2015), Shapley regression values (Lipovetsky and Conklin, 2001) and Tree interpreter (Saabas, 2014).

Shapley Values

SHAP values are based on Shapley values, a solution concept for cooperative games first introduced in Shapley (1953). Shapley values are used to allocate payouts to game agents depending on their contribution to the total payout.

A formal analysis of Shapley values is as follows: Consider a coalitional game with transferable utility, i.e. a pair $G = (N, v)$, where $N = 1, 2, \ldots, n$ is a finite set of agents and characteristic function $v : 2^N \to \mathbb{R}$, with $v(\emptyset) = 0$, associates with each coalition $S \subseteq N$ a real-valued payoff $v(S)$ (worth of the coalition) that is available for distribution among coalition agents. Coalition N is termed *grand coalition*.

The marginal contribution of agent $i \in N$ to coalition S is the difference in the worth of coalition S as a result of agent's i joining the coalition, that is:

$$v(S \cup \{i\}) - v(S)$$

According to Shapley, the payoff of coalition S that should be allocated to each agent i is equal to its average marginal contribution over all possible permutations of the coalition's agents. This payoff constitutes the agent's Shapley value $\phi_i(v)$. More formally,

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} \left(v(S \cup \{i\}) - v(S)\right) \tag{2.35}$$

The Shapley value is proved to be the only value that satisfies the axioms of efficiency, symmetry, additivity and null-player (Shapley, 1953). More specifically:

- Efficiency states that the sum of the Shapley values of all agents is equal to the value of the grand coalition, i.e.:

$$\sum_{i \in N} \phi_i(v) = v(N) \tag{2.36}$$

- Symmetry states that two agents that make the same marginal contributions to a coalition will have the same value:

$$v(S \cup \{i\}) = v(S \cup \{j\}) \Leftrightarrow \phi_i(v) = \phi_j(v) \quad \forall i, j \in N \, \forall S \subseteq N \setminus \{i, j\} \tag{2.37}$$

- Additivity states that, if the characteristic functions v and w of two different games with the same set of agents are added to form a new game, the value of an agent in the new

game is equal to the sum of its value in the aforementioned two games:

$$\phi_i(v+w) = \phi_i(v) + \phi_i(w) \quad \forall i \in N \tag{2.38}$$

- The null-player axiom states that, if an agent i has zero marginal contribution to every coalition, then its value will be zero as well:

$$v(S \cup \{i\}) = v(S) \Leftrightarrow \phi_i(v) = 0 \quad \forall i \in N \, \forall S \subseteq N \backslash \{i\} \tag{2.39}$$

Defining SHAP values

Understanding SHAP values involves a proper understanding of additive feature attribution methods. Additive feature attribution methods utilise what is termed an *explanation model*, which is defined as any interpretable approximation of the original model.

This can be mathematically formulated as follows (Lundberg and Lee, 2017): Let f denote the original prediction model and g denote the explanation model. Explanation models make use of simplified inputs x' which map to the original inputs x through a function h_x: $x = h_x(x')$. Local methods, which explain a prediction $f(x)$ on the basis of a single input x, strive to ensure that $g(z') \approx f(h_x(z'))$ whenever $z' \approx x'$.

In additive feature attribution methods, the explanation model is a linear function of binary variables:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{2.40}$$

where $z' \in \{0,1\}^M$, M is the number of simplified input features, ϕ_0 is the value g would expect to be predicted by f if the input contained no features, and $\phi_i \in \mathbb{R}$ is the effect attributed to each input feature. Summing the effects attributed to all input features yields an approximation of the output $f(x)$ of the original model.

Additive feature attribution methods possess three desirable properties: local accuracy, missingness and consistency.

When the task is to approximate f for a particular input x, **local accuracy** requires that the output of explanation model g for simplified input x' be equal to the output of the original model f:

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i \tag{2.41}$$

If simplified inputs are considered to indicate feature presence, **missingness** requires that features not present in the original input have no effect attributed to them:

$$x'_i = 0 \Rightarrow \phi_i = 0 \tag{2.42}$$

Consistency suggests that changing a model so that a feature has a larger effect on the model does not decrease the effect attributed to the feature in question. More formally: Let $f_x(z') = f(h_x(z'))$ and $z'\backslash i$ denote setting $z'_i = 0$. For any two models f and f', if

$$f'_x(z') - f'_x(z'\backslash i) \geqslant f_x(z') - f_x(z'\backslash i) \tag{2.43}$$

for all inputs $z' \in \{0,1\}^M$, then

$$\phi_i(f',x) \geqslant \phi_i(f,x) \tag{2.44}$$

It can be proved that there is only one explanation model g that complies to the definition of additive feature attribution methods and satisfies all three properties illustrated above:

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!} \left[f_x(z') - f_x(z'\backslash i)\right] \tag{2.45}$$

where $|z'|$ is the number of non-zero entries in z' and $z' \subseteq x'$ denotes all vectors z' in which the non-zero entries are a subset of the non-zero entries in x'. In combined cooperative game theory results, ϕ_i values are known as Shapley values (Lipovetsky and Conklin, 2001).

SHAP values are the solution to (2.45), where $f_x(z') = f(h_x(z')) = E\left[f(z)|z_S\right]$ and S denotes the set of non-zero indices in z'. SHAP values attribute to each input feature the change it induces in the expected prediction of the model when the prediction is conditional on that feature. Ultimately, SHAP values demonstrate how to get to the actual output $f(x)$ starting from the base value $E\left[f(z)\right]$, which would be predicted if no features were known. In the case of non-independent input features or non-linear models, where the order in which input features are added to the base value is important, SHAP values are calculated by averaging ϕ_i values across all possible orders of input features.

The exact calculation of SHAP values is a demanding task, since there are 2^M permutations for M features and, therefore, calculating SHAP values would require 2^M calculations for each single prediction. In related literature, this problem is most commonly addressed by sampling methods (see e.g. Benati et al., 2019; Castro et al., 2017, 2009). In the same spirit, (Lundberg and Lee, 2017) propose specific approximation methods, either model-agnostic or model-specific, in order to accelerate calculations.

Kernel SHAP, the most prominent model-agnostic method, is a combination of another additive feature attribution method, LIME (Ribeiro et al., 2016), and Shapley values. In order to calculate ϕ, LIME minimises the following objective function:

$$\xi = \underset{g \in G}{\arg\min}\, L(f,g,\pi_{x'}) + \Omega(g) \tag{2.46}$$

where L denotes the loss evaluated over a set of samples in the simplified input space, $\pi_{x'}$ is a weighting kernel of the loss function and Ω is a regularisation term, penalising the complexity of the explanation model g.

If terms L, $\pi_{x'}$ and Ω are defined as in the following equations, then SHAP values can be derived from (2.46) by regression:

$$\Omega(g) = 0 \tag{2.47}$$

$$\pi_{x'}(z') = \frac{(M-1)}{(M \text{ choose } |z'|)|z'|(M-|z'|)} \tag{2.48}$$

$$L(f, g, \pi_{x'}) = \sum_{z' \in Z} \left[f(h_x(z')) - g(z') \right]^2 \pi_{x'}(z') \tag{2.49}$$

where $|z'|$ denotes the number of non-zero elements in z'.

This approach has the advantage of significantly reducing computational costs. However, the computed SHAP values are not 100% accurate, while Kernel SHAP also assumes feature independence.

Advantages, Criticisms and Considerations

SHAP has been gaining increasing attention since its inception and is considered a reliable and prominent modern explainability method (see eg. Antwarg et al., 2019; Mokhtari et al., 2019; Rathi, 2019)

Advantages of SHAP may be summarised as follows:

- SHAP has a theoretical foundation in game theory.
- SHAP provides measures of both global and local interpretability: collective SHAP values represent the impact (either positive or negative) each input variable has on the output variable (*global interpretability*), while each prediction can receive its own set of SHAP values which demonstrate the impact of each input variable on that particular prediction (*local interpretability*).
- SHAP provides a fast model-specific implementation for tree-based models.
- The authors offer an implementation of SHAP values in Python (the programming language commonly used in machine learning), in a package that offers easy-to-use insightful visualisation tools.

The main disadvantages of SHAP are as follows:

- SHAP is vulnerable to adversarial attacks; there is empirical evidence that SHAP can be effectively exploited by adversaries to produce harmless explanations for predictions generated by biased classifiers (Slack et al., 2020).
- Kernel SHAP requires considerable time for computations (although much less than using Shapley values). This renders it impractical when the training instances are many.
- Kernel SHAP assumes feature independence; however, there have been attempts to alleviate the effects of such dependence (Aas et al., 2019).

Researchers have also expressed concerns regarding the indiscriminate use of SHAP values. Mittelstadt et al. (2019) highlight the fact that methods for explanatory AI, such as SHAP, are based on models that do not strive to fully capture reality but instead provide a reliable approximation of it. As such, explainability methods are subject to certain limitations, which must be fully understood by researchers to avoid misuse and misleading statements. Kaur et al. (2020), in their research on data professionals, report that, while many participants could not fully comprehend the insights provided by SHAP analysis, they nevertheless used SHAP to determine whether models were deployment-ready.

CHAPTER 3

METHODOLOGY

3.1 Introductory Remarks

As illustrated in Chapter 2, there is an extensive literature on arms races, including the Greek–Turkish one. Related literature is mainly focussed on conventional econometric methods, while researchers have identified critical methodological issues.

Among research methods, an approach scarcely adopted in literature is that of neural networks. After the notable paper of Refenes et al. (1995), only Andreou and Zombanakis (2000, 2011) concern themselves with the use of neural networks to approach the issue of Greek–Turkish arms race. Although all three papers note the advantages of using neural networks, none of them touches upon methological issues which are related to neural network model design and input significance evaluation.

This thesis builds upon the neural network approach to arms races. It strives to investigate and unveil core statistical issues, whilst leveraging the advantageous properties of neural networks and the renewed potential in this field.

3.2 Neural Network Design and Evaluation Process

The process of examining arms races through the use of neural networks can be summarised as follows: A model to predict armaments is designed and trained on available data, until an acceptable performance is obtained. Input significance analysis is then performed on the trained model to establish whether an arms race exists, based on the proposed definition of arms races.

3.2.1 Defining Arms Races

The term ‘arms race’ has been extensively used in related literature. However, it has acquired many different —sometimes contradictory—definitions and thus, a disambiguation of the term is necessary before proceeding further with the analysis.

In the context of this thesis, a less strict definition of bilateral arms races is proposed. An arms race is defined as a specific pattern of arms acquisition, where the armaments level of each

country is primarily explained by factors that refer to the military capabilities of its rival. More specifically, the following assumptions are made:

(a) The armaments level of a country is explained by both internal and external factors.

(b) For an arms acquisition procedure to qualify as an arms race, factors that are directly related to the rival's military capabilities should exert significant influence over the armaments level of each country. It should be said, however, that drawing an absolute line between factors that qualify as related to military capabilities and factors that do not qualify as such can be a challenging task.

(c) The existence of an arms race does not imply the existence of hostility between the two rivals. While the perception of hostility is deeply embedded in arms races literature, an arms race as defined above could as well be the result of an attempt to preserve 'military prestige'.

(d) 'Rapid' or 'abnormal' rates of military growth are not considered decisive features of an arms race. Instead, as Kydd (2000) maintains, constant high levels of military expenditure may be indicators of an arms race per se. In other words, the intensity of military competition may not manifest itself in military growth.

(e) An arms race may even exist when both countries' armaments decrease. The existence of an arms race when armaments decrease may sound counterintuitive considering the notion of 'racing'. However, this particular case may be understood as a situation where adversaries believe that they supersede their rivals and thus decide to reduce spending in order to preserve resources.

Two final notes should be made on the definition: First, this definition helps distinguish between situations where simultaneous changes of armaments are explained by factors pertaining to the adversaries and situations where internal factors drive increases. Second, arms races in the context of this definition do not necessarily bear a negative connotation; they may be a country's best option given its goals and its security environment (Glaser, 2004).

3.2.2 Input and Output Variables

As far as the output variable of the network is concerned, researchers that employ neural networks choose either the defence expenditure as a share of GDP (Andreou and Zombanakis, 2000, 2011) or the change of military expenditure (Refenes et al., 1995). As illustrated in Chapter 2, there is also a wider discussion over what constitutes a suitable proxy of armaments. This thesis adopts

the view expressed by Brauer (2002) that defence expenditure is the best available measure of armaments. More specifically, the output variable used in this thesis is the change of defence expenditure, as in Refenes et al. (1995), since it is more faithful to the original Richardson model and prevents the issue of autocorrelations in level data leading to deceptively good performance.

Choosing input variables is a rather challenging task. A multitude of variables has been used in related literature, since there is no consensus about a specific theoretical framework for variable selection. In this thesis, three different models are developed, each one utilising a different set of input variables:

(a) A model based on the original Richardson model, using the Greek and Turkish defence expenditure of period $t-1$ as input variables (Model A).

(b) A model based on the original paper by Refenes et al. (1995), utilising the same input variables (Model B).

(c) A self-developed model, aiming to resolve methodological issues related to the choice of input variables, particularly correlation (Model C). No a priori theoretical framework is postulated; instead, variables are chosen on a purely ad hoc basis.

Given the small number of available observations for defence expenditure data, the inclusion of a large number of input variables in the self-developed model was avoided, since that would inflate the number of network parameters, thus increasing the danger of overfitting. After all, as Zapranis and Refenes (1999, p. 76) stress, 'given n points we can always find an $(n-1)-$dimensional hyperplane that will provide a perfect fit for the data.' It is the same principle highlighted in the saying recalled by Luterbacher (1975, p. 213): 'with four parameters you can fit an elephant and with eight you can make him [*sic*] wiggle his [*sic*] tail.' Details about the exact variables used are given in Chapter 4.

3.2.3 Data Reliability

The following notes have to be made about concerns over data reliability:

(a) Data used for this thesis were gathered from readily available and reputable sources, such as the World Bank and the IMF. Defence expenditure data in particular were drawn from the SIPRI database, which is reportedly considered a gold standard (Ward, 2020, p. 62).

(b) Regarding defence expenditure data, since there is no apparent way to overcome the margin of error that stems from over- or under-reporting military expenditure, it is assumed that this

deviation from real figures is statistically insignificant or that both Greece and Turkey opt for statistically equal measures of deviation for their expenditure figures. In other words, we consider deviations not to exert significant influence over the results of the models. After all, if the opposite were to be accepted, all related research would be rendered impossible.

3.2.4 Training Process

The process used for neural network model selection is a modified version of the process proposed by Zapranis and Refenes (1999). More specifically:

1. The process starts with the simplest class of models (one hidden layer and one hidden unit)
2. Model parameters are estimated
3. The loss function is calculated on the validation set
4. The model with the least associated validation loss is chosen
5. The process is iterated until there are 10 hidden units in the network; the chosen number of hidden units is the one associated with the least loss

Dataset division

Following common practice, the dataset is divided into three subsets: a training set, a validation set and a test set, comprising approximately 60%, 20% and 20% of all observations respectively. More specifically, the training set comprises years 1963 – 1995, the validation set comprises years 1996 – 2006 and the test set comprises years 2007 – 2018.

The validation set is used for the determination of model hyperparameters and the detection of overfitting, through a process of evaluating model performance for different combinations of hyperparameter values. Since the validation set is used during the process of model fitting, it cannot be used for its evaluation. Using it for evaluation would yield an overly optimistic performance. Thus, a separate test set, which includes 'unseen' data and enables an unbiased comparison of different models, is used.

Scaling

Data often comprises features which have a different range of values. When this unscaled data is used for training, the loss function is likely to exhibit very elongated valleys. During

optimisation using gradient descent, the existence of unscaled data leads to the gradient being steep with respect to some of the parameters, which results in large oscillations in the search space. Scaling ensures that the magnitude of values assumed by the features is at a comparable range.

Following standard practice, scaling is applied to the input data in order to facilitate gradient descent convergence and, thus, facilitate learning. The scaler chosen in this thesis is RobustScaler (scikit-learn developers, 2020). Robust Scaler subtracts the median and then scales data according to Interquantile Range (IQR), i.e. the range between the first and third quartiles (or the 25th and 75th quantiles, respectively). Scaling is performed based on IQR instead of unit variance (as in standardisation) as a means to alleviate the effect of outliers. Mathematically:

$$\text{Scaled Value} = \frac{\text{Original Value} - \text{Median}}{Q_3 - Q_1} \tag{3.1}$$

where Q_3 and Q_1 denote the first and the third quartile respectively.

RobustScaler was fitted on training data and then applied to validation and test data, so as to avert information leakage.

Basic model parameters

All neural networks used in this thesis are fully interconnected multilayer feedforward networks, since the analytical power of this network type is satisfactory and well understood (Bishop, 1995; Rumelhart and McClelland, 1986). The basic parameters used in all models are as follows:

- *Activation functions:* The sigmoid function is used for data in the input layer, while the hyperbolic tangent function is used for the hidden layer. The hyperbolic tangent function resticts the output to a range $[-1, 1]$, which is practically the range assumed by the change in defence expenditure.

- *Optimiser:* Adam optimiser is used; the parameters used for Adam are as follows (see Google developers, 2021): learning_rate (step size) is 0.001, beta_1 is 0.9, beta_2 is 0.999, epsilon is 1e-7

- *Epochs:* The number of training epochs is increased stepwise (at 100-epoch intervals, starting with 100 epochs) until loss starts increasing

- The error measure which was used to train the model was Mean Square Error (MSE), which is a commonly used measure in machine learning applications (Du and Swamy, 2019).

A series of simulations were performed in order to identify a region of statistical stability of the network performance. The network parameters chosen represent a reasonable compromise between maximum performance and stability.

Input significance analysis

As far as input significance is concerned, both Refenes et al. (1995) and Andreou and Zombanakis (2000, 2011) estimate input significance through the calculation of partial derivatives with respect to the inputs. In this thesis, input significance is performed using SHAP (Lundberg and Lee, 2017), which is considered one of the most prominent modern methods. More specifically, the model-agnostic Kernel SHAP is used for the approximation of SHAP values. Increased time for the execution of the algorithm is not a concern in this case due to the small number of training examples and input variables.

Correlation analysis

Kernel SHAP calculates approximate SHAP values assuming feature independence. As such, an analysis of correlations between input variables is essential. If inputs are significantly correlated, their effects cannot be separated during the calculation of SHAP values and, therefore, feature importance calculations are not reliable.

Two measures of correlation are calculated: Pearson's correlation coefficient, which captures linear interactions between variables, and distance correlation (Székely et al., 2007), which is able to capture nonlinear interactions as well. More formally:

Pearson's correlation coefficient

Given a set of pairs $\{(x_1,y_1),\ldots,(x_n,y_n)\}$ consisting of n pairs, Pearson's correlation coefficient r_{xy} is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{3.2}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (respectively for y)

Distance correlation

Given a set of pairs (X_k, Y_k), $k = 1, 2, \ldots, n$ sampled from a pair of random variables (X, Y), let:

$$\alpha_{j,k} = \left\| X_j - X_k \right\| \tag{3.3}$$

$$b_{j,k} = \left\| Y_j - Y_k \right\| \tag{3.4}$$

where $j, k = 1, 2, \ldots, n$ and $\|\cdot\|$ denotes Euclidean norm, be the distance matrices containing all pairwise distances.

Then, let:

$$A_{j,k} := \alpha_{j,k} - \overline{\alpha_{j\cdot}} - \overline{\alpha_{\cdot k}} + \overline{\alpha_{\cdot\cdot}} \tag{3.5}$$

$$B_{j,k} := b_{j,k} - \overline{b_{j\cdot}} - \overline{b_{\cdot k}} + \overline{b_{\cdot\cdot}} \tag{3.6}$$

where $\overline{\alpha_{j\cdot}}$ is the jth row mean, $\overline{\alpha_{\cdot k}}$ is the kth column mean and $\overline{\alpha_{\cdot\cdot}}$ is the grand mean of the distance matrix of the X sample (similarly for b values).

Given that, distance covariance is defined as the non-negative square root of:

$$\mathrm{dCov}^2(X, Y) := \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k} B_{j,k} \tag{3.7}$$

Similarly, distance variance is defined as the non-negative square root of:

$$\mathrm{dVar}^2(X) := \mathrm{dCov}^2(X, X) = \frac{1}{n^2} \sum_{k,l} A_{k,l}^2 \tag{3.8}$$

Distance correlation is then defined as follows:

$$\mathrm{dCor}(X, Y) = \frac{\mathrm{dCov}(X, Y)}{\sqrt{\mathrm{dVar}(X)\, \mathrm{dVar}(Y)}} \tag{3.9}$$

Distance correlation takes values in range $[0, 1]$, where:

- $\mathrm{dCor}(X, Y) = 0$ if and only if X and Y are independent
- $\mathrm{dCor}(X, Y) = 1$ implies that the dimensions of the linear subspaces spanned by X and Y samples respectively are almost surely equal; if it is assumed that these subspaces are equal, then $Y = A + bCX$ for some vector A, scalar b and orthonormal matrix C.

Proceeding to the statistical evaluation of calculated correlation coefficients unveils another oft-overlooked yet critical issue: the choice of a significance level α. The value of α is typically set at 0.05, although values of 0.01 and 0.1 are also commonly used in literature. It should be noted, however, that this choice is merely a convention, bearing no scientific basis (Arrow, 1960; Lehmann and Romano, 2005). Concerns have been raised that significance levels are used in a ritualistic way (see e.g. Keuzenkamp and Magnus, 1995), which results in a distortion of the scientific process and in unreliable conclusions (Wasserstein and Lazar, 2016).

Since the test for the statistical evaluation of calculated correlation coefficients is conducted with a sample size of 33, it is quite likely that the power of the test is low. Taking this low power into account, the approach taken in defining a significance level is the decision-theoretic approach proposed by Kim and Choi (2019). This approach calculates the optimal significance level through an optimisation process, which considers key factors of hypothesis testing: sample size (the power of the test), losses from incorrect decisions, the researcher's prior beliefs for the null (H_0) and alternative (H_1) hypotheses and the substantive importance of the relationship being tested.

More specifically, the optimal significance level is calculated by minimising the expected loss from hypothesis testing. Significance level α and probability of Type II error β are chosen in such a way that the expected loss function is minimised. The expected loss function is defined as

$$PL_1\alpha + (1-P)L_2\beta \tag{3.10}$$

where $P \equiv \text{Prob}(H_0) = 1 - \text{Prob}(H_1)$ is the researcher's prior belief for H_0 and L_i, $i = I, II$ is the loss from Type i error. In the case of a researcher who is impartial between H_0 and H_1, in terms of prior beliefs and losses from incorrect solutions, it is reasonable to set $P = 0.5$ and $L_2/L_1 = 1$ when minimising loss.

When calculating the optimal significance level, a choice should be made for the value of H_1 under which the power is calculated. According to Kim and Choi (2019), this choice should be made as a result of thorough economic analysis, although it is understood that this choice may be entirely subjective or significantly difficult.

Since there is no research regarding the exact effect of correlation levels on input significance analysis using Kernel SHAP, three arbitrary limits of 0.3, 0.4 and 0.5 are examined for Pearson's correlation coefficient. This choice is made in the sense that these correlation levels are considered practically significant for correlation to adversely affect input significance analysis. Through the use of OptSig package in R, the optimal significance levels for the three chosen limits are 0.21, 0.12 and 0.08 respectively.

The same significance levels are also used for distance correlation evaluation, since Kim and Choi (2019) have not implemented a corresponding function for distance correlation. Although these levels might not be optimal, they are certainly more appropriate than the typical 0.05 and 0.01 levels. However, extended research on the issue should be carried out.

Multicollinearity

Variance Inflation Factor (VIF) is used for the detection of any existing multicollinearity between input variables, which could adversely affect Kernel SHAP evaluations.

The calculation of VIF is performed through the following steps: If $X_i, i = 1,2,\ldots,N$ denotes the ith input variable of a neural network model with N input variables, then the VIF for this input variable is:

$$VIF_i = \frac{1}{1-R_i^2} \tag{3.11}$$

where R_i^2 denotes the unadjusted coefficient of determination obtained by regressing input variable i on the remaining input variables via an ordinary least squares regression (OLS).

The value of VIF_i is used to assess the magnitude of multicollinearity. A VIF value of 1 indicates no multicollinearity, while a value of 10 is typically used to denote high multicollinearity (Kutner et al., 2004).

3.2.5 Evaluation

The predictive power of the model is evaluated using three measures: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2). More formally:

Let $(\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N)$ denote values predicted by the network and $(y_1, y_2, \cdots, y_N)$ denote the actual values, where N is the sample size. Then:
Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3.12}$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3.13}$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{3.14}$$

To obtain estimates of minimum acceptable performance, three baseline models are used as benchmarks:

(a) A model which predicts that the change in military expenditure for each year is equal to last year's change, in other words, $\Delta Y_{t+1} = \Delta Y_t$.

(b) A model which predicts that military expenditure remains stable, i.e. $\Delta Y_t = 0$

(c) A model which predicts that the change of military expenditure is permanently equal to the mean change observed on the training set, i.e. $\Delta Y_{t+1} = \overline{\Delta \Upsilon}_{train}$.

3.3 Software

Computations are mainly performed using Python (version 3.7) and R (version 4.0.5) programming languages. The main packages used for this thesis are as follows:

- Jupyter 1.0.0 (Kluyver et al., 2016) for development
- NumPy 1.19.5 (Harris et al., 2020) and Pandas 1.2.1 (McKinney, 2010) for scientific computations and data processing
- Matplotlib 3.3.4 (Hunter, 2007) for data visualisations
- SciPy 1.6.0 (Virtanen et al., 2020) and Statsmodels 0.12.2 (Seabold and Perktold, 2010) for statistical tests
- Scikit-learn 0.24.1 (Pedregosa et al., 2011) and Tensorflow 2.4.1 (Abadi et al., 2015) for neural network design, estimation and evaluation
- SHAP 0.38.1 (Lundberg and Lee, 2017) for input significance analysis
- OptSig 2.1 for R language (Kim, 2020) for the computation of optimal significance levels

Considering the significance of reproducibility in machine learning research (Stodden and Miguez, 2014), all datasets as well as the relevant code used for the design and training of neural network models can be found in the Appendix.

CHAPTER 4

RESULTS

4.1 Greek Defence Expenditure: Analysis

Before proceeding to a detailed analysis of the results, it is significant that an analysis of the output variable (change in Greek defence expenditure) is performed. Figure 4.1 presents the complete time series, Figure 4.2 shows the histogram and the cumulative distribution function, while Table 4.1 presents important descriptive statistics for the complete dataset as well as for the values that comprise the training set.

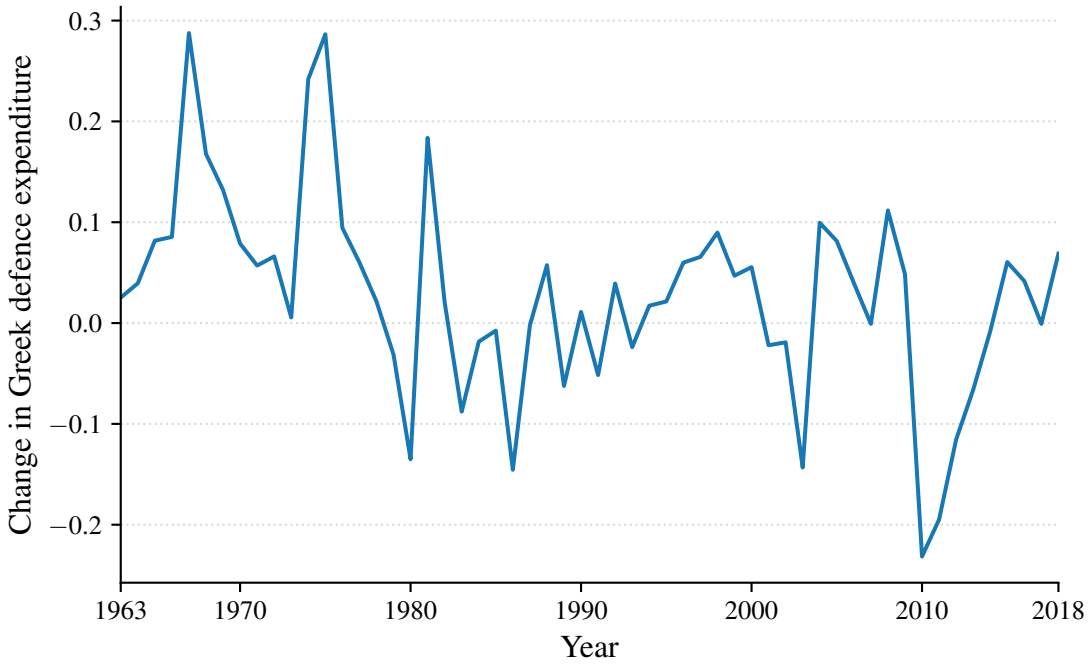

Figure 4.1: Plot of the output variable (Change in Greek defence expenditure) for years 1963–2018

The table and plots provide some interesting insights: The change in Greek military expenditure has a mean of 2.8%, exhibiting significant variance (10.1%) and a very wide range of values (52%). It is characterised by sharp increases (such as those in years 1967–1968, 1974–1975 and 1981) and decreases (such as those in years 1980, 1986, 2003 and 2010), with the overall trend being downward. Moreover, approximately 40% of the values are over 10% or under −10%.

Table 4.1: Descriptive statistics for the output variable (Change in Greek defence expenditure). Statistics are calculated over the complete dataset (years 1963–2018) as well as over the training set only (years 1963–1995).

Statistic	Complete dataset	Training set
Mean	0.028	0.046
St. Dev.	0.101	0.103
Min	−0.232	−0.146
Max	0.288	0.288
Skewness	0.067	0.670
Kurtosis	1.296	0.737

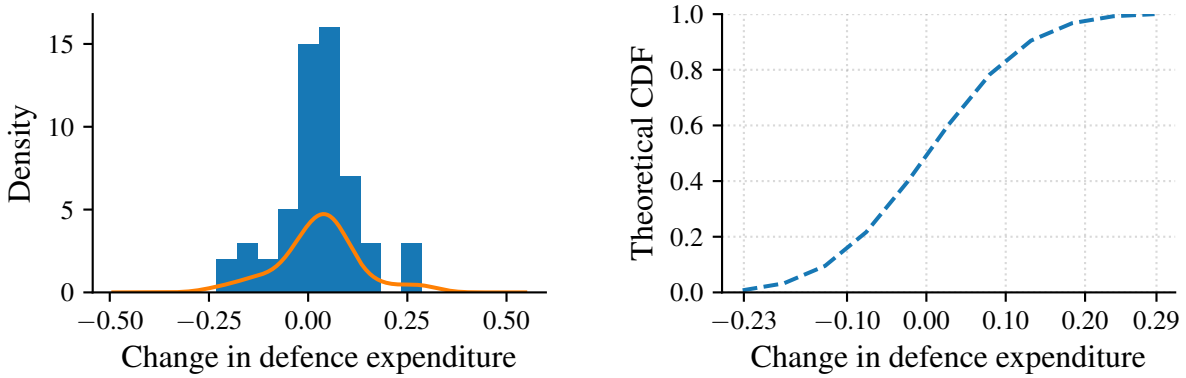

Figure 4.2: Histogram (left) and theoretical Cumulative Distribution Function (right) for the output variable (Change in Greek defence expenditure)

4.2 Baseline Models

The baseline models which were used to assess the minimum acceptable performance for the three models examined in this thesis, yielded the results shown on Table 4.2 when evaluated over the test set. The best values obtained by the baseline models were an RMSE of 10% and an MAE of 7.67%, while values for R^2 were very low (and even negative in two of the models).

Table 4.2: Performance metrics for baseline models

	Performance metric		
Model	RMSE	MAE	R^2
$\Delta Y_{t+1} = \Delta Y_t$	0.1008	0.0766	0.0442
$\Delta Y_t = 0$	0.1058	0.0791	−0.0537
$\Delta Y_{t+1} = \overline{\Delta Y}_{train}$	0.1245	0.0875	−0.4588

4.3 Analysis of Input Variables

Information and descriptive statistics for the variables used in all models are shown on Table 4.3. An apparent observation is that many of the variables are characterised by outliers. Although these outliers may squash the scaled variables to a narrow range, winsorising or removing them was not considered a sensible choice, since they may contain useful information.

Table 4.3: Descriptive statistics for variables of all models

Alias	Variable	Measure	Descriptive Statistics					
			Mean	**St. Dev**	**Min**	**Max**	**Skewness**	**Kurtosis**
Model A								
A1	Turkish military expenditure	Constant 2010 USD [billion]	12.267	6.233	2.532	24.431	−0.104	−1.226
A2	Greek military expenditure	Constant 2010 USD [billion]	6.534	2.210	1.744	10.623	−0.673	−0.170
Model B								
B1	Ratio of Greek to Turkish armed forces personnel	—	0.277	0.049	0.197	0.402	0.776	0.104
B2	Greek military expenditure per person in the armed forces	Constant 2010 USD [10,000s]	37.278	13.884	10.215	74.323	−0.094	0.091
B3	Turkish military expenditure per person in the armed forces	Constant 2010 USD [10,000s]	18.948	11.739	5.478	68.782	1.540	4.509
B4	Greek military expenditure	Share of GDP (%)	0.038	0.010	0.023	0.059	0.516	−0.592
B5	Turkish military expenditure	Share of GDP (%)	0.033	0.008	0.018	0.051	−0.184	−0.318
Model C								
C1	Turkish military expenditure	Share of GDP (%)	0.033	0.008	0.018	0.051	−0.184	−0.318
C2	Lagged change of Turkish military expenditure	—	0.048	0.120	−0.119	0.749	3.636	20.904
C3	Lagged change of Greek military expenditure	—	0.027	0.101	−0.232	0.288	0.095	1.324
C4	Aggregate change of Greek public debt over the previous two years	Share of GDP (%)	5.903	9.981	−6.061	45.351	1.862	4.625
C5	Aggregate change of Greek gross capital formation over the previous two years	Share of GDP (%)	−0.264	4.440	−11.866	10.266	−0.047	0.303
C6	Lagged change of Greek imports	Share of GDP (%)	0.355	2.008	−7.206	6.576	−0.203	4.246

4.4 Neural Network Results

The training parameters chosen for each of the models through experimentation are illustated on Table 4.4. Please note that the notation $S_1 - S_2 - \cdots - S_m$ used in the Units column represents a neural network with m layers, where S_i is the number of nodes in the ith layer. Input layer is counted as layer 1 and layer m is the output layer.

Table 4.4: Neural network parameters for all models

Model	Units	Activation	Optimiser	Epochs
Model A	2–8–1	Sigmoid–Tanh	Adam	400
Model B	5–8–1	Sigmoid–Tanh	Adam	300
Model C	6–7–1	Sigmoid–Tanh	Adam	700

Training the network for each of the models yields the results illustrated on Table 4.5. Models A and B obtain a performance that is only marginally better than that obtained by the baseline models when evaluated over the test set. Model C obtains the best performance among all models, yielding an RMSE of 7.98%, an MAE of 6.29% and an R^2 of 40% when evaluated over the test set. The accuracy of sign prediction for Model C is 66.7%.

An interesting finding is that the validation error is marginally smaller than the training error in all models. This can be explained by the fact that the number of observations, on which the validation error is calculated, is small. The same issue leads to R^2 being negative, since the mean of the validation set contains enough information about the set. However, the fact that the validation set performance exhibits a similar pattern in all models (lower RMSE and MAE than that of the training set, negative R^2) implies that the data points included in the validation set might conform to a different distribution.

Considering related literature, Refenes et al. (1995), in their tests for years 1962–1990, report very high combined R^2 levels for their training and test sets (85.3%), but they do not report a separate R^2 for the test set, due to the number of observations being too small (only five observations). The model used in the original paper is trained for 10,000 epochs, using a topology of 5–32–16–1 neurons, a learning rate of 0.3 and a momentum of 0.2. Using these parameters to train Model B —instead of the parameters used in this thesis— yields a network that is almost perfectly capable of learning the training set (RMSE < 0.03, $R^2 > 0.99$), but completely unable to generalise (RMSE for the validation and test sets consistently over 0.3).

Table 4.5: Performance Metrics for all models

	Performance metric		
Dataset	**RMSE**	**MAE**	**R^2**
Model A			
Training set	0.0891	0.0698	0.2248
Validation set	0.0844	0.0774	−0.5907
Test set	0.1025	0.0864	0.0106
Model B			
Training set	0.0856	0.0652	0.2846
Validation set	0.0781	0.0604	−0.3604
Test set	0.0957	0.0707	0.1373
Model C			
Training set	0.0774	0.0538	0.4139
Validation set	0.0688	0.0485	−0.0550
Test set	0.0798	0.0629	0.4006
Validation & Test set	0.0747	0.0560	0.3408

4.5 Correlations

Since Kernel SHAP assumes feature independence, correlations between variables should first be analysed before proceeding to input significance analysis. Table 4.6 and Table 4.7 present Pearson's and distance correlation coefficients respectively for Model A. Both coefficients are statistically significant at any reasonable level of significance, which constitutes sufficient evidence of a significant relationship between variables A1 and A2. As such, SHAP values are not calculated for Model A, because the results would be unreliable.

Table 4.6: Pearson's correlation matrix for the inputs of Model A

	A1	A2
A1	1.000	0.752 (<.001)
A2		1.000

Table 4.7: Distance correlation matrix for the inputs of Model A

	A1	A2
A1	1.000	0.900 (<.001)
A2		1.000

Table 4.8 and Table 4.9 present Pearson's and distance correlation coefficients respectively for Model B. Out of 10 Pearson's correlation coefficients, only two are statistically insignificant at an 8% and a 12% significance level and only one at a 21% level. Distance correlation coefficients are all statistically significant at an 8% significance level. As in Model A, there is sufficient evidence of significant relationships between the input variables used in Model B and, therefore, calculating SHAP values would also yield unreliable results.

Table 4.8: Pearson's correlation matrix for the inputs of Model B

	B1	B2	B3	B4	B5
B1	1.000	−0.849 (<.001)	−0.545 (.001)	−0.378 (.030)	−0.246 (.169)
B2		1.000	0.604 (<.001)	0.696 (<.001)	0.536 (.001)
B3			1.000	0.078 (.664)	0.520 (.002)
B4				1.000	0.587 (<.001)
B5					1.000

Table 4.9: Distance correlation matrix for the inputs of Model B

	B1	B2	B3	B4	B5
B1	1.000	0.842 (<.001)	0.747 (<.001)	0.379 (.064)	0.398 (.053)
B2		1.000	0.814 (<.001)	0.719 (<.001)	0.588 (<.001)
B3			1.000	0.443 (.021)	0.603 (<.001)
B4				1.000	0.622 (<.001)
B5					1.000

Given the correlation coefficients above for models A and B, the objective that underpinned the creation of Model C was to form a model that would not be affected by correlation issues, at least not as severe as those between variables in models A and B. Table 4.8 and Table 4.9 present Pearson's and distance correlation coefficients respectively for Model C. While still affected by correlation, Model C has relatively less issues than models A and B. Out of 15 Pearson's correlation coefficients, seven are statistically significant at 8%, 12% and 21% levels.

As far as distance correlation is concerned, only five coefficients are significant at an 8% level, eight at a 12% level and eleven at a 21% level.

Table 4.10: Pearson's correlation matrix for the inputs of Model C

	C1	C2	C3	C4	C5	C6
C1	1.000	0.386 (.027)	0.056 (.755)	−0.085 (.636)	−0.399 (.021)	−0.160 (.373)
C2		1.000	0.443 (.012)	−0.051 (.779)	−0.270 (.128)	0.099 (.584)
C3			1.000	−0.338 (.054)	−0.315 (.074)	−0.224 (.210)
C4				1.000	−0.134 (.456)	−0.345 (.049)
C5					1.000	0.359 (.040)
C6						1.000

Table 4.11: Distance correlation matrix for the inputs of Model C

	C1	C2	C3	C4	C5	C6
C1	1.000	0.399 (.108)	0.353 (.182)	0.326 (.288)	0.428 (.036)	0.360 (.164)
C2		1.000	0.440 (.037)	0.250 (.892)	0.412 (.061)	0.360 (.192)
C3			1.000	0.486 (.001)	0.395 (.081)	0.343 (.242)
C4				1.000	0.257 (.802)	0.391 (.097)
C5					1.000	0.450 (.022)
C6						1.000

Since Model C exhibits notably reduced correlation issues compared to models A and B, input significance analysis is performed only on Model C. Before proceeding to the calculation of SHAP values, a test for multicollinearity is performed using Variance Inflation Factor (VIF) values. The results are presented in Table 4.12. Since all VIF values are close to unity and definitely less than 5 or 10, there are no significant multicollinearity issues between the input variables of Model C.

Table 4.12: Variance Inflation Factor (VIF) values for the variables of Model C

Variable	VIF value
C1	1.645
C2	1.876
C3	2.224
C4	1.771
C5	1.516
C6	1.770

4.6 Input Significance Analysis

Input significance analysis for Model C is carried out through the calculation of SHAP values using Kernel SHAP. Figure 4.3 presents an aggregate summary plot, in which features are ranked based on their average absolute SHAP values. Each of these SHAP values represents the average impact each feature has on model output magnitude. The lagged change of Greek defence expenditure (variable C3) emerges as the most influential variable, followed by the lagged change of Greek imports as a share of GDP (variable C6), the aggregate change of Greek public debt over the previous two years as a share of GDP (variable C4) and the Turkish military expenditure as a share of GDP (variable C1).

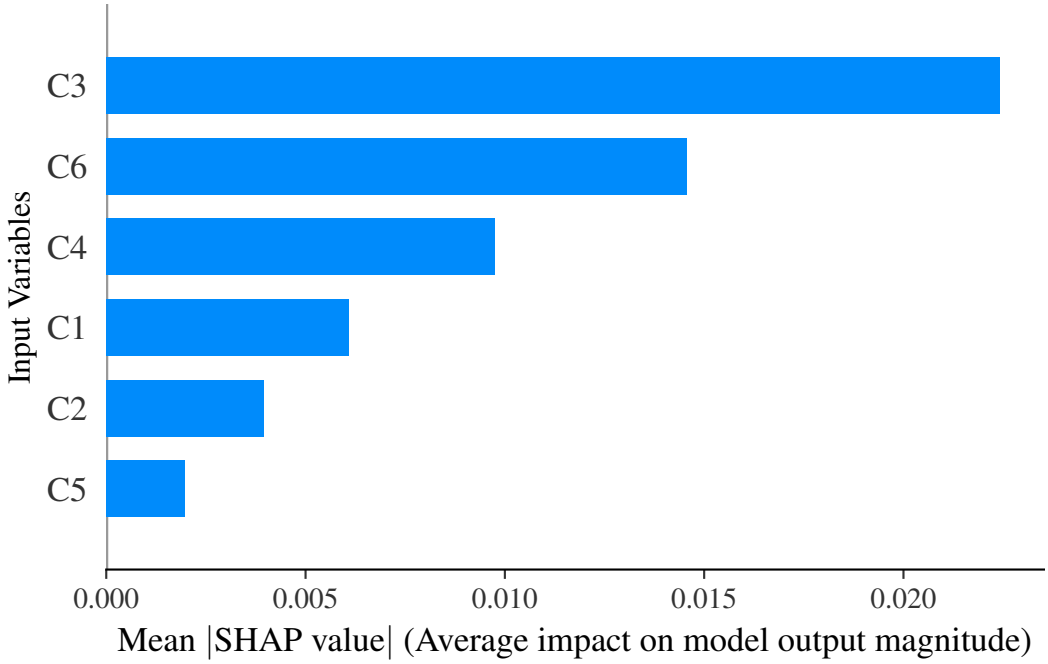

Figure 4.3: SHAP aggregate summary plot for Model C. The input variables of Model C (C1 to C6) are ranked based on their average absolute SHAP values.

The aggregate summary plot does not provide information on the exact way in which

each feature affects the output of the model. For instance, the fact that a feature exhibits a medium average effect on the output could be the result of either a large effect for a few predictions, but no effect in general, or a medium effect for all predictions. For this reason, the SHAP library offers a detailed summary plot, illustrated in Figure 4.4. The detailed summary plot shows the effect of each observation for each feature separately in a concise way. More specifically, the plot consists of many dots, each of which has the following characteristics:

- The vertical location of the dot shows which feature is concerned. Dots in the first row of Figure 4.4 denote observations of variable C3, dots in the second row denote onservations of variable C6, etc. Input variables are ranked based on their average impact (as shown in Figure 4.3).
- The colour of the dot shows whether the value of the feature it depicts was high or low for that particular observation. Blue dots denote low values, while red dots denote high values.
- The horizontal location of the dot shows whether the effect of that specific value caused a higher or lower prediction.

Having said the above, Figure 4.4 shows that high values of change in defence expenditure are associated with high values of change in defence expenditure during the previous year, high values of the lagged change in imports and high values of the aggregate change of Greek public debt over the previous two years. The rest of the variables do not exhibit a notably significant impact on model output. It is also interesting to note that a specific high value of the Turkish military expenditure had a diminishing effect on the change of Greek defence expenditure.

Figure 4.5 shows SHAP dependence plots for variables C3, C6, C4 and C1. Dependence plots enable researchers to understand how each input variable by itself affects the output of the model. They represent the change in the output variable as each of the input variables increases and, in this way, they provide a more detailed view of the interactions presented in the detailed summary plot.

In all plots, the horizontal axis shows the actual (scaled) value from the dataset, while the vertical axis shows the impact on the model output, as measured by the SHAP value. Given the illustrations in Figure 4.5, it appears that the effects of variables C3 and C6 are mostly linear, with a few observations having increased impact, while variables C4 and C6 exhibit nonlinear effects.

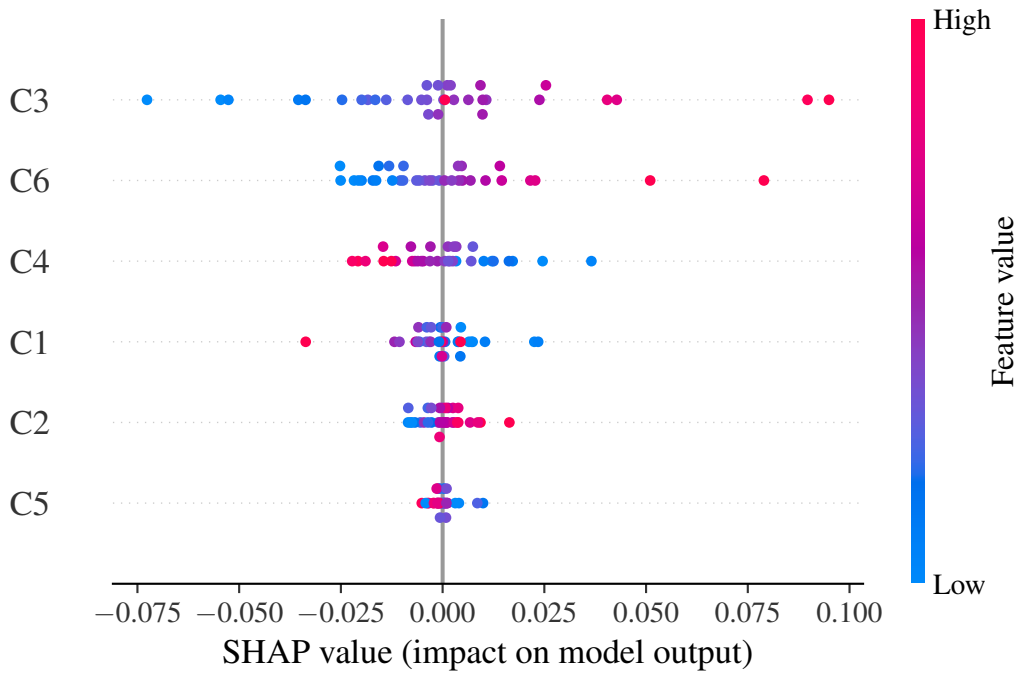

Figure 4.4: SHAP detailed summary plot for Model C. The input variables of Model C (C1 to C6) are ranked based on their average absolute SHAP values. Each dot represents an observation of the variable corresponding to the row in which it is located.

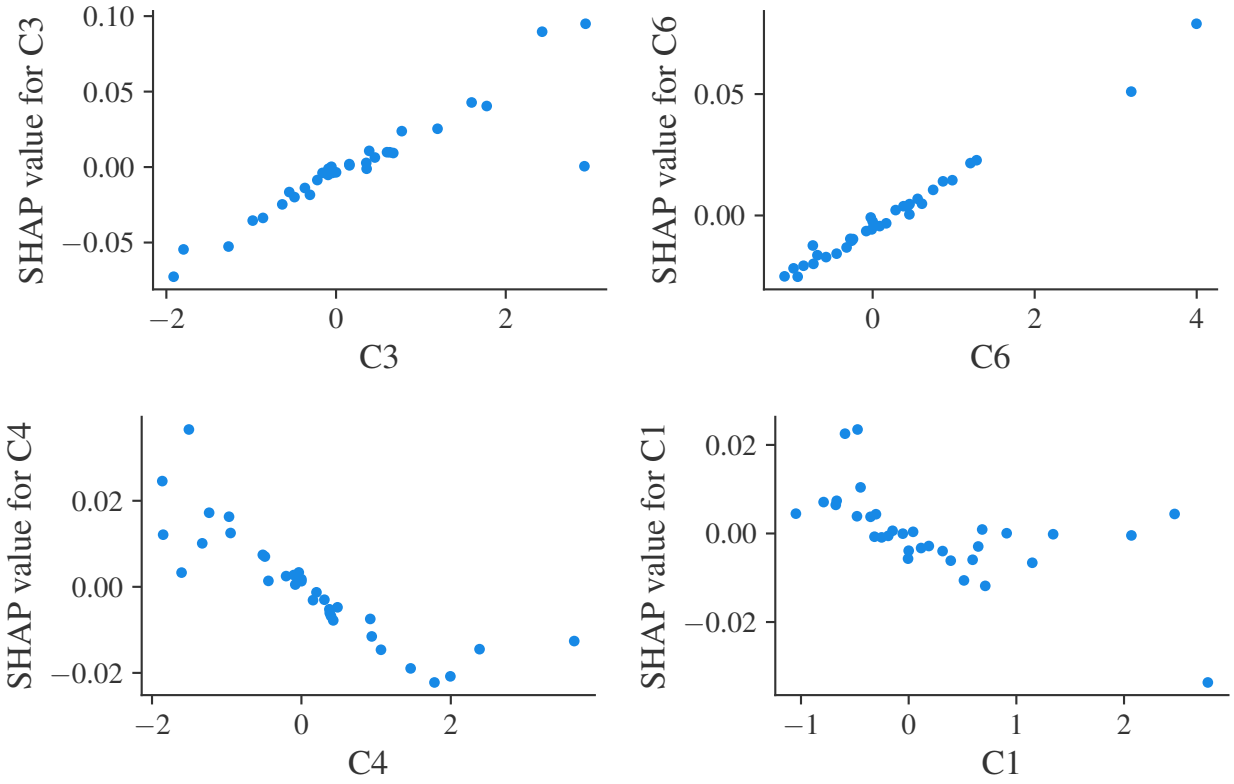

Figure 4.5: SHAP dependence plots for the four most significant variables of Model C based on the mean of their absolute SHAP values

CHAPTER 5

DISCUSSION

The models analysed in Chapter 4 yield particularly interesting results. Among the three models A, B and C that were trained, it is only Model C that exhibits a performance (as measured by RMSE, MAE and R^2) that is substantially superior to that of the baseline models. Models A and B fail to achieve a noteworthy improvement of RMSE or MAE over the baseline models, while their R^2 values are low when evaluated over the test set.

Given the SHAP values of Model C, it appears that the change in Greek defence expenditure is best predicted by internal factors, which suggests that Greece does not engage in an arms race with Turkey. The two Turkey-related input variables of Model C rank only fourth and fifth (out of six) by order of their impact to the output of the model.

The results outlined above are subject to a series of considerations related to research design, methodology and interpretation. A list of such considerations includes variance, correlation, the choice of input variables, data scarcity and reliability, improper understanding of SHAP values and specification bias.

A quick look at the plot and the table of statistics for the change in Greek defence expenditure reveals that the change in spending exhibits considerable variance as well as outlying values. The same applies to input variables used in models A, B and C. Such variance places a constraint on the optimal level of error that can be achieved by any model and should raise suspicion against semmingly perfect predictions. Furthermore, the optimal attainable level of error may in fact not be acceptable for the task at hand.

Analysing how the aforementioned considerations apply to Model C is of great importance. Model C achieves an optimal RMSE of 7.47% and a corresponding MAE of 5.6%. While these values are superior to those reported in earlier studies, they still may not be acceptable depending on the use case: An RMSE of 7.47% translates into a discrepancy of 474.53 million euros when considering Greek defence spending in 2018. On the other hand, given that the variance of the change in Greek military spending is approximately 10%, pursuing an MAE of 1% for example would be an overly optimistic expectation for a predictive model.

As far as R^2 values are concerned, the value of Model C (40%) seems notably inferior to that reported in Refenes et al. (1995). Refenes et al. (1995) report a notably high R^2 value for their combined training and test set (85.3%). The authors note that reporting a separate R^2 for their test set was impossible due to the small number of observations. However, reporting a single

R^2 value for both the training and test sets skews results in the direction of higher R^2 values and is therefore unreliable. Andreou and Zombanakis (2011) do not report an R^2 value, but they do report moderate to high values of the correlation coefficient. These high values reported in previous research establish great expectations for the performance of neural networks, which are seemingly unrealistic for this specific task.

Correlation is also a critical issue. Input variables in Models A and B exhibit significant correlation, which leads to conventional input significance estimation methods yielding unreliable results. Model C is also not unaffected by correlation issues (albeit to a lesser extent). This highlights the difficulty of choosing variables, since most of them are correlated, and the need for input significance methods that are not affected by such issues. Moreover, as Selbst and Barocas (2018) and Kroll et al. (2017) have shown, the mere presence of correlated variables complicates the identification and prevention of bias, even in the case of fully transparent models.

Interestingly enough, all previous related studies that employ neural networks and use partial derivatives as a method for input significance analysis do not examine whether inputs are correlated, although correlation issues affect partial derivatives. The same observation applies for multicollinearity. Kernel SHAP, which is used in this thesis, implies feature independence as well. When a set of correlated features is introduced to the algorithm, it arbitrarily assigns a large weight to one feature in the set and, thus, the remaining features score poorly in terms of their SHAP values.

Evaluating correlation coefficients unveils another crucial issue: the choice of appropriate levels of significance. The decision-theoretic approach by Kim and Choi (2019), which is adopted in this thesis, may provide a more accurate estimate of the appropriate level of significance, but it is subject to the choice made for the value of H_1 under which the test power is calculated.

Another challenging task is the choice of input variables. Input variables in Model C, which exhibited superior performance compared to models A and B, were chosen on an ad hoc basis. As '*post hoc* rationalization of one's findings is very easy' (Brauer, 2002, p. 90), the formulation of a post hoc theoretical framework would be of no practical use. To illustrate how effortlessly such a framework can be developed, it could be said for Model C that imports were used as an indicator of GDP growth (implying that GDP growth translates into more available resources to be devoted to defence, since defence is considered a public good.)

Choosing variables on an ad hoc basis eliminates the need for a theoretical model, but a question inevitably arises: Which variables should be chosen? The final choice of variables is not free of bias, since it is directly influenced by the researcher's beliefs and opinions. Moreover,

a combination of variables completely different to the one chosen may yield comparable or even better results.

The fact that there is no consensus among researchers on the most appropriate variables does not contribute towards the alleviation of confusion. It also appears that the same set of variables would not be appropriate for all countries (Hollist, 1977*b*), as well as that the effects of specific variables may be obscured by complex interactions (McGinnis, 1991). These considerations also imply that a model that incorporates factors applicable to all countries is rather implausible.

The availability and reliability of data is also a prominent issue. In an attempt to tackle these issues, data in this thesis have been drawn from reliable and reputable sources. However, this choice is not unassailable, since reliable sources have also been found to be susceptible to inconsistences (Brauer, 2002). In addition, technical issues, such as currency conversions, corrections for inflation, changes introduced in newer versions of the same dataset or even intentional inaccuracies on the part of governments, may influence the results. The issue is further perplexed by the scarcity of reliable information about the estimation process that has been used in some of the historical data.

The scarcity of data due to their annual frequency also limits the number of statistical methods that can be applied. It also means that common methods that are used when large datasets are available cannot be used as such, at least not without additional considerations. This issue is apparent in earlier research using neural networks: All three previous studies that employed neural networks to study arms races (Andreou and Zombanakis, 2000, 2011; Refenes et al., 1995) opted for a train−test split, without a separate validation set. However, such an approach encourages the use of test data to make decisions about the model and, as such, it leads to an overestimation of model performance.

The absence of standardisation is a great challenge as well. Arms race research is characterised by an absence of standardisation as far as a definition is concerned. The approach chosen in this thesis is that of defence spending dynamics, which is the approach commonly taken in related literature. Alternative approaches, especially these based on qualitative criteria, were avoided since they are rather complicated. However, the spending dynamics approach chooses military expenditure as a measure of armaments, which may fail to capture some aspects of armaments.

An accurate interpretation of results demands a proper understanding of SHAP values. SHAP values attempt to explain the particular neural network model and not the true data generating process. In other words, SHAP produces feature attributions that are true to the

model, but not necessarily true to the data. Therefore, it is assumed that, for the models trained in this thesis, the process they model constitutes a reliable approximation of the true process, and as such, feature attributions are true to the data as well.

As in all models, specification bias is a fundamental concern. In this thesis, the analysis is performed on an already trained model, which is perceived as an accurate representation of the underlying process that defines spending. However, in reality, this process may actually not exist (in the sense of being able to be mathematically modelled) or depart significantly from the process modelled here. This is what is commonly termed *specification bias* (Zapranis and Refenes, 1999). Specification bias may emerge when relevant variables are omitted, irrelevant variables are included, the learning algorithm, model selection method or functional form are inefficient as well as when the model variables are characterised by measurement errors (Zapranis and Refenes, 1999).

Of course, the fact that the models developed may not be a perfectly accurate representation of the phenomena being modelled does not imply that they are not useful. After all, reality, especially when human actors are involved, is particularly complicated. As Box et al. (2009) aptly state:

> All models are approximations. Assumptions, whether implied or clearly stated, are never exactly true. All models are wrong, but some models are useful. So the question you need to ask is not 'Is the model true?' (it never is) but 'Is the model good enough for this particular application?'

In the specific case of defence expenditure modelling, it has already been shown that different models, using a different mix of variables, may be equally capable of explaining military expenditure (Wallace and Wilson, 1978).

The statistical considerations outlined in the previous paragraphs may be the underlying reason for the contradictory results reported by researchers on the issue of the Greek–Turkish arms race. It is also interesting that the results of Model C contradict those attained by Andreou and Zombanakis (2000, 2011) and Refenes et al. (1995), in that Turkey-related factors are not prime predictors of Greek spending.

Furthermore, an interesting fact that became apparent in Chapter 2 is that most approaches taken by researchers revolve around traditional econometrics, whereas other methods, including neural networks, have not been widely used. In the search for the underlying reasons for this, only assumptions can be made: Researchers might be uncertain about the effectiveness of these methods or unfamiliar with them.

Nevertheless, it appears that neural networks are not the optimal solution when the available dataset is small, as in this case. Of course, this applies to many classical methods as well and highlights the need for modern sophisticated algorithms, that are able to circumvent the restrictions posed by limited data.

Interest in arms races and their consequences has also abated since the end of Cold War. This diminished interest in arms race research is not conducive to the emergence of new approaches and to a critical approach to already existing ones. It is also apparent that the wider considerations about the suitability of using quantitative analysis to model human behaviour will not cease to exist, since using quantitative analysis means that results are inescapably obtained through a process of simplifying complex social concepts.

Taking into consideration all the issues outlined above, this thesis highlights critical issues in arms race research and may serve as a solid framework for the study of arms races. Greek policymakers would probably benefit from this analysis, which provides a complete insight into the evolution and structure of Greek defence expediture during the last 60 years. The analysis could also be beneficial to Turkish policymakers, who strive to uncover the underlying mechanism that defines Greek spending and who may perceive increases in Greek spending due to domestic factors as hostile. Irrespective of arms race theory, Model C could also be used merely for its predictive power.

CHAPTER 6

CONCLUSIONS

This thesis examined the defence spending dynamics between Greece and Turkey in the context of arms race research and neural networks. It strived to answer the question whether an arms race between the two countries exists, using neural networks for years 1963–2018. The objective of this research was multidimensional: to provide an overview of neural networks deployment in arms race research; to uncover methodological issues and examine statistical considerations involved when using neural networks to examine the existence of an arms race between rivals; to reignite interest in the use of neural networks and machine learning methods in arms race research.

Three different models, each one including different variables, were developed: A model based on original research by Richardson (1960*a*), a model based on contemporary research by Refenes et al. (1995) and a model developed by the author. 'Arms racing' was defined in the context of defence spending analysis.

The main findings can be summarised as follows:

(a) Arms races are a highly complex issue, especially given the fact that human actors and their decisions are involved.

(b) Greece apparently (given the limitations analysed in Chapter 5) does not engage in an action–reaction relationship with Turkey, since the annual change in Greek defence spending is principally determined by internal factors.

(c) Methodological and statistical issues —such as the correlation between input variables, the choice of proper significance levels and input variables, the availability and reliability of data— exert considerable influence on the results, but they have not been given proper attention by researchers.

(d) Neural network implementations in particular do not produce perfect results when used in arms race research and they are not devoid of most shortcomings that affect traditional methods. However, they have the potential to offer an advantage over traditional methods, although research is needed to define the proper variable selection process and establish algorithms that are able to leverage small datasets.

As illustrated in Chapter 2, researchers have questioned whether the arms race literature actually produced substantial knowledge. This thesis provides a solid framework for researchers, highlighting model design and estimation issues that should always be taken into consideration when using neural networks in arms race research.

Neural networks are definitely not a remedy for all research objectives. However, when used cautiously and given constant advancements in algorithms and input significance techniques, they are capable of providing invaluable insights. This is also valid for the particular case of the Greek–Turkish arms race. Moreover, the models developed in this particular case may also be used merely for their predictive power, besides being used as a medium to examine the existence of an arms race.

The ever-asked question of whether an arms race exists may remain unanswered: Statistics gives no definitive answer to the question, since related research is plagued by a series of methodological and interpretation issues. It is probably this reason that literature presents no definitive conclusion on the issue. Nevertheless, ensuring data quality (and research reproducibility) is much more essential than beating benchmarks by 1% or 2%, which may not add significant value to already existing research.

Limitations and recommendations for future research

This research is bound to a set of limitations, most of which have been outlined throughout this thesis. The most prominent example is the statistical limitations inherent in neural network models, especially the hypothesis of independent and identically distributed training datasets. This assumption implies that all samples stem from the same generative process and this generative process is assumed to have no memory of past generated samples. This hypothesis, especially for time series data, is unfortunately highly implausible. Limitations also apply for the data and most methodological processes employed in this thesis; relevant considerations have been made in previous chapters.

As far as future research is concerned, developments in machine learning during the last decade as well as the increasing availability of computational power translate to a greater potential for all research realms, including arms race research. More specifically, the following directions can be proposed for future research:

(a) In an era of big data and machine learning, it would be interesting to incorporate social and psychological parameters into arms race models. Sophisticated algorithms, such as algorithms for Natural Language Processing, could be applied on data obtained from social media in an attempt to track social sentiment. In addition, data manipulation could be facili-

tated by newer, modern methods of machine learning, taking advantage of the availability of increased computational power.

(b) The digitisation of public administration enables researchers to obtain more detailed data on public expenditure, including military spending data. The availability of highly disaggregated military spending data would enable researchers to define armaments in a stricter sense, thus eliminating concerns about the use of defence expenditure as a proxy for armaments. Of course, more data translates to more opportunities for modelling in general.

(c) Arms race modelling could be applied to means for electronic warfare, which is arguably expected to obtain increasing significance in 21st century military operations.

(d) Arms race modelling has frequently disregarded the theoretical background of arms races, focussing on statistical measures and models. A critical analysis of theoretical aspects, in the light of modern political science theories, would be beneficial to arms races theory.

(e) Arms race modelling could be given a renewed potential by the use of modern, more advanced machine learning models, which are not dependent upon the i.i.d. hypothesis and are also capable of resolving the issues arising from the small number of available observations.

(f) The development and use of methods for input significance analysis which are not affected by correlations between variables would also be particularly beneficial for arms race research.

The points made above do not imply that traditional econometrics ought to be disregarded. Modern econometric methods should also be considered in conjunction with more sophisticated machine learning ones. Emphasis should also be placed on creating algorithms based on much smaller datasets than those commonly used in machine learning research, rather than solely aiming to obtain larger datasets. Focussing on data quality (instead of quantity) and working on standardisation should also consitute prime priorities for machine learning research in general, so that results are statistically robust. Furthermore, researchers should be clear about the methodologies they use and their limitations.

Closing remarks

Greek–Turkish relations have always been particularly complex, fluctuating between periods of escalation and periods of rapprochement. Statistical modelling, whether it refers to traditional regression analysis or to modern machine learning methods, is merely an attempt to uncover and examine such complex relationships, in order to better understand them and provide solutions

to problems. However, in the end, all models are approximations and, thus, research should strive to find the best and most faithful of these approximations. As Richardson (1960*a*) aptly stated for his arms race model, 'The equations are merely a description of what people would do if they did not stop to think.' In the specific case of Greek–Turkish rivalry, settling disputes would undoubtedly be beneficial for both countries, given the detrimental ramifications of these disputes in the recent past and the turbulence of the region in which both countries are located.

REFERENCES

Aas, K., Jullum, M. and Løland, A. (2019), Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. arXiv preprint, arXiv:1903.10464.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. (2015), 'TensorFlow: Large-scale machine learning on heterogeneous systems'. Software available from tensorflow.org.
URL: *https://www.tensorflow.org/*

Abdel-Hamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G. and Yu, D. (2014), 'Convolutional neural networks for speech recognition', *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22(10), 1533–1545. doi: 10.1109/TASLP.2014.2339736.

Adithya, V., Vinod, P. R. and Gopalakrishnan, U. (2013), Artificial neural network based method for Indian sign language recognition, *in* '2013 IEEE Conference on Information Communication Technologies', pp. 1080–1085. doi: 10.1109/CICT.2013.6558259.

Albalate, D., Bel, G. and Elias, F. (2012), 'Institutional determinants of military spending', *Journal of Comparative Economics* 40(2), 279–290. doi: 10.1016/j.jce.2011.12.006.

Anderton, C. H. (1989), 'Arms race modeling: Problems and prospects', *Journal of Conflict Resolution* 33(2), 346–367.

Andreou, A. S. and Zombanakis, G. A. (2000), 'Financial versus human resources in the Greek – Turkish arms race: A forecasting investigation using artificial neural networks', *Defence and Peace Economics* 11(2), 403–426. doi: 10.1080/10430710008404956.

Andreou, A. S. and Zombanakis, G. A. (2011), 'Financial versus human resources in the Greek–Turkish arms race 10 years on: A forecasting investigation using artificial neural networks', *Defence and Peace Economics* 22(4), 459–469. doi: 10.1080/10242694.2010.539858.

Andreou, A. S. and Zombanakis, G. A., eds (2003), *Intelligent Information Systems Applied to Complicated Defence Problems: The Greek – Turkish Arms Race and the Cyprus Issue*, Papazisis, Athens.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016), 'Machine bias', *ProPublica* . Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 11 February 2021].

Antwarg, L., Miller, R. M., Shapira, B. and Rokach, L. (2019), Explaining anomalies detected by autoencoders using SHAP. arXiv preprint, arXiv:1903.02407.

Arrow, K. (1960), Decision theory and the choice of a level of significance for the t-test, *in* I. Olkin, S. G. Ghurye, W. Hoeding, W. G. Madow and H. B. Mann, eds, 'Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling', Stanford University Press, Stanford, pp. 70–78.

Athanassiou, E. and Kollias, C. (2000), Military Tension and Foreign Direct Investment: Evidence from the Greek-Turkish Rivalry, *in* J. Brauer and K. Hartley, eds, 'The Economics of Regional Security: NATO, the Mediterranean, and Southern Africa', Harwood Academic Publishers, Amsterdam.

Athanassiou, E., Kollias, C., Nikolaidou, E. and Zografakis, S. (2002), Greece: Military Expenditure, Economic Growth, and the Opportunity Cost of Defense, *in* J. Brauer and J. P. Dunne, eds, 'Arming the South: The Economics of Military Expenditure, Arms Production and Arms Trade in Developing Countries', Palgrave, New York, pp. 291–317. doi: 10.1057/9780230501256_14.

Avramides, C. A. (1997), 'Alternative models of Greek defence expenditures', *Defence and Peace Economics* 8(2), 145–187. doi: 10.1080/10430719708404874.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R. and Samek, W. (2015), 'On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation', *PLoS ONE* 10(7). doi: 10.1371/journal.pone.0130140.

Balabin, R. M. and Lomakina, E. I. (2009), 'Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies', *The Journal of Chemical Physics* 131(7). doi: 10.1063/1.3206326.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020), 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI', *Information Fusion* 58, 82–115. doi: 10.1016/j.inffus.2019.12.012.

Batchelor, P., Dunne, P. and Lamb, G. (2002), 'The demand for military spending in South Africa', *Journal of Peace Research* 39(3), 339–354. doi: 10.1177/0022343302039003005.

Bellany, I. (1975), 'The Richardson theory of 'arms races': Themes and variations', *British Journal of International Studies* 1(2), 119–130. doi: 10.1017/S0260210500116468.

Benati, S., López-Blázquez, F. and Puerto, J. (2019), 'A stochastic approach to approximate values in cooperative games', *European Journal of Operational Research* 279, 93–106. doi: 10.1016/j.ejor.2019.05.027.

Bishop, C. M., ed. (1995), *Neural Networks For Pattern Recognition*, Clarendon Press, Oxford.

Boulding, K. E., ed. (1962), *Conflict and Defense*, Harper & Row, New York.

Box, G. E. P., Luceño, A. and Paniagua-Quiñones, M., eds (2009), *Statistical Control by Monitoring and Adjustment*, 2nd edn, Wiley, New York.

Brauer, J. (2002), 'Survey and review of the defense economics literature on Greece and Turkey: What have we learned?', *Defence and Peace Economics* 13(2), 85–107. doi: 10.1080/10242690210969.

Brauer, J. (2007), 'Data, models, coefficients: The case of United States military expenditure', *Conflict Management and Peace Science* 24(1), 55–64. doi: 10.1080/07388940601102845.

Brzoska, M. (1981), 'The reporting of military expenditures', *Journal of Peace Research* 18(3), 261–275. doi: 10.1177/002234338101800303.

Buchanan, B. G. (2019), 'Artificial intelligence in finance'. doi: 10.5281/zenodo.2626454.

Busch, P. A. (1970), Mathematical Models of Arms Races, *in* B. M. Russett, ed., 'What Price Vigilance?', Yale University Press, New Haven, pp. 193–233.

Buzan, B. (1983), *People, States and Fear: the National Security Problem in International Relations*, Harvester, Brighton.

Buzan, B. (1987), *An Introduction to Strategic Studies: Military Technology and International Relations*, Macmillan, London.

Caliskan, A., Bryson, J. J. and Narayanan, A. (2017), 'Semantics derived automatically from language corpora contain human-like biases', *Science* 356(6334), 183–186. doi: 10.1126/science.aal4230.

Çandar, C. (2019), 'Η νέα Τουρκία', 'Ο νέος εθνικισμός': 'Οι πολιτικές για την Μέση Ανατολή και οι Κούρδοι ['The new Turkey', 'The new nationalism': The policies regarding the Middle East and the Kurds], *in* H. Millas, ed., 'Η νέα Τουρκία εκ των έσω [The New Turkey from inside]', Sideris, Athens.

Cao, M., Alkayem, N. F., Pan, L. and Novák, D. (2016), 'Advanced methods in neural networks-based sensitivity analysis with their applications in civil engineering', *Artificial Neural Networks – Models and Applications* . doi: 10.5772/64026.

Carlini, N. and Wagner, D. (2017), Adversarial examples are not easily detected: Bypassing ten detection methods, *in* 'Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security', pp. 3–14. doi: 10.1145/3128572.3140444.

Caspary, W. (1967), 'Richardson's model of arms races: Description, critique and an alternative model', *International Studies Quarterly* 2(10), 63–88. doi: 10.2307/3013990.

Castro, J., Gómez, D., Molina, E. and Tejada, J. (2017), 'Improving polynomial estimation of the Shapley value by stratified random sampling with optimum allocation', *Computers & Operations Research* 82, 180–188. doi: 10.1016/j.cor.2017.01.019.

Castro, J., Gómez, D. and Tejada, J. (2009), 'Polynomial calculation of the Shapley value based on sampling', *Computers & Operations Research* 36(5), 1726–1730. doi: 10.1016/j.cor.2008.04.004.

Caulkins, J., Cohen, J., Gorr, W. and Wei, J. (1996), 'Predicting criminal recidivism: A comparison of neural network models with statistical methods', *Journal of Criminal Justice* 24(3), 227–240. doi: 10.1016/0047-2352(96)00012-8.

Chalikias, M. and Skordoulis, M. (2014), 'Implementation of Richardson's arms race model in advertising expenditure of two competitive firms', *Applied Mathematical Sciences* 8(81), 4013–4023. doi: 10.12988/ams.2014.45336.

Clark, J., Koprinska, I. and Poon, J. (2003), A neural network based approach to automated e-mail classification, *in* 'Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI 2003)', pp. 702–705. doi: 10.1109/WI.2003.1241300.

Collier, P. and Hoeffler, A. (2007), 'Unintended consequences: Does aid promote arms races?', *Oxford Bulletin of Economics and Statistics* 69(1), 1–27. doi: 10.1111/j.1468-0084.2006.00439.x.

Constas, D., ed. (1991), *The Greek–Turkish Conflict in the 1990s*, Macmillan, London.

Cusack, T. R. and Ward, M. D. (1981), 'Military spending in the United States, Soviet Union, and the People's Republic of China', *Journal of Conflict Resolution* 25(3), 429–469. doi: 10.1177/002200278102500303.

Cybenko, G. (1989), 'Approximation by superpositions of a sigmoidal function', *Mathematics of Control, Signals and Systems* 2(4), 303–314. doi: 10.1007/BF02551274.

Datta, A., Sen, S. and Zick, Y. (2016), Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems, *in* '2016 IEEE Symposium on Security and Privacy (SP)', pp. 598–617. doi: 10.1109/SP.2016.42.

Deger, S. and Sen, S. (1983), 'Military expenditure, spin-off and economic development', *Journal of Development Economics* 13(1–2), 67–83. doi: 10.1016/0304-3878(83)90050-0.

Demirel, S. (1998), 'The need for dialogue', *Harvard International Review* 21(1), 24–26.

Diehl, P. F. (1983), 'Arms races and escalations: A closer look', *Journal of Peace Research* 20(3), 205–212. doi: 10.1177/002234338302000301.

Dritsakis, N. (2004), 'Defense spending and economic growth: An empirical investigation for Greece and Turkey', *Journal of Policy Modeling* 26(2), 249–264. doi: 10.1016/j.jpolmod.2004.03.011.

Du, K. L. and Swamy, M. N. S. (2019), *Neural Networks and Statistical Learning*, 2nd edn, Springer, London. doi: 10.1007/978-1-4471-7452-3.

Duchi, J., Hazan, E. and Singer, Y. (2011), 'Adaptive subgradient methods for online learning and stochastic optimization', *Journal of Machine Learning Research* 12, 2121–2159.

Dundar, F. (1999), *Türkiye Nüfus Sayımlarında Azınlıklar [Minorities in the Turkish Census]*, Doz Yayınları, Istanbul.

Dunne, J. P., Nikolaidou, E. and Mylonidis, N. (2003), 'The demand for military spending in the peripheral economies of Europe', *Defence and Peace Economics* 14(6), 447–460. doi: 10.1080/1024269032000085215.

Dunne, J. P., Nikolaidou, E. and Smith, R. P. (2005), 'Is there an arms race between Greece and Turkey?', *Peace Economics, Peace Science and Public Policy* 11(2), 1–35. doi: 10.2202/1554-8597.1086.

Dunne, J. P. and Perlo-Freeman, S. (2003*a*), 'The demand for military spending in developing countries', *International Review of Applied Economics* 17(1), 23–48. doi: 10.1080/713673166.

Dunne, J. P. and Perlo-Freeman, S. (2003*b*), 'The demand for military spending in developing countries: A dynamic panel analysis', *Defence and Peace Economics* 14(6), 461–474. doi: 10.1080/1024269032000085224.

Dunne, J. P., Perlo-Freeman, S. and Smith, R. P. (2008), 'The demand for military spending in developing countries: Hostility versus capability', *Defence and Peace Economics* 19(4), 293–302. doi: 10.1080/10242690802166566.

Dunne, J. P. and Smith, R. P. (2007), The Econometrics of Military Arms Races, *in* K. Hartley and T. Sandler, eds, 'Handbook of Defence Economics', Elsevier, Amsterdam, pp. 913–940.

Dunne, P. and Nikolaidou, E. (2001), 'Military expenditure and economic growth: A demand and supply model for Greece, 1960–96', *Defence and Peace Economics* 12(1), 47–67. doi: 10.1080/10430710108404976.

Dwarakish, G. S., Rakshith, S. and Natesan, U. (2013), 'Review on applications of neural network in coastal engineering', *Artificial Intelligent Systems and Machine Learning* 5(7), 324–331.

Eichenberg, R. and Stoll, R. (2003), 'Democratic control of the defense budget in the United States and Western Europe', *Journal of Conflict Resolution* 47(4), 399–422. doi: 10.1177/0022002703254477.

European Commission (2020), *AMECO Online - Gross Public Debt*. Available at: https://dashboard.tech.ec.europa.eu/qs_digit_dashboard_mt/public/sense/app/667e9fba-eea7-4d17-abf0-ef20f6994336/sheet/2f9f3ab7-09e9-4665-92d1-de9ead91fac7/state/analysis [Accessed 23 July 2020].

Ferejohn, J. A. (1976), On the effects of aid to nations in arms races: The Richardson model, *in* D. A. Zinnes and J. V. Gillespie, eds, 'Mathematical Models in International Relations', Praeger, New York, pp. 218–251.

Flores, A. Q. (2011), 'Alliances as contiguity in spatial models of military expenditures', *Conflict Management and Peace Science* 28(4), 402–418. doi: 10.1177/0738894211413064.

Fu, K., Cheng, D., Tu, Y. and Zhang, L. (2016), Credit card fraud detection using convolutional neural networks, *in* A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee and D. Liu, eds, 'Neural Information Processing. ICONIP 2016. Lecture Notes in Computer Science', Vol. 9949, Springer, Cham, pp. 483–490. doi: 10.1007/978-3-319-46675-0_53.

Funahashi, K.-I. (1989), 'On the approximate realization of continuous mappings by neural networks', *Neural Networks* 2(3), 183–192. doi: 10.1016/0893-6080(89)90003-8.

Ganesan, D. N., Venkatesh, D. K., Rama, M. A. and Malathi Palani, A. (2010), 'Application of neural networks in diagnosing cancer disease using demographic data', *International Journal of Computer Applications* 1(26), 81–97. doi: 10.5120/476-783.

Garson, G. D. (1991), 'Interpreting neural network connection weights', *Artificial Intelligence Expert* 6, 47–51.

Gedeon, T. D. (1997), 'Data mining of inputs: Analysing magnitude and functional measures', *International Journal of Neural Systems* 8(2), 209–218. doi: 10.1142/S0129065797000227.

Georgiou, G. M. (1990), 'Is there an arms race between Greece and Turkey? Some preliminary results', *Cyprus Journal of Economics* 3(1), 58–73.

Georgiou, G. M., Kapopoulos, P. T. and Lazaretou, S. (1996), 'Modelling Greek–Turkish rivalry: An empirical investigation of defence spending dynamics', *Journal of Peace Research* 33(2), 229–239. doi: 10.1177/0022343396033002008.

Gevrey, M., Dimopoulos, I. and Lek, S. (2003), 'Review and comparison of methods to study the contribution of variables in artificial neural network models', *Ecological Modelling* 160(3), 249–264. doi: 10.1016/S0304-3800(02)00257-0.

Gibler, D. M., Rider, T. J. and Hutchison, M. L. (2005), 'Taking arms against a sea of troubles: Conventional arms races during periods of rivalry', *Journal of Peace Research* 42(2), 131–147.

Glaser, C. L. (2004), 'When are arms races dangerous? Rational versus suboptimal arming', *International Security* 28(4), 44–84. doi: 10.2307/4137449.

Gluck, M. A. and Bower, G. H. (1988), 'Evaluating an adaptive network model of human learning', *Journal of Memory and Language* 27(2), 166–195. doi: 10.1016/0749-596X(88)90072-1.

Goh, A. T. C. (1995), 'Back-propagation neural networks for modeling complex systems', *Artificial Intelligence in Engineering* 9(3), 143–151. doi: 10.1016/0954-1810(94)00011-S.

Gonzalez, R. A. and Mehay, S. L. (1990), 'Publicness, scale, and spillover effects in defense spending', *Public Finance Quarterly* 18(3), 273–290. doi: 10.1177/109114219001800302.

Goodman, B. and Flaxman, S. (2016), European Union regulations on algorithmic decision-making and a "right to explanation". arXiv preprint, arXiv:1606.08813.

Google developers (2021), *tf.keras.optimizers.Adam | TensorFlow Core v2.4.1*. Available at: https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam [Accessed 22 March 2021].

Granger, R., Ambros-Ingerson, J., Staubli, U. and Lynch, G. (1989), Memorial Operation of Multiple, Interacting Simulated Brain Structures, *in* M. A. Gluck and D. E. Rumelhart, eds, 'Neuroscience and Connectionist Theory', Lawrence Erlbaum Associates, Hillsdale, pp. 95–129.

Graves, A. and Schmidhuber, J. (2009), Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks, *in* D. Koller, D. Schuurmans, Y. Bengio and L. Bottou, eds, 'Advances in Neural Information Processing Systems', Vol. 21, pp. 545–552.

Gray, C. S. (1971), 'The arms race phenomenon', *World Politics* 24(1), 39–79. doi: 10.2307/2009706.

Grigoriadis, I. N. (2011), 'Redefining the nation: Shifting boundaries of the 'other' in Greece and Turkey', *Middle Eastern Studies* 47(1), 167–182. doi: 10.1080/00263206.2011.536632.

Gupta, S., De Mello, L. and Sharan, R. (2001), 'Corruption and military spending', *European Journal of Political Economy* 17(4), 749–777.

Hammond, G. T. (1993), *Plowshares into Swords: Arms Races in International Politics, 1840–1991*, South Carolina Press, Columbia.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H.,

Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T. E. (2020), 'Array programming with NumPy', *Nature* 585, 357 – 362. doi: 10.1038/s41586-020-2649-2.

Heraclides, A. (1980), 'Socialization to conflict: A case study of the ingroup – outgroup images in the educational system of Greece', *The Greek Review of Social Research* 38, 16 – 42.

Heraclides, A. (2012), "What will become of us without barbarians?' The enduring Greek – Turkish rivalry as an identity-based conflict', *Southeast European and Black Sea Studies* 12(1), 115 – 134. doi: 10.1080/14683857.2012.661944.

Heraclides, A. (2019), Greek – Turkish relations and conflict: A bird's eye view, *in* A. Heraclides and G. Alioğlu Çakmak, eds, 'Greece and Turkey in Conflict and Cooperation: From Europeanization to de-Europeanization', Routledge Advances in European Politics, Routledge, New York, pp. 3 – 12.

Heraclides, A., ed. (2010), *The Greek – Turkish Conflict in the Aegean: Imagined Enemies*, Palgrave Macmillan, Hampshire.

Hewitt, D. (1992), 'Military expenditures worldwide: Determinants and trends, 1972 – 1988', *Journal of Public Policy* 12(2), 105 – 152. doi: 10.1017/S0143814X00005080.

Hill, W. W. (1978), 'A time-lagged Richardson arms race model', *Journal of Peace Science* 3(1), 55 – 62.

Hill, W. W. (1992), 'Several sequential augmentations of Richardson's arms race model', *Mathematical and Computer Modelling* 16(8-9), 201 – 212. doi: 10.1016/0895-7177(92)90096-4.

Hinton, G. (2014), rmsprop: Divide the gradient by a running average of its recent magnitude. Lecture 6 of the online course 'Neural Networks for Machine Learning'. Available at: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf [Accessed 2 February 2021].

Hollist, W. L. (1977*a*), 'Alternative explanations of competitive arms processes: Tests on four pairs of nations', *American Journal of Political Science* 21(2), 313 – 340. doi: 10.2307/2110497.

Hollist, W. L. (1977*b*), 'An analysis of arms processes in the United States and the Soviet Union', *International Studies Quarterly* 21(3), 503 – 528. doi: 10.2307/2600235.

Horn, M. D. (1987), Arms Races and the International System, PhD thesis, Department of Political Science, University of Rochester, Rochester.

Hornik, K., Stinchcombe, M. and White, H. (1989), 'Multilayer feedforward networks are universal approximators', *Neural Networks* 2(5), 359–366. doi: 10.1016/0893-6080(89)90020-8.

Huang, G. B. (2003), 'Learning capability and storage capacity of two-hidden-layer feedforward networks', *IEEE Transactions on Neural Networks* 14(2), 274–281. doi: 10.1109/TNN.2003.809401.

Hunter, J. D. (2007), 'Matplotlib: A 2D graphics environment', *Computing in Science & Engineering* 9(3), 90–95. doi: 10.1109/MCSE.2007.55.

Huntington, S. P. (1958), 'Arms races: Prerequisites and results', *Public Policy* 8(1), 41–86.

Ifestos, P. and Platias, A., eds (1992), *Ελληνική Αποτρεπτική Στρατηγική [Greek Deterrence Strategy]*, Papazisis, Athens.

Insel, A. (2019), Η δικαιοσύνη και οι συνέπειές της [Justice and its consequences], *in* H. Millas, ed., 'Η νέα Τουρκία εκ των έσω [The New Turkey from inside]', Sideris, Athens.

International Monetary Fund [IMF] (2020), *Historical Public Debt - Time Series - IMF DATA*. Available at: https://data.imf.org/?sk=c9494a18-c15e-4294-90f4-6bb7b30e95a4&hide_uv=1 [Accessed 22 July 2020].

Intriligator, M. D. (1975), 'Strategic considerations in the Richardson model of arms races', *Journal of Political Economy* 83(2), 339–354.

Intriligator, M. D. and Brito, D. L. (1984), 'Can arms races lead to the outbreak of war?', *Journal of Conflict Resolution* 28(1), 63–84.

Jervis, R. (1976), *Perception and Misperception in International Politics*, Princeton University Press, Princeton.

Kadera, K. M., Crescenzi, M. and Zinnes, D. A. (2020), Richardson and the Study of Dynamic Conflict Processes, *in* N. P. Gleditsch, ed., 'Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences', Vol. 27 of *Pioneers in Arts, Humanities, Science, Engineering, Practice*, Springer, Heidelberg, pp. 45–56. doi: 10.1007/978-3-030-31589-4_5.

Kapopoulos, P. T. and Lazaretou, S. (1993), 'Modelling the demand for Greek defence expenditure: An error correction approach', *Cyprus Journal of Economics* 6(1), 73–86.

Katsaitis, O., Kondylis, K. and Zombanakis, G. A. (2019), 'Concerns on the issue of defence expenditure in the post-crisis Greece', *Security and Defence Quarterly* 24(2), 177–201. doi: 10.35467/sdq/103408.

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H. and Wortman Vaughan, J. (2020), Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning, *in* 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', pp. 1–14. doi: 10.1145/3313831.3376219.

Keskar, N. S. and Socher, R. (2017), Improving generalization performance by switching from Adam to SGD. arXiv preprint, arXiv:1712.07628.

Keuzenkamp, H. A. and Magnus, J. R. (1995), 'On tests and significance in econometrics', *Journal of Econometrics* 67(1), 5–24. doi: 10.1016/0304-4076(94)01624-9.

Keyder, Ç. (2005), A history and geography of Turkish nationalism, *in* F. Birtek and T. Dragonas, eds, 'Citizenship, the nation–state in Greece, Turkey', Routledge, London, pp. 3–17.

Kim, H.-C., Kim, H. M. and Lee, J. (2013), 'The post-coup military spending question revisited, 1960–2000', *International Interactions: Empirical and Theoretical Research in International Relations* 39(3), 367–385. doi: 10.1080/03050629.2013.782305.

Kim, J. H. (2020), 'Decision-theoretic hypothesis testing: A primer with R package OptSig', *The American Statistician* 74(4), 370–379. doi: 10.1080/00031305.2020.1750484.

Kim, J. H. and Choi, I. (2019), 'Choosing the level of significance: A decision-theoretic approach', *Abacus* 57(1), 27–71. doi: 10.1111/abac.12172.

Kingma, D. P. and Ba, J. (2014), Adam: A method for stochastic optimization. arXiv preprint, arXiv:1412.6980.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S. and Willing, C. (2016), Jupyter Notebooks – a publishing format for reproducible computational workflows, *in* F. Loizides and B. Schmidt, eds, 'Positioning and Power in Academic Publishing: Players, Agents and Agendas', IOS Press, pp. 87–90.

Kollias, C. (1991), 'Greece and Turkey: The case study of an arms race from the Greek perspective', *Spoudai* 41(1), 64–81.

Kollias, C. (1994), 'Is there a Greek–Turkish arms race? The view from Athens', *Cyprus Journal of Economics* 7(1), 52–62.

Kollias, C. (1996), 'The Greek–Turkish conflict and Greek military expenditure 1960–92', *Journal of Peace Research* 33(2), 217–228. doi: 10.1177/0022343396033002007.

Kollias, C. (2018), Η αόρατος χειρ και η σιδηρά πυγμή: βίοι παράλληλοι [The invisible hand and the iron fist: parallel lives], *in* C. Grigorakou, ed., 'Proceedings of the 1st Finance Corps Conference of the Hellenic Army', Hellenic Army General Staff/Finance Directorate, Athens, pp. 25–34.

Kollias, C. and Makrydakis, S. (1997), 'Is there a Greek–Turkish arms race? Evidence from cointegration and causality tests', *Defence and Peace Economics* 8(4), 355–379.

Kollias, C. and Paleologou, S. M. (2002), 'Is there a Greek–Turkish arms race? Some further empirical results from causality tests', *Defence and Peace Economics* 13(4), 321–328. doi: 10.1080/10242690212357.

Kollias, C. and Paleologou, S. M. (2003), 'Determinants of the demand for Greek military expenditure', *Defence and Peace Economics* 14(6), 437–445. doi: 10.1080/1024269032000085206.

Kollias, C., Paleologou, S. M. and Stergiou, A. (2016), 'Military expenditure in Greece: Security challenges and economic constraints', *The Economics of Peace and Security Journal* 11(1), 28–34. doi: 10.15355/epsj.11.1.28.

Krebs, R. R. (1999), 'Perverse institutionalism: NATO and the Greco–Turkish conflict', *International Organization* 53(2), 343–377. doi: 10.1162/002081899550904.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G. and Yu, H. (2017), 'Accountable algorithms', *University of Pennsylvania Law Review* 165, 633–705.

Kumar, I. E., Venkatasubramanian, S., Scheidegger, C. and Friedler, S. (2020), Problems with Shapley-value-based explanations as feature importance measures. arXiv preprint, arXiv:2002.11097.

Kutner, M. H., Nachtsheim, C. J. and Neter, J., eds (2004), *Applied Linear Regression Models*, 4th edn, McGraw-Hill/Irwin, Chicago.

Kydd, A. (2000), 'Arms races and arms control: Modeling the hawk perspective', *American Journal of Political Science* 44(2), 228–244.

Lambelet, J. C. (1973), 'Towards a dynamic two-theater model of the East–West arms race', *Journal of Peace Science* 1(1), 1–38. doi: 10.1177/073889427300100101.

Lambelet, J. C. (1986), 'The formal ('economic') analysis of arms races: What—if anything—have we learned since Richardson?', *Conflict Management and Peace Science* 9(2), 1–17. doi: 10.1177/073889428600900201.

Lambelet, J. C., Luterbacher, U. and Allan, P. (1979), 'Dynamics of arms races: Mutual stimulation vs. self-stimulation', *Journal of Peace Science* 4(1), 49–66. doi: 10.1177/073889427900400102.

Lanchester, F. W. (1916), *Aircraft in Warfare: The Dawn of the Fourth Arm*, Appleton, New York.

Larrabee, F. S. (2012), 'Greek–Turkish relations in an era of regional and global change', *Southeast European and Black Sea Studies* 12(4), 471–479. doi: 10.1080/14683857.2012.741843.

Lawrence, S., Giles, C. L., Tsoi, A. C. and Back, A. D. (1997), 'Face recognition: A convolutional neural-network approach', *IEEE Transactions on Neural Networks* 8(1), 98–113. doi: 10.1109/72.554195.

Lebovic, J. H. (1999), 'Using military spending data: The complexity of simple inference', *Journal of Peace Research* 36(6), 681–697. doi: 10.1177/0022343399036006005.

Lehmann, E. L. and Romano, J. S., eds (2005), *Testing Statistical Hypotheses*, 3rd edn, Springer, New York.

Lek, S., Belaud, A., Dimopoulos, I., Lauga, J. and Moreau, J. (1995), 'Improved estimation, using neural networks, of the food consumption of fish populations', *Marine and Freshwater Research* 46(8), 1229–1236. doi: 10.1071/MF9951229.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S. (1996), 'Application of neural networks to modelling nonlinear relationships in ecology', *Ecological Modelling* 90(1), 39–52. doi: 10.1016/0304-3800(95)00142-5.

Lipovetsky, S. and Conklin, M. (2001), 'Analysis of regression in game theory approach', *Applied Stochastic Models in Business and Industry* 17(4), 319–330. doi: 10.1002/asmb.446.

Looney, R. E. and Mehay, S. L. (1990), United States Defence Expenditures: Trends and Analysis, *in* K. Hartley and T. Sandler, eds, 'The Economics of Defense Spending: An International Survey', Routledge, London, pp. 13–39.

Loshchilov, I. and Hutter, F. (2019), Decoupled weight decay regularization, *in* 'International Conference on Learning Representations'.

Lucas, J., Sun, S., Zemel, R. and Grosse, R. (2019), Aggregated momentum: Stability through passive damping, *in* 'International Conference on Learning Representations'.

Lucier, C. E. (1979), 'Changes in the values of arms race parameters', *Journal of Conflict Resolution* 23(1), 17–39.

Lundberg, S. M. and Lee, S.-I. (2017), A unified approach to interpreting model predictions, *in* I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds, 'Advances in Neural Information Processing Systems 30', Curran Associates, Inc., pp. 4765–4774.

Luterbacher, U. (1975), 'Arms race models: Where do we stand?', *European Journal of Political Research* 3(2), 199–217. doi: 10.1111/j.1475-6765.1975.tb00525.x.

Luterbacher, U. (1985), 'The frustrated commentator: An evaluation of the work of Raymond Aron', *International Studies Quarterly* 29(1), 39–49. doi: 10.2307/2600478.

Ma, J. and Yarats, D. (2019), Quasi-hyperbolic momentum and Adam for deep learning, *in* 'International Conference on Learning Representations'.

Majeski, S. J. (1985), 'Expectations and arms races', *American Journal of Political Science* 29(2), 217–245.

Majeski, S. J. and Jones, D. L. (1981), 'Arms race modeling: Causality analysis and model specification', *The Journal of Conflict Resolution* 25(2), 259–288. doi: 10.1177/002200278102500203.

Masters, T., ed. (1994), *Practical Neural Network Recipes in C++*, Academic Press, San Diego.

Matthews, R. (1999), 'Greek–Turkish tensions fuel defence industrialization', *RUSI Journal* 144(1), 52–58. doi: 10.1080/03071849908446355.

Mayer, J., Khairy, K. and Howard, J. (2010), 'Drawing an elephant with four complex parameters', *American Journal of Physics* 78(6), 648–649. doi: 10.1119/1.3254017.

McCubbins, M. D. (1983), 'Policy components of arms competition', *American Journal of Political Science* 27(3), 385–406. doi: 10.2307/2110977.

McCulloch, W. S. and Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of Mathematical Biophysics* 5, 115–133. doi: 10.1007/BF02478259.

McGinnis, M. D. (1991), 'Richardson, rationality, and restrictive models of arms races', *Journal of Conflict Resolution* 35(3), 443–473. doi: 10.1177/0022002791035003003.

McKinney, W. (2010), Data Structures for Statistical Computing in Python, *in* 'Proceedings of the 9th Python in Science Conference', Vol. 445, Austin, pp. 51–56. doi: 10.25080/Majora-92bf1922-00a.

Mintz, A. (1988), *The Politics of Resource Allocation in the U.S. Department of Defense: International Crises and Domestic Constraints*, Westview, Boulder.

Mintz, A. (1989), 'Guns versus butter: A disaggregated analysis', *American Political Science Review* 83(4), 1285–1293. doi: 10.1017/S0003055400088158.

Mintz, A. and Hicks, A. (1984), 'Military keynesianism in the United States, 1949-1976: Disaggregating military expenditures and their determination', *American Journal of Sociology* 90(2), 411–417. doi: 10.1086/228086.

Mishra, A. K. and Desai, V. R. (2006), 'Drought forecasting using feed-forward recursive neural network', *Ecological Modelling* 198(1–2), 127–138. doi: 10.1016/j.ecolmodel.2006.04.017.

Mittelstadt, B., Russell, C. and Wachter, S. (2019), Explaining explanations in AI, *in* 'Proceedings of the Conference on Fairness, Accountability, and Transparency', pp. 279–288. doi: 10.1145/3287560.3287574.

Mokhtari, K. E., Higdon, B. P. and Başar, A. (2019), Interpreting financial time series with SHAP values, *in* 'Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering', CASCON '19, pp. 166–172.

Moll, K. D. and Luebbert, G. M. (1980), 'Arms race and military expenditure models', *Journal of Conflict Resolution* 24(1), 153–185. doi: 10.1177/002200278002400107.

Montaño, J. J. and Palmer, A. (2003), 'Numeric sensitivity analysis applied to feedforward neural networks', *Neural Computing & Applications* 12(2), 119–125. doi: 10.1007/s00521-003-0377-9.

Moosavi-Dezfooli, S., Fawzi, A., Fawzi, O. and Frossard, P. (2017), Universal adversarial perturbations, *in* '2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', pp. 86–94. doi: 10.1109/CVPR.2017.17.

Mourad, M. and Nehme, B. (2019), 'Economic determinants affecting military expenditures: Panel data analysis', *Global Journal of Science Frontier Research: (F) Mathematics and Decision Sciences* 19(3), 11–41.

Moustakis, F., ed. (2003), *The Greek–Turkish Relationship and NATO*, Frank Cass, London.

Nabian, M. A. and Meidani, H. (2018), 'Deep learning for accelerated seismic reliability analysis of transportation networks', *Computer-Aided Civil and Infrastructure Engineering* 33(6), 443–458. doi: 10.1111/mice.12359.

Nikolaidou, E. (2008), 'The demand for military expenditure: Evidence from the EU15 (1961–2015)', *Defence and Peace Economics* 19(4), 273–292. doi: 10.1080/10242690802166533.

Nix, R. and Zhang, J. (2017), Classification of Android apps and malware using deep neural networks, *in* '2017 International Joint Conference on Neural Networks (IJCNN)', pp. 1871–1878. doi: 10.1109/IJCNN.2017.7966078.

North Atlantic Treaty Organization [NATO] (2020), *NATO – Topic: Funding NATO*. Available at: https://www.nato.int/cps/ro/natohq/topics_67655.htm [Accessed 24 August 2020].

Öcal, N. (2002), 'Asymmetric effects of military expenditure between Turkey and Greece', *Defence and Peace Economics* 13(5), 405–416. doi: 10.1080/10242690213511.

Öcal, N. and Yildirim, J. (2009), 'Arms race between Turkey and Greece: A threshold cointegration analysis', *Defence and Peace Economics* 20(2), 123–129. doi: 10.1080/10242690801962254.

Ostrom, C. W. (1978), 'A reactive linkage model of the U.S. defense expenditure policymaking process', *American Political Science Review* 72(3), 941–957. doi: 10.2307/1955113.

Paparas, D., Richter, C. and Paparas, A. (2016), 'Military spending and economic growth in Greece and the arms race between Greece and Turkey', *Journal of Economics Library* 3(1), 38–56.

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B. and Swami, A. (2016), The limitations of deep learning in adversarial settings, *in* '2016 IEEE European Symposium on Security and Privacy (EuroS&P)', pp. 372–387. doi: 10.1109/EuroSP.2016.36.

Passi, S. and Jackson, S. J. (2018), Trust in data science: Collaboration, translation, and accountability in corporate data science projects, *in* 'Proceedings of the ACM Human–Computer Interaction', Vol. 2 (CSCW), Article 136, pp. 1–28. doi: 10.1145/3274405.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* 12, 2825–2830.

Platias, A. (1991), Greece's Strategic Doctrine: In Search of Autonomy and Deterrence, *in* D. Constas, ed., 'The Greek–Turkish Conflict in the 1990s', Macmillan, London, pp. 91–109.

Polak, E. (1991), *Computational Methods for Optimization*, Academic Press, New York.

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W. and Wallach, H. (2018), Manipulating and measuring model interpretability. arXiv preprint, arXiv:1802.07810.

Prechelt, L. (1998), 'Automatic early stopping using cross validation: Quantifying the criteria', *Neural Networks* 11(4), 761–767. doi: 10.1016/S0893-6080(98)00010-0.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn, Cambridge University Press, Cambridge.

Rapoport, A. (1957), 'Lewis F. Richardson's mathematical theory of war', *Journal of Conflict Resolution* 1(3), 249–299. doi: 10.1177/002200275700100301.

Rathi, S. (2019), Generating counterfactual and contrastive explanations using SHAP. arXiv preprint, arXiv:1906.09293.

Rattinger, H. (1975), 'Rüstung in Europa: Aufrüstung, Wettrüsten und andere Erklärungen [Armaments in Europe: Arming, arms races and other explanations]', *Österreichische Zeitschrift für Politikwissenschaft* 4, 231–250. doi: 10.20378/irbo-52935.

Rattinger, H. (1976*a*), 'Econometrics and arms races: A critical review and some extensions', *Journal of Political Research* 4(4), 421–439. doi: 10.1111/j.1475-6765.1976.tb00544.x.

Rattinger, H. (1976*b*), 'From war to war to war: Arms races in the Middle East', *International Studies Quarterly* 20(4), 501–531. doi: 10.2307/2600338.

Rattinger, H. (1984), Empirical Validation of Richardson Models of Arms Races, *in* R. Avenhaus and R. K. Huber, eds, 'Quantitative Assessment of Arms Control: Mathematical Modeling and Simulation in the Analysis of Arms Control Problems', Plenum Press, New York, pp. 179–203.

Reddi, S. J., Kale, S. and Kumar, S. (2019), On the convergence of Adam and beyond. arXiv preprint, arXiv:1904.09237.

Refenes, A. N., Kollias, C. and Zapranis, A. (1995), 'External security determinants of Greek military expenditure: An empirical investigation using neural networks', *Defence and Peace Economics* 6(1), 27–41. doi: 10.1080/10430719508404810.

Riaz, F. and Niazi, M. A. (2017), 'Towards social autonomous vehicles: Efficient collision avoidance scheme using Richardson's arms race model', *PLoS ONE* 12(10), 1–22. doi: 10.1371/journal.pone.0186103.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016), "Why should I trust you?" Explaining the predictions of any classifier, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', San Francisco, USA, pp. 1135–1144. doi: 10.1145/2939672.2939778.

Richardson, L. F. (1919), *The Mathematical Psychology of War*, Hunt, Oxford.

Richardson, L. F. (1960*a*), *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War*, Boxwood, Pittsburgh.

Richardson, L. F. (1960*b*), *Statistics of Deadly Quarrels*, Boxwood, Pittsburgh.

Rider, T. J., Findley, M. G. and Diehl, P. F. (2011), 'Just part of the game? Arms races, rivalry, and war', *Journal of Peace Research* 48(1), 85–100. doi: 10.1177/0022343310389505.

Ruder, S. (2016), An overview of gradient descent optimization algorithms. arXiv preprint, arXiv:1609.04747.

Rumelhart, D. and McClelland, J. (1986), *Parallel Distributed Processing*, MIT Press, Cambridge.

Russell, S. J. and Norvig, P. (2016), *Artificial Intelligence: A Modern Approach (Global Edition)*, Pearson, Harlow.

Saabas, A. (2014), 'Interpreting random forests'. Available at: http://blog.datadive.net/interpreting-random-forests/ [Accessed 24 November 2020].

Saaty, T. L. (1968), *Mathematical Models of Arms Control and Disarmament*, Wiley, New York.

Şahin, H. and Özsoy, O. (2008), 'Arms race between Greece and Turkey: A Markov switching approach', *Defence and Peace Economics* 19(3), 209–216. doi: 10.1080/10242690801972154.

Sample, S. G. (1997), 'Arms races and dispute escalation: Resolving the debate', *Journal of Peace Research* 34(1), 7–22. doi: 10.1177/0022343397034001002.

Saperstein, A. M. (1984), 'Chaos — A model for the outbreak of war', *Nature* 309, 303–305. doi: 10.1038/309303a0.

Sarker, I. H. (2021), 'Machine learning: Algorithms, real-world applications and research directions', *SN Computer Science* 2(160). doi: 10.1007/s42979-021-00592-x.

Sarle, W. S. (2000), 'How to measure importance of inputs?', Available at: ftp://ftp.sas.com/pub/neural/importance.html [Accessed 02 January 2021].

Schelling, T. C. (1966), *Arms and Influence*, Yale University Press, New Haven, CT.

scikit-learn developers (2020), *sklearn.preprocessing.RobustScaler – scikit-learn 0.24.1 documentation*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html [Accessed 22 March 2021].

Seabold, S. and Perktold, J. (2010), statsmodels: Econometric and statistical modeling with Python, *in* '9th Python in Science Conference'. doi: 10.25080/Majora-92bf1922-011.

Seiglie, C. (1992), Determinants of Military Expenditures, *in* W. Isard and C. H. Anderton, eds, 'Economics of Arms Reduction and the Peace Process', Elsevier, Amsterdam, pp. 183–202. doi: 10.1016/B978-0-444-88848-8.50013-0.

Seiglie, C. (2016), 'Openness of the economy, terms of trade, and arms', *Southern Economic Journal* 82(3), 748–759. doi: 10.1002/soej.12033.

Selbst, A. D. and Barocas, S. (2018), 'The intuitive appeal of explainable machines', *Fordham Law Review* 87, 1085–1139.

Senghaas, D. (1990), Arms Race Dynamics and Arms Control, *in* N. P. Gleditsch and O. Njølstad, eds, 'Arms Races: Technological and Political Dynamics', Sage, London, pp. 15–30.

Sengupta, N., Sahidullah, M. and Saha, G. (2016), 'Lung sound classification using cepstral-based statistical features', *Computers in Biology and Medicine* 75, 118–129. doi: 10.1016/j.compbiomed.2016.05.013.

Sezer, O. B. and Ozbayoglu, A. M. (2018), 'Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach', *Applied Soft Computing* 70, 525–538. doi: 10.1016/j.asoc.2018.04.024.

Shapley, L. S. (1953), A value for n-person games, *in* H. W. Kuhn and A. W. Tucker, eds, 'Contributions to the Theory of Games II', Princeton University Press, Princeton, pp. 307–317.

Shrikumar, A., Greenside, P. and Kundaje, A. (2017), Learning important features through propagating activation differences. arXiv preprint, arXiv:1704.02685.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. and Hassabis, D. (2016), 'Mastering the game of Go with deep neural networks and tree search', *Nature* 529(7587), 484–489. doi: 10.1038/nature16961.

Singer, J. D., Bremer, S. and Stuckey, J. (1972), Capability distribution, uncertainty, and major power war, 1820–1965, *in* B. Russett, ed., 'Peace, War, and Numbers', Sage, Beverly Hills, pp. 19–48. National Material Capabilities data set, version 5.0.

Slack, D., Hilgard, S., Jia, E., Singh, S. and Lakkaraju, H. (2020), Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods, *in* 'Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society', AIES '20, pp. 180–186. doi: 10.1145/3375627.3375830.

Smith, R. (1989), 'Models of military expenditure', *Journal of Applied Econometrics* 4(4), 345–359. doi: 10.1002/jae.3950040404.

Smith, R. (1995), The Demand for Military Expenditure, *in* K. Hartley and T. Sandler, eds, 'Handbook of Defence Economics', North Holland, Amsterdam, pp. 69–87.

Smith, R. P. (1998), 'Quantitative methods in peace research', *Journal of Peace Research* 35(4), 419–427. doi: 10.1177/0022343398035004001.

Smith, R. P. (2020), The Influence of the Richardson Arms Race Model, *in* N. P. Gleditsch, ed., 'Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences',

Vol. 27 of *Pioneers in Arts, Humanities, Science, Engineering, Practice*, Springer, Heidelberg, pp. 25–34. doi: 10.1007/978-3-030-31589-4_3.

Smith, R., Sola, M. and Spagnolo, F. (2000), 'The Prisoner's Dilemma and regime-switching in the Greek–Turkish arms race', *Journal of Peace Research* 37(6), 737–750. doi: 10.1177/0022343300037006005.

Solarin, S. A. (2017), 'Determinants of military expenditure and the role of globalisation in a cross-country analysis', *Defence and Peace Economics* 29(7), 1–18. doi: 10.1080/10242694.2017.1309259.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014), 'Dropout: A simple way to prevent neural networks from overfitting', *Journal of Machine Learning Research* 15(56), 1929–1958.

Steiner, B. H. (1973), Arms races, diplomacy, and recurring behaviour: Lessons from two cases, *in* 'Sage Professional Papers in International Studies', number 02-013, Sage, Beverly Hills, pp. 25–34.

Stephanopoulos, C. (1998), 'An Aegean peace', *Harvard International Review* 21(1), 18–23.

Stockholm International Peace Research Institute [SIPRI] (2020), *SIPRI Military Expenditure Database 2020*. Available at: https://www.sipri.org/databases/milex [Accessed 17 July 2020].

Stodden, V. and Miguez, S. (2014), 'Best practices for computational science: Software infrastructure and environments for reproducible and extensible research', *Journal of Open Research Software* 2(1), e21. doi: 10.5334/jors.ay.

Stoll, R. J. (1982), 'Let the researcher beware: The use of the Richardson equations to estimate the parameters of a dyadic arms acquisition process', *American Journal of Political Science* 26(1), 77–89.

Strumbelj, E. and Kononenko, I. (2014), 'Explaining prediction models and individual predictions with feature contributions', *Knowledge and Information Systems* 41(3), 647–665. doi: 10.1007/s10115-013-0679-x.

Sun, Q. and Yu, Q. (1999), 'Determinants of China's military expenditures: 1965–93', *Journal of Peace Research* 36(1), 23–33. doi: 10.1177/0022343399036001002.

Sung, A. H. (1998), 'Ranking importance of input parameters of neural networks', *Expert Systems with Applications* 15(3–4), 405–411. doi: 10.1016/S0957-4174(98)00041-4.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007), 'Measuring and testing dependence by correlation of distances', *The Annals of Statistics* 35(6), 2769–2794. doi: 10.1214/009053607000000505.

Tamura, S. and Tateishi, M. (1997), 'Capabilities of a four-layered feedforward neural network: four layers versus three', *IEEE Transactions on Neural Networks* 8(2), 251–255. doi: 10.1109/72.557662.

Turan, I. and Barlas, D. (1999), 'Turkish–Greek balance: A key to peace and cooperation in the Balkans', *East European Quarterly* 32(4), 469–488.

United Nations Development Programme/Human Development Report Office [UNDP/HDRO] (2019), *Human Development Report 2019 – Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*. Available at: http://hdr.undp.org/sites/default/files/hdr2019.pdf [Accessed 03 September 2020].

United Nations [UN] (2017), 'Concluding Main Part of Seventy-Second Session, General Assembly Adopts $5.397 Billion Budget for 2018-2019, as Recommended by Fifth Committee'. Available at: https://www.un.org/press/en/2017/ga11997.doc.htm [Accessed 4 September 2020].

van der Smagt, P. P. (1994), 'Minimisation methods for training feedforward neural networks', *Neural Networks* 7(1), 1–11. doi: 10.1016/0893-6080(94)90052-3.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. and SciPy 1.0 Contributors (2020), 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python', *Nature Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2.

Wagner, D. L., Perkins, R. T. and Taagepera, R. (1975), 'Complete solution to Richardson's arms race equations', *Journal of Peace Science* 1(2), 159–172.

Wallace, M. D. (1979), 'Arms races and escalation: Some new evidence', *The Journal of Conflict Resolution* 23(1), 3–16.

Wallace, M. D. and Wilson, J. M. (1978), 'Non-linear arms race models', *Journal of Peace Research* 15(2), 175–192. doi: 10.1177/002234337801500205.

Wang, W., Jones, P. and Partridge, D. (2000), 'Assessing the impact of input features in a feedforward neural network', *Neural Computing & Applications* 9(2), 101–112. doi: 10.1007/PL00009895.

Ward, M. D. (2020), Back to the Future: Richardson's Multilateral Arms Race Model, *in* N. P. Gleditsch, ed., 'Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences', Vol. 27 of *Pioneers in Arts, Humanities, Science, Engineering, Practice*, Springer, Heidelberg, pp. 57–71. doi: 10.1007/978-3-030-31589-4_6.

Wasserman, P. D. (1989), *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York.

Wasserstein, R. L. and Lazar, N. A. (2016), 'The ASA statement on p-values: Context, process, and purpose', *The American Statistician* 70(2), 129–133. doi: 10.1080/00031305.2016.1154108.

West, D. M. (2018), *The Future of Work: Robots, AI, and Automation*, Brookings Institution Press, Washington, DC.

White, H. (1992*a*), Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings, *in* H. White, ed., 'Artificial Neural Networks: Approximations and Learning Theory', Blackwell, Oxford, pp. 160–190.

White, H. (1992*b*), Consequences and Detection of Misspecified Nonlinear Regression Models, *in* H. White, ed., 'Artificial Neural Networks: Approximations and Learning Theory', Blackwell, Oxford, pp. 224–258.

White, H. (1992*c*), Learning and Statistics, *in* H. White, ed., 'Artificial Neural Networks: Approximations and Learning Theory', Blackwell, Oxford, pp. 79–80.

White, H. and Gallant, A. R. (1992), There Exists a Neural Network That Does Not Make Avoidable Mistakes, *in* H. White, ed., 'Artificial Neural Networks: Approximations and Learning Theory', Blackwell, Oxford, pp. 5–11.

White, H., Hornik, K. and Stinchcombe, M. (1992), Universal Approximation of an Unknown Mapping and its Derivatives, *in* H. White, ed., 'Artificial Neural Networks: Approximations and Learning Theory', Blackwell, Oxford, pp. 55–78.

White, H. and Stinchcombe, M. (1992), Approximating and Learning Unknown Mappings Using Multilayer Feedforward Networks with Bounded Weights, *in* H. White, ed., 'Artificial Neural Networks: Approximations and Learning Theory', Blackwell, Oxford, pp. 41–54.

Wilson, A. C., Roelofs, R., Stern, M., Srebro, N. and Recht, B. (2017), The marginal value of adaptive gradient methods in machine learning. arXiv preprint, arXiv:1705.08292.

Wohlstetter, A. (1974), 'Is there a strategic arms race? (II): Rivals, but no "Race"', *Foreign Policy* 16, 48–81. doi: 10.2307/1147844.

World Bank (2020*a*), *Armed forces personnel, total*. Available at: https://data.worldbank.org/indicator/MS.MIL.TOTL.P1?locations=GR-TR [Accessed 22 July 2020].

World Bank (2020*b*), *GDP per capita (constant 2010 US$)*. Available at: https://data.worldbank.org/indicator/NY.GDP.PCAP.KD?locations=GR-TR [Accessed 22 July 2020].

World Bank (2020*c*), *Gross capital formation (% of GDP)*. Available at: https://data.worldbank.org/indicator/NE.GDI.TOTL.ZS?locations=GR-TR [Accessed 22 July 2020].

World Bank (2020*d*), *Imports of goods and services (% of GDP)*. Available at: https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS?locations=GR-TR [Accessed 22 July 2020].

Wu, B., Iandola, F., Jin, P. H. and Keutzer, K. (2017), SqueezeDet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops', pp. 129–137.

Xiang, C., Ding, S. Q. and Lee, T. H. (2005), 'Geometrical interpretation and architecture selection of MLP', *IEEE Transactions on Neural Networks* 16(1), 84–96. doi: 10.1109/TNN.2004.836197.

Yang, Y. and Zhang, Q. (1997), 'A hierarchical analysis for rock engineering using artificial neural networks', *Rock Mechanics and Rock Engineering* 30(4), 207–222. doi: 10.1007/BF01045717.

Zapranis, A. and Refenes, A. P. (1999), *Principles of Neural Model Identification, Selection and Adequacy*, Springer, London.

Zeiler, M. D. (2012), ADADELTA: An adaptive learning rate method. arXiv preprint, arXiv:1212.5701.

Zhu, J., Liapis, A., Risi, S., Bidarra, R. and Youngblood, G. M. (2018), Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation, *in* '2018 IEEE Conference on Computational Intelligence and Games', pp. 1–8. doi: 10.1109/CIG.2018.8490433.

Zinnes, D. A. (1980), 'Three puzzles in search of a researcher: Presidential address', *International Studies Quarterly* 24(3), 315–342. doi: 10.2307/2600250.

Zinnes, D. A., Gillespie, J. V. and Rubison, R. M. (1976), A Reinterpretation of the Richardson Arms Race Model, *in* D. A. Zinnes and J. V. Gillespie, eds, 'Mathematical Models in International Relations', Praeger, New York, pp. 189–217.

Zissis, D., Xidias, E. K. and Lekkas, D. (2015), 'A cloud based architecture capable of perceiving and predicting multiple vessel behaviour', *Applied Soft Computing* 35, 652–661. doi: 10.1016/j.asoc.2015.07.002.

Zuk, G. and Thompson, W. R. (1982), 'The post-coup military spending question: A pooled cross-sectional time series analysis', *American Political Science Review* 76(1), 60–74. doi: 10.2307/1960442.

APPENDICES

APPENDIX A

Data Sources and Original Time Series

A.1 Data Sources

The output variable used in this thesis is the annual change in Greek defence expenditure. Military expenditure data in constant 2018 USD were retrieved from SIPRI Military Expenditure database (SIPRI, 2020). The original constant 2018 series was rescaled to 2010 and then the annual change was calculated. Details on the definition of expenditure to which SIPRI (2020) data adhere, as well as on the sources, methods and the calculation procedure employed can be found on https://www.sipri.org/databases/milex/sources-and-methods.

Table A.1 presents detailed information on the sources of each of the input variables used in models A (A1, A2), B (B1 to B5) and C (C1 to C6).

Table A.1: Input variables: Sources and calculation process

Alias	Variable	Calculation process
A1	Turkish military expenditure (constant 2010 USD)	SIPRI (2020) data; the original constant 2018 series was rescaled to 2010
A2	Greek military expenditure (constant 2010 USD)	SIPRI (2020) data; the original constant 2018 series was rescaled to 2010
B1	Ratio of Greek to Turkish armed forces personnel	The related Greek and Turkish series were obtained from Correlates of War Project (Singer et al., 1972) for years 1963–1989 and from World Bank (2020*a*) for years 1990 onwards
B2	Greek military expenditure per person in the armed forces (constant 2010 USD)	Calculations based on military expenditure and armed forces personnel data, as illustrated above for variables A1, A2 and B1
B3	Turkish military expenditure per person in the armed forces (constant 2010 USD)	Calculations as illustrated above for variable B2
B4	Greek military expenditure as a share of GDP	SIPRI (2020) data

(Continued on next page)

Alias	Variable	Calculation process
B5	Turkish military expenditure as a share of GDP	SIPRI (2020) data
C1	Turkish military expenditure as a share of GDP	SIPRI (2020) data
C2	Lagged change of Turkish military expenditure	Calculations as illustrated above for variable A1
C3	Lagged change of Greek military expenditure	Calculations as illustrated above for variable A2
C4	Aggregate change of Greek public debt as a share of GDP over the previous two years	Calculations based on European Commission (2020) data for years 1976–1978 and 2016–2018, and on International Monetary Fund [IMF] (2020) data for the remaining years
C5	Aggregate change of Greek gross capital formation as a share of GDP over the previous two years	Calculations based on World Bank (2020*c*) data
C6	Lagged change of Greek imports as a share of GDP	Calculations based on World Bank (2020*d*) data

A.2 Complete Time Series

Table A.2 presents the full time series of the output variable which was used in all models. Table A.3 and Table A.4 include the full time series of the input variables used in each of the three models.

Table A.2: Original time series of the output variable

YEAR	Annual Change in Milex
1963	0.025266
1964	0.039375
1965	0.081702
1966	0.085420
1967	0.287612

(Continued on next page)

YEAR	Annual Change in Milex
1968	0.167804
1969	0.131957
1970	0.078851
1971	0.057024
1972	0.066070
1973	0.005572
1974	0.241943
1975	0.286399
1976	0.094358
1977	0.060130
1978	0.021438
1979	-0.031282
1980	-0.135105
1981	0.183664
1982	0.020373
1983	-0.087800
1984	-0.018431
1985	-0.007508
1986	-0.145505
1987	-0.002198
1988	0.057408
1989	-0.062377
1990	0.011023
1991	-0.051574
1992	0.039221
1993	-0.023923
1994	0.017227
1995	0.021388
1996	0.059907
1997	0.065634
1998	0.089693

(Continued on next page)

YEAR	Annual Change in Milex
1999	0.046941
2000	0.055441
2001	-0.022020
2002	-0.019061
2003	-0.143283
2004	0.099446
2005	0.081346
2006	0.039723
2007	-0.000794
2008	0.111679
2009	0.048459
2010	-0.231541
2011	-0.194869
2012	-0.115542
2013	-0.066188
2014	-0.008161
2015	0.060482
2016	0.041677
2017	-0.000798
2018	0.068882

Table A.3: Original time series of the input variables used in Models A and B

YEAR	A1	A2	B1	B2	B3	B4	B5
1963	2.532212	1.743595	0.380952	10.897468	6.029075	0.040484	0.037111
1964	2.601814	1.787649	0.368421	10.215139	5.477502	0.038245	0.033873
1965	2.790453	1.858038	0.364583	10.617358	5.813444	0.035736	0.034225
1966	2.923608	2.009843	0.387755	10.578120	5.966547	0.034985	0.035544
1967	2.929702	2.181523	0.343137	12.465845	5.744513	0.035847	0.031440
1968	3.156463	2.808955	0.320755	16.523265	5.955591	0.043457	0.032596
1969	3.528948	3.280309	0.345794	17.731400	6.596165	0.046902	0.032756
1970	3.421243	3.713170	0.330275	20.628723	6.277511	0.047888	0.030793

(Continued on next page)

YEAR	A1	A2	B1	B2	B3	B4	B5
1971	3.794946	4.005957	0.333333	22.255316	7.027677	0.041973	0.032565
1972	4.348634	4.234392	0.295082	23.524403	7.128909	0.041192	0.034998
1973	4.570667	4.514158	0.295082	25.078656	7.492896	0.039565	0.034455
1974	4.846117	4.539311	0.326821	24.670168	8.607668	0.035177	0.033495
1975	5.433277	5.637565	0.324042	30.309490	9.465639	0.048245	0.031930
1976	9.501539	7.252158	0.316781	39.200855	16.269758	0.058473	0.051195
1977	9.982664	7.936457	0.275964	42.669122	14.811074	0.058292	0.049430
1978	9.611899	8.413677	0.242542	44.992926	12.466795	0.059185	0.047136
1979	8.801617	8.594048	0.257975	46.204557	12.207513	0.055754	0.041873
1980	7.809553	8.325207	0.267908	44.519823	11.188472	0.051812	0.033573
1981	8.094282	7.200426	0.259414	38.711968	11.289097	0.046608	0.039003
1982	9.132282	8.522883	0.253711	45.334485	12.324268	0.057357	0.038245
1983	9.983554	8.696524	0.241873	46.755506	12.982516	0.056266	0.042982
1984	9.446062	7.932972	0.214806	44.819049	11.463667	0.051725	0.039376
1985	9.182694	7.786760	0.241718	39.526702	11.267109	0.048354	0.035971
1986	9.742028	7.728293	0.246929	38.449221	11.968094	0.046926	0.035310
1987	10.945899	6.603791	0.234884	32.692035	12.727790	0.041280	0.036380
1988	10.453659	6.589277	0.226394	33.111942	11.892672	0.042555	0.033286
1989	9.207643	6.967553	0.234947	35.012831	10.870889	0.041982	0.029322
1990	10.653745	6.532941	0.257692	32.502192	13.658647	0.037659	0.031487
1991	12.873424	6.604951	0.261378	32.860454	16.740474	0.037983	0.035277
1992	13.233498	6.264307	0.254975	30.557594	16.459575	0.034849	0.037544
1993	13.919571	6.509997	0.295455	31.298062	19.772118	0.036298	0.038706
1994	15.389814	6.354258	0.310496	29.832198	22.434131	0.036001	0.039214
1995	15.046418	6.463724	0.254007	31.377300	18.552919	0.035820	0.040514
1996	15.457469	6.601970	0.292464	32.715410	22.402130	0.031732	0.039016
1997	17.299611	6.997471	0.281109	35.198547	24.462119	0.033119	0.041397
1998	18.029787	7.456746	0.202509	44.839120	21.955415	0.033659	0.041037
1999	18.892226	8.125566	0.210058	47.104728	23.005633	0.035339	0.031845
2000	20.859838	8.506990	0.201617	50.159140	24.797716	0.033605	0.038890
2001	20.187662	8.978629	0.197173	55.016111	24.390071	0.034654	0.036611

(Continued on next page)

YEAR	A1	A2	B1	B2	B3	B4	B5
2002	18.506454	8.780923	0.245377	53.804678	27.825070	0.032514	0.036035
2003	19.690833	8.613548	0.273165	47.431432	29.619183	0.030774	0.037959
2004	18.311711	7.379375	0.273165	40.635328	27.544691	0.024939	0.032960
2005	17.017100	8.113228	0.271104	48.582201	27.625162	0.026057	0.026980
2006	16.401409	8.773207	0.272285	52.221471	26.582510	0.028366	0.024094
2007	17.168637	9.121701	0.263072	56.656530	28.053328	0.027834	0.023595
2008	16.554395	9.114461	0.263072	56.611557	27.049665	0.026793	0.022179
2009	16.792700	10.132356	0.262643	62.933887	27.394290	0.029831	0.021993
2010	17.968348	10.623365	0.233251	74.322527	29.321717	0.032248	0.024893
2011	17.650464	8.163619	0.244202	54.552509	28.802977	0.027270	0.022867
2012	17.801872	6.572783	0.242086	44.305921	29.050052	0.024769	0.020427
2013	18.238764	5.813348	0.240454	39.452649	29.762997	0.024077	0.020246
2014	18.730914	5.428572	0.243065	36.445597	30.566114	0.023577	0.019385
2015	18.867583	5.384269	0.239801	36.640141	30.789137	0.023336	0.018815
2016	19.414806	5.709919	0.287012	38.856202	37.919543	0.024509	0.018224
2017	22.762561	5.947891	0.279199	41.608190	44.458128	0.025431	0.020645
2018	24.431482	5.943142	0.402449	41.574972	68.782324	0.025163	0.020653

Table A.4: Original time series of the input variables used in Model C

YEAR	C1	C2	C3	C4	C5	C6
1963	0.037111	0.065069	0.016879	3.457813	6.940367	0.637355
1964	0.033873	0.027487	0.025266	10.040714	4.875719	0.514081
1965	0.034225	0.072503	0.039375	6.400683	8.328525	1.445589
1966	0.035544	0.047718	0.081702	-6.060714	7.700823	0.300921
1967	0.031440	0.002084	0.085420	-2.459258	-0.327476	-1.398431
1968	0.032596	0.077401	0.287612	4.634895	-5.295779	-0.392454
1969	0.032756	0.118007	0.167804	4.049728	0.656968	0.421067
1970	0.030793	-0.030520	0.131957	4.633547	7.212620	-0.050403
1971	0.032565	0.109230	0.078851	1.666535	4.374353	-0.371839
1972	0.034998	0.145902	0.057024	-0.810699	1.258063	-0.064680
1973	0.034455	0.051058	0.066070	1.827463	5.289126	1.025416

(Continued on next page)

YEAR	C1	C2	C3	C4	C5	C6
1974	0.033495	0.060265	0.005572	-2.989318	10.265656	3.905102
1975	0.031930	0.121161	0.241943	-4.014732	-3.494081	1.168346
1976	0.051195	0.748768	0.286399	2.097054	-11.865816	-0.079866
1977	0.049430	0.050637	0.094358	-4.577992	-0.042562	-0.351321
1978	0.047136	-0.037141	0.060130	-5.993864	-2.673756	-0.148658
1979	0.041873	-0.084300	0.021438	5.525800	-5.428035	-1.197491
1980	0.033573	-0.112714	-0.031282	4.432938	-1.651345	0.874566
1981	0.039003	0.036459	-0.135105	-0.929277	-2.168465	4.903316
1982	0.038245	0.128239	0.183664	4.162577	-7.072480	0.489125
1983	0.042982	0.093216	0.020373	6.783665	-0.826886	-1.108442
1984	0.039376	-0.053838	-0.087800	6.909273	3.622805	0.056975
1985	0.035971	-0.027881	-0.018431	10.750713	0.018292	-0.601881
1986	0.035310	0.060912	-0.007508	13.030579	2.701082	0.156655
1987	0.036380	0.123575	-0.145505	7.081359	0.528334	0.703344
1988	0.033286	-0.044970	-0.002198	5.791455	-8.263206	-0.763243
1989	0.029322	-0.119194	0.057408	9.927156	-3.741431	-0.895625
1990	0.031487	0.157054	-0.062377	7.408664	3.777782	1.539447
1991	0.035277	0.208347	0.011023	16.085417	0.964591	0.511615
1992	0.037544	0.027970	-0.051574	14.861407	1.546389	-0.956570
1993	0.038706	0.051844	0.039221	6.813178	-2.146998	-0.449394
1994	0.039214	0.105624	-0.023923	25.605684	-4.223073	-0.965622
1995	0.040514	-0.022313	0.017227	18.327902	-2.620615	-1.260618
1996	0.039016	0.027319	0.021388	-1.298534	-1.710378	0.965085
1997	0.041397	0.119175	0.059907	3.039830	0.564715	0.526679
1998	0.041037	0.042208	0.065634	0.461728	-0.030712	-0.157313
1999	0.031845	0.047834	0.089693	-3.910824	1.821631	2.919733
2000	0.038890	0.104149	0.046941	-0.545183	1.708827	2.149823
2001	0.036611	-0.032223	0.055441	7.509999	0.650187	6.575777
2002	0.036035	-0.083279	-0.022020	8.174607	1.541627	-1.346072
2003	0.037959	0.063998	-0.019061	-0.071684	-1.075982	-3.115083
2004	0.032960	-0.070039	-0.143283	-5.625014	1.683550	-0.590671

(Continued on next page)

YEAR	C1	C2	C3	C4	C5	C6
2005	0.026980	-0.070699	0.099446	-1.992753	0.557465	-0.455144
2006	0.024094	-0.036181	0.081346	5.935936	-5.275607	0.398307
2007	0.023595	0.046778	0.039723	0.703641	0.844732	2.087596
2008	0.022179	-0.035777	-0.000794	-4.289242	5.031820	3.327875
2009	0.021993	0.014395	0.111679	5.841856	-1.640609	0.964042
2010	0.024893	0.070009	0.048459	23.642029	-8.793201	-7.206487
2011	0.022867	-0.017691	-0.231541	36.834172	-7.463051	1.965233
2012	0.020427	0.008578	-0.194869	45.351377	-3.233241	1.582311
2013	0.020246	0.024542	-0.115542	13.314803	-4.245087	0.824569
2014	0.019385	0.026984	-0.066188	5.580951	-3.503609	0.031941
2015	0.018815	0.007296	-0.008161	20.498312	-0.892133	1.610699
2016	0.018224	0.029003	0.060482	-0.738890	-1.384036	-3.274523
2017	0.020645	0.172433	0.041677	-1.574865	-0.436458	-0.725108
2018	0.020653	0.073319	-0.000798	-0.771041	2.292515	3.233377

APPENDIX B

Model Deployment Code

B.1 Code for Neural Network Implementation

This section hosts the relevant code used for neural network model design, training, and input significance assessment. Code used for data analysis and plotting is not included. Please note that `Network_Data.xlsx` is the datafile containing the data used for model implementation, while `GRC_OUT` denotes the output value. The code is written in Python programming language.

```
# ============== SETUP ==============
# Import essential libraries and define essential functions
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import shap
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import RobustScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

# Import data and save them in a pandas DataFrame
data = pd.read_excel('Network_Data.xlsx', index_col='Year')

# ===== SPLITTING AND SCALING DATA ====
# Separate inputs from the output
Y = data['GRC_OUT']
X = data.iloc[:,0:-1]

# Run train_test_split twice to get a 20% validation and a 20% test set
X_train, X_test, Y_train, Y_test = train_test_split(
    X, Y, test_size = 0.2, shuffle=False
)
X_train, X_val, Y_train, Y_val = train_test_split(
    X_train, Y_train, test_size=0.25, shuffle=False
)  # 0.25 x 0.8 = 0.20

# Scale inputs using RobustScaler
scaler = RobustScaler().fit(X_train)
X_train_scaled = scaler.transform(X_train)
X_val_scaled = scaler.transform(X_val)
X_test_scaled = scaler.transform(X_test)

# ============= TRAINING ==============
# Define model
input_param = len(X.columns)
```

```
model = Sequential()
model.add(Dense(8, input_shape=(input_param,), activation='sigmoid'))
model.add(Dense(1, activation='tanh'))

# Compile model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])

# Fit model
history = model.fit(
    X_train_scaled,
    Y_train,
    epochs=400,
    validation_data=(X_val_scaled, Y_val),
    shuffle=False,
)

# ==== INPUT SIGNIFICANCE ANALYSIS ====
# Initiate SHAP KernelExplainer
train_data = pd.DataFrame(data=X_train_scaled, columns=X_train.columns)
explainer = shap.KernelExplainer(model.predict, train_data)
shap_values = explainer.shap_values(train_data.values)

# Show aggregate and detailed summary plots of significance
shap.summary_plot(shap_values, train_data, plot_type='bar')
shap.summary_plot(shap_values[0], train_data)

# Show dependence plots for specific features
shap.dependence_plot('A1', shap_values[0], train_data, interaction_index=None)
```

B.2 Code for Distance Correlation and Significance Evaluation

Distance correlations and their respective p-values were calculated with the use of the following function (the function was adapted from https://gist.github.com/wladston/c931b1495184fbb99bec):

```
import numpy as np
import copy
from scipy.spatial.distance import squareform, pdist

def distcorr(X, Y):
    '''
    Computes distance correlation values
    '''
    X = np.atleast_1d(X)
    Y = np.atleast_1d(Y)
    if np.prod(X.shape) == len(X):
        X = X[:, None]
    if np.prod(Y.shape) == len(Y):
        Y = Y[:, None]
    X = np.atleast_2d(X)
```

```
    Y = np.atleast_2d(Y)
    n = X.shape[0]
    if Y.shape[0] != X.shape[0]:
        raise ValueError('Number of samples must match')
    a = squareform(pdist(X))
    b = squareform(pdist(Y))
    A = a - a.mean(axis=0)[None, :] - a.mean(axis=1)[:, None] + a.mean()
    B = b - b.mean(axis=0)[None, :] - b.mean(axis=1)[:, None] + b.mean()

    dcov2_xy = (A * B).sum()/float(n * n)
    dcov2_xx = (A * A).sum()/float(n * n)
    dcov2_yy = (B * B).sum()/float(n * n)
    dcor = np.sqrt(dcov2_xy)/np.sqrt(np.sqrt(dcov2_xx) * np.sqrt(dcov2_yy))
    return dcor

def distcorr_pval(Xval, Yval):
    '''
    Computes p-values for distance correlations
    '''
    dcor = distcorr(Xval, Yval)
    greater = 0
    for i in range(10000):
        Y_r = copy.copy(Yval)
        np.random.shuffle(Y_r)
        if distcorr(Xval, Y_r) > dcor:
            greater += 1
    return greater / float(10000)
```

Optimal α values for the statistical evaluation of Pearson's correlation coefficients were calculated using the following code snippet in R:

```
# Load the related library
library(OptSig)

# Compute the optimal alpha for r = 0.3 and n = 33 samples
OptSig.r(r=0.3,n=33,p=0.5,k=1,alternative="two.sided",Figure=TRUE)
```

APPENDIX C

Neural Network Code and Results

C.1 Weight Matrices of Estimated Models

The matrices are N by M, where N is equal to the number of nodes in the originating layer, including bias, and M is equal to the number of nodes in the following layer. The last row of each matrix, separated by a horizontal line, represents bias weights.

MODEL A

Input-to-hidden layer:

$$\left[\begin{array}{rrrrrrrr} -0.569759 & -0.462136 & -0.189277 & -0.131063 & -0.029737 & -0.878172 & 0.358170 & -0.428806 \\ 0.207046 & -0.154160 & -0.055650 & 0.370568 & 0.536060 & 0.538502 & 0.545108 & 0.413171 \\ \hline -0.128960 & 0.132582 & -0.117870 & 0.163147 & 0.119394 & 0.103370 & -0.101611 & -0.118736 \end{array}\right]$$

Hidden-to output layer:

$$\left[\begin{array}{r} 0.640360 \\ -0.296102 \\ 0.568180 \\ -0.130521 \\ -0.633560 \\ -0.133909 \\ 0.258975 \\ 0.234491 \\ \hline -0.119995 \end{array}\right]$$

MODEL B

Input-to-hidden layer:

$$\left[\begin{array}{rrrrrrrr} 0.397452 & 0.279373 & 0.435790 & 0.338786 & 0.466226 & 0.437695 & 0.201164 & -0.066611 \\ 0.337195 & 0.598885 & -0.509573 & 0.014524 & -0.517159 & 0.567260 & 0.389814 & -0.114518 \\ 0.428316 & -0.380692 & -0.055236 & 0.079742 & 0.433928 & -0.766567 & -0.046195 & 0.436510 \\ 0.377701 & 0.529106 & -0.216357 & -0.414180 & -0.397521 & -0.033929 & -0.704164 & -0.270049 \\ -0.726874 & -0.504857 & -0.254894 & -0.725022 & 0.460804 & -0.415824 & -0.450057 & 0.324175 \\ \hline -0.023106 & -0.077291 & 0.085838 & 0.082104 & -0.062818 & 0.106134 & -0.075340 & 0.084105 \end{array}\right]$$

Hidden-to output layer:

$$\left[\begin{array}{r} -0.038068 \\ 0.392638 \\ -0.077942 \\ -0.435447 \\ 0.688233 \\ -0.134616 \\ 0.619090 \\ -0.680338 \\ \hline -0.071759 \end{array}\right]$$

MODEL C

Input-to-hidden layer:

$$\left[\begin{array}{rrrrrrr} -0.316597 & 0.353010 & -0.594194 & -0.683211 & -0.392810 & 0.364488 & -0.148115 \\ -0.130117 & -0.579301 & -0.557802 & 0.086160 & -0.521744 & -0.180149 & -0.172058 \\ -0.201901 & 0.253417 & 0.044809 & -0.202409 & -0.410014 & -0.699423 & -0.621107 \\ -0.103640 & 0.498021 & -0.350541 & -0.307615 & -0.212132 & 0.209884 & -0.046929 \\ 0.040499 & -0.069798 & 0.096153 & -0.290260 & 0.424843 & 0.106820 & -0.252358 \\ 0.440410 & -0.036361 & 0.446742 & -0.022014 & 0.107451 & -0.007716 & -0.092463 \\ \hline 0.006418 & 0.005433 & -0.042525 & -0.029697 & -0.013291 & 0.005507 & 0.036757 \end{array}\right]$$

Hidden-to output layer:

$$\left[\begin{array}{r} -0.330032 \\ -0.046609 \\ 0.543609 \\ 0.417901 \\ -0.394217 \\ 0.648764 \\ -0.678234 \\ \hline -0.024609 \end{array}\right]$$

C.2 Detailed Neural Network Results

This section hosts the complete estimated time series for all models. Training set results are shown on Table C.1, validation set results are shown on Table C.2 and test set results are shown on Table C.3:

Table C.1: Detailed training set results for all models

Year	Real Value	Model A	Model B	Model C
1963	0.025266	0.128593	0.140067	0.054354
1964	0.039375	0.127655	0.141114	0.042196
1965	0.081702	0.125439	0.141638	0.073739
1966	0.085420	0.123223	0.161513	0.089432
1967	0.287612	0.122058	0.127296	0.061033
1968	0.167804	0.115941	0.078038	0.177199
1969	0.131957	0.109598	0.095796	0.120344
1970	0.078851	0.107825	0.076367	0.078068
1971	0.057024	0.102731	0.105528	0.067567
1972	0.066070	0.096556	0.065967	0.069366
1973	0.005572	0.092995	0.071810	0.082439
1974	0.241943	0.090512	0.116332	0.108828
1975	0.286399	0.079431	0.073775	0.236033
1976	0.094358	0.039166	0.058761	0.035267
1977	0.060130	0.032341	0.019596	0.078403
1978	0.021438	0.033166	-0.015126	0.067285
1979	-0.031282	0.038835	0.012717	0.008619
1980	-0.135105	0.047783	0.029465	0.040802
1981	0.183664	0.050302	0.024690	0.094265
1982	0.020373	0.036493	0.004054	0.093142
1983	-0.087800	0.029115	-0.010759	0.033102
1984	-0.018431	0.036492	-0.025199	-0.030315
1985	-0.007508	0.039174	0.000290	-0.003861
1986	-0.145505	0.035121	0.001854	0.014278
1987	-0.002198	0.031608	-0.011423	-0.023380
1988	0.057408	0.035295	-0.023747	0.022901
1989	-0.062377	0.042774	0.000237	0.027338
1990	0.011023	0.034113	0.013967	0.047611
1991	-0.051574	0.017881	0.010892	0.035983
1992	0.039221	0.017297	0.019939	-0.032129

(Continued on next page)

Year	Real Value	Model A	Model B	Model C
1993	-0.023923	0.011259	0.040773	0.032451
1994	0.017227	0.002502	0.033937	0.020987
1995	0.021388	0.004066	0.018610	0.009743

Table C.2: Detailed validation set results for all models

Year	Real Value	Model A	Model B	Model C
1996	0.059907	0.000569	0.046958	0.072132
1997	0.065634	-0.013621	0.037839	0.065277
1998	0.089693	-0.021241	-0.002599	0.057889
1999	0.046941	-0.031287	-0.047082	0.111645
2000	0.055441	-0.046502	-0.012308	0.103828
2001	0.022020	-0.045755	-0.030298	0.134322
2002	0.019061	-0.033239	0.021071	-0.015204
2003	0.143283	-0.039874	0.044676	-0.016685
2004	0.099446	-0.022474	0.038087	0.042065
2005	0.081346	-0.018969	0.025118	0.061109
2006	0.039723	-0.018787	0.023955	0.051673

Table C.3: Detailed test set results for all models

Year	Real Value	Model A	Model B	Model C
2007	-0.000794	-0.026231	0.010137	0.049703
2008	0.111679	-0.021858	0.010111	0.007535
2009	0.048459	-0.029329	0.010806	0.080132
2010	-0.231541	-0.040870	-0.021429	-0.055271
2011	-0.194869	-0.023488	-0.014941	-0.065910
2012	-0.115542	-0.013890	-0.016871	-0.112098
2013	-0.066188	-0.011326	-0.017687	0.015330
2014	-0.008161	-0.011449	-0.013333	-0.018439
2015	0.060482	-0.011888	-0.017414	0.029262
2016	0.041677	-0.017129	0.035108	-0.014053

(Continued on next page)

Year	Real Value	Model A	Model B	Model C
2017	-0.000798	-0.035894	0.015166	0.030062
2018	0.068882	-0.043379	0.012929	0.018667