Vrije Universiteit Amsterdam
School of Business and Economics

Department of Finance

Do Stocks Tell The Story?

Understanding the Contribution of Equity Market Characteristics to Systemic Risk

Master Thesis

submitted for the degree of

MSc. Finance: Duisenberg Honors Programme in Finance and Technology

Muppiri Nirmal Reddy
Student Number : 2566944

Submission Date: August 5, 2021
Supervisor: Iman van Lelyveld

Abstract

Past crises, such as the 2008 Global Financial Crisis, have shown how adverse the consequences of financial instability can be to the real economy. Existing measures of financial instability within the equity market, equate market characteristics to financial stress and are backward looking. Owing to their role in amplification of exogenous market shocks, we contend that equity market characteristics *shape* financial stress instead. We consider three market characteristics (volatility, liquidity and market regime) and four fundamental characteristics (leverage, profitability, valuation and coverage), to forecast systemic stress contribution of the equity market segment. We show that using an encoder-decoder model, one can effectively forecast equity market instability up to 9 weeks in advance, during both crisis and non-crisis periods. This finding is supported by stress testing and sensitivity analysis. Our research helps financial supervision limit the adverse socio-economic costs associated with equity market instability.

Contents

1 Introduction

Following the adverse economic effects and labour market effects of the eurozone crisis, an important priority of the European Central Bank (ECB) has been to ensure economic and financial stability within the European Union (EU). Financial stability is a characteristic of the financial system whereby the financial institutions and the financial markets, are resistant to economic shocks. Consequently, financial institutions and markets are able to fulfil their functions such as financial intermediation, liquidity provision and management of risks. As a result, disruption of such functions, for instance during the sovereign debt crisis, warrants research into how financial instability can be prevented in the future.

Systemic risk, defined by the Bank for International Settlements (BIS) as “a risk of disruption to financial services that is caused by an impairment of all or parts of the financial system and has the potential to have serious negative consequences for the real economy”, is a major driver of financial instability (Caruana, 2010). Put differently, it is the likelihood that financial instability materially impacts the functioning of a financial system. Systemic risk is especially relevant for financial stability in the European context for the following reasons: First, European financial structure is traditionally bank-based (Langfield and Pagano, 2016). This implies that from an economic perspective, European banks play a major role in circulating funds from the lenders to the borrowers. Second, the European financial sector is heavily interconnected between financial institutions, businesses, and households (Peltonen et al., 2019). While the benefit of such interconnectedness is an efficient functioning of the financial system, the major drawback is that a local negative shock in one part of the system is rapidly transmitted and amplified into a systemic event. Given that banks hold a large amount of sovereign debt (Dell’Ariccia et al., 2018), along with the fact that there are significant cross-border exposures and contagion effects (due to the interconnectedness of the financial sector), it is intuitive to realize the systemic consequences of a sovereign default in the form of widespread financial instability that was ultimately observed in the eurozone crisis. Unsurprisingly, measurement of systemic risk with the EU has received considerable attention over the last decade.

Financial markets are an important source for the amplification of systemic risk, especially during periods of crisis. Bavoso (2018) supports this view, particularly in the bond markets, by demonstrating the increased leverage and risk-taking granted by the use of complex, innovative debt transactions. He argues that given the recent surge market-based channels of finance, coupled with such transactions, the existing EU Capital Markets Union (CMU) framework fails to sufficiently validate the dangers associated with market-based finance. The CMU places a strong emphasis on resuscitating the securitization market, without considering the effects of excessive debt creation, which can have catastrophic consequences both at the firm and systemic level.

The relevance of financial markets for systemic risk can also be motivated through a contagion approach. If traders operating in a domestic market have reduced wealth due to a local negative shock, they are motivated to rebalance their portfolios and any physical exposures they have in external markets. They would like to sell their assets, thereby triggering a crash in such markets, despite the domestic and the external markets not necessarily being linked in terms of their fundamentals. Financial markets, therefore play an important role in amplifying a negative shock through wealth effects (Kyle and Xiong, 2001). In addition, European households invest more than 60% of their wealth in pension/insurance/investment funds (European Fund and Asset Management Association, 2020). Given that these funds have high exposure to financial markets, and that households form a major economic agent, methods to address financial market instability have gathered considerable attention.

In view of financial markets' contribution to systemic risk, substantial research has focused on finding contributors of systemic risk within the financial markets. The underlying idea being that monitoring such contributors will (i) provide insight into systemic risk build-up within the financial system and (ii) prevent adverse outcomes for the real economy. To give examples, volatility, which is essentially a measure of fluctuations in asset prices, was found to play a role in systemic risk build-up during the (a) stock market crash of 1987, (b) Asian crisis of 1996-97 and (c) global financial crisis of 2007-08 (Mieg, 2020). Market liquidity, the impact of trading an asset on its price, was found to be inversely correlated with systemic risk, in that, systemic risk increases as the market becomes illiquid (Jarrow and Lamichhane, 2021). If one considers financial instability from a Minsky perspective, one could argue that asset prices (or rather a change in asset prices) contribute to an endogenous build-up of systemic risk (Minsky, 1982). While these contributors play a role in the transmission and amplification of systemic risk, it is not clear whether these contributors drive systemic risk or are a consequence of high systemic risk.

For the scope of this paper, we assume that the contributors amplify systemic risk. The reasoning is as follows: mounting imbalances, lax supervision and increased risk appetite build up systemic risk over time and make the financial system vulnerable to exogenous shocks (fiscal shocks, GDP shocks, failure of systemically important financial institutions, bubble bursts). Once the shock has materialised, increased volatility, decreased asset prices, further disrupt the real economy through wealth effects and decreased credit supply, amplifying systemic risk. In addition, we narrow the scope from financial markets to only the equity market segment. The reasoning is because (a) unlike debt markets or derivative markets, the equity market has received little academic attention, (b) existing measures of equity market instability capture realized financial stress, are therefore backward looking and (c) the amplifying role of market characteristics remains unexplored in the equity market[1]. In this thesis, we use characteristics like liquidity, volatility etc. to forecast systemic risk contribution of the equity market. We capture further amplification of an existing shock solely due to characteristics of the equity market and doing so, we establish a stronger link between characteristics of the financial markets and systemic risk. We consider the European Stock Market, as an attempt to convince ECB's macro prudential supervision that our results are applicable at the EU level. We set the research question as follows:

"How can we relate the characteristics of equity market to systemic risk in the context of the European Union?"

To address this question, we capture fundamental and market characteristics of STOXX 600 Index[2] to forecast the financial stress within equity markets. It is this financial stress that ultimately forms the systemic risk contribution of the equity market. The benchmark chosen for financial stress, is the equity subcomponent of the composite index of systemic stress (CISS-E). We then adopt a Neural Granger Causality approach – i.e., assuming the characteristics cause[3] the effect that is increased systemic stress, we use a non-linear neural network framework to forecast equity contribution to systemic stress purely based on the characteristics. In order to embed conditions[4] needed for Granger Causality into our neural network, we use an encoder-decoder architecture. Consequently, all equity characteristics are temporally encoded into a fixed-dimensional context vector *before* being temporally decoded into our benchmark, CISS-E.

The value-addition of this thesis is to offer a forward-looking estimation of financial stress within the equity markets, conditional on the behaviour of equity market characteristics. The CISS-E, our benchmark, uses aggregation of three components to estimate financial stress: (i) realized volatility, (ii) maximum cumulated loss over a two-year window, and (iii) stock-bond correlation. These components are predominantly backward-looking and don't consider how financial stress is likely to evolve in the equity markets. The CISS-E assumes that characteristics of the equity market (such as realized volatility) *equate* to financial stress, while they should *shape* financial

[1]Economically, the reasoning is justified. As mentioned earlier in the paragraph, once a shock has materialised, increased volatility, decreased liquidity etc. further disrupt the real economy through wealth effects and decreased credit supply, amplifying systemic risk. However, equity market stress indicators fail to consider this "amplification effect" as they are not forward looking.

[2]This index was chosen because it covers 90% of the free-float market capitalization of the European stock market.

[3]The words "cause" or "causality" in the context of this paper refer purely to statistical concept of Granger Causality, whereby market and fundamental characteristics alone are enough to forecast systemic stress contribution of equity markets.

[4]The conditions necessary to establish Granger Causality: (a) cause precedes effect, (b) cause has unique information about future values of the effect and (c) cause and effect are temporally related, are assumed to be met.

stress—this forms the crux of this thesis. In a practical setting, our forward-looking methodology helps financial supervision limit the adverse socio-economic costs associated with equity market instability.

This thesis is structured as follows: First, we will extensively gather concepts relevant to the equity market with a link to systemic risk and justify the chosen benchmark for equity segment contribution to systemic risk. Next, we construct the data that enables these theoretical concepts to be embedded within our model. We then explain the intuition behind model choice and discuss decisions made during its implementation. Finally, we illustrate our results, where we identify the optimal model to answer the research question. The writing style chosen for this thesis is critical and reflective rather than descriptive, as choices and decisions play a pivotal role in the outcomes of machine learning methods.

2 Economic Foundations

The goal of this thesis is to forecast the systemic risk contribution of equity market, using only the market and fundamental characteristics of the equity market. Doing so, we attempt to relate equity characteristics to systemic risk. In order to have an effective forecast, we need to ensure that the characteristics we relate to systemic risk, have a conceptual link to systemic risk[5]. In addition, we also ensure that our choice of benchmark, CISS-E, is justified to reflect the systemic risk contribution of equity market. The objectives of this chapter are therefore: (i) to justify conceptual link between equity characteristics and systemic risk, and (ii) to justify the systemic risk benchmark. We will conclude this chapter with a list of features that capture equity market instability and a financial stability indicator (FSI) that is applicable specifically to the equity market.

Before we continue, an overarching assumption of this thesis is that equity characteristics alone are sufficient to forecast the systemic risk contribution of the equity segment. Given that systemic risk is a culmination of several market segments dynamically amplifying the risk within the economy, the critical reader can argue that characteristics of other market segments also ought to be considered. We contend that this is not necessary as any external[6] amplification of existing equity market instability, is preceded by changes in equity characteristics. Consequently, capturing non-linear dynamics of these changes through a neural network, ought to completely reflect equity market instability.

2.1 Justifying Equity Characteristics: Market Characteristics Relating to Systemic Risk

In this section, we present market characteristics chosen for our analysis and their economic link with systemic risk. The list is by no means exhaustive but captures the multifaceted nature of risk within the financial markets. We do not define "risk within the financial markets" because it's a broad concept with multiple components (applying to several asset classes: bonds, derivatives etc.) and will only be focusing on a few that are relevant to the equity sector, namely: (i) volatility, (ii) liquidity, and (iii) market regime.

[5]Else, we would violate the second condition necessary for Granger Causality (mentioned in Footnote 4).
[6]In this context, "external" refers to other market segments that could influence the equity market.

2.1.1 Volatility

An important sub-component of financial risk is the risk that arises from price movements. It is therefore imperative to consider price volatility as an important market characteristic for this analysis.

Before we explore the link between volatility and systemic risk, it is important to define volatility. The definition plays an important role in the extent to which it contributes to systemic risk. Principally, there are two main schools of thought:

- volatility can be defined as the variation in prices, whereby strong variations are thought to be a form of risk.
- volatility can be defined as the normal, random variation within the random walk of a measure.

The intuition behind the former definition stems from Markowitz's portfolio selection theory (Markowitz, 1952; Markowitz, 1959), while that of the latter stems from the complex world we live in (Taleb and Goldstein, 2012).

In economic literature, volatility is extended to variation in economic figures such as GDP (Campi and Dueñas, 2017; Dapena, 2011) and to parameters that are less quantifiable than stock prices such as humanitarian aid (Hudson, 2015). On the other hand, in mathematical finance, volatility can also be seen as the standard deviation of logarithmic returns for a series of asset prices (Hull, 2015). So, the definition of volatility can also differ based on the field of study. There is also a distinction between types of volatility based on the fundamental characteristics of volatility (irrespective of which previously mentioned definition of volatility we consider). Here, we consider historic volatility, implied volatility and idiosyncratic volatility. Historical volatility refers to the actual price movement observed in a given security (securities) over a past period of time. Implied volatility on the other hand, accounts for future expectations of volatility, and can be defined as the fair value of volatility based on market's expectations of movements over a given period. The importance of idiosyncratic volatility (the volatility of a single security) has also been noted in literature (Ang et al., 2009).

Link with Systemic Risk

Having established these definitions of volatility, we can relate volatility to systemic risk as follows:

- When we think of economies and volatility of capital flows, the systemic consequences of volatility are more severe for developing economies than for developed economies since social costs are higher and the distributional consequences of crises are larger in developing economies (Bekaert et al., 2014; Calvo and Reinhart, 2000).

- One could also argue that low volatility can be predictive of an economic crisis, as shown in a long-term study by Danielsson et al. (2018). The idea here is that conditions of low volatility encourage greater risk-taking behavior in economic agents.

- Finally, the role of volatility can also be seen in the amplification of systemic risk. The idea here is that an economic shock, causes increased price volatility in the equity market, resulting in credit crunches, liquidity shortage, and further financial instability. These are referred to as feedback effects in the real economy.

2.1.2 Liquidity

For the second characteristic, we consider market liquidity - the ease at which securities can be bought or sold at the current market price. Much like volatility, there are several definitions for liquidity but for this research we aggregate the following four characteristics that are representative of a liquid market as discussed by Danyliv et al. (2014). They are:

- **Tightness** whereby tight competition between buyers and sellers results in low spreads for a given security in a given market. Therefore, the tighter the market, the more liquid the market.

- **Immediacy of execution** which refers to the speed at which a large buy or sell order can be executed in a given market. Therefore, the longer it takes to execute a long order, the longer it takes for buyers or sellers to supply liquidity, thereby implying lower market liquidity.

- **Depth** which captures the total number of bids or asks for a given security at varying price levels. The greater the market depth, the lower the price impact of a large order in the given market and the higher the market liquidity.

- **Resiliency** which captures how quickly prices tend to revert back to their fundamental values in a given market. The greater the market resiliency, the more liquid the market.

Link with Systemic Risk

Much like volatility, market illiquidity contributes to the amplification of systemic risk. Dealers who act as counterparties and ensure a liquid market operate with limited capital. If there is a market shock, that imposes a constraint on their funding liquidity, they are forced to reduce their positions, and by extension, provide lower liquidity to the market. This results in a liquidity ‘doom loop’ whereby a reduction in market liquidity pushes prices down, which then forces counterparties to reduce their positions further, which then results in even lower market liquidity. The consequence of this doom loop is the fire sales of securities, which results in increased systemic risk due to price-mediated contagion.

2.1.3 Market Regime

Studies have shown that there exists an interactive effect such that abnormal volatilities and risk contagion in financial markets are exacerbated depending on whether there is a bull or a bear market (Chen et al., 2021; Li and Zakamulin, 2020). Classical econometric approaches rely on pairwise measurements (for instance, Pearson's correlation) to describe relationships within a given network system (Etesami et al., 2017). However, neural networks capture multidimensional non-linear relationships in a given network. Therefore, the inclusion of market regime might be helpful for our analysis.

Important to define, is that when we refer to market regime, we don't mean at solely looking at whether there is a bull or a bear market. It could also be looking at which direction the market is moving in, how strong these trends are and whether there are signs of market reversal.

A potential point of criticism here is that based on this definition, and given that we use neural networks, it could be that we have to ensure stationarity in our data to effectively train our model. In this case, we would have to get rid of trend and we would not capture this interactive effect[7] discussed earlier. In order to check for this issue, we have included an elaborate section in the results, whereby we check whether stationarity improves systemic stress estimation. This was found not to be the case, so the inclusion of this market characteristic makes sense.

Link with Systemic Risk

Market regime is important for systemic risk, because it determines the extent to which systemic risk is materialized, transmitted and amplified. For instance, if there is a strong bull market (or a consistent uptrend), there is room for relatively more market correction, than if there were a bear market. During a bear market, there's more incentive for the policy maker to not allow systemic risk to be materialized into the economy as that would mean a collapse of the financial system.

The intuition behind the characteristics in this subsection is that they exist both in periods of low and high systemic stress. However, their behavior changes depending on the level of stress within the economy, and this is precisely what we want our machine learning model to learn. We will now proceed to fundamental characteristics relating to systemic risk.

[7]Interactive effect of abnormal volatilities and risk contagion in financial markets being exacerbated depending on whether there is a bull or a bear market.

2.2 Justifying Equity Characteristics: Fundamental Characteristics Relating to Systemic Risk

In this section, we present fundamental characteristics of firms chosen for this analysis. While these characteristics have a relatively weak basis in theory than the market characteristics, we sought to include them in our equity characteristics as they reflect the systematic risk of the equity market (Astuty, 2017; Venanzi, 2020). Put differently, in this analysis, fundamental characteristics reflect the risk inherent to the entire equity market segment. We consider four classes of firm performance ratios as fundamental characteristics, namely:

- **Profitability Ratios,** which indicate how efficiently a company generates profit and value for its shareholders.
- **Valuation Ratios,** which indicate the relationship between the market value of a firm and the fundamental values of a firm.
- **Coverage Ratios,** which indicate a firm's ability to meet its financial obligations. The higher the coverage ratios of a firm, the more resilient a firm is to financial distress.
- **Leverage Ratios,** which indicate the level of debt incurred by a firm against accounts in its balance sheet, income statement or cash flow statement.

Systematic risk is relevant to answering our research question, as systemic risk measures are correlated[8] with systematic risk. In particular, Marginal Expected Shortfall (MES) and Conditional Value at Risk (ΔCoVaR), two popular measures of estimating systemic risk, are known to exhibit a strong interconnection with systematic risk (Benoit et al., 2017; Kubitza and Gründl, 2016). There was reported to be a 96% correlation between MES and systematic risk due to the beta factor, and 51% correlation between ΔCoVaR and systematic risk. While these measures are most appropriate for financial institutions, the principal idea that these measures capture the contribution of an institution (or a firm, in our scope) to systemic risk, can be applied to both financial and non-financial firms. So, to sum up: financial performance ratios are linked to systematic risk of a firm, and this systematic risk is correlated with systemic risk measures (MES/ ΔCoVaR). This is the interplay between financial ratios, systematic risk and systemic risk.

[8]To be clear, this correlation is purely intended to illustrate a link between systematic risk and systemic risk in academic literature. This correlation is a conceptual reason as to why fundamental characteristics might be important, but it's not definitive for the forecasting exercise that forms the scope of this thesis (to forecast systemic risk contribution using only equity characteristics).

2.3 Justifying Systemic Risk Benchmark

The objective of this subsection is to ensure that the choice of benchmark FSI is justified to reflect the systemic risk contribution of the equity market. This FSI will serve to evaluate the performance of our model in subsequent sections of this thesis.

Academic literature offers limited FSIs which truly reflect the systemic risk contribution of equity market. When we consider indicators of financial stability specific to the equity markets, the literature is predominantly composed of measures such as volatility, change in equity indices (Gadanecz and Jayaram, 2008) and liquidity (International Monetary Fund, 2006). These measures are equated to systemic risk, while for the purpose of this thesis, we want to study if these measures shape financial stress. This is precisely why they form our equity characteristics. Alternative measures such as non-performing loans are less applicable to non-financial firms for systemic risk contribution, while measures such as equity CDS-spreads[9] to assess risk appetite are more applicable to derivative markets than equity markets. Finally, studies with a similar methodology to this thesis, i.e. those based on Granger-causality, look at propagation of systemic risk (Balboa et al., 2015; Corsi et al., 2018) and focus primarily on the banking sector, so their benchmarks fall outside the scope of this study.

A financial stability benchmark that is particular to the equity markets and does not equate market characteristics to systemic risk, addresses the problems associated with former measures of financial instability. This is precisely what the CISS-E offers (Holló et al., 2012). It considers the following financial stress indicators and computes their arithmetic average to estimate the value of the CISS-E:

- **Realized Volatility** of idiosyncratic equity return, calculated as the weekly average of absolute daily log returns and transformed by its recursive sample cumulative distribution function (CDF).
- **Maximum Cumulated Loss (CMAX)** for stock market index calculated over a two-year moving window and transformed by its recursive sample CDF.
- **Stock-bond correlation** calculated as the weekly average of difference between the 4-year and 4-week correlation coefficients between daily log returns of stock market price index and 10-year German government bond price index.

The consideration of CMAX, a measure originally developed to identify crisis period in international stock markets (Coudert et al., 2006; Sarkar and Patel, 1998), and the stock-bond correlation,

[9]CDS: Credit Default Swaps

to capture flight-to-liquidity and flight-to quality phenomena[10] (Baele et al., 2009), in addition to realized volatility (a market characteristic) is what makes the CISS-E most applicable for the scope of this thesis. Instead of equating market characteristics to systemic risk contribution of the equity market, it considers a combination of investor behavior, market characteristics and realized loss in the equity market to provide an estimate of equity market instability.

In addition, the CISS-E contends that measurement of systemic *stress* (the materialized systemic risk—i.e., the materialized consequences for real economy) is more appropriate than that of systemic *risk*. Going back to the objective of the ECB, to ensure economic and financial stability within the EU, systemic stress estimation is far more valuable than systemic risk estimation. An indication of systemic stress would serve as a measure of *realized* financial (in)stability in the EU, to the ECB. This fits well with the scope of this thesis since we assume that financial instability is materialized into equity market characteristics before it is further amplified through wealth and feedback effects.

The benefit of the CISS-E is that it is one of the few stability indicators that embeds features other than equity characteristics to estimate equity market instability. However, it has certain limitations, which are imperative to address, since other indicators fail to preserve the economic foundations of equity market instability. To that effect, the CISS-E also serves as an effective benchmark since it warrants research that builds on its existing conceptual foundations. Below, we discuss the criticism surrounding the CISS-E and how this paper addresses this criticism. This reflects the value-addition of this thesis.

- A major criticism of the CISS-E is that it is backward-looking and therefore not predictive of forthcoming equity market instability. This can be attributed to the CISS-E being composed of three indicators, all of which are backward looking. Our research addresses this criticism, as we use market characteristics to give a forward-looking indication at how CISS-E is likely to be, using a neural network with an encoder-decoder architecture[11].
- The three indicators that form the CISS-E are all assumed to be equally important in estimating the materialized level of equity market instability (owing to the arithmetic average). This might be inaccurate. Put differently, the fact that the relative importance of each indicator might be different under different circumstances, is unaccounted for. To build on this limitation, we opt specifically for a neural network structure, as it captures the non-linear

[10]These phenomena underpin the idea that during periods of high systemic stress, investors shift funds out of risky stock markets into safer bond markets.

[11]Our neural network is trained such that all equity characteristics are temporally encoded into a fixed-dimensional context vector before being temporally decoded into our benchmark, CISS-E. Consequently, information observed in equity characteristics today, can be used to estimate CISS-E values tomorrow.

dynamics of equity characteristics. In a practical context (i.e., during model serving), incremental (re)training of our neural network with continuous flow of new data, enables us to visualize how feature importance changes over time[12].

- The publication frequency of the CISS-E is low (updated weekly) and with markets operating on higher frequency, there is a need for a higher publication frequency of equity market instability. To address this issue, we operate on a daily frequency for this analysis.

- Finally, the CISS-E methodology identifies critical value thresholds based on two variants of parsimonious econometric regime switching models. A criticism with this approach is that critical values of CISS-E might be less informative than the forecasted progression of systemic stress. Increasing systemic stress over time suggests an increasing vulnerability to exogenous shocks and this is more relevant for financial supervision, especially during crisis periods. We tackle this limitation by implementing sequence-to-sequence learning in our encoder-decoder architecture[13]. Consequently, when we encode our equity characteristics, we encode several weeks of characteristics at a daily frequency. Similarly, when we decode our CISS-E benchmark, we decode weeks of CISS-E values at a daily frequency. Therefore, based on past few weeks of daily equity characteristics data, we are able to say something about the coming few weeks of the CISS-E behavior.

The following paragraph places CISS-E in the context of the Composite Index of Systemic Stress (CISS). The goal is to give the reader an idea of how the CISS (the parent index of the CISS-E) was constructed. We will then outline the focus of studies building on the CISS and why CISS-E remains the best choice as the benchmark for the scope of this thesis.

The CISS is an econometric approach to capturing systemic stress introduced by Holló et al. (2012) in which captures financial stress both at the market level and the systemic level. They propose the structure illustrated in Figure 1 for constructing the CISS. Their methodology is the following: They select a total of 15 stress indicators from different lower-level market segments and transform them into empirical cumulative distribution functions to produce market stress subindices. Then, they compute a dynamic correlation matrix between these subindices using a parametric exponentially weighted moving average model (EWMA). Finally, they construct the CISS by weighting the subindices with cross correlation between 5 financial markets, according to

[12] We perform sensitivity analysis in the Results chapter to illustrate feature importance across all samples. However, before we present the important features, we also outline the methodology used for sensitivity analysis. In this methodology, we illustrate how feature importance can be evaluated on a truncation-by-truncation basis.

[13] To recap, we have a neural network (relevant because it establishes relationship between data, much like a linear regression model), with encoder-decoder architecture (relevant because it enables market characteristics to fully predict equity market instability) and sequence-to-sequence learning (relevant because we want a sequence of market characteristics to predict the progression of equity market instability) embedded into our network.

modern portfolio theory. The idea behind this construction, is that the cross correlation captures the contagion between different markets and the subindices capture the stress within the financial markets.

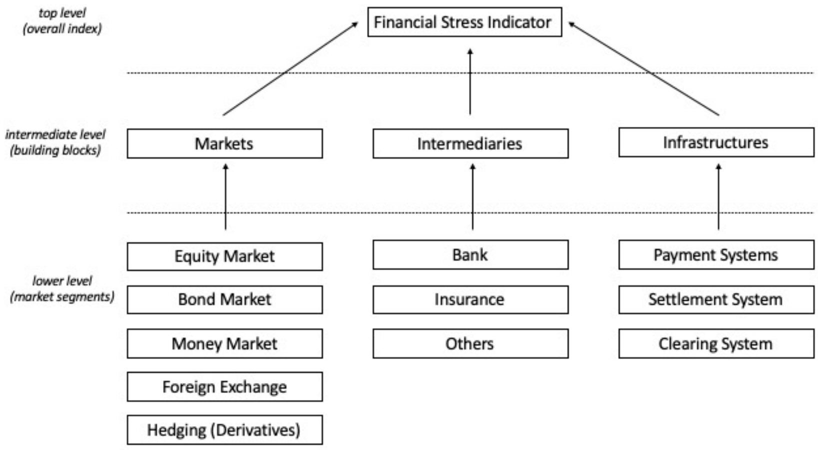

Figure 1: Structure for constructing the Composite Index of Systemic Stress (CISS)

For the scope of this thesis, as far as one market segment is concerned, the equity component of the CISS (CISS-E) seems to be most effective as a benchmark. Studies that have built on the CISS focus more on the improving the dynamic correlation between sectors than the contribution of a specific sector to systemic risk. For instance, Louzis et al. (2012) and Cabrera-Rodríguez et al. (2014) construct a stress index similar to the CISS, for Greece and Columbia respectively. In their methodology, instead of using the EWMA model to estimate the dynamic correlation, they use a multivariate GARCH model. They use principal component analysis (PCA), unlike the CISS, which uses ordered statistics for the first level aggregation (stress indicators to subindices). PCA is sensitive to outliers, implying that the market level aggregations are less robust compared to those of CISS (Holló et al., 2012). CISS therefore has a superior approach in estimating the contribution of a specific sector to systemic risk.

Table 1: Choices of Market and Fundamental Equity Segment Characteristics (Left); Benchmark Financial Stability Indicator (Right)

Equity Characteristics		**Financial Stability Benchmark**
Market	*Fundamental*	
Volatility Liquidity Market Regime	Profitability Leverage Coverage Valuation	CISS-E

Our objectives this section were: (i) to justify conceptual link between equity characteristics and systemic risk, and (ii) to justify the systemic risk benchmark. Having done so, we conclude on the list of features that capture equity market instability, and financial stability indicator (FSI) that is applicable specifically to the equity market segment. This is displayed in Table 1. In the following chapter, we investigate data sources that allow the findings of this chapter to be embedded into our neural network framework.

3 Data

The focus thus far has been on setting the theoretical groundwork. We started with the relevance of this research, justified the economic reasoning behind our choice for equity characteristics, explored their relation to systemic risk and established a benchmark to capture equity market instability. In this chapter, we outline data considerations as a prelude to the implementation of our neural network framework, to ultimately establish a link between equity characteristics and systemic risk in the context of EU. We will discuss data and it's engineering along with the motivation behind our assumptions, choices and decisions.

To ensure that the neural network is trained on fundamental economic concepts underlying systemic risk, maintaining flexible data requirements was a decision made for this thesis. Consequently, instead of selecting specific indicators that relate to systemic risk, we consider indicators that fall into classes of equity characteristics that amplify equity market instability (discussed in the previous chapter). Consequently, we provide such indicators to the neural network. Doing so, we estimate the contribution of equity subsector to systemic risk, while preserving its economic foundations. During training, the neural network then assigns weights to these indicators depending on how the systemic stress evolved in the past – and depending on the model evaluation technique (for example: walk forward cross validation), it can assign these weights differently for different training and testing subsamples. We thereby maintain a flexible view of which indicators (under a market characteristic) contribute to systemic stress at a given instance, but a firm view of market characteristics that are responsible for equity market instability[14].

We consider daily data from the Eurostoxx 600 Index[15] from January 01, 2016 to January 01, 2021. All of our market-specific indicators will be constructed from the Open, High, Low, Close and Volume (OHLCV) quotes aggregated on a daily time frame. The method of aggregation from stocks to index, for market-specific indicators is primarily based on free-float weighting i.e., components are weighted according to the proportion of shares held by the public. As for the fundamental indicators[16], they are retrieved from FactSet. Their method of aggregation from stocks to index

[14]As established in the previous chapter, most of the existing benchmarks of equity market instability equate market characteristics to instability. To some extent, the CISS-E does the same with the inclusion of realized volatility. Consequently, there is a case to be made whether the results of our analysis are accurate. Put differently, there might be a specification error in our approach because some of the independent and dependent variables are the same. By maintaining a firm view of the market characteristics, we urge development of benchmarks of systemic stress in the equity market segment which exclude market characteristics. We will discuss this in further depth in the Results and Discussion chapters.

[15]This Index was chosen because it includes large cap and small cap stocks from all equity subcomponents in the Europe Total Market. More information on the exact construction of the index, refer to "STOXX Index Methodology Guide (Portfolio Based Indices)" (2021).

[16]i.e., indicators which fall under firm-performance classes that ultimately form the fundamental characteristics (discussed in the previous section)

is based on adding fundamental values of all companies included in the index at the end of each quarter—and then computing the fundamental ratios. The ratios are therefore aggregated on a quarterly basis, with values being reported on March, June, September and December during a given year.

Table 2: Snapshot of complete table with all indicators found in Appendix A.1

Characteristic	Indicator Name	Abbreviation	Source	Stress Behaviour
Volatility	Relative Strength Index Stochastic Oscillator Price Rate of Change	RSI SO ROC	Murphy (1999)	+
⋮	⋮	⋮	⋮	⋮

We list all the indicators, equity characteristic classes each indicator falls under, and their abbreviations in Table 6 in Appendix A.1. We present a snapshot of it in Table 2. We present constructor sources for each indicator, which offer instructions on how to recreate the indicators to replicate our study[17]. Stress behaviour corresponds with how each indicator responds to financial stress[18]. A positive sign (+) represents increasing values of the indicator with increasing financial stress and the inverse is true for the negative sign (-). Its inclusion in the data chapter is justified as it is relevant for stress testing later on. Instructions for recreating the CISS-E can be found in the paper by Holló et al. (2012). For this research, daily CISS-E values were retrieved from ECB's Statistical Data Warehouse from January 01, 2016 to January 01, 2021.

An important consideration relevant to the scope of this thesis, is to investigate how fundamental and the market indicators are converted to the same time scale. Ideally, more granular sources of data (i.e., daily market data) ought to be aggregated to less granular periods (i.e., quarterly periods), to not make assumptions about the less granular information (i.e., not make assumptions about the daily behaviour of the quarterly fundamental data). However, this would mean that the results would be limited to 20 data points (5 years x 4 quarters) per indicator, implying too few data points to train our neural network.

Generally speaking, the quantity of data necessary for a research question is proportional to the complexity of the problem at hand. Therefore, if the objective were to compare the contribution of all market segments to systemic risk as a whole, a large dataset would be necessary (in the order of 10^5 datapoints). However, given that our scope concerns the contribution of equity market to systemic risk, we settle for fewer datapoints, simply because our problem is less complex. In addition, the emphasis on economic foundations behind this model, also plays a huge part in reducing the complexity of our model as we assume that market and fundamental characteristics

[17]The link between each indicator and the market characteristic becomes apparent during it's construction.

[18]We derive stress behaviour primarily from the link between equity characteristics and systemic risk, discussed extensively in the former chapter.

wholly drive systemic risk contribution of the equity segment.

The time scale that was ultimately chosen for all inputs was the daily format (so we convert quarterly fundamental ratios into daily ratios), and it can be justified as follows:

- First, given the complexity of our problem, daily aggregation over 5 years leaves a sizeable dataset (around 1500-2000 datapoints) for us to train our neural network.
- Second, given that fundamental characteristics form the undiversifiable market risk, one can argue that market characteristics are more important for the amplification of financial instability. Consequently, introducing a bias in favour of market indicators helps with answering the research question.
- Third, the variance of fundamental indicators, on a quarterly basis is in general smaller than that of the market ratios. Through quarterly aggregation, we would lose information from the daily market indicators, only to combine these indicators with fundamental indicators that are relatively stable over each quarter. From a machine learning perspective, this approach to data handling is a no-win situation.

Another challenge that needed addressing was whether a given year would have 252 days (i.e., we assume a year only consist of trading days) or 365 days. Given that we already introduce a bias towards market data previously, if we are to assume that a year has 252 days, then we would place such a heavy emphasis on the market indicators, that the predictions of our model would not differ despite the exclusion of fundamental indicators. Therefore, we assume that a year has 365 days. In addition, as we established previously in the Introduction, that the households have a strong exposure to capital markets, one can argue that impact of high systemic risk on real economy manifests in the form of decreased household consumption (Parker, 2012), which takes place primarily during weekends (LS et al., 2017). Therefore, there is also an economic reasoning behind including weekends in our analysis. To be critical of this choice, one can argue that since we investigate financial market indicators and not economic activities (such as production, consumption or distribution), our data reflects dissemination of price information, which occurs exclusively on trading days. Consequently, it might be better to assume that a year has 252 days. This is a fair argument and our choice of 365 days in a year poses as a limitation of this study[19].

As for the method of temporal (dis)aggregation, we chose linear interpolation for both market and fundamental data. This implies that discrepancies observed in asset prices between market open and close, especially during weekends is less pronounced than if we only considered trading days (i.e., the weekend effect is diminished). This is an assumption of our methodology. We

[19]We offer concrete suggestions on how to tackle this limitation in the Discussion chapter

tried following the ESS & Eurostat's Guidelines on Temporal Disaggregation (Buono et al., 2018), wherein we interpolate the fundamental features from a quarterly to a daily format, by using market OHLCV[20] data as our reference time series (market indicators cannot be used as references, because we found to have identical features both on the market and the fundamental side). However, this approach was unsuccessful. Given the low variance in quarterly fundamental features, we observed that all fundamental features closely resembled the market OHLCV data – to the point that inclusion of these "referenced" fundamental features in our dataset was redundant.

In this chapter, we have discussed data and its engineering with respect to our choices, assumptions and decisions. In addition, we also present indicators that can be recreated and linked to equity characteristics discussed in the former chapter. The next step is to include these indicators in our neural network framework, to ultimately relate equity characteristics to systemic risk. Since our neural network framework includes both sequence-to-sequence learning and encoder-decoder architecture, we have dedicated a complete chapter to explaining the details of our neural network titled "Model Choice". We subsequently add a follow-up chapter titled "Model Implementation" to explain the specifics entailing implementation of our model. The following chapters are important for the uninformed reader to grasp the structure and construction of our framework, to eventually recreate the results.

[20]OHLCV: Open, High, Low, Close, Volume quotes.

4 Model Choice

The main objective of this chapter is to provide clarity on the functioning of our neural network framework. We will begin with the encoder-decoder architecture, proceed with sequence-to-sequence learning before explaining how the neural networks use these components to relate equity characteristics to systemic risk. We will also demonstrate the idea behind using the attention mechanism and how it addresses problems associated with sequence-to-sequence learning. Given that this is a finance paper as opposed to a computer science paper, the emphasis will be placed on understanding our neural network framework and its mechanisms from an economic perspective.

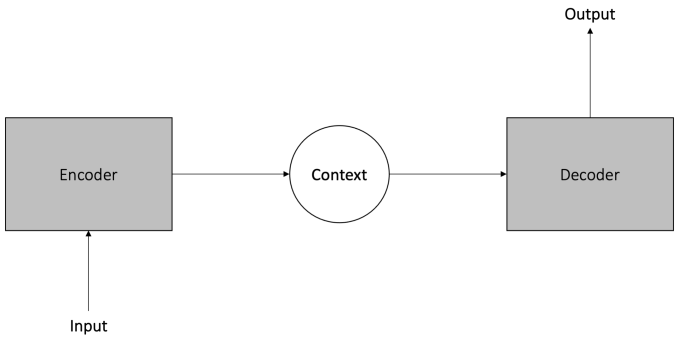

Figure 2: A Basic Encoder-Decoder Model. Grey indicates the involvement of a recurrent neural network.

An encoder-decoder model is a machine learning model that employs the use of recurrent neural networks for sequence-to-sequence prediction problems. Developed originally for machine translation problems such as text summarization and question answering, a basic encoder-decoder model performs the following fundamental functions: it takes an input sequence, encodes this input sequence into a context vector (where all the important information from the entire input is condensed into), which then helps the decoder to predict an output sequence. The encoder and the decoder are each supported by a recurrent neural network denoted in grey. This is illustrated in Figure 2.

The input sequence can be fed into the encoder either in the form of cross-sectional data or time-series data. A good example of cross-sectional data would be stocks that are sorted based on increasing book-to-market ratio, while that of time-series data would be the daily price of a given stock. In either case, inputs must be consistently given on a sequential basis. This property of consistently feeding inputs is highly relevant to the problem of financial stability since it allows for real-time monitoring of systemic risk build-up through market inputs. To encode inputs on a sequential basis in order to subsequently decode outputs from the context vector on a sequential basis, is the core concept behind sequence-to-sequence learning[21]. This is illustrated in Figure 3.

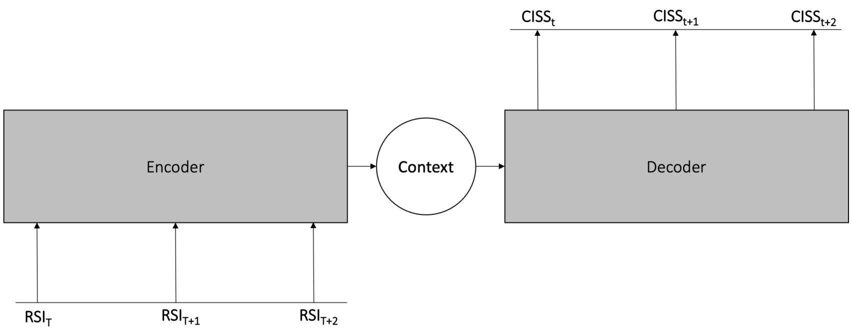

Figure 3: Sequential Encoding and Decoding. Note that 'T' and 't' can be two different points in time, just as 'N' and 'n' can be two different sequence lengths. Each time-step represents a daily value. Grey indicates the involvement of a recurrent neural network.

In Figure 3, we only consider a single indicator. We can progressively add more indicators such that we feed all market and fundamental indicators (discussed in the Data chapter) into our encoder. This is illustrated in Figure 4.

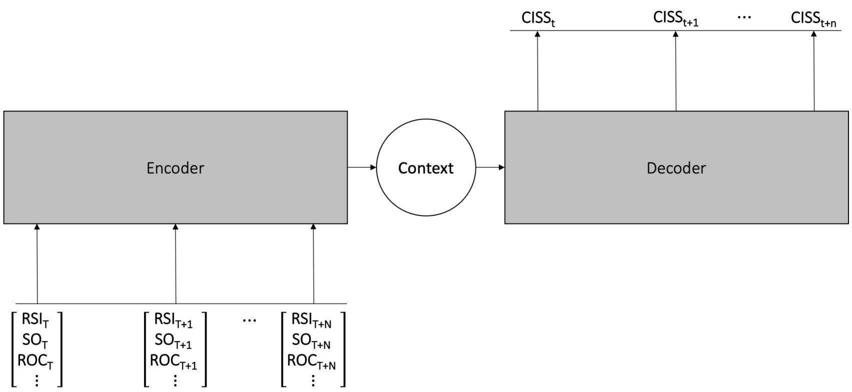

Figure 4: Sequential Encoding with all market and fundamental indicators, followed by Sequential Decoding. Note that 'T' and 't' can be two different points in time, just as 'N' and 'n' can be two different sequence lengths. Each time-step represents a daily value. Grey indicates the involvement of a recurrent neural network.

Given such sequential inputs of data, it would be ideal for information at each time point to

[21] It is important to note that each timestep 'T' in Figure 3 reflects a day in our dataset. With sequence-to-sequence learning, we can encode a sequence of several timesteps to learn market information in the previous weeks, to then decode equity market instability forecast for the upcoming weeks.

be sequentially encoded and forwarded to the subsequent time-point. The reasoning being that if our model were to be applied to real-time systemic risk forecasting, we want to be assured that all the information from previous timesteps is successfully encoded up until the current time-period, such that we obtain the most recent estimates of systemic risk measures. This is precisely how a recurrent neural network (RNN) fits into our framework.

The RNN is a feedforward neural network that has an internal memory (Sherstinsky, 2018) and is highly relevant for this research, since a RNN can read inputs at every time step, remember some information/context through hidden layer activations, which are then passed onto the next time step. Given that this process occurs recursively for every step, in the context of Figure 4, a uni-directional RNN considers information from time-period T while processing information from $T + N$.

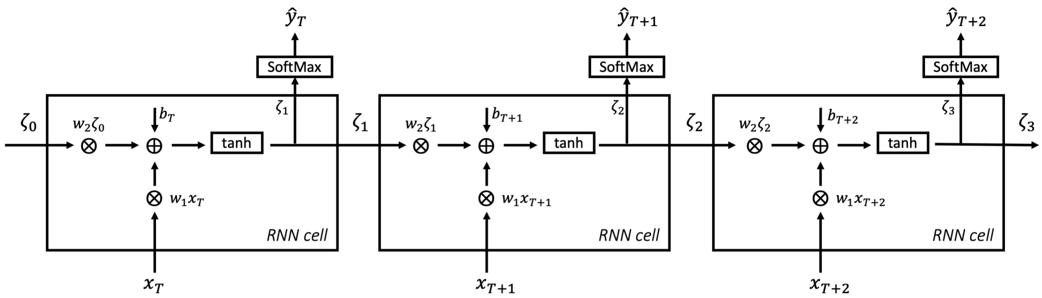

Figure 5: Snapshot of a Recurrent Neural Network

Figure 5 shows three cells within a uni-directional RNN in the context of our encoder-decoder model framework. The hidden states (ζ) progressively encode information supplied at each time point, such that the final hidden state (ζ_3) captures the information from x_T to x_{T+2}. This is the so called 'internal memory' described previously. 'x_T' refers to an n-dimensional column vector representing the market attributes at time-point T (as illustrated in Figure 4). w_1 and w_2 are weights to modulate the input (x) and the hidden state (ζ) at each time point. They can either be the same for each cell (as assumed in Figure 5) or can be tuned to optimize model performance. $\hat{y}_T$ is the output that can be produced from a single RNN cell at each time-point T, but since we opt for an encoder-decoder framework, we postpone this step such that the outputs are only produced after the final hidden state is encoded into the context vector (see Figure 6). This is how the effect of decoding is achieved. 'b_T, b_{T+1} and b_{T+2}' are coefficients shared temporally at each time-step and the point is that all of these characteristics pass through a 'tanh' function (see Equation 1), which is a mathematical activation function used to handle long-term dependencies over the entire input sequences.

$$tanh : g(x_T) = \frac{e^{x_T} - e^{-x_T}}{e^{x_T} + e^{-x_T}} \tag{1}$$

Intuitively, the activation function serves two purposes. First, it ensures that all the inputs are standardized such that the resulting transformation $g(x_T)$ is between -1 and 1. In that sense, it serves as a squashing function. Second, this standardization over each cell allows the neural network to prioritize some information transmitted through the hidden states as being "more important" than other information. Over the course of the entire input sequence, this implies that complex patterns which might be relevant for the decoding process are "learned" by the model.

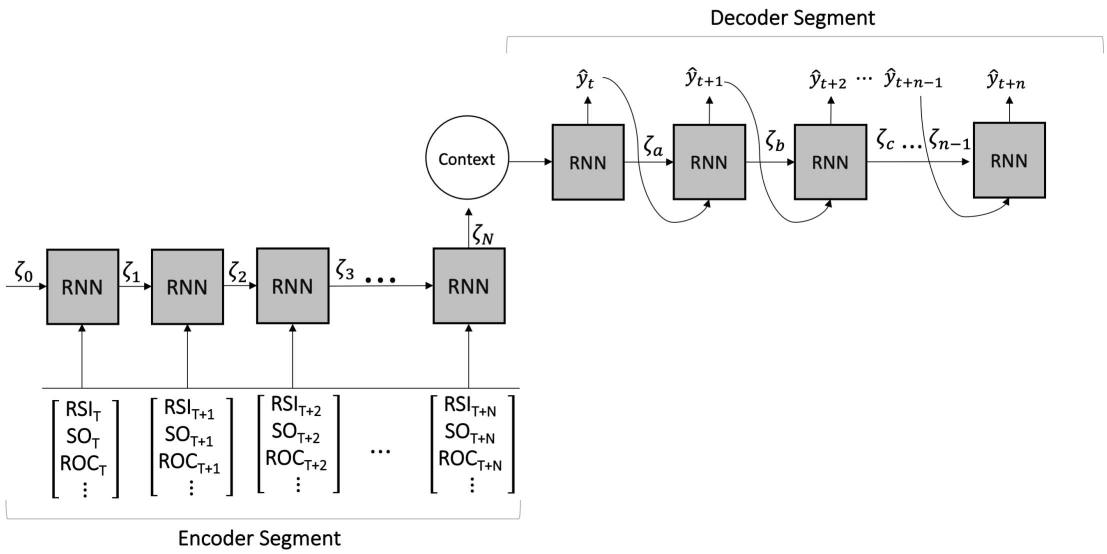

Figure 6: Final Setup of Neural Network Framework. This figure encapsulates the interplay between sequence-to-sequence learning, encoder-decoder architecture and recurrent neural networks (RNN). '$\hat{y}_t$' corresponds to the predicted CISS-E value at a given timestep 't'. 'N' and 'n' denote the encoding and decoding sequence lengths respectively.

The final setup of our neural network framework is shown in Figure 6. Note that inputs (market and fundamental indicators) must be fed at each timestep during the encoding, but during the decoding, the previous outputs form the inputs of the subsequent cells. From an economic perspective, such decoding is logical as our outputs are systemic risk estimates. Due to the procyclical nature of the economy, it is expected that periods of lower systemic risk (for instance, during an economic boom when there is increased public spending) precede subsequent periods of higher systemic risk (during an economic recession). Therefore, it is reasonable that former outputs serve as subsequent inputs during the decoding phase. During the decoding phase, it may be that information from future states is necessary for understanding information from prior periods. For instance, if a spike in CISS-E was observed at $t = 1$, and we want to discern if that spike is an anomaly, we would need information from following periods ($t = 2, 3, \ldots, t$) to reflect that the observation was an anomaly. In this instance, we would employ a bi-directional RNN structure to incorporate information from future states into past periods. This, however, falls outside the scope of this thesis.

When the encoder-decoder model was first proposed for neural translation, the input and the output sequences were linguistic phrases (Cho et al., 2014). An interesting observation in the paper was that as the input sequences got longer, there was a marked deterioration in the performance of the encoder-decoder model. To solve this problem, the attention mechanism was proposed (Bahdanau et al., 2015). They propose the following: Traditionally, last hidden state of the input sequence is used to form the context vector (ζ_N in Figure 6), which is then decoded. Instead, the weighted sum of all the hidden states ($\zeta_0, \zeta_1, ..., \zeta_N$) of the input sequence can be used to form the context vector. The underlying point being, that each hidden state which carries more information about the input sequence has a higher standardized weight thanks to the squashing 'tanh' function (see Figure 5).

For our model, we use a dot-product attention mechanism, which is a more refined attention mechanism (Luong et al., 2015). It is relevant to this research because in addition to calculating the weighted sum of the hidden states, we calculate an alignment score function to determine how well the hidden states of the encoder align with those of the decoder. Once these alignment scores are calculated, they can be standardized through a SoftMax function to sum to 1. Finally, we train our neural network to maximize the alignment score. The alignment score function is given below: ζ_i refers to hidden states within the encoder (From Figure 6, $i = 1, 2, 3 \ldots$) while ζ_j refers to hidden states within the decoder (From Figure 6, $j = a, b, c \ldots$) with $\mathbf{z}$ being a vector of all the hidden states and $\mathbf{T}$ indicating transposition of a vector.

$$Alignment\ Score : f_{att}(\zeta_i \zeta_j) = \mathbf{z^T}_i \mathbf{z}_j \tag{2}$$

The final model framework used for our analysis is summarized in Figure 7. For this analysis, we vary both input and output sequence lengths, while using the model framework in Figure 6 as our baseline model. We then augment the baseline model framework with the attention mechanism as shown in Figure 7 to see if it improves our results. Finally, we justify the best model (i.e., (i) a model with or without attention mechanism and (ii) a model with best performing encoding and decoding sequence lengths) relevant to answer the research question. This will be discussed extensively in the Results chapter.

In this section, the goal was to clarify the interplay between sequence-to-sequence learning, encoder-decoder architecture and recurrent neural networks. We start off with a basic encoder-

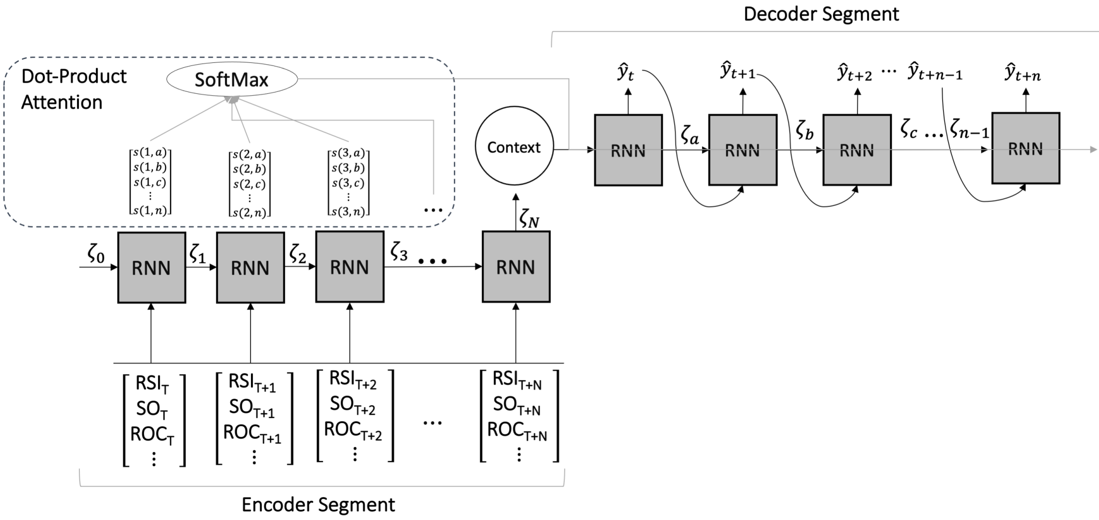

Figure 7: Final Setup of Neural Network Framework with Attention. This figure encapsulates the interplay between sequence-to-sequence learning, encoder-decoder architecture and recurrent neural networks (RNN). '$\hat{y}_T$' corresponds to the predicted CISS-E value at a given timestep 't'. 'N' and 'n' denote the encoding and decoding sequence lengths respectively. $s(1, a)$ is the alignment score of ζ_1 in the encoder with respect to ζ_a in the decoder. This logic extends to all $s(i, j)$, where $s(i, j)$ is the alignment score of hidden states at position 'i' within the encoder with respect to position 'j' at the decoder. For our training, we propose the use of a dot-product to calculate $s(i, j)$.

decoder model, outline the economic considerations relevant for the scope of this thesis, to finalize the model framework used in this thesis. In the following section, we outline the considerations made for implementation of this framework, relevant for the replication of our results.

5 Model Implementation

This section serves as a preamble to the results. We discuss primarily, the implementation of the encoder-decoder model. The objective is to justify our motivation behind our choices and highlight the limitations and assumptions relevant to the implementation of our model. For the reader interested in replicating our research, we also include hyperparameter specifications and model plots used for our analysis.

5.1 Motivation behind RNN structure

At each time step of our encoder-decoder model, there are three choices of RNN units that can be implemented: Basic RNN unit, Long-Short Term Memory (LSTM) unit and Gated Recurrent Unit (GRU). Each unit has a certain number of hidden layers, each with a certain number of nodes, that are responsible for modelling our data. The vector with market and fundamental indicators is passed through this unit at each timestep, to obtain hidden states (as discussed earlier). In our setup, we choose an LSTM cell, as it offers higher accuracy than a GRU and as it preserves information over many timesteps, a known limitation of the RNN (i.e., the vanishing gradient problem). As for the choice of hidden layers, we select 4 layers, and this was established through trial and error. There are more robust methods to estimate the optimal number of hidden layers and they will be discussed in the Evaluation section.

5.2 Preference for Encoder-Decoder Model

To think critically, a policy maker may ask why the use of a encoder-decoder model is warranted, in place of a simple multi-output regression or a simple LSTM (see Figure 5) to predict multiple values of the output sequences. The answer is because this model captures conditions needed for Granger-Causality (i.e. cause precedes effect, cause has unique information about future values of the effect, cause and effect are temporally related) between our input and output sequences. This is not captured in neither the multi-output regressor nor a single LSTM cell. In addition, our architecture captures instantaneous (via Adaptive Moment Estimation (Adam)), non-linear (via Exponential Activation Function (ELU)) causal relationships between the sequences, rendering it superior to the statistical Granger-Causality test.

5.3 Motivation behind Truncation, Sequence Generation & Model Validation

As established in the Data chapter, we consider 365 *days* $\times$ 5 *years* $=$ 1825 *timesteps* each for 81 input features (market and fundamental indicators) and 1 output feature (CISS-E). However, to evaluate a machine learning model, it is necessary to split the available data into training and test data. This implies that instead of using one-long input sequence of size (1825×81) to predict one-long output sequence of size (1825×1), it is necessary to create subsamples or *truncations* such that some samples can be reserved for training our model, while the rest are reserved for testing our model.

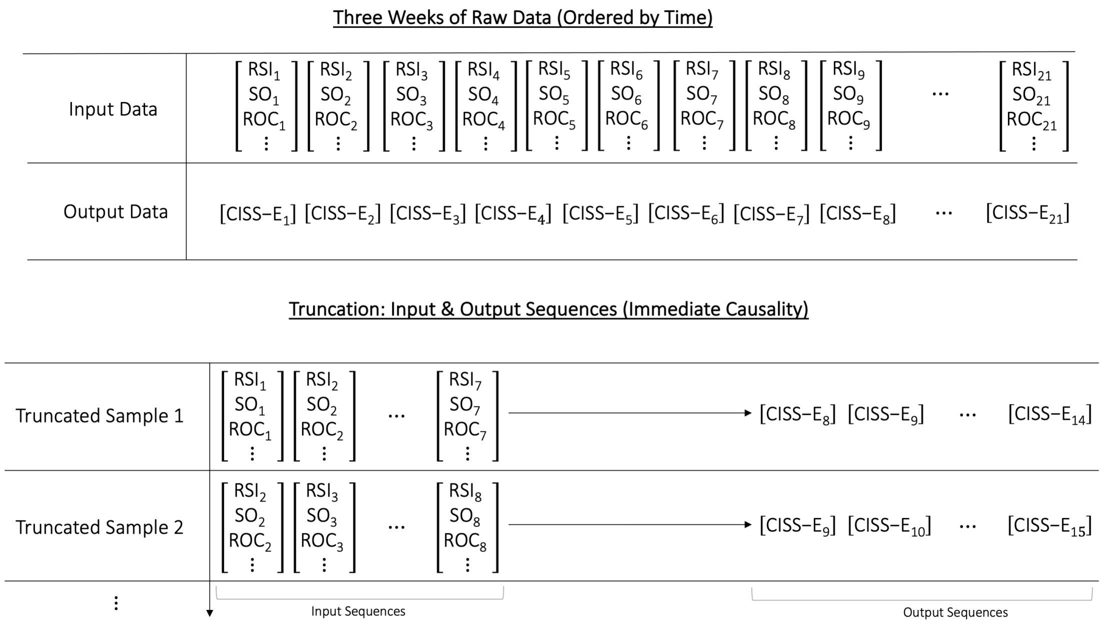

Figure 8: Truncation and Sequence Generation. The raw input and output data, consists of 81 market and fundamental indicators (established in the Data chapter) and 1 benchmark respectively, ordered by time. We assume immediate causality between our input features and output targets (i.e., the first 1×1 vector of the output sequence is exactly one timestep away from the last 81×1 vector of the input sequence). We assume rolling truncations, with our input and output sequence rolled forward one timestep (i.e., one day) until we reach the final input or output vector of our dataset (timestep 21 in the case of three weeks of raw data). We vary the length of input and output sequences in our Results chapter but keep the other assumptions fixed. Each input and output sequence pair forms a truncated sample or a *truncation*.

For the truncations, we make two choices: First, truncations are performed only along the time component (for instance, each input truncation must have 81 features and the number of timesteps must be lower than 1825—this applies to output truncations too) and second, the truncations are performed on a rolling basis. By rolling truncations, assuming that each input sequence is one week long (so each input sequence has 7 time-steps), then input sequence #1 would be from time step 1 to time step 7, while input sequence #2 would be from time-step 2 to time step 8 and so forth. This is also illustrated in the input sequences of truncated samples in Figure 8.

To facilitate *Neural Granger Causality*, we assume that input sequences immediately precede output sequences (Immediate Causality in Figure 8); however, the length of input and output sequences is variable. So, assuming that one week of input information "causes" an observable effect for the CISS-E in the subsequent week, we pair input and output sequences as follows:

- Input sequence #1: 81 indicators, timestep 1 to time step 7 "causes" Output sequence #1: 1 benchmark, timestep 8 to timestep 14;
- Input sequence #2: 81 indicators, timestep 2 to time step 8 "causes" Output sequence #2: 1 benchmark, timestep 9 to timestep 15;
- and so on.

Each input-output sequence pair subsequently forms a truncated sample. This is also illustrated in Figure 8.

Before we justify the lengths of input and output sequences experimented with by our model, we briefly discuss our approach to model validation. From an economic perspective, it has been established that stock market cycle precedes the economic cycle from a range of 6 months to 10 years. However, given the size of our dataset (5 years of daily data) and our methodology of sequence creation, sequence lengths that reflect short-medium term effects (i.e. input and output lengths of 365 days) would need a sizeable portion of our data, allocated to testing the model. From a machine learning perspective, having a smaller training sample could result in overfitting, implying that the predictions of our model are poor when the test data does not resemble our training data (for instance, during crisis periods). We therefore balance economic findings against the practical relevance of our encoder-decoder model.

To strike this balance, the following choices were made: First, we adopt a hold-out validation approach with a train-test split of 80-20% and, second, we focus on short-term effects of market characteristics on equity market instability. We adopt 3 week increments of sequence lengths up to a maximum of 12 weeks for both inputs and outputs (so we have $4 \times 4 = 16$ input and output length combinations: (3,6,9,12 weeks) $\times$ (3,6,9,12 weeks)). The 80-20 split can be motivated by the fact that most of the covid-19 pandemic's effects on the financial markets and CISS-E were observed in early 2020. As a result, most of the test-dataset will contain samples that were infrequently observed during training, suggesting a fair reflection of the potential of our model. In addition, the focus on short-term effects ensures that enough truncations are saved for validation, despite only having 20% of our data for testing. The choice for a hold-out validation approach, is primarily due to reduced computational time and power requirements. Practically speaking, we train the model once (unlike cross validation) and test our model on the covid-19 crisis data. Given that there's an inherent data imbalance between crisis periods and stable periods (i.e., stable periods are longer than crisis periods, in general), our current hold-out specification makes for more robust results, than for instance, cross validation.

5.4 Additional Considerations

- As we consider ELU as our activation function, we want to avoid gradient explosion. This is precisely why we introduce batch normalization—to standardize the inputs to a layer for a given mini-batch such that this batch follows a standard normal distribution ($\mu = 0,\ \sigma = 1$). This is specific to the training phase.
- To implement the encoder-decoder model, we adopt a Time-Distributed Dense layer as opposed to a standard Dense layer for the decoding phase. This implies that our LSTM has the same weights applied to a given timestep during the LSTM un-rolling.
- For model evaluation, we consider mean square error (MSE) as the loss function, with the mean absolute error (MAE[22]) as our evaluation metric. In addition, we also plot value distributions, random train-test samples to gauge our predictions and variation of MAE over training and testing sequences to check for any periodical patterns.
- A summary table with model specifications is presented next. Choices which have not been discussed were found through trial and error—and possible improvements for these choices will be discussed in the Discussion Chapter. All models shown in results were implemented on the Keras API. The optimal model plots (discussed in Results) are available in Appendix A.2, which along with the specifications, allow for efficient recreation of our results.

Table 3: Model Specifications. LSTM structure is the same, both on the encoder and decoder segment.

Hidden Layers:	*4*	Recurrent_Dropout:	*0.2*	Callback:	*Early Stopping*
Activation:	*ELU*	Optimizer:	*Adam*	Monitor:	*Validation Loss*
LSTM Layers:	*1*	Learning Rate:	*0.01*	Mode:	*Minimum*
Dropout:	*0.2*	Clipnorm:	*1*	Patience:	*20*
Batch Size:	*100*	Epoch Size:	*100*		

[22] $MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$, where y_i is the predicted CISS-E value, x_i is the actual CISS-E value and n is the sample size.

6 Results

The objective of this chapter is to present the results of our encoder-decoder model implementation, as an attempt to relate characteristics of equity market to systemic risk in the context of EU. It is structured as follows: First, we execute data pre-processing techniques on a baseline encoder-decoder model specification with fixed input and output sequence lengths to establish data considerations most appropriate for this analysis. Then, we vary the input and output sequences as discussed in the previous chapter, adding the attention mechanism to our model to assess the relevance of market and fundamental indicators in forecasting the CISS-E. Finally, we establish the characteristics of an optimal encoder-decoder model (i.e., relevance of attention mechanism, optimal input and output sequence lengths for our research question) based on our evaluation metric, the MAE. We conclude this section with critical look at our model choice discussing feature importance and stress testing, to emphasize the impact of this research.

6.1 Data Pre-Processing

The objective of trying different pre-processing techniques is to determine the set of assumptions that best suit our specifications (as discussed in the chapter on Model Implementation). This is relevant to this research as optimal assumptions best enable optimal training of our encoder-decoder model. We will select a single specification of input and output sequence lengths (outside the set of 16 specifications mentioned earlier as we don't want our pre-processing to introduce a bias towards one specification). We will then investigate data transformation, stationarity, and their implications for model training. Finally, we conclude this subsection with the final assumptions regarding data transformation and stationarity.

6.1.1 Data Transformations

The specification of input and output sequence is arbitrarily chosen to be 7 weeks and 3 weeks, respectively. We look 7 weeks (49 days) back on market and fundamental indicators to forecast the subsequent 3 weeks (21 days) of CISS-E estimations. For the data transformations, we look at three types, each building on top of each other:

For the first type, we set a maximum limit by scaling each observation in the observation matrix X, such that the maximum of the transformed values is capped to 1. So, the transformation (X_{sc}) is given by:

$$X_{sc} = \frac{X}{max(X)} \tag{3}$$

For the second type, we build on the previous transformation by ensuring that the minimum of transformed values is capped to 0, such that the transformation X_{sc} has a range of [0,1]. The transformation is now given by:

$$X_{sc} = \frac{X - min(X)}{max(X) - min(X)} \tag{4}$$

For the final type, we consider a Quantile Transformation, wherein we use a deterministic, rank-based, inverse normal transformation to approximate expected normal values of market and fundamental indicators, as used by Holló et al. (2012) for the construction of CISS-E. During the back-transformation of ranks, it is important to clip the maximum and minimum rank-values to ensure that they are not transformed to positive and negative infinity respectively. For this paper, we set the minimum and maximum to $1e - 7$ and $1 - (1e - 7)$ quantiles respectively. The transformation (excluding this clipping) can be expressed simply as follows, where F and Φ represent the cumulative distribution functions of an empirical and standard normal distribution respectively:

$$X_{sc} = \Phi^{-1}(F(X)) \tag{5}$$

There are two important points to note for all three types of transformations:

- These transformations must be made prior to creating the *truncations*. Else, given the rolling nature of the truncated sample creation (see Figure 8), certain values might be over represented than others and this may influence the distribution functions during the quantile transformation.
- All transformations must be made exclusively based on information from the training dataset. Else, if these transformations are applied across the entire dataset, then a bias is introduced into model evaluation as information from the test set is incorporated into the training set.

We construct the encoder-decoder model based on the specifications in Table 3 and apply the three transformations discussed above. The training value distributions and training predictions of all transformations are expressed below, in Figures 9 and 10 respectively.

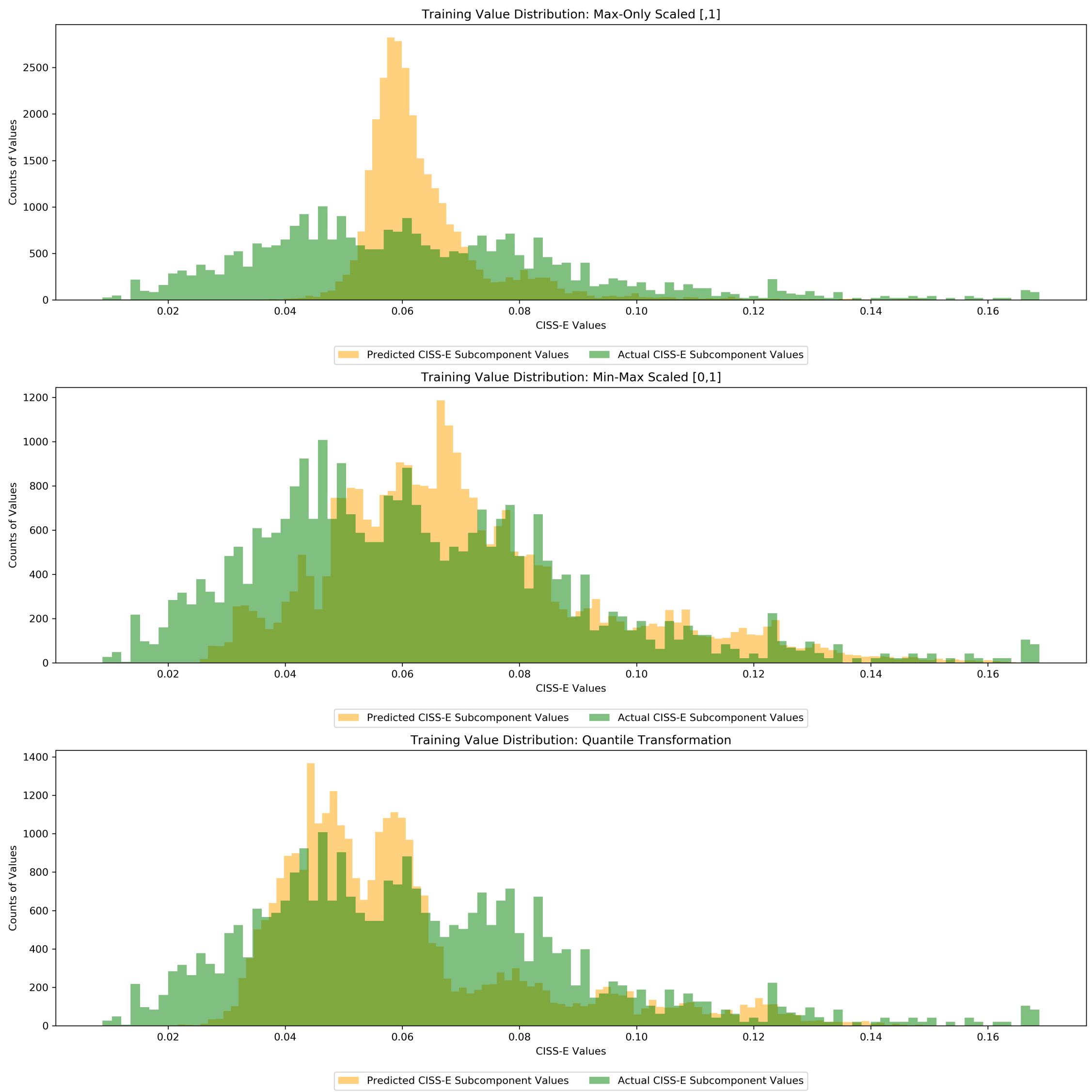

Figure 9: Value distributions of output sequences, for different transformations on the input sequences of the training dataset. Generally speaking, the greater the overlap between predicted and actual CISS-E value distributions, the more appropriate the model. However, given that we estimate equity market instability, we favour overestimation of instability rather than its underestimation. As CISS-E values increase, the optimal transformation therefore ought to have increased overlap between the predicted and actual CISS-E values. This is best observed for Min-Max Transformation.

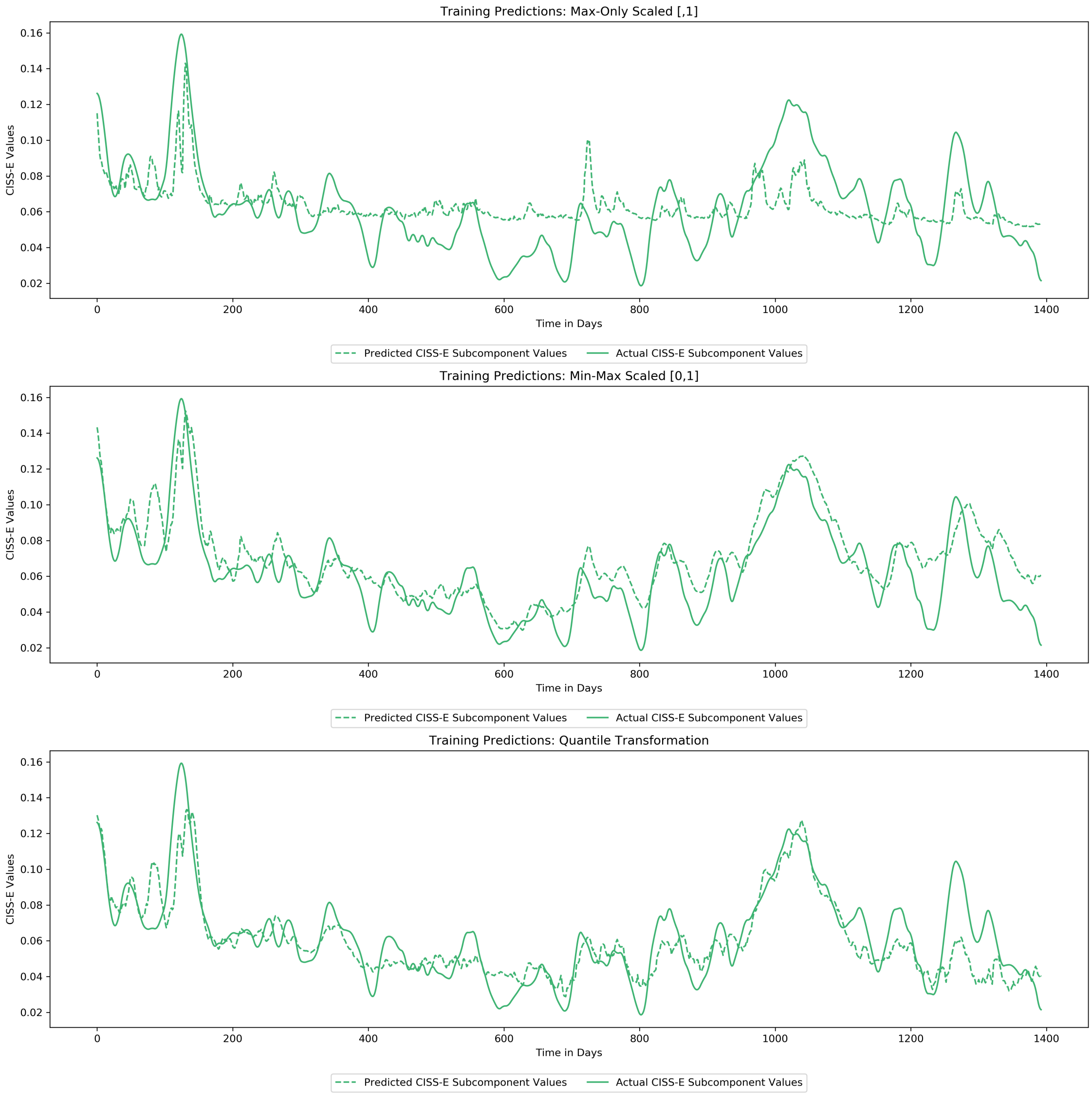

Figure 10: CISS-E predictions for different transformations on the input sequences of the training dataset.

From the training distributions in Figure 9, we observe that the overlap in predicted and actual distributions progressively improve from max-only transformations to quantile transformations. For the min-max transformations, we observe a distinct spike in prediction counts at the CISS-E value of 0.07. For the max-only transformation, we observe a similar spike at the value of 0.06, while for the quantile transformation, we observe two distinct spikes at 0.046 and 0.06, respectively. These observations imply that the min-max transformation tends to overpredict the level of equity market instability, relative to the other transformations. This reasoning is also supported by the fact that the distribution overlap between the predicted and the actual CISS-E values, following the prediction count spike (i.e., for CISS-E values of 0.07 and higher), is the greatest for min-max transformation relative to the other transformations.

From the predictions in Figure 10, we observe that the min-max transformation is the most appropriate to assess equity market instability, especially from days 1200-1400. The phenomenon of over prediction, as noted from the value distributions, is also observed in the predictions of the min-max transformations (days 900-1100). For the most part, we observe that the quantile distribution produces effective predictions. However, the min-max transformation is the most sensitive to changes in equity market instability.

For the plots of training predictions, in Figure 10, one might have expected a 3-dimension plot. As we predict sequences of 21 timesteps on a rolling basis, it would be ideal to understand how predictions vary per sequence and over time. However, as there is a significant overlap between one truncation and the next, visualizing both dimensions comes at the expense of clarity. Therefore, in Figure 10 and for the rest of the paper, we have prioritized for variation in predictions over time. This means that each value of a prediction on a given timestep, is the average predicted value across all sequences that include the exact timestep.

The main take-away from this subsection is that we proceed with the min-max transformation for the remainder of the analysis. In the context of systemic risk, an ideal predictive model is sensitive to changes in systemic stress. It helps to slightly overpredict the forecasted level of financial stress, as that allows financial supervisors to manage the consequences of equity market instability towards the real economy. In the next subsection, we continue with further data pre-processing. We investigate the presence of non-stationarity in our dataset and explore its consequences for our analysis.

6.1.2 Stationarity

Stationarity is relevant to the scope of our research as we perform time-series forecasting (i.e., we forecast CISS-E sequences purely based on market and fundamental indicators). A fundamental assumption of time-series analysis is that the process generating the time-series, retains statistical properties that are constant over time. If this assumption is violated, a time-series is said to be 'non-stationary' and the statistical relationships established by the encoder-decoder model are likely to change over time.

There are contrasting views on whether or not stationarity improves predictions in the machine learning literature. For regression problems with LSTMs, whereby prediction is the sole objective, stationarity in time-series data is useful. For sequence-to-sequence problems, whereby prediction and temporal dependency are both important objectives, RNN based sequence-to-sequence architectures perform well with non-stationary data (Wang et al., 2020). It is for this reason that we investigate stationarity by investigating both unit-root and non unit-root causes of non-stationarity. We study prediction and distribution plots to check if stationarity is relevant for this analysis.

Unit Root Causes

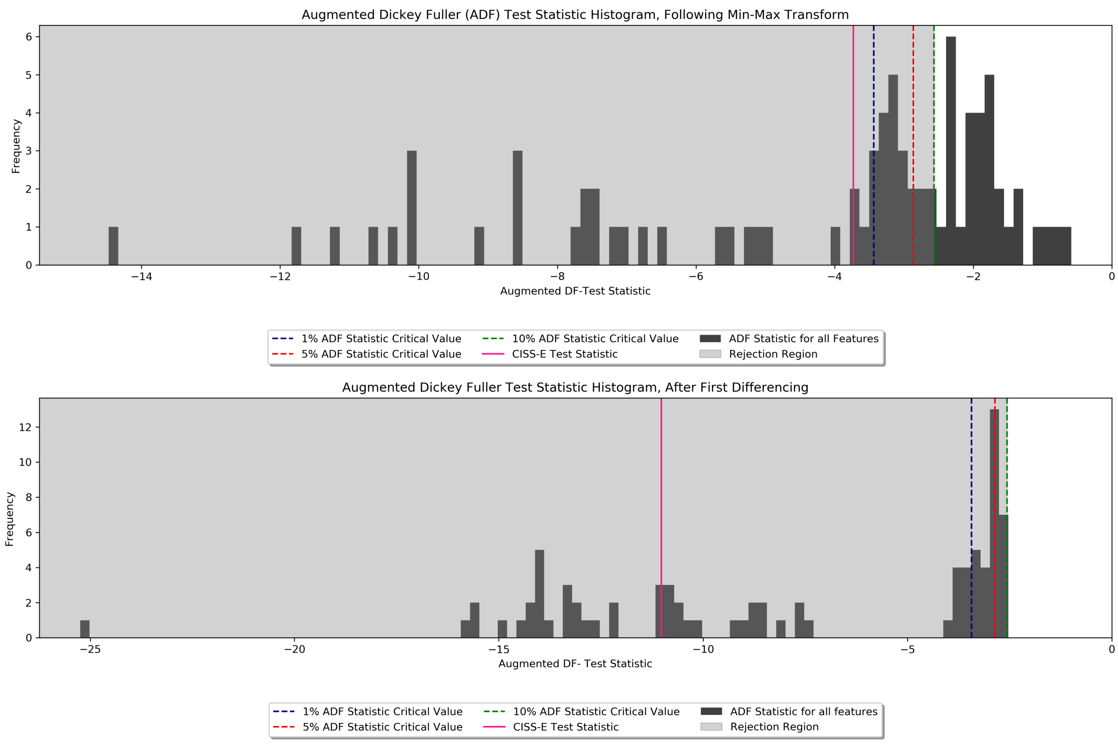

Figure 11: Augmented Dickey Fuller Test Statistics Represented in A Histogram, Pre- & Post-First Differencing. The features are all market and fundamental indicators discussed in the Data section.

To establish whether our dataset suffers from unit-root causes of non-stationarity, we use the Augmented Dickey Fuller (ADF) test on time-series of market and fundamental indicators. The null hypothesis of the test is that a unit root is present in the timeseries and the alternative hypothesis is that there is no unit root present in the time series data. In the Figure 11, we plot the ADF Test-Statistic Histograms. The X-axis represents the ADF test-statistic, which is relevant because the more negative it is, the higher the confidence interval with which we reject the null hypothesis. The Y-axis represents the frequency, which represents the number of indicators (features) falling within a given interval of the test-statistic. We observe that following the Min-Max Transform, there are substantial number of features that fall outside the rejection region (top)—i.e. they have a unit root. However, following first differencing (i.e. computing the series of changes from one time step to the next), we reject the presence of a unit root for all features at a 10% confidence interval.

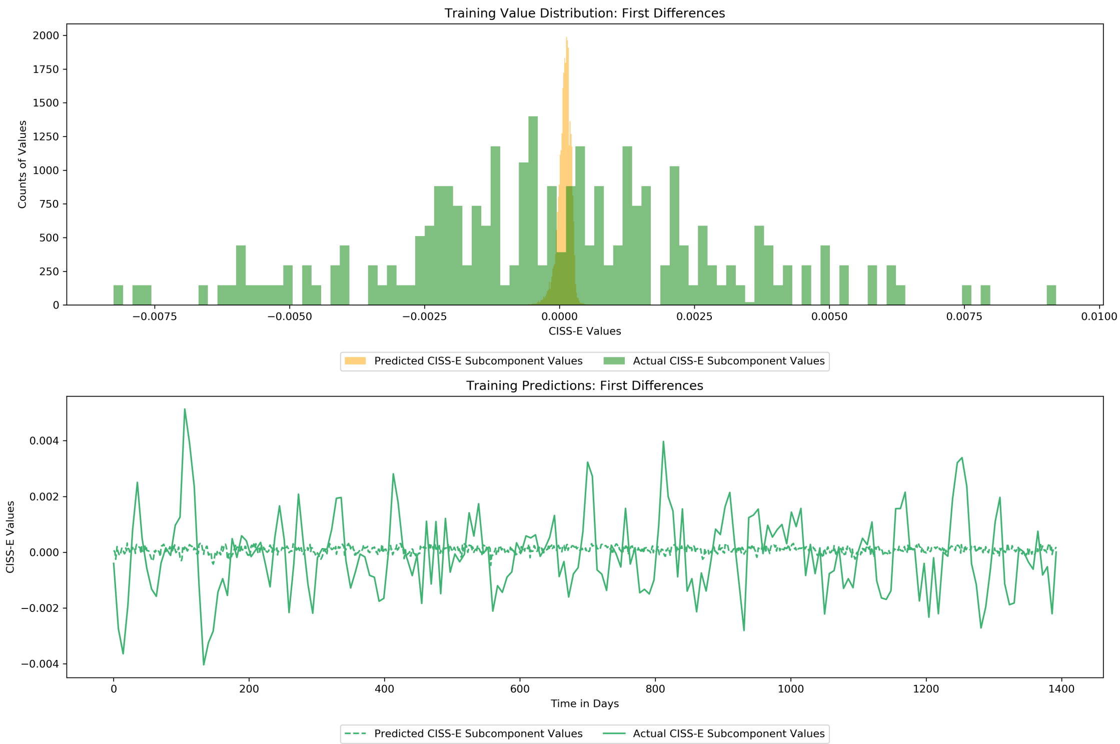

Figure 12: Training Value Distributions & Predictions Following First Differencing

Following first differencing, we plot the training value distributions and the training predictions, to check if first differencing is important for data pre-processing. As we observe in Figure 12, first differencing is ineffective as we ensure stationarity both in the encoder-inputs (features) and the decoder-outputs (targets). When we train the encoder-decoder model and plot the predictions, we check whether changes in features are predictive of changes in targets. This is however, not the objective of this thesis. The purpose of the CISS-E benchmark is to check if the encoder-decoder model makes logical predictions of the contribution of equity market segment to systemic stress. Thus, further analysis is warranted, particularly to ensure stationarity in the features without modifying the targets, for instance through time series decomposition.

Time Series Decomposition

Time Series Decomposition is a useful tool to assess both unit root and non unit root processes in time-series, and is therefore superior to the ADF test in statistical power. The objective of a time-series decomposition is to visualize a given time series as a combination of its level, trend, seasonality and noise components. By separating the noise component from the time series, we exclude non unit-root causes of non-stationarity. Consequently, instead of using features directly, we use the noise associated with each feature, to predict the CISS-E values for each truncated sample. Doing so, we modify the features but do not influence the benchmark.

There are two choices to decompose a given time series: through Additive or Multiplicative Decomposition. Additive decomposition assumes that the original time series can be recreated if its components can be added together. Multiplicative decomposition, on the other hand, assumes that the recreation occurs only if the components can be multiplied with each other. For this thesis, we use the type of decomposition that best matches the seasonality. Multiplicative seasonality assumes that the width of the seasonal pattern increases with the trend, however, additive seasonality assumes that the width of the seasonal pattern is fairly constant over time. A plot (Figure 19 is included in Appendix A.3) to help visualize the two forms of seasonality. To visualize the seasonality, we have plotted all features in our data. For each plot of 81 features, we consider whether the seasonal component is additive or multiplicative. We only consider the training portion for our choices since we want to limit the bias in our model. An example is illustrated in Figure 13.

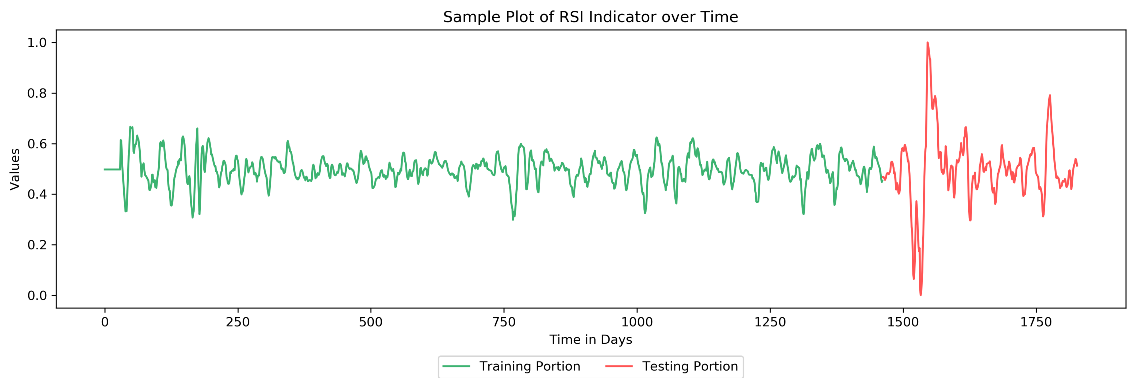

Figure 13: Sample plot to check for type of seasonality.

Based on plots of all features, we opt for additive seasonal decomposition. Figure 20 in Appendix A.3 shows a sample decomposition of one of the features, with the observed time-series being a linear combination of the trend, seasonal and the residual components. This decomposition was repeated for all features. Once the residuals have been extracted, we perform the test-train split and create truncation samples to feed into the encoder-decoder model. We use the truncated samples part of the training dataset to forecast the untransformed CISS-E values. The value distributions and the predictions are displayed in Figure 14.

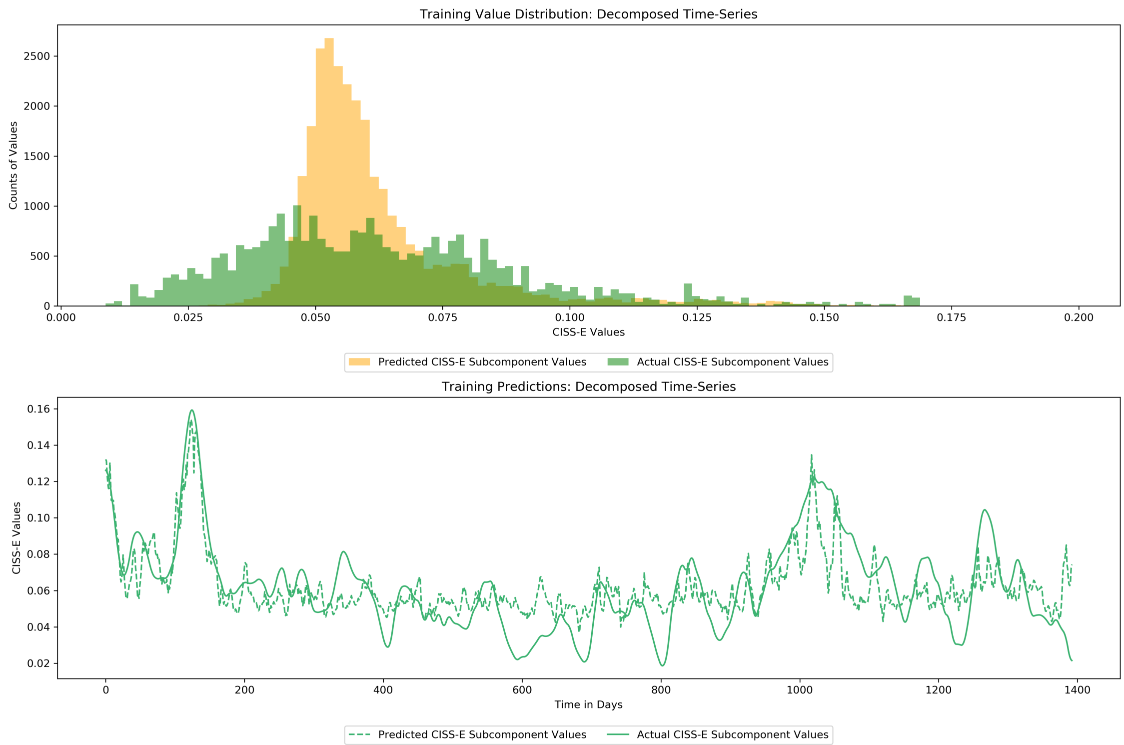

Figure 14: Training Value Distributions & Predictions Following Additive Decomposition.

From Figure 14, we observe that the phenomenon of overprediction, as was previously observed in Figure 10, is diminished. The overlapping area between the actual and predicted CISS-E values, especially for higher CISS-E values (larger than 0.075) is considerably smaller. Instead, we observe a higher sensitivity to stress, as the predictions are more noisy in Figure 14. This is particularly true for smaller spikes of CISS-E. Comparing the value distributions to Figure 9, we observe that the current model specifications (described in Table 3) are more suitable for CISS-E estimation, when stationarity is not ensured within the indicator data. This consideration allows for the accurate forecast of the highest contribution of equity segment to systemic stress.

A critical reader might wonder why such an elaborate section on stationarity is necessary for the goal of this thesis. The reason is because market regime forms an important equity characteristic that is relevant to equity market instability (see all indicators that conform to market regime in Appendix A.1). Most of the indicators that conform to market regime, are trend based and time series decomposition would remove trend-based effects. From the Economic Foundations chapter, we support the importance of market regime to this research, from an academic-literature perspective. The in-depth analysis of stationarity supports the importance of market regime, particularly in relation to equity market instability forecasting, from an empirical standpoint.

6.2 Optimal Model Selection for CISS-E Estimation

Following data pre-processing, we proceed to optimal model selection. As discussed in the Model Implementation chapter, we adopt 3 week increments of sequence lengths up to a maximum of 12 weeks for both inputs and outputs ($4 \times 4 = 16$ input and output length combinations: (3,6,9,12 weeks) $\times$ (3,6,9,12 weeks)). The motivation for having so many specifications, is to successfully predict an output sequence of CISS-E values that is greater than 3 weeks[23]. Afterall, the longer the prediction horizon, the more useful this thesis is for supervision. We start with a baseline encoder-decoder model (see Figure 6), implement the 16 input-output combinations and augment the baseline model with attention mechanism (see Figure 7) to check if it improves our predictions. The results, based on MAE (the chosen evaluation metric from Model Implementation chapter), are displayed in Table 4.

Table 4: Average mean absolute error (MAE) per model, for varying input and output sequence lengths during training and testing.

Input Lengths	**Output Lengths**								**Model Choice**
	3 Weeks		6 Weeks		9 Weeks		12 Weeks		
	Train	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*	
3 Weeks	0.019	0.042	0.019	0.049	0.020	0.041	0.021	0.048	LSTM-Encoder Decoder
6 Weeks	0.019	0.047	0.020	0.039	0.020	0.040	0.020	0.043	
9 Weeks	0.019	0.039	0.019	0.044	0.020	0.037	0.021	0.040	
12 Weeks	0.017	0.030	0.021	0.035	0.020	0.020	0.020	0.041	
3 Weeks	0.018	0.061	0.019	0.093	0.023	0.054	0.021	0.086	LSTM-Encoder Decoder with Attention
6 Weeks	0.013	0.142	0.019	0.099	0.020	0.085	0.020	0.074	
9 Weeks	0.015	0.078	0.018	0.065	0.020	0.062	0.020	0.037	
12 Weeks	0.014	0.085	0.018	0.075	0.018	0.070	0.019	0.138	

On a general note, the MAE values for all specifications seem to be low. The accuracy during training (i.e., periods of low stress) is around 98%[24] while during testing (i.e., a combination

[23] In the Data Pre-Processing section, we adopt a 3 week look-forward period for the data. In this section, we build on that by considering longer look-forward periods.

[24] Accuracy = 1 - Error

of high and low stress) the accuracy varies between 86-98%. The results show promise for the encoder-decoder model as it's relevance for systemic risk has not yet been explored in academic literature[25].

We observe that the errors on the testing dataset are much larger than on the training dataset, for the encoder-decoder model with or without attention. This is to be expected since for the sake of robustness, we include the Covid-19 pandemic in the test set.

On the training dataset, for a given output sequence length, as we increase the input sequence length, we observe that the MAE decreases for both the baseline model as well as the model with global attention. This is expected because the greater the input sequence, the more information encoded and the more information our decoder has to estimate CISS-E sequences. On the test set, we don't observe a similar pattern. This is attributable to the inclusion of Covid-19 crisis in the test set—as a result, this trend is only viable during times of low systemic stress. This conclusion also applies when we consider variable output sequence lengths for a fixed input sequence length.

Evaluating the encoder-decoder model with and without attention mechanism, we observe that the MAE on the training set is similar for either model choice. However, the test MAE is considerably larger in the model with attention. This is an unexpected finding. Given the global attention adds weights to each timestep of the input sequence, it is expected that an encoder-decoder model with attention has a comparative advantage over the baseline model[26]. This finding could be attributed to a possible bias introduced by our data considerations. Given that the initial data pre-processing has only been done on the baseline model, the assumptions chosen for the dataset might be ill-suited for an encoder-decoder model with attention. The inclusion of the covid-19 pandemic in the test dataset might have amplified this bias[27].

To select the optimal model choice, we consider the models with the best performance in the testing set. In addition, we also consider the model with the lowest error discrepancy between the training and the testing set, as that indicates a robust model. Given this criteria, we choose a simple encoder-decoder model with an input sequence length of 12 weeks to forecast equity market instability over the following 9 weeks (see Table 4). We denote the optimal model as Model (9,12) for the remainder of this chapter.

To check whether the predictions are similar during model training and testing, we discuss

[25]Given that MAE ought to be as close to zero as possible, the prediction error of the encoder-decoder model is low. However, we exercise caution towards the interpretation of these numbers. The CISS-E is standardized to a scale between 0 and 1, and as we use 1825 datapoints, the MAE is likely to be low from Footnote 22. That said, we did not specify that the predicted values of CISS-E must be between 0 and 1, so we retain our position regarding the promise of the encoder-decoder model.

[26]Attention adds an extra layer of weights to determine which timestep has the most relevance for the output sequence, and therefore has an informational edge over the baseline model.

[27]We have also given thought to whether or not the model specifications mentioned in Table 3 are appropriate. As the training behaviour for both models is generally similar, we assume that they are fairly appropriate.

the training and testing predictions of Model (9,12), along with its loss behaviour and prediction distributions. This is illustrated in Figure 15.

From the training and testing predictions, we observe that the model is able to capture the trend of equity market instability. This is consistent for both the training and the testing portions. From the loss behaviour, we observe that the training loss approaches 0 and the initial epochs are extremely relevant for training the model.

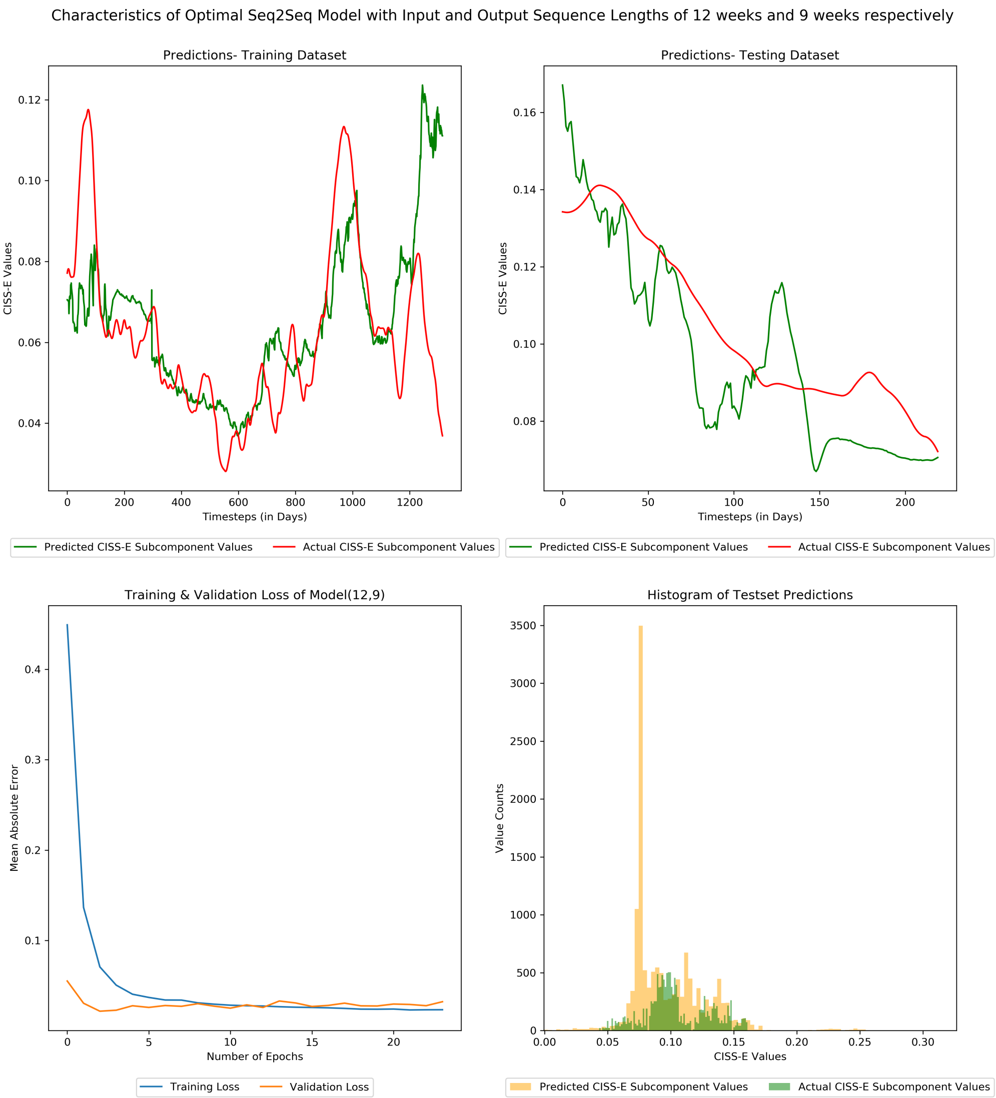

Figure 15: Attributes of Optimal Model

With the validation (or testing) loss, we observe that the loss for 10 epochs is lower than the training loss before slightly increasing for the subsequent epochs. This could be attributed to the choice of a constant learning rate. When the learning rate is too high, it may be that during gradient descent, an iteration passes the local minimum of the loss function, resulting in an increased error. A solution we suggest for future work, is to incorporate learning rate scheduling to tackle this issue. We also observe that the histogram of testset predictions shows a large spike at the CISS-E value of 0.075. This might be resolved with better model specifications, for instance through hyperparameter tuning. This is another suggestion for future work.

6.2.1 A Critical Look at the Optimal Model

In the final part of the results, we highlight the impact of this research. To illustrate how responsive the model is to stress, we conduct a stress test on the optimal model. As volatility is included in the CISS-E calculation by Holló et al. (2012), we also outline the five most influential indicators relevant for CISS-E prediction for our optimal model. If volatility-based indicators seem to be the most influential, it reveals a major limitation of this research[28].

For the stress test, we increase the values of each indicator in the test dataset by 25%, 50% and 75%, depending on the link between its market characteristic class and systemic risk (see Table 6 in Appendix A.1). We then load the trained optimal model to predict CISS-E values based on the "stressed" inputs. The results of this exercise are displayed in Table 16 below.

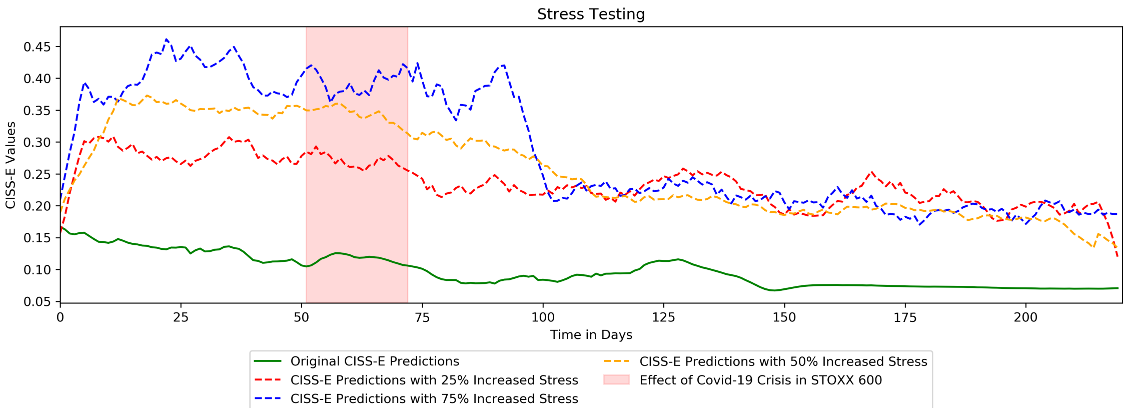

Figure 16: CISS-E Predictions of the Optimal Model with Increasing Levels of Financial Stress

The stress test shows that the market indicators are important to signal increased level of equity market instability, particularly around an impending crisis. During more stable periods (Day 100 and above), we observe an increased level of equity market in the form of a constant vertical shift, regardless of whether the financial stress is increased by 25%, 50% or 75%. During instable periods (Day 0-100), we observe that an increase in financial stress is followed by a corresponding shift in

[28]Solutions to this limitation are highlighted in the Discussion chapter

the CISS-E value estimations. This observation supports the core assumption of the thesis, that the market characteristics *shape* systemic risk due to their amplification and wealth effects. During stable periods, this amplification effect is limited, which explains the constant shift in CISS-E estimations. The optimal model we propose is responsive to equity market stress, particularly in times of financial instability.

In order to identify important features in the optimal model, we perform a sensitivity analysis. The methodology is the following: For each truncation sample, we first compute the predicted CISS-E values '$\hat{y}$' of the optimal model. Then, we perturb each feature (market or fundamental indicator) iteratively by a random normal distribution centered at 0, with scale 0.2 to compute '$\hat{y}_i$' (where $i = 81$: the number of features, i.e., we obtain 81 prediction sequences, per truncated sample). We then measure the effect of perturbation for each prediction sequence, by computing its root mean square error (RMSE) relative to $\hat{y}$. The higher the RMSE, the more the relative importance of a feature. We repeat this process for each truncation sample in both the training and the test dataset (since the results are consistent among either dataset). We calculate the average RMSE per feature, across all truncation samples[29]. The five features with the largest average RMSE are displayed below.

Table 5: Features with the Highest Relevance for the Optimal Encoder-Decoder Model.

Characteristic	Indicator	Perturbation Effect (RMSE)
Volatility	ROC	0.0982
Volatility	RSI	0.0931
Market Regime	ADX	0.0782
Market Regime	MACD	0.0772
Liquidity	MFI	0.0652

From Table 5, we observe that the characteristics with highest relevance seem to be market related, with a good balance among all three market characteristics considered for this thesis. The highest perturbation is observed in indicators relying on volatility, followed by those relying on market regime and liquidity. The findings lend further support to the relevance of market characteristics in forecasting upcoming equity market instability.

[29]To evaluate feature importance on a truncation-by-truncation basis, we recommend skipping this step.

7 Conclusion

The main objective of this thesis was to investigate the amplifying role of equity market characteristics in the build-up of systemic risk. To achieve this goal, we gather market and fundamental characteristics of STOXX 600 Index with a conceptual link to systemic risk. We then construct data that captures these characteristics as indicators and identify the equity component of systemic stress (CISS-E) as our equity market instability benchmark. Assuming the characteristics cause the effect that is increased systemic stress, we use a non-linear encoder-decoder model to forecast equity contribution to systemic stress purely based on the characteristics. Following data transformation and stationarity considerations, we consider 16 encoding and decoding sequence lengths to establish an optimal encoder-decoder model. The criteria for model selection was chosen to be: prediction performance on the test set containing the Covid-19 crisis period, and discrepancy between train and test set prediction performance to assess model behaviour during crisis and non-crisis periods. We determine that the optimal encoder-decoder model meeting the model selection criteria encodes 12 weeks of market and fundamental characteristics to decode the following 9 weeks of equity market instability. To illustrate the impact of this research, we perform a stress test and sensitivity analysis. We find that market and fundamental indicators have the ability to signal an increased level of equity market instability, 9 weeks in advance, particularly around an impending crisis. The sensitivity analysis reveals that market characteristics, namely volatility and market regime, have a higher relative importance in predicting equity market instability than fundamental characteristics. The forward-looking methodology on the European equity market buys time for European macroprudential supervision to limit the adverse socio-economic costs associated with equity market instability.

8 Discussion

Throughout this thesis, we opt for an ongoing evaluation approach. We consider the choices made in each chapter and justify the reasoning behind these choices. In this chapter, we opt for a systemic[30] view of the thesis. We illustrate contribution to existing literature, acknowledge structural limitations of this study and discuss avenues for further research.

Our study contributes to existing literature in three ways. First, it strengthens the knowledge on which equity characteristics are linked to systemic risk by validating the results of earlier studies. It highlights the amplifying role of market characteristics in periods of high systemic stress, and identifies key characteristics (market regime and volatility) responsible for this amplification. Second, it demonstrates the added value of machine learning (ML) to systemic risk and financial instability. It shows that ML models need not always be "black-box" by forecasting financial instability on grounds of strong economic foundations. It also highlights the promise of encoder-decoder models as their relevance for systemic risk has not yet been explored in literature. Third, by studying the equity market segment, it promotes the study of forward-looking approaches to protect financial stability of households.

While this thesis promotes the amplifying role of market characteristics when considering market instability, we acknowledge that the methodology can be improved. There are three main limitations of this study:

- First, we consider indirect causality between market characteristics and financial instability. For each characteristic, we create indicators on a daily basis, which are then used forecast market instability. Instead a more direct approach would be to use definitions of market characteristics, without the use of indicators, to forecast the CISS-E. The justification behind opting for indicators is that it adds greater dimensionality to the encoder-decoder model. This allows for more reliable results, in both training and testing samples, as we observe for the optimal model. To not compromise on reliability, we suggest future studies building on this research, to consider more several types of market characteristics with more granular data (i.e., hourly data as opposed to daily data).

- Second, the machine learning methodology can be tuned to reflect a more fair performance of the encoder-decoder model. For instance, we consider the loss function to be the MSE during the training phase. However, in the context of financial instability, it is far more important to predict high levels of systemic risk than lower levels of systemic risk. Put differently, the ideal loss function penalizes more harshly for underestimation of CISS-E than

[30]Pun intended.

for it's overestimation. The MAE does not consider imbalanced penalties. To address this limitation, we select data pre-processing considerations that are sensitive to high levels of systemic stress. It could be that the encoder-decoder model with attention, does perform better towards high levels of stress, but under-predicts stress during stable periods resulting in the higher test MAE errors observed in Table 4. For studies building on this work, we suggest using a custom loss function to reflect a more fair performance of the encoder-decoder model. We do advise caution though, as using a custom loss function means a subjective answer to "when is the optimization problem of the neural network fully solved?", and this answer needs to be appropriately justified.

- Third, the accuracy of the results might be limited as the CISS-E includes a market characteristic (volatility). Given the Neural Granger Causality setup we employ, the inclusion of volatility in CISS-E means that, to some extent, we perform an autoregressive forecast. Ideally, when we consider systemic risk, we would like to capture the effects of market instability on the broader economy. We would like to consider market characteristics and how they influence determinants of real economy. For studies building on this work, we suggest substituting equity market characteristics of the CISS-E with determinants of real economy that depend on households (for instance, household demand). This approach should establish a stronger connection between equity market characteristics and systemic risk.

References

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics*, *91*(1), 1–23. https://doi.org/10.1016/j.jfineco.2007.12.005

Astuty, P. (2017). *The Influence of Fundamental Factors and Systematic Risk to Stock Prices on Companies Listed in the Indonesian Stock Exchange* (tech. rep.).

Baele, L., Bekaert, G., & Inghelbrecht, K. (2009). The Determinants of Stock and Bond Return Comovements. https://doi.org/10.3386/W15260

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1409.0473v7

Balboa, M., López-Espinosa, G., & Rubia, A. (2015). Granger causality and systemic risk. *Finance Research Letters*, *15*, 49–58. https://doi.org/10.1016/j.frl.2015.08.003

Bavoso, V. (2018). Market-Based Finance, Debt and Systemic Risk: A Critique of the EU Capital Markets Union. *Accounting, Economics and Law*. https://doi.org/10.1515/ael-2017-0039

Bekaert, G., Ehrmann, M., Fratzscher, M., & Mehl, A. (2014). *The Global Crisis and Equity Market Contagion* (tech. rep.). http://www.diw.de/discussionpapers

Benoit, S., Colliard, J. E., Hurlin, C., & Pérignon, C. (2017). Where the risks lie: A survey on systemic risk. https://doi.org/10.1093/rof/rfw026

Buono, D., Elliott, D., Mazzi, G. L., Bikker, R., Frölich, M., Gatto, R., Guardalbascio, B., Hauf, S., Infante, E., Moauro, F., Oltmanns, E., Palate, J., Safr, K., Stoltze, P. T., & Iorio, F. D. (2018). ESS guidelines on temporal disaggregation, benchmarking and reconciliation 2018 edition. *Eurostat Manuals and Guidelines*. https://doi.org/10.2785/846595

Cabrera-Rodríguez, W. A., Hurtado-Guarín, J. L., Morales, M., & Rojas-Bohórquez, J. S. (2014). *A Composite Indicator of Systemic Stress (CISS) for Colombia* (tech. rep.). Banco de la República. Bogotá, Colombia. https://doi.org/10.32468/be.826

Calvo, G., & Reinhart, C. (2000). *Fear of Floating* (tech. rep.). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w7993

Campi, M., & Dueñas, M. (2017). *Volatility and Economic Growth in the Twentieth Century* (tech. rep.).

Caruana, J. (2010). Systemic risk: How to deal with it? *Bankarstvo*, *39*(7-8).

Chen, M., Wang, Y., Wu, B., & Huang, D. (2021). Dynamic analyses of contagion risk and module evolution on the sse a-shares market based on minimum information entropy. *Entropy*, *23*(4). https://doi.org/10.3390/e23040434

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1724–1734. https://doi.org/10.3115/v1/d14-1179

Corsi, F., Lillo, F., Pirino, D., & Trapin, L. (2018). Measuring the propagation of financial distress with Granger-causality tail risk networks. *Journal of Financial Stability*, *38*, 18–36. https://doi.org/10.1016/j.jfs.2018.06.003

Coudert, V., Gex, M., Coudert, V., & Gex, M. (2006). Can risk aversion indicators anticipate financial crises? *Financial Stability Review*, (9), 67–87. https://EconPapers.repec.org/RePEc:bfr:fisrev:2006:9:4

Danielsson, J., Valenzuela, M., & Zer, I. (2018). Learning from History: Volatility and financial crises. *Review of Financial Studies*, *31*(7), 2774–2805. https://doi.org/10.1093/rfs/hhy049

Danyliv, O., Bland, B., & Nicholass, D. (2014). Convenient Liquidity Measure for Financial Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2385914

Dapena, J. P. (2011). Volatility of GDP, Macro Applications and Policy Implications of Real Options for Structure of Capital Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.997143

Datacamp. (2020). Multiplicative vs additive seasonality. https://campus.datacamp.com/courses/arima-models-in-python/chapter-4-seasonal-arima-models?ex=12

Dell'Ariccia, G., Ferreira, C., Jenkinson, N., Laeven, L., Martin, A., Minoiu, C., & Popov, A. (2018). Managing the Sovereign-Bank Nexus. *Departmental Papers / Policy Papers*, *18*(16). https://doi.org/10.5089/9781484359624.087

Etesami, J., Habibnia, A., & Kiyavash, N. (2017). *Econometric Modeling of Systemic Risk: Going Beyond Pairwise Comparison and Allowing for Nonlinearity* (tech. rep.).

European Fund and Asset Management Association. (2020). *Household Participation in Capital Markets: Assessing The Current State and Measuring Future Progress* (tech. rep. September). European Fund and Asset Management Association. https://www.efama.org/Publications/KPI%20Repo

Gadanecz, B., & Jayaram, K. (2008). Measures of financial stability - a review.

Holló, D., Kremer, M., & Lo Duca, M. (2012). *cISS-A cOMPOSItE INdIcAtOR Of SyStEMIc StRESS IN thE fINANcIAl SyStEM* (tech. rep.). http://www.ecb.

Hudson, J. (2015). Consequences of aid volatility for macroeconomic management and aid effectiveness. *World Development*, *69*, 62–74. https://doi.org/10.1016/j.worlddev.2013.12.010

Hull, J. C. (2015). *Options, Futures, and Other Derivatives* (Ninth). http://ebooks.cambridge.org/ref/id/CBO9781107415324A009%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25246403%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4249520%5Cnhttp://link.springer.com/10.1007/978-1-4419-9230-7_2%5Cnhttp://dx.doi.org/10.

International Monetary Fund. (2006). Financial Soundness Indicators: Compilation Guide. http://www.imf.org

Jarrow, R., & Lamichhane, S. (2021). Asset price bubbles, market liquidity, and systemic risk. *Mathematics and Financial Economics*, *15*(1). https://doi.org/10.1007/s11579-019-00247-9

Kubitza, C., & Gründl, H. (2016). *Systemic Risk, Systematic Risk, and the Identification of Systemically Important Financial Institutions* (tech. rep.).

Kyle, A. S., & Xiong, W. (2001). Contagion as a wealth effect. *Journal of Finance*. https://doi.org/10.1111/0022-1082.00373

Langfield, S., & Pagano, M. (2016). Bank bias in Europe: Effects on systemic risk and growth. *Economic Policy*, *31*(85). https://doi.org/10.1093/epolic/eiv019

Li, X., & Zakamulin, V. (2020). Stock volatility predictability in bull and bear markets. *Quantitative Finance*, *20*(7), 1149–1167. https://doi.org/10.1080/14697688.2020.1725101

Louzis, D. P., Vouldis, A. T., & Metaxas, V. L. (2012). Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking and Finance*, *36*(4), 1012–1027. https://doi.org/10.1016/j.jbankfin.2011.10.012

LS, M., BK, H., CCP, E., AM, S., E, V., R, S., & RA, P. (2017). Food Consumption According to the Days of the Week - National Food Survey, 2008-2009. *Revista de saude publica*, *51*, 93. https://doi.org/10.11606/S1518-8787.2017051006053

Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 1412–1421. https://doi.org/10.18653/v1/d15-1166

Markowitz, H. (1952). *Portfolio Selection* (tech. rep. No. 1).

Markowitz, H. (1959). Portfolio Selection Efficient Diversification of Investments. *Journal of Financial Risk Management*. https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1482688

Mieg, H. A. (2020). Volatility as a Transmitter of Systemic Risk: Is there a Structural Risk in Finance? *Risk Analysis*. https://doi.org/10.1111/risa.13564

Minsky, H. P. (1982). The Financial-Instability Hypothesis: Capitalist Processes and the Behavior of the Economy. *Financial Crises: Theory, History, Policy*, *282*.

Murphy, J. J. (1999). John J Murphy - Technical Analysis Of The Financial Markets.pdf. *Pennsylvania Dental Journal*, *77*(2), 33–4. https://books.google.com/books/about/Technical_Analysis_of_the_Financial_Mark.html?id=5zhXEqdr_IcC

Parker, J. A. (2012). LEADS on Macroeconomic Risks to and from the Household Sector. https://doi.org/10.3386/W18510

Peltonen, T. A., Rancan, M., & Sarlin, P. (2019). Interconnectedness of the banking sector as a vulnerability to crises. *International Journal of Finance and Economics*, *24*(2). https://doi.org/10.1002/ijfe.1701

Sarkar, A., & Patel, S. A. (1998). Stock Market Crises in Developed and Emerging Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.76168

Sherstinsky, A. (2018). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena*, *404*. https://doi.org/10.1016/j.physd.2019.132306

STOXX Index Methodology Guide (Portfolio Based Indices). (2021).

Taleb, N. N., & Goldstein, D. G. (2012). The problem is beyond psychology: The real world is more random than regression analyses. https://doi.org/10.1016/j.ijforecast.2012.02.003

Venanzi, D. (2020). Systematic risk of European banks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3644863

Wang, X., Guan, X., Cao, J., Zhang, N., & Wu, H. (2020). Forecast Network-Wide Traffic States for Multiple Steps Ahead: A Deep Learning Approach Considering Dynamic Non-Local Spatial Correlation and Non-Stationary Temporal Dependency. *Transportation Research Part C: Emerging Technologies*, *119*. https://doi.org/10.1016/j.trc.2020.102763

A Appendices

A.1 Complete List of Indicators

Table 6: Complete list of indicators used. We present all the indicators, equity characteristic classes each indicator falls under, and their abbreviations. We also present constructor sources for each indicator, which offer instructions on how to recreate the indicators to replicate our study. Market-specific indicators can be reconstructed through Murphy (1999), A Handbook on Technical Analysis, while fundamental indicators can be reconstructed using Financial Statements of companies included within the STOXX 600. A positive (+) or a negative (-) sign is also given to summarize whether each indicator increases or decreases with increasing equity market stress, relevant for stress testing.

Characteristic	Indicator Name	Abbreviation	Source	Stress Behaviour
	Relative Strength Index	RSI		+
	Stochastic Oscillator	SO		+
	Price Rate of Change	ROC		+
Volatility	Awesome Oscillator	AWE	Murphy (1999)	+
	Average True Range	ATR		+
	Ultimate Oscillator	ULT		+
	Bollinger Bands	BB_down		+
		BB_width		+
	Moving Average Convergence Divergence	MACD		-
	Exponential Moving Average with Different Periods	EMA1		-
		EMA2		-
		EMA3		-
		EMA4		-
Market Regime		EMA5		-
	Simple Moving Average	SMA		-
	Triple Exponential Average	TRIX		-
	Detrended Price Oscillator	DPO		-
	Commodity Channel Index	CCI		-
	Average Directional Movement Index	ADX		-
	Vortex Indicator	VOR		-
	Parabolic Stop and Reverse Indicator	PSAR		-
	Accumulation Distribution Indicator	ADI		-
	On-Balance Volume	OBV	Murphy (1999)	-
	Force Index Indicator	FII		-
	Ease of Movement Indicator	EMV		-
Liquidity	Money Flow Index	MFI		-
	Value Weighted Average Price	VWAP		-
	Lagged EMA Indicators	Diff_EMA1		-
		Diff_EMA2		-
		Diff_EMA3		-
		Diff_EMA4		-
		Diff_EMA5	Murphy (1999)	-
	Average Candle Size	C-L		-
		H-L		-
	Long-Term Debt/Total Equity	LTD/TE		
	Long-Term Debt/Total Capital	LTD/TC		
	Long-Term Debt/Total Assets	LTD/TA		
	Total Debt/Total Assets	TD/TA		
	Total Debt/Total Equity	TD/TE		
	Net Debt/Total Capital	ND/TC		
	Total Debt/Total Capital	TD/TC		
	Net Debt/EBITDA	ND/EDA	Financial Statements of Firms Included in STOXX 600	Not Applicable
	Net Debt/(EBITDA-Capex)	ND/ECAP		
Leverage	Total Debt/EBITDA	TD/EDA		
	Total Debt/EBIT	TD/E		
	EBIT/Interest Expense	E/IE		
	EBITDA/Interest Expense	EDA/IE		
	CFO/Interest Expense	CFO/IE		
	Long-Term Debt/EBITDA	LTD/EDA		
	Net Debt/FFO	ND/FFO		

Table 6: Complete list of indicators used. We present all the indicators, equity characteristic classes each indicator falls under, and their abbreviations. We also present constructor sources for each indicator, which offer instructions on how to recreate the indicators to replicate our study. Market-specific indicators can be reconstructed through Murphy (1999), A Handbook on Technical Analysis, while fundamental indicators can be reconstructed using Financial Statements of companies included within the STOXX 600. A positive (+) or a negative (-) sign is also given to summarize whether each indicator increases or decreases with increasing equity market stress, relevant for stress testing.

Characteristic	Indicator Name	Abbreviation	Source	Stress Behaviour
	Long-Term Debt/FFO	LTD/FFO		
	FCF/Total Debt	FDF/TD		
	CFO/Total Debt	CFO/TD		
	Price/Earnings	P/E		
Profitability	Price/Sales	P/S		
	Price/Book Value	P/BV		
	Price/Cash Flow	P/CF		
	Enterprise Value/EBIT	EV/E		
Valuation	Enterprise Value/EBITDA	EV/EDA		
	Enterprise Value/Sales	EV/S		
	Gross Margin	GM	Financial Statements of Firms Included in STOXX 600	Not Applicable
	Operating Margin	OM		
	Net Margin	NM		
	EBIT Margin	EM		
Coverage	EBITDA Margin	EDAM		
	Return on Assets	ROA		
	Return on Equity	ROE		
	Return on Invested Capital	RIC		
	Free Cash Flow Margin	FCFM		

A.2 Model Plots

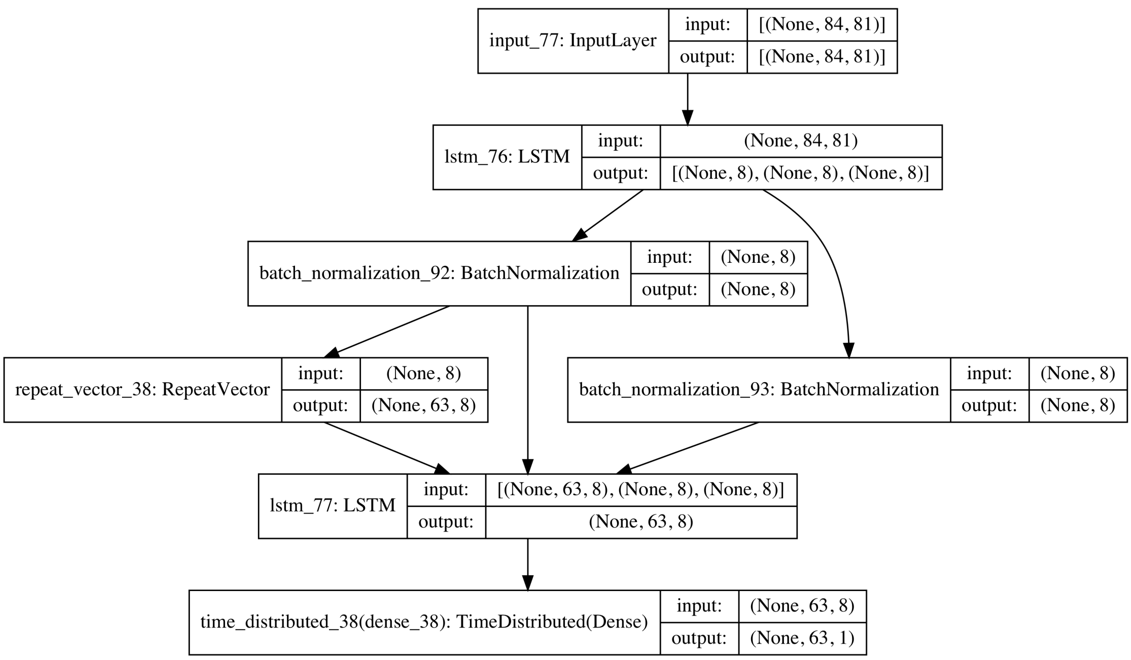

Figure 17: Plot of Baseline Optimal Model. Input and Output Sequence Lengths are 12 weeks and 9 weeks respectively (84 timesteps with 81 market and fundamental indicators predict 63 timesteps with 1 benchmark)

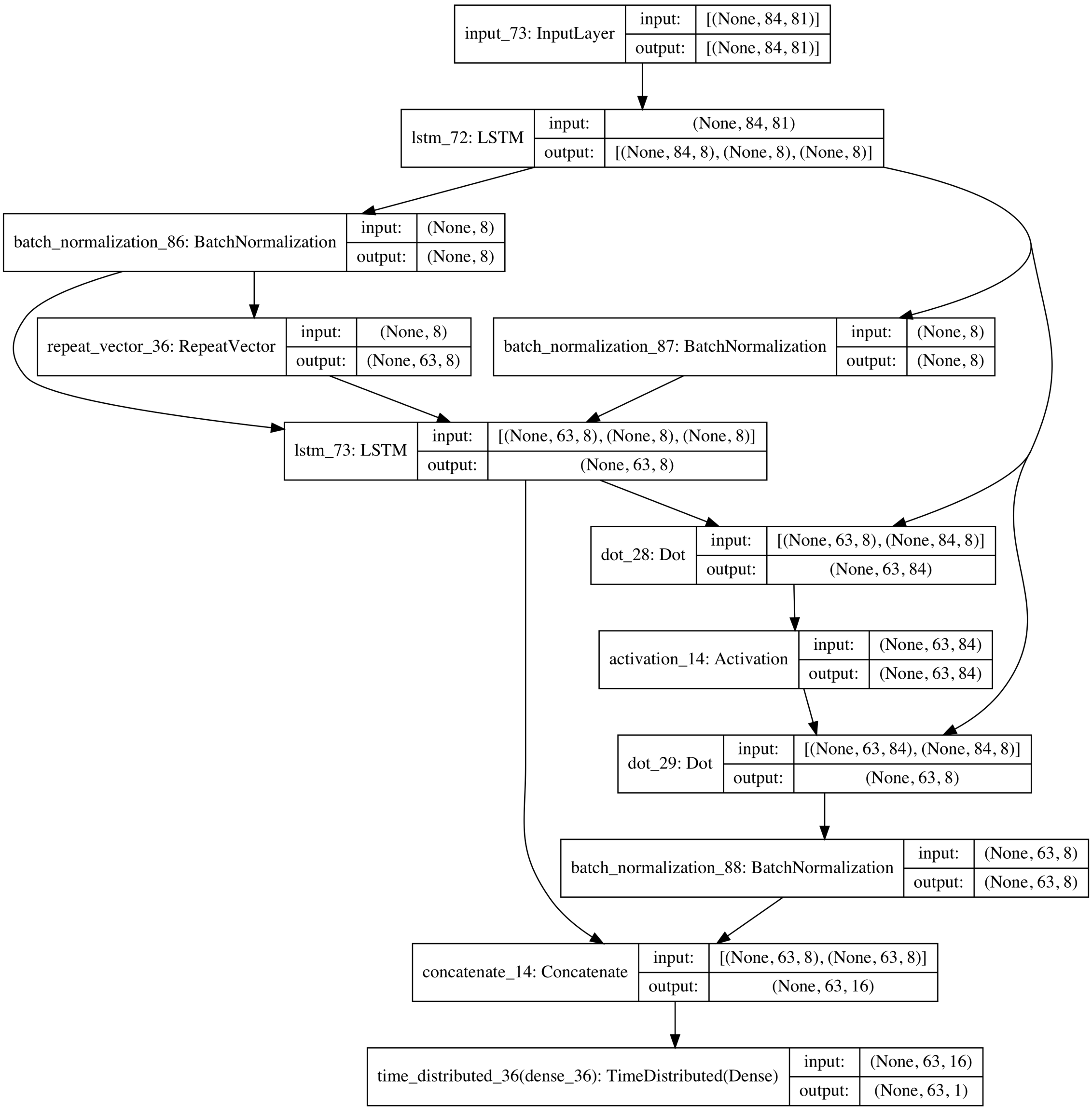

Figure 18: Plot of Optimal Model with Attention. Input and Output Sequence Lengths are 12 weeks and 9 weeks respectively (84 timesteps with 81 market and fundamental indicators predict 63 timesteps with 1 benchmark)

A.3 Time Series Decomposition

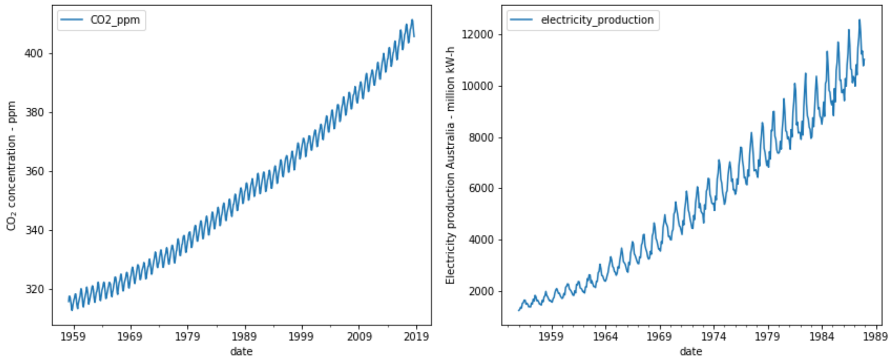

Figure 19: Additive Seasonality vs Multiplicative Seasonality. Source: Datacamp (2020).

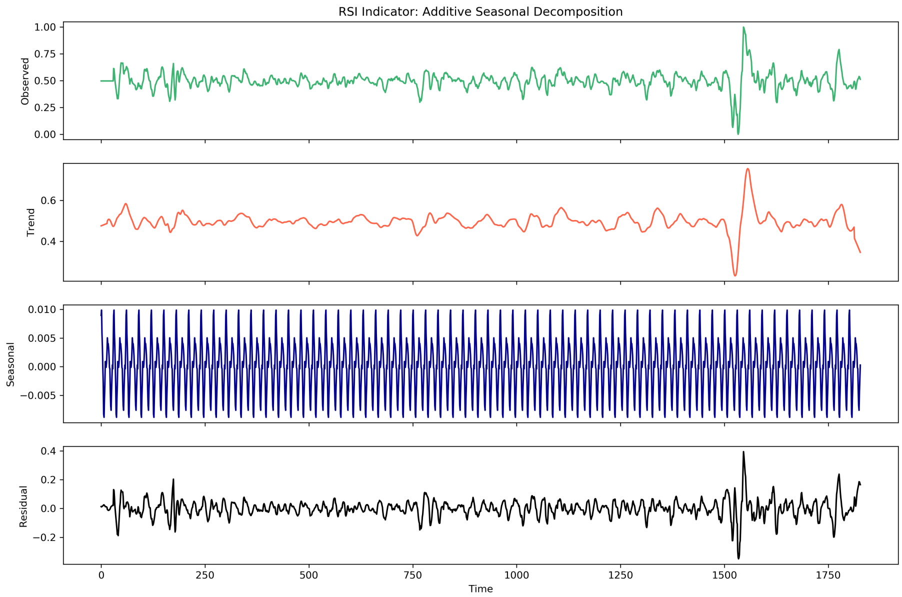

Figure 20: Sample Seasonal Decomposition Plot.