Universidade de São Paulo

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Departamento de Economia

Programa de Pós-graduação em Economia

Área: Economia Aplicada

Forecasting Sovereign CDS returns via Deep Learning

Pedro Antonio Sá Barreto de Lima

Supervisor: Prof Dr. Márcio Poletti Laurini

Ribeirão Preto

2021

Pedro Antonio Sá Barreto de Lima

Forecasting Sovereign CDS returns via Deep Learning

Dissertação apresentada ao Programa de Pós-Graduação em Economia – Área de Concentração: Economia Aplicada, da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências Econômicas.

Universidade of São Paulo – USP

Supervisor: Prof Dr. Márcio Poletti Laurini

Ribeirão Preto

2021

Ficha Catalográfica

Pedro Antonio Sá Barreto de Lima
Forecasting Sovereign CDS returns via Deep Learning – Ribeirão Preto, 2021
46 p. : il.; 30 cm.
Supervisor: Prof Dr. Márcio Poletti Laurini

Dissertação de Mestrado – Universidade of São Paulo – Faculdade de Economia, Admnistração e Contabilidade – FEA/USP – Campus Ribeirão Preto; Departamento de Economia Programa de Pós-Graduação em Economia; Área de Concentração: Economia Aplicada, 2021.

1. Deep Learning. 2. Sovereign Credit Default Swap. 3. Time Series Forecasting 4. Recurrent Neural Networks.
I. Orientador: Prof. Dr. Márcio Poletti Laurini.
II. Universidade of São Paulo – USP – Campus Ribeirão Preto. III. Faculdade de Economia, Administração e Contabilidade. IV. Forecasting Sovereign CDS returns via Deep Learning

"Presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001"

(Portaria Nº 206, de 04/09/2018)

Acknowledgements

To my advisor, Márcio Poletti Laurini, for all the knowledge transferred during the master's degree and for the excellent guidance.

To the members of the board for having accepted the invitation to participate.

To Alex Luiz Ferreira and Fábio Augusto Reis Gomes for their comments on the qualification and pre-defense stages.

To the professors who assisted in my training and the masters colleagues for the exchange of knowledge.

To my family for the encouragement during my studies.

To my girlfriend, for her patience and companionship during this journey.

To CNPq for financial support in the first year of the master's degree.

Abstract

LIMA, P. A. S. B. **Forecasting Sovereign CDS returns via Deep Learning.** Dissertation (Master Degree) - School of Economics, Business and Accounting at Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2021.

This paper is an empirical exercise of forecasting for a class of credit derivatives known as sovereign credit default swaps. Utilizing non-linear and non-parametric machine learning techniques termed Deep Learning, this study was made utilizing daily emerging country data and a set of financial and macroeconomic indicators as features. The non-linear nature of the financial derivative studied here suggests that this novel technique can better capture data behavior compared to a baseline random walk model. A Grid Search Cross-validation is conducted to estimate the hyperparameters of the model. To evaluate the predictive out of sample forecasting is utilized deterministic and statistical metrics concluding that there is a predictive gain utilizing this technique.

Key-words: 1. Deep Learning. 2. Sovereign Credit Default Swap. 3. Time Series Forecasting 4.Recurrent Neural Networks.

Resumo

LIMA, P. A. S. B. **Previsão de retornos de CDS soberanos por meio de Aprendizagem Profunda.** Dissertação (Mestrado) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2021.

Este artigo é um exercício empírico de previsão para uma classe de derivativos de crédito conhecida como *credit default swap* soberano. Utilizando técnicas de aprendizado de máquina não lineares e não paramétricas denominadas Redes Neurais Profundas, este estudo foi feito usando dados diários de países emergentes e um conjunto de indicadores financeiros e macroeconômicos como variáveis preditivas. A natureza não linear do derivado financeiro estudado aqui sugere que esta técnica pode capturar melhor o comportamento dos dados em comparação com um modelo de passeio aleatório, que foi utilizado para ser base comparativa. Uma validação cruzada utilizando *grid search* é conduzida para estimar os hiperparâmetros do modelo. Para avaliar a previsão fora da amostra, são utilizadas métricas determinísticas e estatísticas, concluindo que há um ganho preditivo usando esta técnica.

Palavras-chave: 1. Redes Neurais Profundas . 2. Credit Default Swap Soberano. 3. Previsão para Série de Tempo 4.Redes Neurais Recorrentes.

List of Figures

List of Tables

Contents

1 Introduction

This paper forecasts sovereign Credit Default Swap returns using state-of-the-art statistical learning techniques called Deep Learning, which have common usage for image recognition, language processing, and other computer science applications, Goodfellow et al. (2016). The non-linear nature of the financial derivative that this research explores suggests that these novel techniques introduce a novel way to capture regularities of Sovereign CDS spreads as spillover effects, dependence on global financial markets indicators, and country-specific risk factors. This paper's results corroborate this hypothesis providing a direct one-step ahead forecasting exercise for 7 different emerging market countries with a set of common features. Analyzing point forecasting metrics, such as MAE and RMSE, the deep neural network shows an advantage to the baseline random walk model. To further assess the statistical significance of this forecast, it is shown an analysis utilizing the Model Confidence Set by Hansen et al. (2011) providing evidence of the statistical difference between the benchmark model.

Credit derivatives permit firms to trade credit risk the same way that they negotiate market risk. This paper investigates a specific credit derivative called Credit Default Swap (CDS). CDS is a contract that offers insurance against a default risk from the borrower institution and the lender. The event that set off the contract is called *"credit event"*, Lando (2020). A crucial divergence between corporate and sovereign CDS contracts is related to the nature of the classification of *"credit events"* that triggers the contingent default insurance payment, as explained by Duffie and Singleton (2012). Another critical difference presented in these contracts is the currency that the contract is designated. There is a risk of depreciation of the said currency in the event of default, and this risk is embedded in the price on the same underlying sovereign asset with different price quotes, as shown by Carr and Wu (2007).

In this work, we approach the forecast problem for CDS utilizing a Statistical Learning framework. Statistical Learning methods in general are interested in understanding the conditional distribution of some variable y give some other variables $x = (x_1, ..., x_p)$, as described by Hastie et al. (2009). The approach can be summarized as a search of space of possible functions that combine the data and form the prediction and search for ways to introduce a flexible approach to proper approximate complex DGP forms and control for overfitting risk. Coulombe et al. (2020) has shown that non-linearities improve considerably predictions of macroeconomic features in a data-rich environment.

The remainder of the work is structured as follows. Section 2 presents the Credit Default Swap and the central literature of machine learning application in finance and how it is employed. Section 3 describes the analyzed dataset. In Section 4 the methodology is presented, Section 6 the estimation results, and in the last section, the results conclusions.

2 Literature Review

Credit Default Swap (CDS) is a financial product of the credit derivatives class designed to reallocate credit exposure between two agents from an underlying fixed income product. A CDS works as an insurance against non-payment, becoming an attractive alternative for investors trying to hedge against current exposure or speculators, with the advantage of not maintaining the reference obligation in their portfolio. Another point researched by Oehmke and Zawadowski (2016) shed a light on the motivations and positions in CDS markets. Their research shows that the CDS market serves as a liquidity role for firms with bonds fragmented into many separate issues and also points out cross-market arbitrage links between the two markets.

In the case of sovereign CDS, monitoring their spread can be an acceptable way to perceive market sentiment of the implicit probability of default in the government's debt obligation. The magnitude of this market, as reported by the Bank of International Settlements (BIS) in Aldasoro and Ehlers (2018), was around 3.3 trillion in mid-2013. The size and information it provides about the market perception of sovereign countries' debt obligations is a motivation to work through models to predict more adequately the spreading behavior. That can be perceived not only by the value it has for risk management in financial companies but also for monitoring the spread evolution for macroprudential policy-making purposes.

This literature review is divided into two sections. The papers presented in the first section focus mostly on evaluating the price dynamics of this financial product; others focus on the out-of-sample forecast of the CDS spread. In the second section, we explore the methodology advances in finance using deep learning.

2.1 Sovereign CDS in Financial Literature

As the corporate equivalent of the standard CDS contract, the Sovereign CDS contract is structured along these lines: the insurance for default acquired by the buyer requires a semi-annual premium payment, formulated in basis points per notional amount of the contract, in the alternative for a payment in the event of a pre-specified credit event. A credit event for this derivative is generally characterized as (i) obligation acceleration, (ii) failure to pay, (iii) restructuring, or (iv) moratorium[1]. Because of the nature of this financial product, they are traded over-the-counter, as pointed out by Augustin et al. (2014).

Because of the nature of this derivative and its market size, understanding and analyzing sovereign CDS spreads are the main focus of the literature. Badaoui et al. (2013) put into question if the drivers of the CDS spreads are caused by intrinsic characteristics by the countries bonds that they are based on. Their analysis shows evidence that liquidity constraints are the main key factor of sovereign CDS spreads compared to bonds, decomposing these effects with factor models. This is particularly important for the choice of a proxy for sovereign default risk.

Pan and Singleton (2008) used a combination of CDS contracts at several maturity points to construct a term structure of CDS spreads and delve into the essence of the risk-neutral default arrival and recovery rates implicit in the term structure. There is a theoretical gain in modeling sovereign credit risk in their work exploiting the term structure of CDS spreads, permitting to separately identify both risk loss rate used for the recovery rate and the parameters that define default arrival, capturing different credit events. They show that a single-factor model of the arrival rate following a *lognormal* process captures most of the variations in the term structures of spreads. The unpredictable variation is economically significant with several global event risks, market volatility, and macroeconomic policy.

A major study performed by Longstaff et al. (2011) suggests that most of the variations in sovereign CDS spreads are due to common regional and global elements. This commonality of factors that different associate countries support the dominant role of global investors in the market. The authors also find no evidence of a separate risk premium in sovereign debt markets than fixed income markets and stocks. Fender et al. (2012) uses daily CDS spread data and uses stochastic volatility models, to investigate

[1]Often, trigger events for sovereign entities results in a restructuring or rescheduling of debt.

a group of emerging market countries. Their results indicates that daily sovereign CDS spreads for emerging market sovereigns are closely associated to global and regional international investors risk appetite than country-specific risk factors, aligned with Longstaff et al. (2011). There is a remarkable difference in CDS spread determinants depending on financial market conditions since they discriminate against the dynamics before and after the 2008 financial crisis. During the financial crisis, country-specific factors no more present a relevant part.

Another vital factor in explaining the changes in the spread in countries with a production chain linked to a specific commodity, such as oil, is the price volatility of these commodities. This results in adjustments in market expectations regarding public accounts' financial stability in these countries, as demonstrated by Wegener et al. (2016), which are linked to price shocks. Focusing on emerging markets, Bouri et al. (2018) study the dependence between oil volatility shocks and BRICS sovereign risk revealing an imbalance in the structure of shock transmissions between oil exporter and importers. The previous is more easily affected by positive oil shocks, though the last mentioned is more disadvantageous the shocks.

Ammer and Cai (2011) evaluate CDS and bond pricing dynamics, finding the link amidst sovereign CDS premiums and bond yield spreads for emerging markets. In the context of the Eurozone debt crisis, Kalbaska and Gatkowski (2012) use CDS data to investigate evidence of financial contagion in EU countries during 2005-2010, using Exponentially Weighted Moving Average (EWMA) methodology and Granger-Causality tests. Broto and Pérez-Quirós (2014) take a different approach and fit a dynamic factor model to decompose the sovereign CDS spreads of OECD countries to confirm the existence of clear financial linkages between those economies.

A paper by Arce et al. (2013), looks to what extent sovereign CDS and bond markets indicate the identical information on their prices in the European Monetary Union. The authors find that deviations between these prices are related to counter-party risk, volatility inside the EMU equity markets, QE policy by the ECB, and flight-to-quality. In an attempt to shed light on the predictability of sovereign CDS, Srivastava et al.

(2016) use an error correction model and identify strong evidence of Granger causality in mean from the S&P option market to the sovereign CDS market.

The Sovereign CDS empirical literature focuses on reduced-form latent models for modeling credit risk. Studies capture macroeconomic fundamentals and spread determinants guided by structural models. Others use conditional volatility models to capture volatility clustering behavior in the series in a combination of factor models. The information provided from this literature, about the drivers of this security, serves as a guide for choosing essential features for our predictive model.

2.2 Deep Learning in Finance

The discussion of the inclusion of neural networks into the toolkit used by econometricians and economists in general started with the work of Kuan and White (1994), developing a theoretical framework compatible with the modern theory of estimation and inference. Applications of this methodology is date back to Kaastra and Boyd (1995) using these models to forecast monthly futures trading volume and have mixed results of the forecast accuracy compared to the ARIMA model for different commodities. In macroeconomics, neural networks provides better forecast inflation than traditional time series econometric models, according to Moshiri and Cameron (2000).

Besides those incursions in the economic literature, this kind of modeling lost popularity because of technical limitations at the time. Moreover, In the last decade, a resurgence occurred for a range of factors: advances in algorithmic development that turned neural networks more accessible for practitioners, increasing computational efficiency that made model training achievable, as explained by Cook (2020). Especially within the last few years, these methods have seen an increasing adoption as an noteworthy machine learning model.

In finance, a wide range of machine learning methods is being used to tackle different problems. Huang et al. (2004) presents an extensive survey of corporate credit rating analysis using neural networks and their performance compared to traditional methods. Khandani et al. (2010) apply generalized classification and regression trees to build a credit-risk forecast model that improves prediction of default behavior for individual consumers, offering new insight for systemic risk management in the credit sector. These advanced techniques can offer new insights in portfolio management as a result of the capability to exploit hidden and non-linear data patterns and interaction unseen to any existing financial or economic theory, as stated in Heaton et al. (2017).

Deep Learning, besides been used successfully in different applications in computer science, literature for time series forecasting is still limited, Gamboa (2017) summarise the recent development in the area. The literature on the use of deep learning in finance is still growing, Dixon et al. (2019) works by applying these novel techniques to Spatio-temporal applications. Gu et al. (2019) presents a new factor conditional asset pricing model exploring flexible non-linear functions, using an autoencoder model. For empirical asset pricing, Gu et al. (2018) use a wide range of Machine Learning models to contribute to a new benchmark in risk premia measures of the individual and aggregate market and show that machine learning can expand our knowledge of expected asset returns. However, it is essential to note that, according to the author, "shallow" learning outperforms "deep learning" in this paper. On the other hand, Chen et al. (2019) design a deep neural network estimation for a general non-linear asset pricing model with no-arbitrage constraints that surpass benchmark approaches measured by the Sharpe ratio explained variation and pricing errors.

Bianchi et al. (2019) shows that non-linear methods used in machine learning are highly useful for the prediction of bond excess returns compared to traditional data compression techniques akin to principal component analysis. Aligned with the evidence brought by the previous authors we argue that Deep Learning models can possibly capture non-linear dependencies across a group of assets that are highly correlated and share a large group of covariates as a common factor to explain their co-movements. The flexibility of a non-parametric and purely data driven form of these models in comparison with traditional structural or parametric modelling is the main advantage of this approach, as pointed out by Chen et al. (2019).

Our empirical strategy has the objective of extending the existing literature on the use of Deep learning in finance. My contribution will come from using state-of-the-art Neural Network algorithms and testing the efficiency of this approach.

3 Dataset

3.1 Sovereign CDS

CDS contracts are not traded on an organized exchange and the guidance of the institutional details of the CDS's are given by the International Swaps and Derivatives Association (ISDA). The theoretical use of a CDS is an insurance contract that hedges the owner against the risk of default of a reference entity, Sovereign or Corporate. In the widest definition, is a two-sided agreement where the exposure to risk party pays a periodic fixed premium, most commonly in quarterly periods. The payment agreement between a CDS seller and a buyer is expressed in basis points of the notional value of the contract and is designated as CDS spread. Most features of the Sovereign CDS market are quite similar to their corporate counterpart, with the exception that information asymmetries in the Sovereign market may be minor since the health of a country's economic situation and public finances is common knowledge. Hence public information should be reflected in prices.

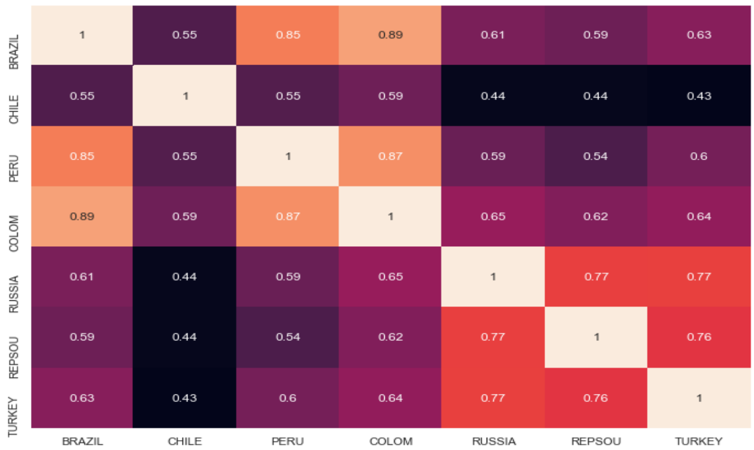
The 5-year maturity Sovereign CDS data used in this research are provided by Bloomberg, which considers different industry sources to aggregates market quotation data. The sample period in the present study extends from January 2007 to December 2019, at a daily traded frequency. The focus of this analysis is 5-year maturity contracts. As reported by Hébert and Schreger (2017), there are problems of inconsistency between Bloomberg data and Markit (a traditional financial service company that provides CDS data). The main issue is that CDS spreads other than five years appear to be unreliable. Last, we will be using spread perceptual changes as the target variable.[1]

The interconnections between markets are revealed during a crisis when the transmission of shocks from one market to another is amplified by extreme volatility. Using the framework of Duffie and Singleton (2012), systemic credit risks are linked with a common vulnerability to an adverse shock. As shown by Uribe and Schmitt-Grohé

[1] Also labeled as "returns", calculated using the log difference.

(2017) emerging markets have a historic background of default and debt crisis, and the effects of these events[2] are shared between them, as shown by Remolona et al. (2008). Hence, we focus on a subset of emerging countries in which spread changes are observed in our chosen period. This results in a sample of 7 emerging market borrowers: Brazil, South Africa, Colombia, Chile, Peru, Russia, and Turkey. Our Sovereign CDS sample is highly correlated, as presented by the table 1. This fundamental statistic gives us the idea that these contracts could share common explanatory elements of their behavior.

Figure 1 – Sovereign CDS Correlations



Source: Bloomberg. Author's elaboration

The data presents a lot of commonalities. Across all countries, they share almost the same quantiles, with minor differences between them. Chile presents the most extreme variations in our set of countries. During the 2008 financial crisis, it's noticeable the extreme changes in CDS returns in our dataset, as seen in table 2. It is important to notice that we treated gaps in our data utilizing Kalman imputation. We set a local linear trend model in state-space form, as described by Durbin and Koopman (2012), and estimate the state vector given the entire sample, replacing the missing values by the model prediction. This entire process is called the Kalman Smoother.

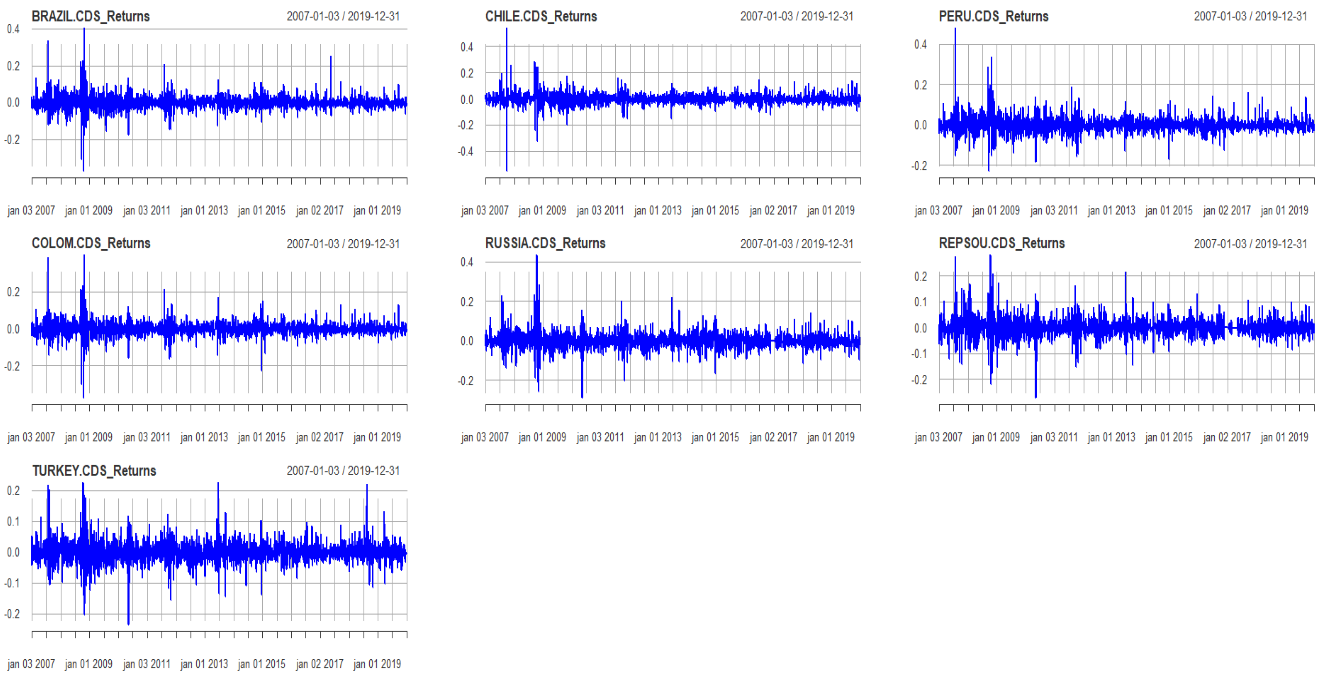
[2]Latin American Debt crisis in 1982, the Mexican Crisis in 1994-1995, the East Asian Crises in 1997, and Russian crisis in 1998

Table 1 – Summary Statistics for Sovereign CDS Data

CDS	N	Mean	St. Dev.	Min	25%	75%	Max
Brazil	3,390	−0.00003	0.035	−0.373	−0.016	0.014	0.405
Chile	3,390	0.0003	0.037	−0.550	−0.016	0.015	0.541
Peru	3,390	−0.0001	0.035	−0.228	−0.016	0.014	0.480
Colombia	3,390	−0.0001	0.034	−0.373	−0.016	0.014	0.398
Russia	3,390	0.0002	0.037	−0.289	−0.017	0.015	0.433
South Africa	3,390	0.0005	0.033	−0.272	−0.014	0.014	0.281
Turkey	3,390	0.0001	0.031	−0.236	−0.015	0.015	0.227

Summary Statistics for Sovereign CDS Data from January of 2007 to December 2019

Figure 2 – Sovereign CDS Returns



Sovereign CDS Data Returns from January of 2007 to December 2019

3.2 Predictors

Prior research substantiates the belief that global elements have a greater explanatory power of CDS spreads changes over country-specific risk, which appears to be almost negligible. These results appear to be stronger when we take into consideration only emerging markets (EMs), which are appeared to be more inclined to worldwide and regional risk premia compared to macroeconomic specific components[3]. This pattern is especially evident during events that generate market stress as an economic crisis. Shocks to financial market volatility and agent's risk aversion effects are more significant than macroeconomic fundamental variables to explain differences between bond spreads of EMs, according to Ciarlone et al. (2009).

Based on this, we constructed a set of predictors, using Bloomberg as a data source. We refer to the level of overall liquid stress of bond and equity markets in US and EU as a measure of global liquidity conditions, Eichengreen and Mody (1998), using Bloomberg Financial Conditions Index (BFCI); The volatility of equities markets represented by VIX is used as a proxy for investors' appetite for risk; US stock market SP 500 index are indicative of market risk; JP Morgan Bond Spreads indexes are used to monitor emerging market debt prices; Libor data is used as a proxy for gauge stress in the financial system.

Most EMs are dependent on commodities as export revenues, hence volatility of commodity prices reflects on the volatility of Sovereign CDS, as shown by Bouri et al. (2017) and the same assumption holds for Oil revenue dependent countries, Bouri et al. (2018). To capture this commodity-specific risk we used Metal, Energy, and Oil prices. We use Citi Group Economic Surprise Index (CESI) as a measure of how economic data is aligned with market expectations. The CESIs is defined as the weighted historical standard deviation of "data surprises" (observed data points versus Bloomberg median expectations) and are calculated daily in a rolling three-month window, providing an aggregation of market misalignment of market expectations and reality.

[3] Longstaff et al. (2011), Fender et al. (2012)

Table 2 – Summary Statistics of Predictors

Statistic	N	Mean	St. Dev.	Min	25%	75%	Max
CESI EUR Index	3,390	−2.679	54.489	−189	−34.7	35.6	149
CESI EM Index	3,390	0.073	28.047	−129.500	−13.800	18.200	65.000
CESI JPY Index	3,390	−1.121	34.550	−96.700	−25.875	22.700	106.500
BFCI EU Index	3,390	−0.851	1.766	−10.833	−1.276	0.146	1.250
BFCI US Index	3,390	−0.549	1.774	−12.564	−0.835	0.495	1.374
VIX Index	3,390	19.258	9.126	9.140	13.402	22.072	80.860
SPX Index	3,390	0.545	17.247	−113.190	−5.880	8.602	104.130
CMDINRGS Index	3,390	−0.018	7.214	−81.280	−3.438	3.600	63.510
CMDIBASS Index	3,390	−0.038	2.662	−12.810	−1.390	1.400	14.230
JPEIPLUS Index	3,390	0.168	2.363	−26.830	−0.810	1.227	13.970
JPEIGLSP Index	3,390	0.019	7.637	−61.130	−3.450	3.130	101.680
JPEIDISP Index	3,390	0.022	7.337	−57.137	−3.310	3.032	97.830
JPEIPLSP Index	3,390	0.022	8.160	−63.850	−3.640	3.378	113.660
CO1 Comdty	3,390	−0.005	1.510	−10.390	−0.750	0.800	10.150
CL1 Comdty	3,390	−0.0001	1.541	−14.310	−0.750	0.780	16.370
EUR003M Index	3,390	−0.001	0.011	−0.118	−0.002	0.001	0.101
US0003M Index	3,390	−0.001	0.023	−0	−0.002	0.002	0

This table contains the summary statistics of predictors collected from January of 2007 to December 2019. Description of the Bloomberg tickets see table 7

4 Methodology

Deep Learning Architectures

We will follow the notation and description presented in Cook (2020) to describe the details of the models used to estimate our results.

4.1 Multilayer Perceptron Models

As defined by Goodfellow et al. (2016), multilayer perceptrons models are the classic deep learning model. The objective of a MLP is to resemble itself to any function $y = f(x)$. This property is determined by the theorem stated by Hornik (1991). A brief explanation of what the theorem states, done by Cook (2020), is that "for any function $f : \mathbb{R}^m \mapsto \mathbb{R}^n$, there exists a neural network with one hidden layer, G, that can approximate f to an arbitrary level of accuracy[1]" (162). Although there are other algorithms with similar properties, neural networks have the advantage of not requiring a restrictive hypothesis, in most cases generalize well, and scale with an increase of the input space.

A multilayer perceptron model is the simplest type of neural network, consisting of multiple interdependent layers of neurons that transforms inputs into target output. The Neural Network that is fully connected contains the so-called *neurons*, which are linear combinations of inputs, adding a constant term, which is transformed through an activation function[2].

$$f\left(\mathbf{x}\boldsymbol{\beta} + \alpha\right)$$

where $\mathbf{x}$ is an n-lenght vector of inputs and $\boldsymbol{\beta}$ are the corresponding weights, and α is a scalar intercept, also known in the literature as 'bias' term.

Neurons are stacked into layers; in this case we will be using a dense layer.

[1] $|f(\mathbf{x}) - G(\mathbf{x})| < \epsilon$

[2] In this work we will use two MLP models with different activation functions:*ReLU(x)* $=$ $max(x, 0)$ and Hyperbolic Tangent: $\frac{e^{2x}-1}{e^{2x}+1}$

Defining $\mathbf{B} = (\beta_1, ..., \beta_p)$ where p equals the number of neurons, suppose that $\mathbf{B}_{(n \times p)}$ and $\boldsymbol{\alpha} = (\alpha_1 ... \alpha_p)$, where the matrix $\mathbf{B}$ sets the weights for each matching input and neurons, and $\boldsymbol{\alpha}$ provides the bias for the latter. Given an n-length input vector $\mathbf{x}$, writing a dense layer with p neurons as,

$$g(\mathbf{x}) = f\left(\mathbf{x}\mathbf{B} + \boldsymbol{\alpha}\right)$$

$$= \begin{bmatrix} f\left(\boldsymbol{x}\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1\right) \\ f\left(\boldsymbol{x}\boldsymbol{\beta}_2 + \boldsymbol{\alpha}_2\right) \\ \vdots \\ f\left(\boldsymbol{x}\boldsymbol{\beta}_p + \boldsymbol{\alpha}_p\right) \end{bmatrix}^T$$

The work focus on higher-order inputs, such as a $m \times n$ matrix of several observations, $\mathbf{X} = (\mathbf{x_1}, ... \mathbf{x_m})'$.

$$f\left(\boldsymbol{X}\boldsymbol{B} + \boldsymbol{\alpha}\right)$$

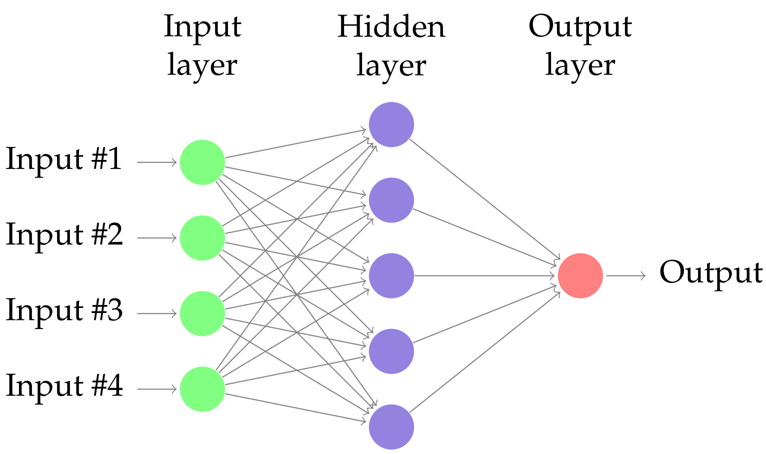
$$= \begin{bmatrix} f\left(x_1\beta_1 + \alpha_1\right) & \dots & f\left(x_1\beta_p + \alpha_p\right) \\ \vdots & \ddots & \vdots \\ f\left(x_m\beta_1 + \alpha_1\right) & \dots & f\left(x_m\beta_p + \alpha_p\right) \end{bmatrix}$$

An MLP that is fully connected with K layers is shaped by grouping dense layers, so that the yield of the previous layer is utilized as an input for the present layer. Let's define k as an index for any layer, then the output of the *k-th* layer is:

$$\begin{aligned} g_k\left(\boldsymbol{X}\right) &= f_k\left(g_{k-1}(\boldsymbol{X})\boldsymbol{B}_k + \boldsymbol{\alpha}_k\right) \\ g_0\left(\boldsymbol{X}\right) &= \boldsymbol{X} \end{aligned}$$

The parameters of the network are all elements $\boldsymbol{B}_k, \boldsymbol{\alpha}_k$ for $k \in (1, ..., K)$. To simplify, define $\theta_k = (\boldsymbol{B_k}, \boldsymbol{\alpha_k})$ and define the last output of the network as $G(\boldsymbol{X}; \boldsymbol{\theta}) = g_K(\boldsymbol{X})$. To be more precise, we are working with a supervised learning problem, where our target is known, $\boldsymbol{Y}$, estimated as $\boldsymbol{\hat{Y}} = \boldsymbol{G}(\mathbf{X}; \boldsymbol{\theta})$. Summarizing the difference amidst the target and the estimation, we define a loss function $\mathcal{L}(\boldsymbol{Y}, \boldsymbol{\hat{Y}})$, in this case the function is $||\boldsymbol{Y} - \boldsymbol{\hat{Y}}||^2$.

Figure 3 – An example of MLP with a Single Hidden Layer



Regularization

One of many problems present in machine learning models is the overfitting into the training set and failing to generalize the model to out of sample data observations. Deep Learning in particular have a great number of parameters, giving the ability to fit different complex data sets. However, all this liberty makes the network prone to overfitting the training set. To address this problem, we introduce a sparsity inclusion to our MLP architecture. Give the proven success and popularity we chosen the *Dropout* technique, Srivastava et al. (2014).

The Dropout algorithm is fairly simple, at every training step, every input neuron has a probability of being "dropped out" determined by $\mathbf{r} \sim \text{Bernoulli}(\pi)$, where $\mathbf{r}$ has a *p-length*. Meaning it will be neglected through this training step, even tho active during the following one. Expressed mathematically:

$$g_k(\boldsymbol{X}) = f_k(\mathbf{r} \odot g_{k-1}(\boldsymbol{X})\boldsymbol{B}_k + \boldsymbol{\alpha}_k)$$

This procedure accomplishes basic two things, according to Cook (2020). First, lessens overfitting by inhibiting high correlated updates of connected neurons. This happens when a neuron updates to compensate for the output of a connected neuron, hence fitting particularities in the data and increasing the chances of overfitting. Dropout curbs the neuron in that layer to turn into exceedingly dependent on the output from any particular neuron in the previous layer. Second, dropout allow us to test a large number of models at once. After all, dropout removes temporarily the neurons from the network, hence to train a network with this regularization technique approximates a model averaging over many sparse networks.

4.2 Recurrent Neural Network

Recurrent Neural Networks (RNNs), initially proposed by Rumelhart et al. (1986), are sequence learner's methods which achieved success in different sets of problems such as natural language processing and speech propagation. RNNs apply an autoregressive function to each input sequence and not impose an autocovariance structure, providing a flexible functional form. Unlike a fully connected network presented before, it forces an arrangement on its inputs and recognize them as a sequence. Writing a basic RNN model with the output of any layer stated as:

$$\begin{aligned} g_t\left(\boldsymbol{x}\right) &= f\left(x_t \boldsymbol{B}_x + g_{t-1}(\boldsymbol{x})\boldsymbol{B}_g\right) \\ g_0\left(\boldsymbol{x}\right) &= \mathbf{0} \end{aligned}$$

where $\mathbf{B}_{x(1\times p)}$ and $\mathbf{B}_{g(p\times p)}$ are matrices of weights, and the resulting $g_t(\mathbf{x})$ is a p-length vector, and $\mathbf{x} = (x_1, x_2, \ldots, x_T)$. All layers divide the same set of weights. Additionally, at the same time each layer $g_t(\mathbf{x})$ takes input from the previous one, they also take outer input from the t-th element in $\mathbf{x}$. Since individual layers include each a new input and considering as an input to the consecutive layer each layers output, $g_t(\mathbf{x})$ is the state of the model at a determined t.

We determined the sequence length needed in a RNN by testing the autocorrelation of the data and using the largest significant lag in an estimated partial autocorrelation function, as suggested by Dixon (2020). The problem is that, even tho traditional RNNs include some time-series behavior, it can suffer from vanishing gradients, according to Bengio et al. (1994). Taking into account this hindrance, it is needed to implement a RNN with Long-Short-Term-Memory (LSTM).

4.2.1 LSTM

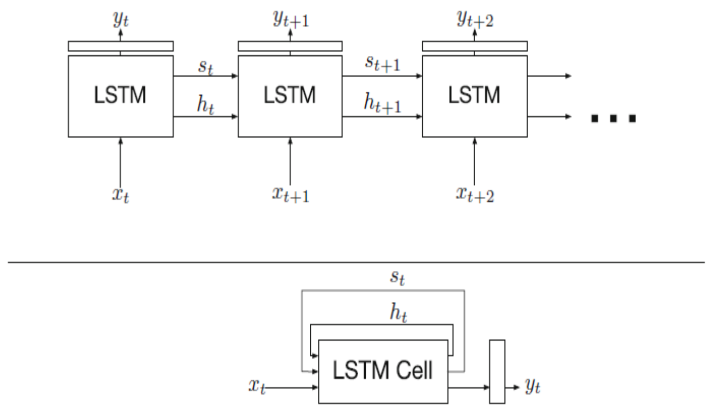
The LSTM proposed by Schmidhuber and Hochreiter (1997) is arranged to find latent variables, incorporating past information of the variable as information. This architecture is one of the most successful commercial AIs for sequential data such as speech recognition, Yu and Deng (2016). As described by Cook (2020), this architecture is a group of expressions that considers the current input; the generated output of the network at the last step, and the last latent process of the network, $\boldsymbol{h}_{t-1}$ and s_{t-1}. The latter carries long-run information, produces $\boldsymbol{h}_t$, and updates the latent variable s_t. Defining $\boldsymbol{Z}_t = [x_t, \boldsymbol{h}_{t-1}]$, and representing the inverse logit function as σ:

$$s_t = \overbrace{\underbrace{s_{t-1}}_{\text{old state}} \odot \underbrace{\sigma\left(Z_t \boldsymbol{B}_d\right)}_{\text{delete selection}}}^{\text{forget}} + \overbrace{\underbrace{\sigma\left(\boldsymbol{Z}_t \boldsymbol{B}_i\right)}_{\text{modification selection}} \odot \underbrace{\tanh\left(\boldsymbol{Z}_t \boldsymbol{B}_c\right)}_{\text{modification magnitude}}}^{\text{modify}}$$

$$\boldsymbol{h}_t = \sigma\left(\boldsymbol{Z}_t \boldsymbol{B}_o\right) \odot \tanh\left(s_t\right)$$

Equation number one updates the LSTM cell memory and it's made of two parts: a forget step, that selects which elements of the cell's memory are to be removed, and a modification one, that determines which shares of the latent process should be changed and the amount of modification. The final expression, $\boldsymbol{h}_t$, is a portrayal of the cell latent process permeated through an output gate based on the present input and last cell output. The use of the particular activation functions in this network fulfills a purpose to preserve the scale of values in the latent and cell output, preventing vanishing gradients or exploding gradients.

Figure 4 – LSTM Architecture



Source: Cook (2020)

4.2.2 GRU

A simpler version of LSTM models and an addendum of RNNs to capture dynamical time-series components is the Gated Recurrent unit (GRU), Cho et al. (2014). It presents a gating mechanism for transmitting a smoothed hidden state, which can be nullified and turn the GRU into a plain RNN or an MLP, as described by Dixon (2020). The main difference of GRU to the LSTM balance the stream of information inside the unit without having a partitioned cell memory. Hence:

$$
\begin{aligned}
\boldsymbol{u_t} &= \sigma\left(\boldsymbol{Z}_t \boldsymbol{B}_u\right) \\
\tilde{\boldsymbol{h}}_{\boldsymbol{t}} &= \tanh\left(x_t \boldsymbol{B}_x + \boldsymbol{B}_h\left(\sigma\left(Z_t \cdot \boldsymbol{B}_r\right) \odot h_{t-1}\right)\right. \\
\boldsymbol{h_t} &= (1 - u_t) \odot h_{t-1} + u_t \odot \tilde{h}_t
\end{aligned}
$$

where $\boldsymbol{u_t}$ is defined as the update gate, deciding how much the unit content updates ; $\tilde{h}_t$ is the candidate activation with an element that forget the dependence of $_{t-1}$. Although it shares a lot of similarities with LSTM, there is no empirical conclusion to which one is better, needing to be evaluated for each case, as points out by Chung et al. (2014).

4.2.3 Regularization

To avoid overfitting in the training of the RNNs we choose a simple technique of weight regularization, an idea shown by Krogh and Hertz (1992) to combat the lack of generalization of neural networks on smaller datasets. The idea is to push weights in the network to zero, where gradients are not significant. In our models we introduced a L1 type of regularization.

5 Estimation

The essential of this data-driven approach centers on finding structure in huge datasets. Out-of-sample forecasting evaluation makes a survey in the ideal sum of regularization and the issue of finding the ideal hyper-parameter choice.

- Training: combine the input with an expected output, until a adequately near coordinate has been found.
- Validation and test: Asses how well the have been trained for out-of-sample prediction.

In this present study, we split the training dataset from 03/01/2007 to 05/25/2017 (80% of the dataset) to fit the model and leave the rest for testing, providing an unbiased evaluation of the fit of the model. Since Deep Learning models have many hyperparameters to tune we need to specify a cross-validation (CV) scheme with the purpose of determine the generalization error of the algorithm and choose the appropriate parameters. Traditional techniques used in Machine Learning such as k-fold CV are not valid for a time-series context, as explained in De Prado (2018). The k-fold CV randomly partition into k equal sized subsamples the original full sample. Of the k subsamples, one is separated as the validation data for testing data, and the remaining $k-1$ subsamples are used as training. This process is repeated k times.

One of the main reasons for this failure is because observations cannot be assumed to be drawn from an IID process, another reason is that testing set is used multiple times leading to selection bias and data leakage. Data leakage is present when the training and validation set have the same information. To avoid data leakage problems, we utilize a CV Block Estimation technique (CV-B), similar to Racine (2000). It differs from k-fold CV by respecting the order sequence in the data and differs from traditional walk forward validation used in time series econometrics by introducing a margin between training and validation, furthermore a margin between the folds used at each interaction in order to prevent data leakage.

We have chosen not to set *ex ante* the hyperparameters in our models. We utilized a grid search CV-B with of hyperparameters that maximizes the CV performance conducting an exhaustive search for a certain parameter space. For the MLP we defined a neuron (p) grid : $[10, 20, 50, 64]$; a Dropout grid ($\mathbf{r}$) : $[0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6]$ and K layer grid : $[3, 4, 5]$. For the RNN models, we define a neuron grid : $[5, 10, 20]$ and a weight regularization grid: $[0, 1e-3, 1e-2, 1e-1, 1e-4, 1e-5]$. This estimation is done for each asset.

As described by Dixon (2020), to build and assess a machine learning, we begin with preparing information of input-output sets $D = \{Y^i, X^i\}_{i=1}^N$. The main goal is to search for a relation of $Y = F(X)$, where we have a loss function $\mathcal{L}(Y, \hat{Y}) = \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2$. The solution of the optimization is:

$$
\begin{gathered}
\min_{B,\alpha} f(B, \alpha) \\
f(B, \alpha) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}\left(Y^{(i)}, \hat{Y}\left(X^{(i)}\right)\right)
\end{gathered}
$$

The loss function is non-convex, having a considerably quantity of local minima and is often hard to find a global minimum. The traditional way to resolve this problem it's calculate the Gradient Descent, an iterative procedure (Ruder (2016)). Defining θ as the bias and weights parameters, where $\theta_k = (\boldsymbol{B}_k, \boldsymbol{\alpha}_k)$. For each of $\theta_k \in \theta$, the gradient of the loss is calculated of the loss, $\nabla_{\theta_k} \mathcal{L}(Y, \hat{Y})$. The negative gradient gives the direction of steepest descent and the direction in which to adjust θ_k to reduce $\mathcal{L}(Y, \hat{Y})$, the following update is performed using the gradient:

$$
\boldsymbol{\theta}_k \leftarrow \boldsymbol{\theta}_k - \gamma \nabla_{\boldsymbol{\theta}_k} L(\boldsymbol{y}; \boldsymbol{\theta})
$$

where γ is the learning rate. The updates are computed for each and every one of the parameters and the loss function is recalculated. These stages are done over up to the fulfilment of the stopping rule. This approach has a high computational cost, for this reason the literature turned to stochastic gradient descent (SGD), as explained by

Deng et al. (2013). This is an alteration of the gradient descent in which the parameters updates are calculated utilizing only a single data point each time. Although this approach solves this problem, the nature of optimization problem stills imposes limitations to the SGD. Due to the loss function non-convex nature it is possible for a SGD procedure to get stuck in saddle points or local minima. More recently, Kingma and Ba (2014) introduced the ADAM algorithm, that adapt the vanilla gradient descent by tuning the learning rate of individual parameters using the estimated first and second moments of the gradient. Ruder (2016) indicates that is the best overall choice of optimizer which justify our use in this work.

6 Results

We estimate the models depicted in Section 4 to make a forecast of one day ahead. This section elaborates the discussion of the results and makes a comparison of all models' forecasting errors. Besides that, we add the forecasting errors of a random walk model to be used as a benchmark. The period utilized for the training and testing of the models covers various different situations of the global economy, such as the 2008 economic crisis and the sovereign debt crisis in Europe, important regime changing events in the market microestructure of the securities presented in this analysis. The final model architecture for each Country are the following:

Table 3 – Model Architecture for each Country

	MLP(ReLu)	MLP(tanh)	RNN	GRU	LSTM
Brazil CDS Returns	HL: 3; Drop: 0.6 ; N: 50	HL : 3 ; Drop: 0.3 ;N: 10	N: 20; L1: 1e-5	N: 20 ; L1: 1e-5	N: 20 ; L1: 0.0001
Chile CDS Returns	HL : 3; Drop: 0.1; N: 60	HL: 4; Drop: 0.1; N: 64	N: 10; L1: 0	N: 10 ; L1: 1e-05	N: 20 ; L1: 0.0001
Peru CDS Returns	HL: 3; Drop: 0.6; N: 64	HL: 3; Drop: 0.5; N: 50	N: 10 ; L1: 1e-5	N: 20 ; L1: 0.0001	N: 20 ; L1: 1e-5
Colombia CDS Returns	HL: 4 ; Drop: 0.3; N : 50	HL: 3; Drop: 0.4; N: 50	N: 5 ; L1: 0.001	N: 5 ; L1: 0.0001	N: 5 ; L1: 1e-5
Russia CDS Returns	HL: 4; Drop: 0.3; N: 20	HL: 4; Drop: 0.4; N: 50	N: 10 ; L1: 0	N: 5 ; L1: 1e-5	N: 5; L1: 0.0001
South Africa CDS Returns	HL: 4; Drop: 0.2; N: 10	HL: 3; Drop: 0.5; N: 50	N: 5 ; L1: 1e-5	N: 20 ; L1: 0.001	N: 20 ; L1: 0.0001
Turkey CDS Returns	HL: 3; Drop: 0.4; N: 50	HL: 5; Drop: 0.3; N: 64	N: 20 ; L1: 0	N: 10 ; L1: 0.001	N: 10 ; L1: 1e-5

HL: Hidden Layers, Drop: Dropout, N: Neurons

Table 4 and 5 displays the root mean squared errors of the forecasts and mean absolute errors. All the figures in bold constitute the best model following the measure of error for the forecast horizon. The figures in the table are reported in basis points. Overall, the MLP with hyperbolic tangent activation function presents the best performance, even compared with the random walk baseline. This result points out that the internal model averaging induced by the sparsity regularization had a better predictive effect than introducing non-linear autocorrelation.

Table 4 – Forecast mean absolute errors

MAE×10000	LSTM	GRU	RNN	MLP(ReLu)	MLP(tanh)	RW
Brazil CDS Returns	187.514	211.574	434.743	147.564	**142.999**	202.944
Chile CDS Returns	**194.751**	288.029	239.108	205.07	197.128	274.911
Peru CDS Returns	241.164	149.706	417.936	142.532	**142.161**	197.458
Colombia CDS Returns	191.606	180.311	196.688	159.444	**155.537**	211.75
Russia CDS Returns	261.785	252.576	220.65	**179.394**	181.803	245.112
South Africa CDS Returns	245.541	244.942	427.808	177.989	**173.092**	249.373
Turkey CDS Returns	188.69	174.173	201.23	176.517	**174.104**	251.487

The table displays the mean absolute errors of the one step ahead forecasts. All the figures in bold constitute the best model following the measure of error for the forecast horizon. The figures in the table are reported in basis points.

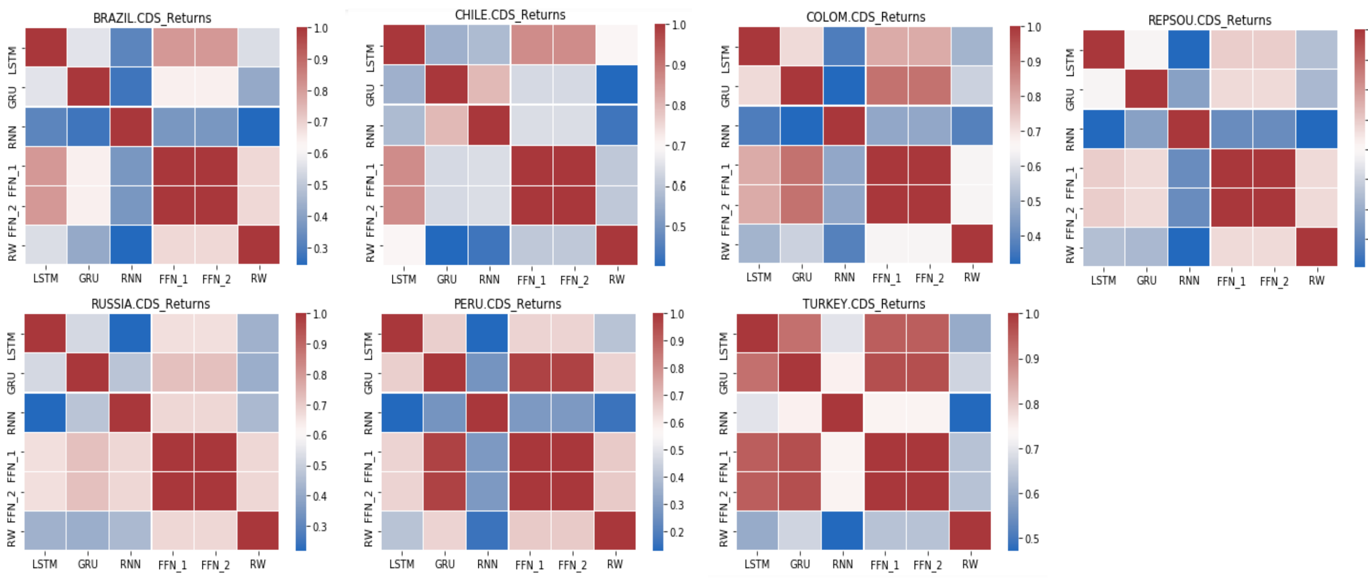
Table 5 – Forecast root mean squared errors

RMSE×10000	LSTM	GRU	RNN	MLP(ReLu)	MLP(tanh)	RW
Brazil CDS Returns	246.727	298.842	632.369	202.72	**199.394**	275.398
Chile CDS Returns	**271.177**	449.303	424.985	308.648	271.846	378.523
Peru CDS Returns	327.155	215.367	659.095	214.504	**213.597**	290.19
Colombia CDS Returns	273.421	252.727	355.028	221.993	**214.05**	292.211
Russia CDS Returns	358.876	349.25	397.972	**254.303**	257.79	347.039
South Africa CDS Returns	325.803	331.512	671.007	243.724	**237.284**	332.597
Turkey CDS Returns	274.651	260.224	352.874	263.982	**257.888**	354.351

The table displays the root mean squared errors of the one step ahead forecasts. All the figures in bold constitute the best model following the measure of error for the forecast horizon. The figures in the table are reported in basis points.

It's important to notice that even if the MLP(tanh) presents better results in comparison with other Deep Learning architectures, when we compare other models with our baseline in majority of cases we get better results. This indicates evidence that non-linear autocorrelation structures can contribute for prediction of returns of a security with a non-linear payoff such as CDS. The comparisons between models utilizing traditional metrics are rather deterministic, i.e, it's not evaluating the difference between the values are statistically significant. We show in the forecast error correlations between the models. The patter in very similar for all countries. The forecasts are all positively correlated putting into question if these models produce statistically significant results.

Figure 5 – Forecasting Error Correlation



To establish this comparison, we utilize the seminal work of the model confidence set, Hansen et al. (2011). Utilizing the MCS allow us to compare a large number of models. The MCS creates a confidence set that incorporates the leading model with probability $(1-\alpha)$ returned by the test. The MCS utilizes bootstrapped samples for the squared error loss function. The confidence set estimates p-values for all models and uses α to select the models to include in the set, also there a cardinal ranking by p-value. The test it's an iteration process that remove models until there is no more rejection of the null. The MCS p-values are presented in the table 6. All the figures in bold constitute the models that lasted in the confidence set with $\alpha = 20\%$. The random walk was removed from all country specific forecasts. The only which were in the confidence set for almost all securities were the MLP, with exception of Chile and Turkey.

Table 6 – Model Confidence Set

MCS p-values	LSTM	GRU	RNN	MLP(ReLu)	MLP(tanh)	RW
Brazil Returns	0	0	0	0	**1**	0
Chile CDS Returns	**1**	0	0	0	0	0
Peru CDS Returns	0	0	0	0	**1**	0
Colombia CDS Returns	0	0	0	0	**1**	0
Russia CDS Returns	0	0	0	**1**	0	0
South Africa CDS Returns	0	0	0	0	**1**	0
Turkey CDS Returns	0	0	**0.33**	0	**1**	0

The table display the model confidence set p-values for the one step ahead forecast. The figures in bold are in the $\alpha = 20\%$ confidence set. Models with p-value of 1 are the best ones, or ones that last in the confidence set.

7 Conclusion

In this present paper we have tested several Deep learning architectures with the objective of capture non-linear effects for forecasting credit derivatives returns. The methods discussed here have been evaluated using Sovereign CDS data of seven developed countries and a set of predictors that incorporate the main drivers identified in the literature. The results can be summarized as follow.

For one-step ahead prediction, the Multilayer Perceptron(MLP) deliver the best forecast 87% of the time, for the set of countries in our sample. When utilizing point metrics we conclude that other non-linear architectures presented in this work have a better performance compared to the baseline random walk model. To verify if this conclusion is statistically significant, we utilize the Model Confidence Set proposed by Hansen et al. (2011) and conclude that the predictions are different and rank the models accordingly, bringing more evidence that non-linearity matters.

Our results are in line with the new evidence of machine learning models in finance. Alternative neural networks modelling techniques gives more flexibility to the forecaster providing a way to capture non-linearity in the data, and provide better results than the traditional benchmark approach in time series econometrics. Although the results shown are promising, this empirical exercise suffers some limitations and that can be explored in an extension of this work. There is no evidence shown that this gains are persistent for longer time frames and that the hyperparameters chosen in this dataset are robust to different regime changes. Another aspect to point out is that extensions may provide a contribution about the drivers of the CDS spreads, showing non-linear relations not captured by previous literature.

Bibliography

Aldasoro, I. and Ehlers, T. (2018). The credit default swap market: what a difference a decade makes. *BIS Quarterly Review*, (June):1–14.

Ammer, J. and Cai, F. (2011). Sovereign CDS and bond pricing dynamics in emerging markets: Does the cheapest-to-deliver option matter? *Journal of International Financial Markets, Institutions and Money*, 21(3):369–387.

Arce, O., Mayordomo, S., and Peña, J. I. (2013). Credit-risk valuation in the sovereign CDS and bonds markets: Evidence from the euro area crisis. *Journal of International Money and Finance*, 35:124–145.

Augustin, P., Subrahmanyam, M. G., Tang, D. Y., and Wang, S. Q. (2014). Credit default swaps–a survey. *Foundations and Trends® in Finance*, 9(1-2):1–196.

Badaoui, S., Cathcart, L., and El-Jahel, L. (2013). Do sovereign credit default swaps represent a clean measure of sovereign default risk? a factor model approach. *Journal of Banking & Finance*, 37(7):2392–2407.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166.

Bianchi, D., Büchner, M., and Tamoni, A. (2019). Bond Risk Premia with Machine Learning. *SSRN Electronic Journal*, pages 1–58.

Bouri, E., de Boyrie, M. E., and Pavlova, I. (2017). Volatility transmission from commodity markets to sovereign cds spreads in emerging and frontier countries. *International Review of Financial Analysis*, 49:155–165.

Bouri, E., Shahzad, S. J. H., Raza, N., and Roubaud, D. (2018). Oil volatility and sovereign risk of BRICS. *Energy Economics*, 70:258–269.

Broto, C. and Pérez-Quirós, G. (2014). Disentangling contagion among sovereign CDS spreads during the European debt crisis. *Journal of Empirical Finance*, 32:165–179.

Carr, P. and Wu, L. (2007). Theory and evidence on the dynamic interactions between sovereign credit default swaps and currency options. *Journal of Banking & Finance*, 31(8):2383–2403.

Chen, L., Pelger, M., and Zhu, J. (2019). Deep Learning in Asset Pricing. *SSRN Electronic Journal*.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078.

Chung, J., Gülçehre, Ç., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555.

Ciarlone, A., Piselli, P., and Trebeschi, G. (2009). Emerging markets' spreads and global financial conditions. *Journal of International Financial Markets, Institutions and Money*, 19(2):222–239.

Cook, T. (2020). Neural networks. In P., F., editor, *Macroeconomic Forecasting in the Era of Big Data. Advanced Studies in Theoretical and Applied Econometrics*, volume 52. Springer, Cham.

Coulombe, P. G., Leroux, M., Stevanovic, D., and Surprenant, S. (2020). How is machine learning useful for macroeconomic forecasting?

De Prado, M. L. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

Deng, L., Li, J., Huang, J.-T., Yao, K., Yu, D., Seide, F., Seltzer, M., Zweig, G., He, X., Williams, J., et al. (2013). Recent advances in deep learning for speech research at microsoft. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8604–8608. IEEE.

Dixon, M. F. (2020). *Machine Learning in Finance: from Theory to Practice*. Springer Nature.

Dixon, M. F., Polson, N. G., and Sokolov, V. O. (2019). Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading. *Applied Stochastic Models in Business and Industry*, 35(3):788–807.

Duffie, D. and Singleton, K. J. (2012). *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Eichengreen, B. and Mody, A. (1998). What explains changing spreads on emerging-market debt: fundamentals or market sentiment? Technical report, National Bureau of Economic Research.

Fender, I., Hayo, B., and Neuenkirch, M. (2012). Daily pricing of emerging market sovereign CDS before and during the global financial crisis. *Journal of Banking and Finance*, 36(10):2786–2794.

Gamboa, J. C. B. (2017). Deep learning for time-series analysis. *CoRR*, abs/1701.01887.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT press.

Gu, S., Kelly, B. T., and Xiu, D. (2018). Empirical Asset Pricing via Machine Learning. *SSRN Electronic Journal*.

Gu, S., Kelly, B. T., and Xiu, D. (2019). Autoencoder asset pricing models. *Available at SSRN*.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.

Hébert, B. and Schreger, J. (2017). The costs of sovereign default: Evidence from argentina. *American Economic Review*, 107(10):3119–45.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, 37(4):543–558.

Kaastra, I. and Boyd, M. S. (1995). Forecasting futures trading volume using neural networks. *The Journal of Futures Markets (1986-1998)*, 15(18):953.

Kalbaska, A. and Gatkowski, M. (2012). Eurozone sovereign contagion: Evidence from the CDS market (2005-2010). *Journal of Economic Behavior and Organization*, 83(3):657–673.

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, 34(11):2767–2787.

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

Krogh, A. and Hertz, J. A. (1992). A simple weight decay can improve generalization. In *Advances in neural information processing systems*, pages 950–957.

Kuan, C.-M. and White, H. (1994). Artificial neural networks: an econometric perspective. *Econometric Reviews*, 13(1):1–91.

Lando, D. (2020). Credit default swaps: A primer and some recent trends. *Annual Review of Financial Economics*, 12:177–192.

Longstaff, F. A., Pan, J., Pedersen, L. H., and Singleton, K. J. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103.

Moshiri, S. and Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19(3):201–217.

Oehmke, M. and Zawadowski, A. (2016). The Anatomy of the CDS Market. *The Review of Financial Studies*, 30(1):80–119.

Pan, J. and Singleton, K. J. (2008). Default and recovery implicit in the term structure of sovereign CDS spreads. *Journal of Finance*, 63(5):2345–2384.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics*, 99(1):39–61.

Remolona, E. M., Scatigna, M., and Wu, E. (2008). The dynamic pricing of sovereign risk in emerging markets: Fundamentals and risk aversion. *The Journal of Fixed Income*, 17(4):57–71.

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

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Schmidhuber, J. and Hochreiter, S. (1997). Long short-term memory. *Neural Computing*, 9(8):1735–1780.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Srivastava, S., Lin, H., Premachandra, I. M., and Roberts, H. (2016). Global risk spillover and the predictability of sovereign CDS spread: International evidence. *International Review of Economics and Finance*, 41:371–390.

Uribe, M. and Schmitt-Grohé, S. (2017). *Open Economy Macroeconomics*. Princeton University Press.

Wegener, C., Basse, T., Kunze, F., and von Mettenheim, H.-J. (2016). Oil prices and sovereign credit risk of oil producing countries: an empirical investigation. *Quantitative Finance*, 16(12):1961–1968.

Yu, D. and Deng, L. (2016). *Automatic Speech Recognition*. Springer.

Appendix

APPENDIX A – Predictors Data

A.1 Predictors Data

A.1.1 List of Predictors

Table 7 – Predictors Data

	DATA	SOURCE	VARIABLE
Global Economic Outlook Data	Bloomberg Financial Conditions Index US	Bloomberg	BFCI US
	Bloomberg Financial Conditions Index EU	Bloomberg	BFCI EU
	Bloomberg Base Energy Commodity Index	Bloomberg	CMDINRGS
	Bloomberg Base Metals Spot Price Commodity Index	Bloomberg	CMDIBASS
	Citi Economic Surprise Index - EUROZONE	Bloomberg	CESIEUR
	Citi Economic Surprise Index - Japan	Bloomberg	CESIJPY
	Citi Economic Surprise Index - Emerging Markets	Bloomberg	CESIEM
	Crude Oil WTI Price	Bloomberg	CL1.Comdty
	Brent Crude Oil Price	Bloomberg	C01.Comdty
Global Financial Data	JP Morgan EMBI Plus Sovereign Spread	Bloomberg	JPEIPLSP Index
	JP Morgan EMBI Plus	Bloomberg	JPEIPLUS Index
	J.P. Morgan EMBI Diversified Sovereign Spread	Bloomberg	JPEIDISP Index
	JP Morgan EMBI Global Diversified Index spread	Bloomberg	JPEIGD Index
	JP Morgan Emerging Markets Bond Index spread	Bloomberg	JPEIGLSP Index
	30-day expected volatility of the S&P 500 Index.	Bloomberg	VIX
	S&P 500 Returns	Bloomberg	SPX.Index
	3 months Euribor rate	Bloomberg	EUR003M
	3 month US Dollar LIBOR interest rate	Bloomberg	US0003M