T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT BİLİM DALI

ESSAYS ON CREDIT SWAP DEFAULT MARKET:
Time Series Predictions with Machine Learning

Ph.D Dissertation

LİNA BAROKAS

ISTANBUL, 2022

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İKTİSAT BİLİM DALI

ESSAYS ON CREDIT SWAP DEFAULT MARKET: Time Series Predictions with Machine Learning

Ph.D Dissertation

LİNA BAROKAS

Advisor: Prof. Sadullah ÇELİK

ISTANBUL, 2022

ABSTRACT

ESSAYS ON CREDIT SWAP DEFAULT MARKET:
Time Series Predictions with Machine Learning

In this dissertation, credit default swap risk premiums of BRICS (Brazil, Russia, China, South Africa) and MINT (Mexico, Indonesia and Turkey) countries are predicted by applying the most recent ground-breaking techniques in machine learning. These techniques include Ridge regression, LASSO estimator and recurrent neural networks, which are Elman's RNN, NARX, GRU and LSTM. The predictive powers of these methods are compared by the mean absolute error, mean squared error and R-squared and these results differ by country and the type of the state-of-the-art forecaster. The contribution of each feature to the prediction is discussed by analyzing the SHAP values. The daily data is collected from 03/01/2012 and ends on 29/11/2019 to avoid COVID-19 effects on the state of the global economy. CDS risk premium for 5-year maturity contracts is used as an output variable and it is predicted by the selected local variables and global variables. In first chapter, in terms of higher R-squared obtained, LASSO outperforms Ridge regression in all selected countries. However, in terms of lower RMSE and the MAE scores, the lowest values are obtained by ridge regression for all selected countries. In second chapter, in all countries except China, the results of NARX outweigh Elman's RNN. In third chapter, GRU outperforms LSTM network for Brazil and Indonesia. LSTM network outperforms GRU in Russia, South Africa, Mexico and Turkey. Both networks perform similarly for China. The results of Turkey model reached the highest forecast accuracy among the selected countries. Overall, recurrent neural networks outperform the ridge regression and LASSO estimator.

Keywords: Credit default swap, Forecasting, Time series, Machine Learning, Recurrent neural networks, Deep learning

ÖZET

KREDİ TEMERRÜT PİYASASI ÜZERİNE MAKALELER:
Makine Öğrenmesi ile Zaman Serisi Tahminleri

Bu tezde, BRICS (Brezilya, Rusya, Çin, Güney Afrika) ve MINT (Meksika, Endonezya, Türkiye) ülkelerinin kredi temerrüt riski, en yeni yöntemler olan makine öğrenmesi ile tahmin edilmiştir. Bu teknikler arasında Ridge regresyonu, LASSO tahmin edicisi ve tekrarlayan sinir ağları (Elman RNN, NARX, LSTM, GRU) kullanılmıştır. Kullanılan metotların tahmin kuvveti, ortalama mutlak hata, ortalama karekök hatası ve anlamlılık katsayısı R ile karşılaştırılmıştır. Sonuçlar ülke ve kullanılan metot bazında farklılıklar göstermektedir. Kullanılan açıklayıcı değişkenlerin, hesaplanan tahminlere katkıları, SHAP değerleri analiz edilerek açıklanmıştır. Çalışmada 03/01/2012'den 29/11/2019 tarihine kadar olan günlük veri kullanılmıştır. Tahmin edilen değişken, 5 yıl vadeli kontratlı kredi temerrüt risk birimi olarak seçilmiştir. İlk bölümün sonucunda, LASSO tahmin edicisi ile daha yüksek R katsayısı elde edilmiştir. Ancak Ridge regresyonu ile daha düşük ortalama mutlak hata ve ortalama karekök hatası elde edilmiştir. İkinci bölümün sonucunda, Çin haricindeki ülkelerde, NARX tekrarlayan sinir ağı, Elman tekrarlayan sinir ağından daha iyi sonuçlar üretmiştir. Son bölümde ise, Brezilya ve Endonezya için GRU tekniği ile LSTM tekniğinden daha iyi sonuçlar elde edilmiştir. Fakat LSTM, Rusya, Güney Afrika, Meksika ve Türkiye için daha iyi sonuçlar çıkarmıştır. Çin için iki yönteminde sonuçları yakındır. Tahmin modellerinin performansını ülke bazlı kıyaslarsak, sinir ağı en yüksek tahmin kuvvetini Türkiye için ulaşmıştır. Sonuç olarak, derin öğrenme yöntemi olan tekrarlayan sinir ağları, makine öğrenmesi olan Ridge regresyonu ve LASSO tahmin edicisinden daha yüksek doğrulukta tahminler hesaplamıştır.

Anahtar Kelimeler: Kredi temerrüt riski, tahmin, zaman serisi, makine öğrenmesi, tekrarlayan sinir ağları, derin öğrenme

ACKNOWLEDGEMENTS

This dissertation is a result of the hard work of many years and I wouldn't be able to achieve this success without the support and love of my family; my father, Mordo Barokas and my mother, Vivi Barokas. I am very thankful to Prof. Sadullah Çelik, as my supervisor, for his continuous encouragement and valuable guidance. I appreciate the help and the contribution of my monitoring jury members, Prof. Arif Orçun Söylemez and Assoc. Prof. Ayben Koy. I am very grateful to my valuable friend Dr. Yasin Kütük for his excellent support on Python and recurrent neural networks. I would like to thank to my friend Dr. Rudy Grosman for sharing his valuable knowledge on programming on R.

Lina BAROKAS

Istanbul, 2022

TABLE OF CONTENTS

LIST OF TABLES ..... **VI**
LIST OF FIGURES ..... **VII**
LIST OF ABBREVATIONS ..... **VIII**
1. INTRODUCTION ..... **IX**
2. LITERATURE REVIEW **1**
2.1. Local features 5
2.1.1. The bond yield and the slope of the yield curve 5
2.1.2. The liquidity risks 7
2.1.3. The local stock market return 8
2.1.4. The exchange rates 9
2.2. Global features 9
2.2.1. Risk, contagion and the volatility in global markets 9
2.1.2. Oil price 10
2.1.3. The spillover effect from the U.S. 11
2.3. Other low-frequency macroeconomic features 12
3. DATA **14**
4. CHAPTER I **17**
4.1. Ridge regression 19
4.2. LASSO (Least Absolute Shrinkage and Selection Operator) 20
4.3. Empirical findings 21
5. CHAPTER II **27**
5.1. Recurrent neural networks (RNN) 27
5.1.1. Elman's RNN 27
5.1.1.1. The forward propagation 28
5.1.1.2. The backward propagation 29
5.1.1.3 Long term dependencies and the vanishing/exploding gradient 30
5.1.2. Nonlinear autoregressive network with exogenous inputs (NARX) 30
5.2. Empirical findings 31
6. CHAPTER II **36**
6.1. Long short-term memory 37

6.2. Gated recurrent unit 39
6.2. The differences of GRU and LSTM 41
6.4. Empirical findings..... 42
7. CONCLUSION..... **49**
REFERENCES..... **51**
APPENDIX..... **60**

LIST OF TABLES

Table 3.1: The explanation of variables

Table 4.1: The estimation results of LASSO and ridge regression

Table 5.1: The estimation results of ELMAN's RNN

Table 5.2: The estimation results of NAR

Table 6.1: The estimation results of LSTM

Table 6.2: The estimation results of GRU

LIST OF FIGURES

Figure 1.1: The process of CDS transaction
Figure 1.2: Total CDS contracts - Notional amounts outstanding (in billions of US Dollars)
Figure 1.3: CDS market outlook in the first half of 2020
Figure 3.1a: The CDS risk premiums for BRICS
Figure 3.1b: The CDS risk premiums for MINT
Figure 4.1: Contours of the error and constraint functions for the LASSO and ridge regression
Figure 4.2a: Ridge regression for BRICS countries
Figure 4.2b: Ridge regression for MINT countries
Figure 4.3a: LASSO for BRICS countries
Figure 4.3b: LASSO for MINT countries
Figure 5.1: The workflow of Recurrent Neural Network
Figure 5.2: The architecture of RNN
Figure 5.3: The architecture of Elman's RNN and NARX
Figure 6.1: The architecture of Long Short-Term Memory
Figure 6.2: The architecture of Gated Recurrent Unit
Figure 6.3: The workflow of LSTM and GRU
Figure 6.4: The feature importance for BRICS countries
Figure 6.5: The feature importance for MINT countries
Figure A.1: Mean square errors for cross validation, ridge (left) and LASSO (right) for BRICS countries
Figure A.2: Mean square errors for cross validation, ridge (left) and LASSO (right) for MINT countries
Figure A.3: The comparison of RNN results by MAE, MSE and R-squared for Brazil
Figure A.4: The comparison of RNN results by MAE, MSE and R-squared for Russia
Figure A.5: The comparison of RNN results by MAE, MSE and R-squared for China
Figure A.6: The comparison of RNN results by MAE, MSE and R-squared for South Africa
Figure A.7: The comparison of RNN results by MAE, MSE and R-squared for Mexico
Figure A.8: The comparison of RNN results by MAE, MSE and R-squared for Indonesia
Figure A.9: The comparison of RNN results by MAE, MSE and R-squared for Turkey
Figure A.10: The estimated loss for Brazil in ELMAN's RNN
Figure A.11: The estimated loss for Russia in ELMAN's RNN
Figure A.12: The estimated loss for China in ELMAN's RNN
Figure A.13: The estimated loss for South Africa in ELMAN's RNN
Figure A.14: The estimated loss for Mexico in ELMAN's RNN

Figure A.15: The estimated loss for Indonesia in ELMAN's RNN
Figure A.16: The estimated loss for Turkey in ELMAN's RNN
Figure A.17: The estimated loss for Brazil in NARX
Figure A.18: The estimated loss for Russia in NARX
Figure A.19: The estimated loss for China in NARX
Figure A.20: The estimated loss for South Africa in NARX
Figure A.21: The estimated loss for Mexico in NARX
Figure A.22: The estimated loss for Indonesia in NARX
Figure A.23: The estimated loss for Turkey in NARX
Figure A.24: The estimated loss for Brazil in LSTM
Figure A.25: The estimated loss for Russia in LSTM
Figure A.27: The estimated loss for South Africa in LSTM
Figure A.28: The estimated loss for Mexico in LSTM
Figure A.29: The estimated loss for Indonesia in LSTM
Figure A.30: The estimated loss for Turkey in LSTM
Figure A.31: The estimated loss for Brazil in GRU
Figure A.32: The estimated loss for Russia in GRU
Figure A.33: The estimated loss for China in GRU
Figure A.34: The estimated loss for South Africa in GRU
Figure A.35: The estimated loss for Mexico in GRU
Figure A.36: The estimated loss for Indonesia in GRU
Figure A.37: The estimated loss for Turkey in GRU

ABBREVIATIONS

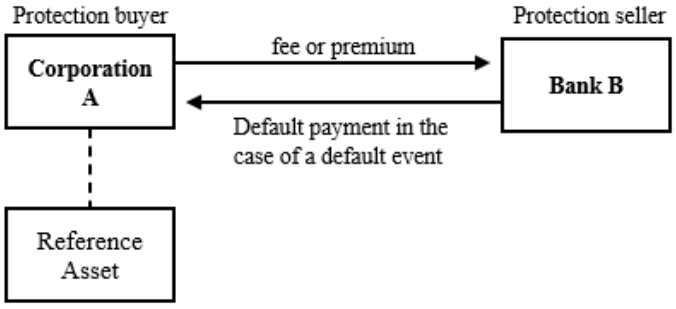
ARIMA	: Autoregressive Integrated Moving Average
BDI	: Baltic Dry Index
BIS	: Bank for International Settlements
BRICS	: Brazil, Russia, India, China, South Africa
CBOE VIX	: Chicago Board Options Exchange Volatility Index
CDS	: Credit Default Risk
CPI	: Consumer Price Index
EMBI	: Emerging Market Bond Index
ISDA	: International Swaps and Derivatives Association
GDP	: Gross Domestic Product
GRU	: Gated Recurrent Unit
LASSO	: Least Absolute Shrinkage and Selection Operator
LIBOR	: The London Inter-Bank Offered Rate
LSTM	: Long Short-Term Memory
MAE	: Mean Absolute Error
MINT	: Mexico, Indonesia, Nigeria, Turkey
MLP	: Multilayer Perceptron
OLS	: Ordinary Least Square
RMSE	: Root Mean Square Error
RNN	: Recurrent Neural Network
RSS	: Residual Sum of Squares

INTRODUCTION

Credit default risk is the possibility of a counterparty's failure to meet its financial obligations. The transfer of the credit default risk let the counterparty to manage its finances more efficiently and to diversify its risk exposure. It is possible to separate the credit risk from the underlying asset and transfer it to another counterparty by credit derivatives. In this case, the owner of the underlying asset and the owner of the credit risk can be different. Credit derivatives are financial contracts which is used to distribute and allocate the credit risk of default to third party besides the lender, it offers a protection in the case of the credit default. It is possible to transfer and reduce the credit default risk but not eliminate. In other words, it lets the investors to hedge their assets against risky events. The use of credit derivatives has gained significance in the last two decades and it has played a crucial role in the spread of the global financial crisis.

The credit derivative market grew rapidly just before the global financial crisis in 2008. The credit default swap (CDS) contracts became the most tradable and liquid product among credit derivatives based on International Swaps and Derivatives Association (ISDA)'s year-end surveys (which year?). CDS is a bilateral contract which provides a protection for the underlying risky asset in exchange for a periodic or on time payments to a protection (CDS) seller. Therefore, in the case of a credit default, the seller is responsible from the covering the loss. The process is illustrated in Figure 1.1, where the corporation A owns the reference asset and wants to buy a protection and Bank B is a protection seller.

Figure 1.1: The process of CDS transaction

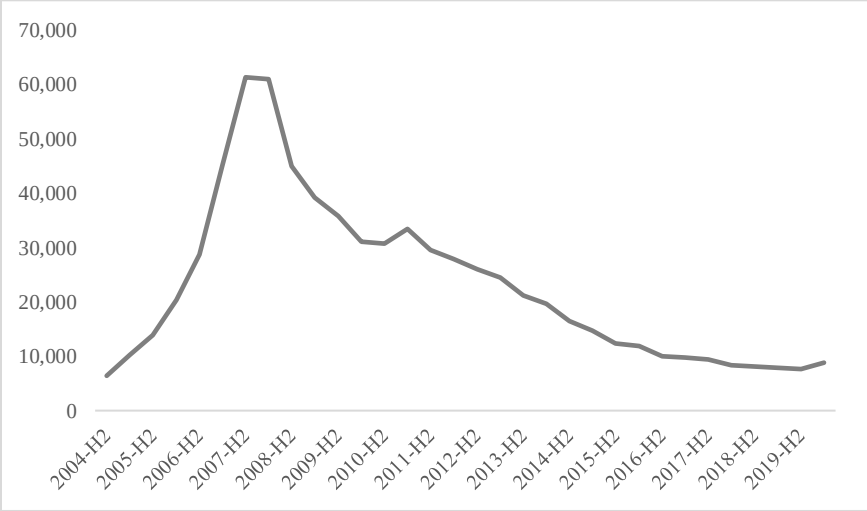


Source: Modified from the plot used in 'An Introduction to Credit Derivatives' by Author.

CDS shows a wide information about the current state of the market and it is seen as the creditworthiness of the underlying asset or the sovereign. It becomes one of the leading indicators to assess sovereign's credit default risk. The CDS risk premiums are revised on daily basis and it reflects the new information in the market rapidly. This is the biggest advantage of CDS relative to other indicators. In addition, the use of CDS helps to reduce the asymmetric information problem in the market.

The growth of CDS market and volume has been accelerated just before the global financial crisis. The total notional amount outstanding in the CDS market increased from 6 trillion US Dollars in 2004 to 60 trillion US Dollars at the end of 2007[1] (see Figure 1.2). This massive growth and the newly designed financial instruments allowed investors to trade credit risk globally. Therefore, the volume of the market has been increased not only because of the demand for transferring or hedging the risk, but also because of the securities traded based on speculation. This risk has been indicated by the Basel Committee regulations since the 1980s. After the global financial crisis, Bank of International Settlements (BIS) and ISDA imposed more regulatory restrictions on CDS trading to enhance the financial system.

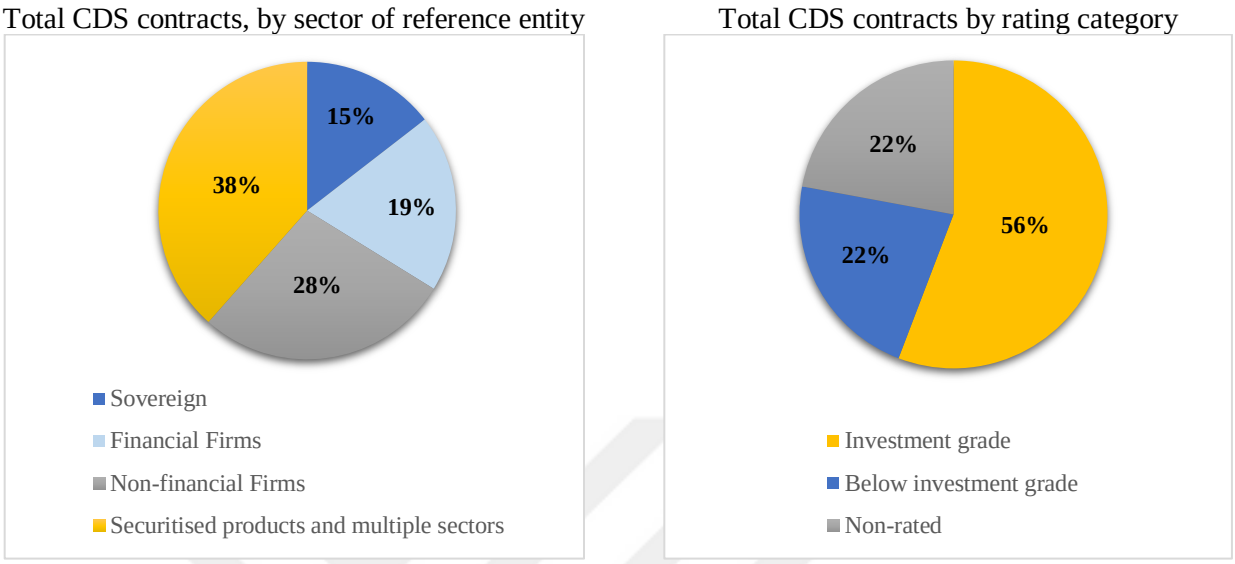
Figure 1.2: Total CDS contracts - Notional amounts outstanding (in billions of US Dollars)



Source: Database of International Swaps and Derivatives Association (ISDA)

[1] According to Bank for International Settlements (BIS) over-the-counter derivatives statistics.

Figure 1.3: CDS market outlook in the first half of 2020



Source: Database of International Swaps and Derivatives Association (ISDA)

The purpose of this dissertation is to predict of Credit Default Swap (CDS) in BRICS (Brazil, Russia, China, South Africa) and MINT (Mexico, Indonesia, Turkey) countries by the application of machine learning algorithms. The selected countries are rapidly growing economies and provide profitable investment opportunities under pro-growth-oriented governing but at the same time they suffer from political instability. The most of the previous studies in the literature aim to investigate the determinants of CDS risk premiums and the studies about prediction is relatively in a limited scope. Therefore, this dissertation makes three contributes to the previous literature. First, the study considers not only the spillover effect from the U.S. but also the spillover effect from Eurozone countries by adding Germany's risk-free rate, the slope of the yield curve, and stock price index return as features. Second, to the best of our knowledge, the Baltic Dry Index (BDI) is included as a feature to predict CDS risk premiums for the first time in the literature. Third, different than other studies, the implementation of the state-of-the-art methodologies has increased the predictive power and the application of recurrent neural networks (RNN) made a methodological contribution to the finance literature.

The literature survey part is divided into three different parts based on the features are used in the previous studies, these are local, global and the other macro variables. Local features are aimed to present the current state of the local economy. Global features are aimed to present the current state of the global economy, the possible spillover and contagion effects from the U.S and Europe. The other macroeconomic features consist of the low-frequency features which is not possible to obtain in daily basis. The data selection process is discussed in the following part. The empirical part aims to discuss and compare the

usefulness and the accuracy of new ground-breaking methodologies in machine learning. In this study, several methodologies are investigated under three major chapters. In first chapter, the regularization techniques in the machine learning, such as ridge regression and LASSO are discussed. In second and third chapters, deep learning architectures which are a sub-field of machine learning are investigated. Elman's recurrent neural network (RNN) and nonlinear autoregressive network with exogenous inputs (NARX) are presented in the second chapter. In the third chapter, the gating techniques to overcome the challenge of vanishing/exploding gradients, which are Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), are discussed. The last part summarizes and concludes the dissertation.

LITERATURE REVIEW

A comprehensive literature analyzes the relation between the CDS risk premiums and financial and macroeconomic variables. The selection of features led us to an enormous discussion in the previous studies. Features differ across studies, and results distinguish by the magnitude of the shocks. We can separate the features used in the literature into three main groups: proxies for the local state of the economy, global features which includes risk and contagion effect, spillover effect from the US and oil price and lastly, low frequency macroeconomic features. Some of the findings are contradictory based on the countries' integration in global financial markets and the state of local economy.

1.1. Local features

1.1.1 The bond yield and the slope of the yield curve

The relationship between CDS risk premium and bond yield has mainly been analyzed using corporate bonds since the underlying principle is the same with sovereign bonds. An investor, who wants to buy protection in case of a risk of credit default, buys CDS. Duffie (1999) shows that the CDS risk premium is equal to the difference between corporate and risk-free bonds when both of them have been seen as measures of credit risk. From an arbitrage strategy perspective, both a risk-free bond and a high-risk bond insured by CDS should have the same yield since both are seen as default-risk free. If an investor buys CDS on this risky bond, the yield of this investment will be a risk-free rate.

Hull, Predescu and White (2004) study the relationship between corporate CDS spread and bond yield with a 5-year maturity and find that the theoretical relation between CDS and bond spread holds. A government bond in local currency is considered a risk-free asset since it is faced with a very low credit default risk. Nonetheless, if there is a one-to-one change between CDS and bond spread, the arbitrage between the two assets is perfect (Duffie 1999: 75, Hull and White 2000: 7). On the other side, Zhu (2006) finds that there could be a divergence from the aforementioned identity because of the liquidity factor. According to the study of Zhu (2006), the equivalence theory of the bond yield and CDS risk premiums holds in the long-run but in the short run, there is a significant deviation from the equilibrium. Blanco, Brennan and Marsh (2005) argue that the bond markets are less liquid than the credit markets based on the

assumption of the pricing of new information in CDS premiums is more rapid than in corporate bond price. Thus, this causes CDS to trade at a higher spread level compared to bond. Collin-Dufresne and Solnik (2001) observe that the corporate yield is closely related to the liquidity of the bond market.

Merton (1974) builds a pricing model for corporate debts based on the risk structure of interest rates. According to the model, the risk-free interest rate is negatively related to credit spreads. An increase in the risk-free rate decreases the present value of the expected future cash flows, which also decreases the probability of the sell option. On the other hand, a rise in risk-free rate raises the expected growth rate of the market value and a higher market value lowers the price of the sell option. Hence, these two effects might lower the CDS spread. Xiao et al. (2004) extend Merton's corporate credit default risk modeling approach to sovereign credit risk to measure sovereign balance sheet risk. They show that most of the theoretical approaches are also applicable for sovereign credit risk. Chan-Lau and Kim (2004) put Merton's theory into the use of sovereign credit risk and examine the relation between sovereign CDS spreads, bonds and equity markets for selected emerging markets. They conclude that CDS risk premiums and bond spreads convergences in some countries. They use the JP Morgan EMBI Bond Index which may not be a good proxy for some of the countries such as Mexico and Turkey.

Besides bond yield, Litterman and Scheinkman (1991) study another important variable, the slope of the yield curve. A steeper yield curve shows an increasing spread difference between long and short-term interest rates, which also indicates a positive upturn in the state of the economy and may result in higher inflation and interest rates. As follows, this causes a higher yield demand by investors and long-term investments become riskier. A flatter yield curve shows a decreasing spread difference between long and short-term interest rates, which indicates uncertain economic situations and weakening in the state of the economy and denotes the end of high economic growth and an increase in fears of a slowdown. In this case, investors demand similar rates for all maturities. The flatter yield curve might happen when the central bank tends to increase policy rates or signal for monetary tightening. An inverse yield curve may occur when short term rates exceed long term rates, if the effect is strong enough. Increasing the policy rate in order to cool down the overheated economy will result in a slowdown in economic growth and it will be followed by a low interest policy again to cope with recession possibility. Therefore, an inverse yield curve indicates that the short term is riskier than the long term and may suggest economic recession. Occurring especially in the countries which heavily depend on foreign exchange income, due to high external debt and current account deficit. These countries are very sensitive to capital outflows and for them, it is difficult to cope with the volatility of exchange rates and price stability problem at the same time due to the possibility of a

high exchange rate passthrough effect on consumer prices. Because of these reasons, the central banks of these countries tend to take unexpected and fast steps on monetary policy.

Liu and Morley (2013) decide to use the spread between a 10-year government bond and 3-month Libor rates and conclude that if there is a deterioration in a country's fiscal position, the 10-year government bond yield will rise as the government attempts to finance the extra borrowings and consequently default risk will increase together with the sovereign CDS spreads. Avinoa and Nnejib (2014) study on European iTraxx CDS index and they preferred to use 10-year and 2-year Euro swap rates for yield curve based on a negative relationship assumption with CDS spreads. Calice et al. (2015) focus on the difference between 10-year and 5-year CDS spreads in Italy, Spain, Portugal, Czech Republic and Poland. Accordingly, the local market yield curve is defined as the difference between 10-year and 5-year government bond yield. They observe a positive shock for Spain and Portugal, while other countries experienced a disconnection between the two markets which might arise from underdeveloped CDS markets and limited arbitrage opportunities. Augustin (2018) investigates the yield curve of 10-year and 1-year CDS premiums. The upward slope shows the dominance of the local shocks and the downward slope shows the dominance of the global shocks. Based on differentiated hypothesis and results, one may expect either a positive or negative relationship between the slope of the yield curve and CDS spreads.

1.1.2 The liquidity risks

The liquidity represents the degree of convertibility of an asset or security to cash in the market, in other words, the speed of the trade of the security, without causing a significant change in its price. The existence of adverse selection, market frictions and any other costs also affect the liquidity of CDS market and increase the spreads in the less liquid markets. As, it is emphasized in the previous section, Zhu defends that liquidity might invalidate the theoretical relationship of CDS and bond spreads being equal. Collin-Dufresne and Solnik (2001) and Blanco, Brennan and Marsh (2005) demonstrate the movements in liquidity premiums may explain a large proportion of the total variation in credit spreads. Acharya and Johnson (2007) find that in the case of an existence of insider trading in the CDS market, causes an adverse selection for uninformed traders. The potential liquidity providers would be less inclined from participating in the market.

The most used variables to measure liquidity are equity volume in both local and global market and bid-ask spread. On the other side, TED spread[2] is a common proxy for the liquidity in the global market. Acharya and Johnson (2007) show that the equity volume is positively related to liquidity risk and it represents the participation in trading by heterogenous investors. Tang and Yan (2007) employ the bid-ask spread and the equity volume separately and both proxies demonstrate that a significant liquidity effect on CDS spreads and liquidity risk has already been priced beyond liquidity level in the CDS market. Lafuente and Serrano (2015) find uni-directional causality from default to liquidity for developed countries. According to the study, illiquidity exhibits a more idiosyncratic nature than default does and illiquidity times are short-lasting. Also, a positive illiquidity risk premium has already been priced in the sovereign CDS market.

However, Kim (2017) finds the bid-ask spread employed by many researchers less reliable as a measure of liquidity if the CDS and liquidity risk are positively correlated. Remolona, Scatigna and Wu (2008) find that the expected loss is largely driven by country-specific sovereign risk fundamentals and market liquidity.

1.1.3 The local stock market return

The option-pricing theory has been suggested by Merton (1974) to define the relationship between bond and equity prices. A declining firm value reflects decreasing equity prices as well as decreasing bond prices. Accordingly, the equity prices and bonds are positively related while bond spreads are negatively related. Thus, there is a negative relation between CDS spreads and stock market return.

Most of the studies prefer to employ the local stock market return to demonstrate the current state of the local economy. A negative relationship between the two variables is confirmed by Chan and Zhang (2009), Longstaff et al. (2011) and Liu and Morley (2013) for the U.S., European countries, Asian countries except for China. Eyssel, Fung and Zhang (2013) also study on China's CDS spreads and their estimates showed that local stock market return is suitable to explain the CDS risk levels. In addition, they show that the changes in CDS risk led to stock market return which may arise from less restrictions, more widespread investors and more efficient information advantage compared to the local stock market. Ma et al. (2018) argue that the CDS markets are driven by the local stock market return especially when there is a good

[2] TED spread is estimated by the difference between 3-months LIBOR based on U.S. Dollars and 3-months Treasury Bill which is calculated by Federal Reserve Bank of St. Louis.

regime. From the magnitude perspective, both Calice et al. (2015) and Augustin (2018) justify that the stock market return on spread changes is stronger compared to other variables. As a result, both the local stock market and sovereign CDS market benefit from the same good news.

Zhang (2003) finds an insignificant relationship between the stock and CDS markets, specifically to explain the default probabilities and recovery pace in the credit market. Moreover, Chan and Zhang couldn't find any cointegration relationship in Japan, Indonesia, Malaysia and the Philippines which might be associated with a non-linear relationship between the markets.

1.1.4 The exchange rates

Longstaff et al. (2011) find that the exchange rate against US Dollar and CDS spreads are positively related in emerging market economies. Liu and Morley (2012) and (2013) employ nominal effective exchange rates and conclude that the exchange rate has the most important effect on CDS spreads in the European Union, the U.S, Japan, and the selected emerging countries. Clark and Kassimatis (2015) and Ma et al. (2018) use bilateral exchange rates and found a positive and strong relationship in selected countries. However, Ma et al. (2018) find insignificant results for China, Russia and Venezuela. The insignificant results in these countries might be based on adopting a pegging currency policy. Yuan and Pongsiri (2014) employ a real effective exchange rate and find statistically insignificant results. Carr and Wu (2007) use currency option implied volatilities and Della Corte, Kozhan and Neuberger (2019) use implied volatilities of currency options and sovereign risk is found to be related to contemporaneously changes in both variables. Calice and Zeng (2019) conclude that the sovereign CDS premiums can predict the movements in the bilateral exchange rates.

1.2. Global Features

1.2.1 Risk, contagion and the volatility in global markets

The Chicago Board Options Exchange (CBOE) Volatility index (VIX) is mostly preferred to define the global risk, contagion and volatility in the previous studies. The index is calculated to produce 30-day expected volatility of the U.S. stock market from mid quote prices of S&P500 Index. The increase of volatility in international markets corresponds to an increase in global risk and uncertainty, therefore VIX and sovereign risk are expected to be positively related.

In some previous studies, the VIX index plays a crucial part in the movements of sovereign credit risk and the variations mostly come from the global factors rather than the local economy measures (Pan and Singleton 2008: 2373, Fontana and Scheiher 2010: 26, Longstaff et al. 2011: 12, Srivastava et al. 2016: 386). Based on Fender, Hayo and Neuenkirch (2012), the country-specific factors are more related to the CDS risk premiums of the emerging markets than the regional and global factors. In addition, they introduce that CDS spreads tend to be affected more during the market stress time. According to Calice et al. (2015), CDS spread tends to move along with the investors' risk appetite. Blommestein, Eijffinger and Qian (2016) find that global and Eurozone factors are the main drivers of the variation in the CDS spreads for Euro area countries.

1.2.2 Oil price

The studies about the relation between oil price and CDS risk premiums have been increased since the oil price collapse in 2014. Guo and Newton (2013) examine the shocks between the CDS risk premiums and the energy market by a regime-switching Vector Autoregressive model. They justify the crucial role of oil price and stock price in explaining the oil price during the risky regimes. Alexandre and de Benoist (2010) and Tandon et al. (2019) document the highest impact for Russia, Argentina, Columbia, Mexico and Venezuela. The magnitude of oil price shock depends on the relevance of oil for a country and the country's position in oil trade, net exporter or importer. Duffie, Pedersen and Singleton (2003) analyze the sovereign yield spreads in Russia and since Russia is an oil exporter country, they conclude that the Brent oil price is an important variable to determine Russia's current account and fiscal deficit. Dai and Serletis (2018) analyze world crude oil production and the CDX index by VAR analysis. The results show that the shocks explain the 90% variation in the CDX index in the long-run. Sharma and Thuraisamy (2013) predict CDS risk return by oil price uncertainty, using out-of-sample prediction for eight Asian countries. Pavlova, E.de Boyrieb and Parhizgaric (2018) examine sovereign risk premiums of ten oil exporting countries by using both crude oil price, Brent oil price and CBOE OVX volatility index and find a directional spillover to sovereign CDS spreads by using VAR model. They also notice an increase in spillover effects in oil-exporting emerging countries after the price collapse in 2014. As a result, the spillover in the sovereign debt market is found to be stronger than in equity markets of oil exporting countries.

Shahzad et al. (2017) predicts the risk transfer from the oil market to the CDS market by rolling-window causality perspective. They focus on asymmetric non-linearities between the oil price and sovereign CDS markets. Oil price is found to be the most important factor to determine CDS risk premiums

in oil-exporting countries. In addition, Sabhka, Perettia and Hmaied (2019) study the nonlinearity of oil shocks on sovereign credit risk for both oil-exporter and oil-importer countries. The results show that the positive news for the oil market affects negatively the sovereign CDS volatility but it tends to stay muted in the stable regime. Chuffart and Hooper (2019) study the sovereign CDS spreads of Russia and Venezuela through a Time Varying Transition Probabilities Markov Switching model. They find that the crude oil price plays a critical role in Venezuela's spreads where this role is less effective on Russia's spreads.

1.2.3 The spillover effect from the U.S.

The most common proxies are the slope of the yield curve, U.S. interest rates and the spread between corporate bonds for the spillover effect from U.S. financial market and the market expectations of the future monetary policy of the Federal Reserve.

If an investor is optimistic about future aspects of the economy, this may stimulate an increase in the slope of the yield curve (McGuire and Schrijvers 2003: 75). As a result, emerging markets may favor from this high demand and it may ease the pressure on the probability of sovereign default, and a decline in emerging market sovereign spreads. By using the long and short-term U.S. government bond spreads, both Zhang (2003) and McGuire and Schrijvers (2003) find a negative relationship with CDS spreads in selected emerging countries. The other studies choose to employ the spread between the 10-year return on U.S. BB-rated industrial corporate bonds and the 6-month U.S. Treasury Bill rate, the spreads between five-year BBB- and AAA-rated bonds and between five-year BB and BBB-rated bonds (Pan and Singleton, 2008; Longstaff et al. ,2011). The 5-year constant maturity U.S. Treasury yield is used by both Longstaff et al. (2011) and Augustin (2018) for global liquidity. The aforementioned studies conclude that CDS risk premiums are mostly driven by the U.S. macroeconomic factors rather than by local market factors for these countries (McGuire and Schrijvers 2003: 74, Pan and Singleton 2008: 2381, Longstaff et al. 2011: 2 and Fender, Hayo and Neuenkirch 2012: 2790). Ma et al. (2018) find that US Treasury yield has a stronger effect during the bad regime than in a good regime. However, Eyssel, Fung and Zhang (2013) use the spread of the 10-year U.S. Treasury yield and the 3-month U.S. Treasury bill rate and the spread of the AAA and BAA U.S. corporate bonds for China and the domestic economic factors seem more related in explaining the CDS risk premiums.

1.3. Other low-frequency macroeconomic features

Some previous studies work on monthly or quarterly data which enabled them to work on different macroeconomic variables such as GDP, industrial production, inflation rate, fiscal policy indicators and debt statistics. The studies which concentrate on how macroeconomic effects impact the CDS spreads directly include Catao and Sutton (2002), Fender, Hayo and Neuenkirch (2012), Liu and Morley (2012) and (2013), Alper, Forni and Gerard (2012) and Ma et al. (2018).

Some of the studies focus on the effect of credit rating changes on CDS spreads and it has been found that the market put greater emphasis on macroeconomic, financial and global variables rather than rating changes (Catao and Sutton 2002: 23, Liu and Morley 2012: 132 and 2013: 345, Ho 2016: 588, and Ma et al. 2018: 15). Catao and Sutton (2002) analyze volatility by separating it into external and domestic policy components. The domestic policy includes government balance-to-GDP and the ratio of monetary base-to-net international reserves. They show that countries facing historically more macroeconomic instability have a tendency to default more than others, and this instability is considered policy-induced, particularly from the procyclicality of fiscal policy. In these countries, the key to improve the country's credit worthiness is to reduce policy instability. Ho (2016) states that current account, foreign debt and international reserves are crucial factors in explaining long-run CDS spreads in emerging countries, but not in the short-run. The international reserve plays a more crucial role than the current account to improve the sovereign CDS risk premiums in the long-run. However, Liu and Morley (2013) find international reserves are statistically insignificant. Ma et al (2018) indicate that a country's sovereign risk is affected through interactions in the international trade, financial market and geopolitical incidence. By employing trade-to-GDP, FDI-to-GDP and GDP variables, they find that open economies which are integrated with the regional economies are facing more notable contagion effect. Smaller countries which have less diversified economies are more sensitive to global crisis.

Surprisingly, both the changes in country ratings and some of the macroeconomic and fiscal variables are found insignificant to explain CDS risk level changes. Alper, Forni and Gerard (2012) use primary deficit-to-GDP ratio, debt-to-GDP and expected deficit growth rate as fiscal drivers to examine sovereign CDS spreads in selected advanced economies. The results of the random effect GLS model show that the fiscal variables have less explanatory power compared to the financial variables and the global variables. Blommestein, Eijffinger and Qian (2016) find that local variables have a lower impact on the CDS risk premiums of four Euro-area countries except for Italy. Augustin and Tedongap (2016) find that

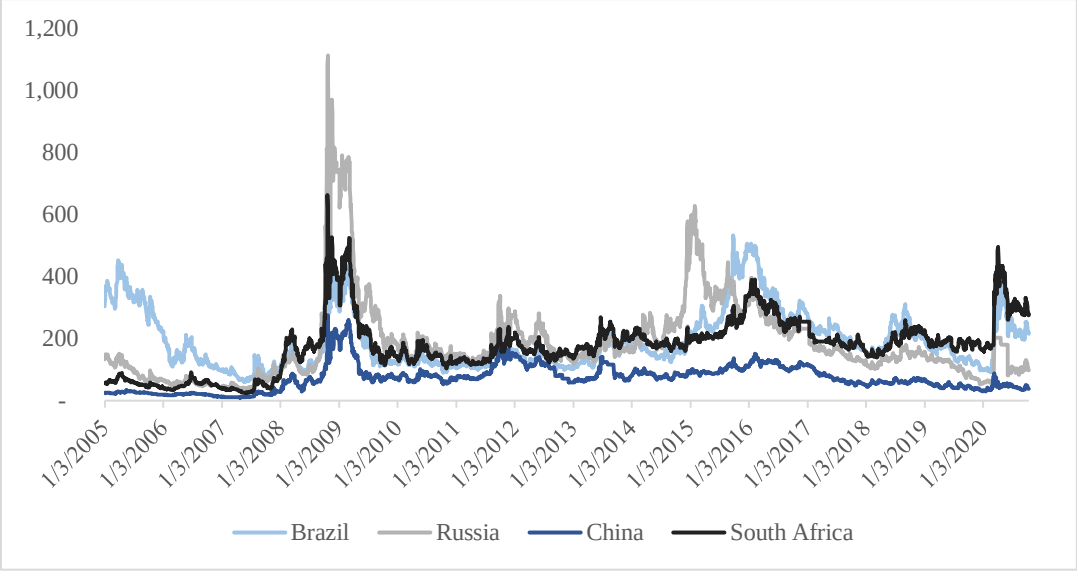
the expected growth and macroeconomic uncertainty are insufficient to explain the variation in the CDS risk premiums.

DATA

The daily data is collected from 03/01/2012 and ends on 29/11/2019 to avoid COVID-19 effects on the state of the global economy. The beginning of the sample was chosen based on the earliest dates for which the data are available for all countries. All data are obtained from Bloomberg. The sample covers BR(I)CS (Brazil, Russia, China and South Africa) and MI(N)T (Mexico, Indonesia and Turkey) countries. India allowed CDS contracts in 2011 with an extra set of rules to prevent a possibility of a financial crisis or the issue of expensive debt by the Reserve Bank of India. The CDS market of India has limited scope compared to the rest of the world, where only the owner of the reference entity has a right to buy CDS and the contract cannot be traded. Therefore, since it is not possible to sell the CDS contract or offset, India has been excluded from the sample. In addition, Nigeria does not take part in the sovereign CDS market, and it is left behind.

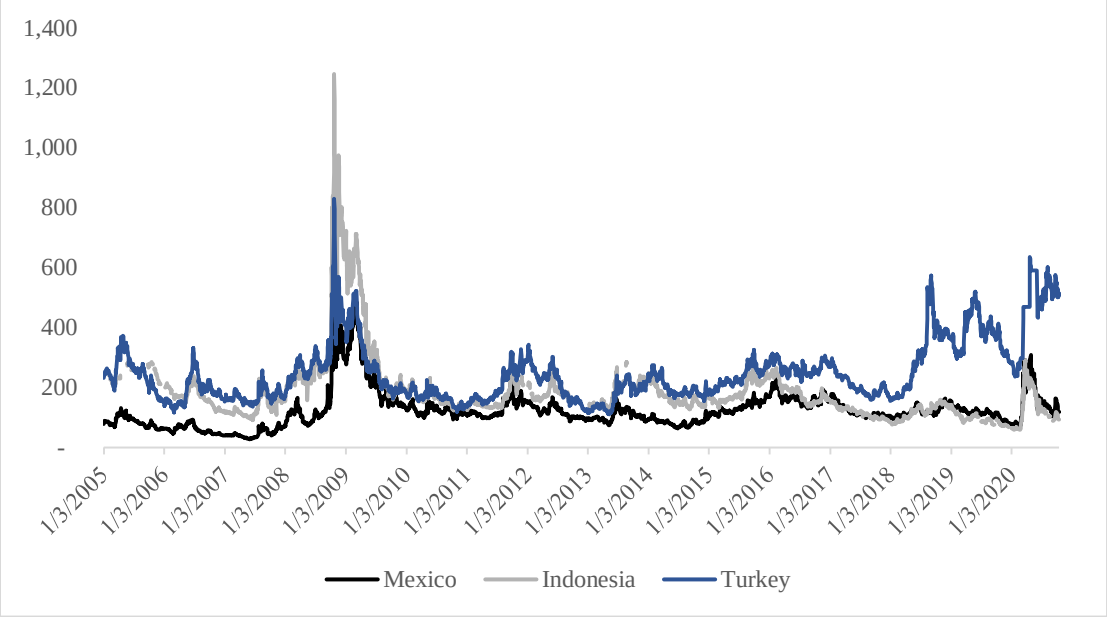
CDS risk premium for 5-year maturity contracts is used as an output variable that is going to be predicted since it faces the highest liquidity in the CDS market compared to 1-year and 10-year maturity CDS contracts. The CDS risk premiums for selected countries are drawn in Figure 3.1 and 3.2.

Figure 3.1: The CDS risk premiums for BRICS



Source: Bloomberg Database.

Figure 3.2: The CDS risk premiums for MINT



Source: Bloomberg Database.

The features are divided into local variables and global variables. Local variables consist of risk-free rate, the slope of the yield curve, the daily change of the stock price index, the daily change of bilateral exchange rate against the U.S Dollar, and the daily change of stock price index volume. The risk-free rate is chosen as a 5-year maturity government bond and the slope of the yield curve is estimated by the difference between a 10-year maturity government bond and a 3-month treasury bill. The daily return of the local stock market and the daily change of the local exchange rate against the U.S. Dollar are used to identify the local state and international competitiveness of the local economy. The liquidity in the local market is proxied by the daily change volume of the stock market.

Since the increased integration of the financial markets, the ability of a country to repay its debt depends also on the state of the global economy. To capture the reallocation of capital and liquidity shifts between financial markets, the variables from the U.S. and Europe equity and fixed income markets are included based on previous studies (Pan and Singleton (2008), Longstaff et al. (2011), Liu and Morley (2012)). To calculate the spillover effect from the U.S., the 5-year constant maturity U.S. treasury yield, the spread of Moody's BAA and AAA corporate bond yield and the daily return of Nasdaq index are chosen. TED spread, the difference between 3-months LIBOR based on U.S. Dollars and 3-months Treasury Bill, is a widely-used proxy for global market liquidity. Germany's 5-year maturity government bond yield, the

spread of the 10-year and 3-month maturity government bond yield and the daily return of DAX and FTSE indexes capture some of the spillover effects from Eurozone countries and the United Kingdom. The CBOE VIX index and oil price are chosen for the global risk and volatility. The Baltic Dry Index (BDI), which is an index of average prices paid for transport of dry bulk materials, is used to capture supply and demand conditions for economic activity. The dry bulk materials are raw materials which can be used in the production of intermediate or final goods. The index is an indicator of production conditions in the future, provides insight into the upcoming economic upturn or downturn. Much more detailed information of respective features is given in Table 3.1.

The dataset is split into two parts as usual in the machine learning literature, the training and the test/validation sets. In the training set, the data covers 70 % of the data in which they start from 03/01/2012 to 20/07/2017. The rest of the data is assigned for the test/validation set starting from 21/07/2017 to 29/11/2019, with the last time point being the beginning of the first emerge of COVID-19 in China.

Table 3.1: The explanation of variables

Variable	Explanation
cds	The 5-year maturity credit default swap risk premium
g5y	The 5-year maturity government bond yield
g10y	The 10-year maturity government bond yield
g3m	The 3-months maturity treasury bill yield
yield	The difference between g10y and g3m for the selected country
exr	Exchange rate against U.S. Dollars
spi	Stock price index
spivol	Stock price index volume
vix	Chicago Board Options Exchange Volatility Index
corp_aaa	Moody's AAA rated corporate bond yield
corp_baa	Moody's BAA rated corporate bond yield
corp_spread	The difference between corp_baa and corp_aaa
ted spread	The difference between 3-months LIBOR based on U.S. Dollars and 3-months treasury bill
oil	Brent oil price
baltic	The Baltic Dry index
nsdq	Nasdaq stock price index
dax	Germany stock price index
ftse	United Kingdom stock price index

Source: Author's definition.

CHAPTER I

In this chapter, we study the shrinkage/regularization methods in machine learning to predict CDS risk premiums. First, we discuss the methodology of linear regression, ridge regression and LASSO. Second, the empirical results are presented.

The linear regression is one of the oldest and widely used approaches for supervised learning. The linear regression assumes a linear relationship between the explained variable *Y* and explanatory variables *X* and predicts a quantitative response *Y* on the basis of a single or single explanatory variable of *X*. This relation can be written as;

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{1.1}$$

where β_0 is an intercept term, β_1 is the slope term and ε is the error term in the linear model.

The linear regression possesses several assumptions which limit the estimation but it is possible to improve these limitations by introducing regularization methods. One of the major limitations is the linearity assumption between the dependent and the explanatory variables. In the real world, the linearity, obtaining a straight-line, is occasional. Linear regression is very sensitive to outliers and in some cases, the outliers have to be removed from the dataset. Second, a multiple regression model, more than one explanatory variable is used, assumes that there is no relation between the explanatory variables, in other words, no existence of multicollinearity. However, this is not the case in general and in the case of multicollinearity, it is hard to separate the individual effect of an explanatory variable on the dependent variable. Third, if the number of observations *n*, is much larger than the number of explanatory variables *p*, then the least squares estimates tend to have a low variance, and it will perform well on test observations. However, if *n* is not much larger than *p*, then it tends to have high variance and a low prediction accuracy, which may cause an overfitting problem. If the number of explanatory variables *p* is bigger than the number of observations *n*, there is no unique solution. It doesn't produce a unique solution and the estimated variance is going to be infinite. The regularization (or shrinkage) techniques help to reduce the estimated variance, as a result, the prediction accuracy improves. Lastly, there is a risk of including irrelevant

variables which may lead to unnecessary complexity. It is possible to omit these variables by setting their coefficients to zero but this is unlikely in the ordinary least square estimation. The regularization (or shrinkage) techniques perform a feature/subset selection by its nature and it excludes or reduces the effect of irrelevant variables on the regression output. By doing so, it reduces the model complexity and prevents the over-fitting problems which arise from the linear regression assumption.

The two most known regularization techniques to shrink the coefficients towards zero are Ridge regression and LASSO (Least Absolute Shrinkage and Selection Operator). Ridge regression is proposed by Heorl and Kennard (1970) and it aims to minimize the coefficient slightly different than the ordinary least squares. Tibshirani (1996) proposes LASSO estimator, which minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Li and Chen (2014) compare the LASSO-based approaches with the dynamic factor model in the out-of-sample forecasting for twenty important macroeconomic variables which cover industrial production, price levels, credit conditions, employment and inflation. As a result, the LASSO-based models outperform the dynamic factor models. De Mol, Giannone, and Reichlin (2008) work on LASSO estimator and Ridge regression in a Bayesian framework. Li, Shi and Gao (2021) predict the monthly S&P500 stock returns by LASSO. Panagiotidis, Stengos and Vravosinos (2018) predict the bitcoin return by LASSO approach and they find that the search intensity on Google and gold price return are the most crucial variables. Zhang, Ma and Wang (2019) predict crude oil prices by LASSO estimator and the LASSO outperforms Ridge regression. Tian, Yu and Guo (2015) predict the corporate failures by LASSO estimator and the results show that the out-of-sample predictive power is superior and outperforms the previous preferred models in the literature. Pereira, Basto and Ferreira da Silva (2016) predict the corporate failures by the logistic LASSO and Ridge regression. Sermpinis, Tsoukas and Zhang (2018) predict the corporate credit ratings by LASSO and compare the results with the benchmark ordered probit model. The prediction accuracy is significantly increased by the application of the LASSO estimator compared to the ordered probit model. Melkumova and Shatskikh (2017) compare OLS, LASSO and Ridge regression to answer the question of how the wine quality score depends on the physicochemical characteristics of the wine. Both the RSS of LASSO and Ridge regression are smaller than the RSS of OLS regression.

1.1 Ridge regression

Ridge regression is proposed by Heorl and Kennard (1970) and it aims to minimize the coefficient slightly different than the ordinary least squares. Recall that the ordinary least squares which aim to minimize the residual sum of squares (RSS), can be estimated as;

$$RSS = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 \tag{1.2}$$

The difference of ridge regression is to add the second term of $\sum_j \beta_j^2$, called as shrinkage penalty. Ridge regression is called L2 regularization technique. The first term, RSS has the same role as in ordinary least squares, while the second term determines the coefficients which can be estimated as zero.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = RSS + \boxed{\lambda\sum_{j=1}^{p}\beta_j^2} \tag{1.3}$$

where $\lambda \geq 0$ is a tuning parameter and it is determined separately. The tuning parameter λ controls the relative impact of these two terms on the regression coefficient estimates. When $\lambda = 0$, the model becomes as same as OLS model, therefore the penalty term has no effect. However, as λ goes towards infinity, the penalty grows, and the ridge regression coefficient estimates will approach zero but it does not set any of the coefficients to zero unless $\lambda = \infty$. The question is how to choose a tuning parameter. The choice of the tuning parameter is based on the smallest cross-validation error received, see appendix A.1 and A.2.

Overall, L2 regulation technique helps to reduce the model complexity which arises from the choice of irrelevant predictors and multicollinearity. Ridge regression includes all predictors p (features) in the final model. The increase in λ will trim the magnitude of the coefficients but it doesn't exclude any predictors from the final model. It is important to note that the shrinkage penalty is only applied to the slope term, not to the intercept.

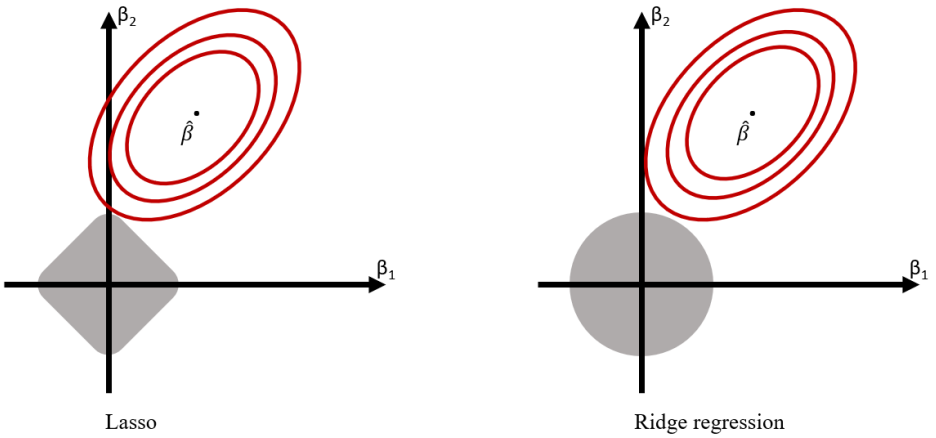
1.2 LASSO (Least Absolute Shrinkage and Selection Operator)

The disadvantage of ridge regression is to keep all predictors in the final model. This doesn't have to be a problem in terms of prediction accuracy, but it may make difficult the interpretation of the model when there is a high number of predictors. Tibshirani (1996) proposes LASSO, a new method for linear models which minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. LASSO is called as L1 regularization technique which helps to overcome this challenge by feature selection. The model is able to produce zero coefficients, and this nature let the model perform a subset selection. The only difference between ridge and LASSO regression is to replace the shrinkage penalty of β^2 by $|\beta|$.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \boxed{\lambda\sum_{j=1}^{p}|\beta_j|} \tag{1.4}$$

LASSO shrinks the coefficient estimates towards zero same as in ridge regression. However, it is able to force some of the coefficient estimates to be exactly equal to zero if the tuning parameter λ is large enough. In this case, LASSO performs a subset/feature selection and it regresses the model which only involves the subset of the predictors. Therefore, the outcome is easier to interpret compared to the one in ridge regression.

Figure 4.1: Contours of the error and constraint functions for the LASSO and ridge regression



Source: Modified from the plot used in 'The Elements of Statistical Learning' by Author.

In Figure 4.1, the solid gray areas are the constraint regions, $|\beta_1| + |\beta_2| \leq s$ and $\beta_1^2 + \beta_2^2 \leq s$, while the red ellipses are the contours of the RSS. The least square estimates are located at the outside of the blue regions. All points located on the ellipses around $\hat{\beta}$, represents the same estimate of the RSS. A wider ellipse corresponds to the increase in the RSS. LASSO and ridge regression coefficients occur at the first point at which an ellipse contacts the constraint region. Since the constraint area of ridge regression is circular, the intersection does not occur on the axis, therefore the ridge regression coefficient estimates will be exclusively non-zero. On the other side, the constraint area of LASSO has intersections at each of the axes, so it will cross the ellipse on the axis where the coefficients will equal zero.

1.3 Empirical findings

The forecasting performance of LASSO and ridge regression are compared based on the mean squared error (RMSE) and the mean absolute error (MAE) and the R-squared in the table 4.1. All results are estimated by the *glmnet* package on *R-Studio*. The RMSE measures the average squared errors between the predicted and the actual values while the MAE ignores the direction of errors by taking the absolute value. The RMSE is more sensitive to outliers since the errors are squared before taking the average. In general, the RMSE is higher than or equal to the MAE (RMSE ≥ MAE). If they are equal, it means that the errors have a similar magnitude. The R-squared shows how much the volatility of the dependent variable can be explained by the selected predictors in the model. The choice of the tuning parameter λ, is based on the smallest cross-validation error received, see appendix A1-2.

Table 4.1: The estimation results of LASSO and ridge regression

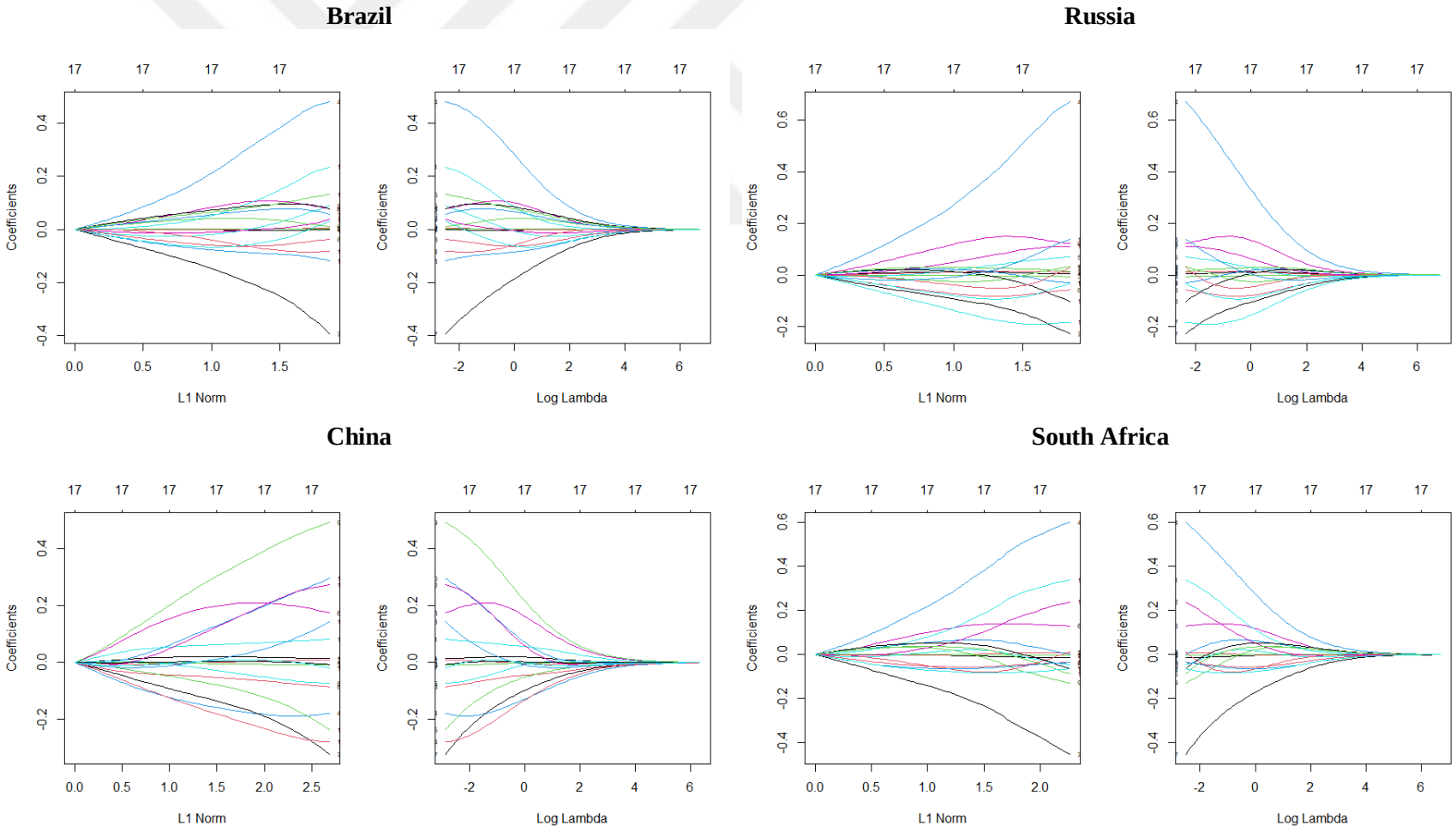
	Brazil	China	Russia	South Africa	Mexico	Indonesia	Turkey
Ridge regression							
RMSE	0.9254	0.8723	0.3839	0.5161	0.8243	0.7824	1.1457
MAE	0.7568	0.7709	0.5061	0.5825	0.7062	0.7572	0.8911
R-sq	0.9590	0.6913	0.9337	0.8869	0.8743	0.8460	0.8859
LASSO							
RMSE	1.5550	1.5281	0.6154	2.1946	1.9061	1.9762	1.7457
MAE	1.0564	1.0033	0.6157	1.1876	1.1134	1.0895	1.1382
R-sq	0.9706	0.7087	0.9605	0.9310	0.9014	0.8874	0.9028
λ	0.0003	0.0004	0.0002	0.0003	0.0003	0.0003	0.0004

Source: Author's estimation.

In terms of higher R-squared obtained, LASSO outperforms Ridge regression in all selected countries. This suggests that the feature/subset selection property of LASSO works better for our data sample. The estimated R-squared for China stands at 0.70 while the R-squared of other countries fluctuates between 0.84 – 0.97. The reason behind this could be because the Chinese CDS market faces more regulatory restrictions compared to the rest of the selected countries in this study. In terms of the RMSE and the MAE scores, the lowest values are obtained by ridge regression for all selected countries. The RMSE scores are higher than MAE scores except Russia and South Africa. It means that the magnitude of the errors is similar in Brazil, China, Mexico, Indonesia and Turkey. From this perspective, Ridge regression, which keeps all features in the model, results better.

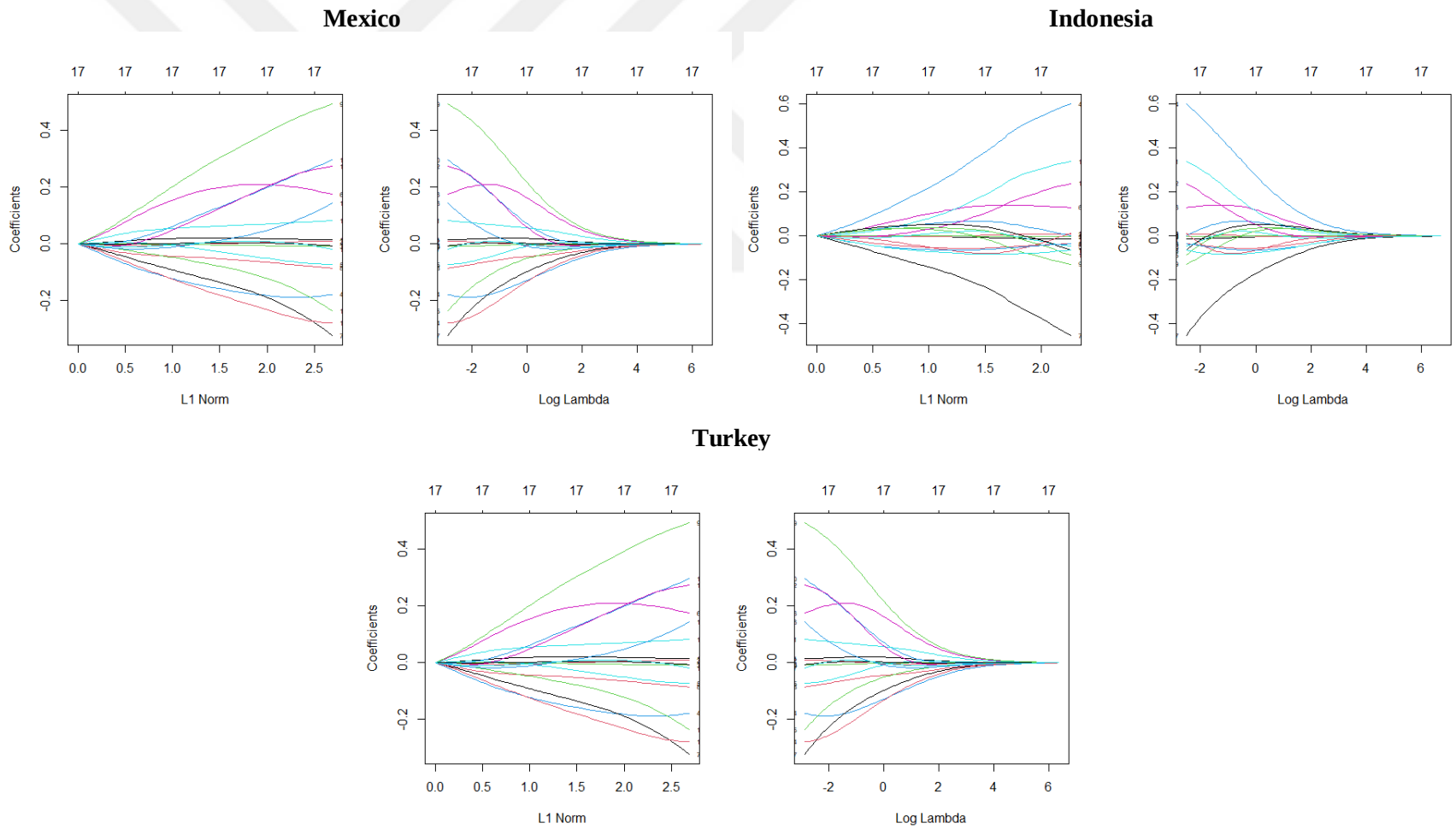
Overall, the results suggest that both models reached strong prediction accuracy and the selected features have a powerful predictive ability. The feature/subset selection doesn't play a crucial role in terms of RMSE and MAE scores but it helps to increase the R-squared (see Figure 4.2a, 4.2b, 4.3a and 4.3b).

Figure 4.2a: Ridge regression for BRICS countries



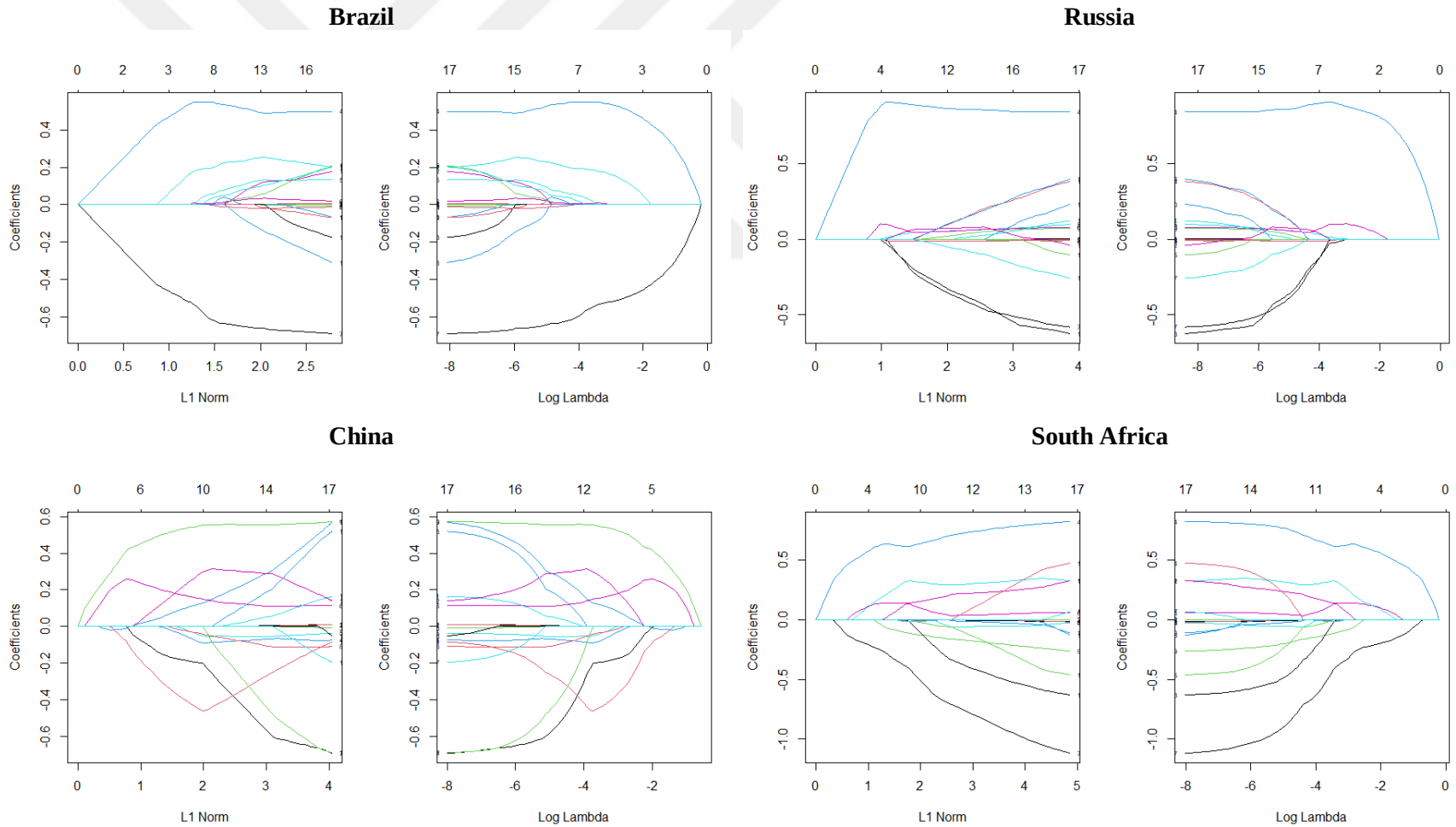
Source: Author's estimation.

Figure 4.2b: Ridge regression for MINT countries



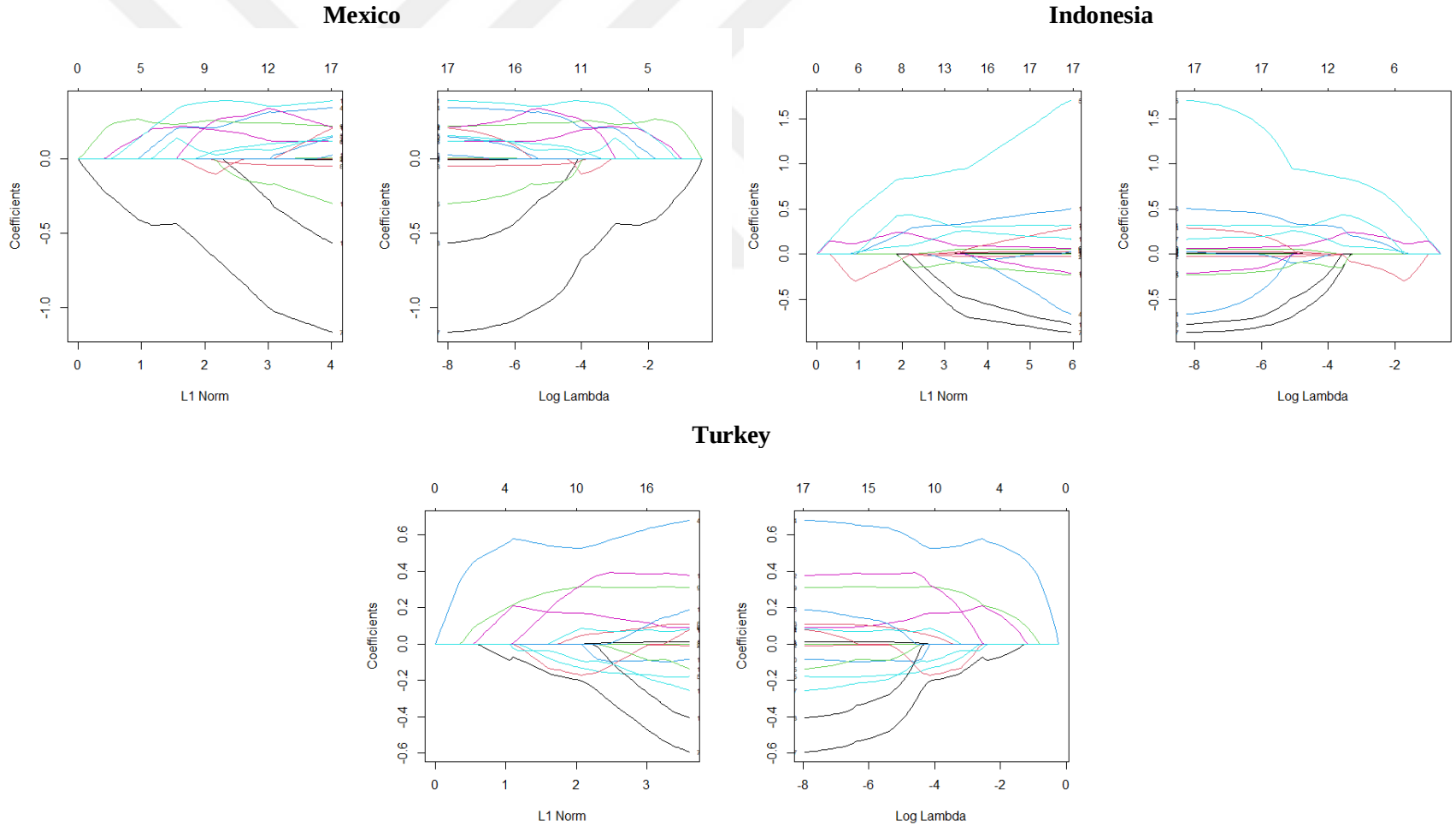
Source: Author's estimation.

Figure 4.3a: LASSO for BRICS countries



Source: Author's estimation.

Figure 4.3b: LASSO for MINT countries



Source: Author's estimation

CHAPTER II

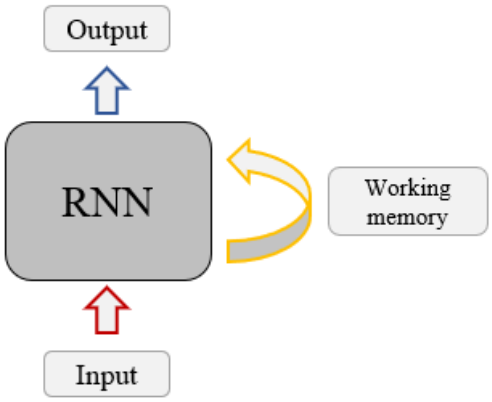
In this chapter, the structure of recurrent neural network models (RNN) has been discussed. Two different types of RNN are estimated for the same dataset to project the CDS risk premiums in the selected countries. These are the simple RNN called Elman and nonlinear autoregressive networks with exogenous inputs (NARX). The studies related to the application of Elman's RNN and NARX are very limited in the previous literature. However, there are only a few examples in other disciplines. Pisoni et al. (2009) and Pisoni et al. (2011) apply NARX network to forecast air pollution levels in Milan. Marcjasz. Uniejewski and Weron (2020) predict the electricity price by NARX.

2. Recurrent Neural Networks

The sequence modelling is to generate a sequence of values by analyzing input variables. The estimated output from the previous step is employed as the input of the current step. In simplified words, it is predicting what comes next. Time series data is one of the most popular and suitable examples for sequential modelling. The time series prediction consists of the study of the present and the past behavior of the system to predict the future (Chandra and Zhang 2012: 116).

Recurrent Neural Network (RNN) is a deep learning algorithm which is specialized in sequential modelling. It is different than the traditional feed forward neural network in terms of processing the input data. The feed-forward neural network processes a fix amount of data to produce a fix amount of output in a time step. On the other hand, RNN processes the data in a sequence, and estimates a hidden state for the next step. This process goes on until the sequence is finalized and the output is estimated. RNN is able to receive a sequence of input values and also it can produce a sequence of outputs. Its architecture's ability to work with a sequence of data makes it very useful to work on time-series data. RNN has a simple structure with a built-in feedback loop and has a memory to remember the knowledge obtained from before and this memory is considered when the output is estimated. The main advantage of RNN is the ability of remembering its previous inputs and takes historical information into account.

Figure 5.1: The workflow of Recurrent Neural Network



Source: Author's drawing

2.1 Elman's Recurrent Neural Network

ELMAN network structure, which is known as one of the first and simplest RNN structure, is proposed by Elman (1990). The network is able to learn the patterns by the recurrent feedback property of the hidden layers. It is a two-layer network which are the input and the hidden layer. The network introduces the recurrent units which aims to transfer the information from the previously hidden layer to the current input layer with a one-step time delay. The information from the previously hidden layer is fed as an input to the current input layer.

2.1.1 The forward propagation

The input vector is represented by x_t and t denotes the time sequence. The corresponding weight matrix with the inputs are denoted by W_x. The hidden layer h_t is a hyperbolic tangent (tanh) function with a range of (-1,1). To estimate hidden state, the input vector x_t is multiplied by the corresponding weight W_x and it is summed up with the information from the previous hidden layer and its weight, $U_h h_{t-1}$, (see Figure 5.2).

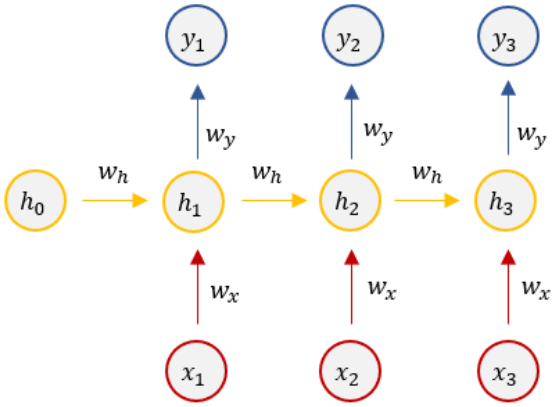
$$h_t = tanh_h(W_x x_t + U_h h_{t-1} + b_h) \quad (2.1)$$

$$y_t = \sigma_{hy}(W_y h_t + b_y) \quad (2.2)$$

The output function is σ_{hy}, where y_t is the output, W_y is the weight of the output and each step is accompanied with a bias unit of b. This is a forward propagation process of an RNN. The error for the time t is estimated by the loss function.

$$L(\hat{y}, y) = \sum_t L(\hat{y}_t - y_t)^2 \tag{2.3}$$

Figure 5.2: The architecture of RNN



Source: Author's drawing.

2.1.2. The backward propagation

The backward propagation was introduced in the 1970s, and it becomes famous after the study of Remulhart, Hinton and Williams (1986). When RNN moves from the input layer to the output layer, called forward propagation. Moving from the output layer to the input layer is called backward propagation. The goal of backward propagation is to minimize the errors produced by loss function during the forward propagation by adjusting or correcting. However, it is not possible to change the input and the actual values of the dependent variable, but it is possible to change the weights of input variables (W_x), and the biases (b_h, b_y) through backward propagation and re-estimate the loss function until we received the minimum value for the errors. The new weights in the backward propagation are estimated by the following formula;

*new weight = old weight – (learning rate * gradient)* (2.4)

2.1.3. Long term dependencies and the vanishing/exploding gradient

As it has shown in the previous section, the estimation of each layer depends on the output from the previous layer. The backward propagation causes a vanishing/exploding gradient problem which is due to the multiplication of a small gradient with the corresponding weight several times. In the existence of a vanishing/exploding gradient problem, the learning process from past events become extremely difficult. The estimated gradient will be very small and as a result, the new weight will be almost equal to the old weight. At the end of this process, the weight of the network is not going to be updated as we want. In other words, the vanishing gradient problem is the forgetting behavior of the neural network (Bengio, Simard and Frasconi 1994: 159, Lin et al. 1996: 1330).

2.2. Nonlinear autoregressive network with exogenous inputs

Elman's RNN is defenseless to the vanishing/exploding gradient problem while the nonlinear autoregressive network with exogenous inputs (NARX) has a practical solution by introducing the short-time lags. NARX adds lags of the features and by doing so it considers the short-term dependencies. Lin et al. (1996) state that the NARX network is also suffering from long-term dependencies however it often performs much better at discovering long-term dependencies than the conventional recurrent neural networks.

The mathematical definition of NARX is same as Elman's RNN structure while only the input vector changes into;

$$y_t = f_y\left(y_{t-1}, y_{t-2,} \dots, y_{t-n}, u_{t-1}, u_{t-2}, \dots, u_{t-k}\right) + \varepsilon_t \tag{2.5}$$

where n and k are the lag order of inputs and outputs respectively. The error term ε_t is the bias between the prediction and the target. NARX has more features than the Elman's RNN because the output is regressed on both the lagged values of the output and the input values. Based on the input variable u_t, the output of the hidden layer can be estimated by;

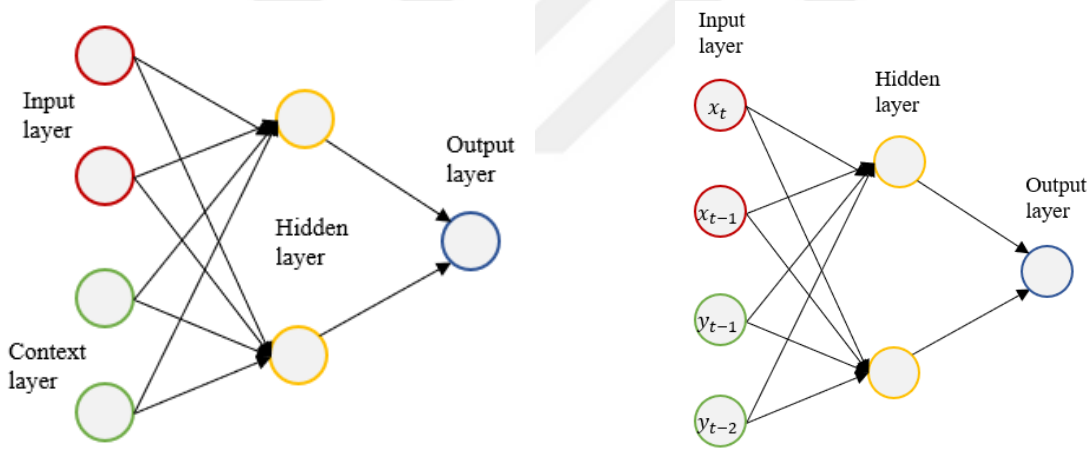
$$h_t = f_h\left[\sum_{t=0}^{k} w_{tk} u_{(t-k)} + \sum_{t=1}^{n} w_{tn} y_{(t-n)} + b_t\right] \tag{2.6}$$

w_{tk} is the corresponding weight between the input and the hidden layer while w_{tn} is the corresponding weight between the hidden layer and the output. The bias term is denoted by b. The mapping function f_h is unknown at the beginning and defined during the training process of the network. Therefore, the output corresponding to the hidden layer h_t can be estimated by;

$$\hat{y}_t = f_y \left[\sum_{t=0}^{j} w_{tj} h_{tj} + b_j \right] \tag{2.7}$$

w_{tj} is the corresponding weight between the hidden layer and the predicted output, b_j is the bias of the predicted output.

Figure 5.3: The architecture of Elman's RNN and NARX



Elman's RNN **NARX**

Source: Author's drawing.

2.3 Empirical Findings

In this section, the results from both networks, Elman's RNN and NARX, are discussed for BRICS and MINT countries. Both networks are run with the same settings in terms of hyperparameters. All data processing is done in *Python 3*, using the packages *NumPy*, *pandas* and *Keras*.

One of the most important hyperparameters is the learning rate λ, which controls the change in network with respect to the estimated error each time the network updates the weights, during the

backpropagation. In other words, it is the pace of the learning of the network to achieve convergence in the training process. If the learning rate is too small, it causes a longer training process and more training epochs, it means that the network is going to make smaller updates to the weights to estimate new errors during the backpropagation. As a result, the network may get stuck. If the learning rate is too large, the network might have an unstable learning process because of the rapid changes, divergent behaviors and fewer training epochs. As a result, the network may converge to a suboptimal solution very fast. Therefore, we ran our networks with four different learning rates based on the traditional default values and most used ones, $\lambda = (0.0001, 0.0003, 0.001, 0.003)$.

The second important hyperparameter is the batch size. The network needs to sample the inputs to train, by this process it can learn the common patterns in the data. If our batch size is 64, this means that the network selects the first 64 observations from the training dataset to train the network and after it continues to train the network again by selecting the next 64 observations. It is important to note that the required memory space is increasing together with the batch size. We decided to employ the network with three different batch sizes since our data period is long enough to work with a higher batch size. The selected batch sizes are 64, 128 and 256. The number of the epoch is fixed at 200 and both networks consist of 50 hidden layers. The number of epochs defines how many times the learning algorithm is going to work through the entire training dataset. This allows the network run again until the errors are minimized.

Our networks run with four different learning rates and three different batches, which gives us twelve different combinations. All countries have twelve different outcomes for both networks. We compared the results by the root mean squared error (RMSE), the mean absolute error (MAE) and the R-squared value in the table 5.1 and 5.2. The distributions of the twelve different outcomes for all networks are illustrated by boxplot in appendix A3-9.

In all countries except China, the results of NARX outweigh Elman's RNN. NARX network is only able to capture a limited number of long-term dependencies compared to the networks that is going to be discussed in the next chapter. However, the results obtained in this chapter suggest that the long-term dependencies play a more important role for all countries except China.

The best prediction accuracy for Brazil, is obtained by employing 0.0003 λ and 256 batch size in NARX, where RMSE and MAE are estimated 0.0006 and 0.0195 respectively with 97.7 % R-squared. The best prediction accuracy for Russia, is obtained by employing 0.003 λ and 256 batch size in NARX, where RMSE and MAE are estimated 0.0002 and 0.0110 respectively with 89.4 % R-squared. The best prediction accuracy for China, is obtained by employing 0.0003 λ and 256 batch size in Elman's RNN, where RMSE

and MAE are estimated 0.0004 and 0.0160 respectively with 92.5 % R-squared. The best prediction accuracy for South Africa, is obtained by employing 0.0003 λ and 256 batch size in NARX, where RMSE and MAE are estimated 0.0004 and 0.0147 respectively with 95.2 % R-squared. The best prediction accuracy for Mexico, is obtained by employing 0.003 λ and 128 batch size in NARX, where RMSE and MAE are estimated 0.0005 and 0.0162 respectively with 94 % R-squared. The best prediction accuracy for Indonesia, is obtained by employing 0.0003 λ and 256 batch size in NARX, where RMSE and MAE are estimated 0.0006 and 0.0188 respectively with 94.6 % R-squared. The best prediction accuracy for Turkey, is obtained by employing 0.0001 λ and 64 batch size in NARX, where RMSE and MAE are estimated 0.0005 and 0.0164 respectively with 99 % R-squared. The results are given in the Table 5.1.The results of the feature importance are going to be discussed at the end of the next section.

Table 5.1: The estimation results of ELMAN's RNN

		Brazil			Russia			China			South Africa		
Batch Size	λ	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq
64	0.0001	0.0108	0.0899	0.5837	0.0027	0.0423	-0.5342	0.0054	0.0615	0.0814	0.0065	0.0734	0.1566
64	0.0003	0.0082	0.0787	0.6855	0.0008	0.0222	0.5433	0.0043	0.0607	0.2696	0.0038	0.0525	0.4992
64	0.001	0.0603	0.2111	-1.3182	0.0023	0.0434	-0.2934	0.0023	0.0387	0.5994	0.0022	0.0356	0.7091
64	0.003	0.0932	0.2743	-2.5816	0.0091	0.0809	-4.1503	0.0060	0.0676	-0.0180	0.0479	0.2120	-5.2537
128	0.0001	**0.0019**	**0.0335**	**0.9286**	0.0022	0.0392	-0.2246	0.0023	0.0405	0.6056	**0.0005**	**0.0169**	**0.9352**
128	0.0003	0.0042	0.0563	0.8382	0.0007	0.0213	0.5913	0.0014	0.0314	0.7693	0.0006	0.0185	0.9267
128	0.001	0.0217	0.1321	0.1648	0.0011	0.0250	0.3560	0.0012	0.0275	0.7979	0.0011	0.0250	0.8567
128	0.003	0.0298	0.1516	-0.1447	0.0007	0.0238	0.5893	0.0025	0.0385	0.5794	0.0030	0.0444	0.6021
256	0.0001	0.0175	0.1060	0.3287	0.0013	0.0300	0.2524	0.0075	0.0721	-0.2755	0.0128	0.0985	-0.6782
256	0.0003	0.0062	0.0708	0.7616	0.0012	0.0275	0.3202	**0.0004**	**0.0160**	**0.9252**	0.0024	0.0401	0.6898
256	0.001	0.0035	0.0541	0.8636	**0.0003**	**0.0132**	**0.8473**	0.0028	0.0447	0.5286	0.0009	0.0241	0.8853
256	0.003	0.0281	0.1478	-0.0819	0.0031	0.0447	-0.7472	0.0019	0.0336	0.6815	0.0025	0.0399	0.6782

		Mexico			Indonesia			Turkey		
Batch Size	λ	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq
64	0.0001	0.0029	0.0435	0.6384	0.0077	0.0762	0.2514	0.0039	0.0487	0.9268
64	0.0003	0.0121	0.1025	-0.5074	0.0029	0.0492	0.7185	0.0058	0.0575	0.8917
64	0.001	0.0010	0.0248	0.8705	0.0056	0.0580	0.4569	0.0210	0.1116	0.6083
64	0.003	0.0018	0.0326	0.7757	0.0034	0.0474	0.6705	0.0267	0.1250	0.5032
128	0.0001	0.0010	0.0255	0.8791	0.0123	0.0981	-0.1956	0.0026	0.0431	0.9515
128	0.0003	0.0008	0.0236	0.8952	0.0032	0.0464	0.6886	0.0156	0.0963	0.7098
128	0.001	0.0013	0.0298	0.8371	0.0010	0.0246	0.9071	0.0096	0.0739	0.8213
128	0.003	0.0075	0.0749	0.0717	0.0017	0.0322	0.8367	0.0153	0.0961	0.7151
256	0.0001	0.0009	0.0241	0.8863	0.0038	0.0473	0.6344	0.0100	0.0755	0.8147
256	0.0003	0.0080	0.0760	0.0040	0.0013	0.0294	0.8779	0.0029	0.0386	0.9467
256	0.001	**0.0005**	**0.0183**	**0.9349**	0.0035	0.0492	0.6602	**0.0020**	**0.0328**	**0.9621**
256	0.003	0.0019	0.0362	0.7598	**0.0007**	**0.0205**	**0.9328**	0.0053	0.0528	0.9021

Source: Author's estimation.

Table 5.2: The estimation results of NARX

		Brazil			Russia			China			South Africa		
Batch Size	λ	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq
64	0.0001	0.0022	0.0430	0.9144	0.0096	0.0952	-4.3977	**0.0006**	**0.0194**	**0.9015**	0.0005	0.0179	0.9348
64	0.0003	0.0028	0.0429	0.8921	0.0003	0.0151	0.8250	0.0013	0.0297	0.7812	0.0198	0.1313	-1.5874
64	0.001	0.0152	0.1079	0.4162	0.0012	0.0295	0.3336	0.0014	0.0296	0.7605	0.0009	0.0257	0.8818
64	0.003	0.0053	0.0585	0.7964	0.0047	0.0578	-1.6265	0.0036	0.0545	0.3836	0.0031	0.0466	0.5984
128	0.0001	0.0008	0.0224	0.9694	0.0009	0.0233	0.5182	0.0046	0.0583	0.2085	0.0009	0.0221	0.8870
128	0.0003	0.0029	0.0466	0.8883	0.0002	0.0123	0.8599	0.0014	0.0321	0.7665	0.0011	0.0290	0.8524
128	0.001	0.0100	0.0890	0.6147	0.0004	0.0166	0.7869	0.0019	0.0382	0.6671	0.0007	0.0207	0.9063
128	0.003	0.0075	0.0778	0.7135	0.0024	0.0455	-0.3713	0.0012	0.0274	0.8020	0.0010	0.0249	0.8653
256	0.0001	0.0095	0.0752	0.6363	0.0010	0.0259	0.4547	0.0057	0.0664	0.0333	0.0032	0.0450	0.5833
256	0.0003	**0.0006**	**0.0195**	**0.9772**	0.0003	0.0125	0.8465	0.0018	0.0371	0.6866	0.0009	0.0236	0.8878
256	0.001	0.0010	0.0255	0.9627	0.0005	0.0194	0.7103	0.0007	0.0204	0.8869	**0.0004**	**0.0147**	**0.9519**
256	0.003	0.0073	0.0719	0.7200	**0.0002**	**0.0110**	**0.8944**	0.0007	0.0222	0.8739	0.0008	0.0211	0.9006

		Mexico			Indonesia			Turkey		
Batch Size	λ	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq
64	0.0001	0.0007	0.0207	0.9190	0.0013	0.0301	0.8743	**0.0005**	**0.0164**	**0.9904**
64	0.0003	0.0011	0.0273	0.8648	0.0008	0.0228	0.9260	0.0017	0.0338	0.9679
64	0.001	0.0032	0.0511	0.6019	0.0009	0.0242	0.9115	0.0009	0.0220	0.9829
64	0.003	0.0024	0.0440	0.7036	0.0012	0.0281	0.8798	0.0039	0.0462	0.9269
128	0.0001	0.0025	0.0422	0.6909	0.0102	0.0894	0.0065	0.0015	0.0282	0.9729
128	0.0003	0.0031	0.0471	0.6208	0.0010	0.0252	0.9055	0.0006	0.0177	0.9890
128	0.001	0.0007	0.0217	0.9100	0.0041	0.0548	0.6029	0.0014	0.0296	0.9741
128	0.003	**0.0005**	**0.0162**	**0.9395**	0.0009	0.0232	0.9153	0.0050	0.0535	0.9075
256	0.0001	0.0062	0.0630	0.2257	0.0055	0.0629	0.4692	0.0038	0.0468	0.9297
256	0.0003	0.0005	0.0181	0.9341	**0.0006**	**0.0188**	**0.9462**	0.0066	0.0697	0.8763
256	0.001	0.0012	0.0256	0.8508	0.0014	0.0302	0.8621	0.0009	0.0226	0.9842
256	0.003	0.0021	0.0421	0.7385	0.0007	0.0218	0.9283	0.0037	0.0448	0.9310

Source: Author's estimation.

CHAPTER III

Recurrent neural networks (RNN) are usually facing the long-term dependency problem because of their sequence modelling nature. To remind, the vanishing or exploding gradient problem is the difficulty of capturing the long-term dependencies because of the multiplicative gradient that can be exponentially decreasing or increasing with respect to the number of layers (Li and Chen 2014: 1002). This problem arises because of the application of backward propagation and in this case, the learning process from the past events becomes extremely difficult.

In the previous chapter, we studied Elman's RNN and NARX. Elman's RNN is defenseless to the vanishing gradient problem while NARX has a practical solution which is adding the short-time lags and introducing variants of time-delay networks (Lin et al. 1996: 1330). However, there are other ways to overcome the challenge that arise from the backpropagation. The most renowned method is the gating technique which allows the RNN to decide when to forget the input and when to remember it for future time steps. The most popular gating techniques are long short-term memory (LSTM) and gated recurrent unit (GRU). These methods are designed to take into account the long-term dependency and they offer a better prediction architecture compared to Elman's RNN and NARX. Meanwhile, they are able to overcome the vanishing or exploding gradient problem.

There are several studies which predict the stock price index by applying LSTM and these studies verify that the predictions obtained by LSTM outperform the other methods (Nelson, Pereira and Oliveira (2017), Fischer and Krauss (2017), Bao, Yue and Rao (2017), Borovkova and Tsiamas (2018), Namini, Tavakoli and Namin (2018) and Tanoe (2019)). Nelson (2017) predicts BOVESPA index by using 15-minute interval observations and reports 53-55% accuracy for the direction of price. Fischer and Krauss (2017) predict S&P500 index by using daily data and comparing LSTM with random forest and logistic regression. The results show that the most profitable trading strategy obtained by LSTM. W. Bao and Rao (2017) predict daily stock prices from six different geographies by LSTM. Borovkova and Tsiamas (2018) pick 22 large-cap U.S. stocks which are traded either on NYSE or NASDAQ and compare LSTM with LASSO and ridge regression. As a result, LSTM outperforms both LASSO and ridge regression. Namini, Tavakoli and Namini (2018) compare LSTM and ARIMA by predicting the stock price indices. Based on the RMSE values, LTSM outperforms ARIMA. Tanoe (2019) predicts the banking industry stocks by applying LSTM, ARIMA and MLP and LSTM outperforms both.

Some other studies focus on predicting other financial variables, such as Bitcoin price, crude oil price, a bond yield. Chih-Hung et al. (2018) and Tandon, Revankar and Parihar (2019) predict Bitcoin price by using LSTM. They obtain significantly less MAE compared to other methods. Gupta and Pandey (2018) predict crude oil prices by applying LSTM. Nunes et al (2019) predict the 10-year Euro government bond yield by LSTM and LASSO. The results of LSTM are similar or better with the results of LASSO.

Some previous studies focus on prediction by GRU. Shen et al. (2018) predict S&P500, DAX and HIS indices by applying GRU and comparing these results with the Support Vector Machine (SVM). They conclude that the GRU network is suitable to study the financial time series prediction. Lin and Sun (2020) predict crude oil prices by different types of GRU architecture. They find that the GRU with multi-layer architecture outperforms the GRU with single-layer architecture.

Several studies compare the predicting performance of LSTM and GRU (Zahara and Ilmiddaviq (2020), Fu, Zhang and Li (2016), Dutta, Kumar and Basu (2020), Minh et al. (2018); and Persio and Honchar (2017)). Persio and Honchar (2017) compare LSTM and GRU by predicting daily GOOGL stock prices. They find that LSTM reaches a higher level of prediction accuracy than GRU. Minh et al. (2018) predict S&P500 index and the two-stream GRU outperforms the GRU and LSTM. Dutta, Kumar and Basu (2020) predict Bitcoin prices and they argue that the simple architecture of GRU works better than LSTM. Zahara, Sugianto and Ilmiddaviq (2020) predict the consumer price index (CPI) for Indonesia by comparing ARIMA, LSTM and GRU. In average, GRU has reduced the MAE by c.10% and LSTM by c.5% compared to ARIMA.

3.1 Long Short-Term Memory

Long short-term memory (LSTM) is introduced by Hochreiter & Schmidhuber, (1997). The previous recurrent neural networks are not capable of working with long term dependencies and LSTM is designed to solve the problem of long-term dependency. The LSTM network is able to learn the long-term dependencies without losing the short-time lag capabilities. In addition, LSTM introduces memory cells to overcome the vanishing gradient problem and gate units to provide a constant error flow through the internal state, instead of vanishing or exploding. The LSTM network consists of three gates, which are input, forget and output gate. Besides the three aforementioned gates, LSTM introduces a cell state which keeps the entire data that has been processed so far. The cell state keeps the long-term memory and it is controlled and modified by the input and the forget gate. The input gate i_t in time period t is;

$$i_t = \sigma_i(W_i x_t + U_h h_{t-1} + b_i)$$

The input gate is a sigmoid function (σ) with a range of (0,1). This function is only able to add memory, it cannot remove or forget the memory since it is never going to be equal to zero. Remember from the previous chapter, Elman's RNN has a *tanh* activation function with a range of (-1,1) which allows the network to forget the memory. At this gate, the input vector x_t is plugged into the function and it is multiplied by corresponding weight *W*, and the information from the previous layer h_{t-1} is also multiplied by its weight and the corresponding bias *b* are added. This process lets the network decide the amount of new information going to be added. At this stage, the network only focuses on the information received from the current input X_t and the short-term memory of the previously hidden layer h_{t-1}.

$$f_t = \sigma_f(W_f x_t + U_h h_{t-1} + b_f)$$

The forget gate f_t is also a sigmoid function with a range of (0,1). At this step, the network decides which part of the long-term memory should be kept or neglected for the next step. Both input and forget gates are selective steps of the network.

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t$$

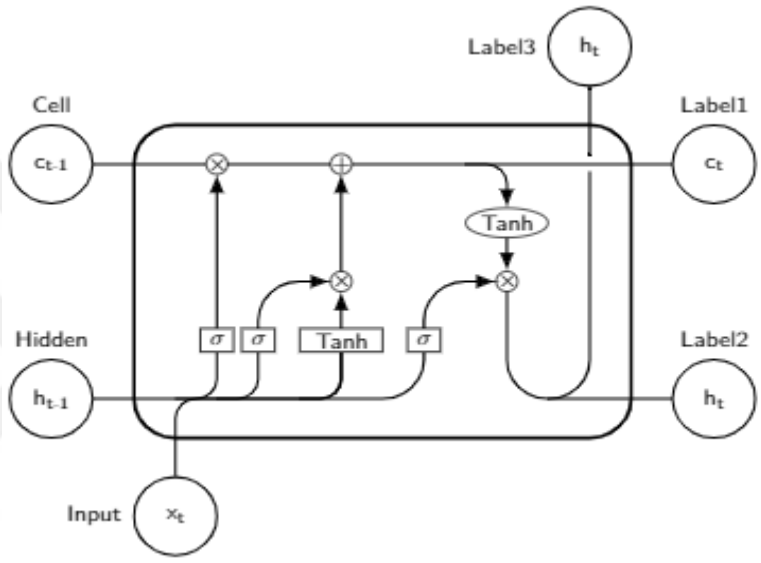
The cell state c_t is able to encode the information by the weights of the input gate and the forget gate, and these weights determine which information is important or not. The information to be forgotten is multiplied by the forget gate and new information is added by the input gate. In other words, the cell decides to keep or override the information in the cell state. For example, if the forget gate is 1, the information is continued to be kept in the cell state, and it is going to be forgotten if the forget gate is 0.

$$o_t = \sigma_o(W_o x_t + U_h h_{t-1} + b_o)$$

$$h_t = o_t \odot tanh(c_t)$$

The output gate o_t is also a sigmoid function with a range of (0.1). At this step, the gate estimates the new short-term memory in the hidden state h_t which is going to be passed to the next layer. Therefore, the hidden state decides what information should be taken to the next sequence.

Figure 6.1: The architecture of Long Short-Term Memory



Source: Author's drawing.

3.2 Gated Recurrent Unit

Gated recurrent unit is introduced by Cho et al. (2014) to overcome the vanishing gradient problem. The network introduces two gates which are an update gate and a reset gate. These gates are aimed to determine which information should be carried to the output. These gates are able to keep information from the beginning of the trained dataset and do not remove irrelevant information. The update gate z_t and reset gate r_t in time period t are;

$$z_t = \sigma_z(W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma_r(W_r x_t + U_r h_{t-1} + b_r)$$

Both update and reset gate are sigmoid function (σ) with a range of (0,1). This function is only able to add memory, it can't remove or forget the memory since it is never going to be zero. The input vector x_t is plugged into the function and it is multiplied by corresponding weight *W*, and the information from the previous layer h_{t-1} is also multiplied by its weight and the corresponding bias *b* are added. The update gate decides how much of the information from previous layers worth remembering to predict future, in other words, it is responsible from the long memory. This process eliminates the problem of vanishing gradient. On the other side, the reset gate decides how much of the past information has to be forgotten from previous layers, in other words, it is responsible for the short memory. The candidate activation vector $\hat{\mathrm{h}}_t$ contains the current memory which is going to be used by the reset gate to keep the relevant past information. It is estimated as;

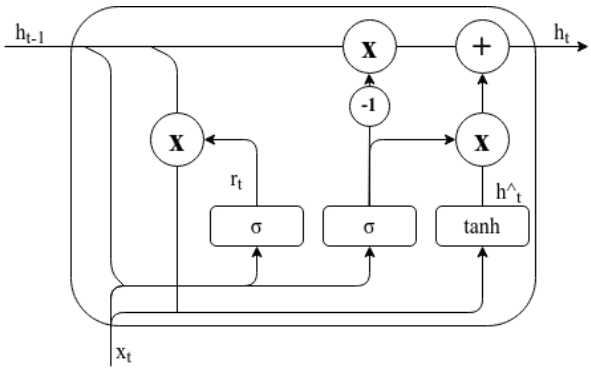
$$\hat{\mathrm{h}}_t = tanh_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

The Hadamart (element-wise) product is denoted by $\odot$ and estimated between the reset gate r_t and the information from the previous layer h_{t-1}. This step decides what information need to be eliminated from the previous hidden layer. In case of facing irrelevant information, the network assigns a number close to 0 to wipe out. If it is 1, it means that all information from the previously hidden layer has to be considered. Lastly, the output vector h_t is estimated holds the memory from the past step and the current memory is estimated by candidate activation vector. In this step, the updated gate z_t determines the amount of information to collect from the previously hidden layer and the candidate activation vector. The output vector h_t is;

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{\mathrm{h}}_t$$

To sum up, GRU is capable of storing and filtering the information through the update and the reset gates. It overcomes the problem of vanishing gradient by only keeping and passing the related information to the next layer.

Figure 6.2: The architecture of Gated Recurrent Unit

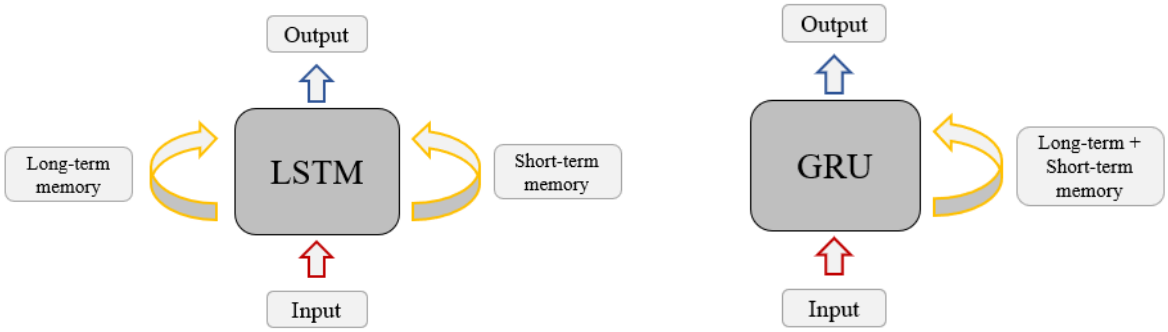


Source: Author's drawing.

3.3 The differences of GRU and LSTM

- The first difference is that GRU network operates with fewer gates compared to LSTM. LSTM has three gates; input, output and forget gate while GRU has two gates; update and reset gate. The simpler design of GRU makes it faster to train because it has fewer gates. But this doesn't mean that the GRU network outweighs LSTM. Sometimes having simplicity or less parameters might decrease the expressiveness of the model.
- Second, GRU doesn't have a separate memory cell or hidden state like LSTM does. GRU keeps both short- and long-term information in the same hidden state while LSTM separates this process by keeping long term dependencies in the cell state and short-term memory in the hidden state.
- Third, the process of choosing information from the previous hidden state to sustain and the process of controlling the new information to be added is controlled by two different gates (input and forget) is a separate process in LSTM network, however, this process runs by the same gate (update gate) in GRU network.

Figure 6.3: The workflow of LSTM and GRU



Source: Author's drawing.

3.4 Empirical Findings

In this section, the results from both networks, LSTM and GRU are discussed for BRICS and MINT countries. Both networks are run with the same settings in chapter 2. We compared the results by the root mean squared error (RMSE), the mean absolute error (MAE) and the R-squared value in the table 6.1 and 6.2. The distributions of the twelve different outcomes for all networks are illustrated by boxplot in the appendix A3-9.

The results obtained in this chapter is more diversified compared to the results in the previous chapter. GRU outperforms LSTM network for Brazil and Indonesia. LSTM network outperforms GRU in Russia, South Africa, Mexico and Turkey. Both networks perform similarly for China. The best prediction accuracy for Brazil, is obtained by employing 0.0001 λ and 125 batch size in GRU, where RMSE and MAE are estimated 0.0027and 0.0423respectively with 89.6 % R-squared. The best prediction accuracy for Russia, is obtained by employing 0.001 λ and 256 batch size in LSTM, where RMSE and MAE are estimated 0.0001 and 0.0095 respectively with 92.1 % R-squared. The best prediction accuracy for China, is obtained by employing 0.0003 λ and 64 batch size in LSTM, where RMSE and MAE are estimated 0.0005 and 0.0174 respectively with 91.6 % R-squared. The best prediction accuracy for South Africa, is obtained by employing 0.001 λ and 64 batch size in LSTM, where RMSE and MAE are estimated 0.0007 and 0.0224 respectively with 90.3 % R-squared. The best prediction accuracy for Mexico, is obtained by employing 0.0001 λ and 128 batch size in LSTM, where RMSE and MAE are estimated 0.0004 and 0.0164 respectively with 94.4 % R-squared. The best prediction accuracy for Indonesia, is obtained by employing 0.001 λ and 256 batch size in GRU, where RMSE and MAE are estimated 0.0006 and 0.0199 respectively with 93.4 % R-squared. The best prediction accuracy for Turkey, is obtained by employing 0.0001 λ and 64 batch size in LSTM, where RMSE and MAE are estimated 0.0006 and 0.0183 respectively with 98.8 % R-squared.

Table 6.1: The estimation results of LSTM

		Brazil			Russia			China			South Africa		
Batch Size	λ	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq
64	0.0001	0.0046	0.0501	0.8218	0.0007	0.0229	0.6217	0.0019	0.0356	0.6787	0.0010	0.0251	0.8717
64	0.0003	0.0146	0.1016	0.4395	0.0006	0.0221	0.6375	**0.0005**	**0.0174**	**0.9164**	0.0040	0.0577	0.4740
64	0.001	0.0487	0.1905	-0.8737	0.0035	0.0557	-0.9685	0.0034	0.0519	0.4216	**0.0007**	**0.0224**	**0.9027**
64	0.003	0.0814	0.2475	-2.1287	0.0055	0.0567	-2.1056	0.0063	0.0652	-0.0685	0.0068	0.0646	0.1053
128	0.0001	0.0068	0.0694	0.7376	0.0003	0.0145	0.8156	0.0011	0.0275	0.8159	0.0029	0.0439	0.6227
128	0.0003	0.0072	0.0676	0.7228	0.0002	0.0109	0.8874	0.0016	0.0330	0.7222	0.0012	0.0258	0.8386
128	0.001	0.0258	0.1397	0.0096	0.0004	0.0164	0.7624	0.0031	0.0503	0.4667	0.0025	0.0410	0.6775
128	0.003	0.0726	0.2359	-1.7928	0.0011	0.0253	0.3576	0.0063	0.0738	-0.0825	0.0041	0.0502	0.4683
256	0.0001	0.0362	0.1531	-0.3903	0.0004	0.0165	0.7639	0.0200	0.1264	-2.4179	0.0023	0.0352	0.7020
256	0.0003	**0.0042**	**0.0494**	**0.8385**	0.0003	0.0133	0.8458	0.0021	0.0358	0.6388	0.0010	0.0241	0.8693
256	0.001	0.0062	0.0641	0.7633	**0.0001**	**0.0095**	**0.9214**	0.0012	0.0302	0.7871	0.0024	0.0433	0.6835
256	0.003	0.0402	0.1734	-0.5456	0.0006	0.0199	0.6873	0.0021	0.0409	0.6471	0.0020	0.0352	0.7448

		Mexico			Indonesia			Turkey		
Batch Size	λ	RMSE	MAE	R-sq	RMSE	MAE	R-sq	RMSE	MAE	R-sq
64	0.0001	0.0007	0.0210	0.9175	0.0013	0.0294	0.8690	**0.0006**	**0.0183**	**0.9885**
64	0.0003	0.0014	0.0310	0.8221	0.0013	0.0285	0.8703	0.0021	0.0354	0.9601
64	0.001	0.0031	0.0467	0.6169	0.0024	0.0417	0.7623	0.0050	0.0478	0.9071
64	0.003	0.0023	0.0382	0.7204	0.0046	0.0583	0.5511	0.0131	0.0814	0.7553
128	0.0001	**0.0004**	**0.0164**	**0.9444**	0.0020	0.0333	0.8093	0.0012	0.0265	0.9778
128	0.0003	0.0015	0.0320	0.8124	0.0013	0.0283	0.8736	0.0009	0.0220	0.9825
128	0.001	0.0011	0.0265	0.8640	0.0014	0.0311	0.8646	0.0025	0.0342	0.9538
128	0.003	0.0024	0.0399	0.7068	0.0026	0.0422	0.7517	0.0143	0.0815	0.7345
256	0.0001	0.0018	0.0353	0.7736	0.0080	0.0714	0.2215	0.0062	0.0603	0.8855
256	0.0003	0.0026	0.0437	0.6827	0.0021	0.0370	0.7974	0.0010	0.0240	0.9817
256	0.001	0.0005	0.0173	0.9359	**0.0008**	**0.0227**	**0.9179**	0.0018	0.0287	0.9662
256	0.003	0.0010	0.0273	0.8701	0.0057	0.0677	0.4455	0.0041	0.0447	0.9240

Source: Author's estimation.

Table 6.2: The estimation results of GRU

		Brazil			Russia			China			South Africa		
Batch Size	λ	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2
64	0.0001	0.0099	0.0867	0.6183	0.0010	0.0284	0.4218	**0.0005**	0.0183	0.9109	0.0013	0.0273	0.8340
64	0.0003	0.0181	0.1169	0.3058	0.0018	0.0391	-0.0392	0.0041	0.0586	0.3016	0.0013	0.0290	0.8263
64	0.001	0.0521	0.2039	-1.0017	0.0025	0.0473	-0.3973	0.0014	0.0295	0.7667	0.0019	0.0371	0.7490
64	0.003	0.0834	0.2550	-2.2083	0.0033	0.0436	-0.8580	0.0042	0.0543	0.2850	0.0058	0.0600	0.2442
128	0.0001	**0.0027**	**0.0423**	**0.8958**	**0.0003**	**0.0129**	**0.8542**	0.0020	0.0388	0.6539	0.0051	0.0639	0.3340
128	0.0003	0.0110	0.0890	0.5777	0.0003	0.0154	0.8063	0.0007	0.0213	0.8880	0.0014	0.0300	0.8174
128	0.001	0.0333	0.1625	-0.2797	0.0003	0.0138	0.8166	0.0030	0.0476	0.4952	**0.0011**	**0.0242**	**0.8587**
128	0.003	0.0742	0.2392	-1.8523	0.0018	0.0389	-0.0263	0.0056	0.0682	0.0393	0.0041	0.0508	0.4643
256	0.0001	0.0327	0.1548	-0.2555	0.0006	0.0190	0.6817	0.0087	0.0818	-0.4841	0.0017	0.0349	0.7824
256	0.0003	0.0032	0.0500	0.8766	0.0005	0.0167	0.7407	0.0008	0.0233	0.8687	0.0018	0.0332	0.7695
256	0.001	0.0079	0.0773	0.6958	0.0007	0.0212	0.6130	**0.0005**	**0.0174**	**0.9176**	0.0086	0.0892	-0.1286
256	0.003	0.0342	0.1632	-0.3155	0.0020	0.0419	-0.1240	0.0061	0.0722	-0.0387	0.0011	0.0281	0.8560

		Mexico			Indonesia			Turkey		
Batch Size	λ	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2
64	0.0001	0.0007	0.0214	0.9128	0.0012	0.0290	0.8844	0.00100	0.02390	0.98150
64	0.0003	0.0021	0.0394	0.7401	0.0008	0.0209	0.9261	0.00450	0.04750	0.91710
64	0.001	0.0007	0.0227	0.9080	0.0017	0.0340	0.8364	0.00610	0.05410	0.88550
64	0.003	0.0023	0.0367	0.7202	0.0055	0.0642	0.4612	0.01560	0.08900	0.70870
128	0.0001	0.0008	0.0226	0.9034	0.0007	**0.0194**	0.9311	**0.00080**	0.02060	0.98420
128	0.0003	0.0007	0.0226	0.9102	0.0009	0.0238	0.9128	0.00120	0.02310	0.97720
128	0.001	0.0010	0.0247	0.8810	0.0017	0.0310	0.8331	0.01330	0.08780	0.75200
128	0.003	0.0013	0.0303	0.8365	0.0031	0.0459	0.6952	0.01790	0.09810	0.66690
256	0.0001	0.0018	0.0326	0.7769	0.0082	0.0743	0.2013	0.00090	0.02310	0.98320
256	0.0003	0.0034	0.0496	0.5764	0.0011	0.0250	0.8950	**0.00080**	0.02200	**0.98430**
256	0.001	**0.0007**	**0.0204**	**0.9182**	**0.0006**	0.0199	**0.9383**	0.00090	**0.01990**	0.98410
256	0.003	0.0025	0.0406	0.6924	0.0008	0.0238	0.9194	0.01640	0.09880	0.69490

Source: Author's estimation

The methodology applied in this dissertation let us examine the importance of features that are more critical in predicting CDS risk premiums in the selected countries. Shapley Additive Explanations (SHAP) values are telling us how much each feature contributes to output either positively or negatively, and are presented in Figure 6.4 and 6.5. The variables are ranked in descending order by their importance. Based on the location of pink color dots, we are able to decide whether the feature has a positive or negative impact on CDS risk premiums.

For Brazil's CDS risk premium, the local features, the 5-year maturity government bond rate and the yield curve are played a more critical role. The pink area is located in the negative side of x-axis for the yield curve of Brazil, which means that the yield curve has negatively correlated with CDS risk premiums for Brazil. For the 5-year maturity government bond rate, the distribution of the pink colored observation a bit ambiguous but since the area is thicker on the negative side, it is more negatively correlated with CDS risk premiums. Both the corporate spread, which shows the capital reallocation and the liquidity shifts between the financial markets and the exchange rate are positively related while the rest of the variables have a relatively smaller impact on CDS risk premium prediction. For Russia's CDS risk premium, the 5-year maturity government bond rate takes the lead and it is followed by Nasdaq and the exchange rate. The DAX index takes place near to top in Russia. It may because of the German investment into Russia which has been reached the highest levels for the past decade. Although Russia's dependence on oil market is very high and it is a net exporter country, the effect of oil on CDS risk premiums is lower than expected. For China's CDS risk premium, Nasdaq is ranked at the top for China and it is not surprising since it is a well-known fact that the giant U.S. tech companies are interested in investing in China. Even though the 5-year maturity government bond rate ranked near the top in all selected countries, it falls back in China. However, instead of the government bond rate, the exchange rate has a high and positive impact on China's CDS prediction. The global features, Nasdaq, DAX, the 5-year constant maturity U.S. treasury yield, the German yield curve and VIX index have a noteworthy impact on China's CDS risk premiums. For South Africa's CDS risk premium, the oil price is the most important feature, which shows that increased uncertainty in the oil market will affect the CDS market more. The oil price is followed by the local features, the exchange rate and the 5-year maturity government bond rate.

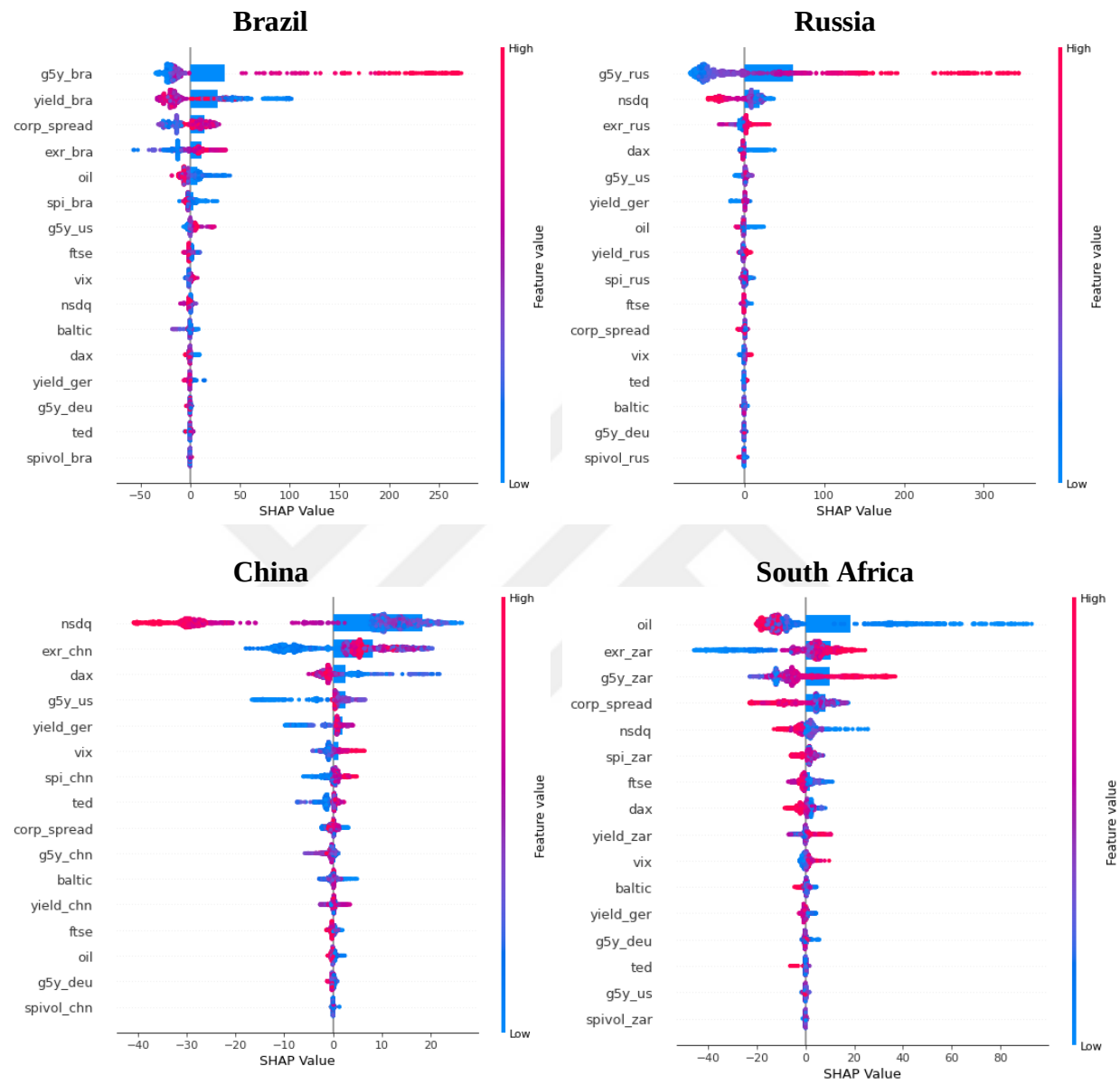
For Mexico's CDS risk premium, the oil price is the most important feature for Mexico, where it is negatively related to CDS risk premiums. Mexico, an oil exporter country, will face a risk of debt crisis more than other countries in case of declining oil prices and this leads to an increase in CDS risk premiums. But the oil price is also the most important feature for Indonesia even though it is an oil-importer country, and oil price didn't come forward for Russia, since it is one of the biggest oil-exporter countries in the global economy. Our findings show similarity with the study of Wang et al. (2020). Therefore, we conclude

that the impact of oil prices on countries varies according to the economic stability of each country, rather than being a net oil-exporter or importer. The negative relationship between CDS risk premiums and oil price are also confirmed by Naifar, Shahzad and Hammoudeh (2020) and Sabhka, Perettia and Hmaied (2019). The oil price is followed by the corporate spread, the 5-year maturity government bond rate, and the exchange rate. In addition, DAX index ranked near the top in Mexico, because Germany is one of the main investing countries in Mexico and the biggest trading partner in The European Union. Germany's investment in Mexico corresponds to near 9% of the total FDI flow to Mexico in 2019.[3] For Indonesia's CDS risk premium, Nasdaq Index took place at the top in predicting CDS in Indonesia. It is a well-known fact that giant U.S. technology companies are interested in investing in Indonesia in recent years. The Indonesian government has been improving the investment environment to attract more U.S. companies to establish in Indonesia through regulatory reforms such as a job creation law under the Omnibus law. Thus, the spillover effect from the U.S. is quite important for the Indonesian CDS market. Nasdaq Index is followed by the local features, such as the yield curve, the stock price index, and the 5-year maturity government bond rate. For Turkey's CDS risk premium, unsurprisingly, the exchange rate is the most important variable for Turkey, which has a long history of currency crisis. One of the major studies that link sovereign debt defaults and currency crises is Reinhart (2002), which states that sovereign default events are mostly followed by a currency crisis. The exchange rate is followed by the 5-year maturity government bond rate, the corporate rate and the stock price index. One can argue that Turkey's CDS risk premium is affected mostly by local developments. Based on the instability in the political and economic environment in Turkey, the outcome is quite significant and foreseeable.

Overall, the selected countries are affected by both the local and global features. The 5-year maturity government bond rate is ranked at the top for Brazil and Russia, and it takes place at the top for other countries except for China. The exchange rate is only ranked at top for Turkey, but it is front for other countries except for Indonesia. The spillover effect from the global economy is non-negligible for all countries. Especially, Nasdaq plays an important role for China, Russia and Indonesia. The oil price is essential for predicting CDS risk premiums of South Africa and Mexico. Brazil and Turkey seem to be affected more by local factors compared to others.

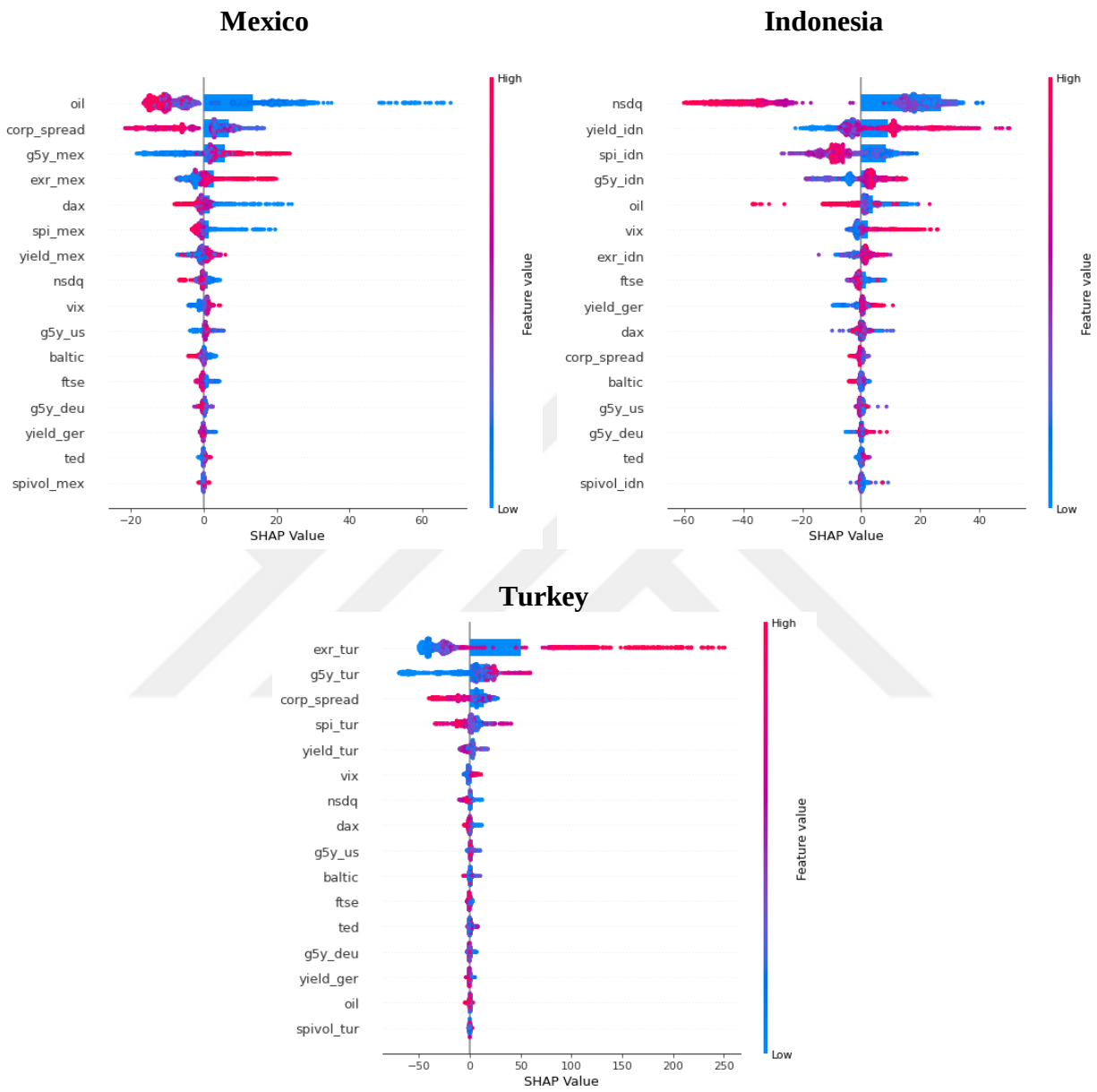
[3] According to the statistics of Secretariat of Economy of Mexico.

Figure 6.4a: The feature importance for BRICS countries



Source: Author's estimation

Figure 6.4b: The feature importance for MINT countries



Source: Author's estimation

7. CONCLUSION

This dissertation investigates various methodologies to predict CDS risk premiums in BRICS and MINT countries. The daily data is collected from 03/01/2012 and ends on 29/11/2019 to avoid COVID-19 effects on the state of the global economy. CDS risk premium for 5-year maturity contracts is used as an output variable that is going to be predicted since it faces the highest liquidity in the CDS market compared to 1-year and 10-year maturity CDS contracts. The features are divided into local variables and global variables. Local variables consist of risk-free rate, the slope of the yield curve, the liquidity, the daily change of the stock price index, the daily change of bilateral exchange rate against the U.S Dollar, and the daily change of stock price index volume. Global features are the 5-year constant maturity U.S. treasury yield, the spread of Moody's BAA and AAA corporate bond yield, the daily return of Nasdaq, DAX, FTSE indexes, TED spread, CBOE VIX index, oil price and the Baltic Dry Index.

First chapter focuses on the two most known regularization techniques, Ridge regression and LASSO (Least Absolute Shrinkage and Selection Operator) to predict CDS risk premiums. The regularization techniques perform a subset selection by its nature and it excludes or reduces the effect of irrelevant variables on the regression output. By doing so, it reduces the model complexity and prevents the over-fitting problems which arise from the linear regression assumption. In terms of higher R-squared obtained, LASSO outperformed Ridge regression in all selected countries. This suggests that the subset selection property of LASSO works better for our data sample. In terms of the RMSE and the MAE scores, the lowest values are obtained by ridge regression for all selected countries. From this perspective, Ridge regression, which keeps all features in the model, results better. The results suggest that both models reached strong prediction accuracy and the selected features have a powerful predictive ability. The subset selection doesn't play a crucial role in terms of RMSE and MAE scores but it helps to increase the R-squared.

Second chapter focuses on predicting CDS risk premiums by two basic type of recurrent neural network which are Elman's RNN and NARX. Elman's RNN is a two-layer network which are the input and the hidden layer. Elman's RNN is defenseless to the vanishing/exploding gradient problem while the nonlinear autoregressive network with exogenous inputs (NARX) has a practical solution by adding the lags of the features and by doing so it considers the short-term dependencies. In all countries except China, the results of NARX outweigh Elman's RNN. NARX network is only able to capture a limited number of

long-term dependencies. However, the results obtained in this chapter suggest that the long-term dependencies play a more important role for all countries except China.

Third chapter introduces the gating technique to overcome the challenge arises from the backpropagation. The gating technique allows the RNN to decide when to forget the input and when to remember it for future time steps. The most popular gating techniques are LSTM and GRU. These techniques are designed to take into account the long-term dependency. GRU outperforms LSTM network for Brazil and Indonesia. LSTM network outperforms GRU in Russia, South Africa, Mexico and Turkey. Both networks perform similarly for China. SHAP values let us see how much each feature contributes to output either positively or negatively. The selected countries are affected by both the local and global features. The spillover effect from the global economy is non-negligible for all countries.

REFERENCES

Acharya, V., and Johnson, T. (2007). Insider trading in credit derivatives. Journal of Financial Economics, 84(1), 110–141. https://doi.org/10.1016/j.jfineco.2006.05.003.

Alexandre, H., and de Benoist, A. (2012). Oil Prices and Government Bond Risk Premiums. Lahore Journal of Business, 1(1), 1–21. https://doi.org/10.35536/ljb.2012.v1.i1.a1.

Alper, C. E., Forni, L., and Gerard, M. (2012). Pricing of Sovereign Credit Risk: Evidence From Advanced Economies During the Financial Crisis. IMF Working Papers, 12(24). https://doi.org/10.5089/9781463931865.001.

Augustin, P. (2018). The term structure of CDS spreads and sovereign credit risk. Journal of Monetary Economics, 96, 53–76. https://doi.org/10.1016/j.jmoneco.2018.04.001.

Avino, D., and Nneji, O. (2014). Are CDS spreads predictable? An analysis of linear and non-linear forecasting models. International Review of Financial Analysis, 34, 262-274. https://doi.org/10.1016/j.irfa.2014.04.001.

Bao, W., Yue, J., and Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one, 12(7), e0180944. https://doi.org/10.1371/journal.pone.0180944.

Bengio, Y., Frasconi, P., and Simard, P. (1994). The problem of learning long-term dependencies in recurrent networks. In IEEE International Conference on Neural Networks, 1183-1188. https://doi.org/10.1109/icnn.1993.298725.

Blanco, R., Brennan, S., and Marsh, I. W. (2005). An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. The Journal of Finance, 60(5), 2255-2281. https://doi.org/10.1111/j.1540-6261.2005.00798.x.

Blommestein, H., Eijffinger, S., and Qian, Z. (2016). Regime-dependent determinants of Euro area sovereign CDS spreads. Journal of Financial Stability, 22, 10-21. https://doi.org/10.1016/j.jfs.2015.11.004.

Borovkova, S., and Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. Journal of Forecasting, 38(6), 600–619. https://doi.org/10.1002/for.2585.

Calice, G., and Zeng, M. (2019). The term structure of sovereign credit default swap and the cross-section of exchange rate predictability. International Journal of Finance and Economics, 26(1), 445–458. https://doi.org/10.1002/ijfe.1798.

Calice, G., Mio, R., Štěrba, F., and Vašíček, B. (2015). Short-term determinants of the idiosyncratic sovereign risk premium: A regime-dependent analysis for European credit default swaps. Journal of Empirical Finance, 33, 174-189. https://doi.org/10.1016/j.jempfin.2015.03.018.

Carr, P., and Wu, L. (2009). Variance risk premiums. The Review of Financial Studies, 22(3), 1311-1341. https://doi.org/10.1093/rfs/hhn038.

Chan, K. C., Fung, H. G., and Zhang, G. (2009). On the relationship between Asian credit default swap and equity markets. Journal of Asia Business Studies.https://doi.org/10.1108/15587890980000414.

Chandra, R., and Zhang, M. (2012). Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction. Neurocomputing, 86, 116-123. https://doi.org/10.1016/j.neucom.2012.01.014.

Chan-Lau, J. A., and Kim, Y. S. (2004). Equity Prices, Credit Default Swaps, and Bond Spreads in Emerging Markets. IMF Working Papers, 04(27). https://doi.org/10.5089/9781451844559.001.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). https://doi.org/10.3115/v1/d14-1179.

Choudhry, M. (2013). Credit Derivatives. An Introduction to Credit Derivatives, 65–85. https://doi.org/10.1016/b978-0-08-098295-3.00004-8.

Collin-Dufresne, P., and Solnik, B. (2001). On the term structure of default premia in the swap and LIBOR markets. The Journal of Finance, 56(3), 1095-1115. https://doi.org/10.1111/0022-1082.00357.

Dai, W., and Serletis, A. (2018). Oil Price Shocks and the Credit Default Swap Market. Open Economies Review, 29(2), 283-293. https://doi.org/10.1007/s11079-017-9454-z.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?. Journal of Econometrics, 146(2), 318-328. https://doi.org/10.1016/j.jeconom.2008.08.011.

Della Corte, P., Kozhan, R., and Neuberger, A. (2021). The cross-section of currency volatility premia. Journal of Financial Economics, 139(3), 950-970. https://doi.org/10.1016/j.jfineco.2020.08.010.

Di Persio, L., and Honchar, O. (2017). Recurrent neural networks approach to the financial forecast of Google assets. International journal of Mathematics and Computers in simulation, 11, 7-13. https://doi.org/10.1016/j.procs.2020.03.049.

Duffie, D. (1999). Credit swap valuation. Financial Analysts Journal, 55(1), 73-87. https://doi.org/10.2469/faj.v55.n1.2243.

Duffie, D., Pedersen, L. H., and Singleton, K. J. (2003). Modeling sovereign yield spreads: A case study of Russian debt. The Journal of Finance, 58(1), 119-159. https://doi.org/10.1111/1540-6261.00520.

Dutta, A., Kumar, S., and Basu, M. (2020). A gated recurrent unit approach to bitcoin price prediction. Journal of Risk and Financial Management, 13(2), 23. https://doi.org/10.3390/jrfm13020023.

Eyssell, T., Fung, H. G., and Zhang, G. (2013). Determinants and price discovery of China sovereign credit default swaps. China Economic Review, 24, 1-15. https://doi.org/10.1016/j.chieco.2012.09.003.

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. https://doi.org/10.1016/j.jbankfin.2012.06.017.

Fischer, T., and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054.

Fontana, A., and Scheicher, M. (2016). An analysis of euro area sovereign CDS and their relation with government bonds. Journal of Banking and Finance, 62, 126-140. https://doi.org/10.1016/j.jbankfin.2015.10.010.

Fu, R., Zhang, Z., and Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, pp. 324-328, IEEE. https://doi.org/10.1109/YAC.2016.7804912.

Guo, B., and Newton, D. (2013). Regime-dependent liquidity determinants of credit default swap spread changes. Journal of Financial Research, 36(2), 279-298. https://doi.org/10.1111/j.1475-6803.2013.12011.x.

Gupta, V., and Pandey, A. (2018). Crude oil price prediction using LSTM networks. International Journal of Computer and Information Engineering, 12(3), 226-230. https://doi.org/10.5281/zenodo.1316680.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). Unsupervised Learning. *Springer Series in Statistics*, 1–101. https://doi.org/10.1007/b94608_14.

Ho, S. H. (2016). Long and short-runs determinants of the sovereign CDS spread in emerging countries. Research in International Business and Finance, 36, 579-590. https://doi.org/10.1016/j.ribaf.2015.07.001.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), https://doi.org/1735-1780. 10.1162/neco.1997.9.8.1735

Hochreiter, S., Bengio, Y., Frasconi, P., and Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.A Field Guide to Dynamical Recurrent Networks. https://doi.org/10.1109/9780470544037.ch14.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67. https://doi.org/10.1080/00401706.1970.10488634.

Hull, J. C., and White, A. D. (2000). Valuing credit default swaps I: No counterparty default risk. The Journal of Derivatives, 8(1), 29-40. https://doi.org/10.3905/jod.2000.319115.

Hull, J., Predescu, M., and White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. Journal of Banking and Finance, 28(11), 2789-2811. https://doi.org/10.2139/ssrn.2173171.

Kassimatis, E. C. K. (2015). Macroeconomic Effects on Emerging-Markets. Economic Review, 79, 43-50. http://dx.doi.org/10.1016/j.jfs.2015.06.002.

Kim, K. (2017). Liquidity basis between credit default swaps and corporate bonds markets. International Review of Economics and Finance, 48, 98-115. https://doi.org/10.1016/j.iref.2016.11.013.

Lafuente, J. A., and Serrano, P. (2015). On the compensation for illiquidity in sovereign credit markets. Journal of Multinational Financial Management, 30, 83-100. https://doi.org/10.1016/j.mulfin.2015.03.003.

Lee, J. H., Shi, Z., and Gao, Z. (2021). On LASSO for predictive regression. Journal of Econometrics. https://doi.org/10.1016/j.jeconom.2021.02.002.

Li, J., and Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. International Journal of Forecasting, 30(4), 996-1015. https://doi.org/10.1016/j.ijforecast.2014.03.016.

Lien Minh, D., Sadeghi-Niaraki, A., Huy, H. D., Min, K., and Moon, H. (2018). Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. IEEE Access, 6, 55392–55404. https://doi.org/10.1109/access.2018.2868970.

Lin, H., and Sun, Q. (2020). Crude oil prices forecasting: an approach of using CEEMDAN-based multi-layer gated recurrent unit networks. Energies, 13(7), 1543. https://doi.org/10.3390/en13071543 .

Lin, T., Horne, B. G., Tino, P., and Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks, 7(6), 1329-1338. https://doi.org/10.1109/72.548162.

Litterman, R. B., Scheinkman, J., and Weiss, L. (1991). Volatility and the Yield Curve. The Journal of Fixed Income, 1(1), 49–53. https://doi.org/10.3905/jfi.1991.692346.

Liu, Y., and Morley, B. (2012). Sovereign credit default swaps and the macroeconomy. Applied Economics Letters, 19(2), 129-132. https://doi.org/10.1080/13504851.2011.568390.

Liu, Y., and Morley, B. (2013). Sovereign credit ratings, the macroeconomy and credit default swap spreads. Brussels economic review, 56(3/4), 335-348.

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. https://doi.org/10.3386/w13658.

Ma, J. Z., Deng, X., Ho, K. C., and Tsai, S. B. (2018). Regime-switching determinants for spreads of emerging markets sovereign credit default swaps. Sustainability, 10(8), 2730. https://doi.org/10.3390/su10082730.

Marcjasz, G., Uniejewski, B., and Weron, R. (2020). Beating the Naïve—Combining LASSO with Naïve Intraday Electricity Price Forecasts. Energies, 13(7), 1667. https://doi.org/10.3390/en13071667.

McGuire, P., and Schrijvers, M. (2003). Common factors in emerging market spreads. BIS Quarterly Review, December.

Melkumova, L. E., and Shatskikh, S. Y. (2017). Comparing Ridge and LASSO estimators for data analysis. Procedia engineering, 201, 746-755. https://doi.org/10.1016/j.proeng.2017.09.615.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance, 29(2), 449-470. https://doi.org/10.1016/0006-8993(74)90991-3.

Naifar, N., Shahzad, S. J. H., and Hammoudeh, S. (2020). Dynamic nonlinear impacts of oil price returns and financial uncertainties on credit risks of oil-exporting countries. Energy Economics, 88, 104747. https://doi.org/10.1016/j.eneco.2020.104747.

Nelson, D. M., Pereira, A. C., and de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In 2017 International joint conference on neural networks, 1419-1426. https://doi.org/10.1109/IJCNN.2017.7966019.

Nunes, M., Gerding, E., McGroarty, F., and Niranjan, M. (2019). The Memory Advantage of Long Short-Term Memory Networks for Bond Yield Forecasting. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3415219.

Pan, J., and Singleton, K. J. (2008). Default and recovery implicit in the term structure of sovereign CDS spreads. The Journal of Finance, 63(5), 2345-2384. https://doi.org/10.1111/j.1540-6261.2008.01399.x.

Panagiotidis, T., Stengos, T., and Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. Finance Research Letters, 27, 235-240. https://doi.org/10.1016/j.frl.2018.03.016.

Pavlova, I., De Boyrie, M. E., and Parhizgari, A. M. (2018). A dynamic spillover analysis of crude oil effects on the sovereign credit risk of exporting countries. The Quarterly Review of Economics and Finance, 68, 10-22. https://doi.org/10.1016/j.qref.2018.03.003.

Pereira, J. M., Basto, M., and da Silva, A. F. (2016). The logistic lasso and ridge regression in predicting corporate failure. Procedia Economics and Finance, 39, 634-641. https://doi.org/10.1016/S2212-5671(16)30310-0.

Pisoni, E., Farina, M., Carnevale, C., and Piroddi, L. (2009). Forecasting peak air pollution levels using NARX models. Engineering Applications of Artificial Intelligence, 22(4-5), 593-602. https://doi.org/10.1016/j.engappai.2009.04.002.

Pisoni, E., Farina, M., Pagani, G., and Piroddi, L. (2011). Environmental Over-Threshold Event Forecasting using NARX Models. IFAC Proceedings Volumes, 44(1), 10559-10564. https://doi.org/10.3182/20110828-6-IT-1002.01327.

Reinhart, C. M. (2002). Default, currency crises, and sovereign credit ratings. The World Bank Economic Review, 16(2), 151-170. https://doi.org/10.1093/wber/16.2.151.

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. https://doi.org/10.1002/ijfe.357.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536. https://doi.org/10.1038/323533a0.

Sabkha, S., De Peretti, C., and Hmaied, D. M. (2019). International risk spillover in sovereign credit markets: an empirical analysis, Managerial Finance. https://doi.org/10.1108/MF-11-2017-0490

Sermpinis, G., Tsoukas, S., and Zhang, P. (2018). Modelling market implied ratings using LASSO variable selection techniques. Journal of Empirical Finance, 48, 19-35. https://doi.org/10.1016/j.jempfin.2018.05.001.

Shahzad, S. J. H., Naifar, N., Hammoudeh, S., and Roubaud, D. (2017). Directional predictability from oil market uncertainty to sovereign credit spreads of oil-exporting countries: Evidence from rolling windows and crossquantilogram analysis. Energy Economics, 68, 327-339. https://doi.org/10.1016/j.eneco.2017.10.001.

Sharma, S. S., and Thuraisamy, K. (2013). Oil price uncertainty and sovereign risk: Evidence from Asian economies. Journal of Asian Economics, 28, 51-57. https://doi.org/10.1016/j.asieco.2013.06.001.

Shen, G., Tan, Q., Zhang, H., Zeng, P., and Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. Procedia computer science, 131, 895-903. https://doi.org/10.1016/j.procs.2018.04.298.

Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE International Conference on Machine Learning and Applications, 1394-1401. https://doi.org/10.1109/ICMLA.2018.00227.

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. https://doi.org/10.1016/j.iref.2015.10.047.

Sutton, B., and Catão, L. (2002). Sovereign Defaults: The Role of Volatility. IMF Working Papers, 02(149). https://doi.org/10.5089/9781451856903.001.

Tandon, C., Revankar, S., and Parihar, S. S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. International Journal of

Information Management Data Insights, 1(2), 100035. https://doi.org/10.1016/j.jjimei.2021.100035.

Tang, D. Y., and Yan, H. (2007). Liquidity and Credit Default Swap Spreads. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.891263.

Tanoe, V. (2019). Banking Industry Stocks Analysis Using Econometric and Deep Learning Approach. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3446853.

Tian, S., Yu, Y., and Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. Journal of Banking and Finance, 52, 89-100. https://doi.org/10.1016/j.jbankfin.2014.12.003.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Wang, J., Sun, X., & Li, J. (2020). How do sovereign credit default swap spreads behave under extreme oil price movements? Evidence from G7 and BRICS countries. Finance Research Letters, 34, 101350. https://doi.org/10.1016/j.frl.2019.101350.

Wu, C. H., Lu, C. C., Ma, Y. F., and Lu, R. S. (2018). A new forecasting framework for bitcoin price with LSTM. In 2018 IEEE International Conference on Data Mining Workshops, 168-175. https://doi.org/10.1109/ICDMW.2018.00032.

Xiao, Y., Gray, D. F., Lim, C. H., … Gapen, M. T. (2004). The Contingent Claims Approach to Corporate Vulnerability Analysis: Estimating Default Risk and Economy-Wide Risk Transfer. IMF Working Papers, 04(121). https://doi.org/10.5089/9781451854411.001.

Yuan, C., and Pongsiri, T. J. (2015). Fiscal austerity, growth prospects, and sovereign CDS spreads: The Eurozone and beyond. International Economics, 141, 50-79. https://doi.org/10.1016/j.inteco.2014.12.001

Zahara, S., Sugianto, & Ilmiddaviq, M. B. (2020). Consumer price index prediction using Long Short Term Memory (LSTM) based cloud computing. Journal of Physics: Conference Series, 1456(1), 012022. https://doi.org/10.1088/1742-6596/1456/1/012022.

Zhang, F. X. (2003). What Did the Credit Market Expect of Argentina Default? Evidence from Default Swap Data. Finance and Economics Discussion Series, 2003(25), 1–43. https://doi.org/10.17016/feds.2003.25.

Zhang, Y., Ma, F., & Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?. Journal of Empirical Finance, 54, 97-117. https://doi.org/10.1016/j.jempfin.2019.08.007.

Zhu, H. (2006). An empirical comparison of credit spreads between the bond market and the credit default swap market. Journal of Financial Services Research, 29(3), 211-235. https://doi.org/10.1007/s10693-006-7626-x.

APPENDIX

Table A.1: Descriptive statistics

Series	N	Mean	St.Dev	Min	Max
cds_bra	2,057	210.500	88.840	100.200	533.3
exr_bra	2,057	3.003	0.731	1.696	4.259
spi_bra	2,057	64,448	16,870	37,498	109,581
spivol_bra	2,057	390,004.042	127,046,284	70,110,300	1,423,400,100
g3m_bra	2,057	9.708	3.006	4.684	15.830
g1y_bra	2,057	9.740	2.946	4.315	16.290
g5y_bra	2,057	10.940	1.867	8.444	17.040
g10y_bra	2,057	11.110	1.995	6.396	16.840
yield_bra	2,057	1.399	1.814	-2.693	6.032
exr_rus	2,057	51.900	14.930	28.990	82.680
spi_rus	2,057	1,862	429.200	1,237	3,009
spivol_rus	2,057	37,408,824,497	32,850,349,584	4,274,850,193	269,164,000,000
g3m_rus	2,057	8.026	1.978	5.443	15.590
g1y_rus	2,057	7.997	1.949	5.380	17.610
g5y_rus	2,057	8.350	1.797	6.040	18.380
g10y_rus	2,057	8.486	1.483	6.340	16.060
yield_rus	2,057	0.461	1.163	-2.851	3.060
cds_chn	2,057	84.250	26.450	36.14	150.500
exr_chn	2,057	6.474	0.298	6.041	7.179
spi_chn	2,057	2,808	577.100	1,950	5,166
spivol_chn	2,057	18,590,744,702	12,736,363,758	4,258,064,000	85,713,280,000
g3m_chn	2,057	2.804	0.596	1.85	5.000
g1y_chn	2,057	2.888	0.524	1.66	4.233
g5y_chn	2,057	3.291	0.462	2.394	4.560
g10y_chn	2,057	3.511	0.434	2.633	4.711
yield_chn	2,057	1.399	1.814	-2.693	6.032
cds_zar	2,057	203.100	46.290	122	392
exr_zar	2,057	12.140	2.290	7.44	16.870
spi_zar	2,057	43,949	6,473	28,972	55,065
spivol_zar	2,057	123,109,342	57,108,997	9,800,365	595,526,015
g3m_zar	2,057	6.430	0.941	4.575	8.087
g1y_zar	2,057	5.811	2.454	1.237	8.050
g5y_zar	2,057	7.600	0.815	5.392	9.903
g10y_zar	2,057	8.347	0.719	6.02	10.360
yield_zar	2,057	1.917	0.581	0.618	3.289
cds_mex	2,057	121.200	28.280	64.500	231.9
exr_mex	2,057	16.380	2.852	11.98	21.95

Source: Author's estimation.

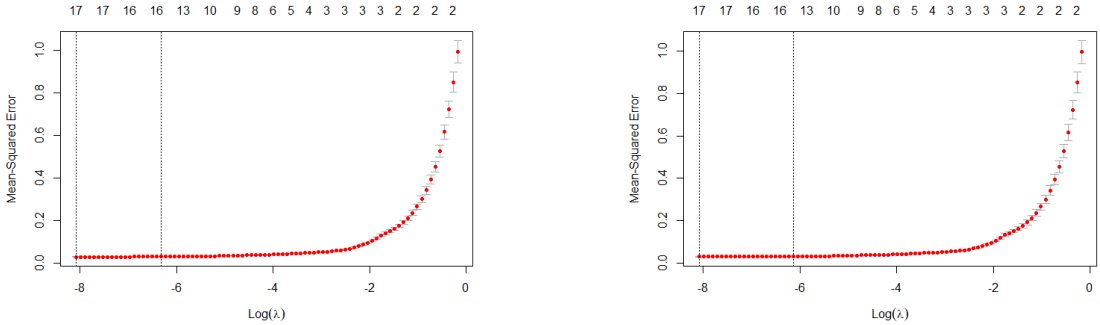
Table A.1: Descriptive statistics (cont'd)

Series	N	Mean	St.Dev	Min	Max
spi_mex	2,057	44,151	3,329	36,549	51,713
spivol_mex	2,057	197,973,483	98,798,731	6,975,142	1,759,230,675
g3m_mex	2,057	5.107	1.972	0.000	8.629
g1y_mex	2,057	5.273	1.847	2.965	8.634
g5y_mex	2,057	5.997	1.224	3.943	9.015
g10y_mex	2,057	6.513	0.962	4.415	9.239
yield_mex	2,057	1.406	1.324	-0.966	9.028
cds_idn	1,816	153.400	45.330	71.14	285.600
exr_idn	2,057	63.500	5.626	48.700	74.390
spi_idn	2,057	5,202	754.300	3,655	6,689
spivol_idn	2,057	5,584,779,451	3,213,574,307	644,025,000	33,521,349,800
g3m_idn	2,057	6.508	0.644	4.216	6.840
g1y_idn	2,057	6.239	1.006	3.615	8.751
g5y_idn	2,057	6.998	1.101	4.475	9.631
g10y_idn	2,057	7.363	1.014	5.047	9.833
yield_idn	2,057	0.856	1.244	-1.793	3.330
cds_tur	2,057	245.800	83.820	111.000	574.400
exr_tur	2,057	3.203	1.344	1.742	6.884
spi_tur	2,057	842.100	144.800	515.600	1,208.000
spivol_tur	2,057	780,243,564	362,757,220	109,005,310	2,586,539,506
g3m_tur	2,057	10.670	2.63	4.050	13.010
g1y_tur	2,057	11.750	4.996	4.590	28.03
g5y_tur	2,057	11.380	3.96	5.530	27.200
g10y_tur	2,057	11.050	3.306	6.100	22.690
yield_tur	2,057	0.383	2.458	-3.990	9.680
vix	2,057	15.230	3.841	9.140	40.740
corp_aaa	2,057	4.056	0.302	3.350	4.700
corp_baa	2,057	4.903	0.318	4.290	5.470
ted	1,937	0.300	0.106	0.100	0.680
g5y_us	2,057	1.629	0.609	0.543	3.092
oil	2,057	76.11	26.190	27.880	126.200
corp_spread	2,057	-0.846	0.264	-1.320	-0.400
nsdq	2,057	5,272	1,652	2,711	8,705
dax	2,057	10,320	1,947	5,969	13,560
ftse	2,057	6,744	582.200	5,260	7,877
baltic	2,057	1,055	407.700	290	2,518
g5y_deu	2,057	0,019	0.459	-0.935	1.115
g10y_deu	2,057	0,721	0.667	-0.716	2.054
g3_deu	2,057	-0.408	0.338	-1.035	0.107
yield_deu	2,057	1.128	0.432	-0.008	2.055

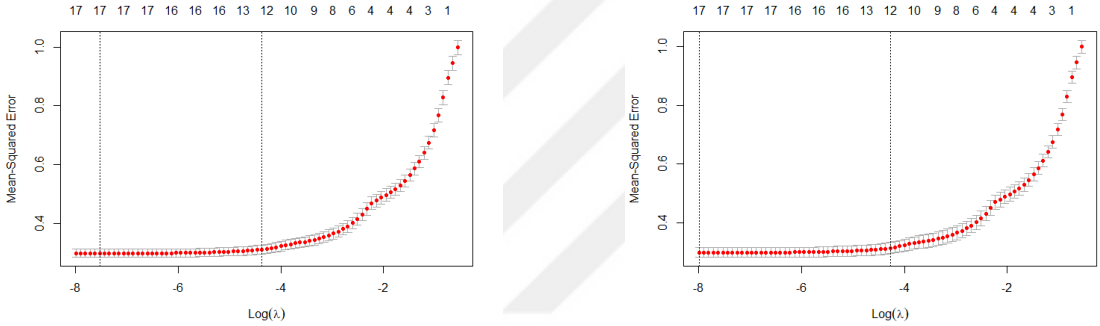
Source: Author's estimation.

Figure A.1: Mean square errors for cross validation, ridge (left) and LASSO (right) for BRICS countries

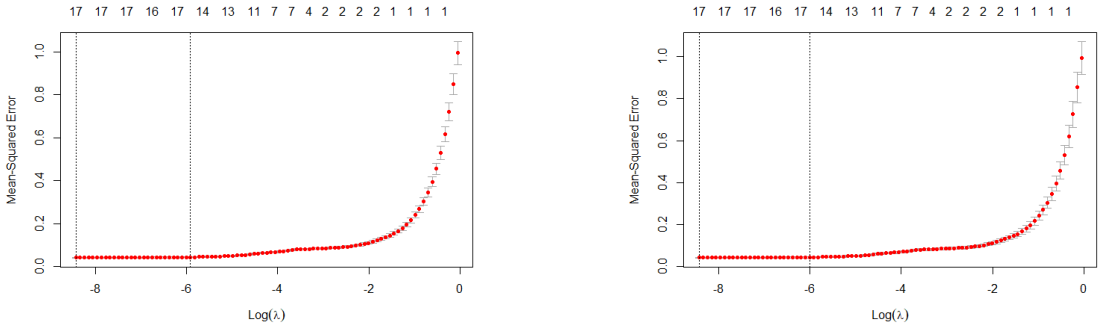
Brazil



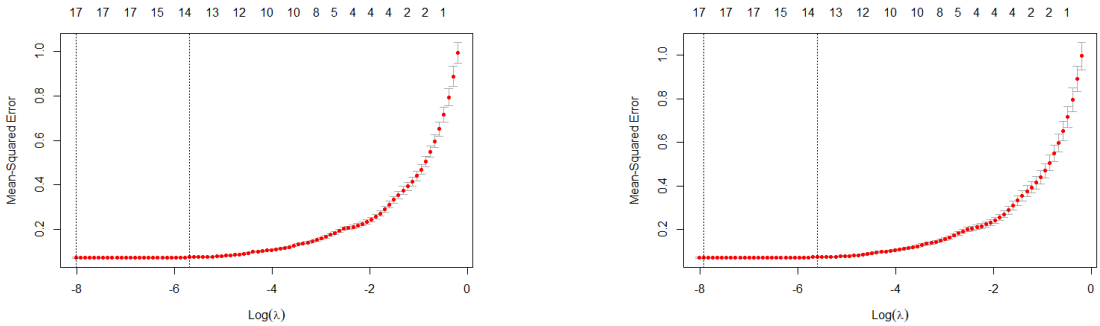
China



Russia

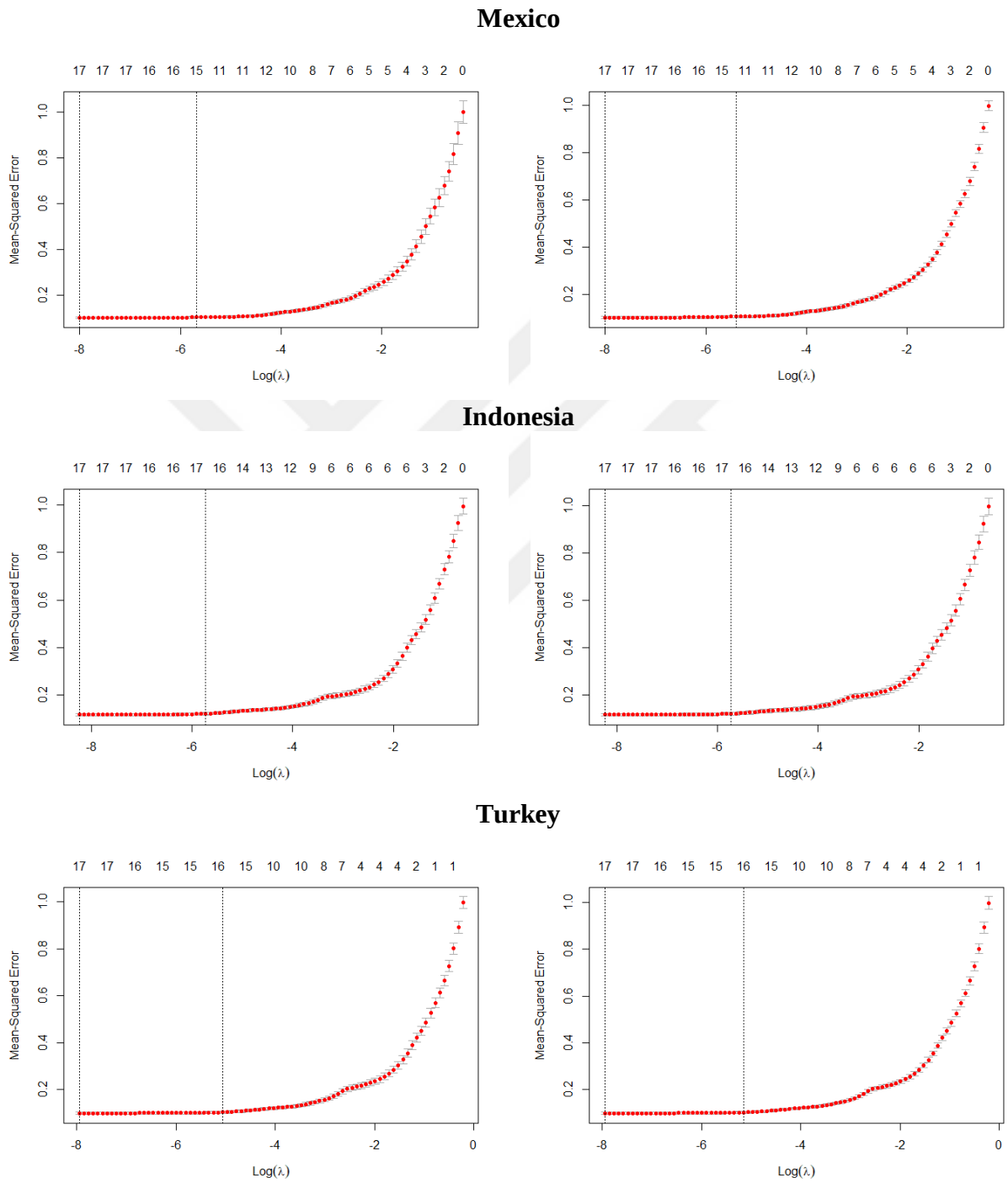


South Africa



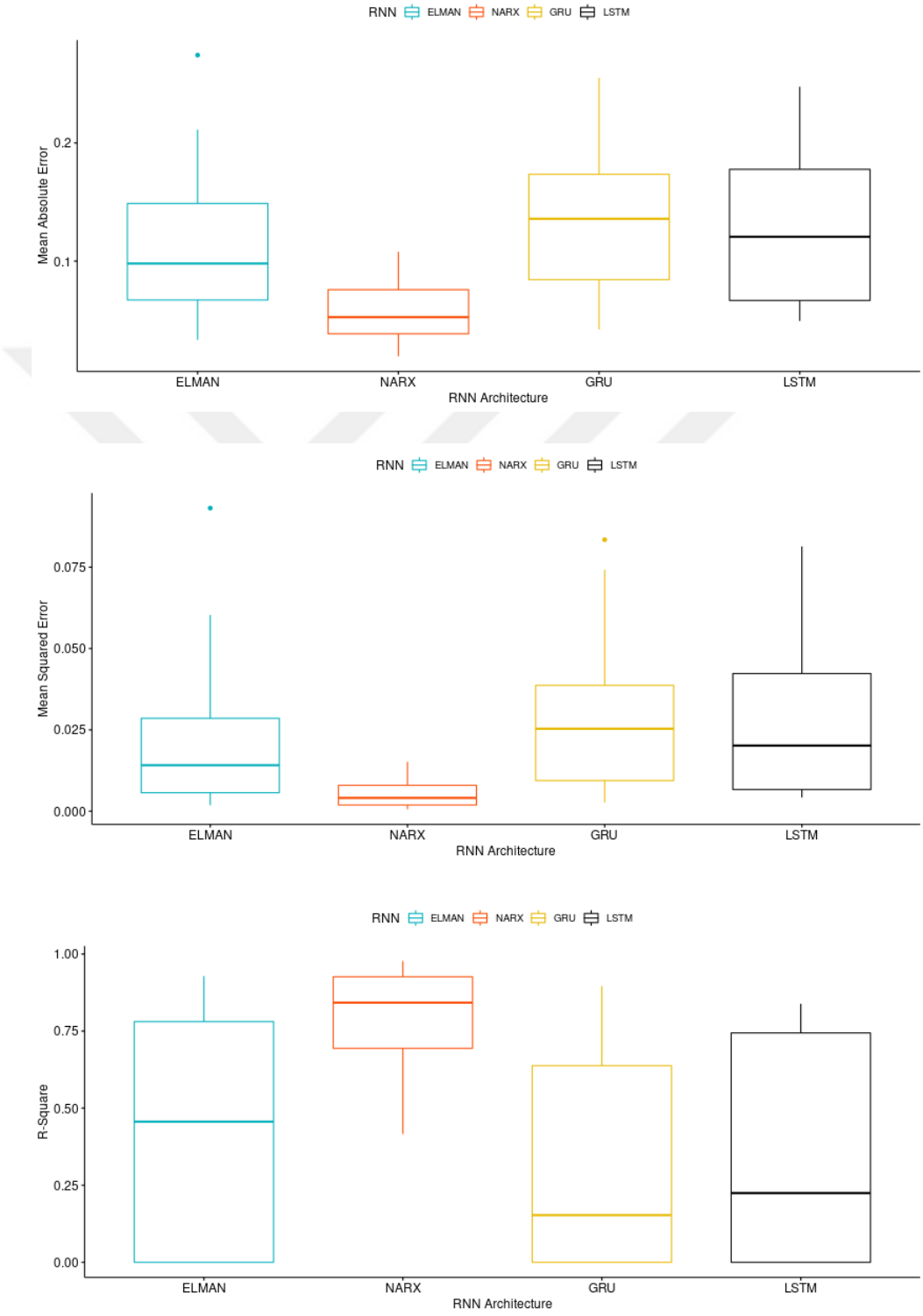
Source: Author's estimation.

Figure A.2: Mean square errors for cross validation, ridge (left) and LASSO (right) for MINT countries



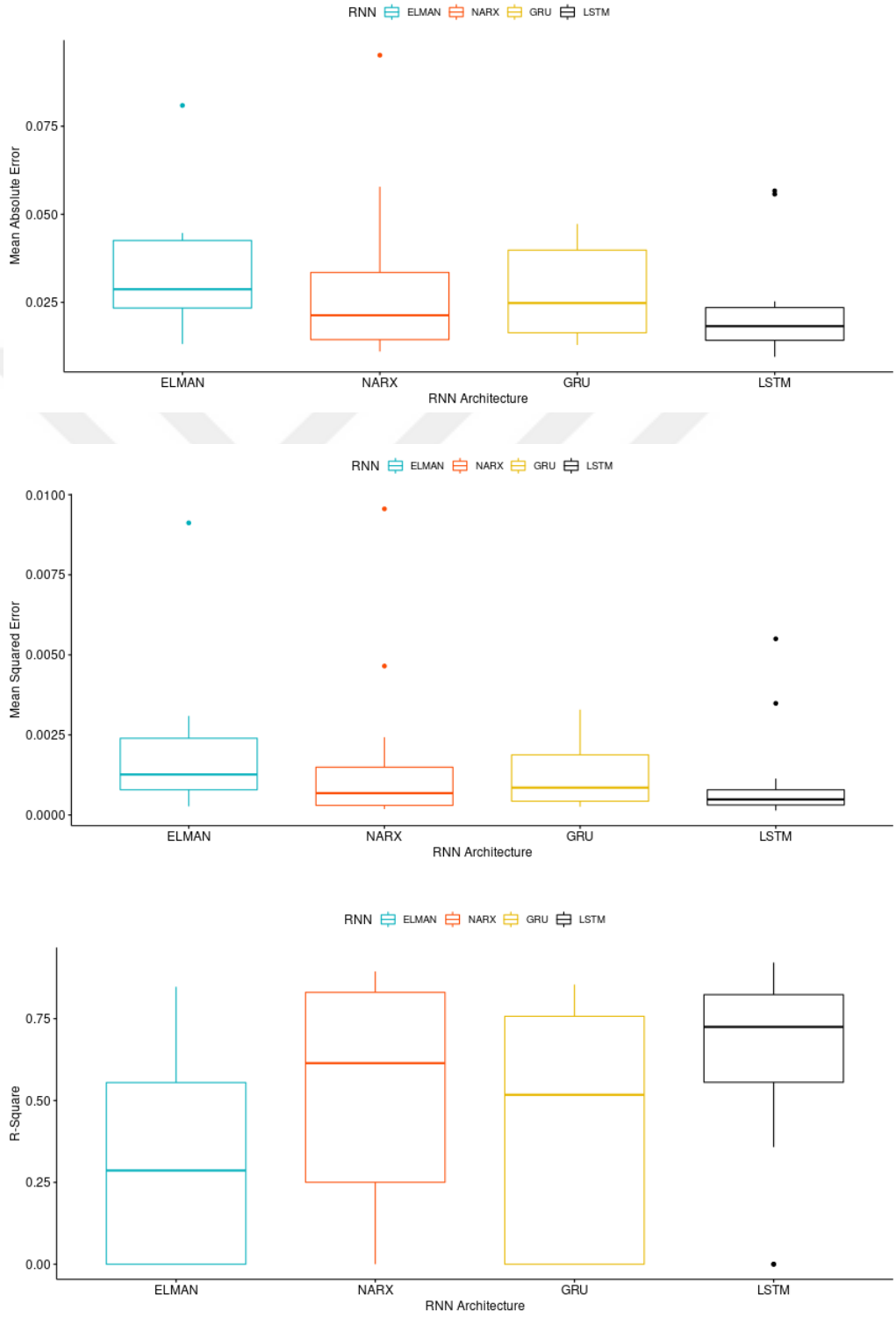
Source: Author's estimation.

Figure A.3: The comparison of RNN results by MAE, MSE and R-squared for Brazil



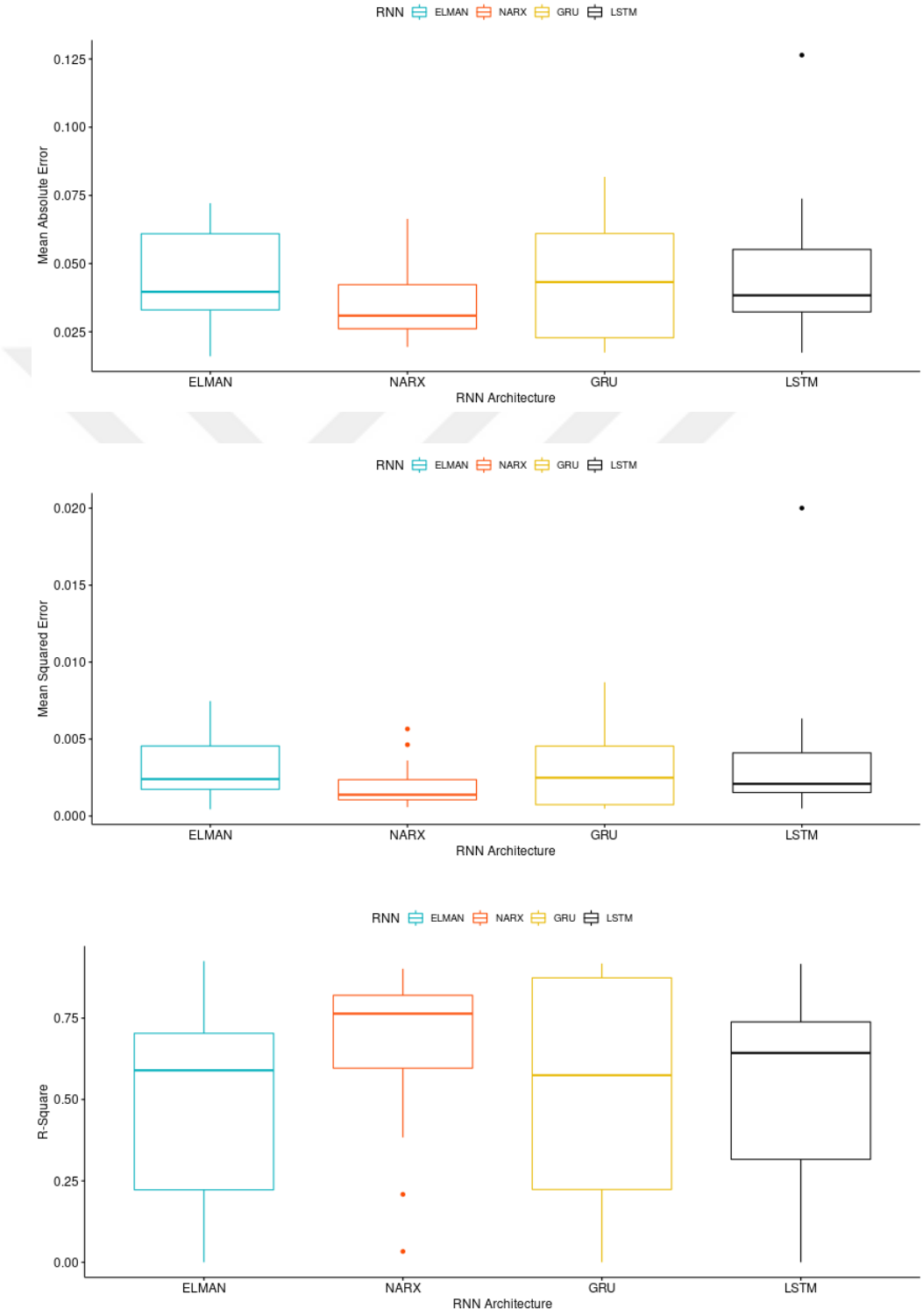
Source: Author's estimation.

Figure A.4: The comparison of RNN results by MAE, MSE and R-squared for Russia



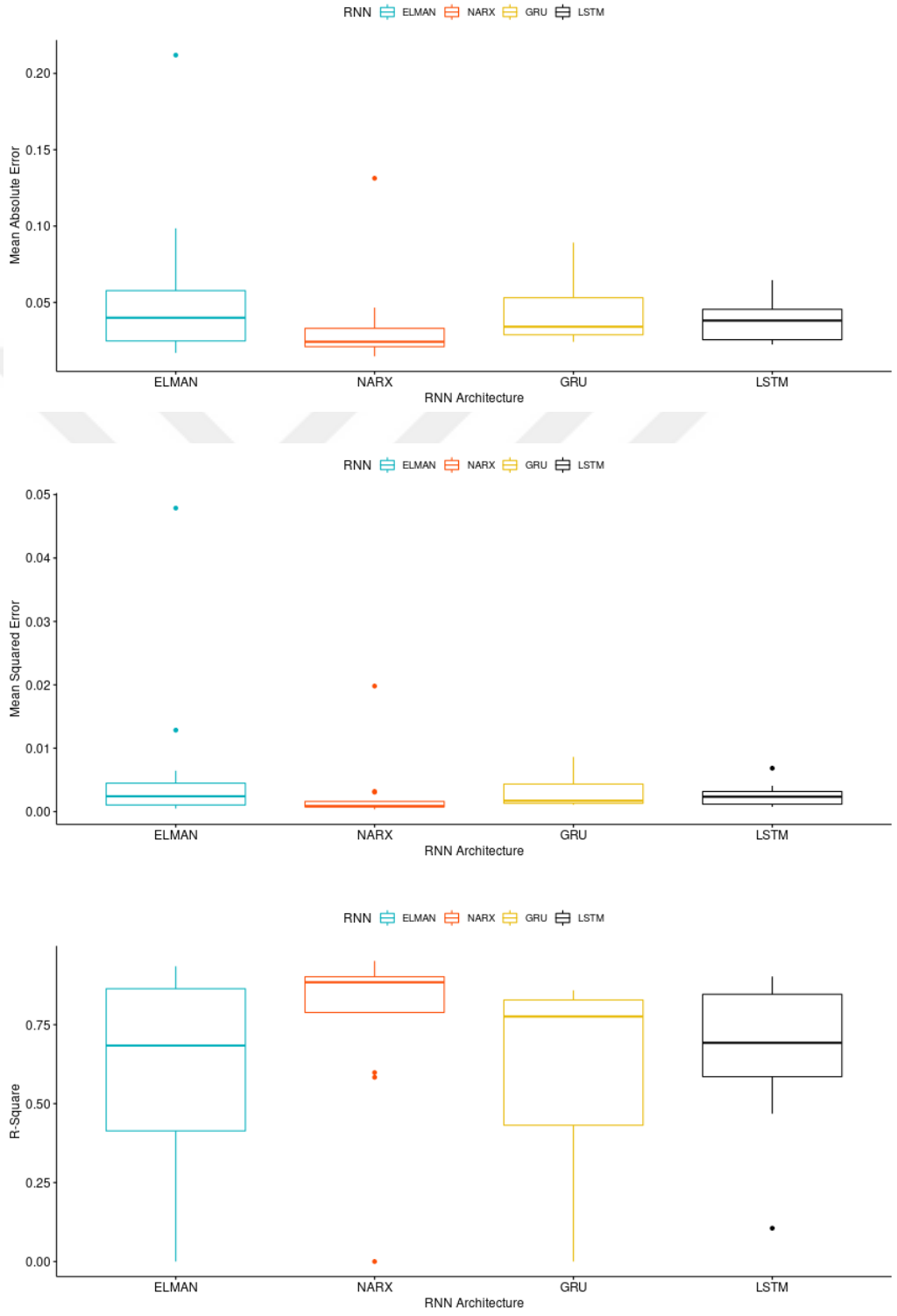
Source: Author's estimation.

Figure A.5: The comparison of RNN results by MAE, MSE and R-squared for China



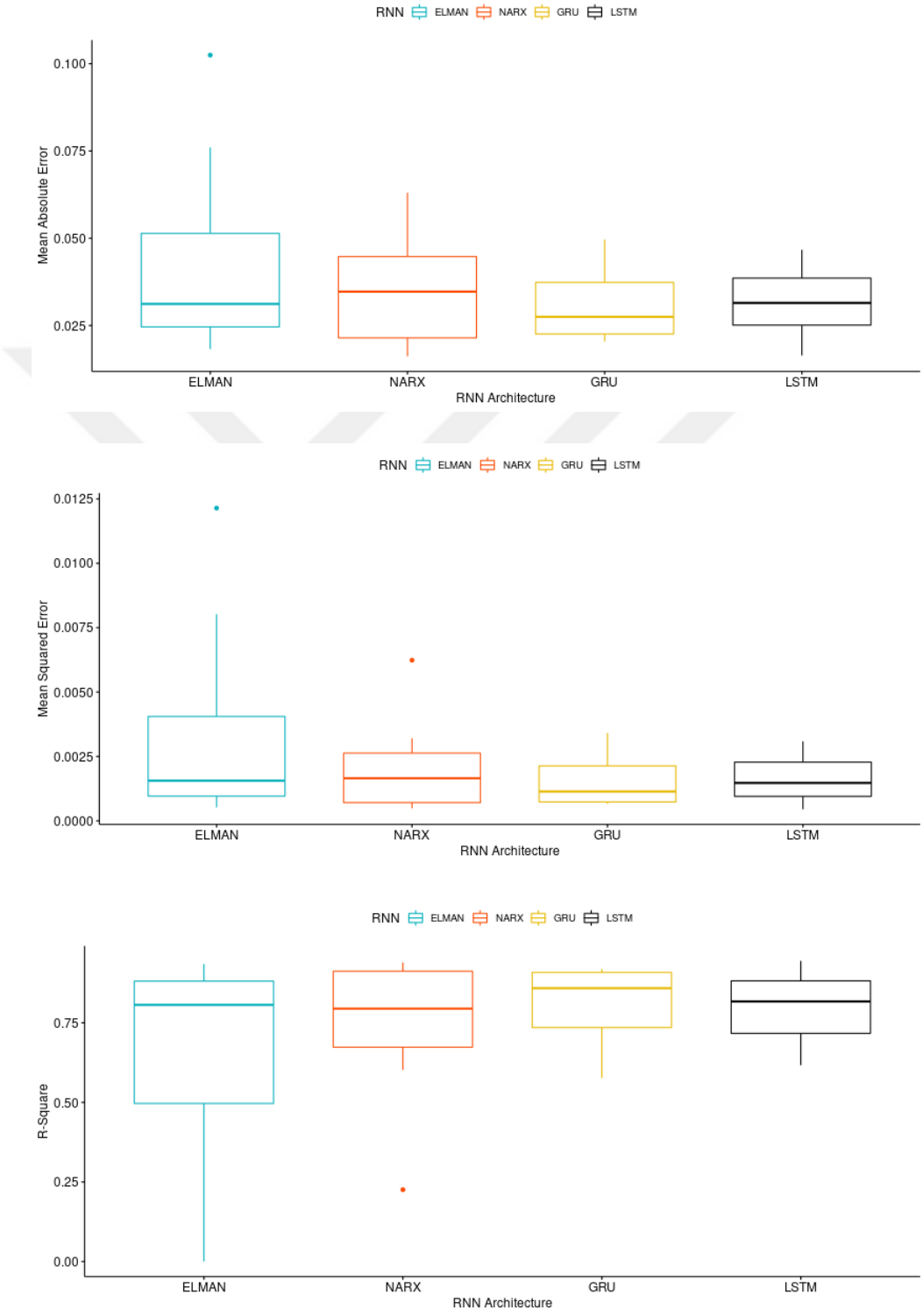
Source: Author's estimation.

Figure A.6: The comparison of RNN results by MAE, MSE and R-squared for South Africa



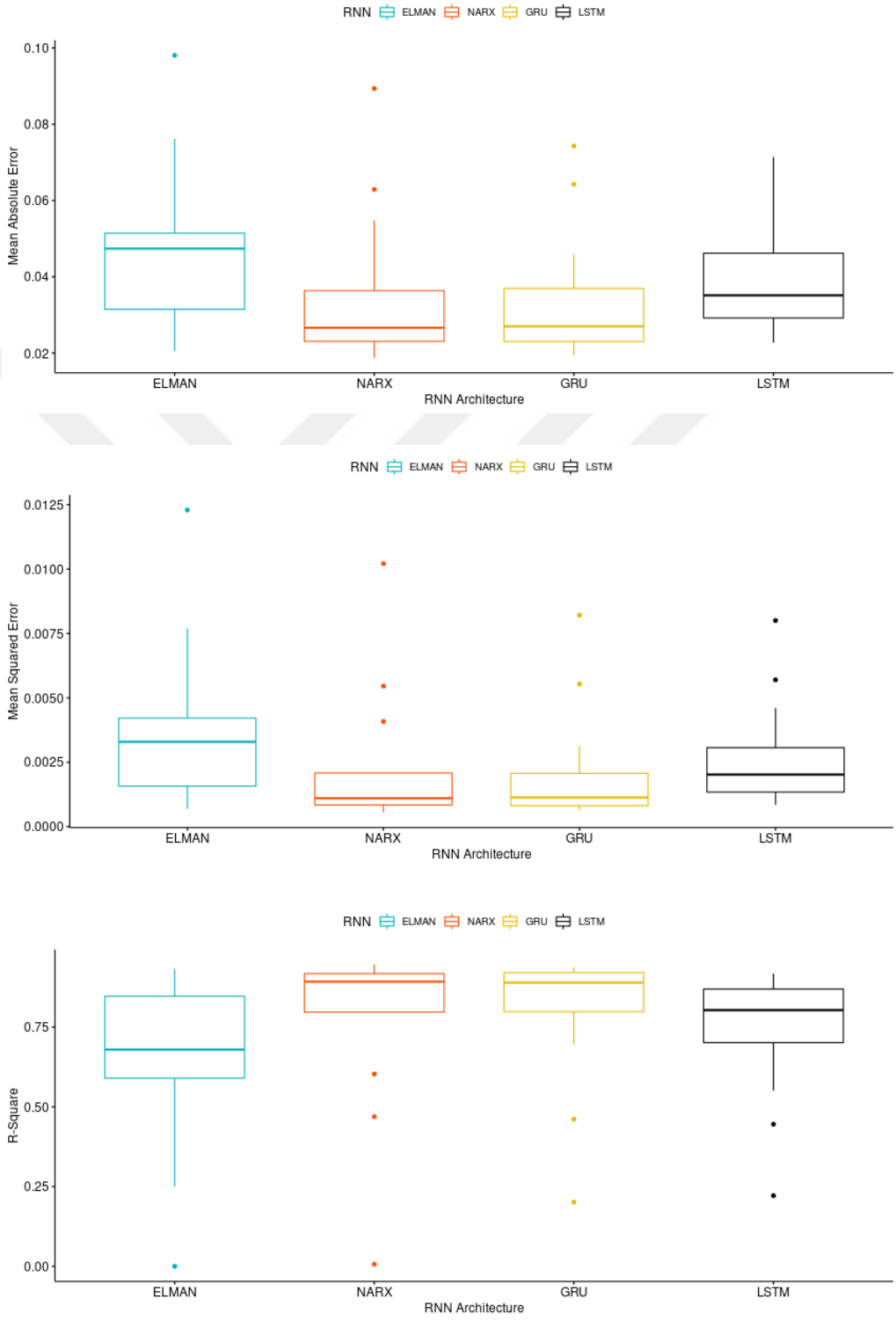
Source: Author's estimation.

Figure A.7: The comparison of RNN results by MAE, MSE and R-squared for Mexico



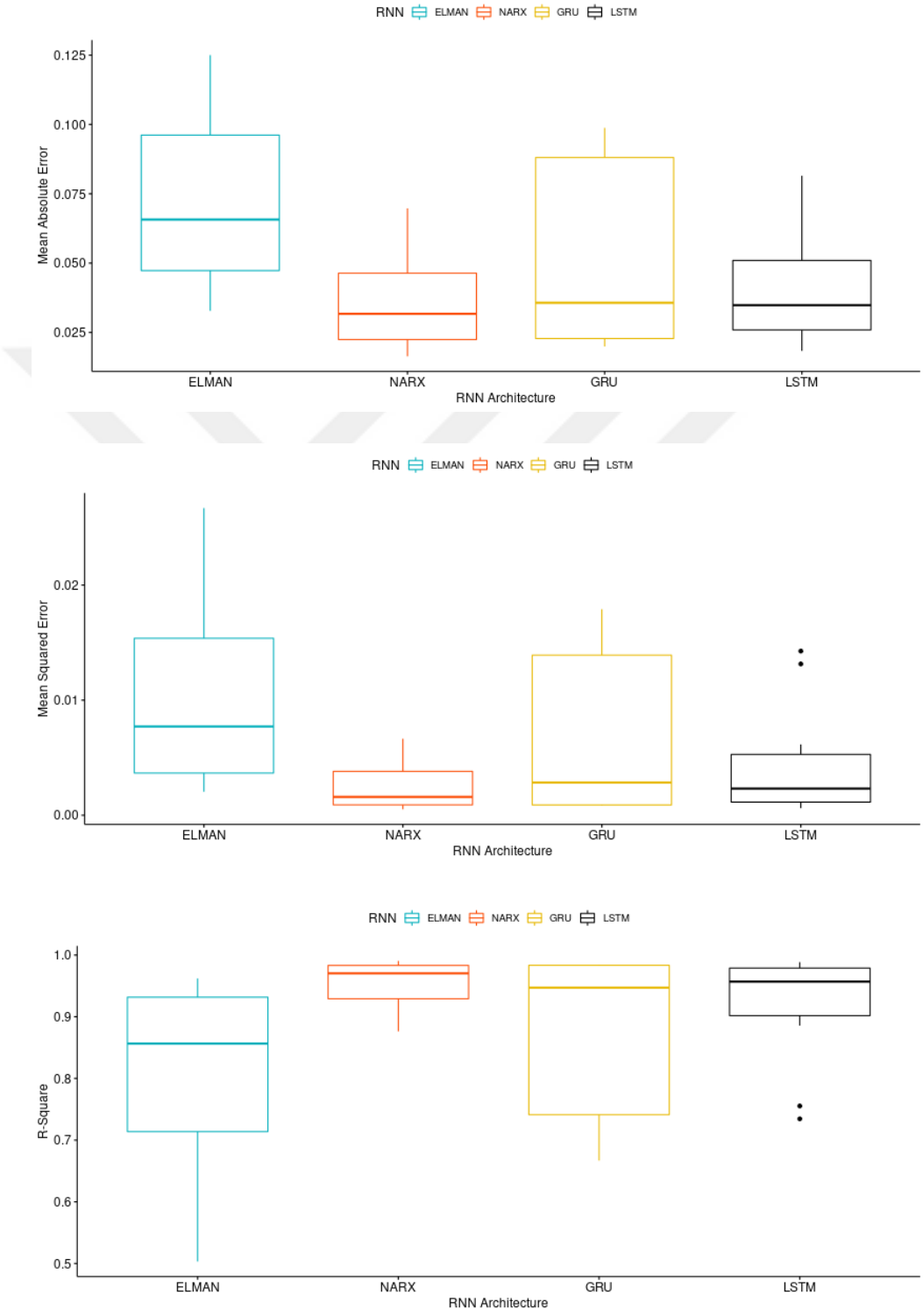
Source: Author's estimation.

Figure A.8: The comparison of RNN results by MAE, MSE and R-squared for Indonesia



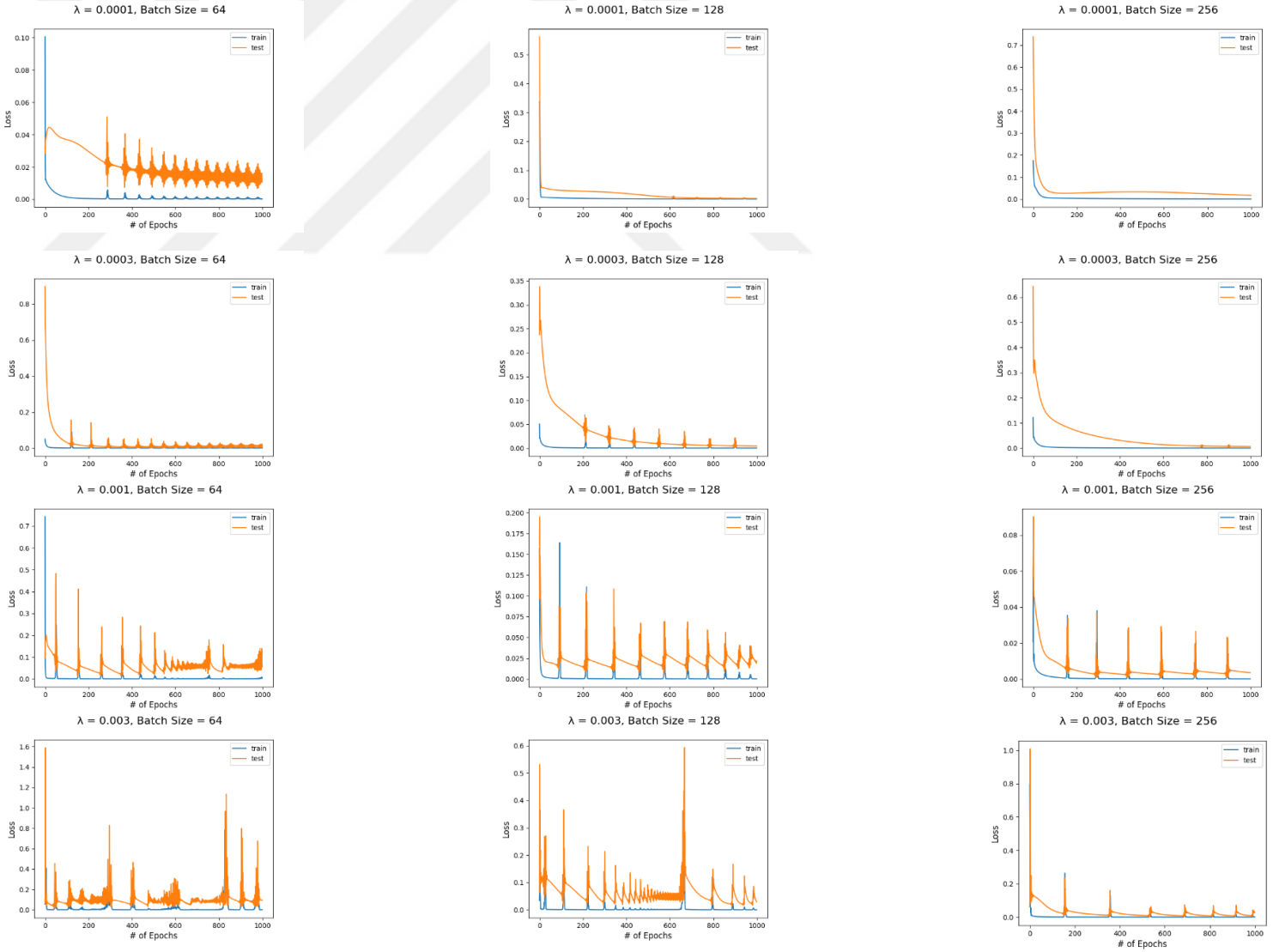
Source: Author's estimation.

Figure A.9: The comparison of RNN results by MAE, MSE and R-squared for Turkey



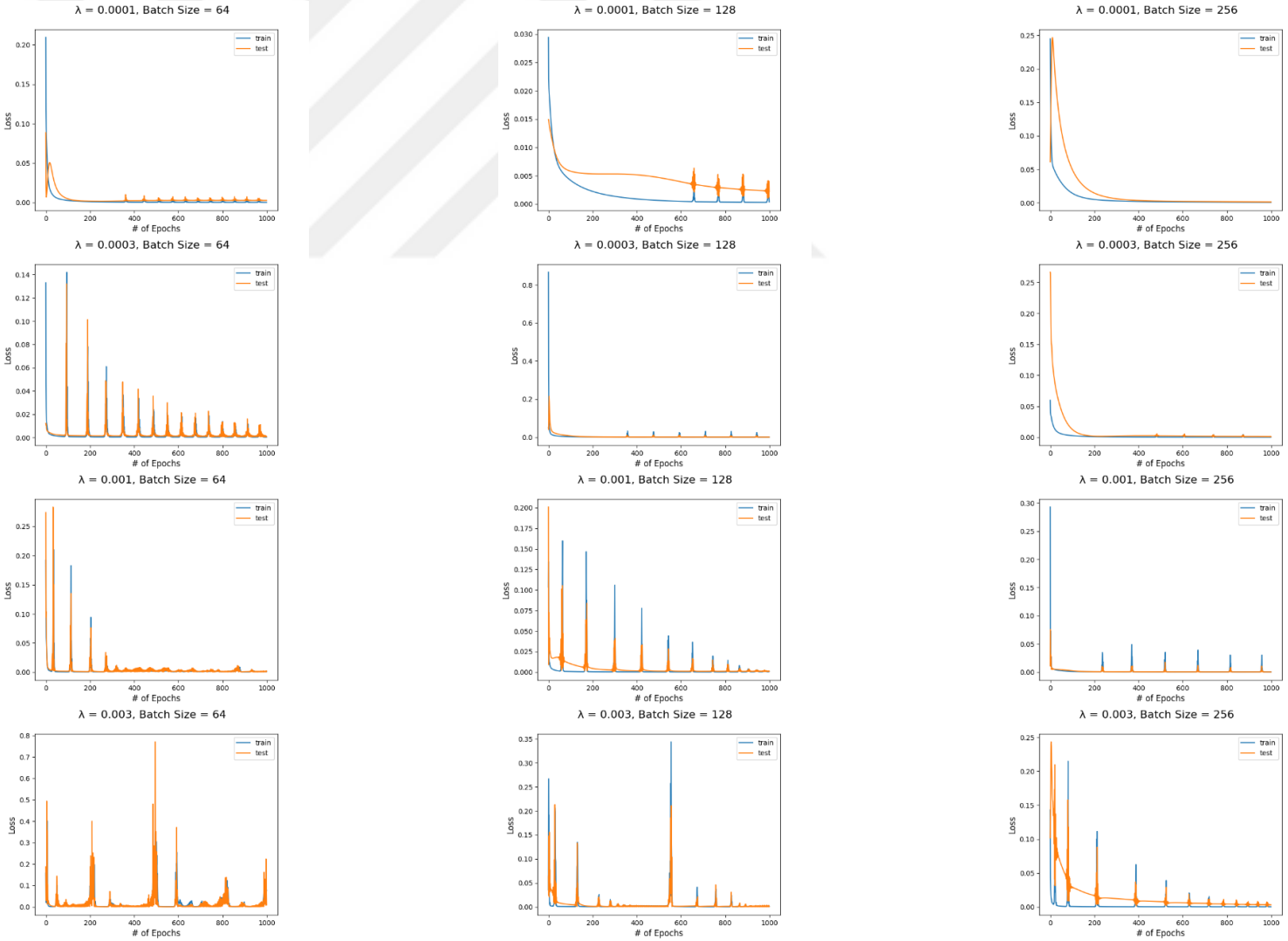
Source: Author's estimation.

Figure A.10: The estimated loss for Brazil in ELMAN's RNN



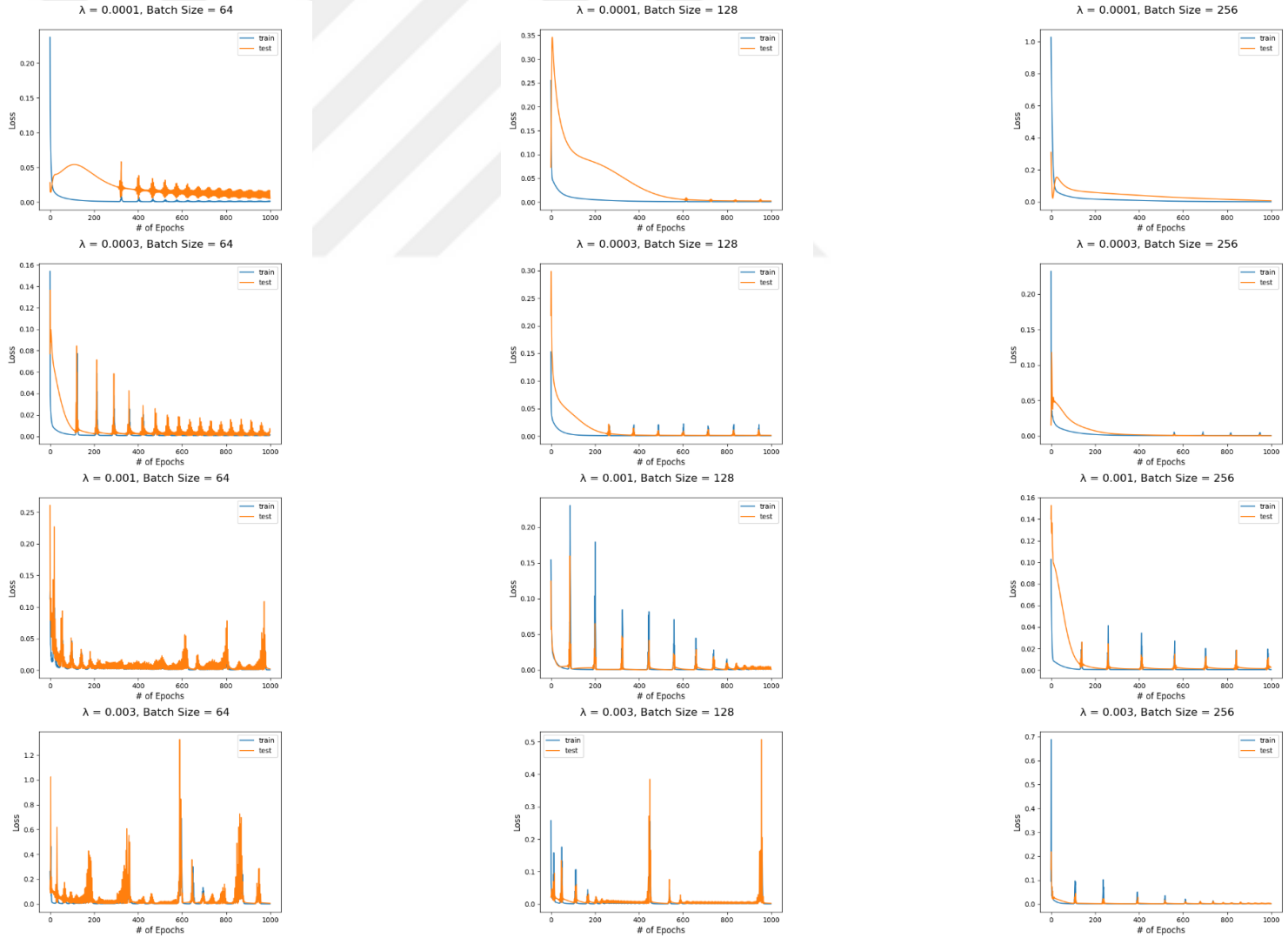
Source: Author's estimation.

Figure A.11: The estimated loss for Russia in ELMAN's RNN



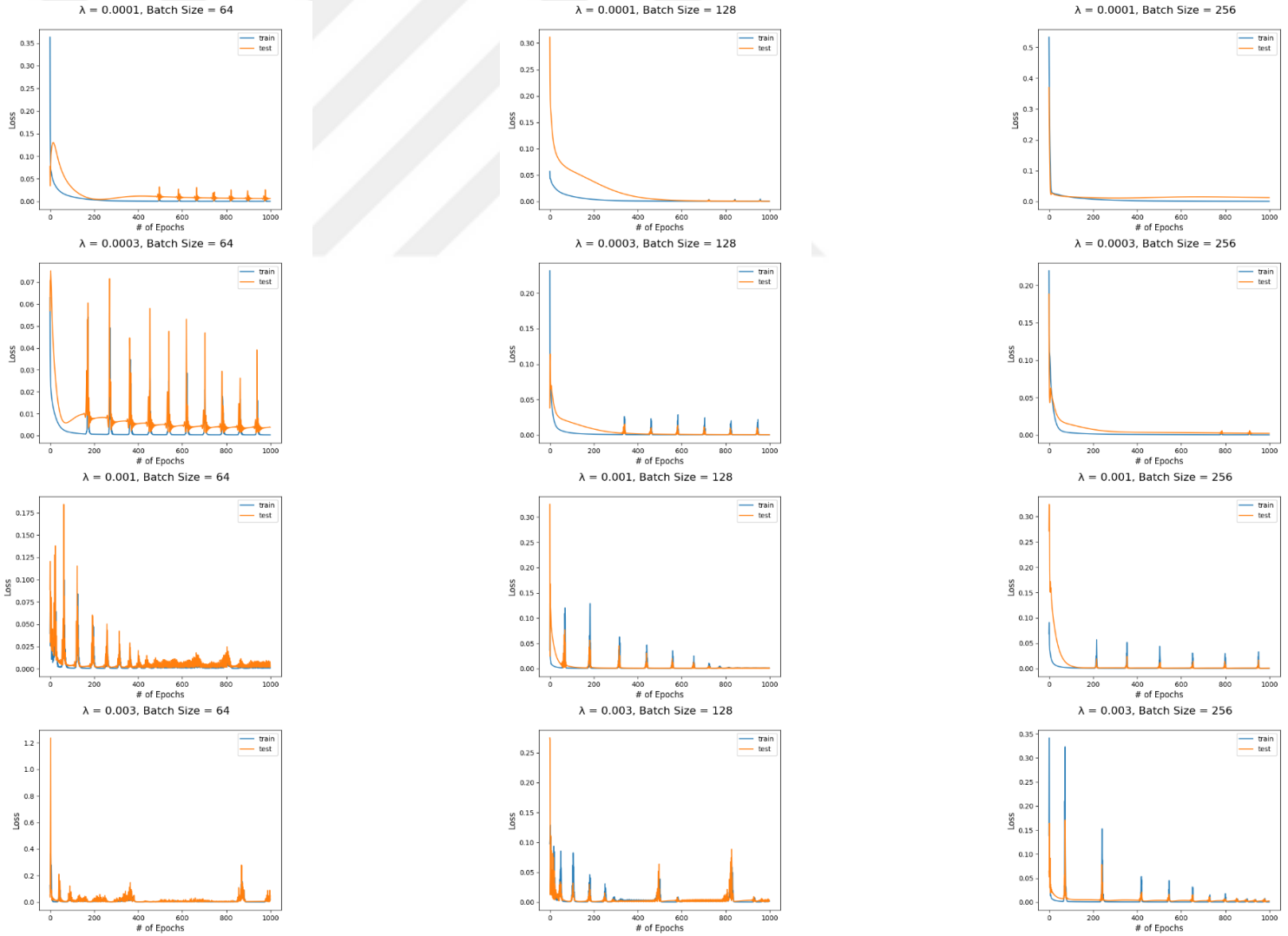
Source: Author's estimation.

Figure A.12: The estimated loss for China in ELMAN's RNN



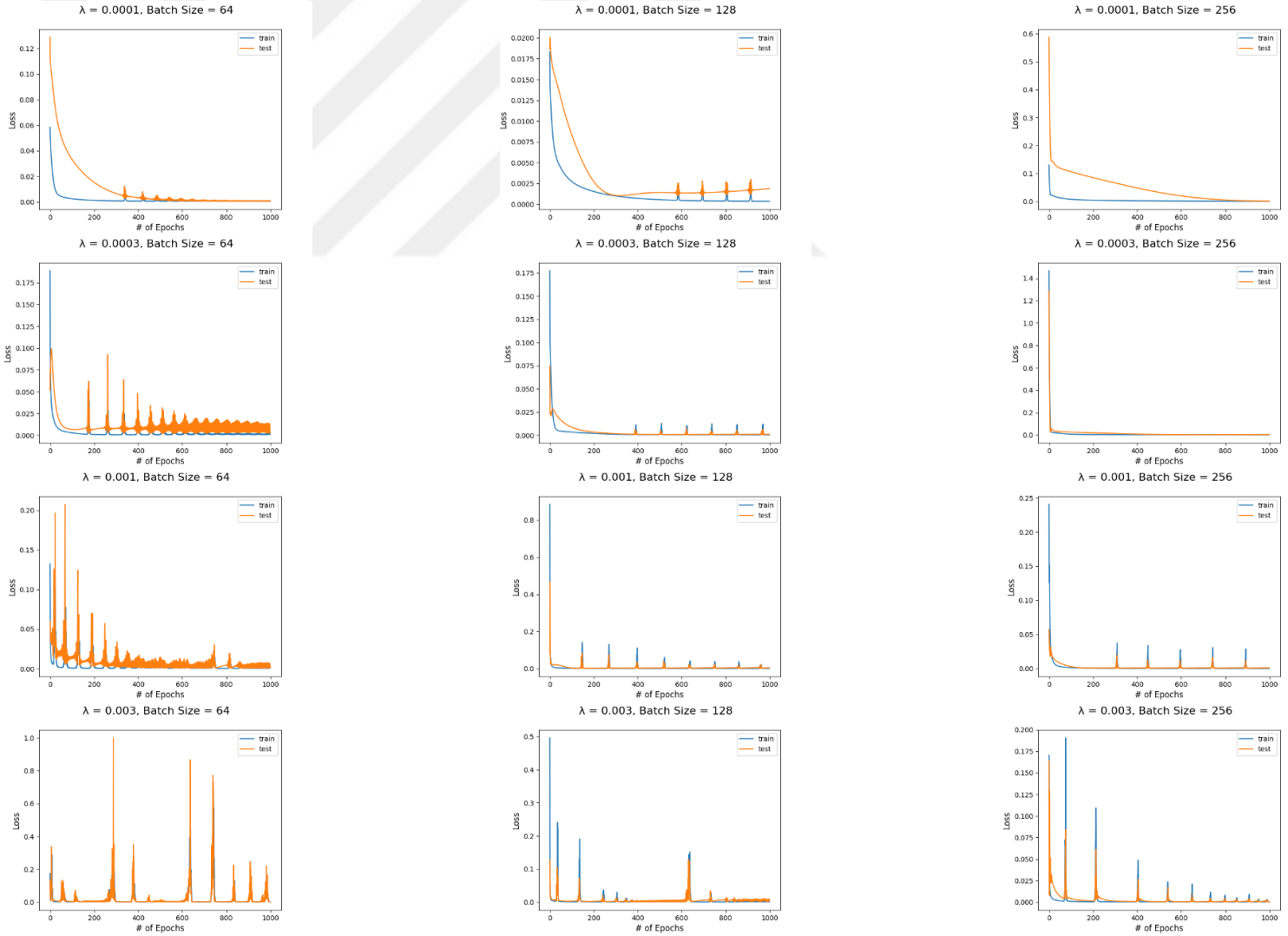
Source: Author's estimation.

Figure A.13: The estimated loss for South Africa in ELMAN's RNN



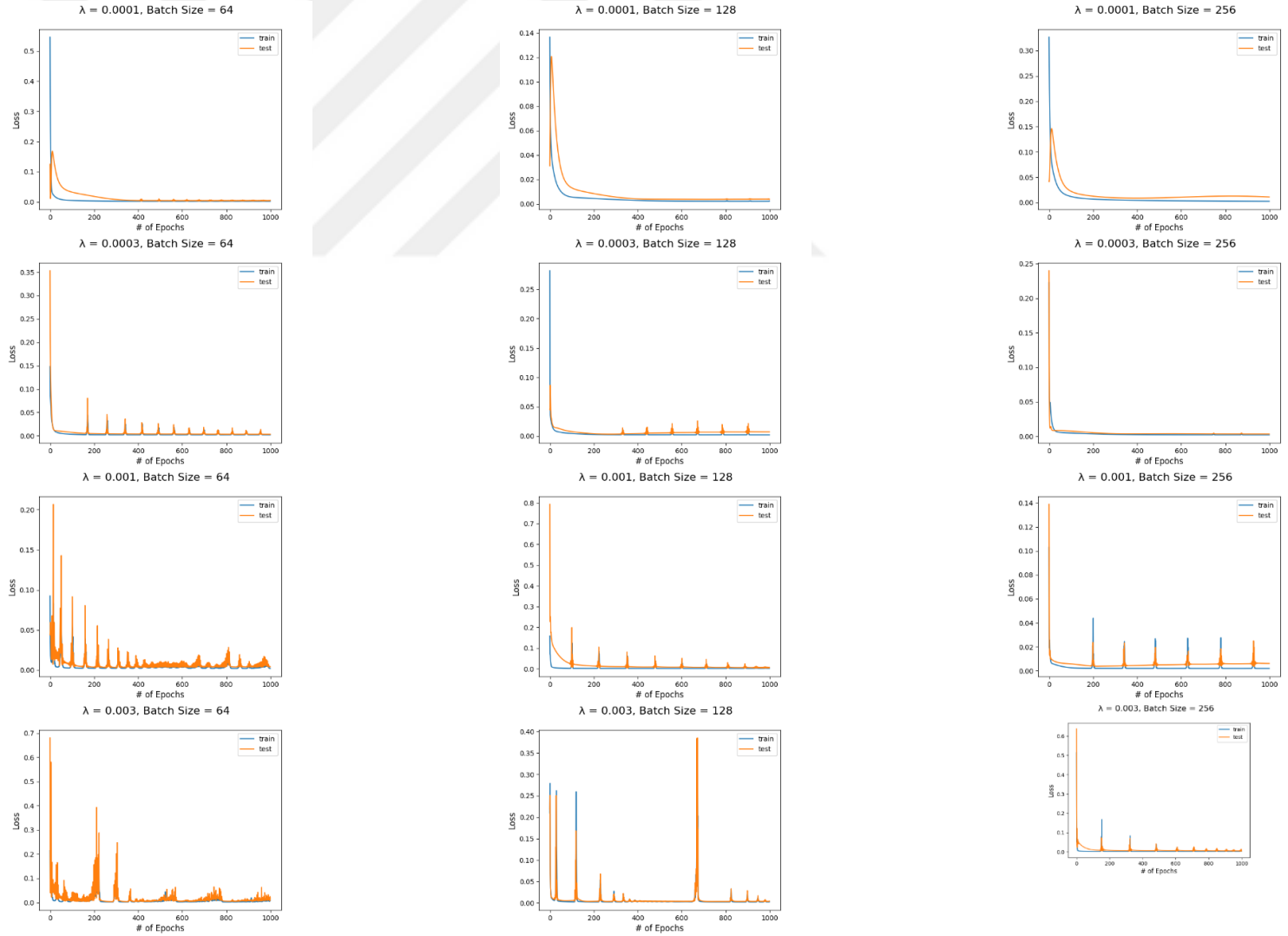
Source: Author's estimation.

Figure A.14: The estimated loss for Mexico in ELMAN's RNN



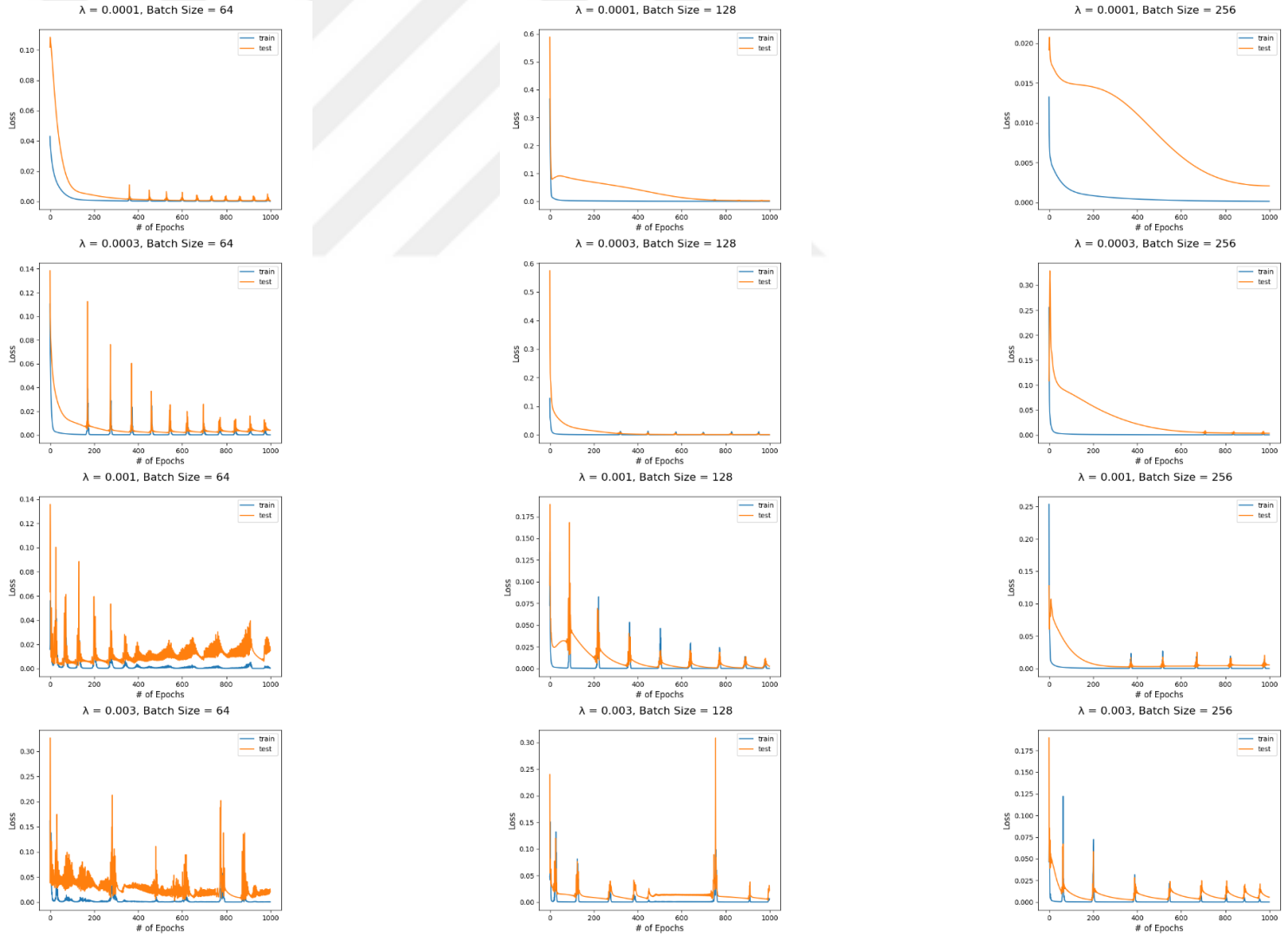
Source: Author's estimation.

Figure A.15: The estimated loss for Indonesia in ELMAN's RNN



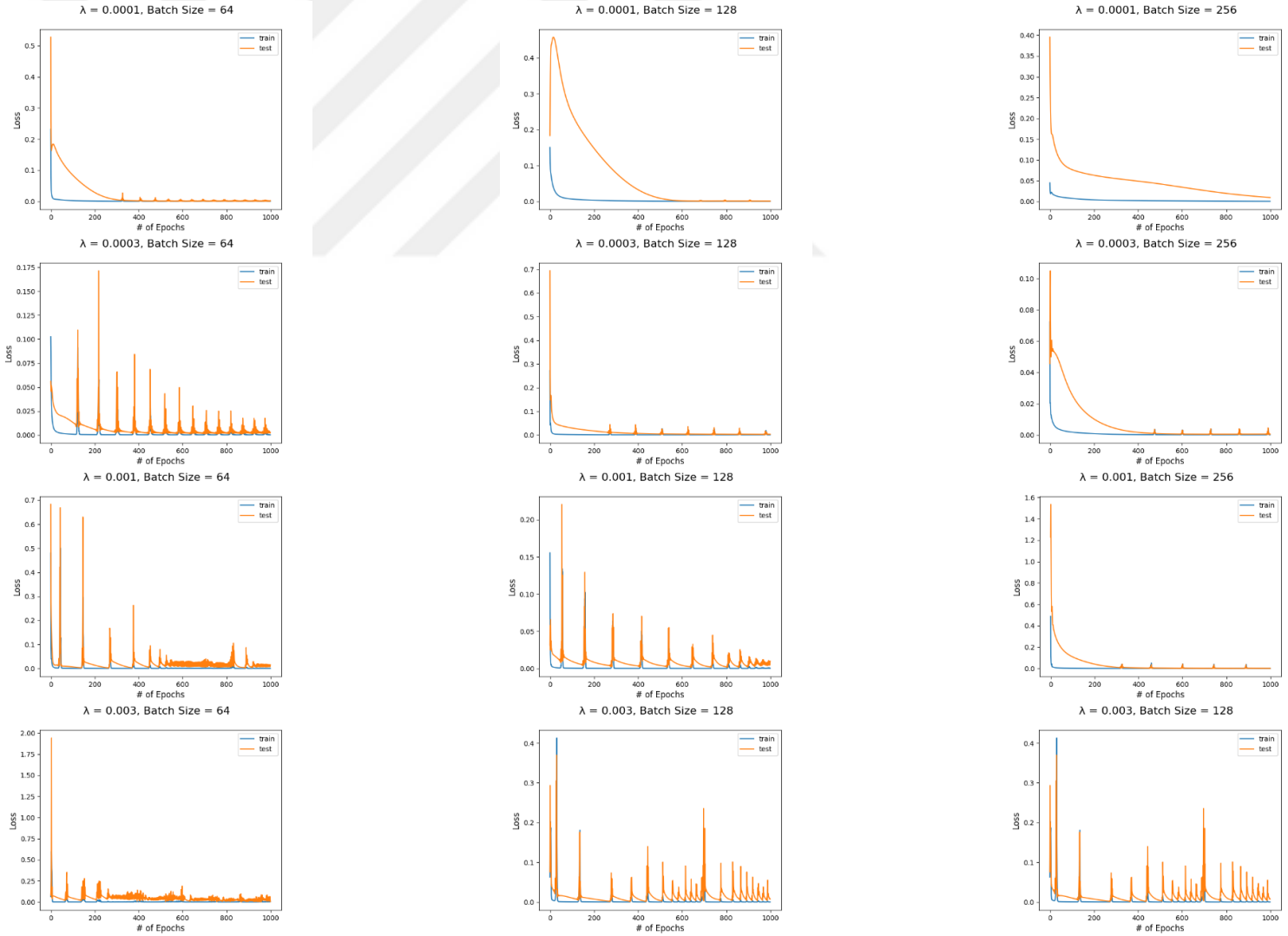
Source: Author's estimation.

Figure A.16: The estimated loss for Turkey in ELMAN's RNN



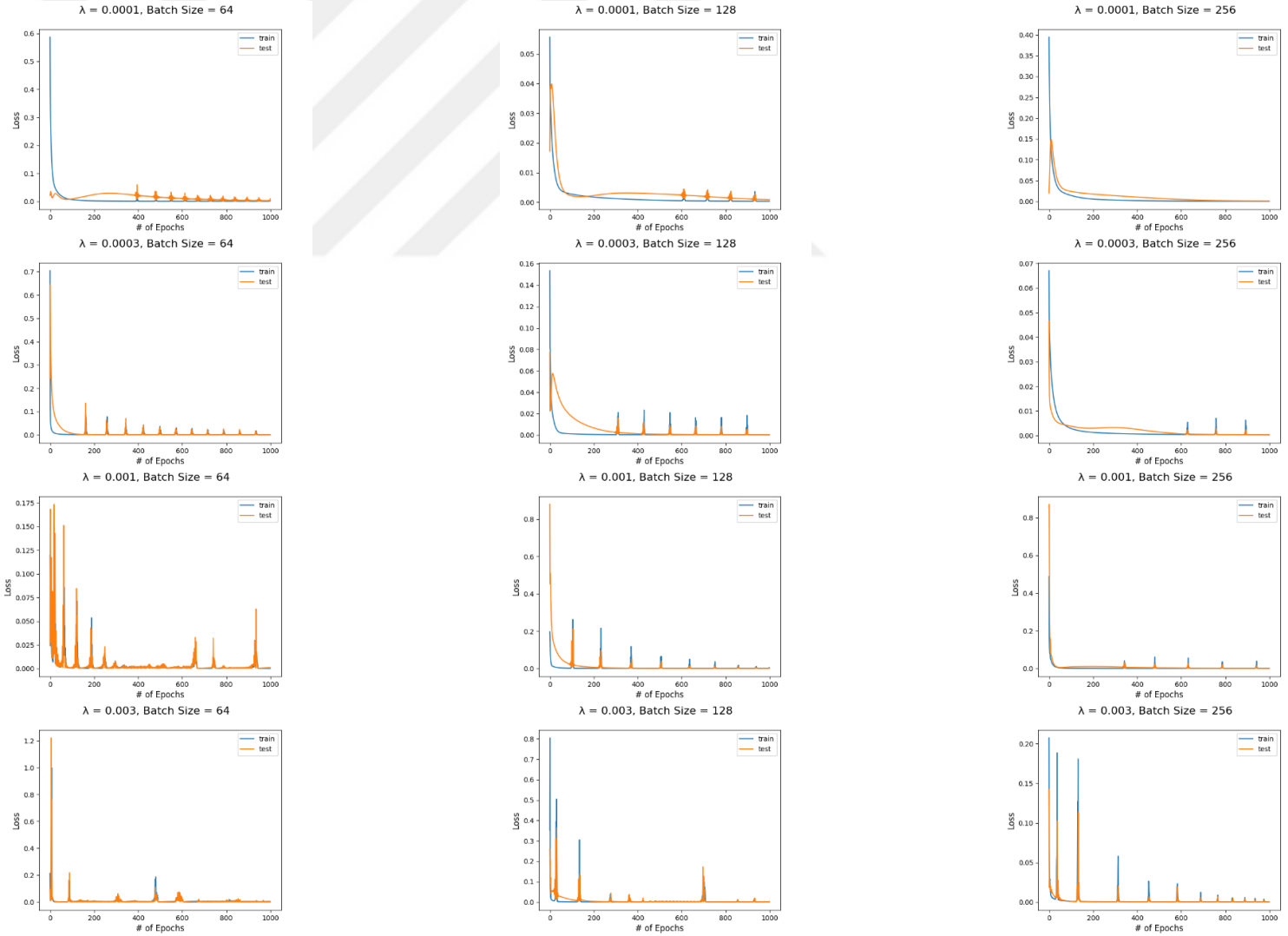
Source: Author's estimation.

Figure A.17: The estimated loss for Brazil in NARX



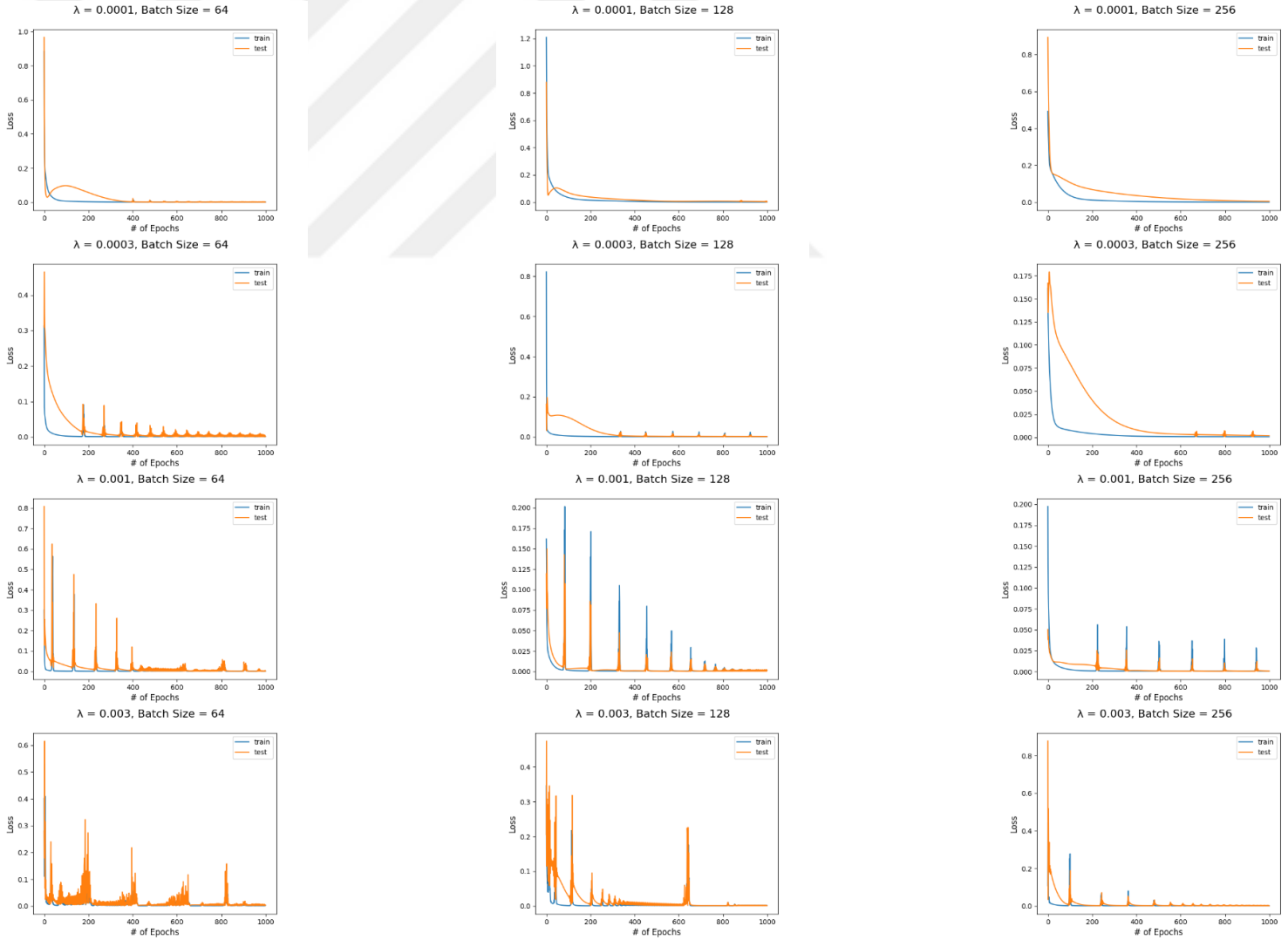
Source: Author's estimation.

Figure A.18: The estimated loss for Russia in NARX



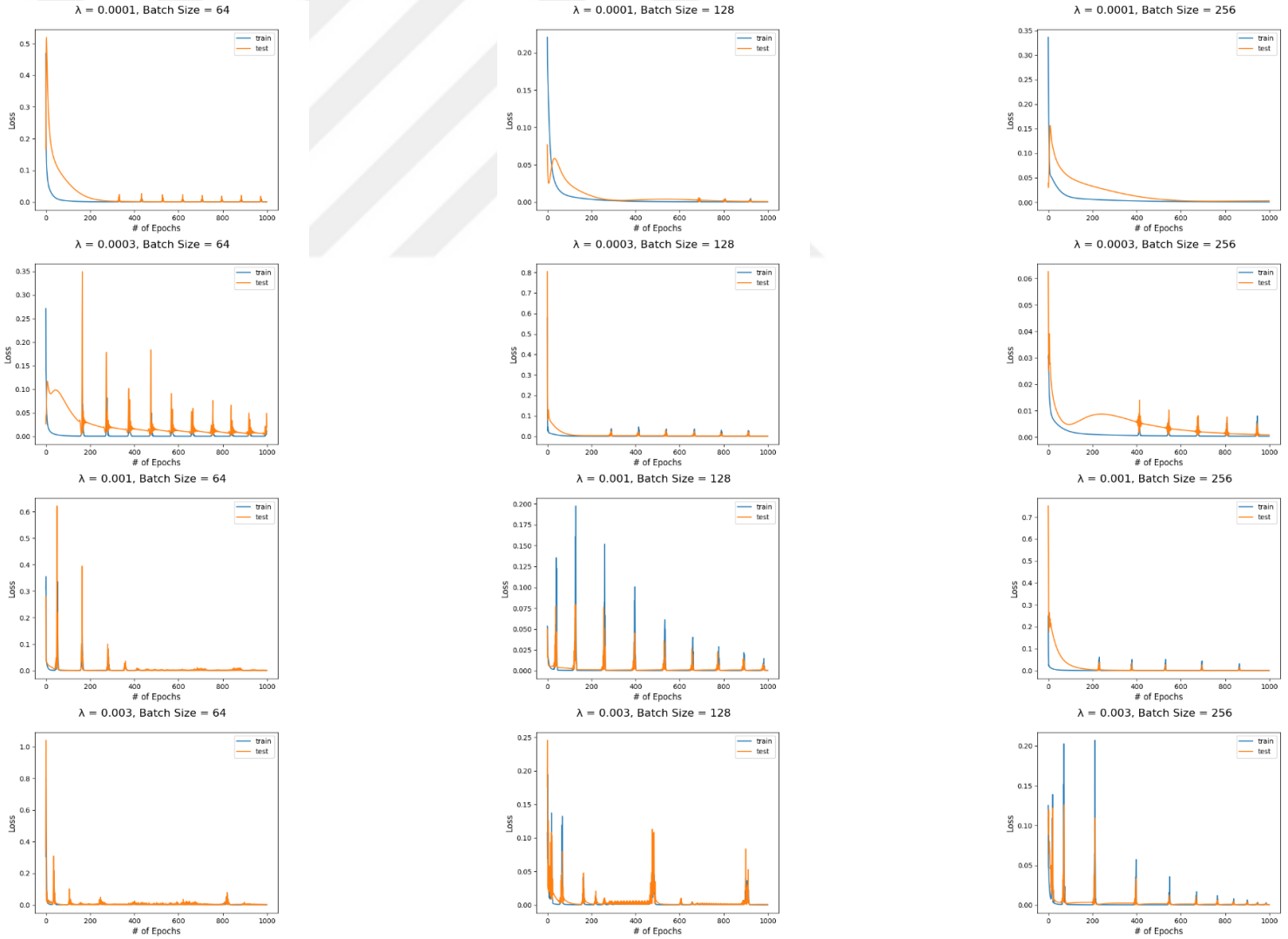
Source: Author's estimation.

Figure A.19: The estimated loss for China in NARX



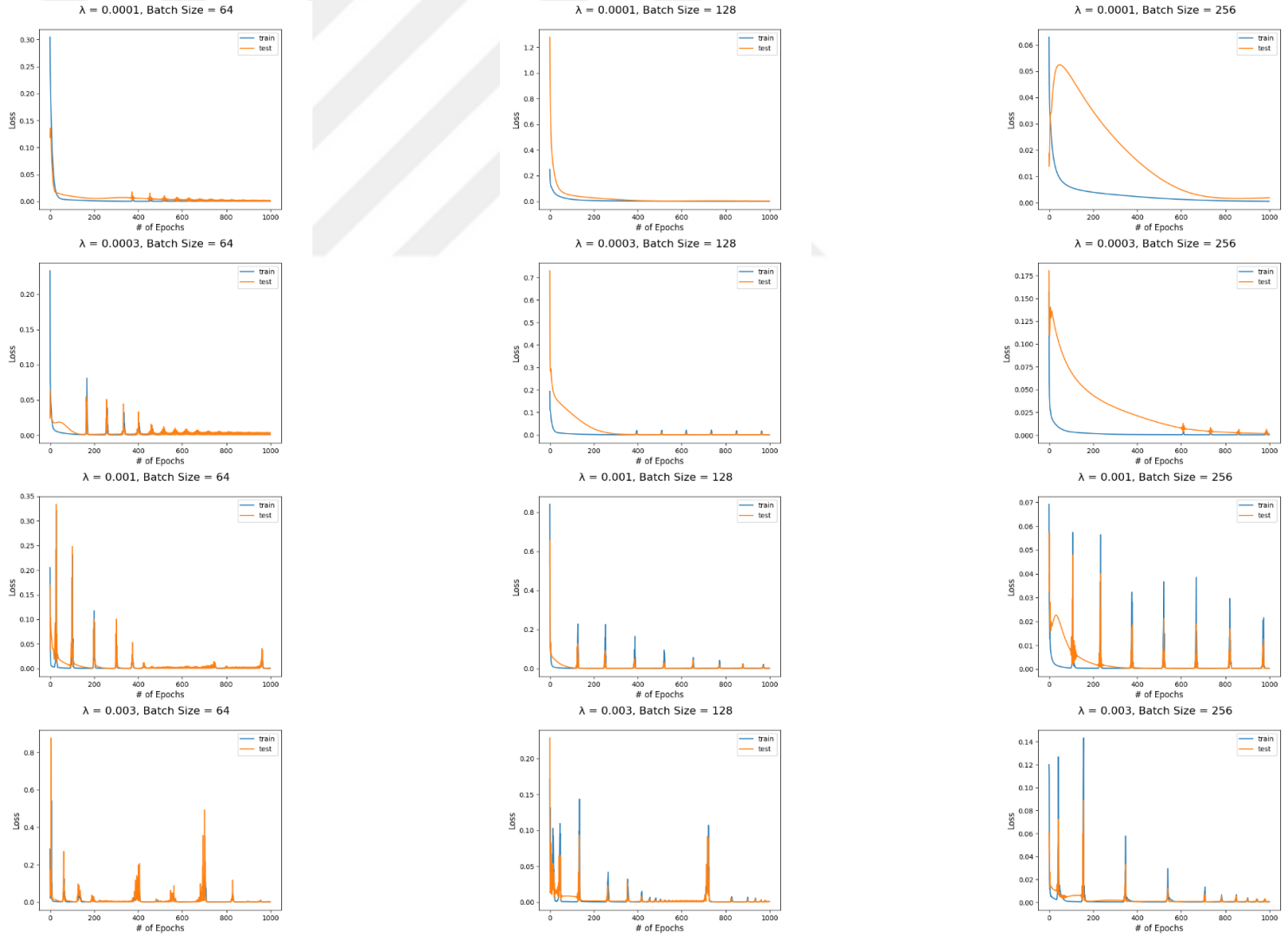
Source: Author's estimation.

Figure A.20: The estimated loss for South Africa in NARX



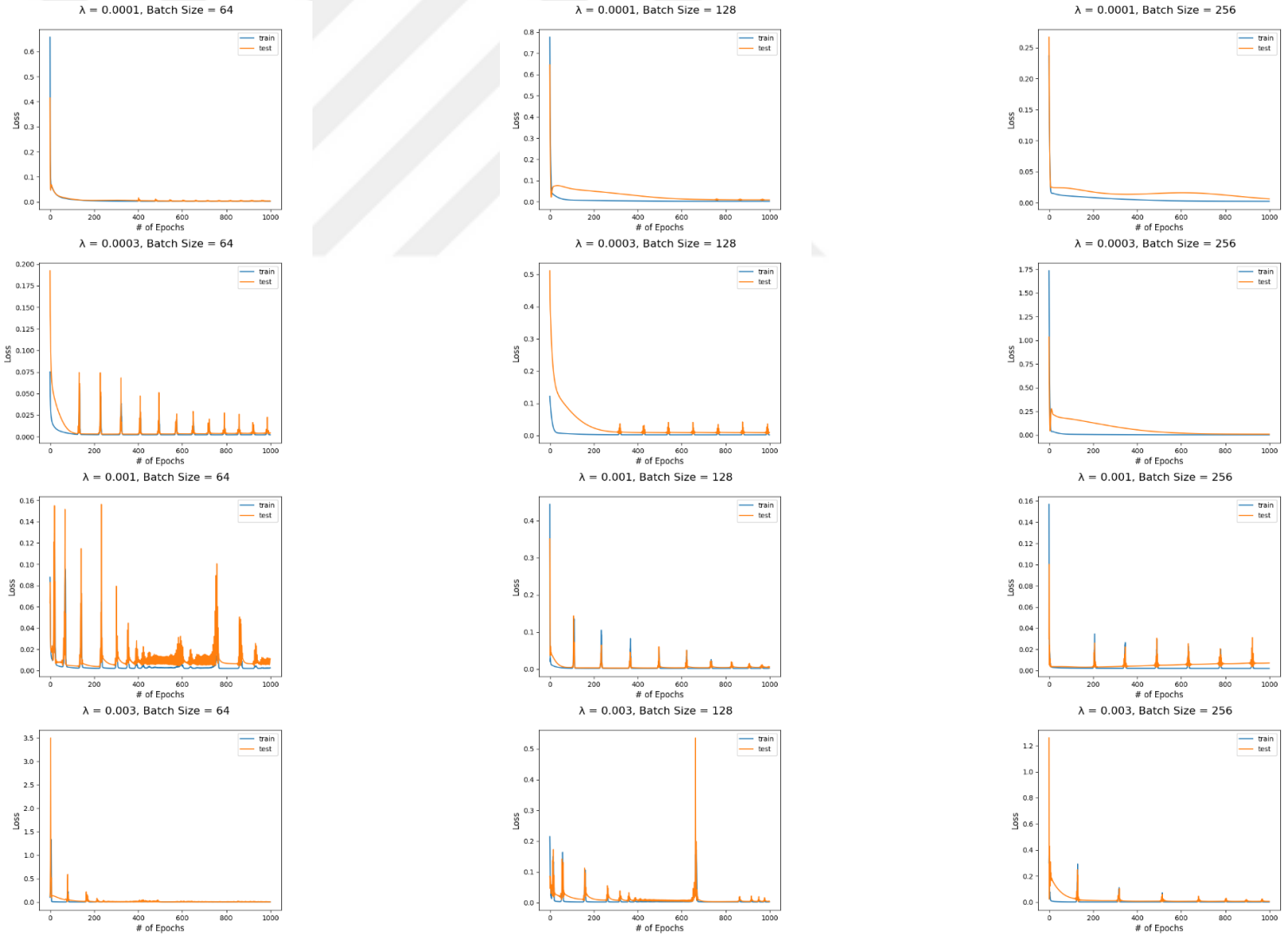
Source: Author's estimation.

Figure A.21: The estimated loss for Mexico in NARX



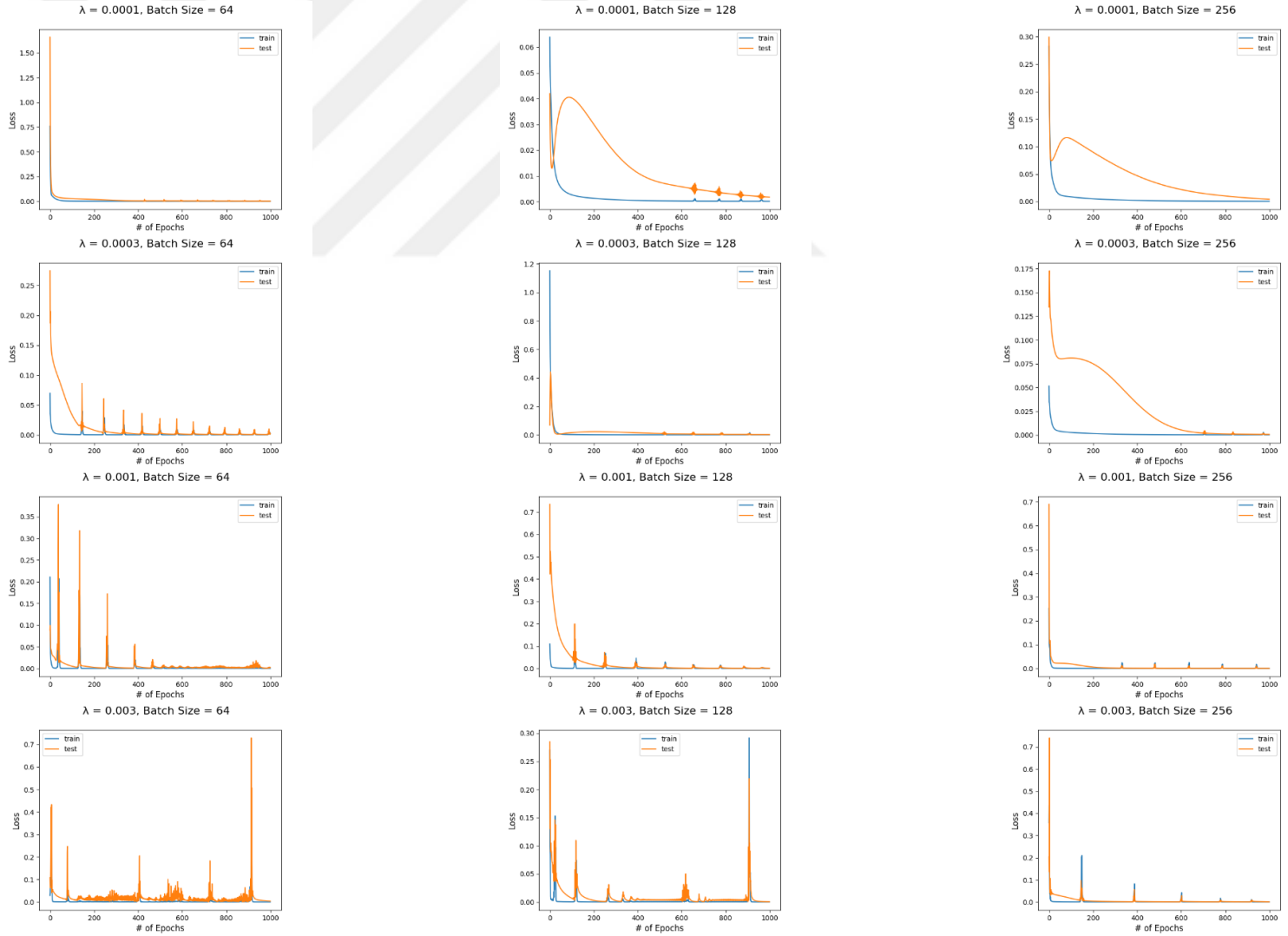
Source: Author's estimation.

Figure A.22: The estimated loss for Indonesia in NARX



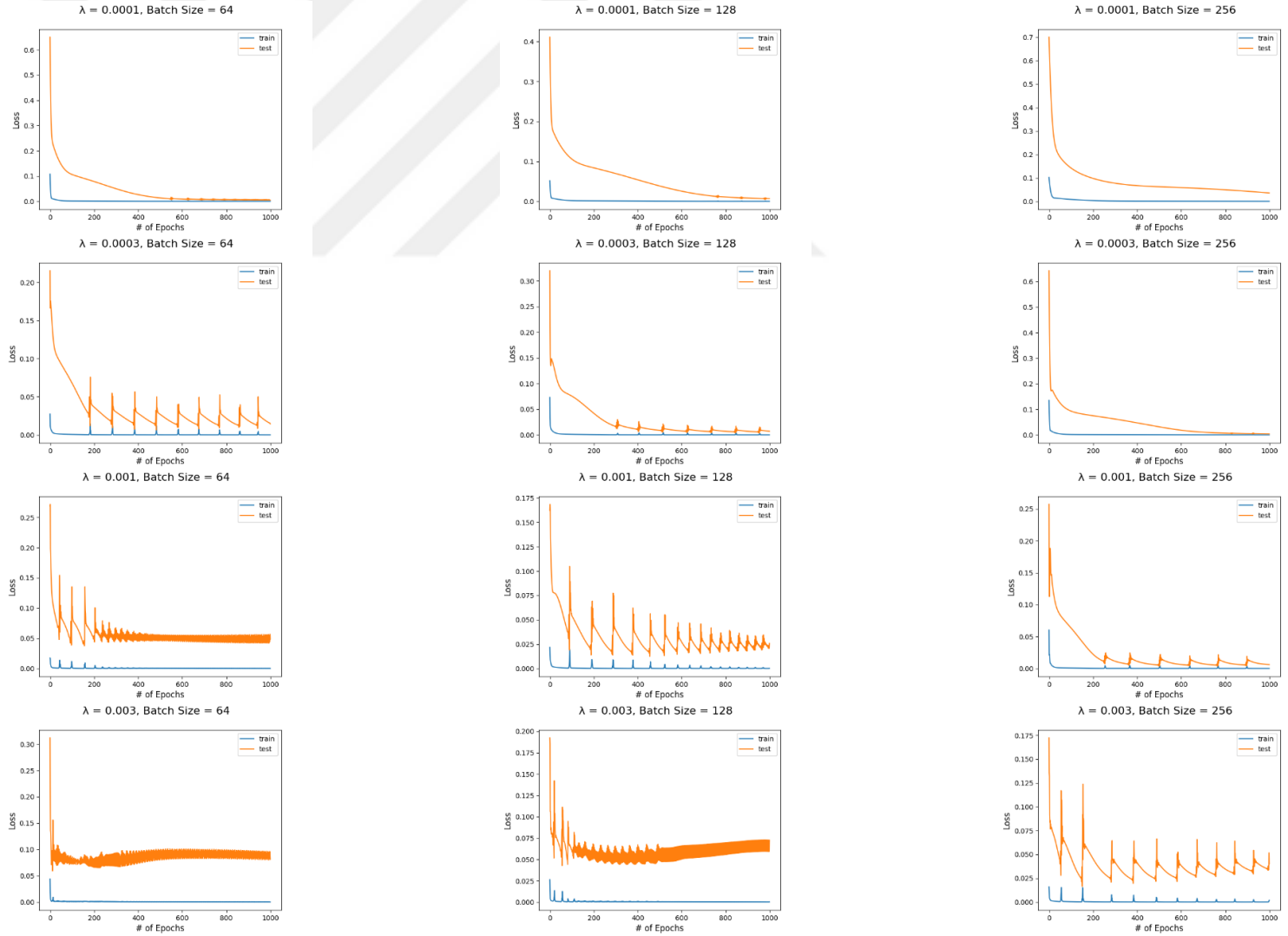
Source: Author's estimation.

Figure A.23: The estimated loss for Turkey in NARX



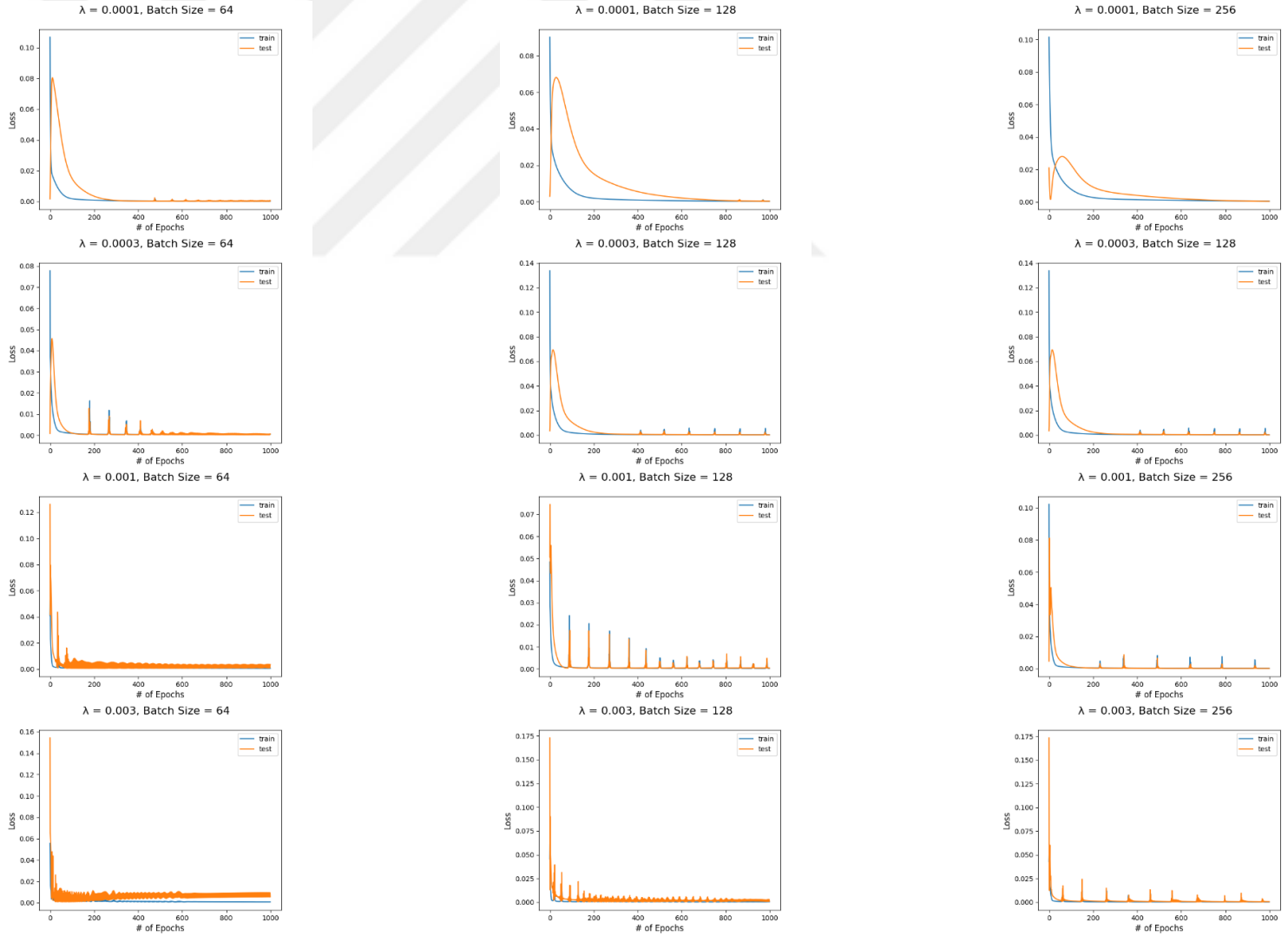
Source: Author's estimation.

Figure A.24: The estimated loss for Brazil in LSTM



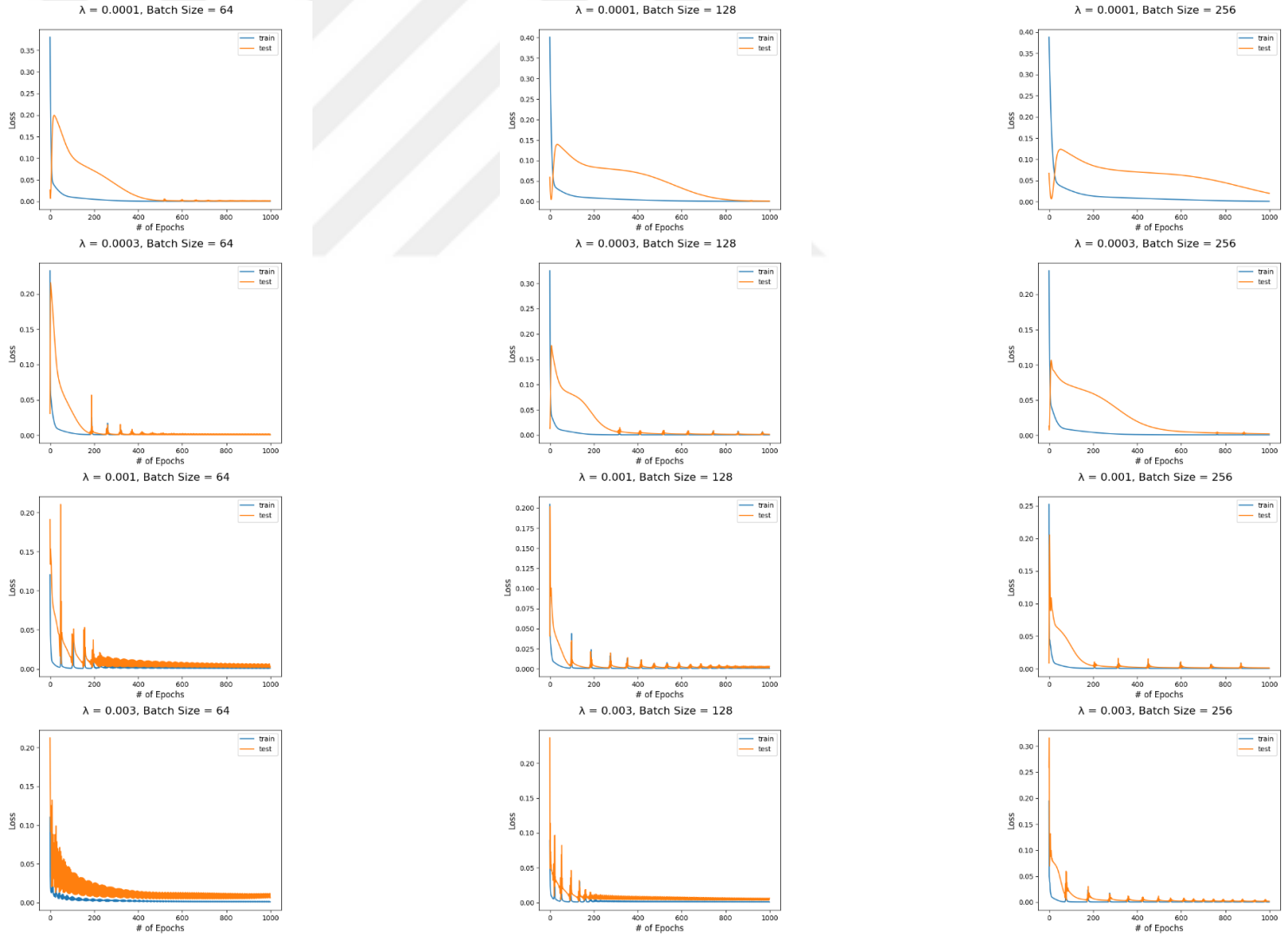
Source: Author's estimation.

Figure A.25: The estimated loss for Russia in LSTM



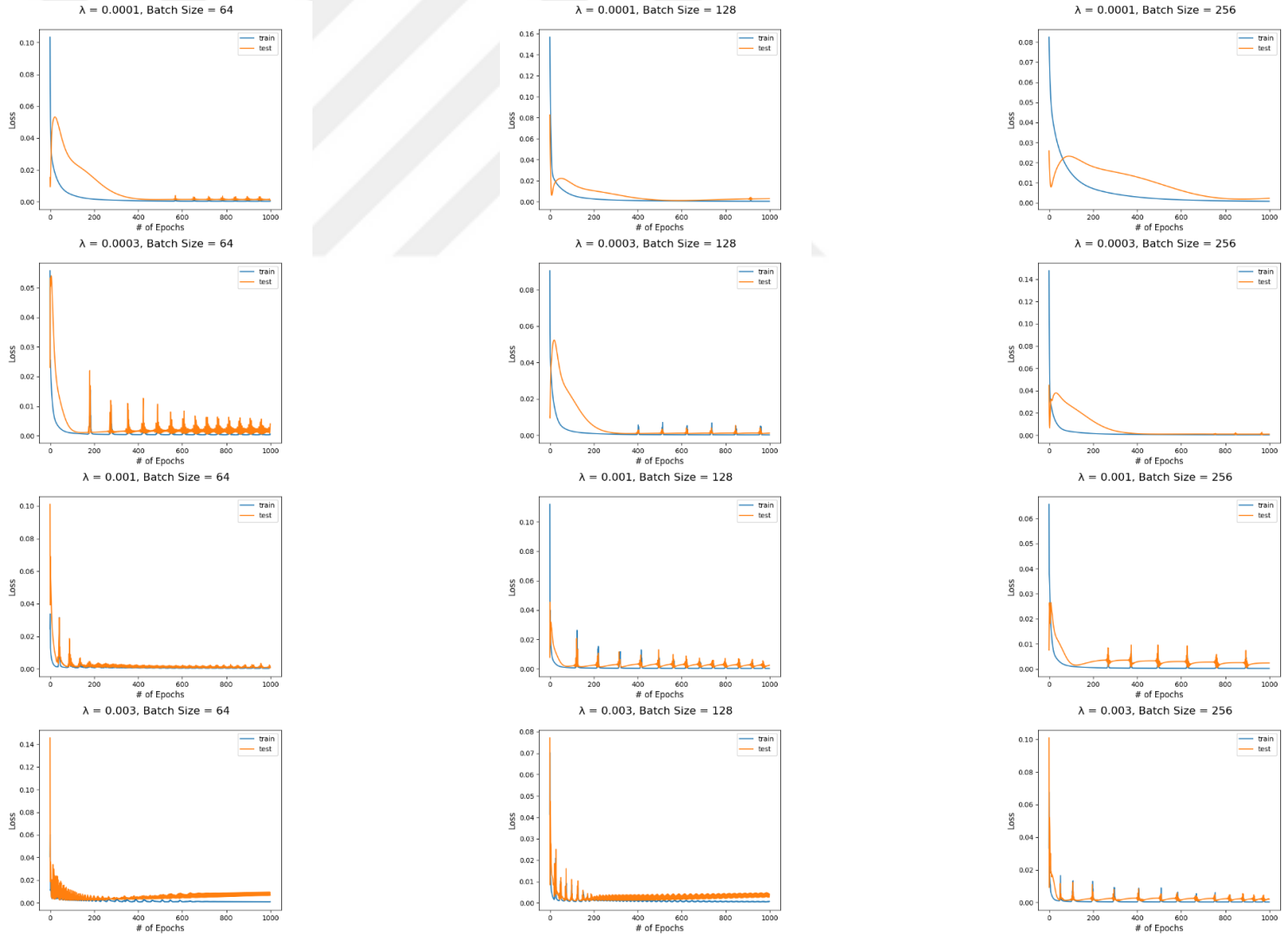
Source: Author's estimation.

Figure A.26: The estimated loss for China in LSTM



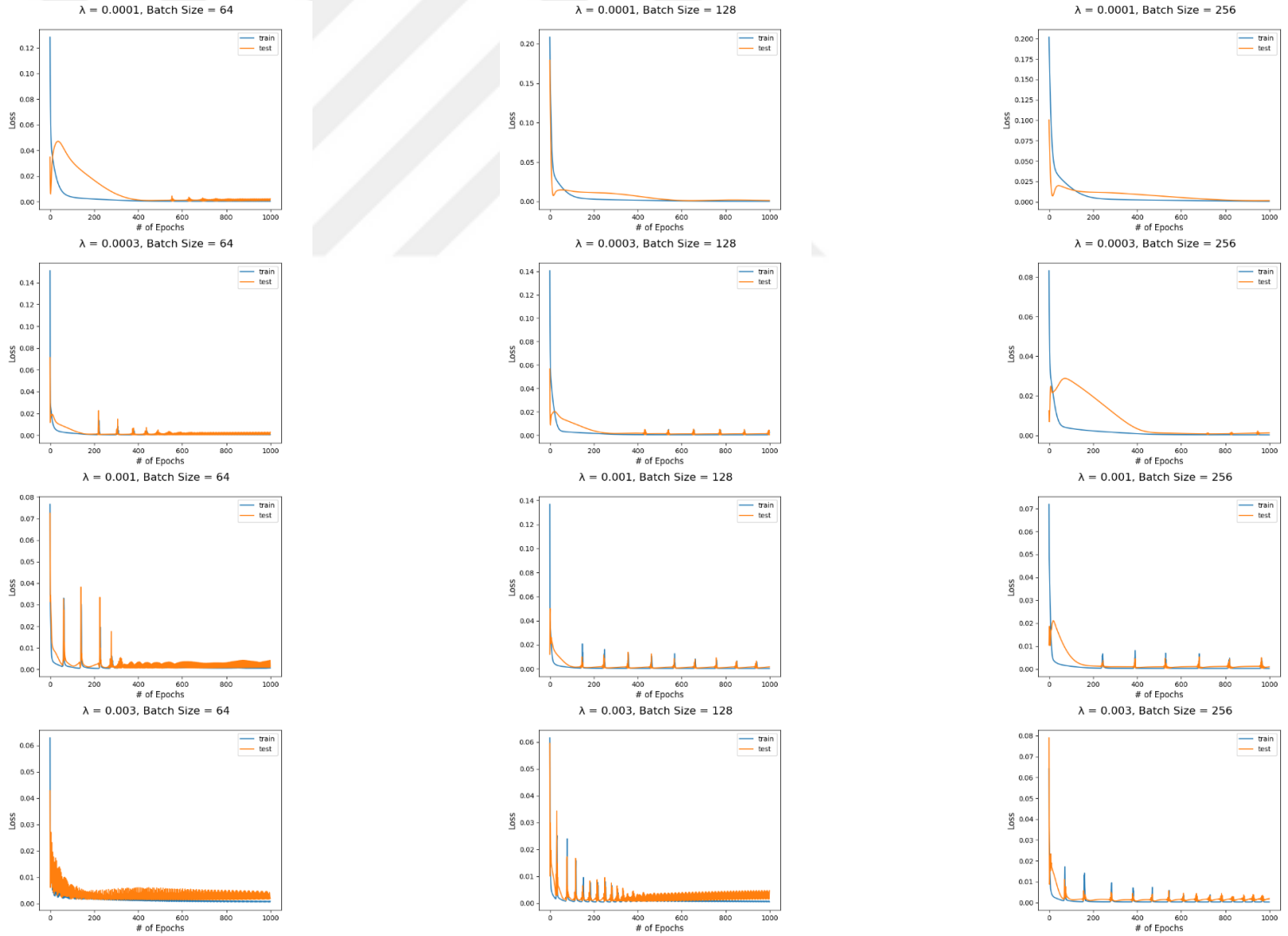
Source: Author's estimation.

Figure A.27: The estimated loss for South Africa in LSTM



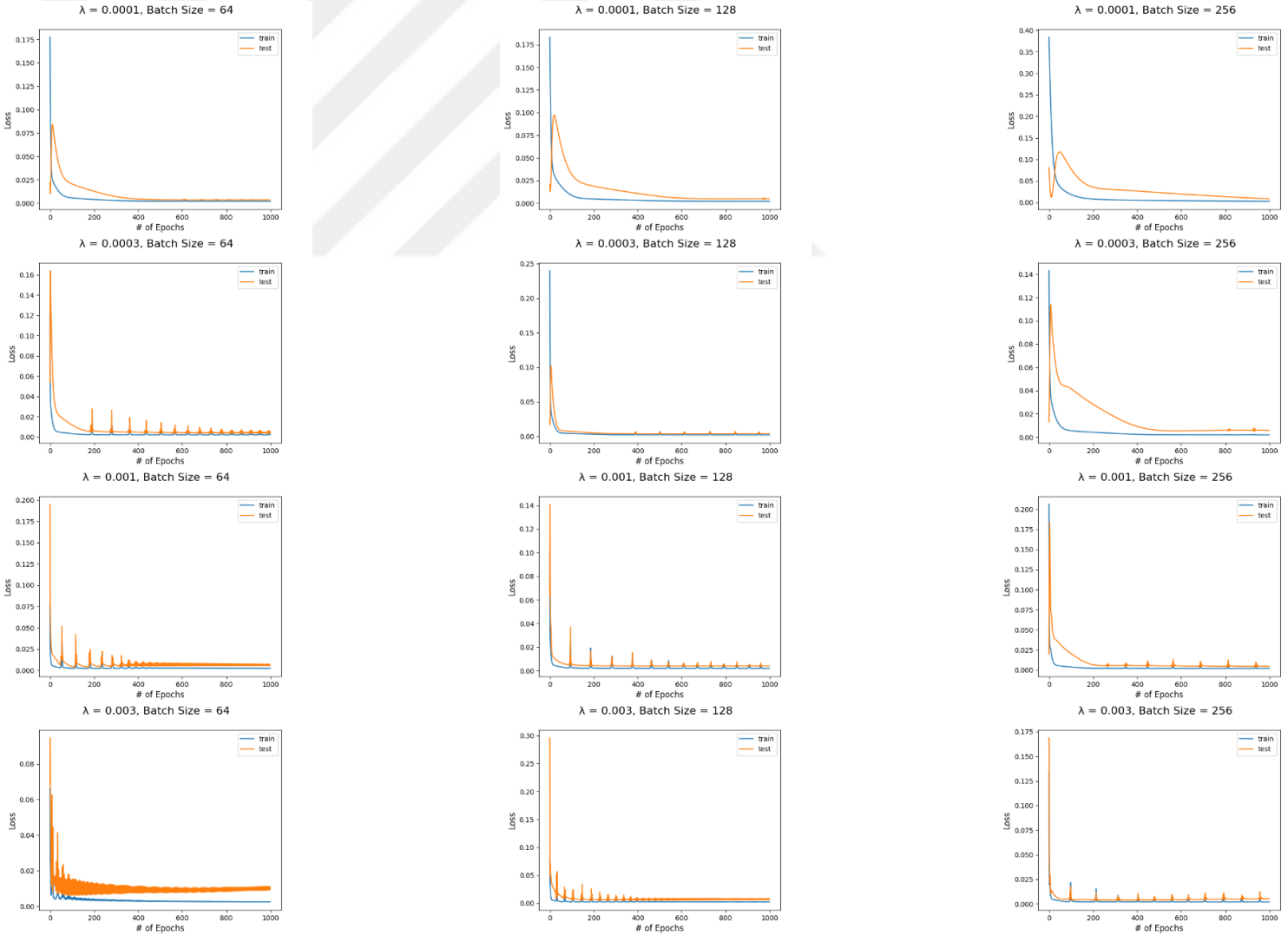
Source: Author's estimation.

Figure A.28: The estimated loss for Mexico in LSTM



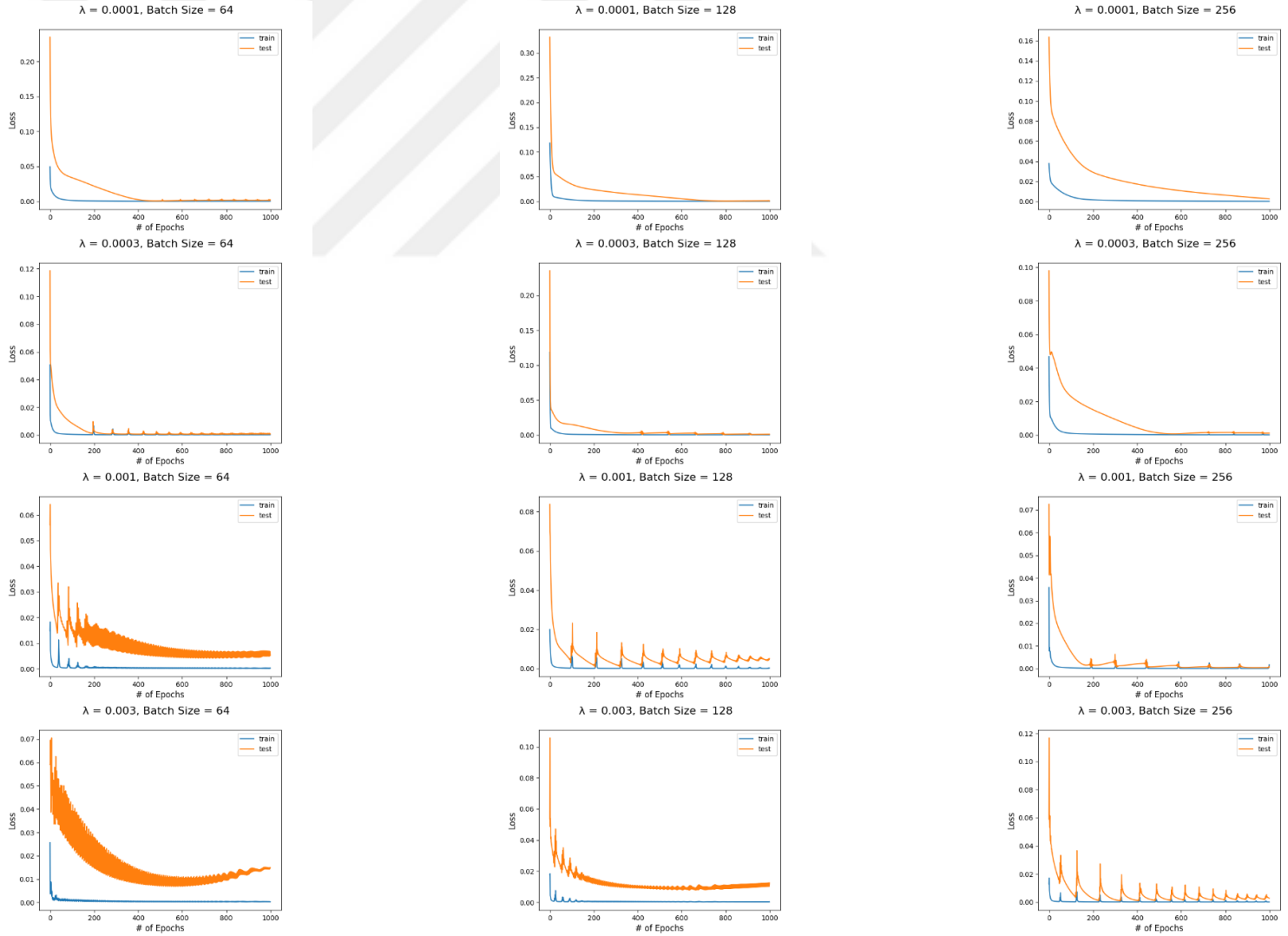
Source: Author's estimation.

Figure A.29: The estimated loss for Indonesia in LSTM



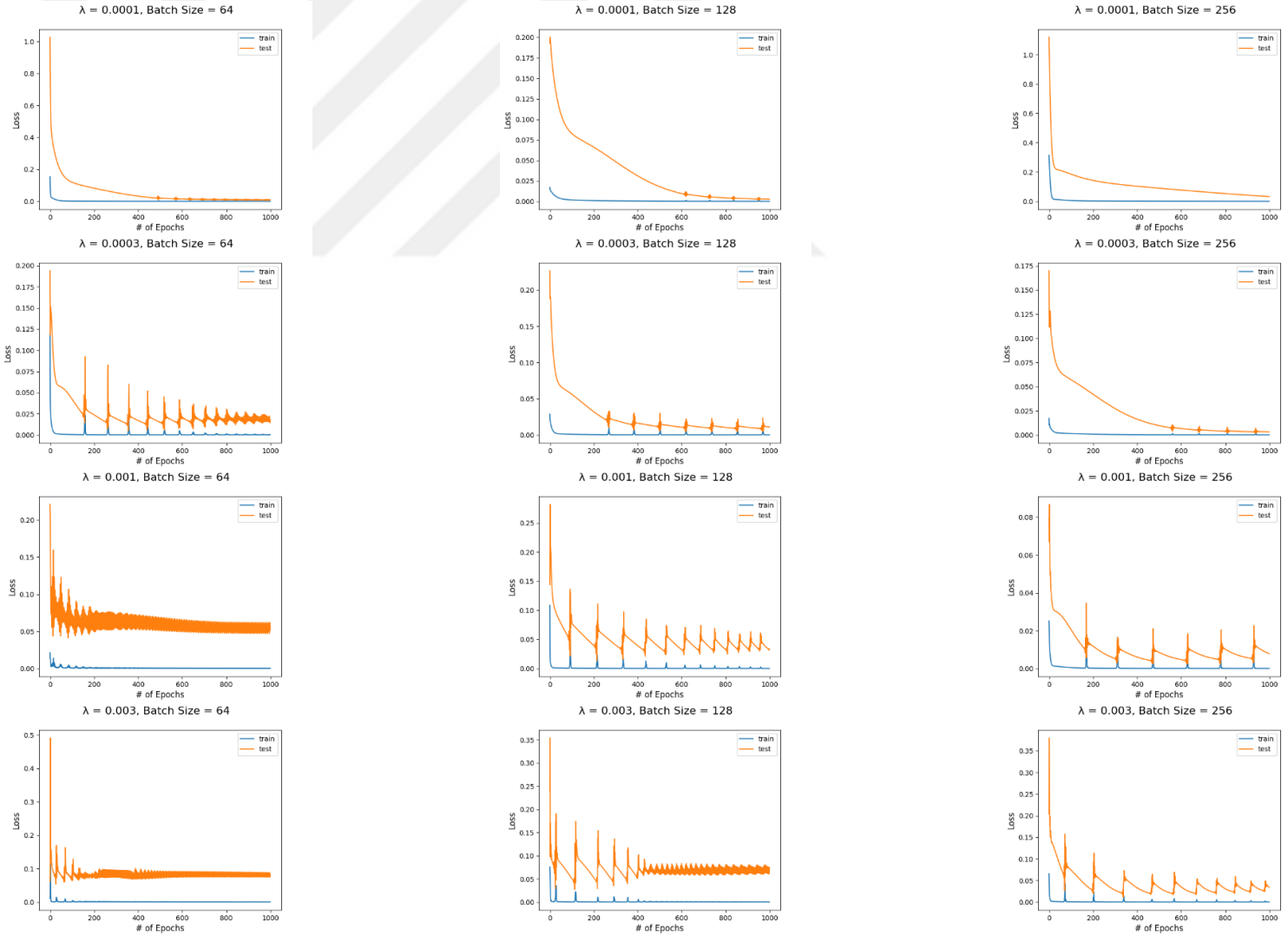
Source: Author's estimation.

Figure A.30: The estimated loss for Turkey in LSTM



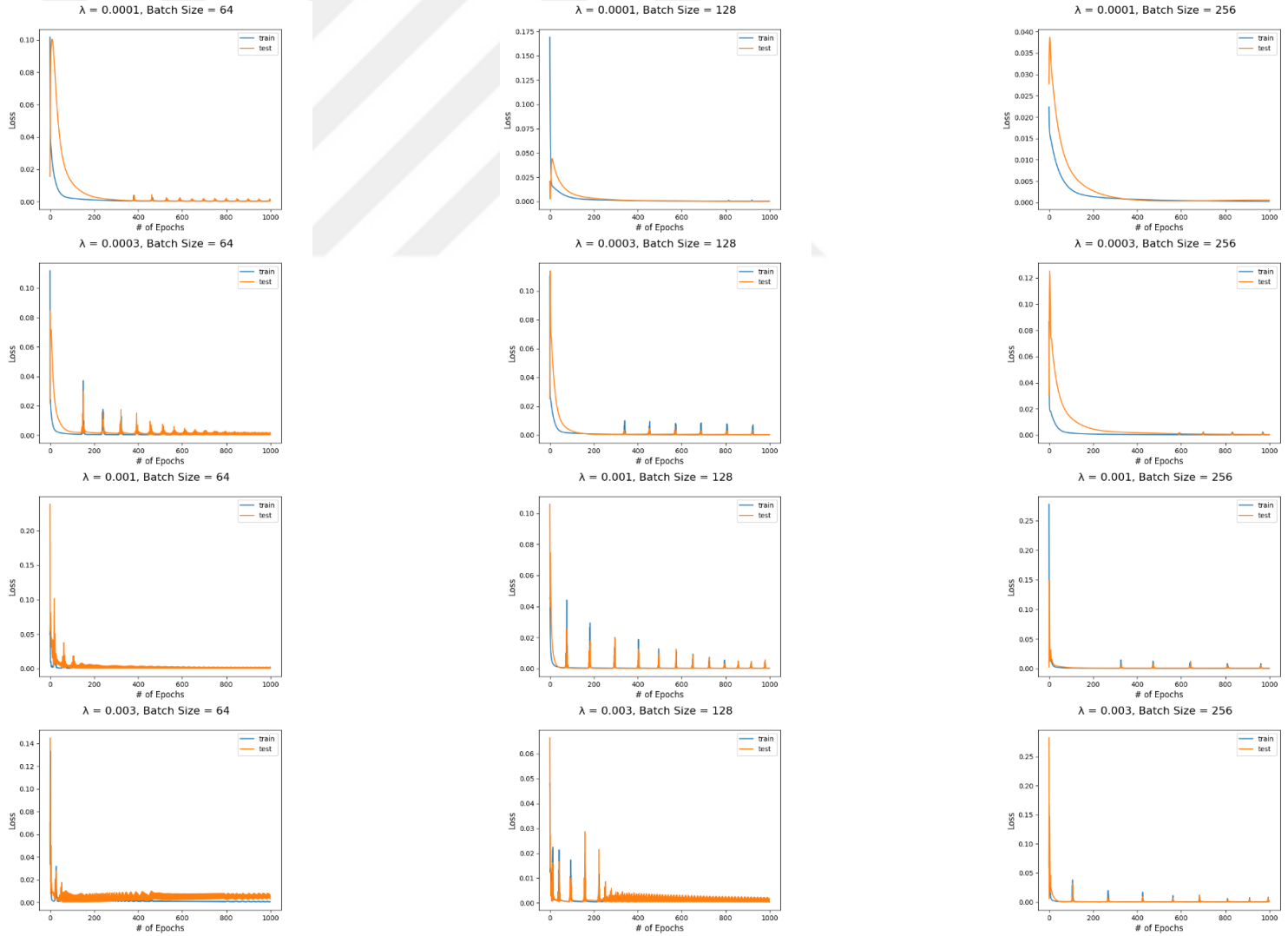
Source: Author's estimation.

Figure A.31: The estimated loss for Brazil in GRU



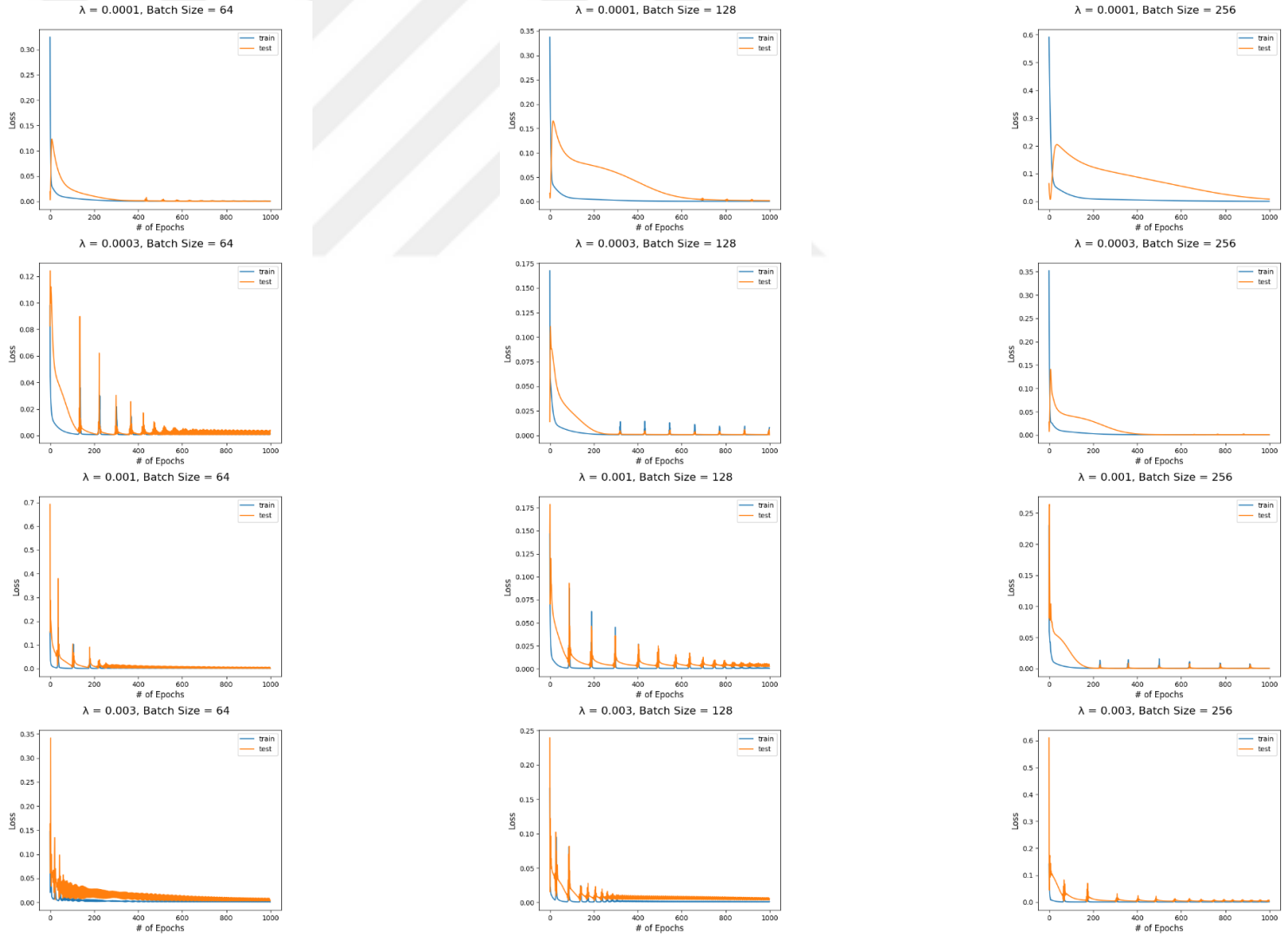
Source: Author's estimation.

Figure A.32: The estimated loss for Russia in GRU



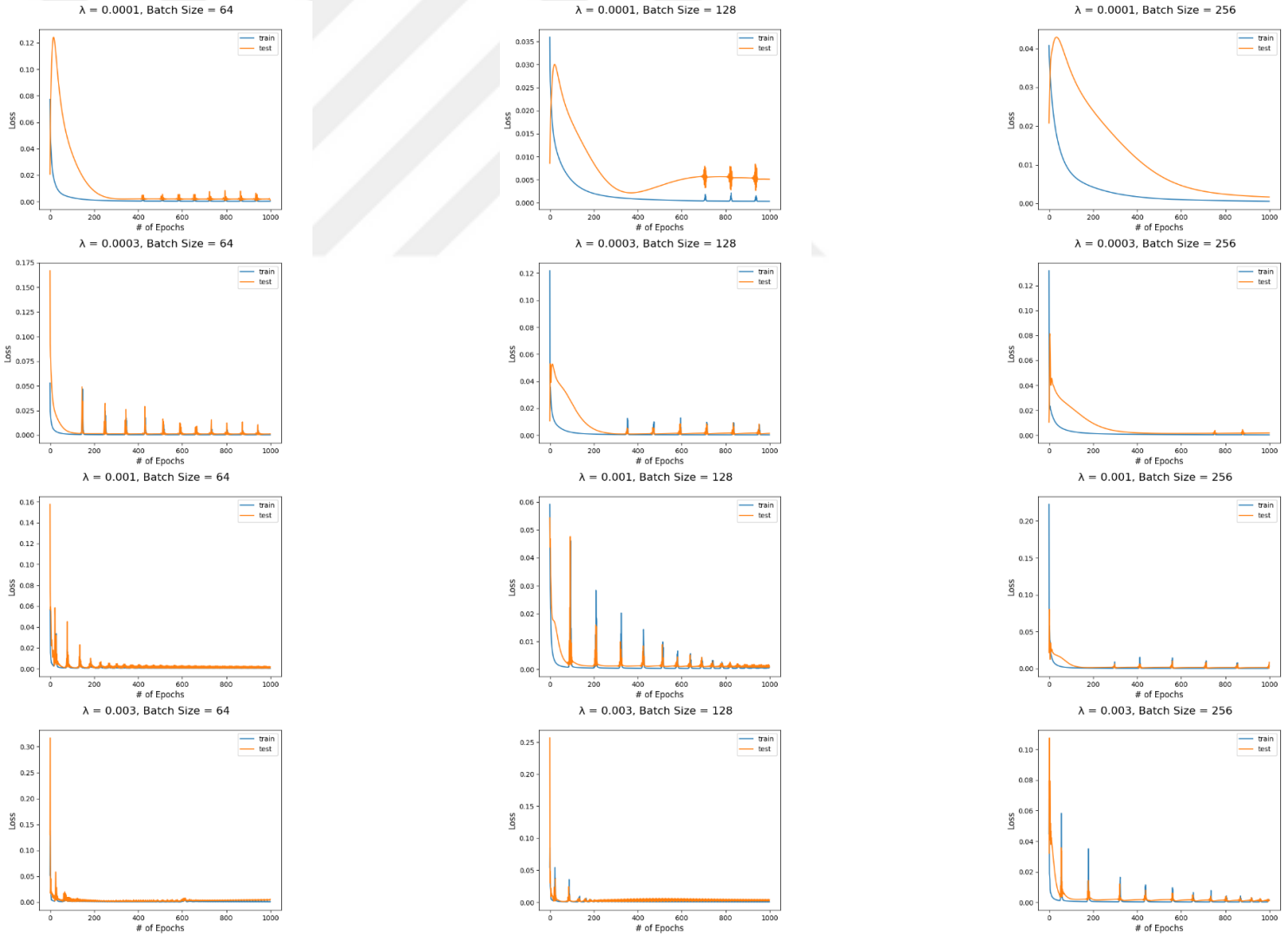
Source: Author's estimation.

Figure A.33: The estimated loss for China in GRU



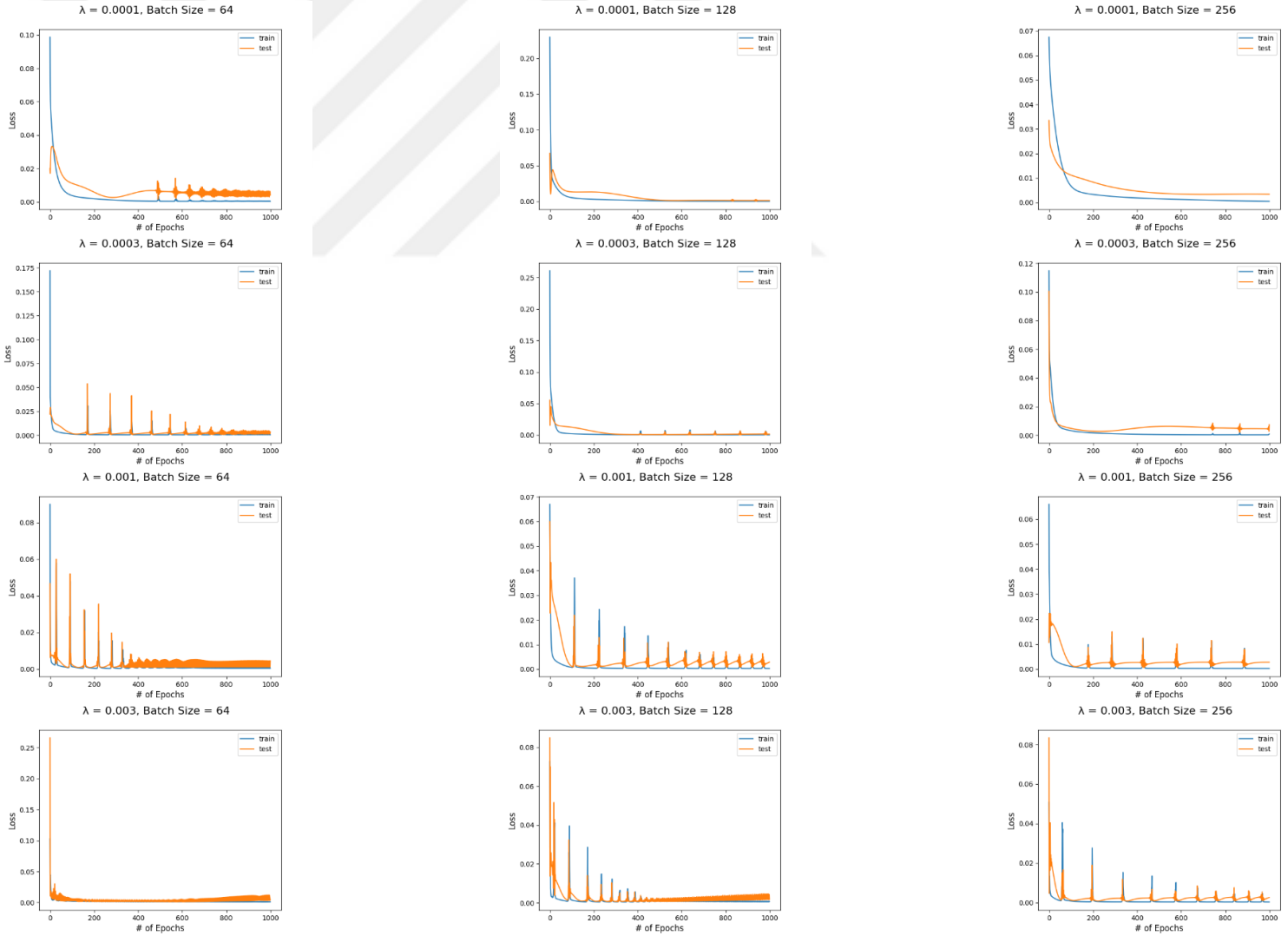
Source: Author's estimation.

Figure A.34: The estimated loss for South Africa in GRU



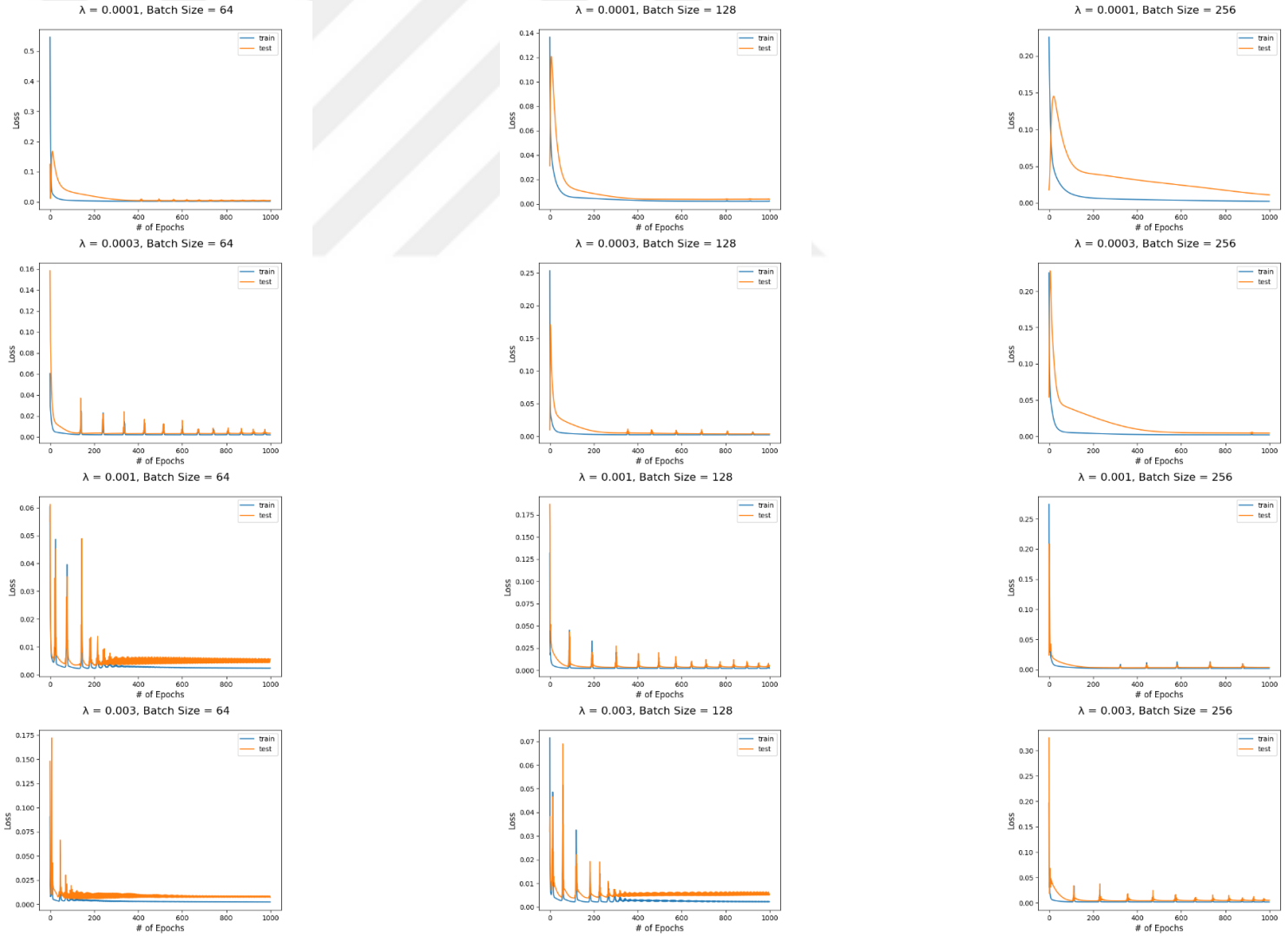
Source: Author's estimation.

Figure A.35: The estimated loss for Mexico in GRU



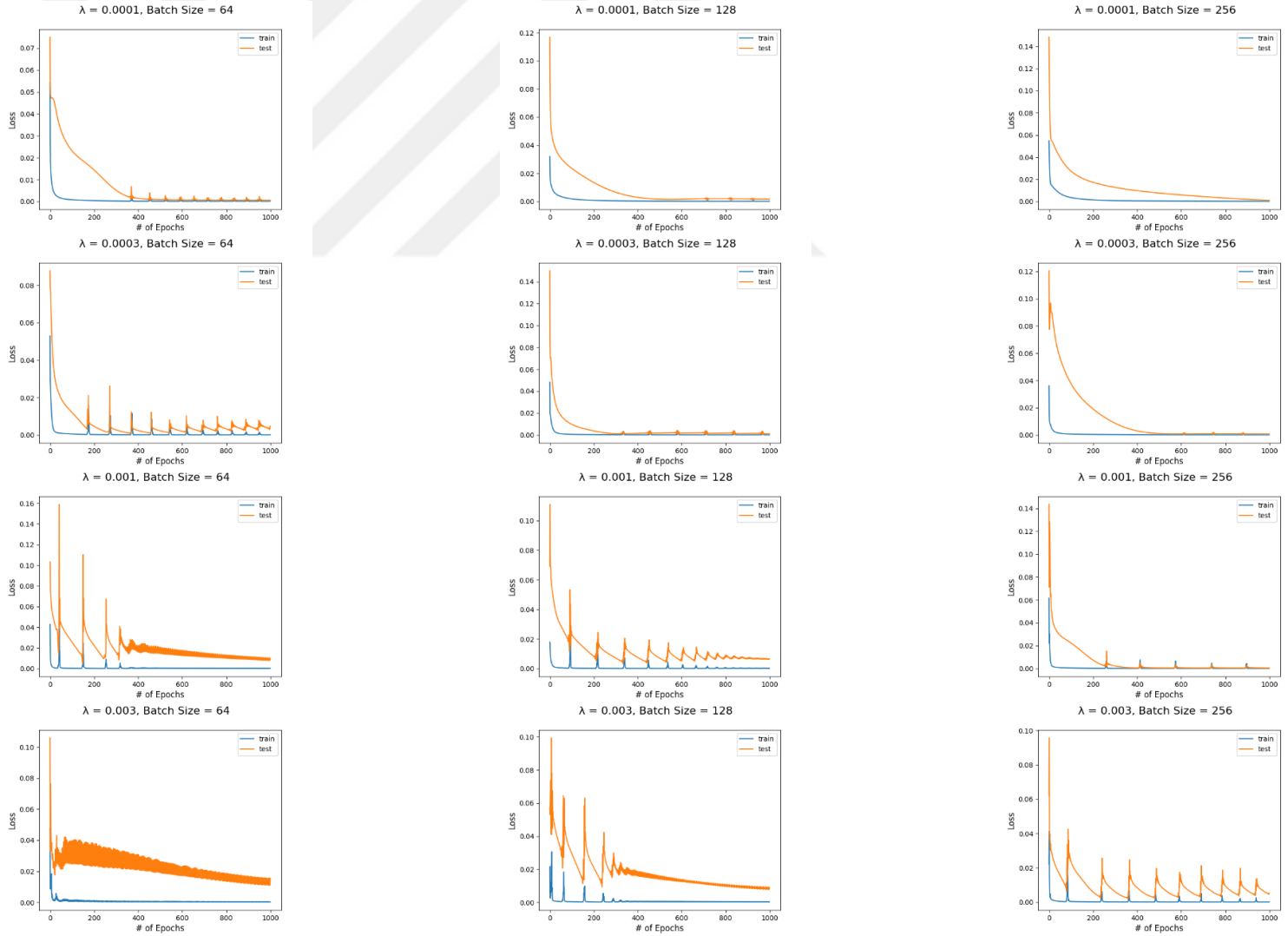
Source: Author's estimation.

Figure A.36: The estimated loss for Indonesia in GRU



Source: Author's estimation.

Figure A.37: The estimated loss for Turkey in GRU



Source: Author's estimation