MÄLARDALEN UNIVERSITY
SWEDEN

School of Education, Culture and Communication

Division of Applied Mathematics

MASTER THESIS IN MATHEMATICS / APPLIED MATHEMATICS

Stochastic Model Identification and Model Metrics with Deep
Learning Applications

by

Berk Alp Yilmaz

Master Thesis in Applied Mathematics

DIVISION OF APPLIED MATHEMATICS

MÄLARDALEN UNIVERSITY

SE-721 23 VÄSTERÅS, SWEDEN

MÄLARDALEN UNIVERSITY
SWEDEN

School of Education, Culture and Communication
Division of Applied Mathematics

Master thesis in mathematics / applied mathematics

Date:

2020-01-28

Project name:

Stochastic model identification and model metrics with deep learning applications

Author:

Berk Alp Yilmaz

Supervisor(s):

George Fodor

Cosupervisor(s):
Olha Bodnar

Reviewer:

Milica Rančić

Examiner:

Anatoliy Malyarenko

Comprising:

30 ECTS credits

Contents

List of Figures

Abstract

Each economic event has a dynamic effect on the market due to decision complexity for agents valuating rare events. The more complex is the event, the more it takes for the markets to incorporate the new information. The goal of this thesis is to explore the dynamic effects of economic events on equity returns and statistical indicators. Furthermore we use train artificial neural networks to classify price run-ups around economic events. Applications will allow explanation of the effects of a economic events on equity returns and its statistical properties.

Keywords: Event study, Abnormal Returns, ARIMA, Kalman Filter, Deep Learning

Chapter 1

Introduction

1.1 Background & Motivation

Financial instruments, from simple loans to complex derivatives are today an essential enabling factor in any economic activity. Continent-wide economies and also our everyday lives depend on how projects such as roads, hospitals, schools, factories and houses can be effectively financed. The financial market effectiveness is always related to *information* and *risks*: a financial market can perform an optimal distribution of the available investment resources if all relevant information is known by all players and it is translated into prices at low costs (low friction). If the market would not be efficient, some actors could win financial gains by arbitrage. Eugene Fama, a Nobel prize winner in economics, introduced the *Efficient Market Hypothesis* (EMH), which states that no substantial gains can be made in any efficient market and consequently the remaining gain variations are stochastic and contain no gain-inducing information.

EMH created large debates both at theoretical and practical level. An investment fund company Renaissance Technologies, (there can be found many such examples), had substantial gains even during the 2008 crisis by using investment strategies based on mathematical concepts, thus contradicting the predicted EMH. In any case, it can be observed that time series from markets are clearly not purely stochastic in the same way as throwing dice. A typical market behavior is that at any given time the next value will be stochastic (as EMH predicts) but will not depart much from the existing value. A large deviation usually will not be obtained not frequently, but only after a path consisting of several smaller deviations. This was an observation done already in 1900 in a PhD thesis of Louis Bachelier, an observation that was lost and rediscovered after many years. Mathematically, this behavior is described by the so called Wiener processes describing Brownian motions in physics. Today there are available powerful mathematical tools to identify the characteristics of such a market using recursive models in cases when normal assumptions (e.g. non-stochastic volatility) are both fulfilled

and not fulfilled.

However, Wiener and other recursive processes do not describe accurately the behavior of the market, specifically under unexpected critical events such as dynamic structural breaks. For example the stock market crashes in 1929, 1987 and 2008 had a rather sudden downwards trend. Other market trends such as when a new technology is introduced (Machine Learning, Artificial Intelligence or e-commerce recently), give to the market a specific behaviour that is not described by traditional models, but it has a dynamics that is slower compared to a typical market crash. This was observed for example in the research of Benoit Mandelbrot[38]. Analyzing and classifying market changes that depart from the typical statistical market models can give substantial benefits in predicting and recognizing the effects of specific economic events on the market. It is the goal of this thesis to model and identify non-standard economic events that influence the market in a way that cannot be well described by traditional statistical models. Since such effects are hard to be identified by observations, we use machine learning methods trained with dynamical and other models that capture these non-standard economic events. We verify the correctness of our models using actual market data. The following example illustrates the usefulness of the method.

1.2 Example

A illustrative example of the interconnections between different markets and economic events is the collapse of US housing market. It can be seen that in Figure 1.1 a prediction using statistical modeling (blue line) will not describe correctly the actual market behavior of the "New one family" data series, which stands for houses sold in USA (black line). Such departures are often defined in the literature as structural breaks[5]. The change of the dynamics of the underlying market can be clearly seen from the graph as a significant departure from the fitted statistical model (like a breaking point from its structure). Even if the upcoming event is well known, namely the fraudulent rating of the housing market, it would still have been difficult to predict the effects of a bubble on the housing market and the related markets. If a similar event had occurred in the past and its effects analyzed mathematically it would be a useful tool for prediction of risks.

Measuring the effects of this change on other markets is however not as easy as it seems by this graph. Therefore we use more powerful methods described further in this thesis.
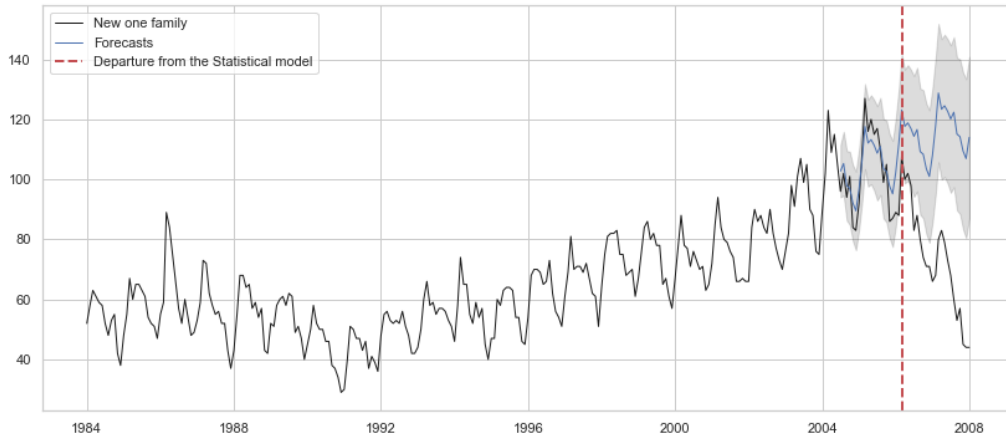
Figure 1.1: Effect of a economic event on a statistical model

1.3 Research Problem

Stochastic processes in financial markets are often described as statistical models of time series such as ARIMA or GARCH, being in essence dynamical models with stochastic components. Many of these methods have assumptions on perseverance of statistical properties of the underlying time series, sometimes obtained by different transformations. If the models have reasonable assumptions and desirable goodness of fit measurements, these models are considered to be successful, thereby the pricing/analysis are deemed to be correct.

However, the market can move in unexpected ways that contradict these assumptions. In these cases the models give local forecasting errors, most often the differences are hard to detect. This thesis focuses on the classification, analysis and predictions of effects of such events using a method of high sensitivity: classification of model deviations using artificial neural networks.

1.4 Research Methodology

The research methodology is based on using traditional econometric modeling for both normal time series (e.g. stock prices) and time series after unexpected events. A labeling of the events is done, where normal cases are labeled as zero and the unexpected events are graded subjectively after their believed degree of disruption on the market. To show the effects of unexpected events, various analysis is done. For some effects, which are complicated to capture with statistical models, deep learning methods are used. The quality of the network training

is estimated again with cross validation. The results from various labeling and analysis is presented with goodness of fit measures if possible.

The methodology is based on identification of a traditional statistical model and comparing this model with effects of known non-standard events. Specifically the work is split into three main parts:
(1) Understanding the data & identifying the right statistical model.
(2) Analyzing & measuring the effects of economic events.
(3) Training neural networks with various inputs in an attempt to train the effects of economic events to the neural network.

To see how would the time series evolve without an unexpected economic event, we need to use right statistical models that capture most of the movements in the time series. Evaluating the models is done through various goodness of fit: methods to measure the performance of statistical models. Using goodness of fit, we will compare different statistical models and evaluate performances on different financial instruments.

Figure 1.2: Sudden drops in the ABB stock prices

Example The plots given in Figure 1.2 show prices of ABB stock in 2016. Numerous events have effected the market price of the company, with two price drops. To point out the significant economic events, two biggest outliers of returns are shaded red (outliers of squared returns). Two shaded areas are possibly the result of disappointing revenue figures in the quarterly reports of the company[1]. One of the biggest price increases of the company is just after the first shaded area (August 2016) but is not detected since it was an increase over many days and not a single jump. This is a good example that shows the effects of irregular events on the price of a company: whether it is a single day drop or a continuous loss over days, the effects of these events are rather complicated to analyze. The movements can be caused by

---

[1]https://new.abb.com/investorrelations/financial-calendar

both internal (company-specific) and external factors (different markets, political events etc.). Analyzing these effects requires careful examination and consideration of many aspects.
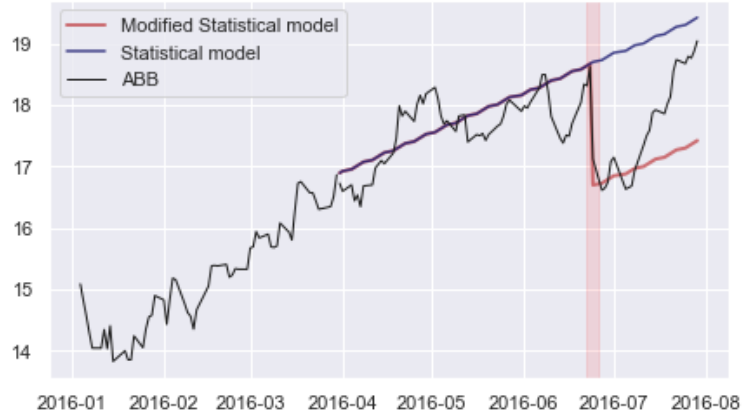
Figure 1.3: Modified ARIMA to fit into event

We fit a statistical model (ARIMA) to forecast the price, seen in Figure 1.3. Even though the forecast isn't very successful, there is a clear deviation from the model around the economic events. A straightforward approach to simulate the economic event is to manually change the predicted prices (Red in Figure 1.3) as demonstrated in the figure. Even though this method can't be used before the event takes place, this intervention results in a better fit of the model, which we will use to simulate price movements similar to these events.

1.5 Literature Review

Literature often defines these events as "structural breaks" where the "breakdates" are the moments where models give big forecasting errors. This definition only has a meaning if discussed with an underlying model. There are many popular methods that tests the existence of structural breaks[6], [8], [27]. Existence of multiple tests shows that even defining or locating these breaks is not a straightforward task.

Estimating & analyzing impacts of these events on the properties of time series require more complicated methods than tests used to locate them, specially when the systems are complex ones, like financial markets. Buyuksahin and Robe use a unique, non-public dataset to show that the cross-market effects between commodity and equity prices increases with hedge fund activity, indicating a complicated relationship between the two markets [12]. Correlations have increased after 2008 and increases with financial distress indicators (TED spreads). The

paper shows that these changes in these connections are hard to measure and predictive powers are unstable when even non-public information is used, since there are many dynamic factors to consider. This signals a potential use for deep learning methods and shows the complicated effects of a single structural break.

To analyze these connections, Bekiros et al. [9] use complex networks and entropy-based centrality measurements for these markets before and after the economic crisis. Resulting correlation and entropy measurements show significant disparity between pre and post crisis. This means that last financial crisis created significant differences in the structural properties of the financial systems, possibly due to realization of diversification benefits of holding commodity futures by financial investors.

However, Daskalaki, Skiadopoulos challenge these diversification benefits using spanning tests that comply with the preferences of both mean-variance (MV) and non-MV investors [17]. Furthermore, pre-2008 era is also tested separately, for testing different market dynamics and demonstrating the effects of holding commodities over turbulent periods. Even though commodities can serve as an inflation hedge, traditional portfolios show better results in the in-sample tests only with the exception of 2005-2008 commodity boom period. It is found that commodities are beneficial to only non-mean variance investors and these results are not preserved out-of sample. Authors tested the pre-2008 era separately, indicating the importance of structural changes of the asset prices when considering portfolio risk and hedging.

In Narayan, Narayan and Sharma, the authors examined cross-market relationships between futures and commodity markets for a different purpose: for prediction and testing profitable technical trading strategies [45]. For relationship between futures and the spot, work of Stoll and Whaley is used: ($F_t = S_t e^{(r-d)(T-t)}$) [56][2]. This shows a change in futures price should create instantaneous effects on its underlying commodity. Using this relationships and a predictive regression model with structured breaks proposed by Bai and Perron [8], authors predict oil, gold, silver and platinum (75% of total trading volume in commodity markets) daily prices. Results show a strong evidence of predictability. Using different technical rules similar to Ratner and Leal [49], authors demonstrate profitable strategies that conclude possible profits are highest in the oil and silver markets using linear models. Results from dynamic trading strategies however are somewhat different. With dynamic strategies, not only rankings based on profitability are different but also platinum and gold are unprofitable. Both technical trading rules reveal that profits were lowest during the financial crisis but more importantly, are regime-dependent (structural breaks matter). Since structural breaks effect the profits from such strategies, models that predict these structural breaks can generate economic profits for practitioners using these technical trading methods, allowing them to exit trades before anticipated economic events in order to avoid possible loss situations.

Adams and Glück go deeper and investigate the reasons behind the change of the depend-

---

[2] Where F_t is the future price and S_t the spot price, $r - d$ is carrying cost, $T - t$ is time remaining.

ence between equity and commodity markets after the financial crisis [1]. Paper shows risk spillovers from equity to commodities were nonexistent before the crisis. Papers show state that there are co-movements even in unrelated assets. Authors believe increasing dependence of the two markets is the result of change in the investment style of the participants, hypothesizing that commodities became a bigger part of portfolios, and commodities became just another category inside a universe of stocks. The term "financialization" is used for increased investment in commodities. Similar view of Cheng and Xiong [14] who report that traders "treat commodity futures as an asset class just like stocks and bonds" further supports this hypothesis. Another structural break test specific for correlation structures has been used, proposed by Galeano and Wied [24].

These papers revolve around the fact that connections resulting in co-movements between equity and commodity markets have increased since the financial crisis, which was nonexistent around the 1990s. There are multiple effects that effect the power of these connections such as financial distress, category (precious metals are the most isolated) and even the amount of hedge fund activity in the market.

Marin and Ribas studied the price-run up before particular economic events (M&A announcements) and found significant abnormal returns [40]. Mitchell, Pulvino and Stafford also studied abnormal returns using different methods (for calculating returns) and found differences both before and after the events [43]. Tang and Xu find that particular price run-ups are caused by unreported insider trading: not all run-ups are caused by insider trading, rumors and strategic targets can influence sophisticated investors to figure out M&A targets [57]. Abnormal returns before the events show the possible predictability of these events while abnormal returns after the events show the possibility of identifying the after-effects of the events on the underlying time series. We will attempt to do both using deep learning methods.

Some researchers used machine learning to predict M&A activity as well: Routledge, Sacchetto and Smith used logistic regression and firm's financial statements [50]. Papers found alignments with some words inside the statements have predictive power for M&A activities. We will focus on prices/returns and different financial indicators such as low/high-opening/closing prices and volume information to see if structural breaks or particular economic events have particular run-up effects.

Besides from a focus on economic events, it is relatively easy to find research that focuses on financial price prediction using neural networks. Philip Widegren investigated the accuracies of feed-forward and recurrent networks for asset and future price movements forecasting and concluded that recurrent networks do not outperform feed-forward networks significantly: it is not necessary to create more complex networks if the features are simple [61]. For more complex inputs however, accuracy increased with model complexity. We will focus on feed-forward networks since this work focuses on predictability rather than accuracy. Hansson, Magnus compared recursive traditional models with recurrent networks (long term short memory networks) and concluded that even though outputs are similar, networks outper-

form with regards to change of direction classification and vastly outperform when a trading strategy is implemented [28]. Gustaf Tegnér found out that even though networks outperform the statistical models, instability of these models raises questions about the applicability [58].

Chapter 2

Time Series concepts

The term time series simply implies series of data ordered in time, typically with regular intervals. Time series analysis are used in variety of scopes, from index tracking to sales forecasting and noted as one of the most effective ways of making predictions [21]. It is common to look at time series graphs (Figure 2.1) to track the performance of an investment or to evaluate investment decisions. We will focus on financial context, particularly in time series analysis of security prices.

Time series data can give important insight about an equity. An investor might be interested in various attributes of the time series such as seasonality and trend. It would be common sense to think that market price of an company would be sensitive to commodity prices if it uses a lot of raw material. An manager might be interested in making reliable forecasts to take correct decisions. This shows the importance of tools used to analyse time series.

Figure 2.1: Time Series of Swedbank Stock Price

2.1 Some Time Series properties

Physics has been a big inspiration for economists when doing financial modelling, particularly for its success at predicting future behavior of material objects [55]. These aspirations are easy to come across: Wiener processes has connections to Brownian motion, the random motion of molecules in liquid. Jump diffusion models uses diffusion processes to simulate after-effects of sudden movements in stock prices [46]. Unlike laws of nature, financial systems change based on the models markets use. Such ever changing dynamics require newer models. Delicacy of financial modelling and its enormous effect on the society has pushed well-known financial engineers to warn professionals into better understanding of models and its underlying assumptions when doing time series analysis [20]. Knowing assumptions is critical to understand the risks, when taking decisions with these models. Some of the important concepts relevant to our topic are homoscedasticity, heteroscedasticity, serial dependence and stationarity.

2.1.1 Homoscedasticity

Homoscedasticity (from Ancient Greek homo "same" and skedasis "dispersion") of a financial time series imply that variance of the data is constant over time. Computational and mathematical processes are simpler to analyze with homoscedastic time series (or under the assumption of homoscedasticity). In Figure 2.2 is an example of a Homoscedastic process, where prices (y axis) increase over time (x axis).
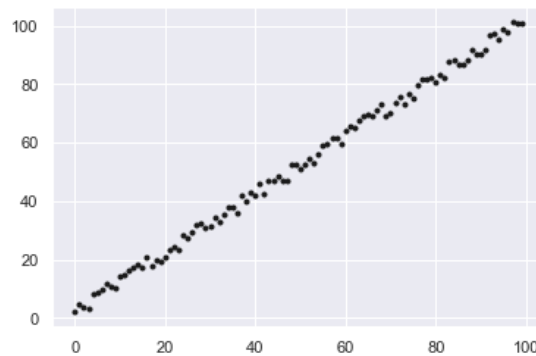
Figure 2.2: Homoscedasticity

2.1.2 Heteroscedasticity

Heteroscedasticity is the absence of Homoscedasticity (Figure 2.3). Volatility of heteroscedastic data changes over time which causes problems in simple regression analysis methods:

15

If the variance increases with price (y axis) then a simple linear regression will fit more to higher prices since distances are higher. This raises concerns using models that assume homoscedasticity since errors will be correlated with time (x axis) and/or the fitted models might be open to errors.
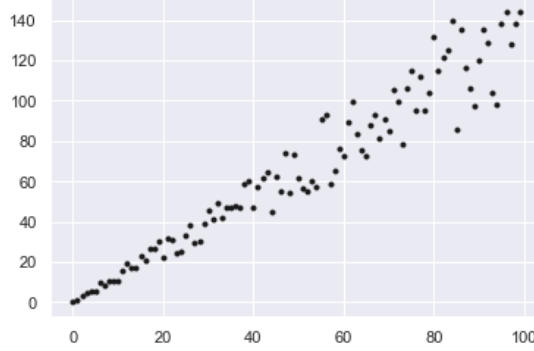
Figure 2.3: Heteroscedasticity

Variance of a financial time series is also called volatility. Variance of random variable X is its dispersion from the mean μ, calculated by:

$$\text{Var}(X) = \text{E}\left[(X - \mu)^2\right].$$

Volatility is an important topic in risk management. Overall volatility has decreased after the 2008 crisis [18]. Volatility is a strong predictor of financial crises: prolonged periods of low volatility leads to higher financial sector leverage [16]. Figure 2.4 shows the rolling standard deviation of Swedbank price with a 50 day window, indicating the time-dependent behaviour of variance in sample. However, it is hard to conclude heteroscedasticity of the population since it is a property of whole population variance.

2.1.3 Serial dependence

When there is a linear dependence between different time steps of the time series, measurement of this dependence helps us explain the structure and perform prediction based on previous time steps. Such statistical dependence is calculated through the autocorrelation function.

Autocorrelation function states how different steps in the time series are related to each other. It is a simple correlation of the series with the lagged series. The definition may vary over different fields. Our definition of autocovariance is as follows [2]:

$$\gamma_k = \text{Cov}\left(x_t, x_{t+k}\right) = E\left[(x_t - \mu)(x_{t+k} - \mu)\right]. \tag{2.1}$$

16
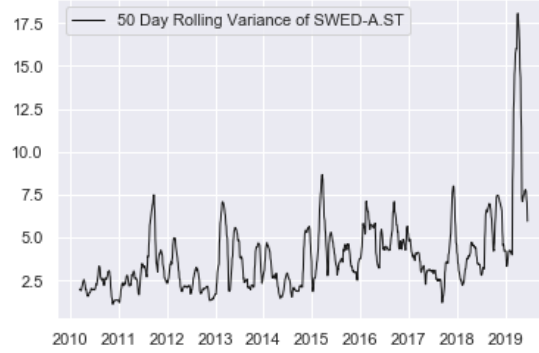
Figure 2.4: 50 days rolling volatility of Swedbank prices

Here μ is the mean of the time series and γ_k is the autocovariance at lag k. For easier analysis, we need to take into account the variance of time series. The definition is parallel to the logic between covariance and Pearson correlation coefficient:

$$\rho_\ell = \frac{\text{Cov}\,(r_t, r_{t-\ell})}{\sqrt{\text{Var}\,(r_t)\,\text{Var}\,(r_{t-\ell})}} = \frac{\text{Cov}\,(r_t, r_{t-\ell})}{\text{Var}\,(r_t)} = \frac{\gamma_\ell}{\gamma_0}$$

ρ_k is called theoretical Autocorrelation Function (ACF). In financial time series, ACF (autocorrelation function) in different time lags are plotted to understand the characteristics of the series and help determining which model to use. Since Sin wave is periodic, its autocorrelation shows resemblance to the wave itself (Figure 2.5). Blue shaded is the % 95 confidence interval indicating the high autocorrelations error outside the area are statistically significant.
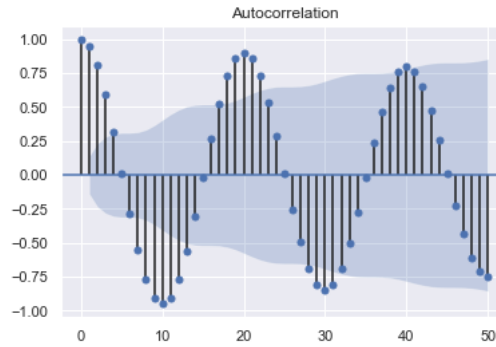
Figure 2.5: Autocorrelation of a sin wave

2.1.4 Stationarity

A time series is called stationary if its unconditional joint probability distribution does not change when shifted in time. Data drawn from a stationary model is easier to analyze and often non-stationary data are transformed to be stationary before analysis. When considering long time intervals, considering complete stationarity can be very restrictive, hard to verify empirically and often weak form is considered [59]. Weak stationarity is when mean and auto-correlation of the data is constant over time. Common causes of violation can be trend or seasonality which can be eliminated before starting the analysis. Common test used for stationarity is the Augmented Dickey-Fuller test, with null hypothesis of unit root existent and alternative hypothesis of stationarity.


```
Slice  0 S:     0 E: 2378  DF:[TestStat&CritVal: -1.68, -3.43,P-val:0.44:Stat= FF ]


Slice  1 S:     0 E:  475  DF:[TestStat&CritVal: -1.67, -3.44,P-val:0.45:Stat= FF ]
Slice  2 S:   475 E:  950  DF:[TestStat&CritVal: -1.11, -3.44,P-val:0.71:Stat= FF ]
Slice  3 S:   950 E: 1425  DF:[TestStat&CritVal: -2.04, -3.44,P-val:0.27:Stat= FF ]
Slice  4 S:  1425 E: 1900  DF:[TestStat&CritVal: -0.32, -3.44,P-val:0.92:Stat= FF ]
Slice  5 S:  1900 E: 2375  DF:[TestStat&CritVal: -1.55, -3.44,P-val:0.51:Stat= FF ]
```

Figure 2.6: Stationarity test on Swedbank stock prices

Daily returns of financial stocks are often assumed to be weakly stationary. As an example we perform stationarity tests on Swedbank prices. First line in the test output is the test for whole series. Stock price is split into 5 equal parts to test stationarity in different date ranges. Smaller p-values indicate smaller probability of an unit-root. All slices (Figure 2.6) in the price have high p-values, thereby we fail to reject the null hypothesis that the prices have a unit root.

Common and direct approach in financial time series analysis is to analyze returns or log returns of the time series. Let P_t be the price of an asset at time t. Our definitions are as follows.

The one-period simple return R_t is

$$R_t = P_t/P_{t-1} - 1,$$

the k-period simple return $R_t(k)$ is

$$R_t(k) = P_t/P_{t-k} - 1.$$

The one-period log return is

$$r_t = \ln(1 + R_t) = \ln(P_t) - \ln(P_{t-1})$$

and the k-period log return is

$$r_t(k) = \ln(1 + R_t(k)) = \ln(P_t) - \ln(P_{t-k}).$$


```
Slice  0 S:     0 E: 2377  DF:[TestStat&CritVal: -9.29,  -3.43,P-val:0.00:Stat=TRUE]

Slice  1 S:     0 E:  475  DF:[TestStat&CritVal:-13.42,  -3.44,P-val:0.00:Stat=TRUE]
Slice  2 S:   475 E:  950  DF:[TestStat&CritVal:-11.84,  -3.44,P-val:0.00:Stat=TRUE]
Slice  3 S:   950 E: 1425  DF:[TestStat&CritVal:-22.78,  -3.44,P-val:0.00:Stat=TRUE]
Slice  4 S: 1425 E: 1900  DF:[TestStat&CritVal: -6.55,  -3.44,P-val:0.00:Stat=TRUE]
Slice  5 S: 1900 E: 2375  DF:[TestStat&CritVal:-15.82,  -3.44,P-val:0.00:Stat=TRUE]
```

Figure 2.7: Stationarity test on Swedbank Returns

We construct one-period returns and log-returns of the same time series and perform the stationarity test as before (Figure 2.7 and 2.8). These two time series now pass the stationarity test for all slices, indicating that these series are, since their properties are more stable, easier to analyze, than the price series itself (Figure 2.6). All tests on returns reject the existence of a unit root. This shows overall stabilizing effect of performing on returns rather than prices

19

```
Slice  0 S:     0 E: 2377  DF:[TestStat&CritVal:-25.26, -3.43,P-val:0.00:Stat=TRUE]


Slice  1 S:     0 E:  475  DF:[TestStat&CritVal:-11.53, -3.44,P-val:0.00:Stat=TRUE]
Slice  2 S:   475 E:  950  DF:[TestStat&CritVal:-12.15, -3.44,P-val:0.00:Stat=TRUE]
Slice  3 S:   950 E: 1425  DF:[TestStat&CritVal:-14.56, -3.44,P-val:0.00:Stat=TRUE]
Slice  4 S:  1425 E: 1900  DF:[TestStat&CritVal: -6.67, -3.44,P-val:0.00:Stat=TRUE]
Slice  5 S:  1900 E: 2375  DF:[TestStat&CritVal:-16.00, -3.44,P-val:0.00:Stat=TRUE]
```

Figure 2.8: Stationarity test on Swedbank log returns

itself.

Calculating returns in time series analysis is also referred as differencing in statistics, and resulting time series of this transformation are referred as integrated values. Differencing and integrated values are used in ARIMA models.

2.2 Time Series models

The market model is used in events study and takes into consideration the effect of external factors. The ARIMA models studied in the academia, used in variety of financial tasks including algorithmic trading & risk management. We will use market model to calculate abnormal returns, and ARIMA to generate Monte Carlo simulations. To explain them, we will start with autoregressive (AR) and moving average (MA) models.

2.2.1 Autoregressive (AR) model

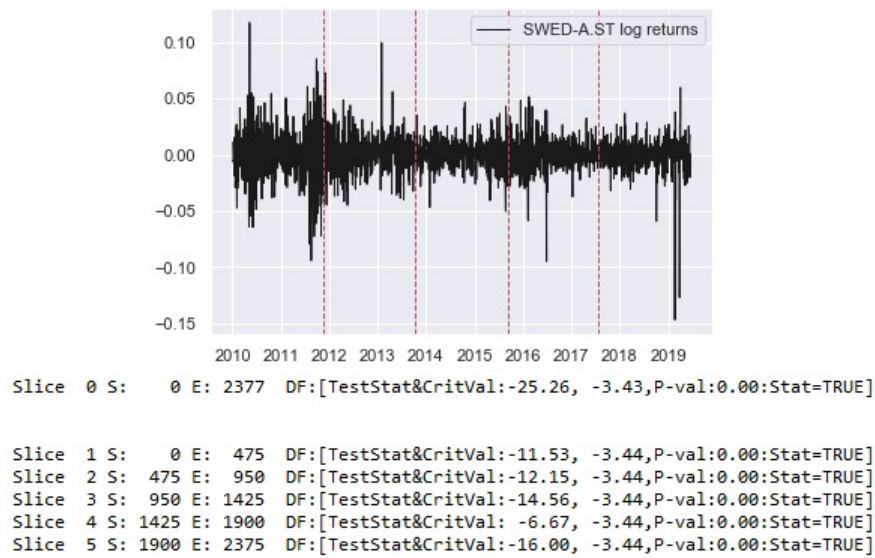Autoregressive model assumes that future values are a linear combination of previous values. The model AR(p) is defined as follows:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

.

Order of the model p simply is the amount of previous data taken into account for predictions. c is a constant, φ_i is the parameters (multipliers of previous time steps), ε_t is random error, X_{t-i} are the previous values. Note that an occurrence of a big error term will forever effect the expected value of the time series.

Figure 2.9: Generated AR(1) and AR(2) processes

In Figure 2.9 is the simulated AR(1) and AR(2) processes with φ_1 bigger than 1 are shown. Parameter above 1 causes these series to have positive trends and thus are not stationary. Since these generated AR(2) process depends on previous two values, its ACF plot (similar to sin wave's ACF plot in Figure 2.5) in Figure 2.10 shows significant autocorrelation for steps 2 and 3.
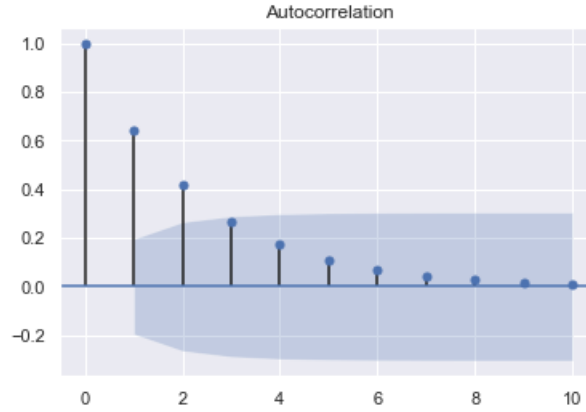
Figure 2.10: ACF plot test of AR(2) Process

2.2.2 Moving Average Model

Moving average model takes into account the past error terms. The definition is:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where q order is the order of the model, θ_i are the parameters, μ is a constant and ε_t are error terms.

2.2.3 ARMA

Autoregressive-Moving average (ARMA) models use AR models to correlate the time series to its lagged values, while using MA model to model error terms of next step as a linear combination of its lagged error values. We simulate an ARMA process in Figure 2.11 as an example. Usually referred to as ARMA(p,q) where p denotes the order of AR part while q denotes the order of MA part:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

2.2.4 ARIMA

ARIMA models are used when data shows non-stationarity. Data is "integrated" until the data becomes viable for the use of ARMA model. ARIMA model of (p,d,q) is an ARMA model

22

Figure 2.11: Simulated ARMA(1,2) Process

integrated (d) times. In financial time series integration refers to the modelling of returns instead of the stock prices itself. Integration of first order can be shown as:

$$y'_t = y_t - y_{t-1}$$

Identification of parameters can be done by the Box-Jenkins method, which has three steps:
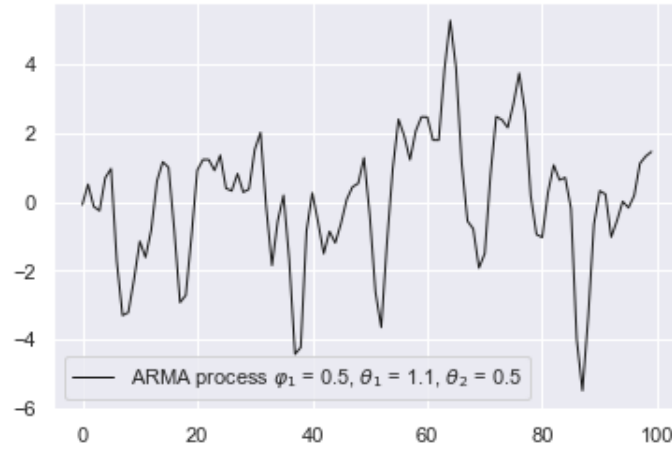(1) Identifying underlying stationary series by differencing and checking stationarity. Once a stationary series is found, ACF plots are used to find the number of parameters.
(2) Estimating parameters, which can be done by maximum likelihood estimation or non-linear least-squares estimation.
(3) Checking the model to see if it is a good fit. Model checking is often done by checking stationarity, autocorrelations and normality of the residuals.

The Box Jenkins method requires manual interaction for modelling and predicting future values. We use pmd-ARIMA package in Python to identify ARIMA models, similar to R's auto-arima function. This allows us to fit many time series in quick succession, without worrying if the model is a perfect fit, which is rarely the scenario. The script tests the stationarity values of integrated values to pick integration level, then fits different ARIMA models and picks the best one based on AIC[1] values. Output of the pmd-ARIMA function can be seen in Figure 2.12. Found model is an ARMA(1,2) model with φ_1=0.47, θ_1 =1.10 and θ_2=0.51 (small deviation of parameters from the generated series is caused by the random term). Similar to the Box-Jenkins method, similar tests can be run for considering how good this model fits the data. In Figure 2.13 it can be seen how the fit was for this particular ARMA model.

---

[1] Defined later in Eq.2.7

23

The residuals show stationarity, normality and no autocorrelation which indicate a good fit, as expected since it was a generated ARMA process.

```
Fit ARIMA: order=(2, 0, 2) seasonal_order=(0, 0, 0, 1); AIC=28511.170, BIC=28554.432, Fit time=2.317 seconds
Fit ARIMA: order=(0, 0, 0) seasonal_order=(0, 0, 0, 1); AIC=45757.083, BIC=45771.504, Fit time=0.114 seconds
Fit ARIMA: order=(1, 0, 0) seasonal_order=(0, 0, 0, 1); AIC=34601.211, BIC=34622.842, Fit time=0.280 seconds
Fit ARIMA: order=(0, 0, 1) seasonal_order=(0, 0, 0, 1); AIC=nan, BIC=nan, Fit time=nan seconds
Fit ARIMA: order=(1, 0, 2) seasonal_order=(0, 0, 0, 1); AIC=28509.178, BIC=28545.230, Fit time=1.336 seconds
Fit ARIMA: order=(1, 0, 1) seasonal_order=(0, 0, 0, 1); AIC=29971.324, BIC=30000.166, Fit time=1.281 seconds
Fit ARIMA: order=(1, 0, 3) seasonal_order=(0, 0, 0, 1); AIC=28511.171, BIC=28554.433, Fit time=1.815 seconds
Fit ARIMA: order=(2, 0, 3) seasonal_order=(0, 0, 0, 1); AIC=28513.178, BIC=28563.651, Fit time=4.537 seconds
Fit ARIMA: order=(0, 0, 2) seasonal_order=(0, 0, 0, 1); AIC=nan, BIC=nan, Fit time=nan seconds
```

Figure 2.12: Model identification using pmd-ARIMA on the Simulated ARMA Process

In Figures 2.6 and 2.7 it can be seen that the time series of Swedbank returns (returns are integrated price series) are stationarity, but not the prices. According to Box-Jenkins method, appropriate model should be on the integrated series, which indicates parameter d to be 1. Model fit on Swedbank prices is an ARIMA(1,1,2) model that gives φ_1=0.84, θ_1=-0.78, θ_2=-0.10.
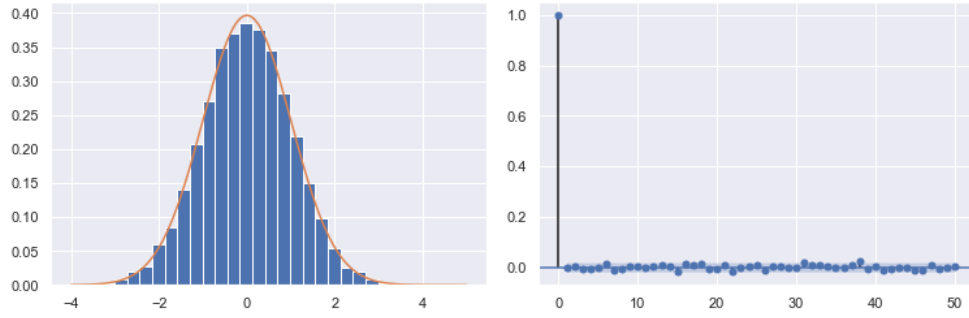
Figure 2.13: Normality and ACF plots of residuals of the simulated ARMA process

After estimating the parameters, we check the model fit in the Swedbank returns in Figure 2.14. To illustrate an ideally fit model, we perform a similar procedure in a simulated ARMA model in Figure 2.13. Residual checking indicates stationarity of residuals however there are significant autocorrelations in multiple steps (25-26) of Swedbank model residuals, indicating a bad fit. In the normality plot, Swedbank has some extreme values (around -30 in x axis), which are caused by outliers in returns. We state that these can be the effects of structural breaks in such model, and elaborate in later sections.

Even though frequent occurrence of unexpected events, that violate normality assumptions, are well-known [19], many institutions use risk management practices (e.g. VAR) that assume normally distributed returns. Peso crisis in 1994 is a good example: according to normal distribution assumptions, it was a 44-standard-deviation event, that would happen once in many billions of years [54]. Fat tails not only create misleading results in risk management

practices but also cause misvaluation in pricing models as well (e.g. option pricing) [30]. Fat-tailed distributions that are empirically relevant are used often to overcome such challenges when simulating financial time series. Stable distributions, jump diffusion, and stochastic volatility are some of the most widely used approaches, however, implementing them is harder than the normal distribution [54].
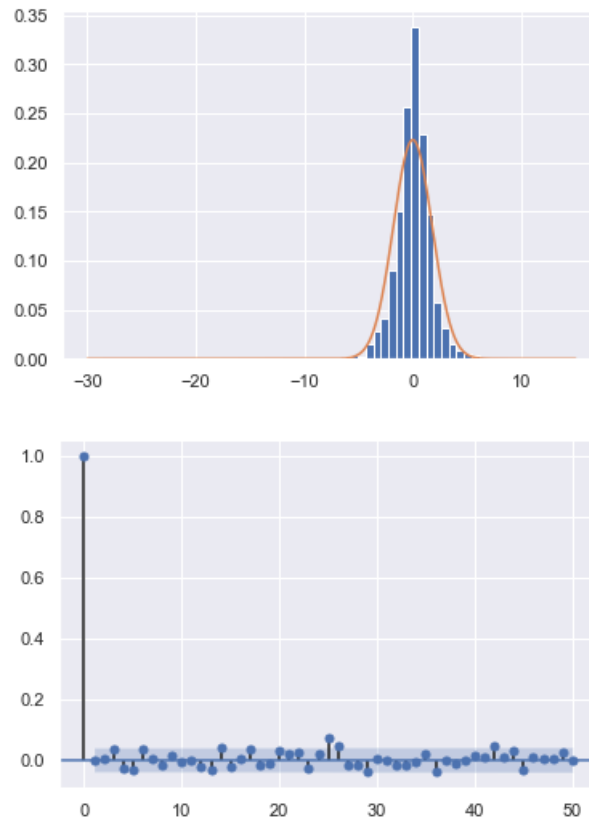

Figure 2.14: Normality and ACF plots of ARIMA residuals of Swedbank stock prices

2.2.5 Market Model

Market model is a simple regression model, assumes that returns of a firm are effected by various sources. To represent these sources, a time series that represents overall markets are used, such as OMX30[2], S&P500 or Dow Jones Industrial Average(DJIA)[3]. In Figure 2.15 common downtrend after 2008 can be seen in time series of Swedbank and OMX30, illustrating the logic behind the market model. Market Model's predictive power is higher compared to post-crisis era, signaling reduced stock-picking benefits from a risk perspective due to share of reduced idiosyncratic risk[4] [18].
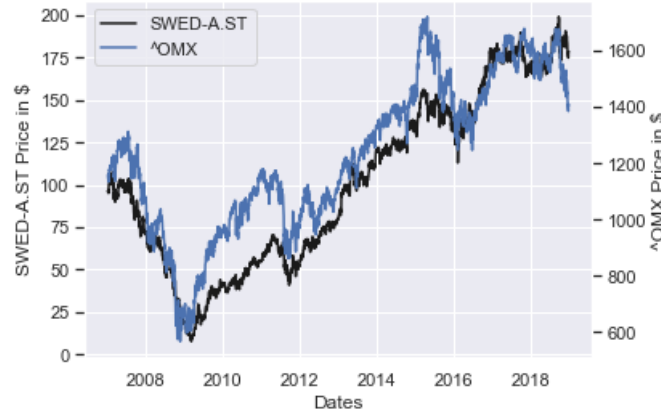

Figure 2.15: Swedbank and Index prices

The model is similar to capital asset pricing model (CAPM)[10] or single index model and even though these terms are used interchangeably, market model does not use risk free rates, making its use simpler and more suitable for the scope of our work. The market model is as follows:

$$R_t = \alpha + \beta * R_{mt} + e_t \tag{2.2}$$

R_{mt} is the market return on time t, e_t is the error on time t (assumed to be normally distributed) and R_t is the expected return of the stock on time t. Individual firms sensitivity to the market are measured by β and firm specific returns are represented by α. Its coefficients are usually measured by a linear least-squares regression between the firm returns and market returns. Its estimations are a sensitive subject since it has variety of applications such as calculating portfolio weights, economic returns, net present value, cost of equity and cost of capital. Unless noted otherwise, ordinary least squares (OLS) is used to estimate expected returns of the market model used in this thesis is as follows:

---

[2]OMX Stockholm 30, represents Swedish Stock Market

[3]We use S&P500 for US stocks, since DJIA has quantitatively important flaws [53].

[4]Firm-specific risk

26

$$\min \sum_{t=1}^{T} (R_t - (\alpha + \beta * R_{mt}))^2. \tag{2.3}$$

There have been many articles that focused on β estimation techniques. Brown and Warner (1985) looked at possible reasons behind problems by studying characteristics of daily stock prices when studying events (such as autocorrelation, departure from normality, excess returns and OLS biases) of using daily prices [11]. To overcome such issues, Mergner and Bulla uses various approaches, including GARCH/ARCH applications and Markov chains and concludes that Kalman filter approach is the preferred model [42].

The estimation of coefficients also requires consideration of an appropriate time horizon. Hollstein et al. suggested periods longer than 12 months should be used for lower prediction errors [29]. Estimations with different rolling windows on Swedbank price (index: OMX30) suggest that mean absolute error for different time horizons are negligible (Figure 2.16).

```
  5 Days Rolling Beta, MAE: 0.00812 Avg α: -0.0001, Avg β: 0.6228
 15 Days Rolling Beta, MAE: 0.00764 Avg α: -0.0002, Avg β: 0.5899
 30 Days Rolling Beta, MAE: 0.00755 Avg α: -0.0002, Avg β: 0.5750
 45 Days Rolling Beta, MAE: 0.00754 Avg α: -0.0002, Avg β: 0.5672
 60 Days Rolling Beta, MAE: 0.00755 Avg α: -0.0002, Avg β: 0.5654
 90 Days Rolling Beta, MAE: 0.00756 Avg α: -0.0002, Avg β: 0.5681
180 Days Rolling Beta, MAE: 0.00756 Avg α: -0.0002, Avg β: 0.5755
360 Days Rolling Beta, MAE: 0.00757 Avg α: -0.0002, Avg β: 0.5780
720 Days Rolling Beta, MAE: 0.00790 Avg α: -0.0002, Avg β: 0.5608
```

Figure 2.16: Rolling windows of high ranges fit better to overall prices

Dadakas et al. (2016) suggest a range less than 90 days due to structural breaks that create misrepresentation of systematic risk[5][15]. We test this by plotting estimated β's. Around structural breaks, low window β's changed way faster compared to window of 360 trading days, indicating a possibility of better fit of model around structural breaks (see Figure 2.17, where blue and red shades denote positive and negative outliers, respectively).

To test how different β estimations give better predictions around outliers, we take a window of (-60,5) days around the outliers and calculate mean absolute error (MAE) for each estimation. 30 days model estimation has the lowest MAE. Even though long β estimations result in overall better price prediction performance, lower windows give better results around outliers (see Figure 2.18). We allocate after outlier performance to adaptation of the models, since a regression with lower price inputs will give higher weights to the price during the event day, incorporating new information faster. Performance differences before the outlier could be because of insider trading activities or complicated investors that predict economic events before the general public. Such trades before the event days will influence the price and again, low time horizons will capture them better than high time horizon estimations.

Market model is an important tool in event study. Event study is a statistical method that focuses on measuring the effects of a particular event on stock price. Methods were invented by

---

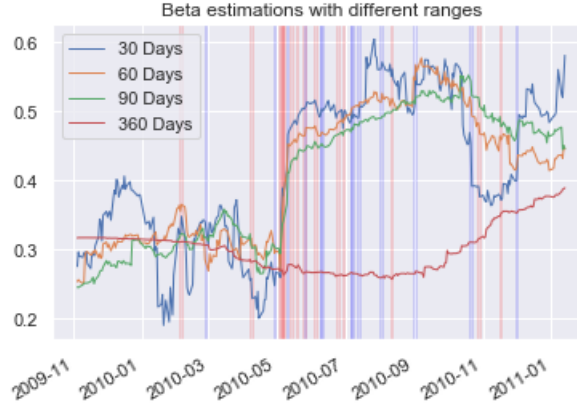[5]Market risk, ie. risk cannot be diversified

Figure 2.17: Estimations of β with different rolling windows

```
Est. window    5 days, before outlier MAE: 0.0130, after outlier MAE: 0.0148
Est. window   15 days, before outlier MAE: 0.0129, after outlier MAE: 0.0148
Est. window   30 days, before outlier MAE: 0.0129, after outlier MAE: 0.0147
Est. window   60 days, before outlier MAE: 0.0131, after outlier MAE: 0.0150
Est. window   90 days, before outlier MAE: 0.0132, after outlier MAE: 0.0152
Est. window  180 days, before outlier MAE: 0.0134, after outlier MAE: 0.0154
Est. window  360 days, before outlier MAE: 0.0135, after outlier MAE: 0.0157
Est. window  720 days, before outlier MAE: 0.0145, after outlier MAE: 0.0167
```

Figure 2.18: Rolling windows of low ranges fit better around outliers

Fama, Fisher, Jensen, and Roll in 1969 [23] where they analyzed stock splits and concluded that market is efficient when it comes to incorporating this particular information. General methodology is explained and exampled in multiple articles[36], [7], [41]. Methodology starts by identifying the event and calculating "abnormal returns" around the events. In research, abnormal returns are unusual returns defined as the difference of expected and occurred returns. A simple definition, where AR_t, Y_t, $\hat{Y}_t$ stands for abnormal, actual and predicted returns in time t respectively is given as follows:

$$AR_t = Y_t - \hat{Y}_t. \tag{2.4}$$

Even though multi-factor models [6] are more suitable for prediction [26], the market model and the mean return model (ie. constant return model) are popular to calculate abnormal returns (AR_t). Abnormal returns are usually defined with the market model, or the CAPM. There are different definitions of abnormal return available that include α in the abnormal returns when using CAPM or the market model[7]. We choose to stick to the definition used in research, that does not include α. Abnormal returns using market model are calculated as follows:

---

[6] Market model is also referred as the single factor model

[7] https://investinganswers.com/, https://www.investopedia.com/, https://www.nasdaq.com/

$$AR_t = R_t - \alpha - \beta_t{}^8 R_{mt}$$

According to the methodology, abnormal returns, or its various levels (eg. cumulative abnormal returns, average abnormal returns) are tested with various tests (eg. t-test) to check statistical significance of "altered" prices. This methodology is regarded as a good tool when evaluating corporate finance decisions from the perspective of a retail investor [13].

Using this methodology, Marin Ribas [40] analyzed mergers and acquisition (M&A) activities and found significant abnormal returns. Sevindik and Gokgoz [52] analyzed M&A activities in emerging markets and found that mean cumulative abnormal returns (CAR)'s have changed after the crisis (some countries had positive CAR's pre-crisis, and after it became negative and vice-versa). Bagliano, Favero and Nicodano studied abnormal returns in Italy and concluded that insider traders start trading much earlier (measured as days before the public announcement) than their counterparts in US, when and assessed if such methods can be used by supervisory authorities [7].

Price movements after merger announcements are an important indicator of investor sentiment, and often signals the success of the merger. Mergers can add value to a company, thus increase its stock price in a lot of ways: economics of scale, efficient use of resources with new management, accessing intellectual property, entering new markets, inefficient asset liquidation, operation and/or know-how synergies, tax cuts, larger product portfolios. Mergers can fail to generate value, and even destroy value: underestimated integration costs, overpaying the target companies, overestimated synergies, hostile takeovers, culture assimilation, due diligence failures are some risks that can cause a loss in value in companies. Research mentioned before shows that before and after some economic events, in this case M&A's, there are significant unusual price movements. One of the important conclusions of these papers is that due to fundamental economical differences between different markets, mergers can create or destroy value in average, but also length of abnormal returns before the announcement dates are different (insider trading behaviour is different).

---

[8]Note that this beta is time dependent, as we take rolling windows to estimate it

2.2.6 Kalman Filter

Kalman filter is an recursive algorithm that is widely used in statistics and engineering since Rudolf E. Kalman published his paper in 1960 [33]. Kalman filter first estimates current state variables, then based on previous errors, gives higher weights to estimates with higher probabilities. Points with high errors, such as outliers in prices, take lower weight when estimating the parameters of the model. An important point of Kalman filter is that it takes the parameters to be stochastic variables dependent on time, instead of constants like in OLS. The algorithm used in this paper is implemented for the following set of equations[9]:

$$x_{t+1} = A_t x_t + b_t + \text{Normal}(0, Q_t).$$
$$z_t = C_t x_t + d_t + \text{Normal}(0, R_t).$$
$$(2.5)$$

The first equation is called the state equation and the second, observation equation, and variables have pre-defined names:

x_t is the state vector, A_t is the transition matrix, b_t is the state offset, Q_t is state transition matrix, z_t is the observation vector, C_t is the observation matrix, d_t is the observation offset, R_t is the observation covariance matrix. After the set of matrices and vectors are defined, expectation maximization (EM) algorithms are used for estimating the parameters used in Equation (2.5). Objective function of the EM algorithm is:

$$\max P(z_{0:T-1}; \theta)$$

where θ is the set of parameters to be estimated by the EM. If we define

$$L(x_{0:T-1}, \theta) = \log P(z_{0:T-1}, x_{0:T-1}; \theta)$$

then the EM algorithm works an iterative process, first by finding states and estimating parameters for each step:

$$P(x_{0:T-1} | z_{0:T-1}, \theta_i).$$

Now that we have the parameters estimated for previous steps, it is time for predicting future steps:

$$\theta_{i+1} = \arg\max_{\theta} E_{x_{0.T-1}}[L(x_{0:T-1}, \theta) | z_{0:T-1}, \theta_i].$$

The algorithm requires initial values for states. Since our sample spaces are quite big, convergence after the initial values was satisfactory.

---

[9]https://pykalman.github.io/class_docs.html

There are various studies that use Kalman filter: [48] used Kalman filter to calculate hedge fund's exposure and found the algorithm superior to rolling linear regression. [22] uses Kalman filter in pairs trading, to calculate the spread. [62] used Kalman filtering in corporate finance: to predict financial distress. We use the algorithm for β and α estimations. Following the market model in Equation (2.2), the observation equation is:

$$R_t = \alpha_t + \beta_t(R_{mt}) + \varepsilon, \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

and the state equations are[10]:

$$\alpha_{t+1} = \alpha_t + \eta_t, \eta_t \sim N\left(0, \sigma_\eta^2\right),$$
$$\beta_{t+1} = \beta_t + \varepsilon_t, \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right).$$

To align with the Equation (2.5), reformulated state equations are:

$$\left[\begin{array}{c} \alpha_{t+1} \\ \beta_{t+1} \end{array}\right] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]\left[\begin{array}{c} \alpha_t \\ \beta_t \end{array}\right] + \left[\begin{array}{c} \eta_t \\ \varepsilon_t \end{array}\right]$$

Reformulated observation equation becomes:

$$R_t = \left[\begin{array}{cc} 1 & R_{mt} \end{array}\right]\left[\begin{array}{c} \alpha_t \\ \beta_t \end{array}\right] + [\varepsilon_t].$$

We will use Kalman filter as an alternative to rolling OLS estimation when calculating abnormal returns, as a more precise alternative to OLS for training the neural networks.
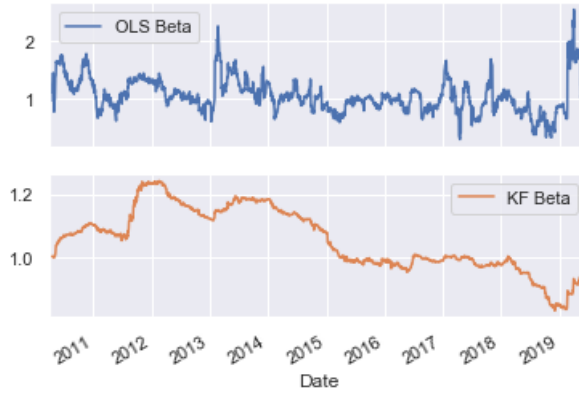

Figure 2.19: Beta estimations

[10]σ_η and σ_ε are the variance of random terms of corresponding state equations.

31

Rolling beta estimations using a window of 60, and the Kalman filter estimations for Swedbank and OMX30 as index, are plotted in Figure 2.19, analysis shows that MAE of OLS is 0.0072 and MAE of KF is 0.0054, which indicates that Kalman filter is a preferable approach giving a better fit.

2.3   Goodness of fit

On estimating the predictive power of the models, we often look how predictions are different from the observations. For this, variety of quantitative methods are used. Some parameters in the aforementioned models can be decided with rather more qualitative methods that require careful manual investigation, and can result in human error. When facing large number of samples, quantitative methods are necessary. Some error measures and methods that are used in estimation are briefly explained in the subsections that follow.

2.3.1   Mean Squared Error (MSE)

Mean squared error is the average of squared differences between predictions and observations.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{2.6}$$

where n is the number of availabe data points and $(Y_i - \hat{Y}_i)$ are the errors.

2.3.2   Mean Percentage Error (MPE)

Mean Percentage error is similar to MSE however scaled by actual terms to provide the practitioner a more unbiased measure. In case of financial time series, often the trend is positive which would create differences in MSE even for similar MPE errors.

$$\text{MPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{Y_i - \hat{Y}_i}{Y_i}.$$

Taking percentages is particularly useful to scale the errors according to magnitude of underlying observations, or also comparing analysis made on different datasets with different ranges. If there is a consistent bias in the predictions, MPE can be used since it does not take absolute or squares of the errors.

2.3.3 Mean Absolute Error (MAE)

Mean absolute error is quite parallel to the MPE, where taking absolutes enables us to measure average distance from observations. It is a common measure of forecast error in time series analysis [31]. Since square is not taken, each error has same weight contribution to the summation. Formula is as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|.$$

2.3.4 Akaike information criterion

Named after Hirotugu Akaike, Akaike information criterion (AIC) is used for model selection. The AIC value is defined as follows [3]:

$$\text{AIC} = 2k - 2\ln(\hat{L}), \tag{2.7}$$

where k is the number of parameters and $\hat{L}$ is the maximum value of the likelihood function for the model. Penalizing the high number of parameters in the model, AIC also deals with over-fitting of the models. Model with the lowest AIC value is preferred.

2.3.5 Bayesian information criterion

Bayesian information criterion (BIC) is similar to AIC and also penalizes the overfitting.

$$\text{BIC} = \ln(n)k - 2\ln(\hat{L}),$$

where k is and $\hat{L}$ are defined in the Section 2.3.4. Model with the lowest BIC value is preferred.

2.4 Deep Learning Concepts

Artificial neural networks (ANN) are computational machines that can perform complicated tasks. Roots of ANN's go back to 40's Hebbian learning. They were originally designed to solve problems in a similar way to a human brain: architecture of the networks is similar to the synaptic connections between the neurons. Operating these networks requires significant processing power, especially for a large artificial network. Thanks to recent developments in technology, neural networks became quite popular. Deep learning methods are used intensively for various applications from image and speech recognition to medical diagnosis.

Networks consist of multiple "layers" which consist of groups of nodes. Nodes are connected to each other with corresponding connections, called the "edges". First layer is called the input layer, last layer is called the output layer and the middle layers are called the hidden layers. Networks with many hidden layers are called deep neural networks. The information is sent to the input layer, usually traverses multiple times through the hidden layers until the information reaches the output layer. In every transfer, the signals are manipulated through edges that can perform computational tasks. The class of neural networks used in this paper is the feed-forward networks. In a simple feed forward neural network, the nodes are calculated through simple summations and multiplications. Value of every node is a product of its connections and corresponding weights. Figure 2.20 illustrates a network with a single hidden layer, 3 inputs and 3 outputs.
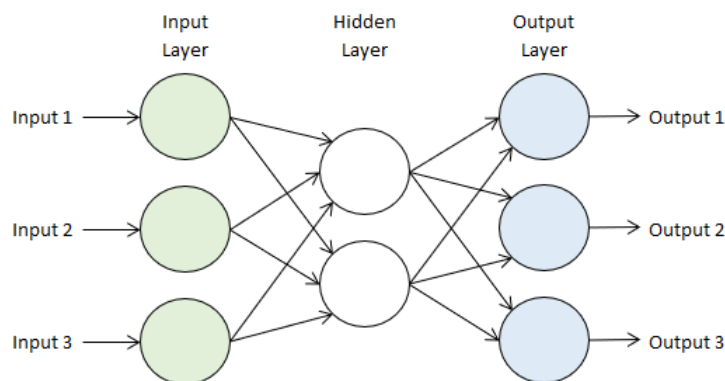
Figure 2.20: Design of a neural network

2.4.1 Feature scaling

Non linearity in inputs and outputs can permit the networks to work properly. For example, values that deviate much from the mean can create biases in the network. Scaling techniques are used to bring the data to a common scale in order to achieve better results by stabilizing the distributions of layer inputs and outputs. We explain min-max and standardization scaling methods briefly, where x' are the new values and x are the inputs.

Min-max normalization brings the data to [0, +1] range by subtracting the minimum from every input and dividing by the inputs range. Formula can be given as: $x' = \frac{x - \min(x)}{\max(x) - \min(x)}$. Since min/max values are the main parameters and financial data often has extreme values, it can be expected that most of the min-max scaled financial data will be in a narrow interval.

Standardization uses standard deviation to divide every values instead of the range, so instead of scaling primarily on extreme values, overall deviation is considered. Formulated by $x' = \frac{x - \bar{x}}{\sigma}$, where σ is the standard deviation of the data, the result has zero-mean and unit variance. This method is used widely in various machine learning algorithms [44].

2.4.2 Activation functions

Due to wast number of nodes in a neural network, activation functions are used to filter and manipulate the active nodes. Models that use these functions are easier to train and often achieve better performance. One of the most popular activation function in research is the rectifier (rectified linear unit, ReLU)[25]. Since this function does not allow negative values, negative inputs might not effect weights at all, as seen as Figure 2.21. It is defined by

$$g(x) = \max(0, x). \tag{2.8}$$

To overcome the problem of ignoring negative values, a similar, widely used function is used. Leaky rectified linear unit (LReLU) [35] is illustrated in Figure 2.22 and defined by:

$$g(x) = \begin{cases} x & \text{if } x > 0 \\ 0.01x & \text{otherwise.} \end{cases} \tag{2.9}$$

There are other activation functions, like parametric ReLU, which is similar to the LReLU but uses a paramater intead of 0.01 multiplier. parametric ReLU is illustrated in Figure 2.23 and defined by:

$$g(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise.} \end{cases} \tag{2.10}$$
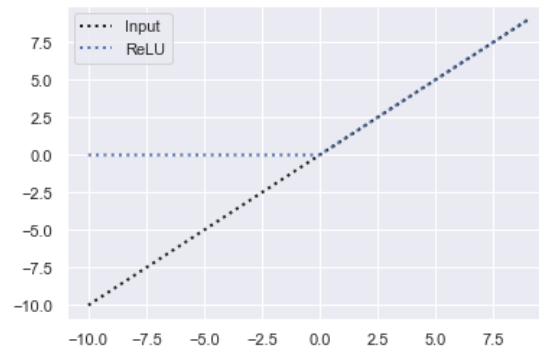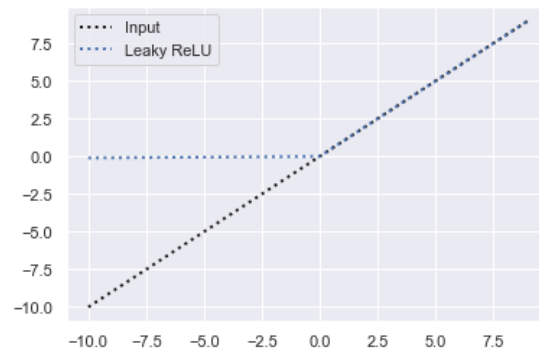
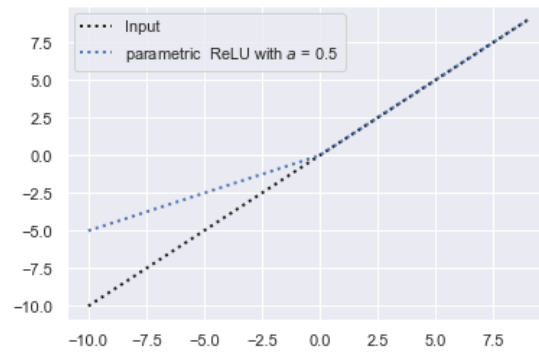Figure 2.21: ReLU

Figure 2.22: LReLU

Figure 2.23: Parametric ReLu

2.4.3 Formulation of Neural Networks

Mathematical representation of the networks is quite straightforward. Value of every node is a sum of its incoming connections:

$$z = \mathbf{w}^T \cdot \mathbf{x} + b, \tag{2.11}$$

where z is the value of the node, w^T, x and b are the weights, incoming signals and the bias, respectively. Outgoing signal, $\hat{y}$ is processed through the selected activation function $f(z)$:
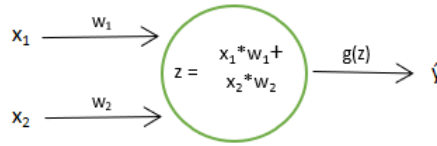
$$\hat{y} = g(z)$$

Figure 2.24: Output calculation of a single node

Inputs are transferred through these calculations from input to output. Weights are determined during the training process to minimize a selected loss function. Loss functions used in this thesis are picked from the goodness of fit measures described before.

2.4.4 Training of Neural Networks

Determining weights in a neural network is often called training of the network. There is a variety of optimization methods, many are versions of gradient descent algorithms. Figure 2.25 shows an example of a gradient descent problem algorithm in a linear regression optimization where MSE in was minimized. We start the algorithm by assigning random coefficients and running the model $y = a + bx$. Parameters then changed in small amounts until desired fitting is achieved Figure 2.26[11].

Size of the steps is called the learning rate. Objective function, in this case MSE is called the loss function and steps (or slopes which are actually simplification of small derivatives) are called gradients. Stochastic gradient descent algorithms choose the observations as inputs to be picked randomly. Inputs chosen in every step of a stochastic gradient descent algorithm are called batches, and there exists several methods to normalize these batches for increased accuracy. Even though these techniques significantly improve the training processes we lack the effective understanding of the reason of its success [51].

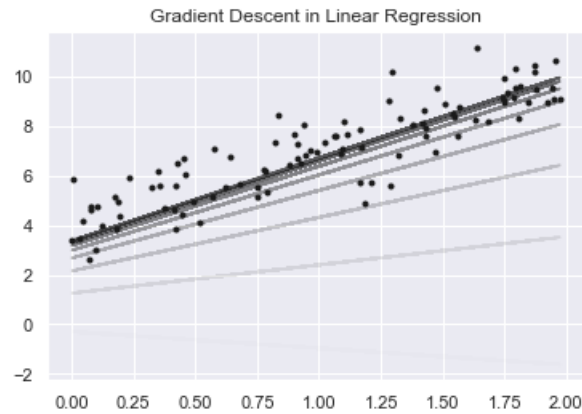[11]x-axis denotes the number of iterations

Figure 2.25: Gradient descent in linear regression
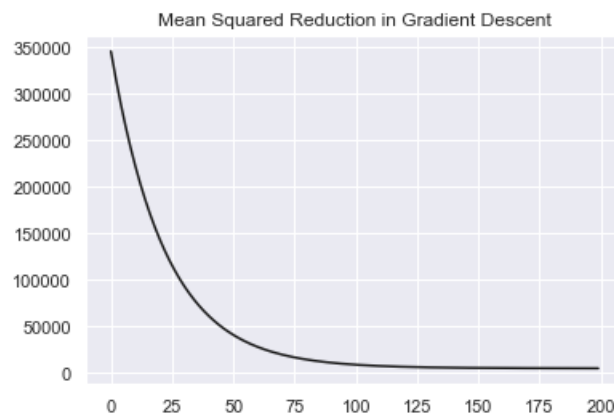
Figure 2.26: Error Reduction in iterations of gradient descent

2.4.5 Backpropagation

Algorithms start by allocating weights randomly and calculating the error of the model. Then back propagation is used to measure effect of every weight on the error, after which the weights are updated to reduce the error function. Gradient descent with backpropagation does not guarantee global minimums but a local minimum since error functions are non-convex.

Chapter 3

Data analysis

There are multiple tests to locate structural breaks. Chow test [27] is one of the frequently used, which tests coefficients of the linear regression on two slices of data. There are other tests as well, having different properties: Andrews and Fair, can be applied to non-linear heterogeneity equations [6] while Bai and Perron tests the null hypothesis of a particular number of structural breaks [8].

Chow test focuses on the changes of parameters of a linear regression, assuming the linear regression variables are stationary over time, therefore causing inference problems when testing non-stationary variables [4]. Similarly, existence of economic events create differences in AR-IMA parameters as well. We plot stationarity tests on prices and returns, then proceed to show how ARIMA parameters change: big changes in price follow big parameter changes.

According to strong efficient markets hypothesis, prices include both public and non-public information and it is not possible to gain superior gains. Aforementioned literature shows that this is not always the case, price run-ups before information is public show that insider traders (or sophisticated investors that were able to anticipate these events) made superior profits. By graphing how the prices would look if it was driven only by the market (with the market model), we see effect of company-specific events and test the existence of statistically significant returns. To come up with an unbiased methodology, we define outliers by their magnitude.

$$magnitude = (R_i - \mu_{sample})/\sigma_{sample}, \tag{3.1}$$

where R_i, μ_{sample}, σ_{sample} stand for return occurred during event i, mean and std dev of sample data respectfully. Since these outliers create big errors[1], they are structural breaks in the prices. Thereby, we assume that every outlier has an underlying economical event. We believe this is an reasonable assumption since it would be unreasonable to think that big outliers can occur without any underlying reason.

---

[1]Range of two sigmas around the mean includes 95% of values in the normal distribution[47]

3.1 Structural breaks and outliers in data

To show effects of economic events on ARIMA predictions, we start by analyzing Swedbank stock prices. Even though Swedbank percentage returns have negative excess kurtosis[2] (Figure 3.1), extreme values occur more than expected by a normal distribution. These outliers are easy to spot on a QQ plot:

Figure 3.1: Distribution and outliers of percentage returns

These outliers can be an important reason of the non-stationarity. In Figure 3.2 red areas indicate the density of outliers. Blue line are the p-values, resulting from a rolling stationarity test. More outliers included in the test, gives higher p-values. Particularly in 2011 where outliers are concentrated, tests that include this area gave high p-values. In we observe year 2018, which has less outliers, the p-values were quite low. We exclude outliers from our dataset to test stationarity without them.

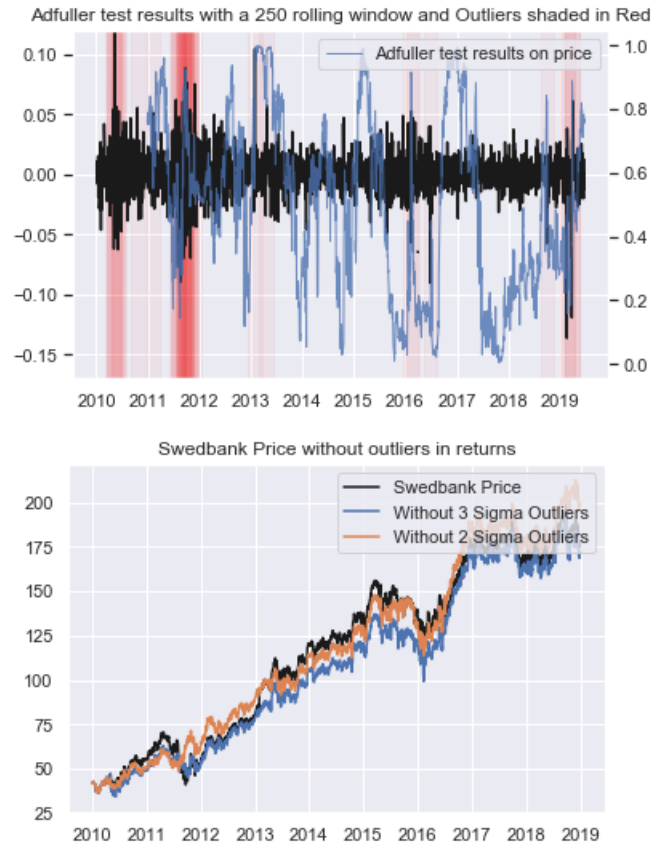[2]Zero Excess Kurtosis indicates a platykurtic distribution[34]

Figure 3.2: Outliers shaded in red with *p*-values of stationarity test

We exclude returns above two and three standard deviation from the mean of returns, recalculate the prices and do the same tests. *p*-values of stationarity tests are 0.558 and 0.292 for the two and three sigma excluded of returns, respectively. Rolling test results show no reasonable improvement (Figure 3.3) in *p*-values.

Even if outliers are the main reasons the tests fail to reject the existence of unit root, excluding just the break dates was not enough to achieve stationarity.

Therefore we decide to go deeper on how outliers effect ARIMA models. We start by looking at residuals, then the parameters. In Figure 3.4, blue spikes show high errors in daily forecasts. As expected, outliers create big forecasting errors, due to nature of financial time series out of sample error is way higher than in-sample error.
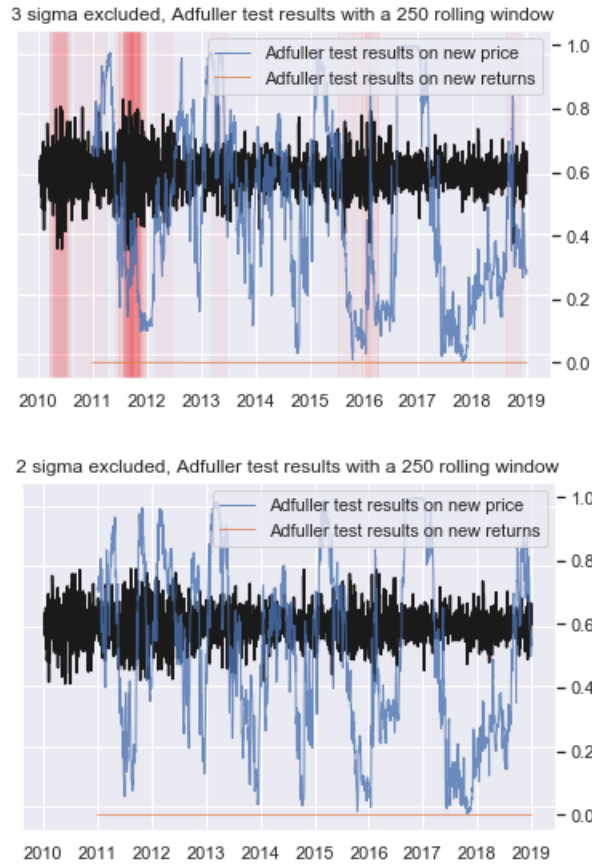
Figure 3.3: P-values of stationarity tests on time series with filtered outliers

These graphs show that economic events that effect these time series so suddenly often have big effects on the identified models. In this case ARIMA models fail to predict the occurrence of the economic events, which results in big prediction errors. These events also create differences in parameters of ARIMA models since parameters are decided on loss functions dependent on errors. To test the effects of structural breaks on parameters, we fit ARIMA models with different ending dates as shown in Figure 3.5.

Resulting changes show that ARIMA models, identified over different time periods on the same time series, have significant differences not only in parameters, but also in number of parameters (Fig. 3.5). To see how parameters evolve over time, we split the data into half and take a small step further into the test set in every iteration.

Figure 3.4: ARIMA errors on outliers

To see clearly the effect of outliers, we do the same iterative process but with a different step size and a larger data range in Figure 3.6. This fixes the type of the model into (1,1,1), making it easier to see the evolution of coefficients. Following the idea of the Chow test, we can see that outliers create big differences of the parameters of identified ARIMA models.
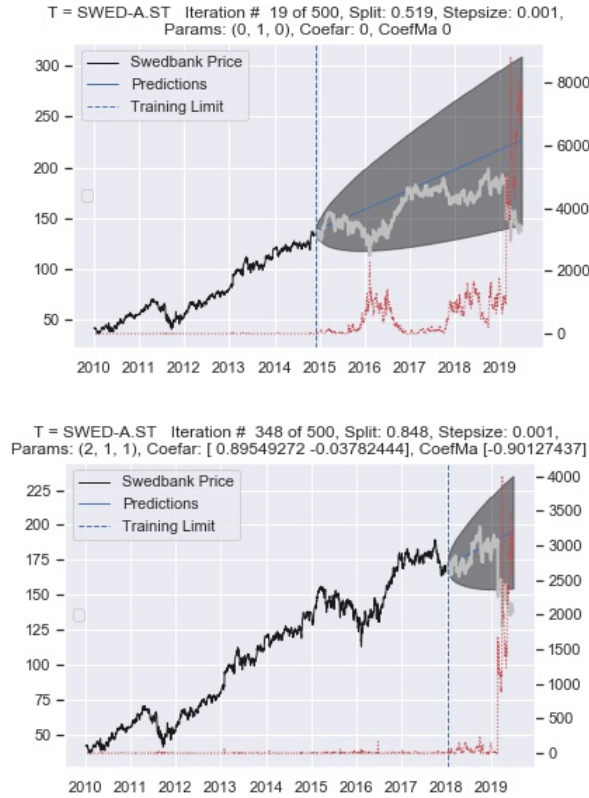
Figure 3.5: Iterations to identify ARIMA parameters over outliers

3.2 Abnormal Returns

To show effect of abnormal returns, we start by showing the returns "generated" by the market, using the market model. Using daily percentage returns from 2010 (Swedbank), we estimate a β, and α of 0.0003. Actual prices differ a lot from the prices generated using the market model. To show how prices would move according to the market model, we plot both. Price differences are visible, especially when big losses incurred in Swedbank.

Subtracting the returns calculated with market model from actual returns, we calculate abnormal returns. Using these returns and the same starting price, we "de-trend" the effect of the market from the prices. Prices calculated by using only abnormal returns are plotted in Figure 3.8. This can be an approximation of how economic events specific to Swedbank moved its overall price: a portfolio of 1.1 index shorts with a long Swedbank stock would look like this if we would ignore α in the market model.

To see average effects of outliers in prices, we plot histograms of abnormal returns around
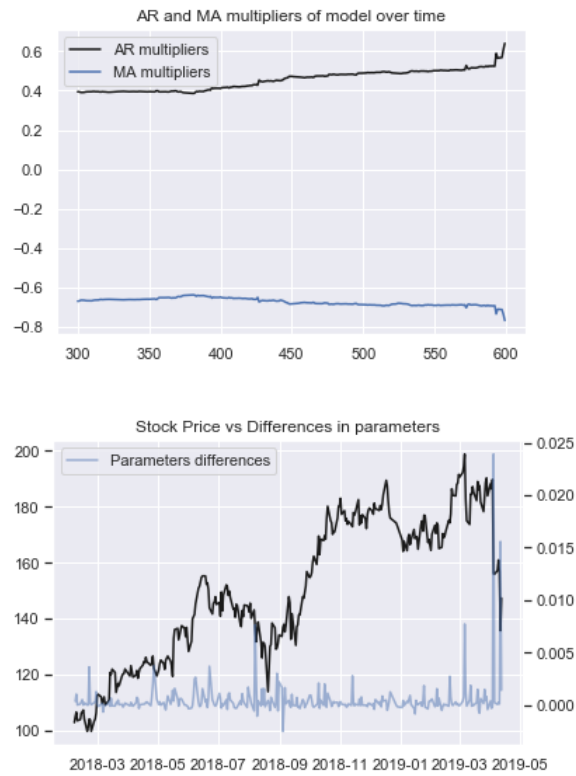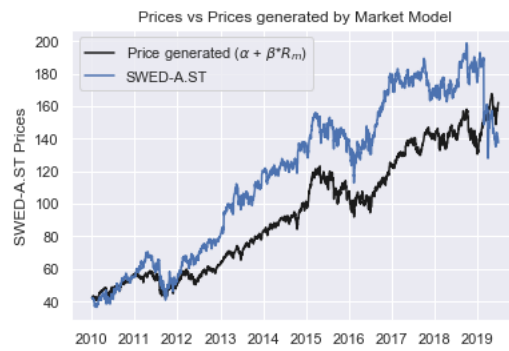
Figure 3.6: Paramater evolution

Figure 3.7: Prices vs market model prices

outliers. Outliers selected to be put in histograms are outliers bigger than $(\mu + 2 * \sigma)$, or for
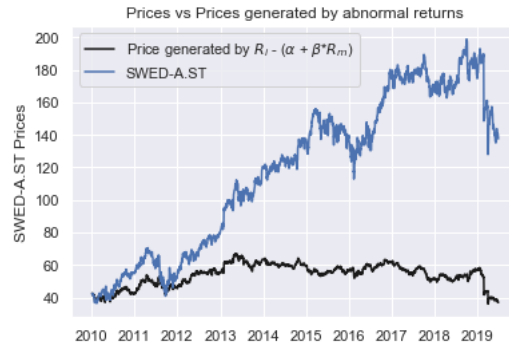
Figure 3.8: Actual prices vs de-trended prices

negative outliers, smaller than $(-\mu - 2 * \sigma)$.[3] Positive and negative outliers are plotted (Fig. 3.9) separately to check possible price run-ups. If there is significant insider trading, their run-ups should be positive and negative, respectively.

---

[3]μ is for average, σ is for std. deviation of percentage returns
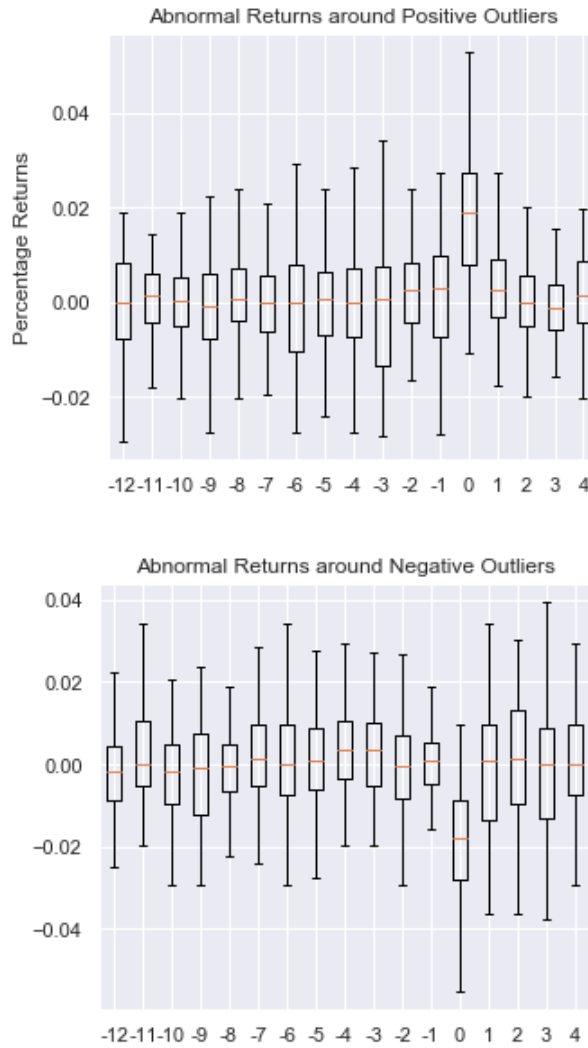
Figure 3.9: Boxplots of abnormal returns around outliers

To test the significance of abnormal returns, we calculate p-values for Welch's and Mann-Whitney U test, comparing the abnormal return of each day with the returns of entire Swed-bank returns [60], [39]. With $\alpha=0.05$, tests show no persistent difference thus we conclude that price run-ups are not statistically significant. Note that Welch's t-test assume both populations are normally distributed and independent, which in this case, are not.
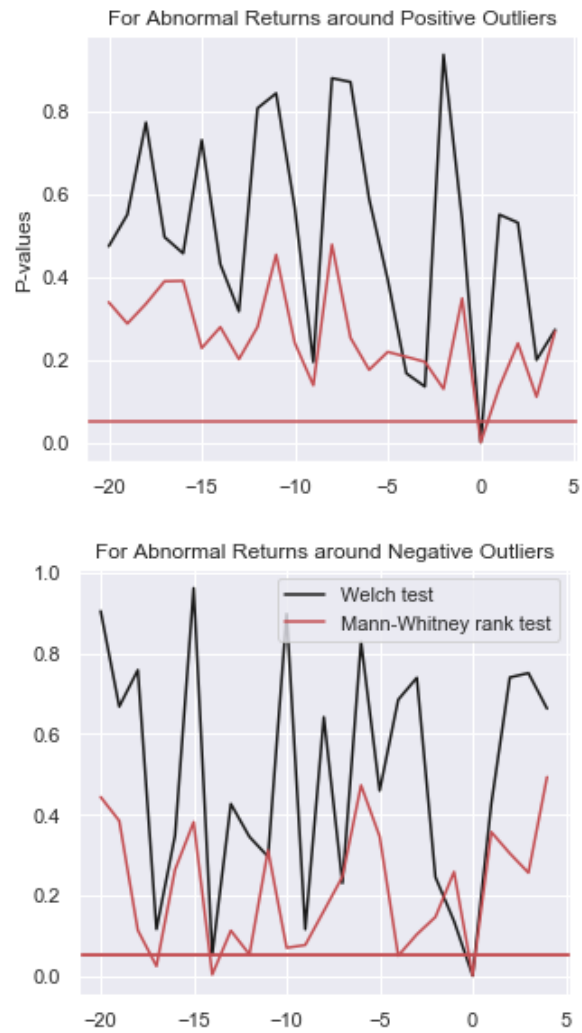
Figure 3.10: Welch's *t*-test on abnormal returns

Chapter 4

Events Studies with ANN

We have shown how outliers effect predictions in financial time series and how outliers in returns create big deviations from expectations. Such economic events, particularly the financial crisis in 2008 has pushed regulators to enforce risk management practices in financial institutions. To include these unusual events in risk management, extreme value theory, stress testing, scenario testing practices are commonly used [54]. However, we will look from a different angle and attempt to classify their effects with neural networks. We start by showing a simple example, attempt to predict events with neural networks, then try to classify effects of similar events with networks.

4.1 Detection

As mentioned before in Chapter 3, there are numerous tests to test the existence of outliers. Following the definition in Chapter 3, we first define the outliers and train the neural network illustrated in Figure 4.1 in order to test its capability to detect this definition.
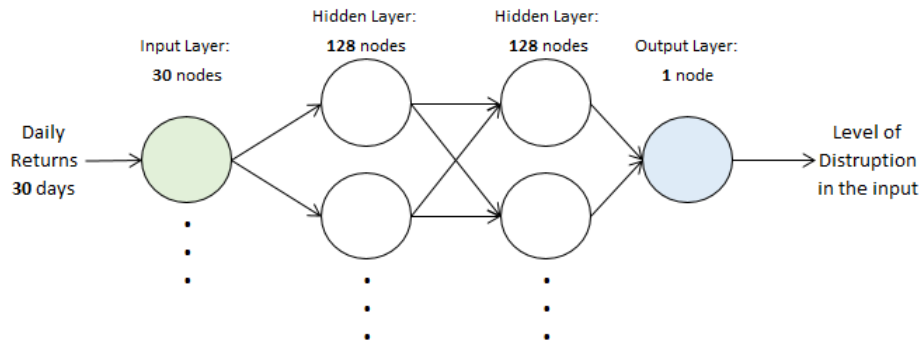
Figure 4.1: Structure of the neural network used in this section

Before trying classification on specific events, we demonstrate the method on a simpler dataset to show our steps. We mark dates with big price changes in Swedbank price similar to Equation 3.1. Figure 4.2 shows how big are detected price changes. Red line around 0.08 is the limit for 5 sigma. Price changes passing this limit are shaded red in the price plot.
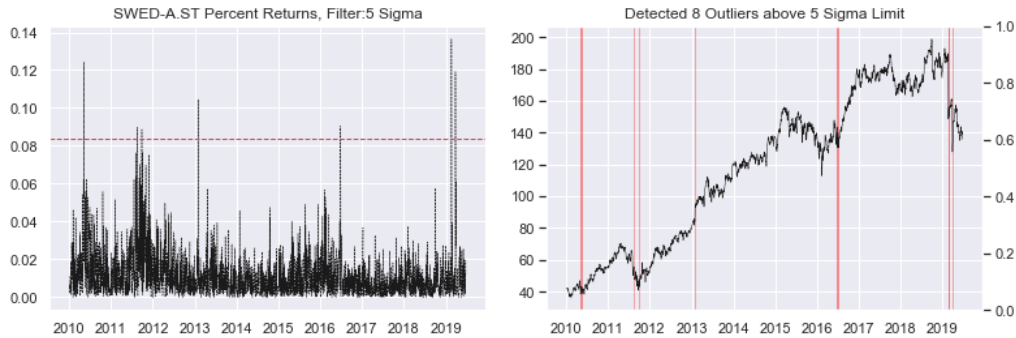
Figure 4.2: Outliers are detected by measuring deviation from the mean

In this example we want to show that a neural network can measure such simple calculation

and filtration. Thereby we scale the inputs of the network based on how big the outliers are. Magnitudes of outliers can be seen in Figure 4.3.
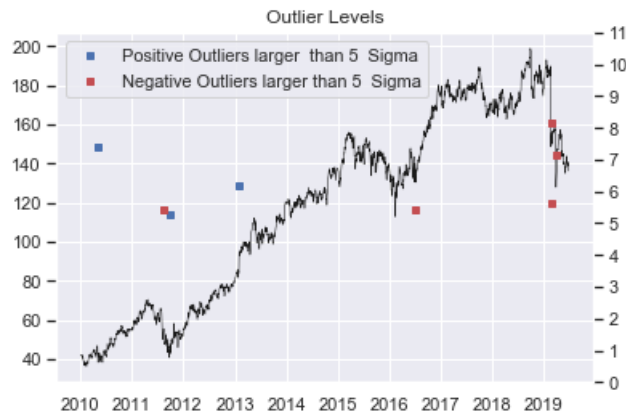
Figure 4.3: Resulting time series is the objective of the neural networks.

Results in Figure 4.4 show that outliers can be detected using neural networks. The accuracy is not good: it is visible from the graph, the network tends to give higher results then actual values and some false positives. This application can be extended to include error terms of statistical models instead of percentage returns, which could allow flexibility of detecting customly defined structural breaks in different models.
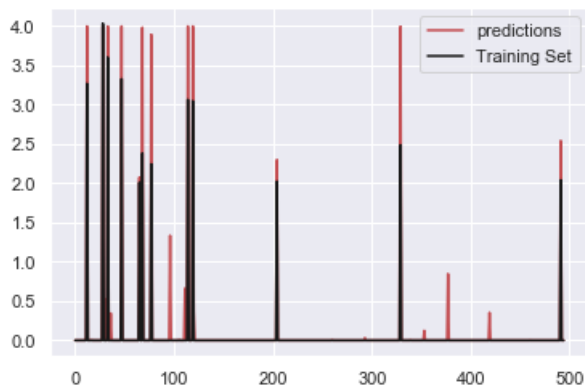
Figure 4.4: Resulting ANN outlier detection.

4.2 Pattern identification

According to research, financial asset movements have changed since the financial crisis [9] [17] [43]. We demonstrated a similar change in Figure 3.6 by showing differences in the parameters of identified models. We try to train the network to see if similar economic events create similar pattern changes that could be captured by the neural network.
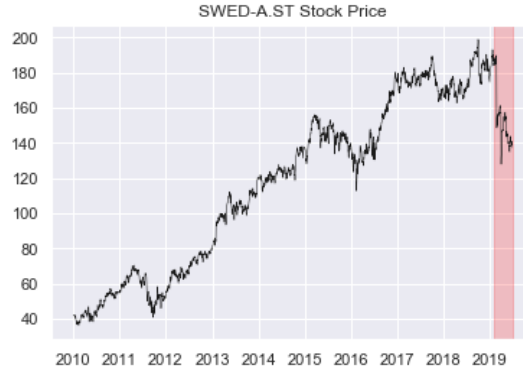
Figure 4.5: Red shaded area is not included in the first identification

However networks require much larger training samples and frequent occurrence of outliers (as seen in Figure 4.4) which eliminate the possibility of training on real data. Instead of using real data, we identify ARIMA models before and after the outlier (we exlude the red area in Figure 4.5) and simulate stock prices as an input to our networks in Figure 4.6. Identified model for the time series is ARIMA(4,1,3) while after the outlier identified model becomes ARIMA(1,1,2).
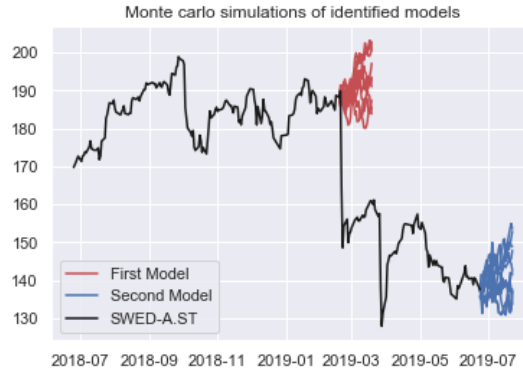
Figure 4.6: Monte Carlo simulations of different models

Since simulations have different parameters, we expect the neural networks to recognize the

patterns. To train the network according to different patterns, we mark the simulated series with a binary classification: returns for the first model as 0 and rest are labeled as 1. We use percentage returns as inputs and classification as outputs to train the network.

Figure 4.7: Classification of simulation of different ARIMA models, in sample

ANN successfully captures in-sample differences of the 30 day return simulations with a sample size of 400 paths, 200 for each model (Figure 4.7). To define a success measure, we define an output above 0.5 as an after-outlier simulation. In our 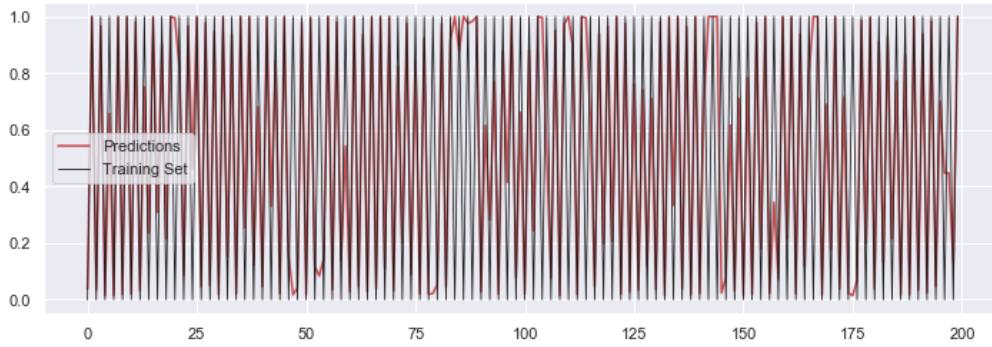training set, a half of the paths result in minor error: 10 false positives[1] and 7 false negatives. Out of sample, 26 paths were false positive and 47 paths were false negative, totalling a 127 correct classification out of 200 paths.

To increase accuracy, we increase the number of epochs and simulated paths. As expected, number of correct classification improves. In sample size of 4000 results 3980 correct classification while out of sample is 3388 out of 4000, demonstrating its ability to recognize patterns generated models with different parameters. We use trained network on stock prices of Swedbank.

Consistent positive output of ANN after the big drops in the stock price show that this network is able to recognize different patterns in returns (Figure 4.8), caused by the structural breaks. To apply this network to the available data, we take rolling mean of the ANN outputs to smooth its high oscillation, filter scores below 0.95: this results in scores that have averaged more than 0.95 in the past 30 days. To compare results with actual outliers, we shade outlier dense dates.

Smoothing and filtration cleaned the output quite well (Figure 4.9). Note that last positive value (2019) in this graph is our "training set". Same scoring and detection is visible in 2015 and 2018, where there were only single outliers in the scanning input and ANN output gave a rolling average above 0.95: successfully detecting patterns created by outliers. Even though its success is visible from the graph, measuring its performance is not an easy task.

---

[1]Paths generated by the first model were classified as the second model by the ANN
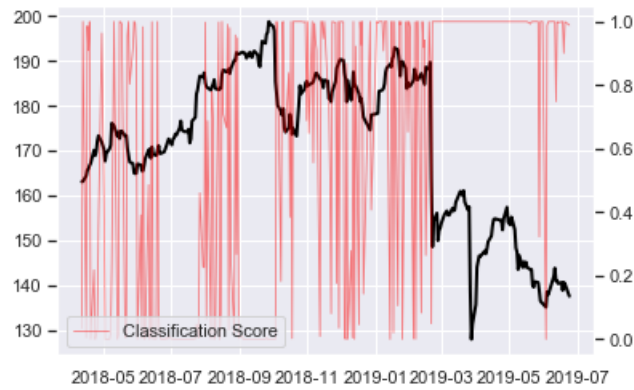
Figure 4.8: Classification on Swedbank percentage returns
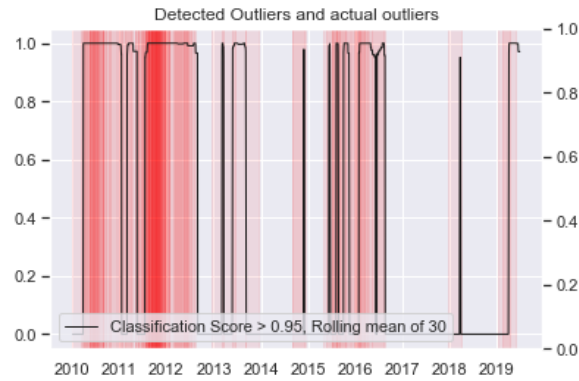
Figure 4.9: ANN classification vs actual outliers

Firstly, frequent occurrence of outliers eliminate the possibility of checking if a particular out-lier is detected. Even though we included two biggest outliers (totals a drop bigger than 15 std deviations) in the training, network was able to detect smaller outliers, raising questions about its sensitivity. Over-sensitivity issues could be double-checked with the number of false positives but it is not a reliable measure in this case since there are none in this particular experiment.

Secondly, measuring correct number of false negatives is not possible: outliers are quite close to each other and once an outlier is detected, ANN output for the next 30 days is expected to be positive, eliminating the chance of detecting a false negatives that are next to any true positives.

4.3 Outlier detection using Monte Carlo simulations

Experiments in Section 4.1 failed to capture outliers because of the small training set: experiments in Section 4.2 captured outliers however failed to give accurate results in regards to their location because of the network architecture. Therefore we change the way our networks work and use Monte Carlo simulations in this section in order to create big dataset for sufficient training of the network. To preserve properties of the outliers, we did not use a scaler and used ReLU activation function. High number of inputs and outputs required longer training, we increased number of epochs to 200. Structure of the network is shown in Figure 4.10.

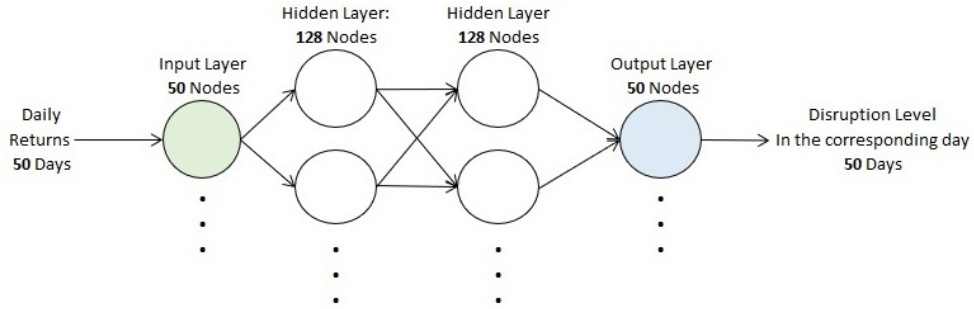Figure 4.10: Structure of the network used in this section

In order to create samples, we used ARIMA identification on the Swedbank prices and created 6500 paths of price movements. Identified model is an ARIMA (3,1,4) model with AR coefficients 0.67, -0.47, 0.50 and MA coeffiecients -0.67, 0.45, -0.48, 0.90 for respective 1, 2, 3, and 4th parameters, and a drift constant of 0.06. High order of coefficients required a high daily simulation for each path: we simulated 50 days for every path. To include outliers, we created manual drops in the stock prices and simulated rest of the price paths: just subtracting a number from stock prices would violate the autoregressive nature of prices. Figure 4.11 shows an example of a normal simulation and a simulation with a manual interaction. Prices behave similarly to a jump diffusion process because of the autoregressive components in the ARIMA: there is a mean-reverting tendency after the drops.

To be able to train the all areas of neural network, outliers of all possible different levels and outliers from all of the inputs are necessary. We have created manual drops homogeneously (as seen from the mean of MC simulations in Figure 4.11 in the simulated paths and manual drops up to 15 std deviations. Every level of disruption was simulated 500 times, totalling 6500 price paths. Since we have 50 output nodes, unlike in previous sections, we have to label the prices/returns, not the entire path. In order to do so, we scan all simulated prices and calculate the daily levels of disruption. Figure 4.12 shows manually dropped 2 paths and their corresponding labeling as an example to illustrate how we trained our neural network.

Results were far better than in previous two experiments, thanks to the Monte Carlo simula-
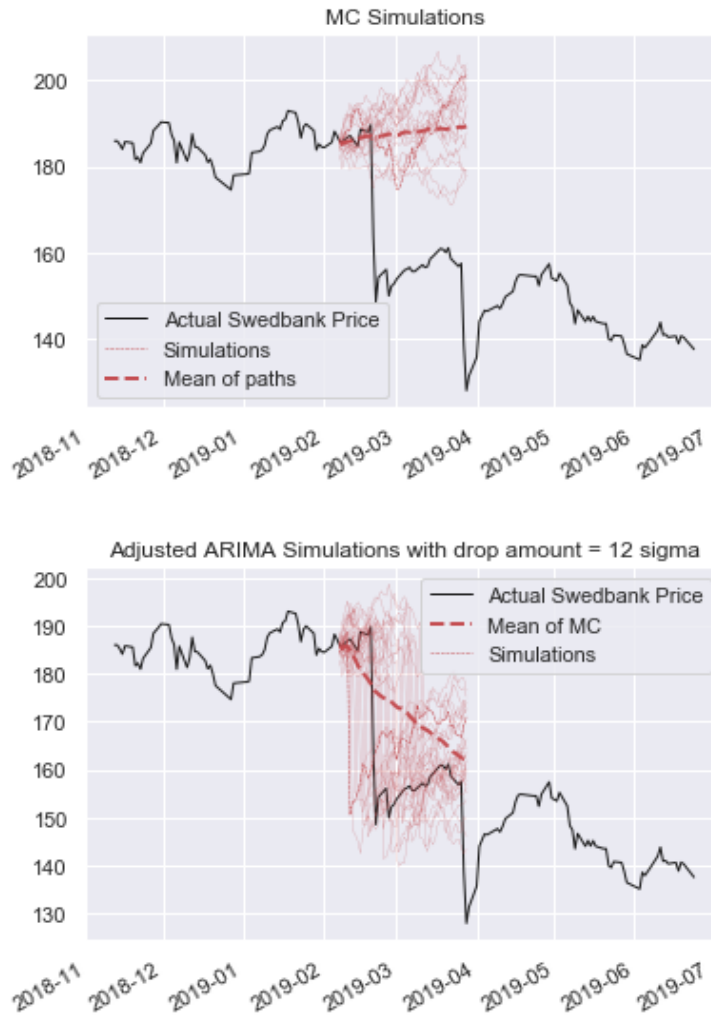
Figure 4.11: ARIMA simulations and simulations with manual drops

tions that provided the sample size necessary for training. Trained network is able to pinpoint locations of the outliers but also measure their magnitude according to Equation (3.1). Biggest measurement error was 0.19, which occurred in 2010-05-10, where the stock price was 43.18 Swedish Kronas, and previous Adj Closing price in 2010-05-07 was 38.4. This was a 12% return and a 7.14 sigma jump where our network gave 7.33, 0.19 higher than the actual value. In this sample of 2380 ANN outputs, all outputs were in the 0.1 / -0.1 range of the actual targets except 5 days (see Figure 4.13), demonstrating the usefulness of Monte Carlo simulations in training neural networks.
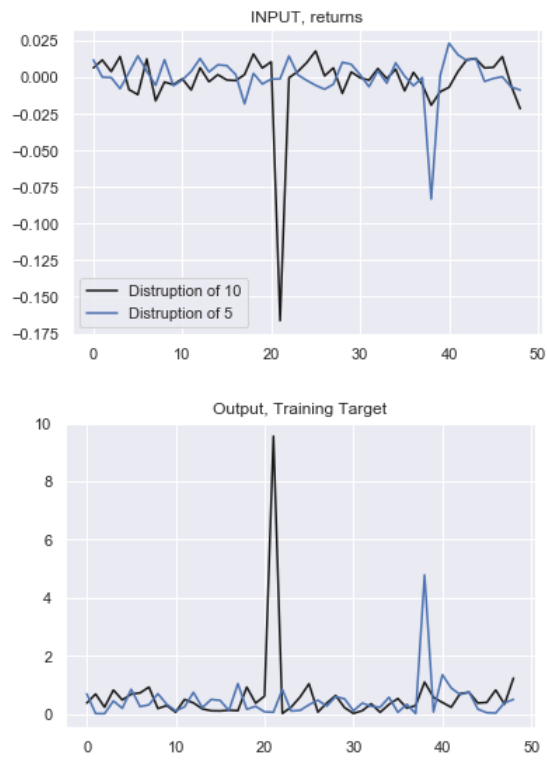
Figure 4.12: Manual drop and labeling of 2 generated paths
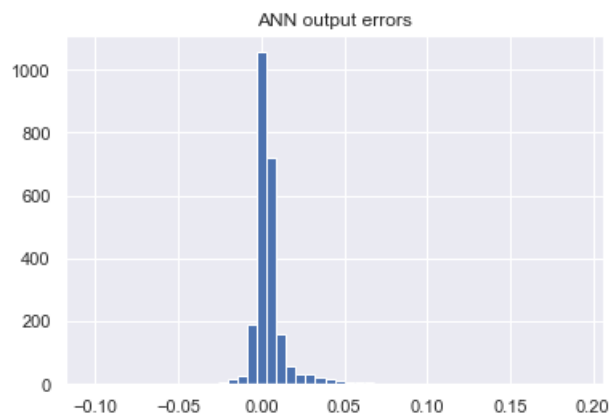
Figure 4.13: Histogram of errors of ANN used for predicting the magnitudes of outliers

57

4.4 Classification with technical indicators

According to strong efficient market hypothesis, prices reflect all information, public and non-public and it is not possible to achieve superior returns [37]. Price run-ups towards particular economic events in aforementioned articles show that some investors were able to predict these events, trading the securities and effecting the prices, and thus making profits. Some technical analysts were able to foresee securities fraud, in case of Enron and Worldcom, due to high trading volumes [32]. If there are any technical indicators that would detect abnormalities, neural networks might classify them. Therefore, we include additional information to train our network. Structure of our network is shown in Figure 4.14.
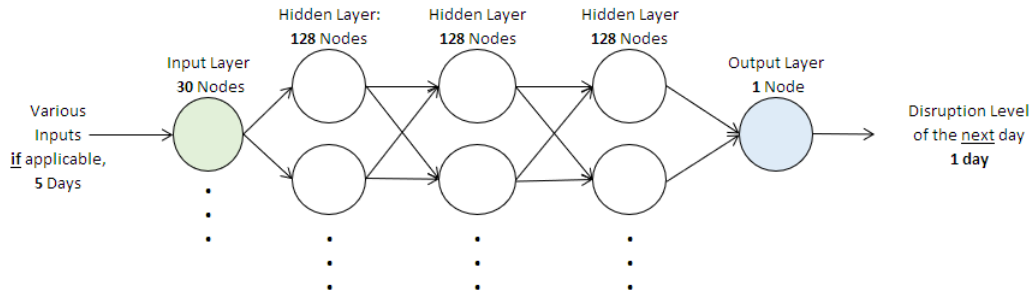
Figure 4.14: Structure of ANN for price run-up recognition

We include, 5, 10 and 30 days of moving averages of adjusted close prices, 5 days rolling standard deviation, previous 5 days of: returns, differences between daily high/low and opening/closing prices, trading volumes and adjusted closing prices. To train the network according to information available before, thus testing abnormalities preceding events, we shift the magnitude vector by one day.

Network failed to recognize shared patterns (if there are any) before the outliers both in sample and out of sample (See Figure 4.15). This result was anticipated: if classification of such patterns were an easy task, participants in the market would take advantage and market would transfer this information to the price and event might not appear as an outlier.
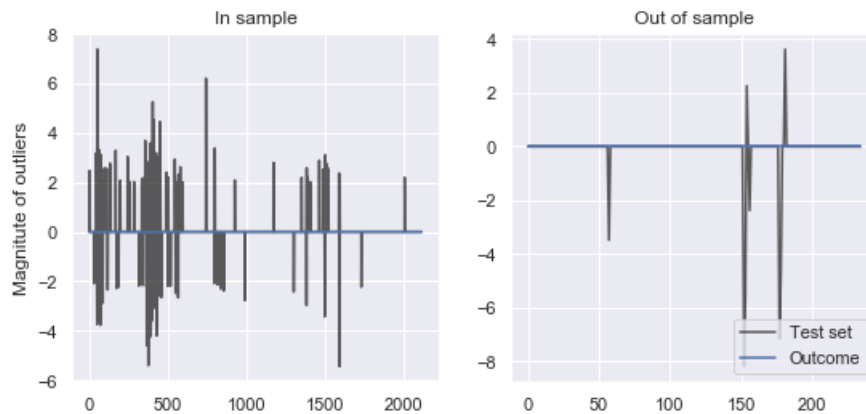
Figure 4.15: Result of classification of preceding patterns

4.5 Classification of abnormal returns before outliers

ANN failed to recognize the difference in abnormal patterns in stocks before events. After this result, we change the structure of the network and try to classify abnormal return patterns before the outliers. In this section, we use abnormal returns, in groups of 5 days. We mark each group of 5 with binary classification, where a group of 5 before an outlier is marked 1. Rest of the return groups are marked 0. Structure of the new network is shown in Figure 4.16.
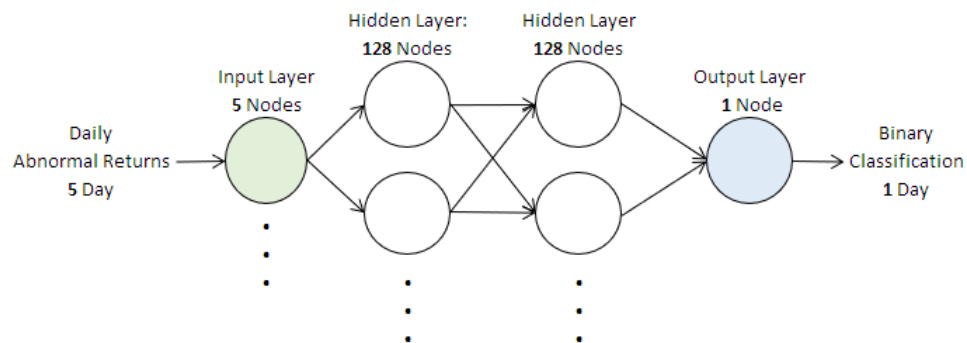
Figure 4.16: ANN structure, classifying abnormal returns

This structure will allow us to teach the network common patterns before outliers, if any are existent. Healthier way to do this would be to feed intraday data with a much larger equity sample, preferably from the same market and equity classes (small/high cap, industrial etc.) since similar markets and companies are found to have similar price behaviors before

economic activities in the aforementioned research. Output of the network are shown in Figure 4.17.
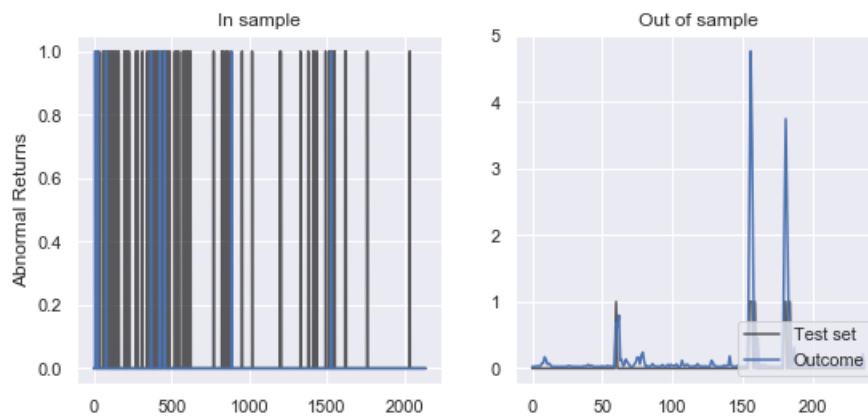
Figure 4.17: Output, Classifying Abnormal Returns

After many iterations to train the network, we have seen a single non zero output of the network, as illustrated above. We find this output unsatisfactory because of its bad fit (visible blue lines on the plot to the left are false positives) in sample. Assuming outputs above 0.5 are positives, network gave only 9 true positives, 7 false positives and 106 false negatives. These 9 true positives show that 9 of the outliers actually had a similar price pattern for 5 days, but this pattern was shared for other 7 false positives where it wasn't followed by an outlier. Thereby common patterns might exist around outliers, but this network has failed to recognize the patterns shared by all of them. A network that would succeed in such task could be used to scan prices and look for possible outliers or could be used by regulatory authorities to look for possible insider trading activities. Possibly, bigger sample sizes or focusing on particular economic events can help us detect shared patterns. We select a particular economic event in the following section.

4.6 Classification of specific events

In research, abnormal returns are often used to analyze effects of quarterly reports, management changes, dividend announcements and price-runups measure insider trading activities related to M&A activities. Past experiment showed us that even if there are shared patterns before the events that create outliers in the price, neural networks failed to recognize them. In this experiment, we focus only on a specific kind of economic event with a bigger sample space: price movements both before and after the M&A announcements. We get the announcements from Mergr.com and prices via the Yahoo finance[2]. Selected companies (tickers) are: Apple (AAPL), Adobe (ADBE), Amazon (AMZN), Salesforce.com (CRM), Facebook (FB), Alphabet (GOOG), Intel (INTC), Microsoft (MSFT), Oracle (ORCL) and VMware (VMW). The details of the percentage returns of the selected data are as follows:

	AAPL	ADBE	AMZN	CRM	FB	GOOG	INTC	MSFT	ORCL	VMW
Count	2634	2634	2634	2634	1783	2634	2634	2634	2634	2634
Mean	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.001
Median	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Std	0.016	0.019	0.021	0.022	0.023	0.015	0.016	0.015	0.015	0.023
Min	-0.12	-0.19	-0.12	-0.12	-0.18	-0.08	-0.09	-0.11	-0.11	-0.21
25%	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01
50%	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
75%	0.010	0.010	0.012	0.012	0.011	0.008	0.009	0.008	0.008	0.012
Max	0.088	0.127	0.267	0.181	0.296	0.160	0.109	0.105	0.101	0.167
Skewness	-0.06	0.097	1.271	0.620	1.272	0.689	0.125	0.043	-0.17	-0.39
Kurtosis	3.842	9.286	15.61	6.471	22.65	11.15	4.167	7.169	7.058	10.73

Method and the neural network used in this section is parallel to the previous one. Instead of previous 5 days, we mark a series of returns starting from 9 previous days to 5 preceeding days to 1, and all other series with 0. Doing so will help neural network to capture any existing patterns, not only before or after the event days, but also during the event day. Structure of the network is show in Figure 4.18.

Next step is to calculate abnormal returns. Since smaller rolling windows give better estimates around outliers, we choose to use a window of 60 days for the OLS estimation. Index used is the S&P500.

Using the abnormal returns and the binary classification of the occurence of M&A activities, we have trained the network for price behaviour of particular economic events. Unlike the previous experiment we have a bigger sample size with a focus on a specific event. However, the network was unable to capture any patterns, indicating absence of any shared movement, or rather, the network is not suitable for such task. Output for the test set is shown in Figure 4.19.

---

[2]https://mergr.com/ and https://finance.yahoo.com/
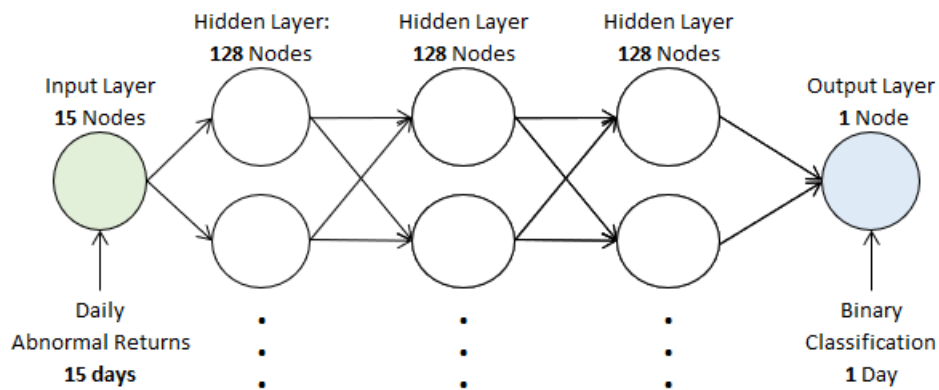
Figure 4.18: Structure of neural network for classification of specific events

Therefore, we change the way we calculate abnormal returns: instead of using the market model and a rolling linear regression, we use Kalman Filter approach to estimate market model coefficients and calculate abnormal returns.

Using the Kalman filter to estimate abnormal returns (Beta's and abnormal returns for respective companies can be seen in Figure 4.21) did not change the result: the shared return patterns generated by the acquirer companies are unable to be recognized by the network, as seen in Figure 4.20. Therefore we were unable to prove the existence of similar price run-ups of the acquirer companies.

Figure 4.19: Economic event classification, OLS approach

Figure 4.20: Economic Event classification, Kalman filter approach

Figure 4.21: Betas and Abnormal returns of selected companies

Chapter 5

Conclusion

We have analyzed various effects of economic events on stock prices. Economic events can create outliers in returns, cause non-stationarity and non-normality and in turn can be a cause of big parameter differences in statistical models as we have demonstrated with applications in ARIMA, Market Model and Kalman Filter. We labeled the econom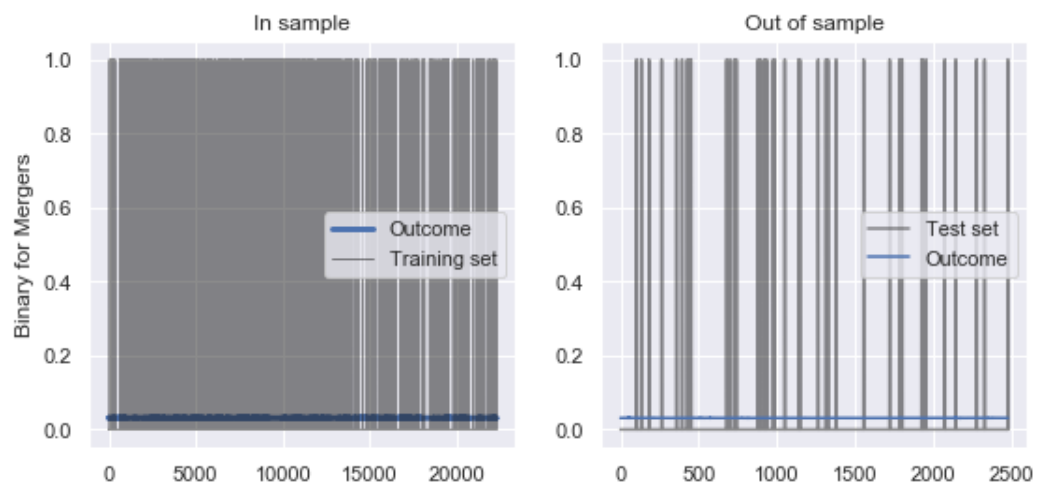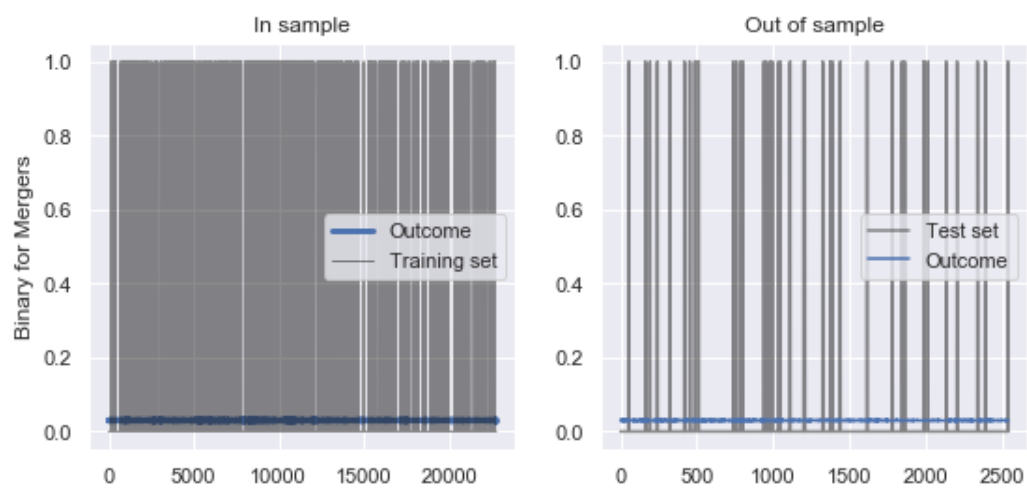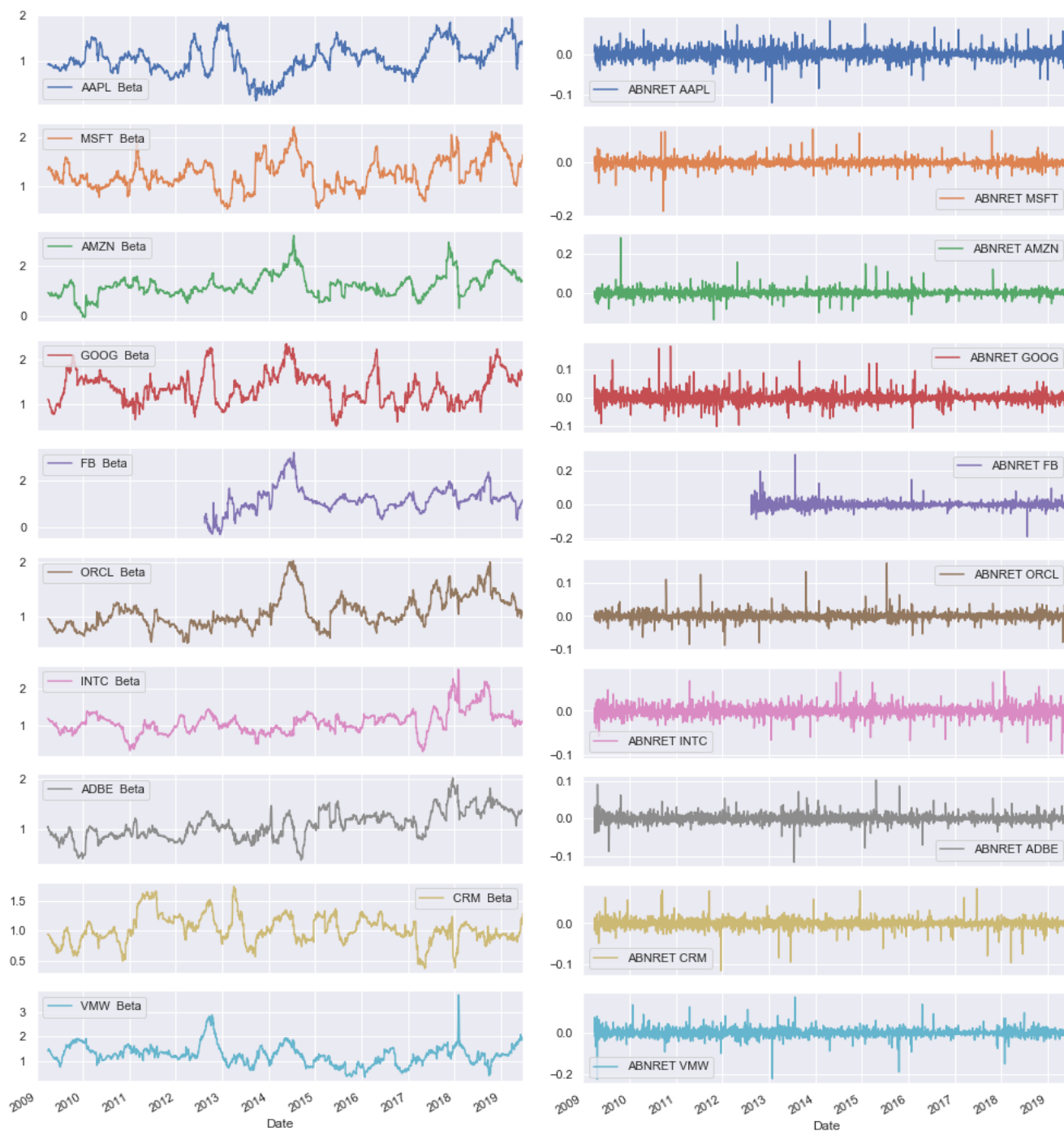ic events with a straight-forward measure of its magnitude and trained deep learning algorithms. The trained network failed to measure outliers accurately. Monte Carlo simulations, which created the necessary sample size, increased accuracy when detecting outlier locations.

According to the research, specific events (M&A) can cause price-run ups in abnormal returns and some investors were able to predict economic events using technical analysis [32]. Using the market model and the Kalman Filter, we calculated abnormal returns to analyze event specific returns. In separate experiments, we used technical indicators and abnormal returns for training the neural networks, however deep learning methods failed to capture any shared patterns around these economic events. Therefore we conclude that either Artificial neural networks are not suitable for learning complex behaviour of returns around economic events or training requires a large sample set.

For future work, we propose three main scopes. Firstly, in scope of Modern Portfolio Theory, Kalman filter can be used to estimate CAPM parameters and the resulting portfolio can be compared with other methods. Second is to use Kalman filter to generate MC simulations to see if it can improve the detection of outliers, or to use ANN spot more complex structural breaks to demonstrate its usefulness compared to traditional methods. Third and lastly, we encourage the use of different models to calculate abnormal returns and usage of more complex neural networks with larger data-sets, to test the possibility of detecting economic events prior to their occurrence or spotting insider trading activities, as such tools can be used to generate economic profits or can be used by regulatory authorities to spot insider traders.

5.1 Summary on reflections of objectives of the thesis

As mentioned earlier, our work was split into three main parts:
(1) Understanding the data & identifying the right statistical model.
(2) Analyzing measuring the effects of economic events.
(3) Training neural networks with various inputs in an attempt to train the effects of economic events to the neural network.

In literature review we mainly have research about last two main parts. For (2): we showed different approaches for detecting structural breaks, analyzing their effects in different scopes and with different methods. For (3): we also showed different uses of neural networks parallel to our scope and how particular events were studied in the scope of events study in the literature. Later on we explained how different models used in literature for events studies in Time Series Concepts chapter, parallel to our step (1). We consider our literature review comprehensive as we dove deep in all 3 steps to find relevant research sensitive to our research scope. We have demonstrated broad knowledge in our field of study and showed in-depth insight into current research and development, in according to objective 1.

In Time Series Concepts Chapter, we have used statistical tests to understand our dataset and explain its statistical properties. We identified different models along with their goodness of fits. We also tested different models with different ranges for a better fit and did extra literature review, as output of this step would be sensitive for us in step (2). We showed substantially deeper knowledge in areas in our scope for objective 1 and demonstrated deeper methodological knowledge in the scope of our research, in according to objective 2.

In Data Analysis Chapter, we further analyzed our dataset, to explain the effects of economic events, using the tests and models used in previous sections. We have studied structural breaks, defined our own break definition and illustrated effects of economic events on stationarity, normality, returns and on the parameters of statistical models. Furthermore we tested the existence of statistically significant abnormal returns around economic events, concluding this chapter with a traditional events study method. We integrated new knowledge from to the next section, demonstrating the ability to deal with complex issues and situations parallel to our scope, in according to objective 3.

In Event Studies with ANN Chapter, we used neural networks with different designs and different inputs (note that also we used different models used these inputs) to test if networks can detect economic events and patterns. Neural networks successfully detected economic events and usage of MC simulations increased its accuracy. Networks were able to capture shared patters after the event. As expected, networks failed to capture accurately common patterns prior to the event dates, even when we used technical indicators, focused on a single event (M&A) or used Kalman filter to calculate abnormal returns. We have identified new questions based on new findings and demonstrated significant ability to carry out advanced tasks and

evaluated the work with appropriate measurements, in parallel to objective 4.

We have successfully illustrated and used our methods in all (3) steps of our work. Even though our methods failed to capture the shared patterns prior to economic events, we have successfully demonstrated methodology of event studies with flexible approaches: usage of different models to calculate abnormal returns and different ways to analyse and classify them (box plots, Welch's test and ANN). We showed significant ability in international contexts, by experimenting in both national and international data and providing conclusions. We showed significant ability in different contexts clearly, by writing about different areas and presenting the paper to both technical and non-technical audiences, in parallel to objective 5. We have demonstrated significant ability to make judgements based on scientific aspects (via usage of scientific methods), societal aspects (we showed the effect of Efficent Markets in introduction) and ethical issues (by proposing methods for regulatory authorities), in parallel to objective 7.

Bibliography

[1] ADAMS, Z., AND GLÜCK, T. Financialization in commodity markets: A passing trend or the new normal? *Journal of Banking & Finance 60* (2015), 93–111.

[2] ADHIKARI, R., AND AGRAWAL, R. K. An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613* (2013).

[3] AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 6 (December 1974), 716–723.

[4] ALLARO, H. B., KASSA, B., AND HUNDIE, B. A time series analysis of structural break time in the macroeconomic variables in ethiopia. *African Journal of Agricultural Research 6*, 2 (2011), 392–400.

[5] ANDREWS, D. W. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society* (1993), 821–856.

[6] ANDREWS, D. W., AND FAIR, R. C. Inference in nonlinear econometric models with structural change. *The Review of Economic Studies 55*, 4 (1988), 615–640.

[7] BAGLIANO, F. C., FAVERO, C. A., AND NICODANO, G. Insider Trading, Traded Volume and Returns. Working papers 26, Former Department of Economics and Public Finance "G. Prato", University of Torino, Oct. 2011.

[8] BAI, J., AND PERRON, P. Estimating and testing linear models with multiple structural changes. *Econometrica* (1998), 47–78.

[9] BEKIROS, S., NGUYEN, D. K., JUNIOR, L. S., AND UDDIN, G. S. Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *European Journal of Operational Research 256*, 3 (2017), 945–961.

[10] BLACK, F. Capital market equilibrium with restricted borrowing. *The Journal of business 45*, 3 (1972), 444–455.

[11] BROWN, S. J., AND WARNER, J. B. Using daily stock returns: The case of event studies. *Journal of financial economics 14*, 1 (1985), 3–31.

[12] BÜYÜKŞAHIN, B., AND ROBE, M. A. Speculators, commodities and cross-market linkages. *Journal of International Money and Finance 42* (2014), 38–70.

[13] CAMPBELL, J. Y., CHAMPBELL, J. J., CAMPBELL, J. W., LO, A. W., LO, A. W., AND MACKINLAY, A. C. *The econometrics of financial markets*. Princeton University press, 1997.

[14] CHENG, I.-H., AND XIONG, W. Financialization of commodity markets. *Annu. Rev. Financ. Econ. 6*, 1 (2014), 419–441.

[15] DADAKAS, D., KARPETIS, C., FASSAS, A., AND VARELAS, E. Sectoral differences in the choice of the time horizon during estimation of the unconditional stock beta. *International Journal of Financial Studies 4*, 4 (2016), 25.

[16] DANIELSSON, J., VALENZUELA, M., AND ZER, I. Learning from history: Volatility and financial crises. *The Review of Financial Studies 31*, 7 (2018), 2774–2805.

[17] DASKALAKI, C., AND SKIADOPOULOS, G. Should investors include commodities in their portfolios after all? new evidence. *Journal of Banking & Finance 35*, 10 (2011), 2606–2626.

[18] DE, I., AND NUTMAN, J. Decomposition of total volatility: How important is idiosyncratic risk? *Finance Professionals' Post, Publication of New York Society of Security Analysts* (2015).

[19] DELCEY, T. Samuelson vs fama on the efficient market hypothesis: The point of view of expertise. *Œconomia. History, Methodology, Philosophy*, 9-1 (2019), 37–58.

[20] DERMAN, E., AND WILMOTT, P. The financial modelers' manifesto. *Available at SSRN 1324878* (2009).

[21] DONG, Y., LI, S., AND GONG, X. Time series analysis: An application of arima model in stock price forecasting. In *2017 International Conference on Innovations in Economic Management and Social Science (IEMSS 2017)* (2017/04), Atlantis Press.

[22] ELLIOTT, R. J., VAN DER HOEK*, J., AND MALCOLM, W. P. Pairs trading. *Quantitative Finance 5*, 3 (2005), 271–276.

[23] FAMA, E. F., FISHER, L., JENSEN, M. C., AND ROLL, R. The adjustment of stock prices to new information. *International economic review 10*, 1 (1969), 1–21.

[24] GALEANO, P., AND WIED, D. Multiple break detection in the correlation structure of random variables. *Computational Statistics & Data Analysis 76* (2014), 262–282.

[25] GOODFELLOW, I., BENGIO, Y., AND COURVILLE, A. *Deep learning*. MIT press, 2016.

[26] GÖRANSSON, L., AND TZAKOV, I. P. An evaluation of asset pricing models in the swedish context - is carharts four-factor model more suitable than its predecessors for explaining the swedish stock exchange?

[27] HANSEN, B. E. The new econometrics of structural change: dating breaks in us labour productivity. *Journal of Economic perspectives 15*, 4 (2001), 117–128.

[28] HANSSON, M. On stock return prediction with lstm networks. Master's thesis, Lund University, 2017.

[29] HOLLSTEIN, F., PROKOPCZUK, M., AND WESE SIMEN, C. How to estimate beta? Tech. rep., Hannover Economic Papers (HEP), 2017.

[30] HULL, J. C. *Options futures and other derivatives*. Pearson Education India, 2003.

[31] HYNDMAN, R. J., AND KOEHLER, A. B. Another look at measures of forecast accuracy. *International journal of forecasting 22*, 4 (2006), 679–688.

[32] INSTITUTE, C. *CFA Program Curriculum 2020 Level I Volume 1*. CFA Curriculum 2020. Wiley, 2019.

[33] KALMAN, R. E. A new approach to linear filtering and prediction problems. *Journal of basic Engineering 82*, 1 (1960), 35–45.

[34] KOKOSKA, S., AND ZWILLINGER, D. Crc standard probability and statistics tables and formulae, student edition, mathematics/probability. *Statistics (London: Taylor and Francis)* (2000).

[35] MAAS, A. L., HANNUN, A. Y., AND NG, A. Y. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (2013), vol. 30, p. 3.

[36] MACKINLAY, A. C. Event studies in economics and finance. *Journal of economic literature 35*, 1 (1997), 13–39.

[37] MALKIEL, B. G. Efficient market hypothesis. In *Finance*. Springer, 1989, pp. 127–134.

[38] MANDELBROT, B. B., AND HUDSON, R. L. *The (mis) behaviour of markets: a fractal view of risk, ruin and reward*. Profile books, 2010.

[39] MANN, H. B., AND WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* (1947), 50–60.

[40] MARIN, C. G. Stock price run-up before m&a announcements:evidence of insider trading or market anticipation? Master's thesis, University of Amsterdam, 2015.

[41] MCWILLIAMS, A., AND SIEGEL, D. Event studies in management research: Theoretical and empirical issues. *Academy of management journal 40*, 3 (1997), 626–657.

[42] MERGNER, S., AND BULLA, J. Time-varying beta risk of pan-european industry portfolios: A comparison of alternative modeling techniques. *The European Journal of Finance 14*, 8 (2008), 771–802.

[43] MITCHELL, M., PULVINO, T., AND STAFFORD, E. Price pressure around mergers. *The Journal of Finance 59*, 1 (2004), 31–63.

[44] MOHAMAD, I. B., AND USMAN, D. Standardization and its effects on k-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology 6*, 17 (2013), 3299–3303.

[45] NARAYAN, P. K., NARAYAN, S., AND SHARMA, S. S. An analysis of commodity markets: what gain for investors? *Journal of Banking & Finance 37*, 10 (2013), 3878–3889.

[46] PITMAN, J., AND YOR, M. A guide to brownian motion and related stochastic processes. *arXiv preprint arXiv:1802.09679* (2018).

[47] PUKELSHEIM, F. The three sigma rule. *The American Statistician 48*, 2 (1994), 88–91.

[48] PUNALES, A. Time-varying coefficient models and the kalman filter: applications to hedge funds. *Ryerson University Masters dissertation* (2011).

[49] RATNER, M., AND LEAL, R. P. Tests of technical trading strategies in the emerging equity markets of latin america and asia. *Journal of Banking & Finance 23*, 12 (1999), 1887–1905.

[50] ROUTLEDGE, B. R., SACCHETTO, S., AND SMITH, N. A. Predicting merger targets and acquirers from text. Tech. rep., November 2016, working Paper.[Online]. Available: pdf/mastelatest. pdf, 2013.

[51] SANTURKAR, S., TSIPRAS, D., ILYAS, A., AND MADRY, A. How does batch normalization help optimization? In *Advances in Neural Information Processing Systems* (2018), pp. 2483–2493.

[52] SEVINDIK, İ., AND GÖKGÖZ, F. A comparative abnormal return analysis of mergers and acquisitions in the emerging markets. *MANAGEMENT FROM AN EMERGING MARKET PERSPECTIVE* (2018), 131.

[53] SHOVEN, J. B., AND SIALM, C. The dow jones industrial average: the impact of fixing its flaws. *The Journal of Wealth Management 3*, 3 (2000), 9–18.

[54] SIMONS, K., ET AL. Model error. *New England Economic Review* (1997), 17–28.

[55] SORNETTE, D. Physics and financial economics (1776–2014): puzzles, ising and agent-based models. *Reports on progress in physics 77*, 6 (2014), 062001.

[56] STOLL, H. R., AND WHALEY, R. E. The dynamics of stock index and stock index futures returns. *Journal of Financial and Quantitative analysis 25*, 4 (1990), 441–468.

[57] TANG, Z., AND XU, X. What causes the target stock price run-up prior to m&a announcements? *Journal of Accounting and Finance 16*, 6 (2016).

[58] TEGNÉR, G. Recurrent neural networks for financial asset forecastings. Master's thesis, KTH Royal Institute of Technology, 2018.

[59] TSAY, R. S. *Analysis of Financial Time Series*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2010.

[60] WELCH, B. L. The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika 34*, 1-2 (01 1947), 28–35.

[61] WIDEGREN, P. Deep learning-based forecasting of financial assets. Master's thesis, KTH Royal Institute of Technology year, 2017.

[62] ZHUANG, Q. Application of kalman filtering in dynamic prediction for corporate financial distress. In *Kalman Filters-Theory for Advanced Applications*. IntechOpen, 2017.