Identifying Bubbles in Financial Data

Cullen, Conor Dave, Ishen Fullerton, Max Losty, Stefan Pearson, Beth

February 22, 2019

Project Supervisor
Dr. David Greenwood

Abstract

This study aims to identify the presence of a financial bubble and analyse its lifespan. This investigation into bubbles encompasses a range of asset classes and considers the distance between instantaneous price and fundamental value. By modelling time series stochastically, a stationary, mean reverting process is formulated by use of the Ornstein-Uhlenbeck (OU) equation. The Savitzky-Golay convolution filter removes white noise from this process; meanwhile the exponential moving average of the time series acts as a proxy for fundamental value. Here it is shown that no conclusive relationship between peak distance from fundamental value and mean reversion time was found. However, autocorrelation plots demonstrated evidence that markets trend in the short run and revert to the mean in the long run. Although the model does not determine whether we are in a bubble, the findings can be combined with market indicators to provide a holistic approach that can ascertain whether markets might be in a boom.

Contents

1 Introduction **1**

2 Literature Review **1**

3 Methods **2**
- 3.1 Proxy to the Fundamental Value . . . 2
- 3.2 Model Specification . . . 3
- 3.3 Testing for Stationarity . . . 3
- 3.4 Filtering White Noise . . . 4
- 3.5 Measuring Peak Height and Mean Reversion Time . . . 4
- 3.6 Skewness and Kurtosis Statistics . . . 5
- 3.7 Autocorrelation . . . 5
- 3.8 Determining Parameters . . . 5
  - 3.8.1 Savitzky-Golay Filter . . . 6
  - 3.8.2 Conditions for the Savitzky-Golay Filter . . . 6
  - 3.8.3 Other Key Parameters . . . 7

4 Results and Analysis **7**
- 4.1 Commodities . . . 7
  - 4.1.1 Gold . . . 8
  - 4.1.2 Crude Oil . . . 8
- 4.2 Equity . . . 8
  - 4.2.1 DJIA . . . 9
  - 4.2.2 S&P500 . . . 9
- 4.3 Gross Domestic Product . . . 9

5 Discussion and Conclusion **10**
- 5.1 Comments on the Model . . . 10
- 5.2 The Effectiveness and Stability of Markets . . . 11
- 5.3 Further Steps . . . 12

A Methods **15**
- A.1 Proxy to the Fundamental Value . . . 15
- A.2 Model Specification/Testing for Stationarity . . . 16
- A.3 Filtering White Noise . . . 17
- A.4 Measuring Peak Height and Mean Reversion Time . . . 18
- A.5 Determining Parameters . . . 18

B Results **22**
- B.1 Dataset Statistics . . . 22
- B.2 Commodities . . . 22
  - B.2.1 Gold . . . 22
  - B.2.2 Crude Oil . . . 26
- B.3 Equity . . . 28
  - B.3.1 DJIA . . . 28
  - B.3.2 S&P500 . . . 31
- B.4 GDP . . . 34

1 Introduction

The Efficient Market Hypothesis (EMH) states that it is impossible to consistently beat the market since share prices reflect all available information [1]. This implies that higher returns can only be obtained by chance or by purchasing riskier investments. However, markets can behave unpredictably when traders act irrationally and invest emotionally. This is where humans do not always act as 'rational agents' and the 'wisdom of crowds' becomes distracted. Misguided investing can cause the formation of a bubble, where investors put a large demand on an asset, causing the price to increase far above what the asset is actually worth. When a bubble 'pops', it causes a crash in the market, which affects all market participants and incurs significant losses. Being able to foresee these crashes enables investors to know when to sell their stocks and avoid these large losses by setting up a trading strategy. Timing is inherently crucial in financial markets.

As mentioned above, the EMH dictates that the asset's current price is reflected by 'all available information' and hence should be approximately equal to its intrinsic or fundamental value. Therefore, it theorises that bubbles are rare, or at least short-lived. Depending on the extent of this information, there are three forms of the EMH. The strong form encompasses all private and public information, the semi-strong form considers only public information whilst the weak form simply relates to historical price and returns. Current evidence suggests that the market tends to be positively autocorrelated in the short run and negatively autocorrelated in the long run [2].

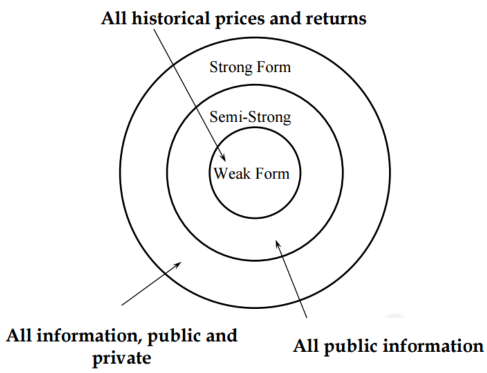

Figure 1: Forms of the Efficient Market Hypothesis [3].

The aim of this project is to examine whether market crashes can be detected by analysing the distance from the instantaneous price to the fundamental value of the asset. Consequently, this project attempts to evaluate the real-world relevance of the EMH. To achieve this aim, a series of objectives are laid out:

1. Developing a proxy to the unknown fundamental value,
2. Measuring the deviation from this proxy to real price over time and
3. Assessing the properties of how this distance varies over time.

If the distance above the fundamental value becomes large over a sustained period of time, this would indicate it is likely there is a bubble present and the probability of collapse is high. This is because conventional market wisdom tells us that an asset cannot price itself indefinitely above its rate of returns and eventually the market will correct itself. Likewise, the opposite is true; an asset priced below its fundamental value should eventually see a price increase and accordingly is an investment opportunity. This report mainly focuses on the former situation where assets are overpriced.

2 Literature Review

Asset pricing dictated by financial markets plays a fundamental role in society. For example, company valuations are based on respective stock prices set by the market. This considered, there is a substantial amount of existing research into every facet of the financial market. Therefore, the review of existing literature must be aimed into attempts at modelling bubbles and crashes. Specifically, looking into market trends, how proxies for fundamental values are developed, and the use of stochastic differential equations to model the evolution of price residuals. This is done to gauge the relevance of existing techniques and solutions to any models considered, and what sort of results to expect.

The paper by Bouchaud et al. backs up claims that markets follow a general trend over the medium-term and eventually revert back to a long run average [2]. In addition, prices tend to be within a factor of 2 and take several years to equilibrate. A study by Branch and Evans investigates how to generate bubbles and crashes as a response to fundamental shocks [4]. It considers the use of recursive learning algorithms based on least-squares method to create a simple linear asset pricing model. Another paper by Hui and Gu employs the use of an alternative proxy based on household income to model the fundamental value of housing [5]. This is used to estimate and analyse the known housing bubble over 2004-08.

A thesis by Hillebrand explores the 'distance' between the fundamental proxies and current price via the use of different mean reversion models [6]. Furthermore, a study conducted by Barndorff-Nielsen and Shephard considers and expands upon the popular OU mean reversion models for financial assets [7]. The paper by Yu investigates the issue of estimation bias for the mean reversion parameter in continuous time models [8]. It concluded that the quality of bias approximation is improved when a bias formula with a nonlinear correcting term is considered. However, this only applies to the case of the univariate OU process. Another thesis paper by Thierfielder investigates the use of different OU processes to simulate interest rate models [9]. It found a bivariate model gives the best representation of an interest rate model.

A different approach to modelling asset price bubbles is investigated through the use of martingales in a paper by Kardaras et al. [10]. Specifically, the effect of asset price bubbles on the pricing of derivatives is examined. However, the strict analytical approach of martingales is not preferred to the flexible, numerical approximations of the stochastic differential equation approach in any of this study's models. This raises the dilemma of prioritising accuracy and analytics against computational capability.

3 Methods

3.1 Proxy to the Fundamental Value

The moving average of the price (or returns) time series is used as a proxy to the recent or long-term trends of the fundamental value (i.e. the unknown market value). Suppose that $\{Z_t\}$ is a purely random process with zero mean and variance σ_Z^2. Then a process $\{X_t\}$ is said to be a moving average process of order q,

$$X_t = \mu + \epsilon_t + p_1\epsilon_{t-1} + ... + p_q\epsilon_{t-q}, \tag{1}$$

where μ is the mean of the series, the $p_1, ..., p_q$ terms are the parameters of the model, and the $\epsilon_t, \epsilon_{t-1}..., \epsilon_{t-q}$ terms are the white noise error terms or random shocks [11]. A simple moving average (SMA) is calculated by adding the closing prices for a number of time periods and then dividing this total by the same number of periods. In Eq. 1, the parameters $p_1, ..., p_q$ are therefore the closing prices and q is the 'window size' parameter (number of periods), which leads to the equation

$$X_t = \frac{p_1 + p_2 + ... + p_q}{q}[12]. \tag{2}$$

The purpose of the SMA is to reduce noise in a time series, so trends become clearer. An example is set up in Appendix A with a simple time series of the Dow Jones Index ranging from 2003 to the start of 2019 (a smaller time period of the larger dataset) in Fig. 2a and an example with two SMAs in Fig. 2b, showing short-term and long-term trends. The SMA is used to quickly identify uptrends or downtrends in an asset price time series. However, SMA lags the original price time series so that trends can only be seen after a period of roughly $q/2$ days [13]. Also, the SMA relies only on historical data. If the EMH holds, then using only historical data will tell us nothing about the future direction of asset prices [12].

To reduce the lag on the time series, an improved moving average process called the exponential moving average (EMA) is used as an alternative. The improvement can be seen when comparing Fig. 2b to Fig. 2c. This is similar to the SMA except instead of assigning equal weighting to all values it gives more weighting to recent price data, hence recent trends are more noticeable. It does this by introducing a decay/smoothing parameter, α, where $0 < \alpha < 1$. If the window size is given by q, then

$$\alpha = \frac{2}{q+1}, \tag{3}$$

and the process is then

$$X_t = (1-\alpha)X_{t-1} + \alpha p_t, \tag{4}$$

where $X_{t_0} = p_{t_0}$ and p_t is the price at time t [13].

Considering the deviation from the time series p_t to its EMA X_t, the residual at time t is thus given by

$$\epsilon_t = X_t - p_t. \tag{5}$$

3.2 Model Specification

The evolution of the residual from Eq. 5 is described by a stochastic differential equation (SDE) that is believed to be stationary and mean reverting. A statistic test to prove stationarity is provided in Section 3.3. To be stationary means that the residual is a stochastic process whose parameters are the mean and variance, which do not change over time. Furthermore, mean reversion is based on the idea that asset prices return to the long run mean of the time series after unexpected events called shocks which are a hypothesised cause of deviations. Thus, this stochastic process is described by a mean reverting OU process, given by the equation

$$d\epsilon_t = \theta(\mu - \epsilon_t)dt + \sigma dW_t \tag{6}$$

where ϵ_t is the residual series, μ is the mean value supported by characteristics of the market in question, σ is the volatility caused by shocks (i.e. the 'moving' standard deviation), θ is the rate at which the shocks dissipate (rate of mean reversion), and dW_t is a Wiener process [9]. The Wiener process, or Brownian motion, as used in Eq. 6, is described as a continuous-time stochastic process W_t for $t \geq 0$ with the following properties [9]:

1. $W_0 = 0$,
2. The function $t - W_t$ is almost surely continuous everywhere and
3. W_t has independently normally distributed increments i.e. $W_t - W_s \sim N(0, t-s)$ for $0 \leq s < t$.

The OU process essentially removes any trends in the time series by converting it into a stationary process. The coefficient θ controls the forcing of the residual ϵ_t back towards the mean μ. This is related to the smoothing parameter α from Eq. 4 by the simple equation

$$\theta = 1 - \alpha. \tag{7}$$

An example of a graph showing the process can be seen in Fig. 3.

3.3 Testing for Stationarity

The OU process is a unique SDE as it is a Markov process, a Gaussian process and a mean reverting process by defintion. The OU is not always stationary, which would mean the process would drift away from the mean over time. The line where the value of $d\epsilon_t$ is equal to zero defines the proxy to the fundamental value, but only if the process is stationary. This can be seen in the example given by Fig. 3, where the process oscillates about the zero line.

To test for stationarity, the Augmented Dickey-Fuller (ADF) test is introduced [14]. The purpose of an ADF test is to test for the presence of a unit root (which indicates non-stationarity) in an autoregressive time series by testing the null hypothesis. The alternative hypothesis is that the process is stationary. The more negative the test statistic, the stronger the rejection of the hypothesis that there is a unit root at a specific confidence level, and thus provides a strong belief that the process is stationary. An SDE has a unit root if 1 is a root of the characteristic equation.

The more basic Dickey-Fuller Ordinary Least Squares (OLS) regression of Δp_t (asset price) against the lagged p_{t-1} gives the regression equation

$$\Delta p_t = \alpha + \beta t + \gamma p_{t-1} + e_t, \tag{8}$$

where α is a drift constant, β is the coefficient on a time trend and e_t are the residuals of the OU process [14]. If there is evidence of autocorrelation in the residuals (i.e. there is a degree of similarity between a given time series and a lagged version of itself over successive time intervals) then ADF OLS regression must be used instead of the original OLS regression. The Durbin-Watson test statistic is used to give evidence of positive autocorrelation, given by

$$\frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=2}^{T} e_t^2}, \tag{9}$$

where T is the total time period [15]. If this is close to zero then there is evidence of autocorrelation. The ADF statistic uses the linear lag model of order q for the time series

$$\Delta p_t = \alpha + \beta t + \gamma p_{t-1} + \delta_1 \Delta p_{t-1} + ... + \delta_{q-1} \Delta p_{t-q+1} + e_t, \tag{10}$$

where $\Delta p_t = p(t) - p(t-1)$. The test is carried out with the null hypothesis $\gamma = 0$ against the alternative hypothesis $\gamma < 0$. Also, the coefficients are set to $\alpha = 0$ and $\beta = 0$ to model a random walk without drift. The value of the test statistic is then

$$ADF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}, \tag{11}$$

where $SE(\hat{\gamma})$ denotes the standard error of the estimated parameter. Critical values for the ADF test statistic are given by:

- -2.567 at the 10% significance level
- -2.862 at the 5% significance level
- -3.432 at the 1% significance level

3.4 Filtering White Noise

The time series data contains white noise which can distract from the main trends of the data and therefore should be removed before trend analysis can take place. For this reason, a convolution filter is applied to the time series to smooth the data. The Savitzky-Golay filter is chosen because it increases the signal-to-noise ratio while keeping distortion to a minimum [16]. Comparing Fig. 4a to Fig. 4b, it can be seen that if the EMA is used to smooth the results once more, there is a large degree of lag on the data. It is noted that using a Savitzky-Golay filter generally maintains the correct location and height of the peaks (see Fig. 4b).

This filter takes polynomials of n^{th} order and uses them to fit to the data points in the time series. If n is small, then there is not sufficient smoothing of the data and if n is too large then too much data may be lost as a result of overfitting. Sixth-order polynomials fit the data to a desirable smoothness and therefore $n = 6$ is chosen. The number of convolution coefficients is also a parameter of the filter which must be balanced to fit the data in unison with the polynomial order. Here 201 convolution coefficients is chosen. The Savitzky-Golay filter is explained in more depth in Section 3.8.

3.5 Measuring Peak Height and Mean Reversion Time

The smoothed version of the OU process is an undulated sinusoid with peaks of different heights and varying oscillation times. In order to analyse the characteristics of the OU process, the height of each peak is calculated and compared with the corresponding mean reversion time. The mean reversion time is found by computing the difference between the time at which the peak occurs and the next time at which the process crosses the mean (or zero) line. The data points for the peak heights and the mean reversion time are then plotted to visualise and investigate the relationship between them. Data for peak heights and mean reversion times can then be collected for each asset to be considered. Graphs in the results are referred to as HT graphs, meaning Height versus Time.

3.6 Skewness and Kurtosis Statistics

Skewness is a measure of how symmetrical or non-symmetrical a dataset is. If skewness is between -0.5 and 0.5 the datapoints are fairly symmetrical. If the skewness is less than -1 or greater than 1 the datapoints are highly skewed. Kurtosis is a measure of the combined weight of the tails of a dataset relative to the rest of its distribution. The kurtosis decreases as the tails become lighter and increases as the tails become heavier. There should not be much emphasis on these statistics as they depend highly on the sample size and can be misleading [17]. However, measuring the skewness and kurtosis of the peak heights of the data can provide some insight into the characteristics of mean reversion when bubbles arise.

3.7 Autocorrelation

Evidence of autocorrelation is provided by the Durbin-Watson statistic. The autocorrelation function is now derived for the purpose of detecting trends in the data [18]. Given measurements $Y_1, Y_2, ..., Y_N$ (i.e. the values retrieved from the OU process in Eq. 6 with N being the total time period, after filtering white noise) at time $T_1, T_2, ..., T_N$, an autocorrelation function of lag k is

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})}, \tag{12}$$

where $\bar{Y}$ is the mean of $Y_1, ..., Y_N$. To produce an autocorrelation plot, the horizontal axis as time lag τ ($\tau = 1, 2, 3, ...$) is plotted against the vertical axis as the autocorrelation coefficient (a value between -1 and 1),

$$R_\tau = \frac{C_\tau}{C_0}, \tag{13}$$

where C_τ is the autocovariance function [19],

$$C_\tau = \frac{1}{N}\sum_{t=1}^{N-\tau}(Y_t - \bar{Y})(Y_{t+\tau} - \bar{Y}), \tag{14}$$

and C_0 is the variance function,

$$C_0 = \frac{\sum_{t=1}^{N}(Y_t - \bar{Y})^2}{N}. \tag{15}$$

The autocorrelation plots show the correlation of the series with itself lagged by τ amount of days. A coefficient of -1 shows a perfect negative correlation and 1 shows a perfect positive correlation. A coefficient close to zero means that there is no relationship between the series and its lag. For the series to be random as the EMH suggests the autocorrelation should be near zero for all time-lag separations. Negative autocorrection indicates higher than average returns will be followed by lower than average returns and shows that the series is mean reverting. Mean reversion in stock prices may reflect market inefficiency. However, stock price mean reversion does not necessarily contradict the EMH [20].

3.8 Determining Parameters

The numerical simulations run on the model depend on certain varying key parameters. The reason for this variation is to identify an ideal combination of values for these parameters that are capable of consistently providing an accurate representation of the time series and its respective properties over all datasets.

3.8.1 Savitzky-Golay Filter

A key feature of the model is the Savitzky-Golay convolution filter which smooths out the detrended OU process. Any results obtained are derived from this smoothed time series. The key parameters of interest relating to the Savitzky-Golay filter are the number of convolution points, m, and the order of the moving polynomial, n. The number of convolution points is a subset of datapoints from a dataset used to fit a moving polynomial of n^{th} order. One way to think about this is to consider the subset of sample datapoints as an analysis interval over which a fitting polynomial is calculated. The analysis interval calculates the best fit to the datapoints in the interval for the n^{th} order polynomial. Every analysis interval has a central point at which the polynomial over the interval is evaluated; the value of the evaluated polynomial providing the smoothed approximation to that specific central point.

The approximation is the mean-squared approximation error for the set of datapoints in the analysis interval. Each side of the central point has b data points, referred to as the 'half-width' of the analysis interval. The number of convolution points over any analysis interval is $m = 2b + 1$ and therefore m must always be odd. A smoothed approximation to the next datapoint is obtained by shifting the analysis interval to the right by one sample and repeating the polynomial fitting and subsequent evaluation at the central point. However, the use of the analysis interval means that on each side of the dataset, there will be b points on which the calculations cannot be performed. A way around this is to artificially extend the data by adding copies of the first b points at the start and last b points at the end in reverse order. This allows the smoothing approximation to be calculated for the first and last b points of the dataset.

3.8.2 Conditions for the Savitzky-Golay Filter

A condition for uniqueness of solutions requires at least as many convolution points as coefficients in the polynomial approximation [21]. That is, the condition $m \geq n$ is required. If n and m are both large and b is close to n, the least-squares approximation becomes ill-conditioned. Hence, any parameters chosen must satisfy these conditions. The notation for a combination of m convolution points and fitting polynomial of order n is (m,n). A justification of the parameter choices follows:

1. The number of convolution points chosen will be applied as the general parameter value between datasets. Doing so keeps the key results for datasets consistent in in the hope of identifying any general trends. The Dow Jones example in Appendix A is the dataset chosen to optimise the number of convolution points in order to sufficiently smooth the detrended OU process. The plots of the convolution filter fitted to the OU process for parameter combinations (201,5) and (1001,5) are displayed in Fig. 6a and Fig. 6b. Here, the order of polynomial was chosen by trial and error. In order to gauge the best smoothing, plots of the fitted convolution filter are visually inspected. The best fit is achieved with a choice of 201 convolution points. Lower than this at 101 convolution points, the curve doesn't perform a sufficient smoothing. Considering the choice of 201 convolution points in Fig. 6a, we see that a sufficient smoothing of the OU process is possible while still providing a good representation of the regions of high volatility. The case of (1001,5) in Fig. 6b illustrates the effect of over-smoothing and there is subsequently little representation of the interesting regions of volatility. Hence, m is set at 201, giving $b = 100$.

2. The order of moving polynomial will also be applied as the general parameter value between datasets. As mentioned before, $m \geq n$ and large n and m with b close to n will result in ill-conditioning of the method. As m is large, n must be small and hence n is kept smaller than 10; n must also be at least 2 because otherwise the polynomials would be straight lines. Plots of convolution filter fitting to the detrended OU process are visually inspected to gauge the best smoothing fit as before. The value of the convolution point parameter is set at the optimal general value of 201. The polynomials of orders 2, 4, 6, 8 are plotted in Fig. 7a, 7b, 4b and 7c. The fit of order 2 to the time series shown in Fig. 7a provides a competent representation of the high variance in volatility but does not capture the majority of the lower shocks from fundamental values. It also fails to demonstrate the differences in height. A polynomial of order 4 is considered in Fig. 7b. The filtered time series provides a better representation of the high variances in volatility with taller peaks and deeper crashes. The polynomial of order 6

provide a more desirable fit for the filter to the OU process in Fig. 4b. Areas of high volatility are better fitted by the filter, representing peaks and nadirs more accurately. A superior fit is also achieved for timeframes of lower volatility. The polynomials of orders 8 and above are subject to erroneous behaviour as shown in Fig. 7c. Accordingly, a polynomial of order 6 is chosen.

3.8.3 Other Key Parameters

1. The short run for the EMA is the numerical value for the dataset size divided into an arbitrary number of equally sized timeframes. When calculating the fundamental value through the EMA, the data is consistent as each point has been calculated with an equal window size and hence an equal number of points, so the contribution of an equal number of trading prices, with recent prices holding more weight, is considered. The dividing parameter was selected to be 50 as the dataset is partitioned into different windows of interest. If the number for the dividing parameter is greater than 50, it partitions the dataset into too many windows and runs the risk of separating regions of interesting behaviour; the rise and crash of a bubble may be split into regions of rise and crash separately for example. If this number is lower than 50, the window sizes will be too large and consider the behaviour of prices that may not reflect current behaviour trends.

2. The border functionality allows us to restrict further analysis exclusively to price peaks above a certain height threshold. This means only the largest peaks in a given window of interest are considered, i.e. peaks most indicative of a bubble. Any trends and interesting behaviour can then be examined for these peaks in isolation. This border height is varied according to the volatility present in the window under analysis.

4 Results and Analysis

This project looks to test the model described in Section 3 against a range of asset classes sourced from historical datasets. These asset classes include commodities, equity, currencies and macroeconomic measures. The latter includes data on a variety of measures that could potentially provide interesting socio-economic insights. However, for the purposes of this analysis the study focuses on gold and oil; the Dow Jones Industrial Average (DJIA) Index and the Standard and Poor's 500 (S&P500) Index; as well as the United States' annual Gross Domestic Product. A list of start dates and closing frequencies (data recorded at daily, monthly or quarterly intervals) for the datasets considered is provided in Tab. 1. Currencies are excluded from the analysis for reasons pertaining to the nature of their price dynamics. Aside from the multitude of monetary variables that influence foreign exchanges, they do not always operate in a free market. For instance, the European Exchange Rate Mechanism was put in place from 1979 to 1999 as a system that reduced exchange rate variability amongst countries in the European Economic Community. Data for the DJIA and S&P500 are taken from [22], and all other data used is from [23]. A bubble is identified if its distance from the fundamental value on the OU process is above a specified boundary condition.

4.1 Commodities

The dataset contains price data on two commodities: gold and crude oil, which span back to April 1968 and January 1946 respectively. Of the two, the price dynamics of oil are more complex. Oil is in such high demand in the 21st century and is consumed at such an unsustainable rate that it lends itself vulnerable to demand and supply shifts. This is before considering the external geopolitical factors that influence the price of oil, most pertinently ongoing affairs in the Middle East or Venezuela and current relations between the United States and Russia. Additionally, more and more modernised economies are actively seeking alternative energy solutions while oil resources continue to diminish. On the other hand, conventional economic wisdom regards gold as a 'safe haven', whereby in times of economic uncertainty investors direct their capital from riskier markets to gold reserves. Of course, the opposite is often true in times of greater economic confidence. Nevertheless, the theory supposes that the fundamental value of gold will rarely change. This idea suggests that the above models will be quite able to measure trends away from fundamental value and detect the occurrence of bubbles.

4.1.1 Gold

As detailed in Section 3, a proxy to the fundamental value of an asset class is developed by considering the EMA. This is displayed in Fig. 8a with a short run parameter of 265 trading days. A detrended time series of the residuals from proxy to price is then plotted alongside the convolution filter applied for smoothing purposes. Fig. 8b shows this for Savitzsky-Golay parameters (201,6). The peaks above the zero line for the convolution filter (201,6) are shown in Fig. 8c. The HT graphs for gold (101,6) and (201,6) showing the height of peaks versus time to next collapse are plotted for border cropping 100 in Fig. 8e and 8f. The effects of an increased number of smoothing coefficients is displayed in the HT graphs where a choice of the parameters (101,6) results in peaks of a higher height than in the case of (201,6). This is found to be the case for all HT graphs per dataset. The corresponding Durbin-Watson, ADF, skewness and kurtosis statistics are shown in Tab. 2.

From Fig. 8c, there are three significant peaks above a distance of 150 units from the fundamental value which have a high magnitude of collapse. These occur roughly at day 3000, day 11300 and day 12250, which correspond to dates 10/01/1979, 07/25/2011 and 03/16/2015 respectively. Notably, demand for gold reserves surged at the same time the global economy suffered in the late 2000s, affirming its 'safe haven' status. In 2013, the price of gold dropped dramatically for reasons subject to much debate. Nonetheless, one hypothesis suggests that this crash developed as confidence began to return to the wider economic markets.

From the autocorrelation plot shown in Fig. 8f, the correlation coefficient is positive for $\tau \leq 500$ and then is negative between 500 and 1125 days. This indicates that the process is negatively correlated for this window of lag times and should mean revert. After 1125 days the process oscillates between positive and negative coefficients as τ increases.

4.1.2 Crude Oil

For the other commodity, crude oil, the same method is followed as in the case of gold. A proxy to the fundamental value of an asset class is first developed by an EMA. This is displayed in Fig. 9a for the short run parameter of 17 months. We note here that this relates to a total of 204 days which is similar to the case of gold. A detrended time series of the price residuals from fundamental proxy to price is then plotted alongside the convolution filter applied to smooth the time series. This is shown in Fig. 9b for Savitzky-Golay parameters (201,6). The peaks above the zero line for convolution filter (201,6) are shown in Fig. 9c. The HT graph for crude oil (201,6) showing the height of peaks versus time to next collapse is plotted for border cropping parameter 0 in Fig. 9d. A smaller value for the border cropping parameter is chosen to exhibit interesting behaviour as heights of peaks are lower than in the case of gold. The Durbin-Watson, ADF, skewness and kurtosis statistics for crude oil are shown in Tab. 2.

From Fig. 9c, there are two significant peaks that have a high magnitude of collapse, which occur at roughly month 410 and month 725, corresponding to dates 03/01/1980 and 06/01/2006 respectively.

The autocorrelation plot for oil, Fig. 9e is much more chaotic than that of the others which reflects the volatility of oil price. The coefficient is closer to zero at shorter lag times although the frequency at which the correlation changes from positive to negative remains roughly the same for all lag times. This shows that the oil time series has frequently occurring periods of mean reverting behaviour.

4.2 Equity

Equity is one's degree of ownership in any asset after subtracting all debts or liabilities associated with that asset. There are two major widely followed American stock market indices: the DJIA Index and the S&P500 Index. The difference between the two is that the DJIA is price-weighted, meaning the sum of the component stock prices are divided by a divisor, so is disproportionately affected by changes in the larger companies it is comprised of [24]. On the other hand, the S&P500 has no such weighting system. Equity prices contain what is known as a rational bubble if investors are willing to pay more for the stock than they know is justified, because they expect to be able to sell it at an even higher price in the future, making the current high price an equilibrium price.

4.2.1 DJIA

The fundamental proxy is developed by an EMA in Fig. 10a with a short run parameter of 297 days. The detrended time series of the price residuals is shown in Fig. 10b alongside the convolution filter, shown in isolation in Fig. 10c with the detected peaks. The HT graph for DJIA (201,6) showing the height of peaks is shown in Fig. 10d with border size 0 and Fig. 10e with border size 1000. The Durbin-Watson, ADF, skewness and kurtosis statistics for the DJIA are shown in Tab. 2.

From the smoothed version of the OU process shown in Fig. 10c, there are three significant peaks above a distance of 2000 units from the fundamental value which have a high magnitude of collapse. These occur roughly at day 10000, day 12000 and day 14500, which correspond to dates 9/21/1999, 9/5/2007 and 8/9/2017. Bubbles that were most likely to have occurred around these dates are the Russian Insolvency Crisis of 1998 and the US Subprime Mortgage Crisis of 2007-2008.

In the autocorrelation plot for DJIA, Fig. 10f, the coefficient is greater than 1 apart from the period 1125-1375. This suggests that the time series will trend up until around 1250 days at which point the series should mean revert. A similar pattern can be seen in the plot for S&P500 (see Fig. 11f) where the coefficient is positive until 1000 days and stays negative up to 1400 days.

4.2.2 S&P500

For the other equity index, S&P500, the same method used for analysing DJIA is considered. The EMA is shown in Fig. 11a with short run parameter of 347 days. The detrended OU process is shown in Fig. 11b, along with the smoothed results, shown in isolation in Fig. 11c. The HT graphs with border size 0 and 100 are shown in Fig. 11d and Fig. 11e respectively.

From the smoothed version of the OU process shown in Fig. 11c, there are three significant peaks above a distance of 200 units from the fundamental value which have a high magnitude of collapse. These occur roughly at day 12450, day 14000 and day 16000, which correspond to dates 28/06/1999, 25/08/2005 and 07/08/2013 respectively. Hence, the bubbles that were most likely to have occurred at these dates are the Dot Com bubble in the late 90s and the US housing bubble in early/mid 2000s.

The S&P500 autocorrelation plot shows a similar pattern to the plot for DJIA where the coefficient is positive until 1000 days and stays negative up to 1400 days. This indicates that the process is mean reverting for this window of lag time values.

4.3 Gross Domestic Product

Gross Domestic Product (GDP) is a monetary measure of the market value of all the final goods or services of a country - including personal consumption, government spending and business investment - within a specific time period. This time period is usually set to 12 months. It is a strong economic indicator of a nation's overall economic activity; generally as the GDP of a country grows, the standard of living also improves. Thus, GDP provides a strong assessment of the rate of growth of the economy, taking the impact of inflation into account.

The data for GDP is relatively small due to its low frequency, so the EMA seen in Fig. 12a has a short run parameter of 5 quarters. The detrended results are shown in Fig. 12b. The process is proved stationary by its ADF statistic, although some small level of drift seems to be present, as seen in Fig. 12c. The HT graph with border 0 is shown in Fig. 12d, which is has the highest degree of randomness out of all HT graphs.

Significant peaks from the GDP OU process as seen in Fig. 12c roughly occur at quarters 100, 125 and 225, corresponding to the second quarter of 1972, the third quarter of 1978 and the third quarter of 2003 respectively. The latter date could be an indicator of the Dot Com bubble, which has also been identified in the S&P500 results.

The autocorrelation plot for GDP shows that the coefficient is positive for $\tau \leq 900$ days and thereafter alternates between positive and negative as τ increases. This is similar to the autocorrelation plot for oil in that it has regularly occurring periods of mean reverting behaviour.

5 Discussion and Conclusion

The statistics shown in Tab. 2 reveal Durbin-Watson statistics that are close to zero for every OU process simulated, meaning there is evidence of positive autocorrelation in every result. Hence, the ADF statistics are valid and are all more negative than the 1% critical value of -3.432, so there is also significant evidence of stationarity.

The relationship between the peak distance from the fundamental value and the mean reversion time appears fairly random from analysis of the scatter plots. There are similarities for the plots of all the asset classes. The lower peak heights have a greater variance in the mean reversion time. This means that for relatively small distances above the fundamental value, it is difficult to determine whether there will be a mean reversion shortly or if the distance will keep growing and mean revert much later. It can also be seen that while lower heights can sometimes have long mean reversion times of over 1500 days as well as shorter times, the higher peaks tend to only have mean reversion times of a shorter nature; less than 500 days. This implies that for a price that is high above the fundamental value, where a bubble may occur will have a fast rate of mean reversion. All of the asset classes have a skewness of above 1 with the exception of GDP. This suggests that the peak heights are positively skewed, and in turn this implies that prices that are considerably high above the fundamental value will have a fast rate of mean reversion, thus have a higher probability of collapse. However, this fact is only relevant for data with a daily closing frequency (large sample size), as the skewness is drastically dependent on the size of the dataset. The kurtosis statistics don't show any overall trend so can be disregarded.

The removal of noise from the OU series is important since it avoids the results picking up on false signals and allows for a clearer picture of the overall trend. However, parts of the original data that has been filtered out could potentially contain useful information that is not considered when analysing the peak heights and mean reversion times. The convolution filter parameters were carefully chosen according to the best fit to the data which means that these unconsidered fluctuations in the time series are kept to a minimum. It should be noted that the parameters were chosen according to the best fit for the Dow Jones data and these parameters may not necessarily be the best fit for the other data sets. Since the time series are relatively similar in nature it is assumed that the parameters are a sufficient fit for all data sets used in the methods.

The autocorrelation plot results provide evidence that markets will trend in the short-term and revert to the mean in the long-term. This shows that it may be possible to beat the market for a short time however investors cannot consistently beat the market in the long-term. The fundamental value will eventually be returned to; however, short-term market trends are much harder to predict. This advocates for the adoption of alternative financial models such as agent-based models which can better reflect the irrationality of human investors.

5.1 Comments on the Model

The model reliably tackles the first two objectives in creating a proxy for fundamental value and measuring the distance to market price which provides a strong platform from which the relationship between the two can be assessed. Initially the fundamental value was established with a SMA which was improved by weighting more recent price data. After techniques to streamline and substantiate the model have been applied, including the removal of noise from the data and confirming the stationarity of the proxy, the residual values of the market price against the proxy were mapped out to offer a way to understand what correlation exists. This model not only allows historical bubbles to be labelled with a decent level of accuracy but also presents a way to analyse their characteristics.

The EMA that is used in the model acts as a reliable proxy to approximate the fundamental value, which remains an unknown quantity. However, the EMA lags the data to some degree so this can carry over when producing the OU processes. Making the window length as small as possible has minimised this effect to some degree, although the model won't be able to predict the exact dates of crashes because of this. The effect of minor variations window size are downplayed by a property of the EMA, which gives more weight to information obtained from recent prices.

With the OU process it is easy to see large deviations from the fundamental value given that the mean is zero, but it comes with some limitations. Although it detrends the residuals, the white noise terms are difficult to remove. These are natural elements of stock markets as a result of shocks, referring to information that confuses genuine underlying trends. Removing these terms is a difficult task, but the use of the Savitzky-Golay filter is a useful method of smoothing the data to notice the trends more easily. This technique is commonly known as kernel smoothing and here is used instead of the EMA as it is shown to have a better fit given the correct parameters. The use of the Savitzky-Golay filter forms an integral part of the model; as such, other filters could be considered. One other filter that could be considered is the Hodrick-Prescott filter used in to model bubbles and crashes [4] .

Collectively, the results of the investigation show that the instantaneous price does not always reflect the fundamental value of the asset. This indicates that the EMH does not always hold however the findings prove that long-term the price will mean revert back to its intrinsic price. The methods used in this project to investigate the distance form the fundamental value provided little insight into the formation of bubbles however further methods could be used to gain more knowledge about market bubbles and crashes as further discussed in Section 5.3. Current theories regarding short-term and long-term behaviour of markets are supported by the findings of this report as the results provide further evidence that the market will trend on the short-term and mean revert on the long-term.

5.2 The Effectiveness and Stability of Markets

It has been observed that price fluctuations are not simply a result of new information but market participants' reactions to previous price fluctuations. Investors are not merely concerned with ascertaining the long-term fundamental value of an asset but instead what others are willing to pay in the short term. As a result, participants endeavour to predict how others might react to fluctuations and buy or sell accordingly. This points to the desire for investors to profit from a bubble by riding along with the appreciating prices before selling their assets before the subsequent crash. The irrational nature of this behaviour can be justified as a prudent decision in each investor's local environment but a global imprudence for the entire market. In addition to the recent dubiety associated with how strongly EMH governs the financial market is the manner in which the financial landscape has developed. Before 1970 financial crises were relatively rare due to the severe regulations on banks, insurance companies and other financial institutions. These regulations were stripped back and an immense expansion in the value of globally traded financial assets between 1980 and 2010 followed, referred to as the finalisation of the market. More and more complex financial securities were introduced, each adding its own uncertainty, leading to the stock of financial assets increasing from 71% of world GDP to 226% in 2012 [25]. The innovation of financial products paired with the ease at which they can now be traded through recent technological advancements has accelerated the behaviour and volatility of markets.

Finally, markets have also been affected by the evolution in investment culture as the industry and the competition has grown. Investment indicators that were once sworn by have been allowed to enter what was considered unhealthy territory. Capital Asset Pricing Model (CAPM) is heavily used in fundamental price analysis and depends on such indicators like Price-to-Earning ratio or dividend yield. These estimations have reached levels significantly beyond historical norms during the bubbles of recent years. This evolution has brought heuristic techniques to investment strategy as participants rely on their own experience and knowledge alongside these indicators to gain an edge, accumulating yet more potential for irrational behaviour by market participants. Low interest rates stimulate market participants to invest their money outside of savings accounts which will deliver poor returns and allow companies to borrow money at an equally cheap rate. This increased liquidity of available money, termed credit expansion, accelerates price volatility as the number of transactions rises. When these transactions rise enough to warrant an increase in demand and hence and increase in price.

Although the model does not provide any definitive findings to suggest we are in a bubble, a more holistic approach which includes other market indicators like interest rates and market liquidity can provide a confident assessment of whether market price and fundamental values have a divergent or convergent nature.

5.3 Further Steps

The common approach to identifying bubbles in economics is to identify when the market value differs from its fundamental value. A different approach that would be interesting to explore in further work is the concept of strict martingale measures that arise naturally when modelling financial bubbles [26]. Martingale measures are stochastic processes for which any given variable has an equal conditional expectation to the next variable in the sequence. In other words, a martingale is the mathematical model of a fair game. In a finite horizon case for a Markov decision process, findings show that a bubble is present if and only if the fundamental price is a strict martingale and the original stock asset is also a martingale. With this further condition that defines a bubble, it may be possible to combine the current model and martingale measures to improve the accuracy when identifying bubbles.

The sensitivity of the market is known to be larger for some assets, especially in the short term, and might be better modelled with an increased probability of larger fluctuations. This could be explored by substituting the Weiner Process that uses random walks which are normally distributed for the Cauchy Process that uses random jumps.

Of course, alternative models that attempt to approximate fundamental value exist. In the case of equities, a common approach that tackles this sort of fundamental analysis is to look at a stock's dividend returns. This school of thought supposes that an investor's confidence in a stock is underpinned by a rate of return on their capital. One such model is the so-called Dividend Discount Model (DDM) [27]. The DDM predicts a company's stock price based on the theory that its present day price is worth the sum of all of its future dividend payments (discounted to present value). However, the DDM was not practical for the purposes of this study as fails to generalise over different asset classes and relies on dividends data that is not easily available.

The introduction of irrationality into our model would be possible by integrating a variable to account for the herding nature of less informed and sophisticated investors. However, the best approach to modelling irrationality would require a very different model consisting of individual agents with different investor strategies. This would allow investigation into the emergent behaviour that arises due to locally prudent decisions combining into globally imprudent effects.

References

[1] Investopedia, "Efficient market hypothesis: Emh ." https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp, (Accessed: 18/02/2019).

[2] J. P. Bouchaud, S. Ciliberti, Y. Lempérière, A. Majewski, P. Seager, and K. S. Ronia, "Black was right: Price is within a factor 2 of Value." Capital Fund Management, Paris, 2017.

[3] Y. Lin, "Predicting the stock market and defeating the efficient market hypothesis." https://seekingalpha.com/article/3877416-predicting-stock-market-defeating-efficient-market-hypothesis-part, (Accessed: 21/02/2019).

[4] W. Branch and G. Evans, "Learning about risk and return: A simple model of bubbles and crashes." International Monetary Fund, Washington, DC, 2009.

[5] E. Hui and Q. Gu, "Study of guangzhou house price bubble based on state-space model." International Journal of Strategic Property Management, 13(4): 287-298, 2009.

[6] E. Hillebrand, "Mean reversion models of financial markets." Stanford, Universität Bremen, 2003.

[7] O. Barndorff-Nielsen and N. Shephard, "Non-gaussian ou based models and some of their uses in financial economics." Journal of the Royal Statistical Society, Series B (Statistical Methodology), Vol. 63(2): 167-241, 2001.

[8] J. Yu, "Bias in the estimation of the mean reversion parameters in continuous time models." Journal of Econometrics, Vol. 169(1): 144-122, 2012.

[9] C. Thierfelder, "The trending ornstein-uhlenbeck process and its applications in mathematical finance." University of Oxford, 2(1-2): 7-9, 12/04/2015.

[10] C. Kardaras and D. Kreher and A. Nikeghbali, "Strict local martingales and bubbles." The Annals of Applied Probability, Institute of Mathematical Statistics, Vol. 25(4): 1827-1867, 2015.

[11] W. Palma, "Long-memory time series: Theory and methods." Wiley Series in Probability and Statistics, 3(4): 57-58, 2007.

[12] A. Hayes, "Simple moving average - sma definition." https://www.investopedia.com/terms/s/sma.asp, (Accessed: 10/12/2018).

[13] G. Efstathopoulos, "Python for finance: Moving average trading strategy." https://www.learndatasci.com/tutorials/python-finance-part-3-moving-average-trading-strategy/#, (Accessed: 10/12/2018).

[14] D. Dickey and W. Fuller, "Distribution of the estimators for autoregressive time series with a unit root." Journal of the American Statistical Association, Wiley and Sons, 74(366): 427–431, 1979.

[15] J. Durbin and G. Watson, "Testing for serial correlation in least squares regression, i." Biometrika, 37(3–4): 409–428, 1950.

[16] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures." Analytical Chemistry, 36(8): 1627–39, 1964.

[17] SPC Excel, "Are the skewness and kurtosis useful statistics?." https://www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics, (Accessed: 18/02/2019).

[18] G. E. P. Box and G. Jenkins, "Time series analysis: Forecasting and control." Holden-Day, 1976.

[19] C. Chatfield, "The analysis of time series: Theory and practice." Springer, 4(1): 60-63, 2013.

[20] L. Spierdijk and J. Bikker, "Mean reversion in stock prices." DNB Working Paper, no. 343, 2012.

[21] R. Schafer, "On the frequency-domain properties of savitzky-golay filters." Hewlett-Packard Laboratories, Palo Alto, California, 2010.

[22] S. Williamson, "Daily closing values of the dow jones average in the united states." https://www.measuringworth.com/datasets/DJA/index.php, (Accessed: 30/01/2019).

[23] Federal Reserve Bank of St. Louis, "Economic research." https://fred.stlouisfed.org/, (Accessed: 15/02/2019).

[24] Investopedia, "What's the difference between the dow jones industrial average and the s&p 500?." https://www.investopedia.com/ask/answers/difference-between-dow-jones-industrial-average-and-sp-500/, (Accessed: 14/02/2019).

[25] International Monetary Fund, "Asset bubbles: Re-thinking policy for the age of asset management." https://www.imf.org/external/pubs/ft/wp/2015/wp1527.pdf, (Accessed: 22/02/2019).

[26] M. Herdegen and M. Schweizer, "Strong bubbles and strict local martingales." International Journal of Theoretical and Applied Finance Issue 19, 2016.

[27] Investopedia, "Dividend discount model." https://www.investopedia.com/terms/d/ddm.asp, (Accessed: 22/02/2019).

Appendix A Methods

A.1 Proxy to the Fundamental Value

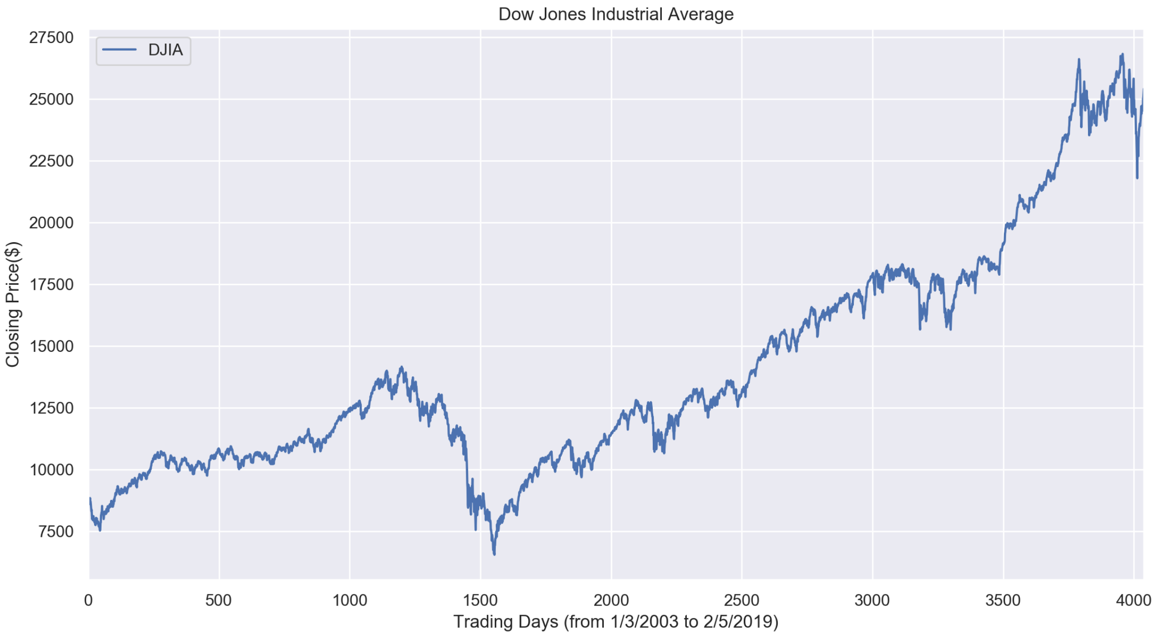

Figure 2a: Figure depicting a simple price time series of the Dow Jones Industrial Average (DJIA), using the closing prices of days from 1/3/2003 to 2/5/2019. Data is collected from [22].

Figure 2b: Example of two simple moving average (SMA) processes applied to the DJIA dataset. The graph shows the SMA with a window size of 100 days and the SMA with a window size of 500 days.

Figure 2c: Example of two exponential moving average (EMA) processes applied to the DJIA dataset. The graph shows the EMA with a window size of 100 days and the EMA with a window size of 500 days.

A.2 Model Specification/Testing for Stationarity

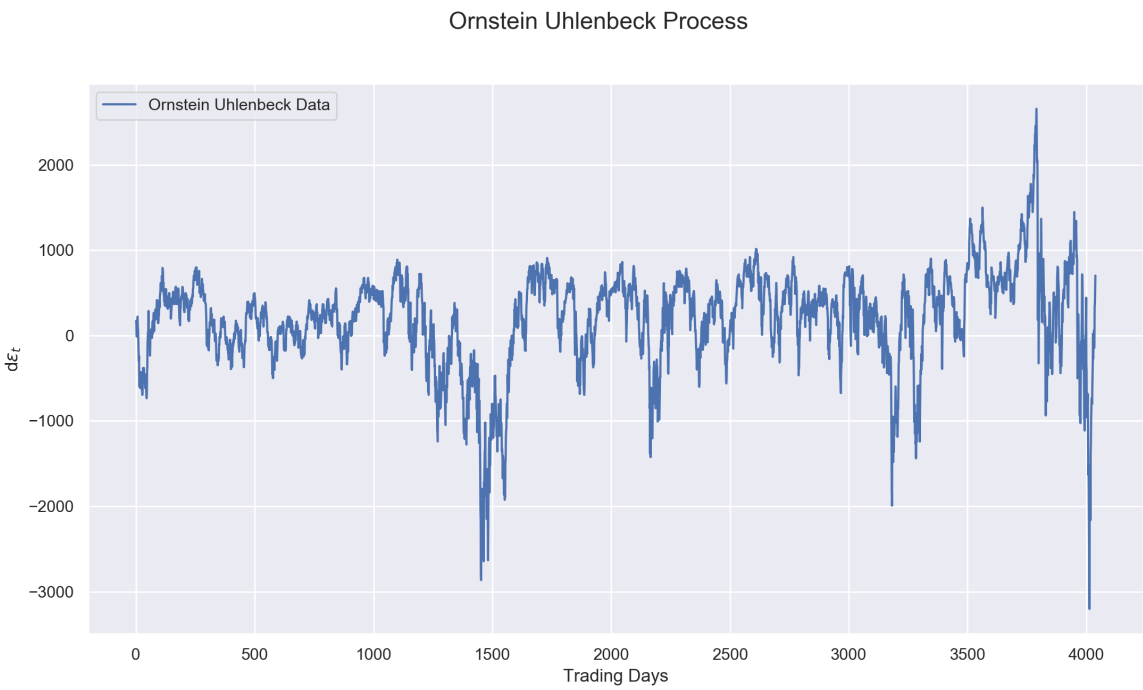

Figure 3: Example of an OU process which describes the evolution of the residuals of the DJIA data with the EMA. Here, residuals have been taken from an EMA with a window size of 100 days.

A.3 Filtering White Noise

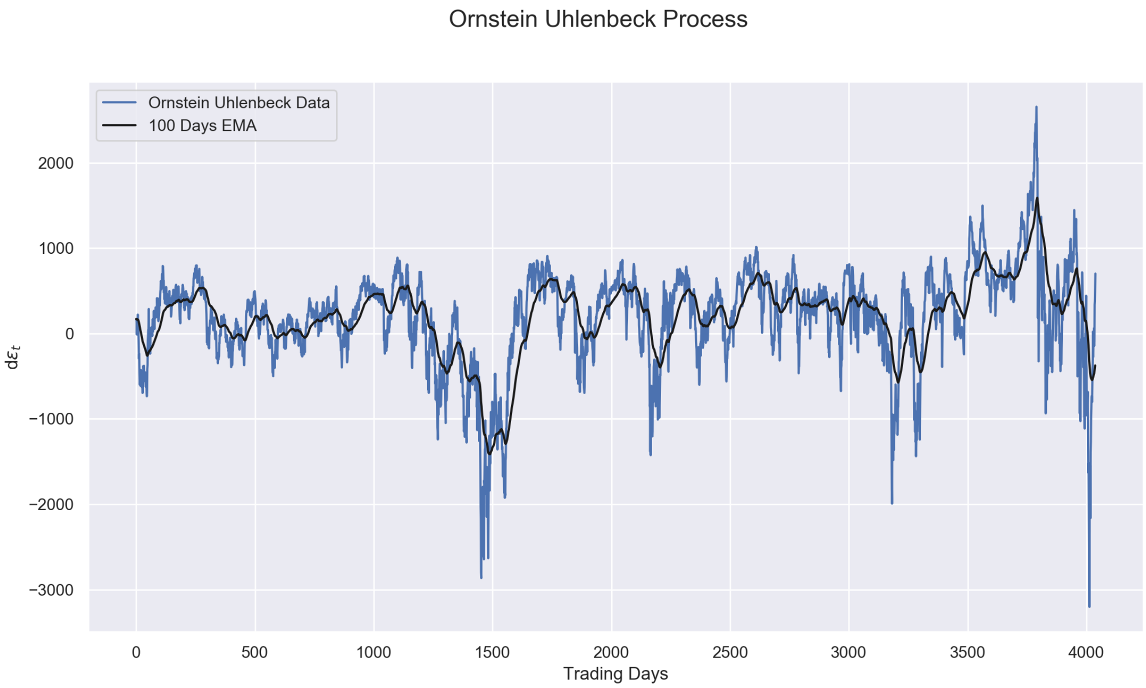

Figure 4a: Filtering of the OU data for the DJIA dataset, again using an EMA with a window size of 100 days.

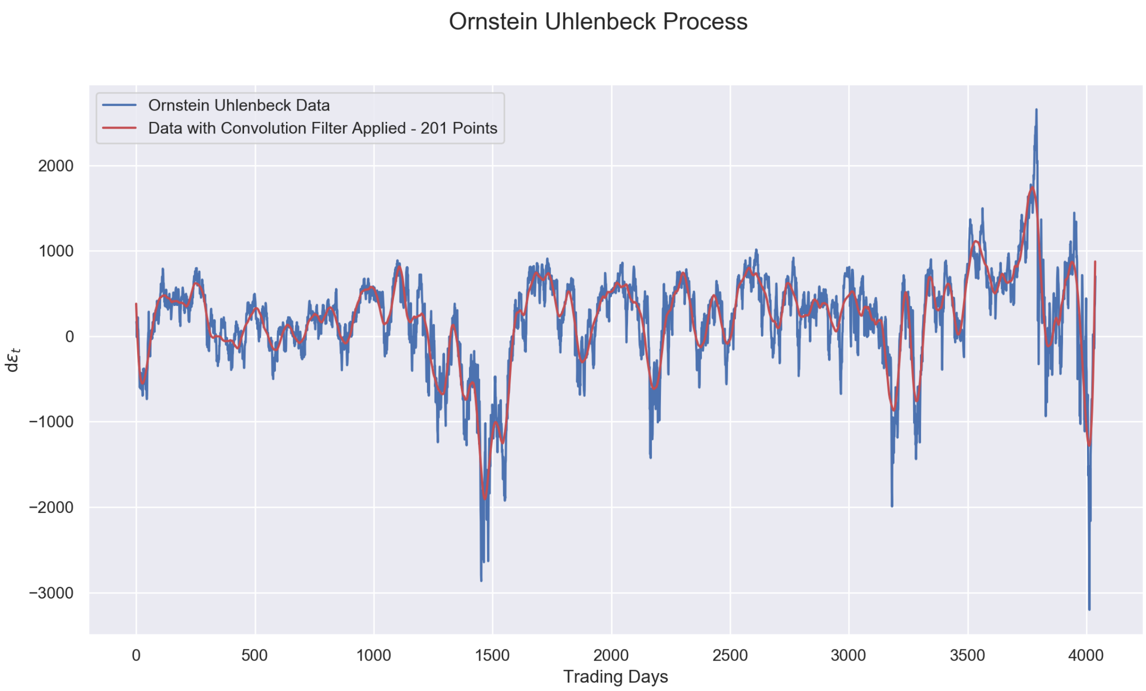

Figure 4b: Filtering of the OU data for the DJIA dataset using the Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree. Compared to the EMA, peak heights are more consistent with the OU process.

A.4 Measuring Peak Height and Mean Reversion Time

Figure 5: Figure depicting detection of peaks above the fundamental value line for the DJIA dataset with 201 convolution points and a fitting polynomial of sixth degree.

A.5 Determining Parameters

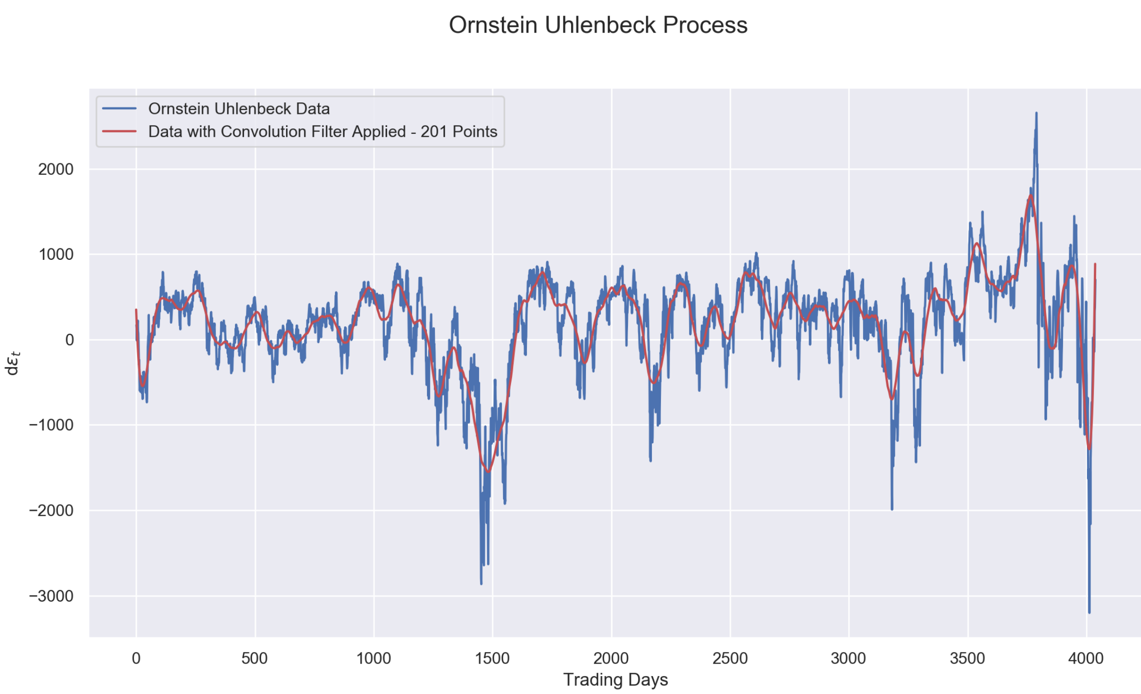

Figure 6a: Filtering of the OU data for DJIA using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of fifth degree.

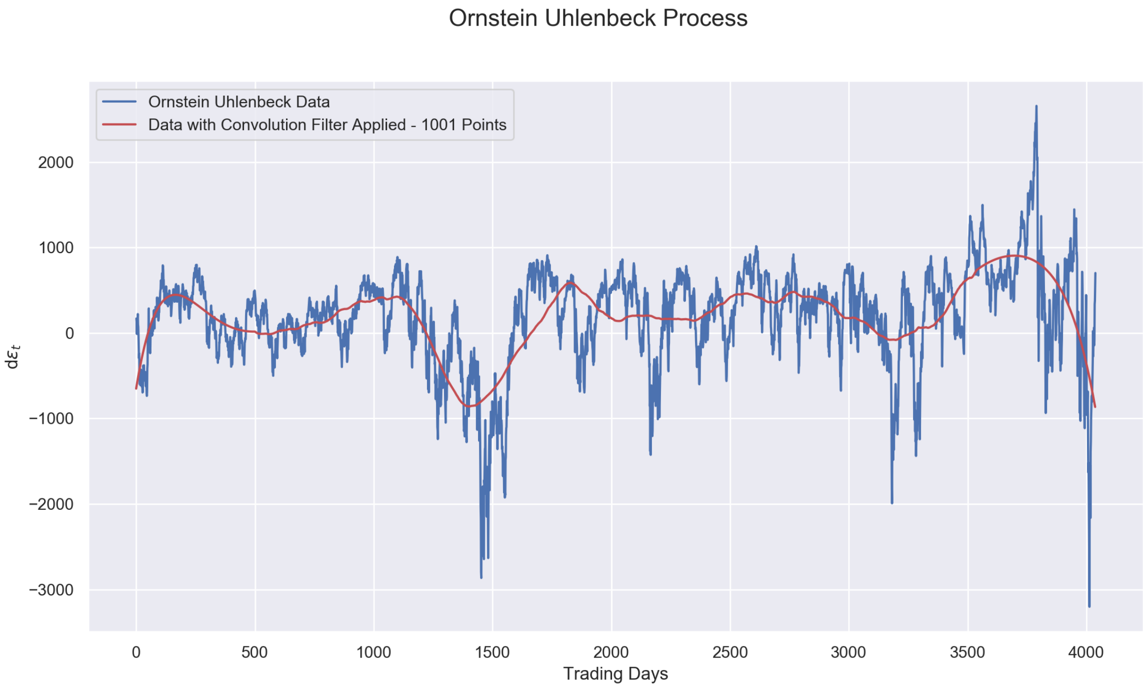

Figure 6b: Filtering of the OU data for DJIA using a Savitzky-Golay filter with 1001 convolution points and a fitting polynomial of fifth degree. The fit here is clearly over-smoothed.

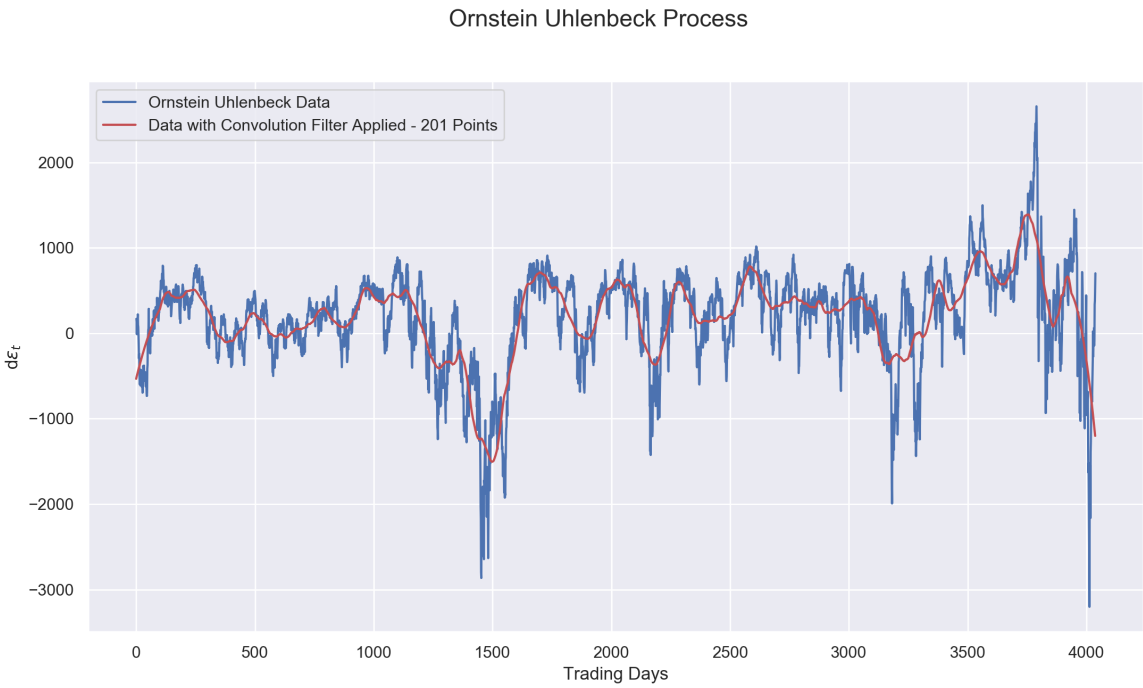

Figure 7a: Filtering of the OU data for DJIA using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of second degree.

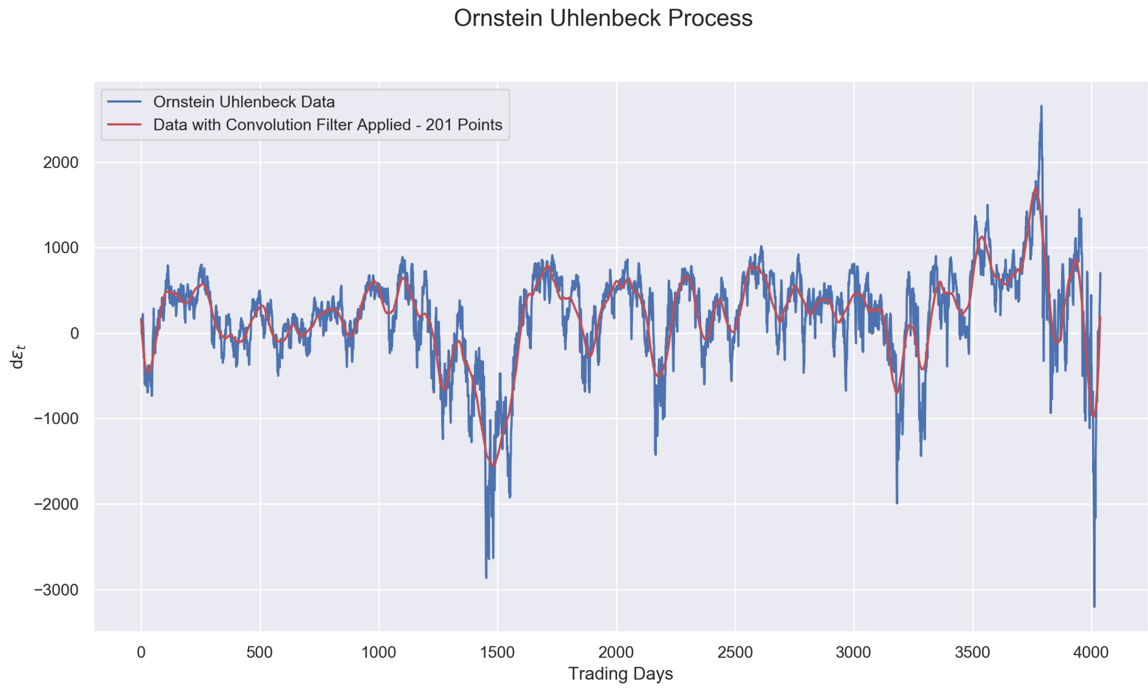

Figure 7b: Filtering of the OU data for DJIA using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of fourth degree.

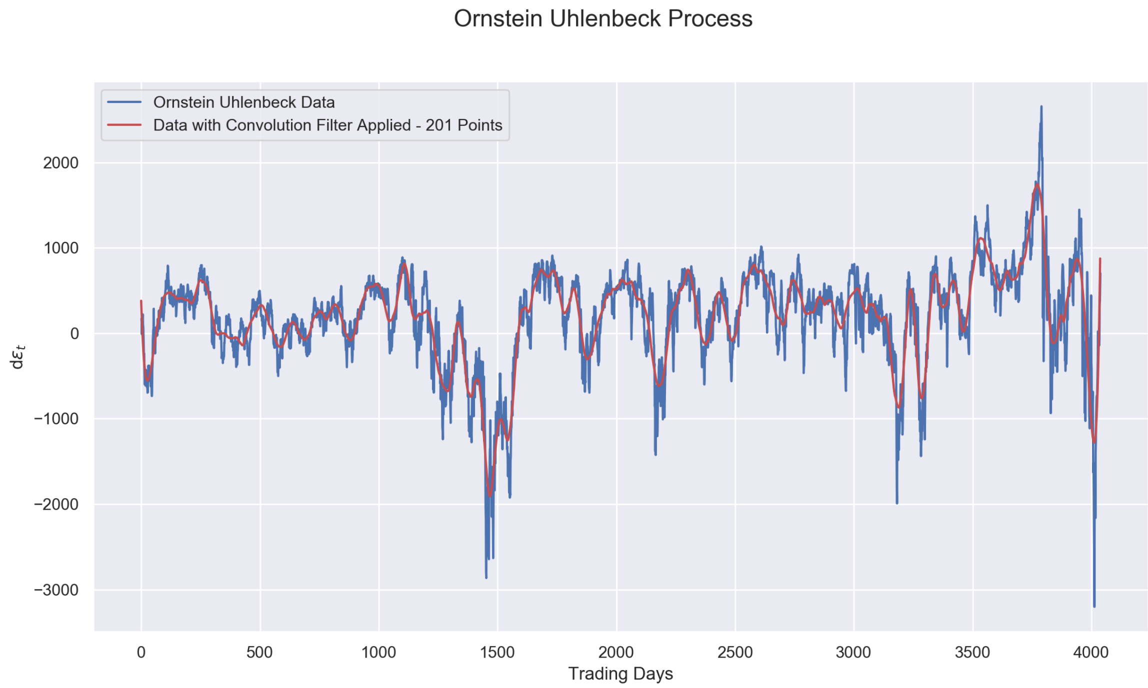

Figure 7c: Filtering of the OU data for DJIA using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree.

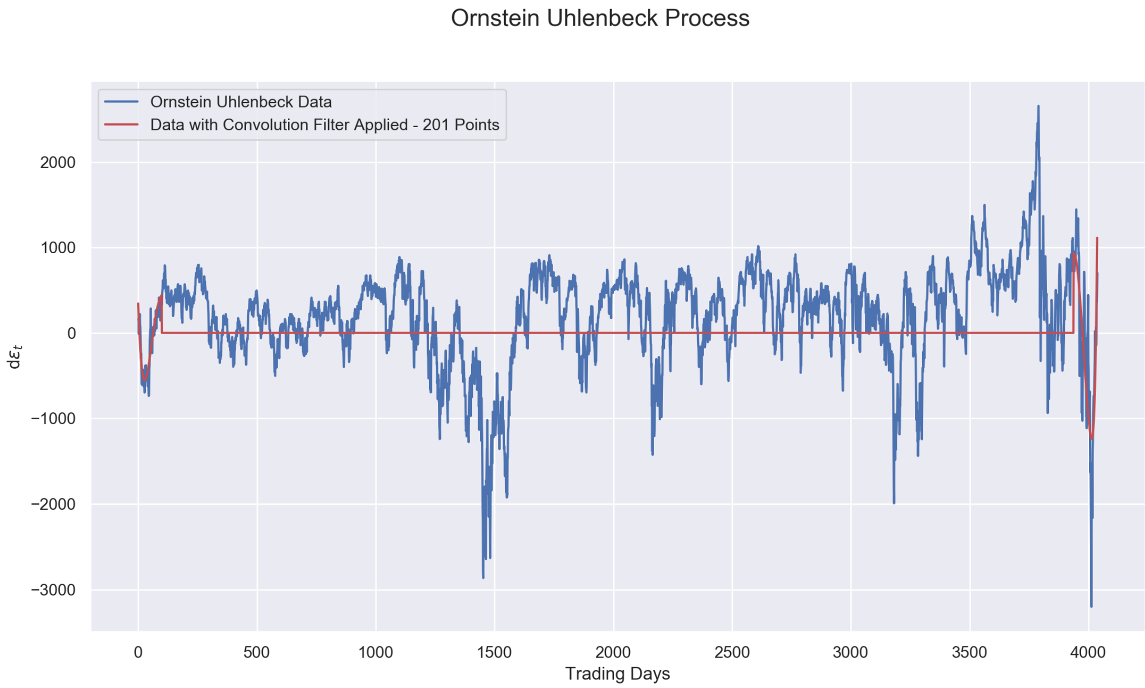

Figure 7d: Filtering of the OU data for DJIA using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of eighth degree. These parameters produce an erroneous fit for the data.

Appendix B Results

B.1 Dataset Statistics

Table 1: Start dates and frequency at which prices are closed for the datasets considered.

Asset Statistics		
Asset Class	Start Date	Closing Frequency
Gold	01-04-1968	Daily
Crude Oil	01-01-1946	Monthly
DJIA	04-01-1960	Daily
S&P	03-01-1950	Daily
GDP	01-04-1947	Quarterly

Table 2: Dataset Statistics.

Asset Statistics				
Asset Class	Durbin-Watson Statistic	ADF Statistic	Skewness	Kurtosis
Gold	0.015	-7.066	2.172	4.886
Crude Oil	0.161	-8.899	1.202	0.147
DJIA	0.067	-8.172	1.734	3.633
S&P	0.012	-7.247	1.001	-0.245
GDP	0.086	-4.501	0.265	-0.963

B.2 Commodities

B.2.1 Gold

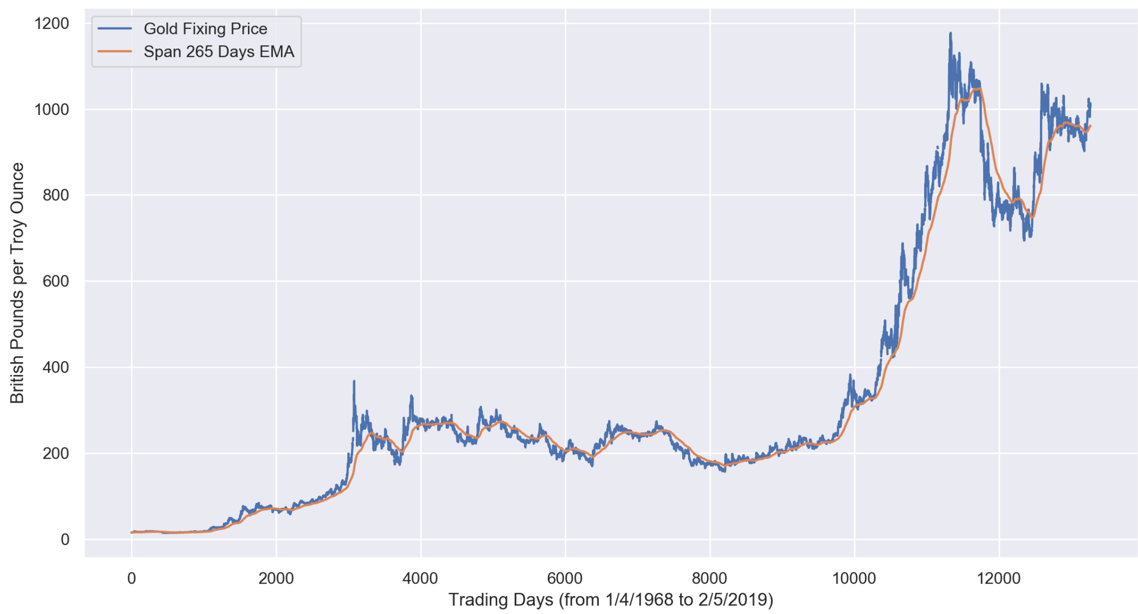

Figure 8a: The exponential moving average (EMA) process applied to the gold time series of daily closing prices from 1968 to 2019. The graph shows the EMA with a window size of 265 days.

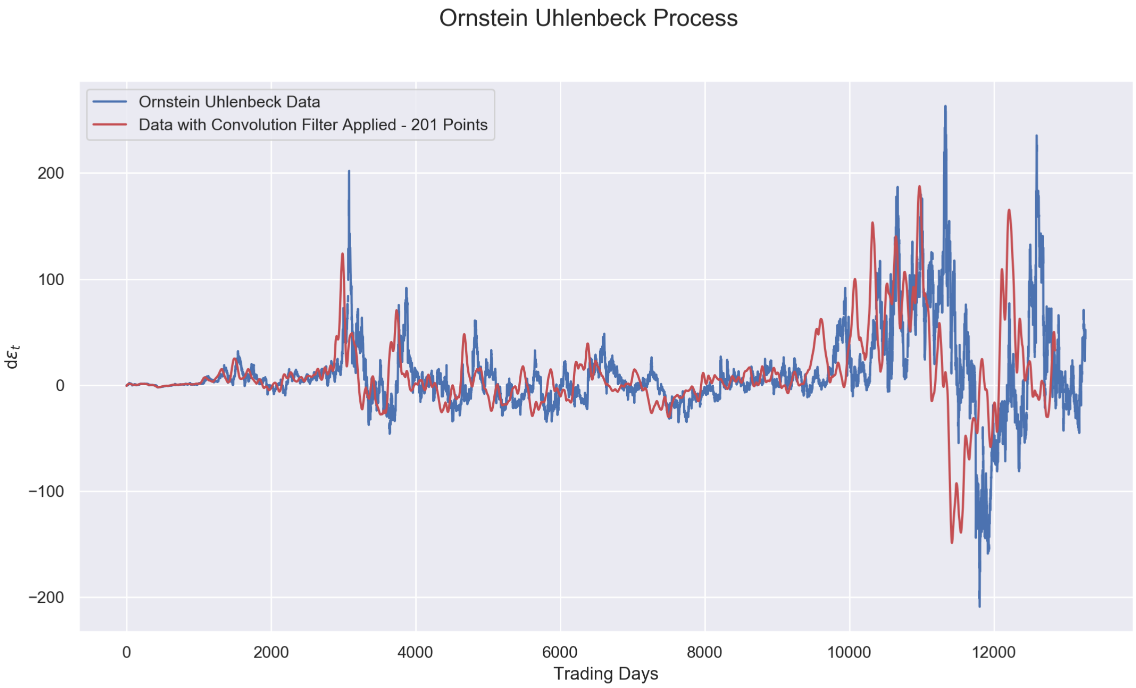

Figure 8b: Filtering of the OU data for gold using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree.

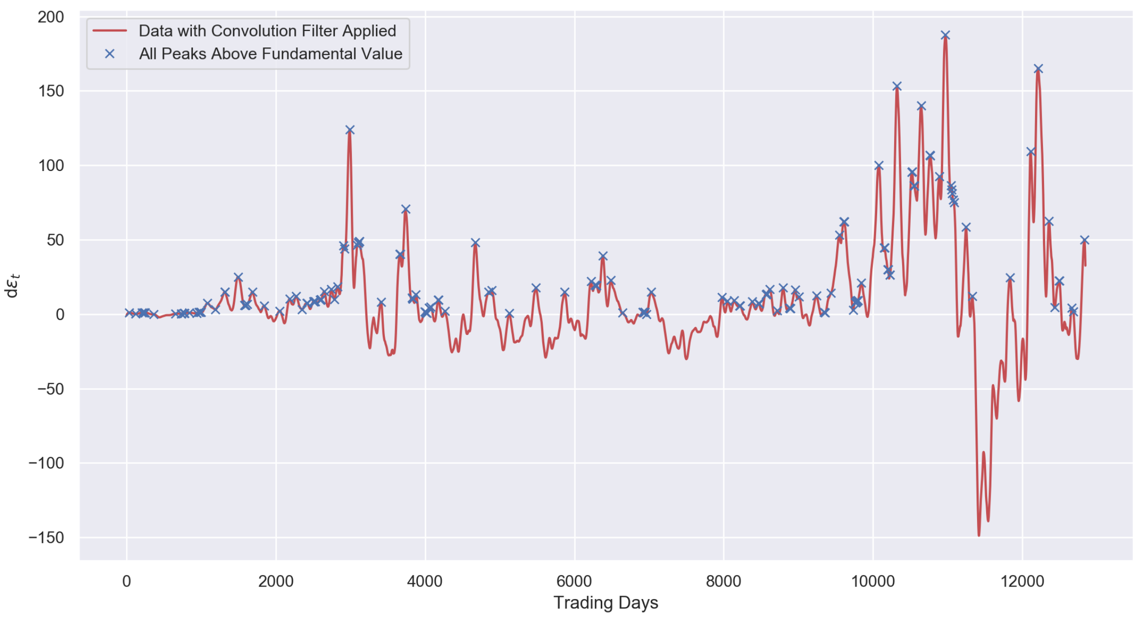

Figure 8c: Figure depicting detection of peaks above the fundamental value line for gold with 201 convolution points and a fitting polynomial of sixth degree.

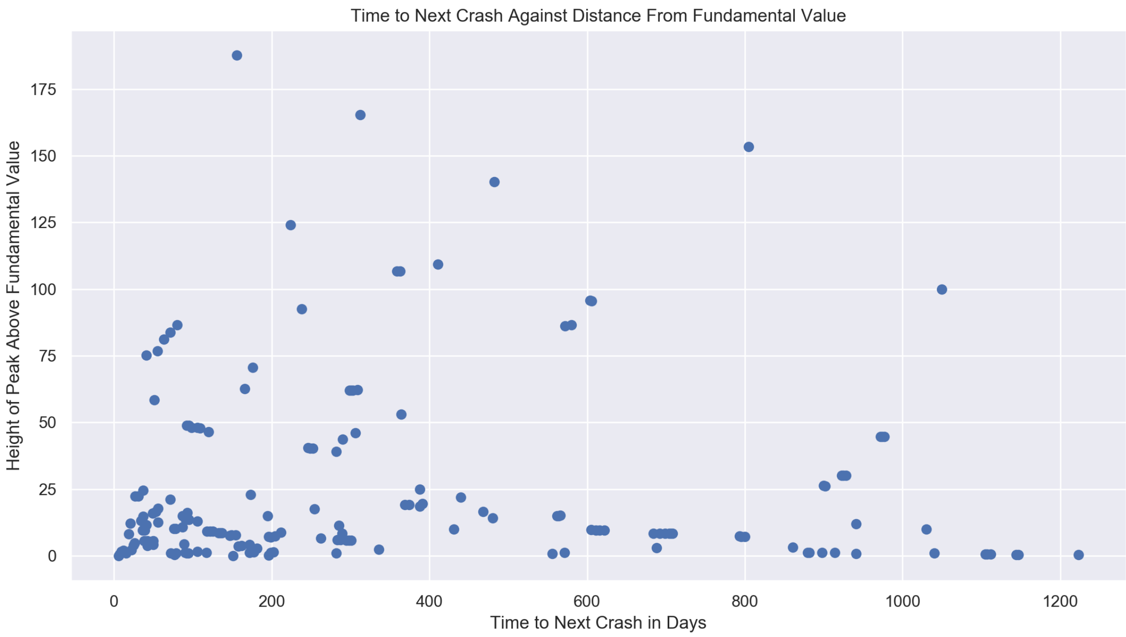

Figure 8d: Graph plotting time to next crash against the height of all peaks above the fundamental value for gold with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 0. The skewness statistic for gold is 2.172, meaning the plot is positively skewed.

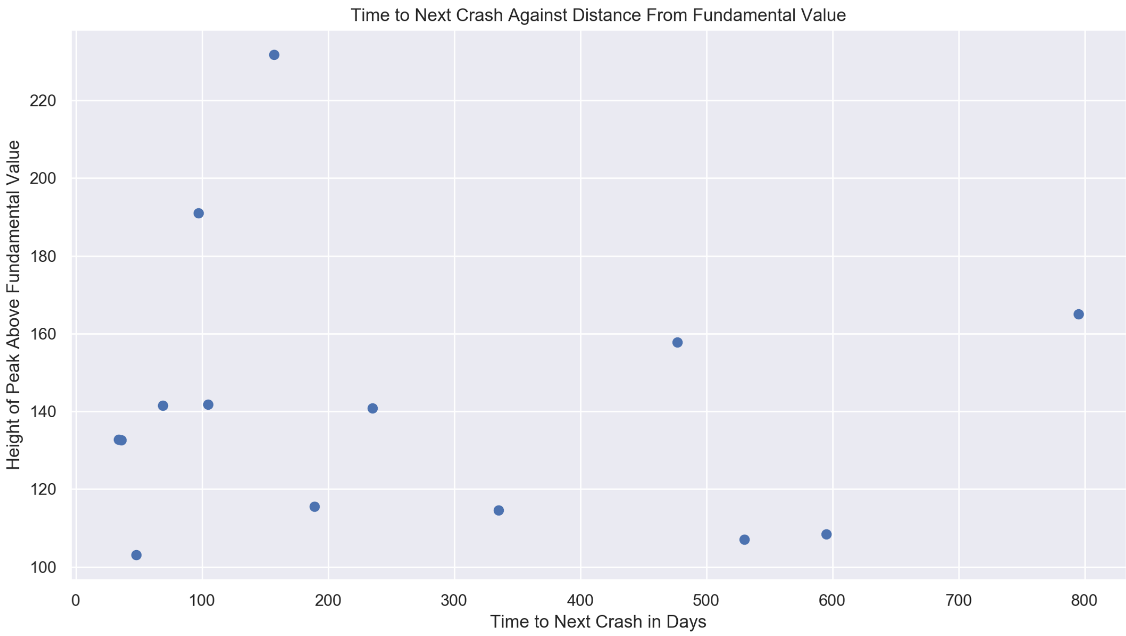

Figure 8e: Graph plotting time to next crash against the height of all peaks above the fundamental value for gold with 101 convolution points, a fitting polynomial of sixth degree and border cropping of 100.

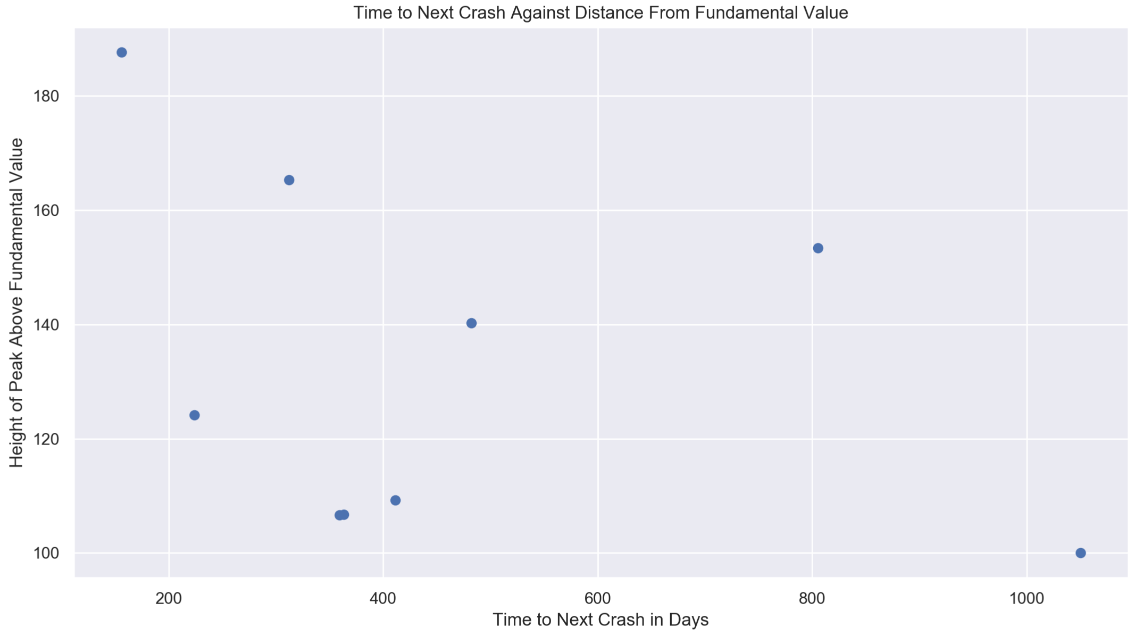

Figure 8f: Graph plotting time to next crash against the height of all peaks above the fundamental value for gold with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 100.

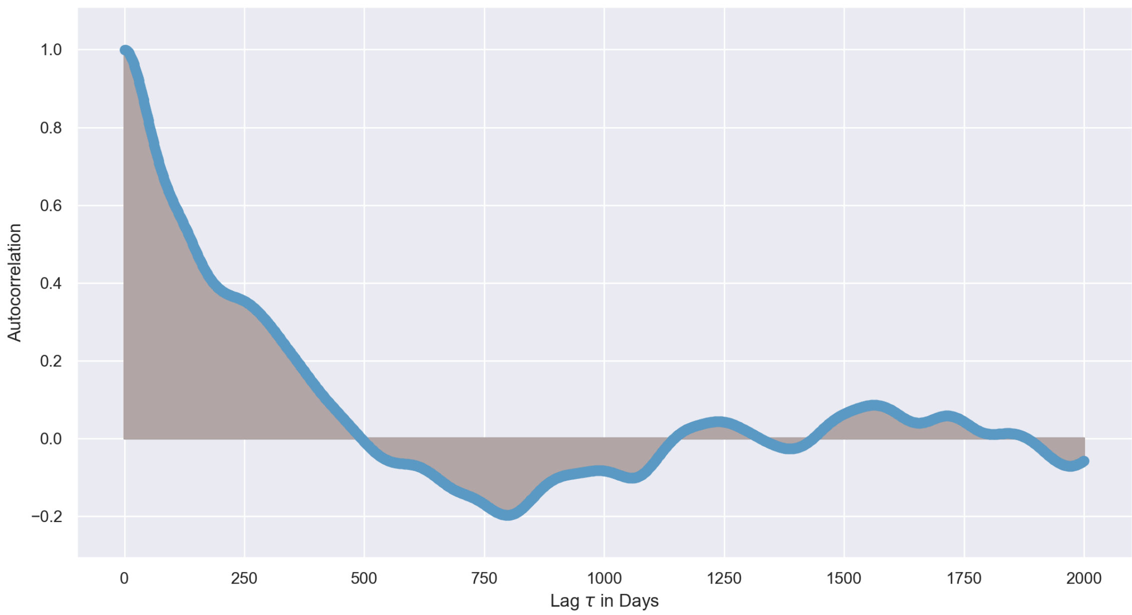

Figure 8g: Autocorrelation plot for gold over a total of 2000 lags. The correlation coefficient is positive between for lag less than 500 and then is negative between 500 and 1125 days. This indicates that the process is negatively correlated for this window and should mean revert. After 1125 days the process oscillates between positive and negative coefficients as tau increases.

B.2.2 Crude Oil

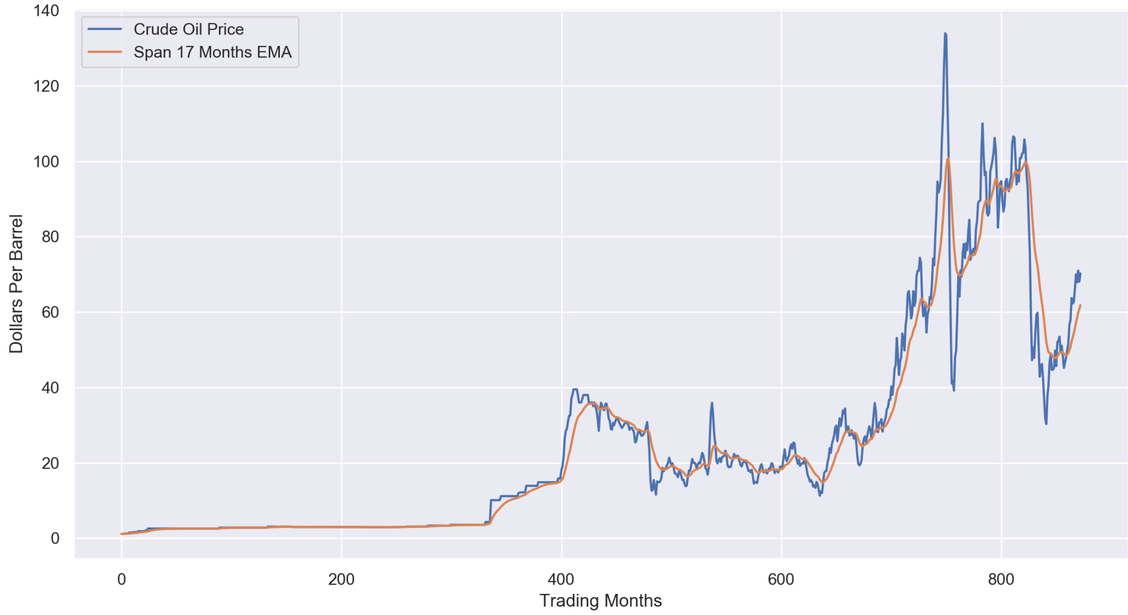

Figure 9a: The exponential moving average (EMA) process applied to the crude oil time series of monthly closing prices from 1946 to 2019. The graph shows the EMA with a window size of 17 months.

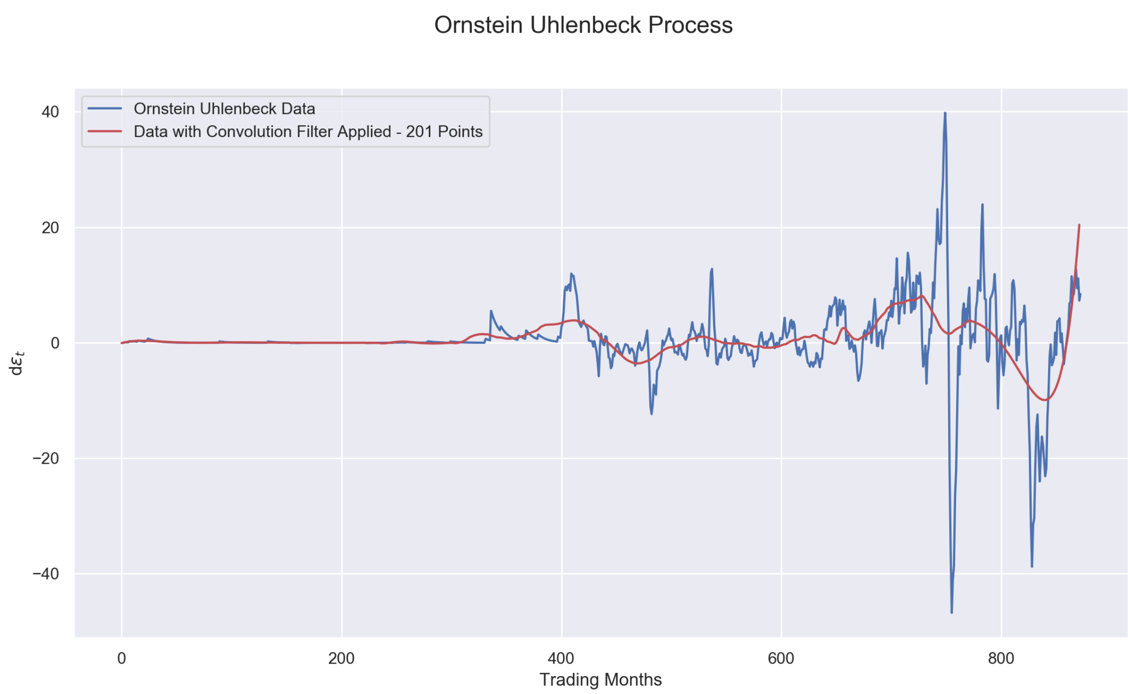

Figure 9b: Filtering of the OU data for crude oil using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree.

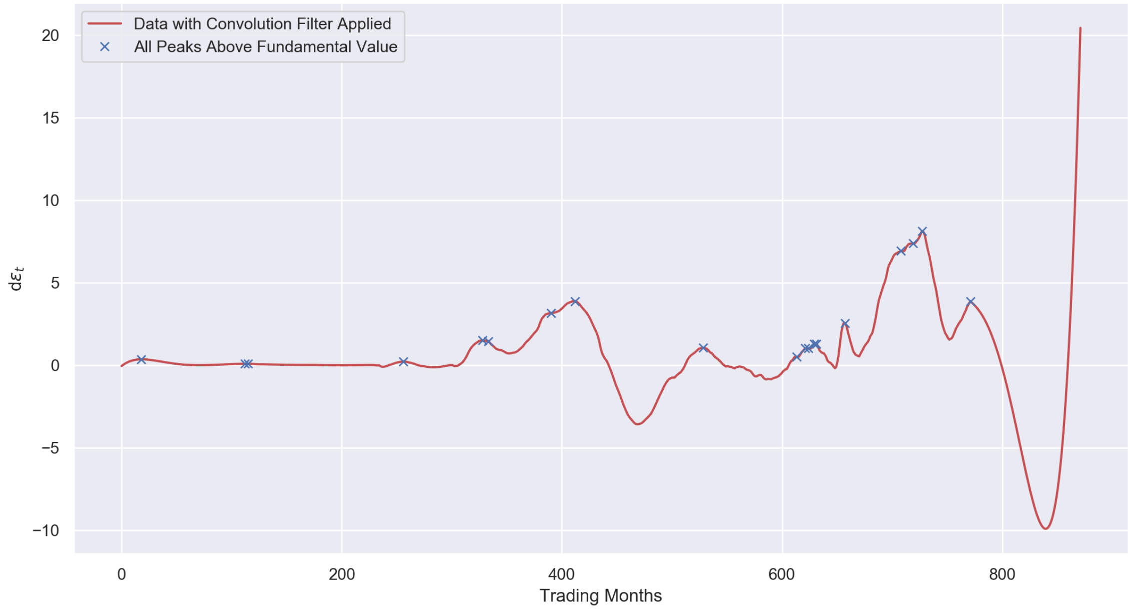

Figure 9c: Figure depicting detection of peaks above the fundamental value line for crude oil with 201 convolution points and a fitting polynomial of sixth degree.

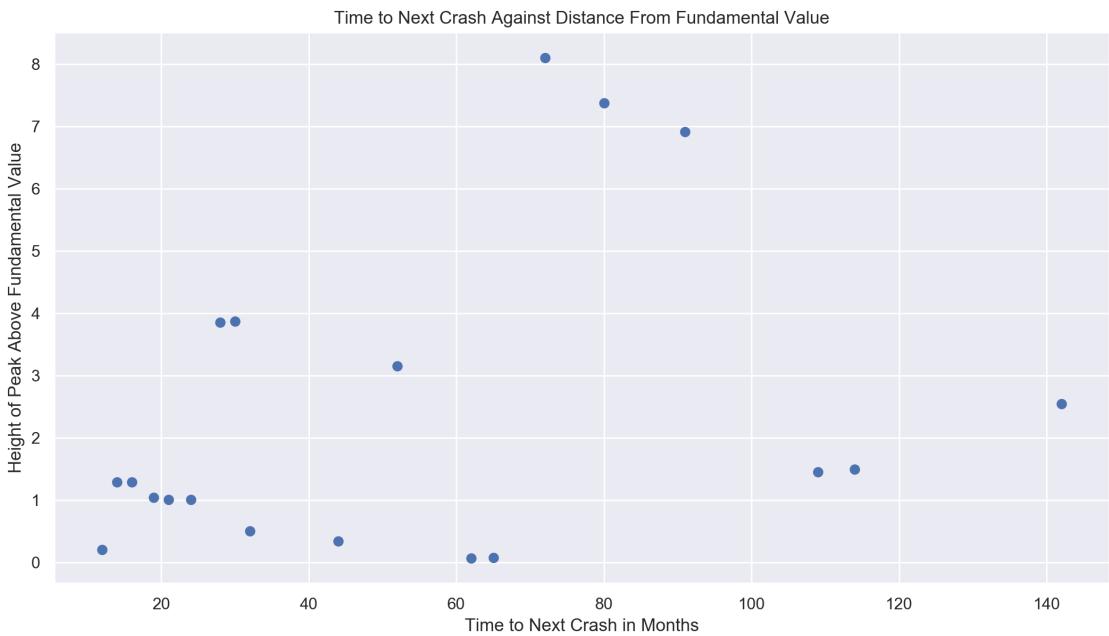

Figure 9d: Graph plotting time to next crash against the height of all peaks above the fundamental value for crude oil with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 0.

Figure 9e: Autocorrelation plot for oil over a total of 1000 lags. The coefficient is closer to zero at shorter lag times although the frequency at which the correlation changes from positive to negative remains roughly the same for all lag times. This shows that the oil time series has frequently occurring periods of mean reverting behaviour.

B.3 Equity

B.3.1 DJIA

Figure 10a: The exponential moving average (EMA) process applied to the DJIA time series of daily closing prices from 1960 to 2019. The graph shows the EMA with a window size of 297 trading days.

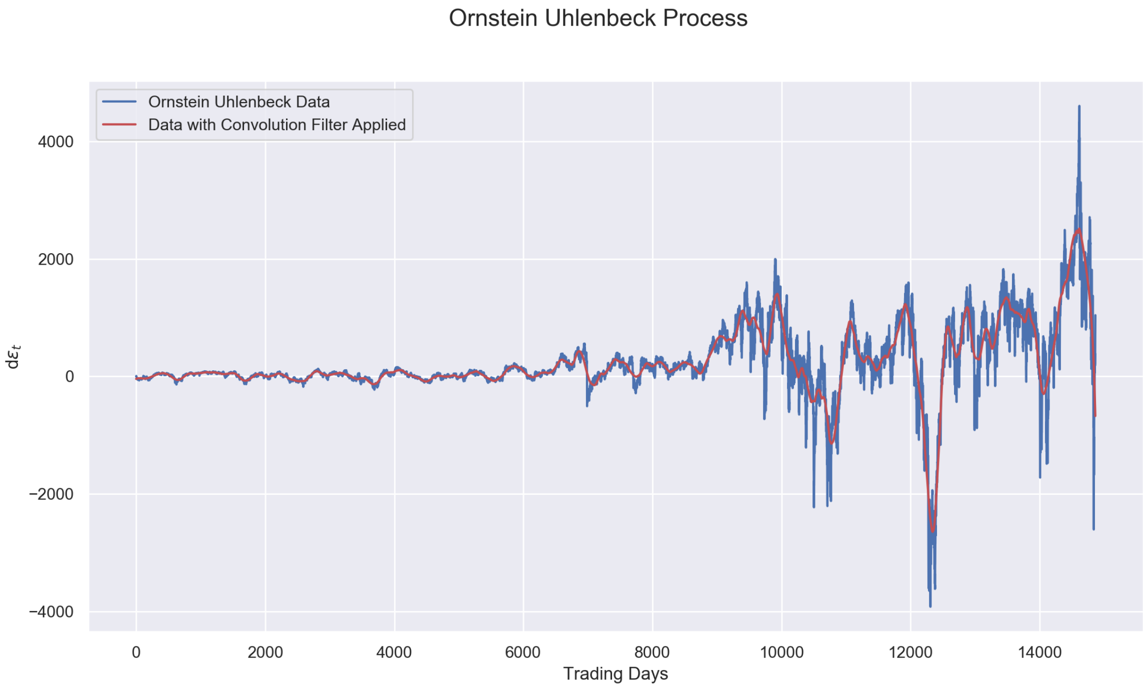

Figure 10b: Filtering of the OU data for the entire DJIA dataset using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree.

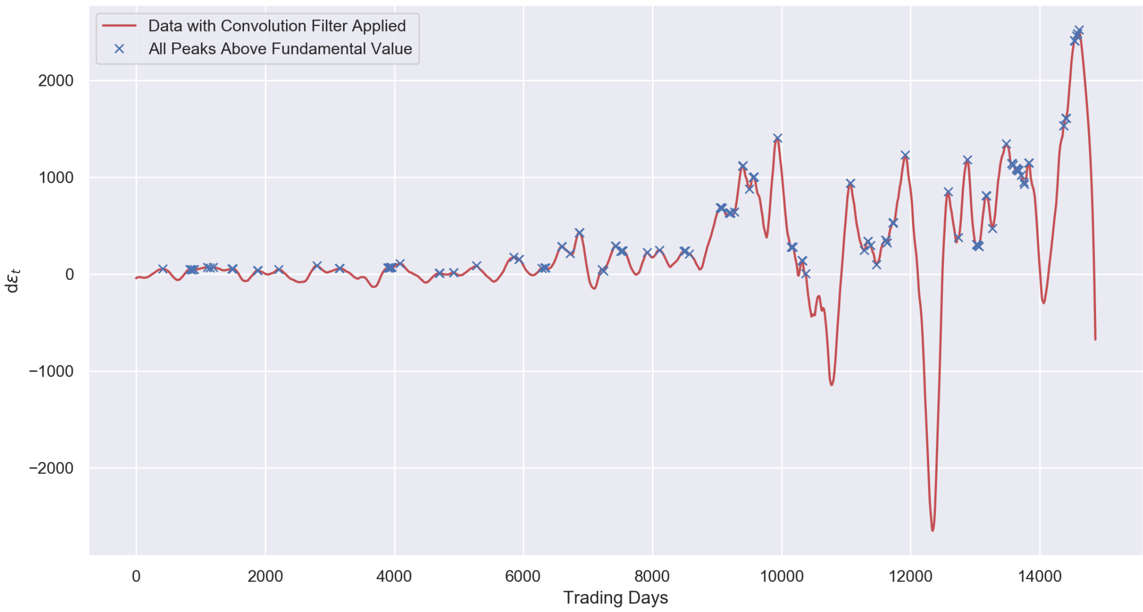

Figure 10c: Figure depicting detection of all peaks above the fundamental value for DJIA with 201 convolution points and a fitting polynomial of sixth degree.

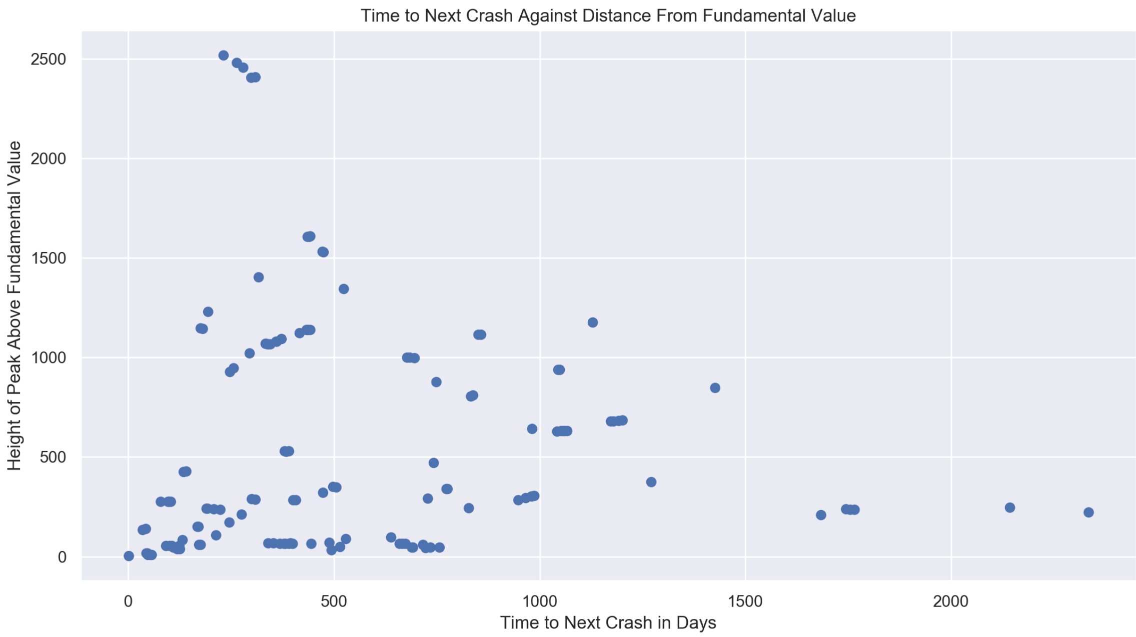

Figure 10d: Graph plotting time to next crash against the height of all peaks above the fundamental value for DJIA with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 0.

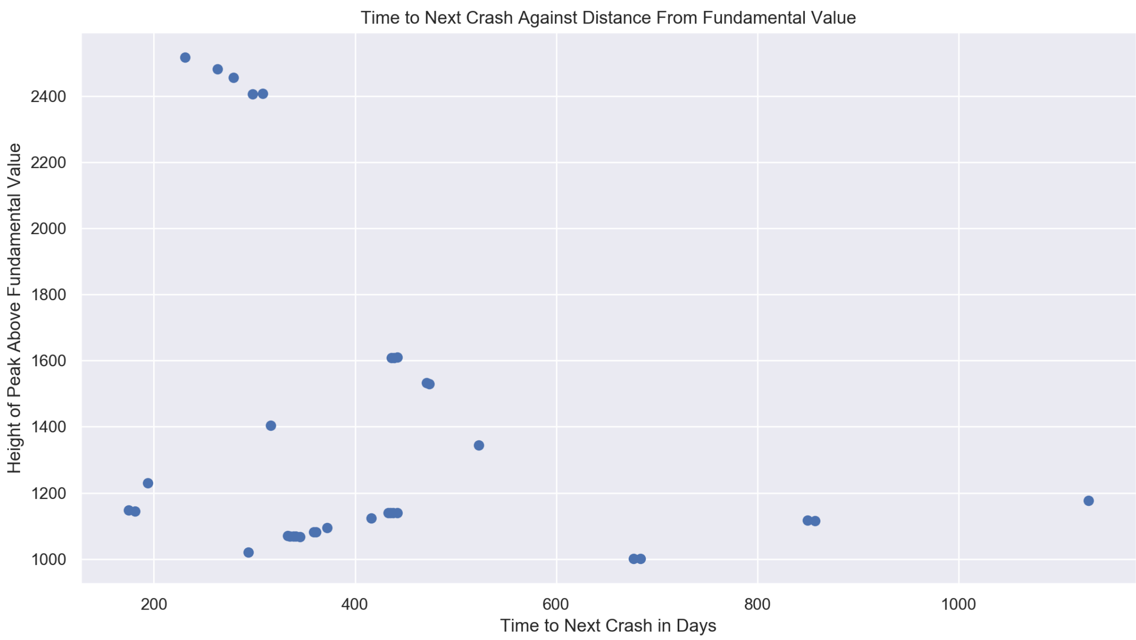

Figure 10e: Graph plotting time to next crash against the height of all peaks above the fundamental value for DJIA with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 1000.

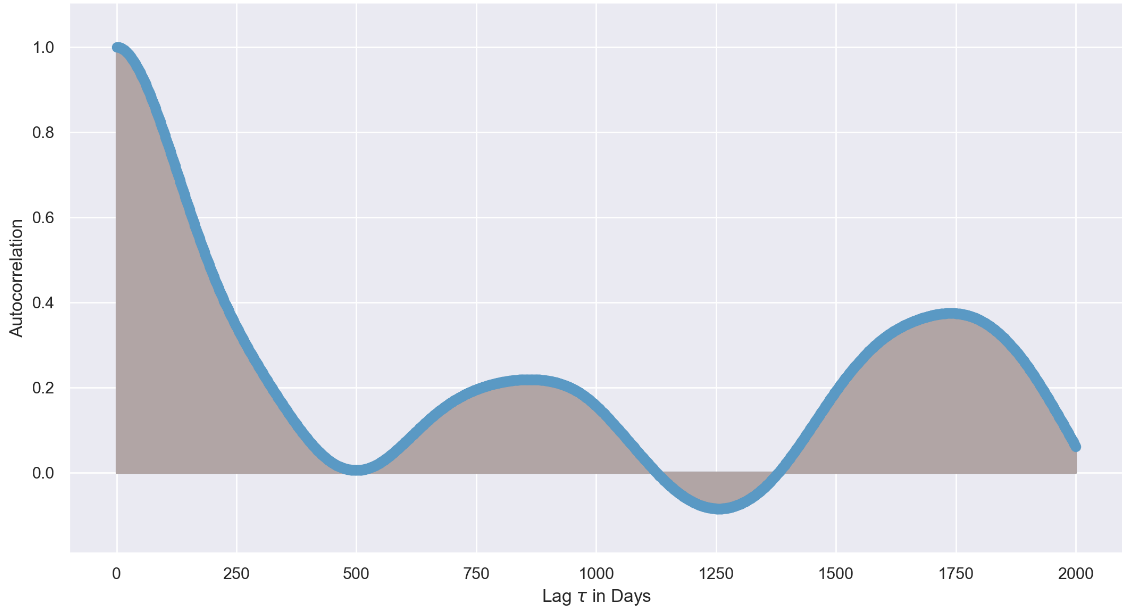

Figure 10f: Autocorrelation plot for DJIA over a total of 2000 lags. The correlation coefficient is greater than 1 apart from the period 1125-1375. This suggests that the time series will trend up until around 1250 days at which point the series will mean revert.

B.3.2 S&P500

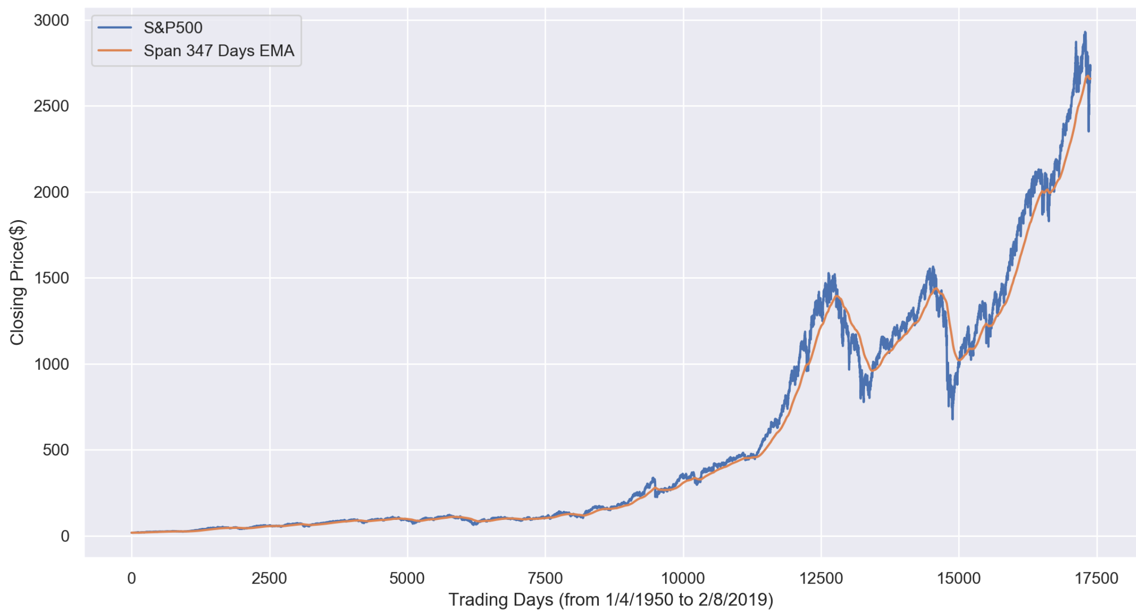

Figure 11a: The exponential moving average (EMA) process applied to the S&P500 time series of daily closing prices from 1950 to 2019. The graph shows the EMA with a window size of 347 trading days.

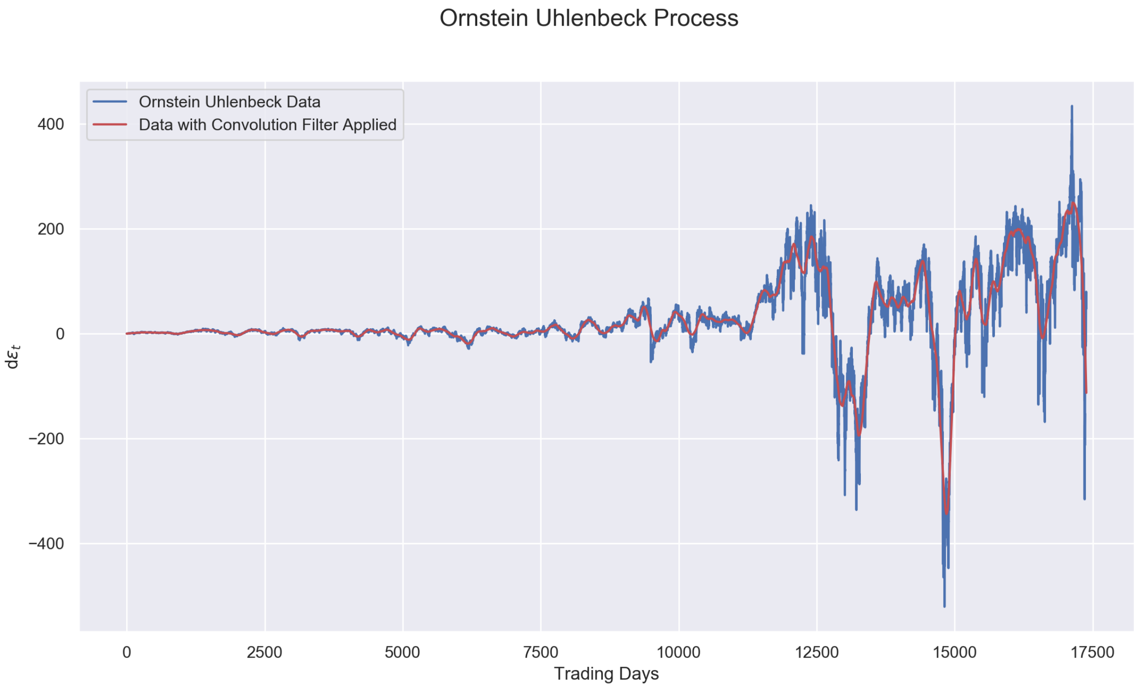

Figure 11b: Filtering of the OU data for the entire S&P500 dataset using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree.

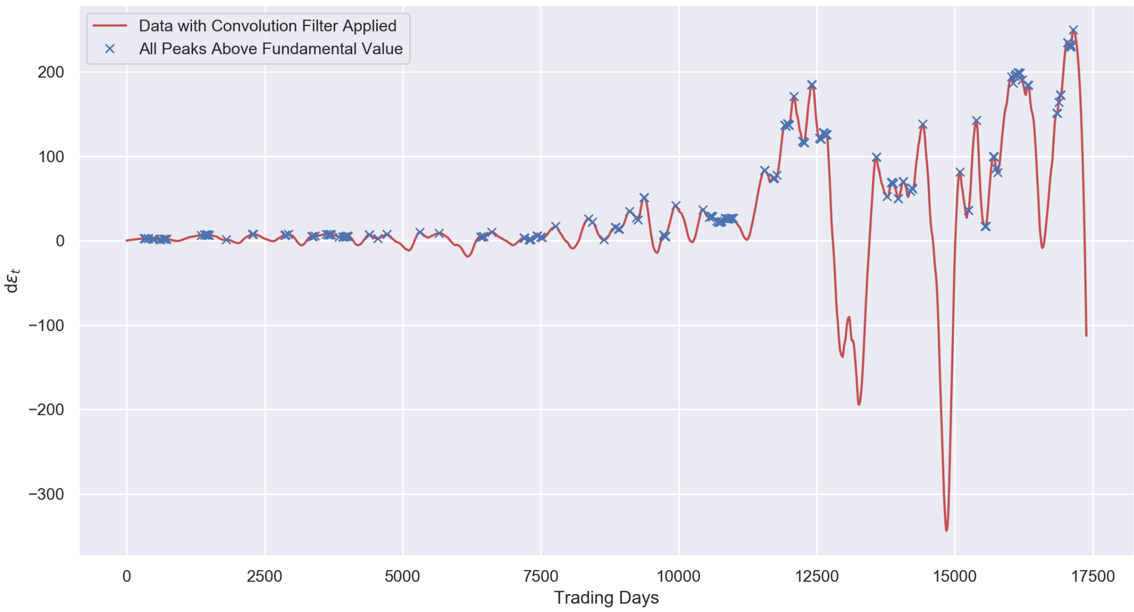

Figure 11c: Figure depicting detection of all peaks above the fundamental value for the S&P500 dataset with a 201 convolution points and a fitting polynomial of sixth degree.

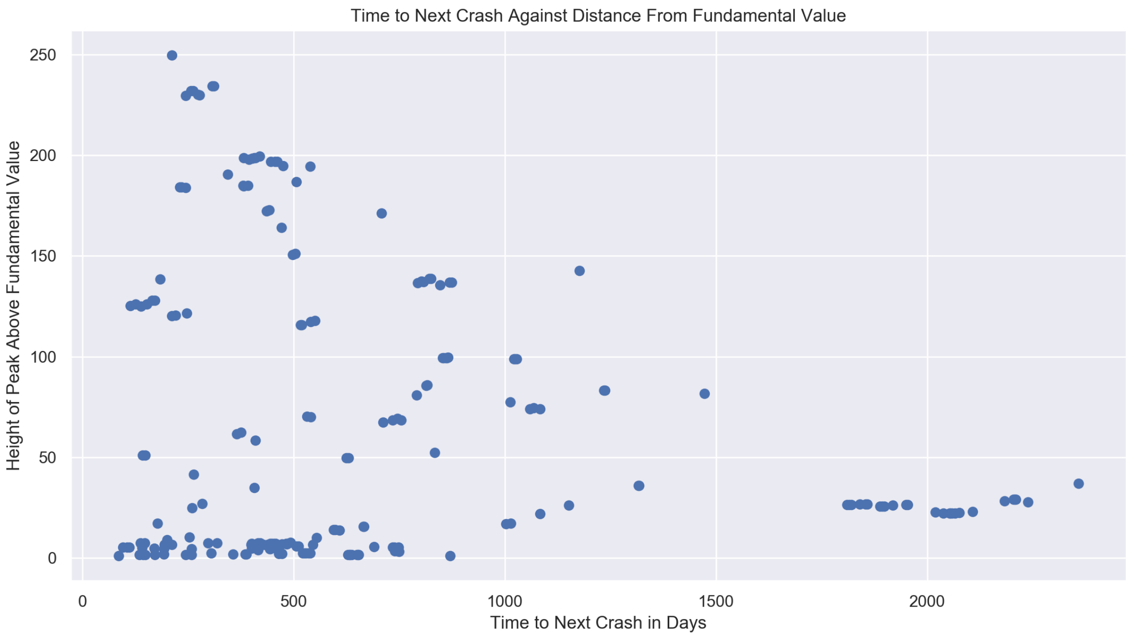

Figure 11d: Graph plotting time to next crash against the height of all peaks above the fundamental value for S&P500 dataset with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 0.

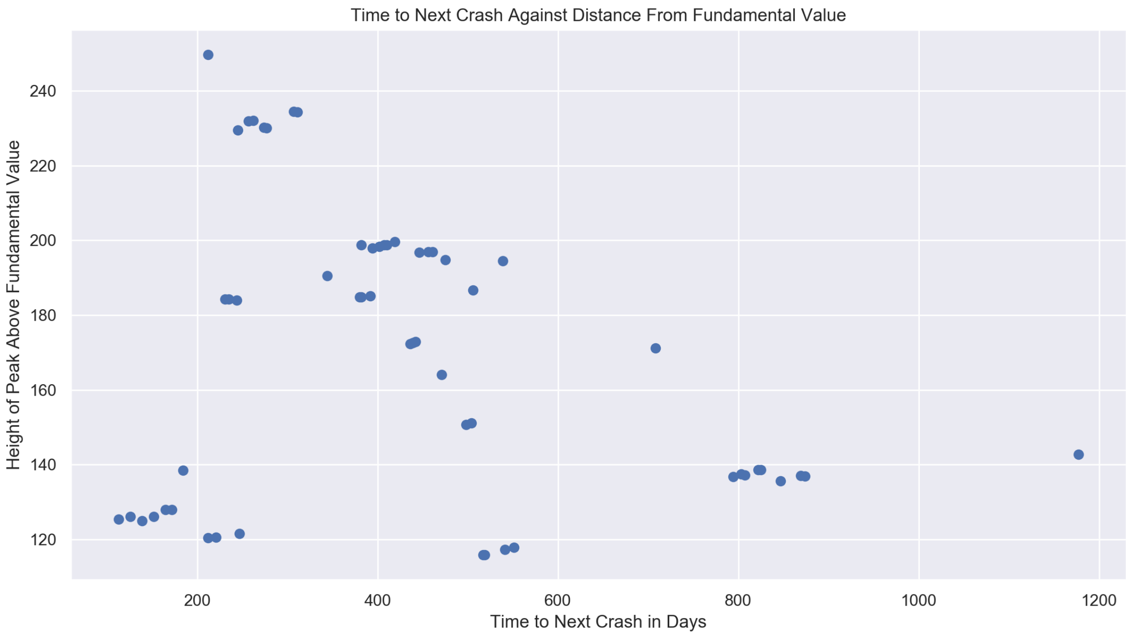

Figure 11e: Graph plotting time to next crash against the height of all peaks above the fundamental value for S&P500 dataset with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 100.

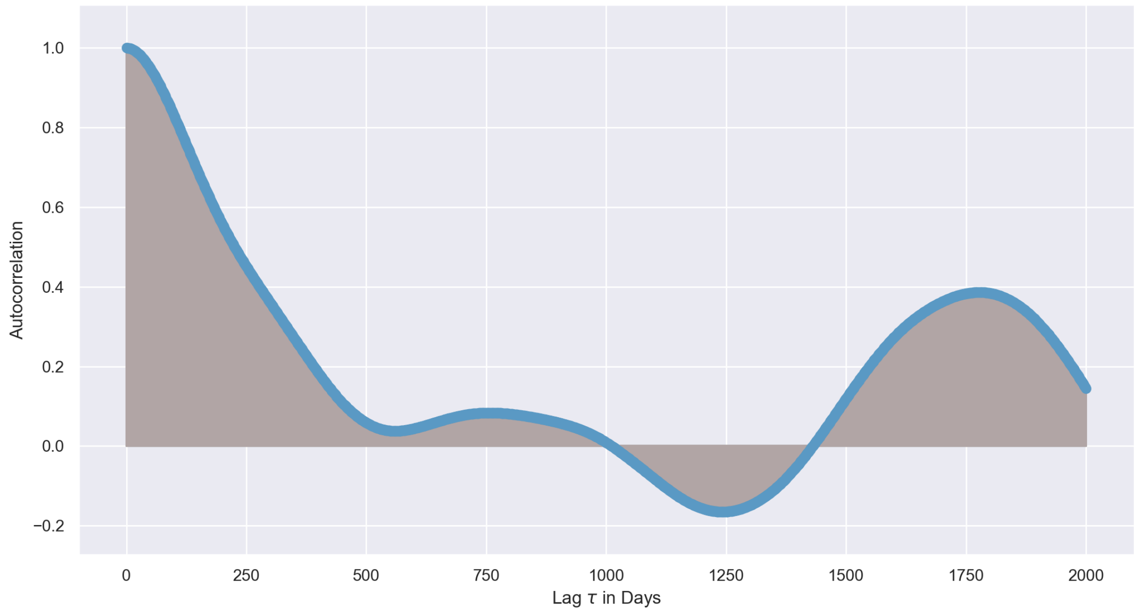

Figure 11f: Autocorrelation plot for SP500 over a total of 2000 lags. The autocorrelation coefficient is positive until 1000 days and stays negative up to 1400 days.

B.4 GDP

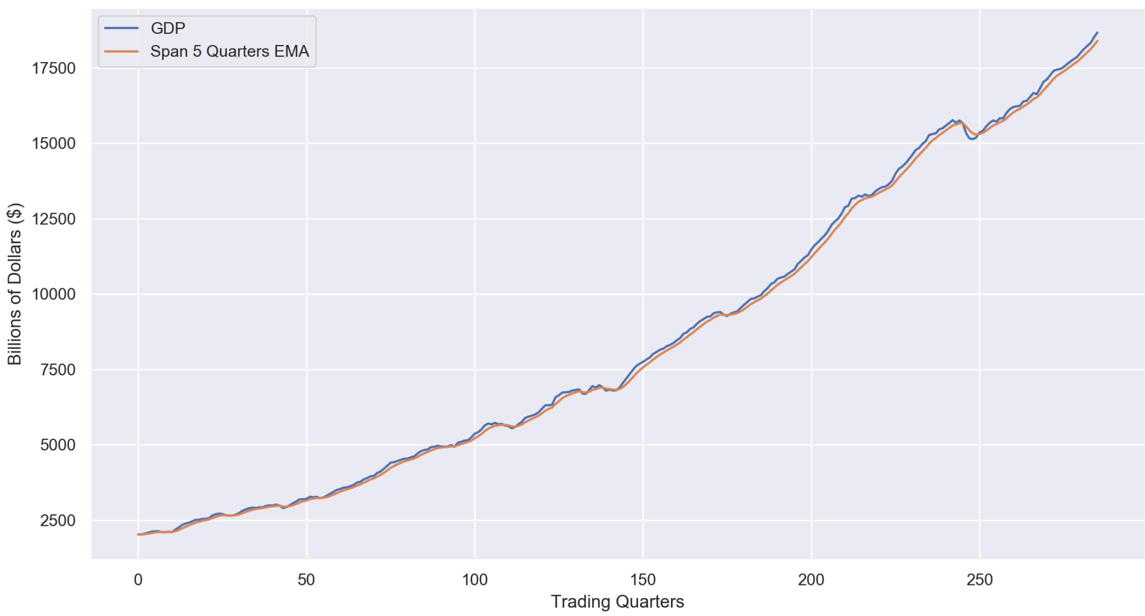

Figure 12a: The exponential moving average (EMA) process applied to the GDP time series of quarterly closing prices from 1947 to 2018. The graph shows the EMA with a window size of 5 quarters.

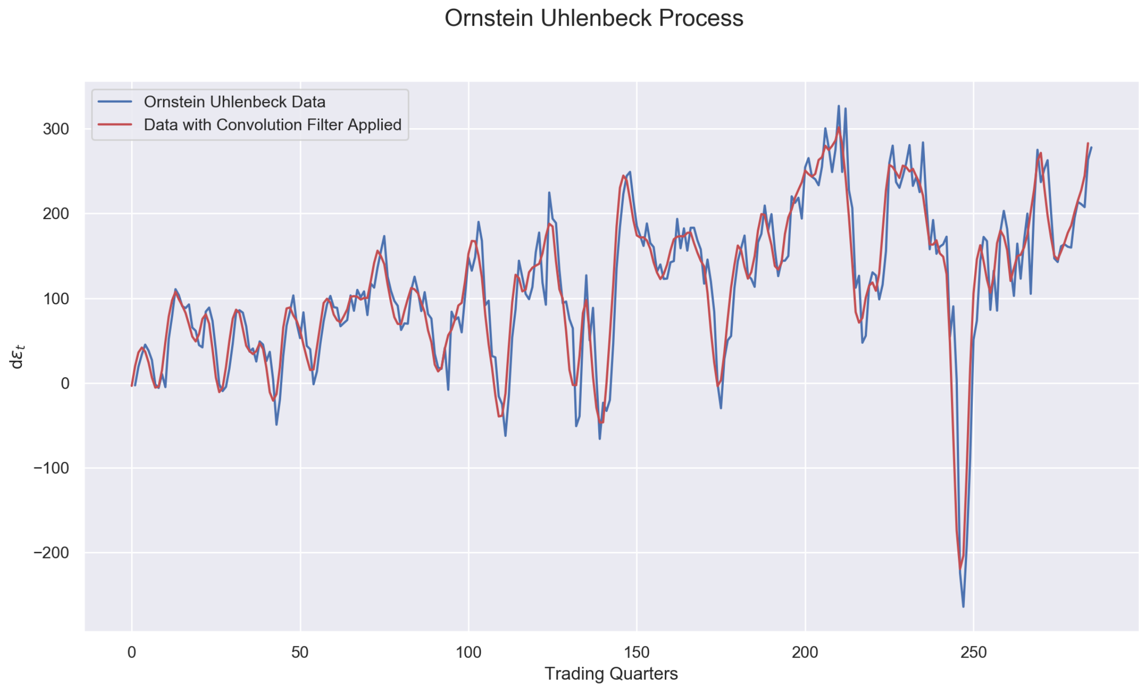

Figure 12b: Filtering of the OU data for the GDP dataset using a Savitzky-Golay filter with 201 convolution points and a fitting polynomial of sixth degree.

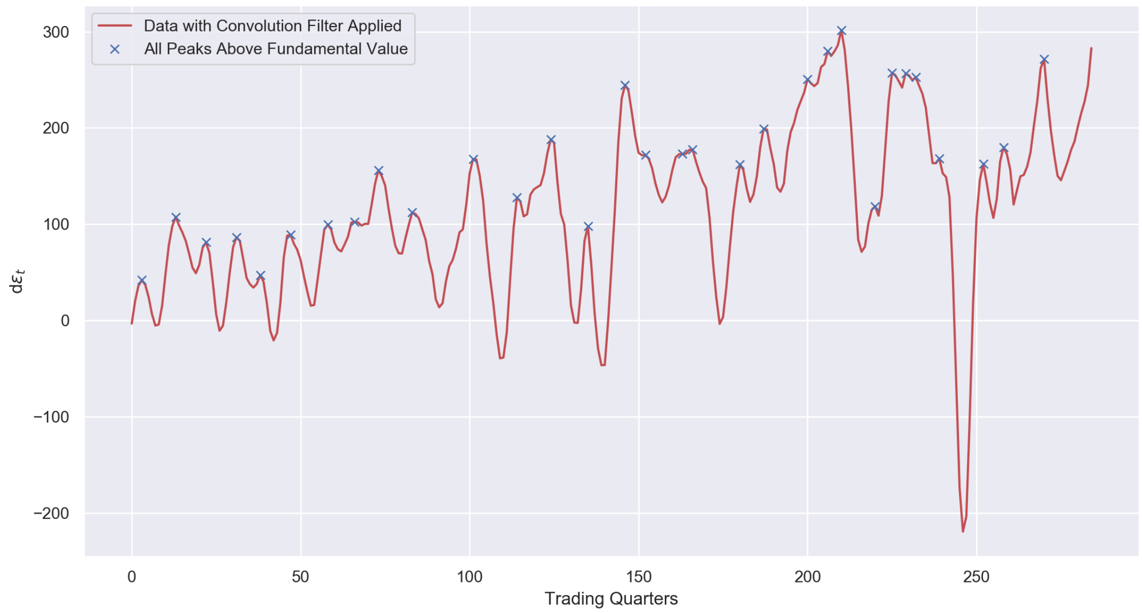

Figure 12c: Figure depicting detection of all peaks above the fundamental value for the GDP dataset with a 201 convolution points and a fitting polynomial of sixth degree.

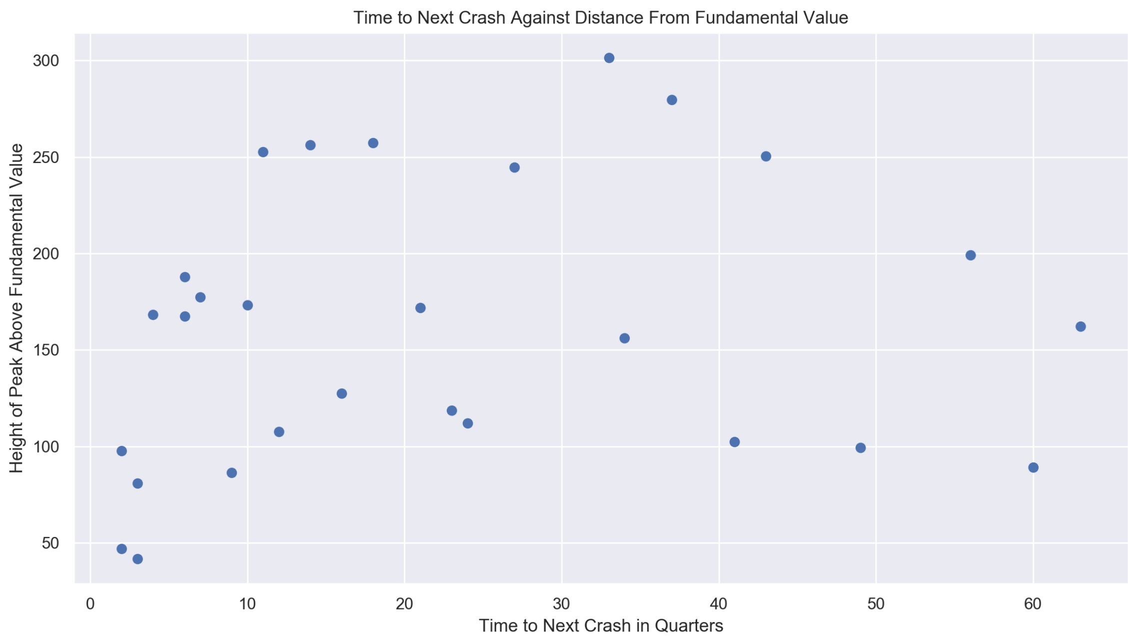

Figure 12d: Graph plotting time to next crash against the height of all peaks above the fundamental value for GDP dataset with 201 convolution points, a fitting polynomial of sixth degree and border cropping of 0.

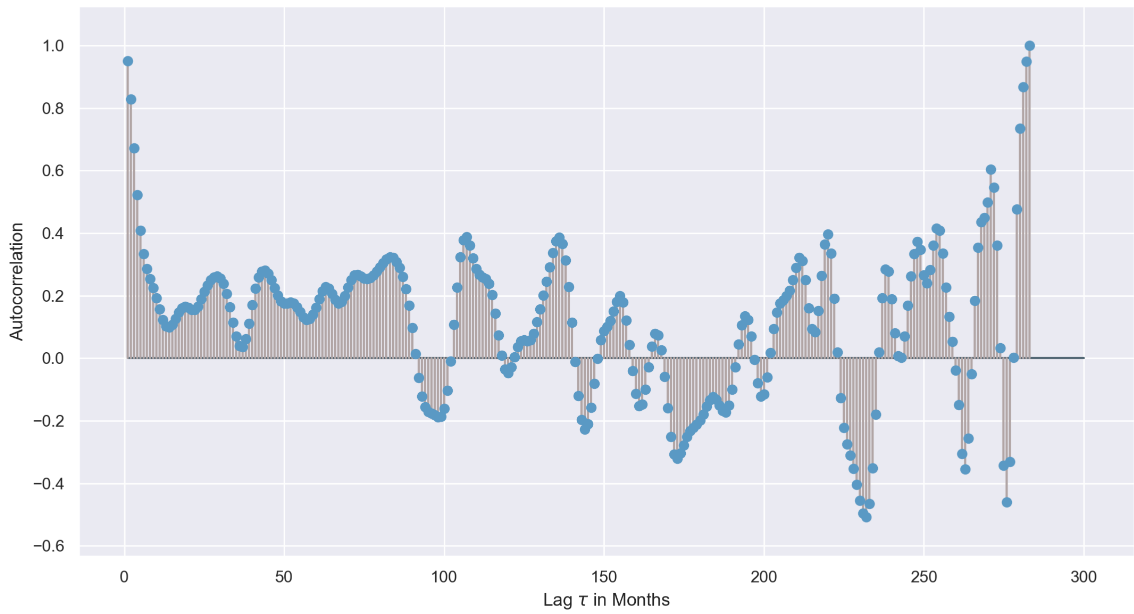

Figure 12e: Autocorrelation plot for GDP over a total of 300 lags. The coefficient is positive for lag less than 900 days and thereafter alternates between positive and negative as the lag increases.