University of Cape Town

FTX5003W
Minor Dissertation in Finance

For
Master of Commerce
Specialising in Finance in the field of Investment Management
(CMO31FTX07)

Using Deep Learning to Characterise Weak Signals in Global Equity Markets: a Case Study of COVID-19

By
Keegan G. Clarke
(CLRKEE001)

Supervisor: Associate Professor Chun-Sung Huang

University of Cape Town

September 2023

Plagiarism Declaration

1. I certify that I have read and understand the Commerce Faculty Ethics in Research Policy. http://www.commerce.uct.ac.za/Pages/ComFac-Downloads
2. I certify that I have read the General Rules and Policies Handbook (Handbook 3) regarding Student Rules of Academic Conduct: RCS1.1 to RCS3.2 and Rules Relating to examinations G20.1 to G22.2.
3. I certify that I have read and understand the document, "Avoiding Plagiarism: A Guide for students".
4. This dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.
5. This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed, and has been cited and referenced.
6. I authorise the University of Cape Town to reproduce for the purpose of research either the whole or any portion of contents in any manner whatsoever.

Student Number	CLRKEE001
Student Name	Keegan Garreth Clarke
Signature of Student	Signed by candidate
Date	10 September 2023

Abstract

This study examines the use of deep learning to identify and characterise anomalous events and their preceding weak signals in equity price data. Particular interest is placed on Gray Rhino events, indicated by the presence of progressively stronger signals prior. The market behaviour prior to and during the COVID-19 pandemic on G-20 equity markets provides a useful context to this end. Existing literature has examined the effects of the pandemic on these markets but has yet to provide conclusive insights into the development of the major equity crash. We compare the existing literature concomitantly to our rigorous application of event study methodology, identifying the presence and effects of signals prior to the market crash. In addition, we develop and deploy a novel Anomaly Characterisation Process (ACP). The ACP utilises an ARIMA time-series model to transform equity price time-series for the extraction of relevant information, whereby subsequent fits of GJR-GARCH and deep-undercomplete-autoencoder models are deployed. Resultantly, measures of dispersion and atypicality are produced which allow for effective and clear characterisation of the degree of typicality of the equity prices and their movements. This innovative method demonstrates efficacy in detecting both point and contextual anomalies. When applied in the context of COVID-19, the findings suggest that different event types can be distinguished successfully with this novel approach through the identification of weak signals. Notably, these insights of the ACP in conjunction with those of the event study suggest that the COVID-19 market crash is consistent with a Gray Rhino event and not a Black Swan event. We briefly demonstrate that these insights can be used by market participants to improve risk-adjusted returns via ACP-informed risk-mitigation techniques.

Acknowledgements

I must express my deepest gratitude to my family and friends for their forbearance and support during this time. I am also grateful to Associate Professor Chun-Sung Huang for his expansion of my capacity for thoughtful analysis. To MG Ferreira is tremendous gratitude for insights into mathematical modelling and codification which proved invaluable to my success. Moreover, towards Simon Sylvester and Rezco is immense appreciation.

Finally, I am most grateful to God by whom "*all things were made*".

Table of Contents

Table of Figures

Table of Tables

1. Introduction

The negative impact of COVID-19 (COVID) and consequent risk mitigation responses are broadly known. For equity investors and researchers it is unknown whether the market rout it triggered was foreseeable ex-ante. To this end, the COVID-19 pandemic has been commonly referred to by the Black Swan metaphor, a term coined by Taleb (2004, 2007) to describe an "outlier" or rare event with "extreme impact" that is only retrospectively predictable. Used to describe the 2008 Global Financial Crisis or GFC (CFI, 2020, Swango, 2020), it is since frequently used in classifying analogously called six sigma events[1]. Taleb (2007) characterised unforeseeable Black Swans from comparable extreme events that are "model-able", defining the decisive difference as predictability. Placing extensive emphasis on outliers of a random variable unaligned with any smooth density function, Hammond (2016) implied limitations of conventional modelling techniques in this pursuit. Sornette (2009) alludes to the limitations of trying to predict Black Swans, by definition unpredictable in type, using probabilistic risk assessments on historical data. It is preferable to forecast types that can be modelled utilising current signals and warnings (Glette-Iversen and Aven, 2021). Further references discussing similar extreme events in financial markets are noted in Pace and Calabrese (2021) and Arian et al. (2020).

Ansoff (1975) described what he termed weak signals, defining them as "warnings, events and developments which are still too incomplete to permit an accurate estimation of their impact and/or determine their full-fledged responses". Rousseau et al. (2021) define it as forewarning that will in due course develop into an event, discussing difficulty in identification since their presence is oftentimes obscured amongst noisy data. A further challenge is the rarity of weak signals alongside a potentially prodigious degree of uncertainty. Weak signals may or may not develop into strong signals or a trend. The tactical ambiguity lies in determining when sufficient information has been assimilated for action to be taken, versus waiting for greater certainty and risking waiting too long (Rousseau et al., 2021). With a view to resolution, Ansoff (1975) suggests applying a form of continual monitoring and having a progressive, emerging response – reminiscent of the adaptive risk analysis of Glette-Iversen and Aven (2021). However, questions regarding the resolution of this problem remain.

The concept of weak signals is incorporated in the metaphor Gray Rhino, a term defined as high impact events occurring after a series of neglected or ignored warnings – coined by Wucker (2016). Approaching from a distance, the Gray Rhino is not acted on until it has arrived with detrimental consequences. The question arises as to whether there are extreme events whose weak signals are not acted on due to incomprehensibility or whether an absence of such weak signals actually exists (and thus the metaphor moot). Considering this uncertainty, deliberation exists as to whether COVID-19 pandemic was indeed a Black Swan or a foreseeable event. Amongst the proponents of the latter is Taleb (2020), who argues that the event was both predicable and foreseeable.

Literature referring to Gray Rhino phenomena include Ferguson (2020), Shore (2020) and Zheng et al. (2021). However, the Gray Rhino metaphor in the specific topic of equity price time-series is yet to be identified. Per definition, the COVID equity market rout is classifiable as a Black Swan event only if it was unpredictable. The presence of preceding weak signals suggests Gray Rhino classification to the exclusion of a Black Swan. Hence, identification of the preceding weak signals is fundamental to this distinction. As Gray Rhinos may be acted upon in anticipation, they provide opportunities for profit and prevention of losses.

[1] Examples of Black Swans include the Black Death pandemic (1347-1351), World War I (1914-1918), the Stock Market Crash of 1929 and the Great Depression, Black Monday of 1987, the Asian Financial Crisis of 1997, the Savings and Loan Crisis during 1986-1995, the "Dotcom" crash of 2000-2002 and the Great Financial Crisis of 2007-2008.

Coupling the Gray Rhino metaphor with the definition of weak signals leads to the necessity of defining an outlier or anomaly. Hawkins et al. (2002) provide a classic definition of an outlier: "an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism". Assuming these weak signals follow this definition then any such outliers present in equity markets are generated by a different mechanism. Since changes in equity prices are affected by transactions of the stock, this mechanism is likely the action of some market participants upon information that the majority does not (yet) consider relevant. The detection of outliers or anomalies relates to the identification of weak signals, defined by Cao (2021) as the quantitative characterisation, classification and forecasting of unusual and dynamic patterns related to constituents of financial markets. Thus, by identifying anomalies we distinguish weak signals and thereby determine Gray Rhino events.

Therefore, we investigate and characterise weak signals in global equity markets antecedent to the COVID equity market rout (COVID rout) in order to establish if it is Gray Rhino event. To do so we utilise event study methodology whilst concurrently developing and deploying a novel anomaly characterisation process to effectively characterise price movements according to their typicality, and thus identify atypical behaviour (including weak signals). Providing ourselves with a broad basis, we sample the G-20 countries' equity markets and focus the event study analysis upon the COVID rout itself and the period immediately preceding. In contrast, we deploy the developed anomaly characterisation process upon samples of larger time periods[2]. The research questions are as follows:

- What are the effects on the global equity markets of key information concerning COVID-19 up until and as it is declared as pandemic?
- Can the event study methodology identify any weak signals in the equity price series as anomalies?
- Can the typicality of equity price data be effectively described by an anomaly characterisation process?
- Can an anomaly characterisation process detect the anomalous price movement in the data prior to the COVID-19 crash?
- Can an anomaly characterisation process be used to detect the weak signals of a Gray Rhino event in the equity price series?

Chapter 2 encompasses a literature review, beginning with the impact of COVID-19 on equity markets, followed by event study methodology. Thereafter, investigation into more advanced techniques of anomaly characterisation and detection informs the development of the anomaly characterisation process. Chapter 3 delineates the research methodology of our study, including sections on both event study and the anomaly characterisation process. Chapter 4 reports our results and their analysis, with their discussion subsequent in Chapter 5. Finally, Chapter 6 contains our conclusions and recommendations for further study.

[2] Refer to Section 3.1 for explicit definition of the samples.

2. Literature Review

In this Chapter, we begin with review of COVID's impact upon equity markets, followed by examination of event study methodology. The evaluation of machine learning approaches in anomaly identification concludes.

2.1 The Impact of COVID-19 on Equity Markets

The COVID pandemic (COVIC) significantly impacted the globe. The World Health Organization (WHO) announced the first case located in Wuhan, China on 31st December 2019. The Chinese New Year exacerbated the viral outbreak, rapidly spreading to other cities in China. Poor understanding of the limited information available pertaining to the new virus initially caused the severity of the public health risk to be underestimated, as well as the underestimation of the infectiousness of the virus (Mohapatra et al., 2020), resulting in exponential increase of COVID cases (Al-Qudah and Houcine, 2021) . On 30th January 2020 WHO declared a Public Health Emergency of International Concern or PHEIC (GLOPID-R, 2020), and after further deliberation finally declared a pandemic in a press release on 11th March 2020 (WHO, 2020). The ensuing panic resulted in containment and preventative measures such as partial or full lockdowns being implemented in most countries. Naturally, this had pronounced impacts upon much of the world's equity markets. These ramifications sparked significant interest in the literature.

Our review pays particular attention to those that follow the form of event studies. Since the novel work by Fama, Fisher, Jensen and Roll (1969), the event study methodology has been used many times in finance. Illustrating both the appeal and ability of the method, it is still used contemporaneously. This methodology involves isolating the effect of a particular event expected to have some positive or negative influence on the price of a financial security. This is observed in returns deviating from expected returns (should the effect not have occurred), and thus are a result of the market accommodating the event's information.

Typically, a market model is specified whereby these abnormal returns are subsequently calculated as the deviation from returns projected by the model - the expected returns (Peterson, 1989). These are oftentimes cumulated over a specified event window to determine an estimate of the total effect. Mathematical expression of these terms is included in Section 3.2.1.1.

Liu et al. (2020) employed event study methodology to explore the effect of COVID on 21 stock market indices utilizing daily closing prices for the period 21 February 2019 to 18 March 2020. They studied panel data and derived expected returns through market models estimated via the ordinary least squares regression method. The Dow Jones Industrial Average (DJIA) was used as the market proxy. The event day was chosen as 20 January 2020, where it was first public knowledge that the disease could be transmitted amongst persons. The event window consists of a total of 35 trading days with 5 subdivisions, with model estimation periods set to lengths of 120, 150 and 180 trading days. Their results suggest that the COVID outbreak had a significant negative effect on all studied markets. Asian countries experienced a greater magnitude of the effect, and a positive relationship between the number of confirmed cases and the abnormal returns. This suggests a market contagion effect, whereby a fairly localised market impact is exported to global equity markets. Such an effect may be exploited as a weak signal for Gray Rhino events. Furthermore, Liu et al. (2020) found increased volatility of stock market returns across the analysed panel. The negative yet heterogenous impact on the global equity market was confirmed by other studies below.

He et al. (2020) studied the effects of COVID on the Chinese stock market, paying special interest to industry sectors. The event day was specified as January 23, 2020 – the day Wuhan began its period of closure from the rest of China. The estimation period was set as 160 days prior to the event and the event window was set as 5 trading days around the event date (totalling 11 days).

Certain sectors such as transportation, environmental and mining industries experienced a large negative effect but other sectors, such as manufacturing, information technology and public management, experienced a strong positive effect. Such heterogenous impact may signal that some market participants acted on new information previously unavailable or obscured. The heterogeneity suggests either the effect is localised to the group of interest, or most participants have not yet accurately interpreted the implications. Further investigation into heterogeneity of industry groupings in the global market is required.

Observing another emerging market, Polemis and Soursou (2020) assessed the immediate impact of COVID on Greek energy companies. Findings suggest that the majority of the companies under observation exhibited a rapid, short-term response whereafter they returned to normal behaviour with most of the significant Abnormal Returns[3] (AR) being seen prior to the lockdown as investors acted on their expectation of it. However, when examining the Cumulative Average Abnormal Returns (CAAR) for the 3 windows, all CAARs calculated were insignificant. This suggests that the interpretation of obscured information presents with different responses, with some participants acting on the information and others not. Polemis and Soursou (2020) employed a market-model, whereby they used 141 observations (trading days) for parameter specification. Their event day was the day of the announcement of the lockdown of Greece, 23 March 2020. They specify a primary event window of 10 days around the event window for a total of 21 days, from 10 days prior to the event to 10 days after. To account for further possible information leaks, they specify two secondary event windows around the same event date, of 20 days and 50 days respectively, with totals of 41 and 51 days respectively. With day zero indicating the event day, these event windows range from day -20 to day 20 and day -50 to 50 respectively.

Singh et al. (2020) applied traditional event methodology to equity indices representative of the G-20 equity markets, finding evidence that all markets reacted negatively to COVID – except Turkey. When measured across all markets the impact progressively developed larger negative magnitudes. Moreover, an unfolding pattern of negative effects was felt first in emerging markets, leading to a consensus whereby the developed markets record sudden sharp declines of large magnitude. Their event date is selected as 20 January 2020, with the event window 58 days prior to and including this date, subdivided into four 10-day, one 11-day and one 7-day subperiods. The estimation window was the 150 days immediately prior to the event date. Their cumulative average abnormal return results show that there was an initial negative reaction, which is curiously followed by minor positive abnormal returns. Thereafter, a much larger negative reaction occurs for a longer subsequent period, suggesting that some market participants reacted to the news first, followed by the remaining participants later.

Using MSCI indices, Harjoto et al. (2020) assessed the effects of COVID in terms of two events, an economic shock and an economic stimulus. The former is proxied by the 11 March 2020 WHO announcement classifying COVID as pandemic and the latter by the Federal Reserve Bank announcement on 9 April. These two events' effects were examined across developed countries excluding the United States, the United States in isolation, emerging markets, and large cap and small cap US firms. They used the S&P500 Index as the market proxy for US large cap and small cap, and the MSCI ACWI equity index for all other groups (6 groups). Their primary conclusion is that the COVID shock was of a greater negative magnitude in Emerging markets than in Developed markets. On removing the US from the developed market group, however, they find that Developed ex-US experienced a greater negative response. Finally, US Small cap eclipsed the negative shock of US Large cap groupings. This provides further evidence that unusual price movements in different markets may be weak signals.

[3] Mathematical definition of event study terms is present in Section 3.2.1.

Zhang et al. (2020) measured the volatility of daily returns (in terms of standard deviation) from US, Italy, Mainland China, Spain, Germany, France, United Kingdom, Switzerland, South Korea, Netherlands, Japan, and Singapore stock markets. Daily data was taken up to 27 March 2020, and the WHO declaration of the pandemic on 11 March 2020 is utilized as day 0. They find a positive relationship between the number of cases and the level of daily volatility in the markets. For all markets (excluding mainland-China) a harmonic mean increase of approximately 1.9x was experienced between February and March 2020. China presented with above average volatility for the month of February, but this was of a lesser magnitude as the remaining markets. This localised volatility increase may be a weak signal. They also present a change in market connectedness after this announcement, measured as before and after correlation heatmaps. Post announcement, Eastern Asian and South-eastern Asian markets increased in connectedness, but this grouping decreased in their connectedness with all other markets, suggesting relational changes that may be exploited in order to detect weak signals. The researchers also made use of graph theory's minimum spanning tree (MST) to examine changes on this sample, which corroborates the clustering of the Asian markets post-announcement. In terms of connectedness, they find evidence that, prior to COVID, the US and the Chinese mainland and Singapore markets are relatively disconnected from each other in terms of behaviour.

Al-Awadhi et al. (2020) use a modified approach to study events by utilizing panel data regression controlled for firm-specific characteristics. They measure the pandemic's effect on the Chinese stock markets of the Hang Seng Index (HSI) and the Shanghai Stock Exchange Composite Index (SSE). They find a stronger negative relationship between daily stock return and the growth in cases of COVID than in the growth of deaths, however both are significant. This provides evidence of market participants acting on unclear information in a manner which can be detected through the rate of change i.e., returns. Utilising the Industry Classification of Listed Companies[4], they find that information technology and medicine manufacturing outperformed the remaining market. In contrast beverages and the transportation sectors experienced a greater negative effect than the remaining market during the viral outbreak.

Further evidence for heterogenous event effects is found by Yan and Qian (2020), who examined the effect of COVID on consumer stocks through event study methodology upon the Shanghai Stock Exchange Consumer 80 Index. They conclude the onset of COVID had a "significantly negative impact on consumer stocks at the very beginning of the crisis, especially on the second day…", with no further evidence after 15-30 days. They suggest this short-term nature is a result of market intervention by government alongside resilience displayed by consumer equities in China (arising from several structural differences, such as strong online shopping platforms and mature logistics systems). This short-term rapid response may be exploited in detecting signs of a Gray Rhino event. Their experiment is designed with 22 January 2020 chosen as the event day, primarily due to it being the first official conference by the State Council Information Office, but also due to the penetration of COVID into 24 of the 31 Chinese provinces by this date. Their estimation period is 140 trading days prior to event day. Following the usual zero indexation of the event day, the authors set event windows from day -2 to day 2, day -5 to day 5, day -10 to day 10, day -15 to day 15 and day -30 to day 30. Abnormal returns are estimated using the constant mean model, whereby the arithmetic average return is estimated over the estimation period. The authors present both the ARs and CARs, but only apply t tests to the CARs.

Most recently, Al-Qudah and Houcine (2021) follow the methodology of Liu et al. (2020) applying it to the six WHO regions, including the countries with the highest number of positive COVID cases in each WHO region in the earlier phases of the pandemic. The estimation period was an average of 220 trading days per region, with an average post-event period of 129 trading days (subdivided into 5 subperiods). The event date was specified as 21 January 2020. Similar to the

[4] As defined by the China Securities Regulatory Commission (CSRC).

aforementioned studies, they likewise find significant negative effects from COVID. However, they observe equity markets in the Western Pacific region to react more rapidly relative to the remaining regions, enduring a greater cumulative fall. Furthermore, the researchers find that the reaction of the South African stock market to COVID was delayed in comparison to USA, China, and Russia. This provides further evidence that reactions to COVID were not uniform.

For convenience, Table 6 in Appendix 1 provides a summary of methodological specification of the contemporaneous literature employing event study methodology. This leads the reader to a discussion of event study methodological literature.

2.2 Event Study Methodological Review

2.2.1 Background and Definition

In the generic Market Model approach to an event study, expected returns are calculated after training a linear model upon data preceding the event. Any anticipated effect of the event is to be observed within the event period (Peterson, 1989). This avoids information leakage whilst limiting the effect of model inaccuracy; if the model is unreasonably inaccurate just prior to the event then it is unlikely that any event effects detected through it will be accurate. Any inaccuracy may be cumulated, further necessitating the limitation of the length of the event window. Once the effects of the event have been estimated, parametric and nonparametric tests are applied to determine whether these estimates are valid estimates of the effect.

2.2.2 The Market Model

The market model assumes that there is an observable market portfolio (similar to CAPM) for the estimation of the market return $R_{m,t}$ and thereby an estimate of the systematic risk β_i for security i. This model is typically estimated using Ordinary Least Squares[5] (OLS). Furthermore, the model assumes that it sufficiently captures most of the 'normal' returns for the security so that abnormal returns can be separated. Violation of the OLS assumptions present challenges in assessing the efficacy of hypothesis test results. This is problematic for event studies as the return generating process is mis-specified and thus event effects may not be accurate (Al-Athari and Al-Amleh, 2016).

We correct for any potential violation of OLS assumptions by deploying Generalised Least Squares (GLS), accounting for undesired structure in the linear regression model's residuals (Aitken, 1936). This will also help in economic markets whose systemic risk is not fully captured in a single factor. In addition, we deploy nonparametric test statistics to provide further rigor in the assessment of statistical significance.

2.2.3 Key Considerations in Application

The event study methodology has faced several criticisms, particularly regarding improper implementation resulting in flawed research design. Circumnavigating this, care must be taken to ensure that (1) the three assumptions are fulfilled[6] and (2) the research design specification does not contain errors in its implementation (McWilliams and Siegel, 1997). We use McWilliams and Siegel (1997) explanation of the assumptions. These are (i) market efficiency, (ii) events being unexpected by market participants (so that effects can be attributed to the event), and (iii) controlling for confounding effects. Key considerations in this respect include sufficient sample size whilst providing an explanation of abnormal returns that is linked to the researcher's original hypotheses – failure of which may bring into question the rigor of deductions.

Tests for Estimated Effects and Event Induced Variance

Financial event studies regularly utilise the classical approach of hypothesis testing. Generalising, the null hypothesis is set to 'there is no effect from the studied event E', and the alternative set to 'there is an effect from event E'. As per common knowledge that most financial data is non-normally distributed, often exhibiting excess kurtosis and skewness, the use of Z-tests in parametric tests (MacKinlay, 1997, McWilliams and Siegel, 1997, Peterson, 1989) has typically been abandoned for the Student's t-distribution (Corrado, 2011) since it is less affected by distributional deviances. Per Seiler (2000), even basic yet correct application of the event study method is sufficiently capable

[5] The OLS specification of linear regression assumes normality of errors, homoscedasticity, variables that are multivariate normal, lack of or low levels of multicollinearity, estimation of coefficients is using outlier free-data, and a linear relationship between dependent and independent variables (DeFusco et. al. 2004, Al-Athari and Al- Amleh, 2016)

[6] This does not exclude some being reasonably relaxed.

at providing the desired insights – in the absence of event-induced variance. Of all the tests they assessed, the nonparametric rank and standardized cross-sectional methods are recommended due to fairly high levels of resilience to this problem. This latter parametric test follows a student's t-distribution.

In addition, prolonged event windows may result in difficulty controlling for confounding events whilst making reconciliation with the assumption of market efficiency obstreperous (McWilliams and Siegel, 1997). Ryngaert and Netter (1990) show empirically that the effect of an event will typically be captured by a brief event window. If information leakage is expected, then the event window should begin prior to the event to account for it. The length of this should be justified, as well as the length of any sub-windows.

2.3 Machine Learning and Characterisation of the Anomalous

To detect anomalies, we must first define what they are and review the literature relevant to the detection of their presence within time-series. As it is of interest to the study, we pay particular attention to those involving deep learning. Consequently, additional definitions and specification is provided for clarity.

The financial domain has applied machine learning (ML) for many years. It has been one of the primary sources of quantitative techniques for finance, for example Ordinary Least Squares regression and its derivations are techniques classified as machine learning. Artificial neural networks (ANNs) are a technique of machine learning recently extended by deep learning (DL)[7], which consists of multiple layers of ANNs. Consequently, single layer ANNs are labelled as shallow networks. Contemporaneously, deep learning appears to hold the greatest potential for large-scale innovation across finance[8]. As the domain of machine learning is broad, we have limited our study to the domain of DL and its application to a financial problem. Specifically, the characterisation of anomalous price movements in order to characterise Gray Rhino events and distinguish them from other kinds. In order to do so, techniques are learnt from their application in other domains.

2.3.1 Definition of an Anomaly

The concept and definition of anomaly can be traced back to Kuhn (1970) as "New and unsuspected phenomena" interrupting the process of "normal science". An outlier can also be defined as an instance unlike the consensus of the data (Chandola et al., 2009). These agree with contemporaneous literature (Hawkins et al., 2002, Chandola et al., 2009, Chalapathy and Chawla, 2019) and we utilise the definition of Hawkins et al. (2002) as our primary definition of an anomaly[9].

Anomaly detection is clarified by Pang et al. (2021) as the operation of detecting data that is materially different from the greater dataset. They also provide description of the unique challenges faced by anomaly detection processes, which we summarise: First, the anomaly and its characteristics are unknown until it occurs. Second, they are rare. Anomalies are thus heterogeneous – they are not only dissimilar to the consensus but also towards other anomalies. However, this does not indicate they are not categorizable (see Section 2.3.2). Nevertheless, this heterogeneity remains within these sub-categories. Finally, the rarity of anomalies regularly translates into severe class-imbalance between anomalous and typical data.

[7] For a brief discussion overview of ML in finance, refer to Section 2 of Ozbayoglu et al. (2020), and Section 3 for extensive discussion regarding various applications of deep learning.

[8] For example, Cao (2019) and PWC FSI (2018) predict future success of investment firms and investment professionals on their ability to strategically incorporate AI (DL) techniques.

[9] There are several aliases for 'anomaly': 'outlier' 'discord', 'novelty', 'discordant observations', 'exceptions', 'aberrations' and 'surprises'. The aliases of preference in this work are the noun 'anomaly', and adjectives 'atypical' and 'anomalous'.

Generally, the detection of anomalies requires first knowing what is regular or typical to the data and then to define what constitutes sufficient deviation in characteristics so as to not be part of the pervading distribution and thus be an anomaly. We bring notice to the fact that anomaly detection is first a process of data characterisation and then only can the actual detection occur. Furthermore, this characterisation is essentially a determination of how typical or atypical the data is in comparison to the remaining data i.e., what is typical. Therefore, characterisation of Gray Rhino events will necessarily require a process that characterises anomalous data effectively.

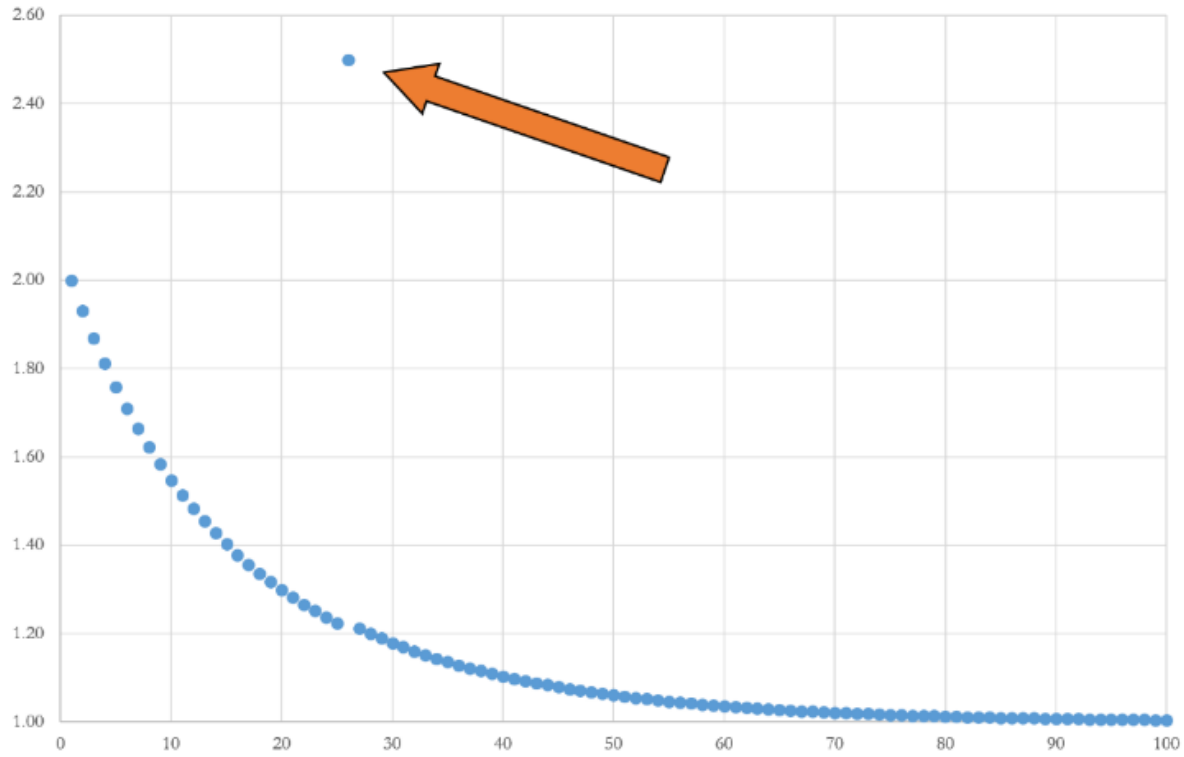
2.3.2 Types of Time-Series Anomalies

There are several broad classifications of anomalies in time-series data with inherent overlap between the categories.

Point Anomalies

Much of the literature tends to concentrate on point anomalies. As evident in Figure 1, these are single instances that deviate from the data's consensus (Chandola et al., 2009). Commonly, they signify random irregularities which lack interpretation (Chalapathy and Chawla, 2019) though this is not necessarily the case.

Figure 1: Example of a Point Anomaly



Contextual or Conditional Anomalies

Contextual anomalies are instances that do not suit the setting they lie within, are conditional on the surrounding instances and would not be anomalous if placed within a different context (Song et al., 2007). Context is typically defined according to the time and space where the anomalous instance features (Chalapathy and Chawla, 2019), but may also be defined according to their 'behavioural features'. For example, the events of 1987's Black Monday as narrated by Bernhardt and Eckblad (2013) may signify a contextual anomaly in that its behavioural features were entirely dissimilar from the bull run of the previous 5 years.

Collective or Group Anomalies

Certain instances typical a singular basis may become anomalies when grouped with other instances (Chalapathy and Chawla, 2019). Some forms of market manipulation represent collective anomalies, for example, 'wash sales' whereby the buyer and seller are the same market participant

with no change in ownership (Putniņš, 2012). What would otherwise be entirely non-anomalous is anomalous due to placement by the exact same party.

2.3.3 Statistical and Traditional Anomaly Detection

Of the approaches to detect such anomalies, there are numerous statistical methods. The majority model the typical behaviour by estimating a statistical distribution that the empirical data may follow. Anomalies are then detected through a threshold of distributional deviance using labelled anomalies (Buda et al., 2018). However, we exclude these from our analysis due to limitations arising from the parametric nature (Ji et al., 2021) as well as the potential lack of suitability arising from class imbalance[10].

2.3.4 Learning Types

In contrast, we may deploy ML based anomaly detection. Three general types of machine learning anomaly detection (Buda et al., 2018) are discussed below.

In supervised learning, labelled examples (i.e. anomaly or non-anomaly) are inputted to model whereby it must learn the characteristics associated with the given labels (Smith and Frank, 2016). This type often follows the approach of training a classifier requiring that every instance in the (large) training dataset be accurately labelled (Buda et al., 2018). Frequently used approaches include K-means clustering and other explicit classification approaches (Ji et al., 2021). Creation of this dataset is problematic, and deep supervised classifiers tend to display suboptimal performance when there is class imbalance (Chalapathy and Chawla, 2019). Alternatively, semi-supervised learning entails creating a training dataset that is partially labelled (Chalapathy and Chawla, 2019). These models are trained to classify typical data (Buda et al., 2018).

Unsupervised learning does not need a dataset that is labelled, due to the defining assumption that most data points are normal (Patterson and Gibson, 2017, Buda et al., 2018). According to Chalapathy and Chawla (2019) they have been shown to surpass other methods, such as support vector machines, principal component analysis and isolation forests, whilst found to be more effective than other types in anomaly detection (Zhou and Paffenroth, 2017, Chalapathy et al., 2017). Furthermore, financial returns data are often stochastic and thus structural components can be construed by noise and non-stationarity (Giles et al., 2001). Considering the nature of the problem is to characterise anomalies, by definition rare and heterogenous in form, and to do so amongst noisy data, unsupervised learning is therefore more appropriate by forgoing the need to label the dataset whilst having been demonstrated to have efficacious results. The question arises as to what kind of unsupervised learning is appropriate.

2.3.5 Deep Learning

A common direction is to detect anomalies by distinguishing between different instances through classifying a single class or one-class classification (Pang et al., 2021). For example, one-class support vector machines is a technique broadly deployed (Schölkopf and Smola, 2002). However, this kind of shallow learning approach is largely surpassed by the more powerful deep learning models.

Chalapathy et al. (2019) combine a neural network with the one-class approach to detect anomalies (OC-NN), outperforming other approaches in complex image and sequence datasets. Furthermore, several feature extraction models combined with simple one-class models have appeared (Andrews et al., 2016, Erfani et al., 2016). More contemporaneously, deeper forms of neural networks as well as their extension recurrent neural networks (RNNs) have garnered much attention in the deep-learning anomaly detection arena. In the multivariate space, Tuor et al (2017b, 2017a) applied deep simple recurrent neural networks for insider threat detection and cyber

[10] Class imbalance occurs when a specific class has substantially less instances than the remaining classes.

anomaly detection, respectively, followed by Assendorp (2017) who applied LSTMs to detect anomalies in industrial machine's multivariate sensor data. Others applied various configurations of LSTMs on various forms of multivariate time-series data (Nucci et al., 2018, Hundman et al., 2018, Nolle et al., 2018). This illustrates the potential of using neural networks in anomaly characterisation.

A common approach is to apply the autoencoder (AE) model-concept in some form, such as Kieu et al. (2018, 2019) and Fu et al. (2019) who used grouped convolutional denoising AEs to detect engine faults in aircraft. Prior to this, Zhou and Paffenroth (2017) deployed AEs modified according to principles of robust principal component analysis (RCPA). A natural development of this is to combine the AE concept with a RNN (Guo et al., 2018). For example, Zhang et al. (2018) developed and tested three variants of AE-based structures with convolutions, LSTM layers and attention mechanisms. Another variant of AE that is garnering interest takes the form of a variational autoencoder (VAE), such as (Ikeda et al., 2018) whereby anomalies are characterised according to dimensions. However, the difficulty of successfully implementing RNNs remains a practical challenge.

2.3.6 Similar Deep Learning Anomaly Detection Studies

Yeh et al. (2016) mention that similarity search and density-based clustering are the most utilized methods for anomaly detection. Kieu et al. (2019) proposes the use of neural network based AEs in time-series data whereby the input is condensed into a hidden representation by the model and then reconstructed (replicated), with the reconstruction being compared to the inputs data[11]. Anomalies or outliers are ignored in this process, enabling their identification through increased model error. Findings show that a single AE also advances precision, and in comparison to ensembles of AEs, often displays superior performance in non-sequential data. Consequently, an adaption was implemented with two "recurrent sparsely-connected neural network autoencoder ensemble frameworks", such that an independent framework coexists with a shared framework of multiple sparsely connected RNN-based AEs. The independent network trains multiple AE separately whilst the shared network trains multiple AE jointly. The reconstruction error is supplied to the multiple AEs for identification of anomalies. Findings showed efficacy, outclassing of controls and newest techniques.

Similarly, Malekia et al. (2021) utilized a deep LSTM-AE to identify anomalies in online time-series streaming data, notably as those above a specific error threshold. They identify and address the problem of evolving anomalies, whereby incidences presently identified as typical may be identified as atypical in upcoming scenarios. The study employed the use of a bespoke algorithm designated Enhanced LSTM Autoencoder (ELSTMAE). The combined objective of ELSTMAE is to reduce noise in the unsupervised data for detection of outliers in online streams of big data and then to specify the amount of anomalous data points. Additionally, the approach employed a probabilistic methodology to identify whether data points were outliers. It proved effective in identifying evolving anomalies, as well as rapid shifts in streaming data.

Niu et al. (2020) researched utilization of LSTM-based VAE generational adversarial network (GAN) to problems encountered when undertaking the mapping of latent space for anomaly identification in streamed industrial sensor time-series data. In this configuration three aspects are trained in unison, namely the encoder with the mapping capacity, the GAN components of the generator and the discriminator. Anomalies are identified by the variance between reconstruction of the VAE component and the outcomes from the GAN's discriminator. Findings show results more accurate and quicker in detecting anomalies in these data types than experimental control models.

[11] More exact description of AEs can be found in Section 3.2.2.2.

Our literature survey revealed a paucity on machine learning derived anomaly detection in financial data for equity stock price characterization purposes. More common are attempts at detecting fraud, such as the detection of money laundering exports fraud through deployment of AEs by Ebberth et al. (2016) or the utilization of LSTM models by Xiao and Jiao (2021) to detect fraud in credit card transaction data.

Nevertheless, the literature survey indicates that the use of an AE-type deep learning model shows great efficacy in characterising and detecting anomalous data.

2.3.7 Autoencoders

According to Aggarwal (2018), dimensionality reduction is intimately associated with the detection of anomalies due to such instances being hard to encode and decode without considerable error. As a result, AEs can be used to detect anomalies. A common method for anomaly detection in non-financial domains (Ozbayoglu et al., 2020), they are considered to be centre of unsupervised Deep-learning Anomaly Detection (DAD) models per Chalapathy and Chawla (2019). Specified as a two-phase process, whereby the first phase, the encoder, performs a kind of dimensionality reduction to encode higher-dimensional data into lower-dimensional representations. This is then decoded by the decoder to give the model's output. Some information is necessarily lost in the encoding, leading to a reconstruction error which can be used to characterise the data according to its degree of atypicality. More typical data points will be reconstructed with greater accuracy than atypical ones. Chalapathy and Chawla (2019) indicate these models assume a high prevalence of normal data in the training dataset, inherently struggling to reconstruct abnormal data – thereby lending themselves to anomaly detection.

Pang et al. (2021) suggests AEs can generally be trained to remember noise, potentially leading to ineffective anomaly detection due to insufficient discrimination between the noise and anomalies– resulting in unusable reconstruction error. Chandola et al. (2009) defines noise as something that negatively encumbers data analysis. It is crucial to address this matter directly as the noise in equity time-series is itself a characteristic of the data (Chen et al., 1986, Moews et al., 2019, Au Yeung et al., 2020). Removal of this noise may thus lead to misrepresentation of the data and prevent effective anomaly characterisation. Therefore, care must be taken when building an AE to avoid the model learning noise.

2.3.8 Dispersion and Anomaly detection

Although our aim is to utilise DL in anomaly detection, modified application of traditional financial tools may likewise prove useful. In portfolio risk management it is common to model the volatility of the assets, forecast the model's expectation and thereby assess the risk that the assets or the portfolio is exposed to (Brooks and Persand, 2003, Andersen et al., 2005). Moreover, many institutions then go on to incorporate this into measures such as Value-at-Risk (Alexander, 2008b). However, these processes are most often focussed on characterising and determining the movement of the prices in terms of their dispersion or volatility. Increased volatility is understood to present increased risk as there is greater chance the price will move unfavourably. Alexander (2008a) provides extensive depth of analysis on techniques that are used practically in performing this volatility estimation and prediction. Routinely, practitioners and researchers alike use Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models to model volatility. Introduced by Engle (1982) and Bollerslev (1986), they capture the time varying dispersion of a returns process in terms of conditional volatility. Per Alexander (2008a) the asymmetric versions of this process generally capture this time-varying volatility with greater efficacy. Of particular interest to us is the GJR-GARCH model, named after its authors Glosten et al. (1993). We suggest this model, that is adept at capturing conditional volatility, can be used alongside the techniques described in previous sections to characterise anomalous data. We expect to utilise this dispersion estimation upon information extracted from equity price data that pertains to the data's typicality

and not the volatility of the prices. Additionally, comparison of the AI techniques with the well-known GJR-GARCH model will serve to provide additional insights.

Having scrutinized the available literature, we move to the employed research methodology.

3. Research Methodology

3.1 Samples, Data and Pre-Processing

3.1.1 Sample construction

In order to ensure broad coverage of major global equity markets, the Group of Twenty (G-20) members were chosen. Spain was also included since it is a permanent guest to the organisation. Table 1 details the benchmark equity indices which were chosen for each G-20 member. Brazil (IBOV Index) was omitted as the data available was not of sufficient quality to utilize in the analyses. Due to its recognisability in literature and in practice, daily-frequency data was utilized for both the Event Study and the Anomaly Characterisation Process.

Table 1: Indices' Identification

B-Ticker	Country	Index Name	Abbreviation
MERVAL Index	Argentina	S&P Merval	MERV
AS51 Index	Australia	S&P/ASX 200	AXJO
SPTSX Index	Canada	S&P/TSX Composite	OSPTX
HSI Index	China	Hang Seng Index	HSI
SHSZ300 Index		Shanghai Shenzhen CSI 300 Index	CSI300
N100 Index	European Union	Euronext 100	EURONEXT 100
CAC Index	France	CAC 40	CAC40
DAX Index	Germany	Deutscher Aktienindex	GDAXI
NIFTY Index	India	Nifty 50	Nifty50
JCI Index	Indonesia	IDX Composite	IDXC
FTSEMIB Index	Italy	FTSE Milano Indice di Borsa	FTSEMIB
NKY Index	Japan	Nikkei 225	NI225
MEXBOL Index	Mexico	Indice de Precios y Cotizaciones Mexico	MXX
IMOEX Index	Russia	MOEX Russia	MOEX
SASEIDX Index	Saudi Arabia	Tadawul All Share	TASI
JALSH Index	South Africa	JSE FTSE All-Share Index	JALSH
KOSPI Index	South Korea	KOSPI Composite Index	KOSPI
IBEX Index	Spain	IBerian IndEX	IBEX 35
XU100 Index	Turkey	BIST 100	XU100
UKX Index	United Kingdom	FTSE 100 Index	FTSE100
INDU Index	U.S.A.	Dow Jones Industrial Average	DJIA
SPX Index		S&P 500	SP500

China contributed 16% and 17% of World GDP in 2019 and 2020 respectively (World Bank Data, 2022) and is the geographic origination of the COVID-19 virus. China also defines regulatory differences between the so-called onshore (A-shares) and offshore (H-shares) Chinese equity (Hui et al., 2013). Consequently, an additional index has been added to increase the samples' representation of the Chinese equity market. Both the Heng Seng Index (H-shares) and the Shanghai Shenzhen CSI 300 Index (A-Shares) have been included in the sample. Tables containing the B-Tickers of all equity securities used in Industry[12] and Supersector[13] groupings are found at

[12] https://github.com/keegangclarke/Estudy/blob/ce8b721d4b3900ad4fb6a50eb9636d3b595e4e7f/sample_data/table_4a_industry_tickers.csv

[13]https://github.com/keegangclarke/Estudy/blob/ce8b721d4b3900ad4fb6a50eb9636d3b595e4e7f/sample_data/table_4b_supersector_tickers.csv

the permalinks provided in the footnotes. Table 2 indicates the equity country classification, per FTSE Russell and S&P, and includes the dates of the first lockdowns in the respective countries.

Table 2: Regional Classification of Markets

B-Ticker	Country	UN Geoscheme Sub-region	Equity Country Classification[14]	First Lockdown
MERVAL Index	Argentina	South America	Frontier	2020/03/19
AS51 Index	Australia	Australia and New Zealand	Developed	2020/03/31
SPTSX Index	Canada	North America	Developed	2020/03/18
HSI Index SHSZ300 Index	China	Eastern Asia	Secondary Emerging	2020/01/23
N100 Index	E.U.	Europe	Not Applicable	2020/03/09
CAC Index	France	Western Europe	Developed	2020/03/17
DAX Index	Germany	Western Europe	Developed	2020/03/16
NIFTY Index	India	Southern Asia	Secondary Emerging	2020/03/25
JCI Index	Indonesia	South-eastern Asia	Secondary Emerging	2020/03/28
TFTSEMIB Index	Italy	Southern Europe	Developed	2020/03/09
NKY Index	Japan	Eastern Asia	Developed	None
MEXBOL Index	Mexico	Central America	Advanced Emerging	2020/03/23
IMOEX Index	Russia	Eastern Europe	Secondary Emerging	2020/03/28
SASEIDX Index	Saudi Arabia	Western Asia	Secondary Emerging	2020/03/09
JALSH Index	South Africa	Southern Africa	Advanced Emerging	2020/03/26
KOSPI Index	South Korea	Eastern Asia	Developed	None
IBEX Index	Spain	Southern Europe	Developed	2020/03/14
XU100 Index	Turkey	Western Asia	Advanced Emerging	2020/04/23
UKX Index	U.K.	Northern Europe	Developed	2020/03/23
INDU Index SPX Index	U.S.A.	North America	Developed	2020/03/19

3.1.1.1 Event Study

The event study analysis uses the daily frequency time series data of the equity-indices' daily closing prices and their constituents' closing prices. Implicitly, these are grouped according to their geographical region – the domicile of the index. For example, the SPX Index is listed on several USA exchanges and contains the 500 largest publicly listed companies in the USA. Consequently, the geographic region and grouping of the SPX Index is the USA. The total sample totals 4157 instruments, grouped according to geographic region. Out of the 4090 instruments of usable quality, 2231 equity securities were able to be grouped according to their Industry Classification Benchmark supersectors. The remaining 1926 stocks were not included due to a lack of ICB classification data. Sample sizes for each supersector, along with an indication of their respective industries, are provided in Table 3.

Table 3: Sample Sizes of ICB Groups

Industry	Sample	Supersector	Sample
Technology	137	Technology	137
Telecommunications	60	Telecommunications	60

[14] The Equity Country Classification is the FTSE Russell classification as at 2020/03.

Health Care	165	Health Care	165
Financials	354	Banks	133
		Insurance	96
		Financial Services	125
Real Estate	146	Real Estate	146
Consumer Discretionary	325	Automobiles and Parts	77
		Consumer Products and Services	92
		Media	27
		Retail	63
		Travel and Leisure	66
Consumer Staples	175	Food Beverage and Tobacco	128
		Personal Care Drug and Grocery Stores	47
Industrials	393	Construction and Materials	118
		Industrial Goods and Services	275
Basic Materials	284	Basic Resources	183
		Chemicals	101
Energy	104	Energy	104
Utilities	88	Utilities	88
Total	2231	Total	2231

3.1.1.2 Event study pre-processing

From the daily closing price data, log-returns were calculated such that the Y variable is the log-return for the closing price P of index constituent i at day t

$$Y_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$$

The corresponding X variable for the beta of the linear regression equation is the log-return of the relevant market j, which was proxied by the relevant index constituent, so that for index price level P at day t

$$X_{j,t} = \ln\left(\frac{P_{j,t}}{P_{j,t-1}}\right)$$

Each index constituent i is contained within a group j, with sample size N. These groups are discussed in Section 3.1.1 Sample Construction, and summarised in Table 2 and Table 3.

3.1.1.3 Anomaly Characterisation Process

The price-level time-series of the indices was used for the ACP – as detailed in Table 2. The total dataset was split into "train" and "test" samples in order to ensure that the model would not memorise the data during the learning process. The model trained its parameters using the "train" dataset; its performance was then tested using the "test" dataset. Performance of the model upon these datasets was compared in order to monitor the learning process and to diagnose any detected problems. This was repeated cyclically until the model failed to improve its performance for several successive iterations whereafter its learning was terminated.

In order to hone the model's sensitivity to anomalies, two known anomalous periods were removed from the training data. These were the periods of the Global Financial Crisis (GFC) and COVID. This was not essential, as the model's undercomplete structure makes it implicitly robust against

learning anomalies. However, there may be marginal improvements in the sensitivity of the model by further reducing the number of anomalous points in the training data. The total dataset or full sample spanned from 2008/01/01 to 2022/02/22. The model was only exposed to data from 2010/01/01 to 2020/01/12.

3.2 Methodology

3.2.1 Event Study

3.2.1.1 Basic Schematic

This fundamentals of the event study methodology were discussed in Section 2.2, and in accordance with this discussion the methodology was modified where appropriate to suit the objectives. Typically, the market model of Sharpe (1964) was used as the return generating process and is specified as:

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \tau_{i,t}$$

$R_{i,t}$ is the rate of return for security i at time t. α_i is the intercept term of security i. β_i is the systematic risk of the security i. $R_{m,t}$ is the return of market portfolio m. $\tau_{i,t}$ is the error term of the characteristic line, with $E[\tau_{i,t}] = 0$. The study made use of the market model for the purposes of calculating the abnormal return and each model was trained according to Ordinary Least Squares (OLS) or Generalized Least Squares (GLS), where appropriate. GLS fortified the accuracy of specification and improved the determination of the expected returns from which the abnormal returns, AR_{it}, were calculated. GLS models were only specified when the OLS showed (1) autocorrelation of error or (2) heteroscedastic errors. This ensured that problems of mis-specified regression models would be avoided. See Section 3.2.1.3 Modifications of Estudy2, below, for further elaboration on procedures and techniques implemented.

The abnormal returns were calculated as the difference between the actual rate of return of security i at time t, and the expected rate of return given its relation to the factors such that:

$$AR_{i,t} = \tau_{i,t} = R_{i,t} - E[R_{i,t} \mid R_{m,t}] = R_{i,t} - (\alpha_i + \beta_i R_{m,t})$$

And with variance:

$$S^2_{AR_{i,t}} = \frac{1}{M_i - 2} \cdot \sum_{i=1}^{h} (AR_i)^2$$

Where, M_i denotes number of non-missing returns for security i, with the estimation period of length h. The 2 accounts for the two parameters in the operation of the market model that must be estimated in order to calculate abnormal returns[15].

Each security's overall reaction to the event was then computed as the cumulative abnormal returns for security i over the event window of length H and is specified below as:

$$CAR_i = \sum_{t=1}^{H} \tau_{i,t}$$

[15] If using a different model, this would need to correspond to the number of parameters of the model.

Abnormal returns were then averaged across a group j in order to increase robustness due to greater sample size[16]. Let N denote the cross-sectional sample size, distinct from the longitudinal length H. Then for group j, the observed effect for day t is estimated as the cross-sectional average abnormal return as follows:

$$AAR_{j,t} = \frac{1}{N}\sum_{j=1}^{N} \tau_j$$

Let E_D denote the event with $\{E_D : D = 1,\ 2,\ 3,\ 4\}$. D denotes a particular information event chosen for the study. The cumulative average abnormal return for event E_D is calculated for group j across the longitudinal event window H as:

$$CAAR_{j,E_D} = \sum_{t=1}^{H} AAR_j$$

$AAR_{j,t}$ and $CAAR_{j,E_D}$ will then be tested for statistical significance.

3.2.1.2 Tests of Statistical Significance

In lieu of the traditionally used t-tests, a parametric and a non-parametric test of significance was utilised per return type. As detailed below, these tests addressed some of the problems associated with the traditional t-tests, such as event-induced variance, which can lead to weak test power generating erroneous results.

Average Abnormal Returns

The parametric test used for AARs was the standardized cross-sectional method, proposed by Boehmer et al. (1991) and reviewed by Seiler (2000). It is robust against additional variance induced by the event; however, it still assumes cross-sectional independence. The null hypothesis for the test is:

$$H_0: \quad E(AAR) = 0$$

This test hybridises the standardized-residual method of Patell (1976) and the ordinary cross-sectional method (Boehmer et al., 1991). The former normalises the residuals and the latter utilises a t-test which employs the event-day cross-sectional standard deviation as its measure of variance, disregarding estimates from the estimation period. Consequentially, the standardized cross-sectional method test statistic Z_{BMP} follows a Student's t-distribution with $h - 2$ degrees of freedom.

The abnormal returns $AR_{i,t}$ are first standardised through an adjusted standard deviation:

$$SAR_{i,t} = \frac{AR_{i,t}}{\text{adjusted } S_{AR_{i,t}}}$$

By incorporating the forecast-error into the standard error $S^2_{AR_t}$, the adjusted standard error of the abnormal returns is:

[16] See Section 3.1.1 Sample Construction for explanation of the groups under study.

$$\text{adjusted } S^2_{AR_{i,t}} = S^2_{AR_t}\left(1 + \frac{1}{M} + \frac{\left(R_{m,t} - \overline{R_{m,t}}\right)^2}{\sum_{t=T_s}^{T_e} R_{m,t} - \overline{R_{m,t}}}\right)$$

Where $\overline{R_{m,t}}$ represents the average of the market returns in the estimation window. Then the summation of standardized abnormal returns $ASAR_t$ on day t is calculated as:

$$ASAR_t = \sum_{i=1}^{N} SAR_{i,t}$$

Where the standard deviation S_{ASAR_t} is the square-root of the variance which is given as:

$$S^2_{ASAR_t} = \frac{1}{N-1}\sum_{i=1}^{N}\left(SAR_{i,t} - \frac{1}{N}\sum_{l=1}^{N} SAR_{l,t}\right)$$

The test statistic for the standardized cross-sectional method for N equities is then given by:

$$Z_{BMP,t} = \frac{ASAR_t}{\sqrt{N}\cdot S_{ASAR_t}}$$

The nonparametric test used for the AAR was the modified rank test of Corrado and Zivney (1992). Through the standardisation of returns, this test is robust to event-induced volatility - unlike Corrado's (1989) original rank test. Furthermore, through implicitly considering cross-sectional correlation, the result of this operation is the calculation of order statistics for a uniform distribution, with an expected value of 0.5. With test statistic $Z_{RANK,t}$, the null hypothesis is given as:

$$H_0: \quad E\left(Z_{RANK,t}\right) = 0.5$$

Abnormal return for security i on day t is first standardised by the standard deviation to arrive at:

$$SAR_{i,t} = \frac{AR_{i,t}}{S_{AR_{i,t}}}$$

The rank $U_{i,t}$ of $SAR_{i,t}$ is then standardised for missing abnormal returns for security i on day t such that:

$$U_{i,t} = \frac{\text{rank}\left(SAR_{i,t}\right)}{(1 + M_i)}$$

M_i denotes the number of non-missing $AR_{i,t}$. The variance of $U_{i,t}$ for estimation period h is denoted as S_u^2 and is given by:

$$S_u^2 = \frac{1}{h}\sum_{t=h_s}^{h_e}\left(\frac{1}{M_t}\sum_{i=1}^{M_t}\left(U_{i,t} - \frac{1}{2}\right)\right)^2$$

Where M_t denotes the number of non-missing returns in the cross-section of N equities on day t. Finally, the test statistic for the modified rank test is calculated as:

$$Z_{RANK,t} = \frac{1}{\sqrt{N}} \cdot \frac{\sum_{i=1}^{N}\left(U_{i,t} - \frac{1}{2}\right)}{S_u}$$

Cumulative Average Abnormal Returns

The parametric test applied is the Crude Dependence Adjustment Test by Brown and Warner (1985). This test utilizes the entire sample for the estimation of variance, thus avoiding its underestimation and improving the test's power. However, this method of construction does not take into consideration the potential existence of unequal variances among observations in the time-series dependence test. The statistic follows a Student's t-distribution with $h - 2$ degrees of freedom. The null hypothesis for the test is:

$$H_0: \quad E\left(CAAR_j\right) = 0$$

For group j, with h_s and h_e indicating the start and end of the estimation window, respectively. t_{CAAR_j} , denoting the test statistic, is given as:

$$t_{CAAR_j} = \frac{CAAR_j}{\sqrt{h_s - h_e} \cdot S_{AAR_j}}$$

With variance:

$$S^2_{AAR_j} = \frac{1}{M-2}\sum_{t=h_s}^{h_e}\left(AAR_{j,t} - \overline{AAR_j}\right)^2$$

With $\overline{AAR_j}$ denoted as

$$\overline{AAR_j} = \frac{1}{M}\sum_{t=h_s}^{h_e} AAR_{j,t}$$

The nonparametric test is the Rank test proposed by Cowan (1992), which was developed as an augmentation of the rank test by Corrado (1989). The test is well-specified in the absence of abnormal returns symmetry (around event date). The distribution approximates a normal distribution; thus, the statistical values are standard normal. Although the test ignores the serial dependence of abnormal returns, it is stable to event-induced variance increases. This test treats the estimation period with number of days h and the event period with number of days H as a single time-series such that $h + H = T$. For this equation, h begins and ends on days h_s and h_e, respectively; H begins and ends on days H_s and H_e, respectively; and T begins and ends on days T_s and T_e, respectively. The statistic is given as:

$$Z_{RANK} = \frac{\overline{K_{E_D}} - \lfloor 0.5T \rfloor}{S_{RANK}}$$

With variance calculated as:

$$S^2_{RANK} = \frac{\sum_{t=T_s}^{T_e}\left(\overline{K_{H,t}} - \lfloor 0.5T \rfloor\right)^2}{T}$$

Where $\overline{K_{E_D}}$ denotes the mean rank across N securities and H days for event E_D, and $\overline{K_{H,t}}$ denotes the mean rank of N securities on day t. The mean rank of T is the expected rank and must be an integer – hence the floor function $\lfloor 0.5T \rfloor$. For example, if $T = 235$ then $\lfloor 0.5T \rfloor = 118$.

3.2.1.3 Software Implementation

The statistical analysis of the Event Study's is performed in R[17]. A survey of available packages disclosed Estudy2[18] to be well suited to the study's application. All event-study-specific parametric and nonparametric statistical tests, as well as some functions such as market model specification, are implemented through Estudy2.

Modifications of Estudy2

To ensure rigor, modifications[19] to the package were warranted. Specifically, the usual process used by Estudy2 to estimate simple linear regression models was enhanced. Ordinarily, parameters are estimated through Ordinary Least Squares (OLS) estimator. However, when specifying a linear regression model, the model must satisfy several assumptions in order to be the Best Linear Unbiased Estimator (BLUE) according to the Gauss-Markov Theorem. If satisfied, then the OLS will provide the lowest sampling variance within the class of linear unbiased estimators. Of these, the two that are of interest in this study are (1) the residuals should not be correlated:

$$\mathrm{Cov}(\tau_i, \tau_j) = 0, \qquad \forall i \neq j$$

and (2) the residuals should be homoscedastic:

$$\mathrm{Var}(\tau_i) = \sigma_\tau^2 < \infty, \quad \forall i \in \mathbb{Z}$$

The Market Model is a simplistic linear model and thus may at times violate these two assumptions, where systemic risk may not be sufficiently captured by single-factor models. Re-specifying each of the Market Models (per the sample in Section 3.2.1) based on the characteristics of each economy and accounting for idiosyncrasies of individual securities would require extensive analysis and additional research. This was unnecessary, however, when the two assumptions (error randomness and homoscedasticity) were tested for violation across the thousands of estimated Market Models. If either one or both of the assumptions were violated, this was then accounted for via the Generalized Least Squares (GLS) estimator (Aitken, 1936). This estimator allows for error autocorrelation and/or heteroscedasticity to be modelled when estimating the regression coefficients such that the linear regression model is still BLUE. Failure to account for violation results in the standard errors of OLS estimates being biased in an unknown direction and unreliable t tests.

As recommended by Lyon and Tsai (1996), the Koenker (1981) studentized Breush-Pagan test is used for heteroscedasticity and the Breusch-Godfrey test (Breusch, 1978, Godfrey, 1978) for autocorrelation. The specification of GLS models was automated by applying these tests to each OLS model, with GLS model specification triggered by positive test results.

3.2.1.4 Events and Event Windows

Event study methodology, as investigated in Section 2.2, assumes market efficiency in its weak-form. In order to reconcile this method with this assumption, event windows are necessarily short in length, aiding in minimising possible confounding effects (McWilliams and Siegel, 1997). Brevity

[17] Our code is available at our GitHub repository: https://github.com/keegangclarke/Estudy.git

[18] https://cran.r-project.org/package=estudy2

[19] The modified code is available at our GitHub repository: https://github.com/keegangclarke/estudy2_gls.git

of event windows is demonstrated to be empirically adequate, however sufficient time for reaction to the informational event is imperative, particularly when the sample contains securities affected by multiple time zones (Ryngaert and Netter, 1990). Per Brown and Warner (1985) and Harjoto et al. (2020) an 11 day event window and a 250 trading day estimation window were deployed. Event window spans 5 days prior until 5 days after event day and the estimation window spans 256 days prior until 6 days prior to event day. All windows are zero-indexed to the event day in question and consider the aforementioned concerns. Figure 2 illustrates event time, where the event window for event 1 is denoted via the square brackets a [-5, 5], and the event day is t=0:

Figure 2: Timeline for Event 1

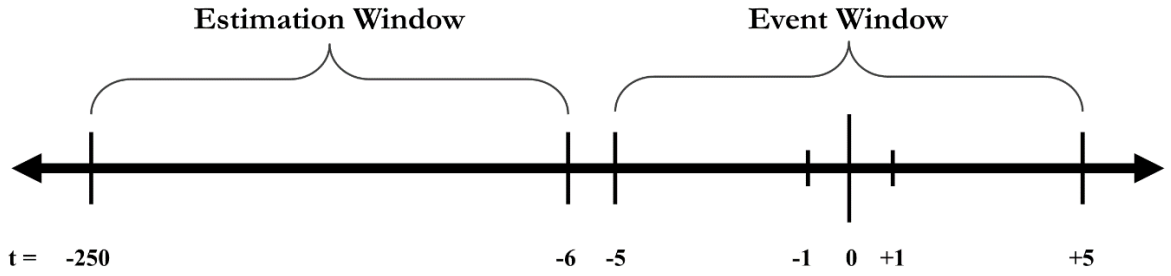


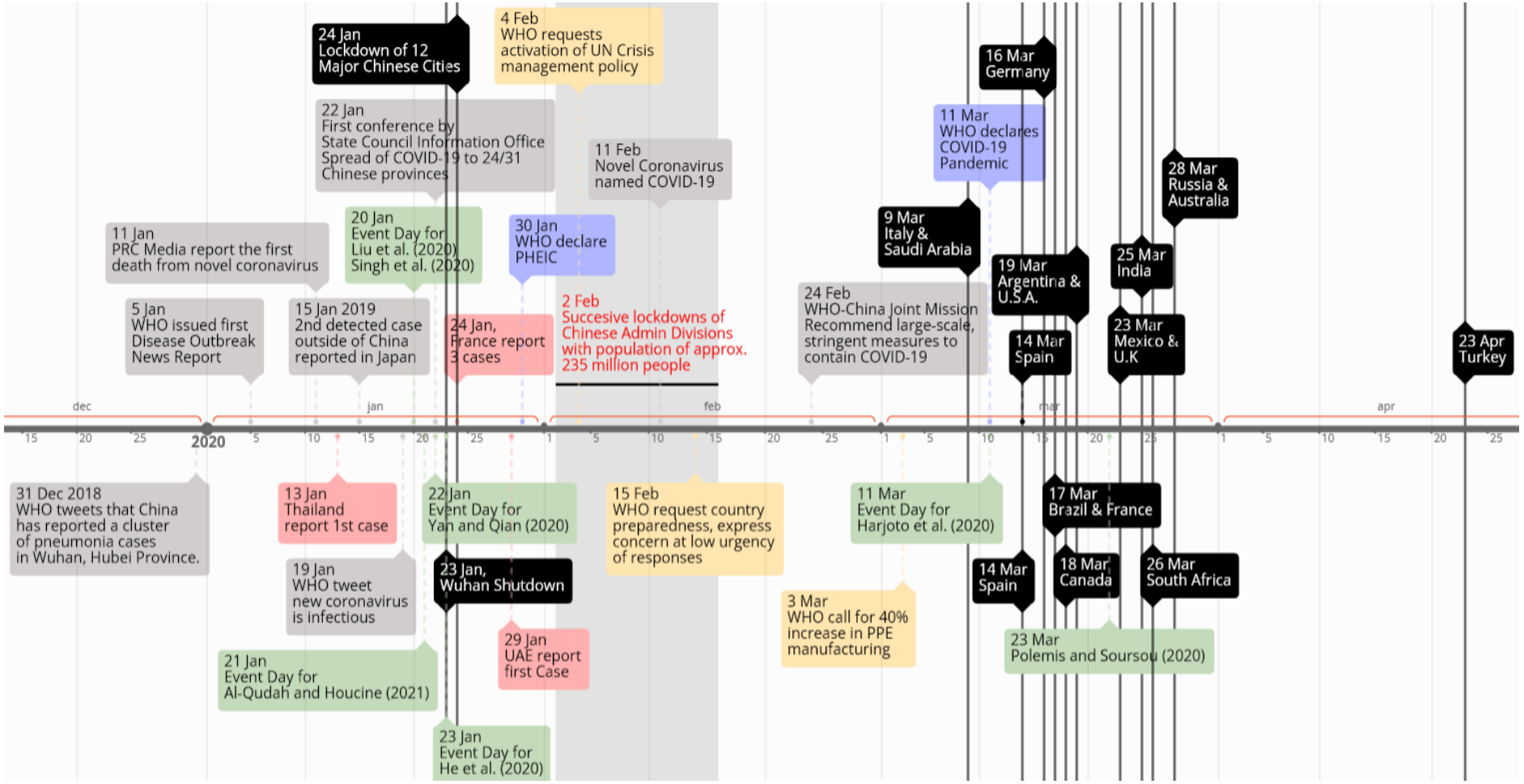
Figure 3 presents a summary of when key information became public[20]. The consensus event day in the literature was the month of January, with the modal day being January the 20th (Liu et al., 2020, Harjoto et al., 2020, He et al., 2020, Polemis and Soursou, 2020, Singh et al., 2020, Yan and Qian, 2020, Al-Qudah and Houcine, 2021). The event window was in general subdivided thereafter. These discussed authors have chosen their event dates with clear consideration of the factors provided. We aim to detect weak signals whilst estimating the effect of COVID-19 across the G-20 economic zones, thus 4 event dates (E_1 to E_4) are assumed. These are in ascending order regarding the likelihood that the event would have signalled the coming pandemic. Premised in this idea is that some market participants treat the events as warning signals and act by changing their portfolios. The observed price-effect from buying and selling provides warning signals to the remaining market participants.

E_1 is defined as 13 January 2020, the day that the first case of coronavirus was detected outside of the Peoples Republic of China (PRC). This is assumed to be the earliest possible indication of the virus, signalling weakly that the virus is likely to spread across the globe. The first reported death on 11 January 2020 is contained within event window and is considered an additional informational event.

E_2 is defined as 30 January 2020, the day the WHO declared a Public Health Emergency of International Concern (PHEIC), the sixth time such a declaration has been undertaken by the WHO. Contained within this event window is 23 January 2020, the day of the first official recognition of coronavirus by the State Council Information Office (SCIO). At this time the virus had spread to 24 of the 31 provinces of China. On 24 January 2020 Wuhan was placed under lockdown. These informational disseminations were signals of the coming impact of the virus a posteriori and thus are considered to be representative weak signals.

[20] A dynamic figure of the timeline can be found at https://time.graphics/line/702377

Figure 3: Timeline of Key Signalling Events of COVID-19



In Figure 3: black boxes with horizontal lines indicate a lockdown, green boxes indicate an informational event, blue indicate key status changes of COVID, red indicate first cases of respective countries, yellow indicate WHO requests which signalled the organisation's expectations, grey indicate information that was not giving a clear signal in itself (and thus requires the corroborating signals to be significant).

E_3 is defined as 24 February 2020, the day the WHO-China Joint Mission began to recommend large-scale, stringent measures to contain COVID-19. This signalled the impact of the virus on global industry and the large shocks then placed on global demand and supply.

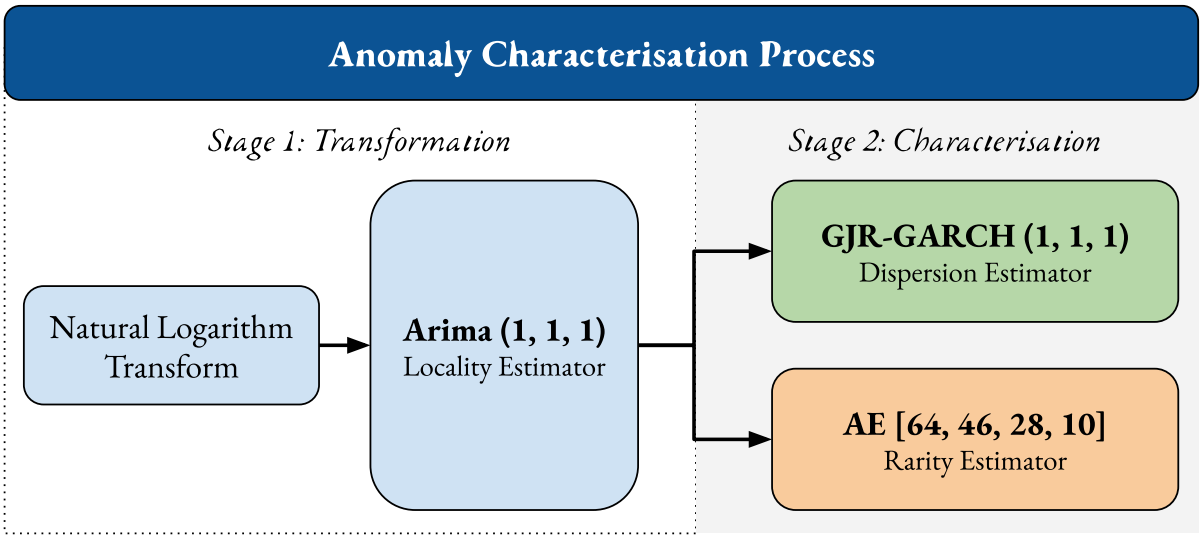
E_4 is defined as 9 March 2020, the day Italy and Saudi Arabia undertook strict lockdown measures following the tone set by the PRC. This can be reasonably assumed to have been a signal to the markets of the upcoming pervasiveness of the disease. Furthermore, the WHO declared COVID-19 to be a pandemic on 11 March 2020.

For convenience, Table 7, containing a summary of the salient facts regarding the event study specification, is available in Appendix 2.

3.2.2 Anomaly Characterisation Process

If a price movement is atypical, then, for this paper, it is determined that the univariate process will be the method by which it is defined as being atypical. This process is defined as the Anomaly Characterisation Process (ACP). The process has two primary stages, as illustrated in Figure 4. Stage 1 involves transforms of the data so that time-varying structure is removed. This is done in order to enable Stage 2 models to characterise the data as anomalous through their different methods.

Figure 4: Diagram of Anomaly Characterisation Process



3.2.2.1 Stage 1: Transformation

Basic transform

The daily closing price-levels P are log-transformed so that for index j at day t, the dependent variable $P_{j,t}$ is expressed as:

$$Y_t = \ln(P_{j,t})$$

This removes the well-known, long-term exponential trend of equity-price series for modelling purposes.

Locality (μ, ε) Estimator: $ARIMA(1, 1, 1)$

An $\mathrm{ARIMA}(1, 1, 1)$ is fitted on Y_t, where according to common convention $p = 1$, $d = 1$, and $q = 1$. The model fits an autoregressive coefficient $\mathrm{AR}(p = 1)$ and a moving average coefficient $\mathrm{MA}(q = 1)$, which capture the growth-trend and smoothed local-trend aspects of the equity-price series (Box et al., 2015). In addition, the data is initially 1st-order differenced $(d = 1)$ such that:

$$y_t = Y_t - Y_{t-1}$$

This differencing provides the absolute change y_t between observations. It is well suited since it is typical of equity-price series to be non-stationary (Schmitt et al., 2013). However, the primary purpose of this differencing is the information it extracts from the prices and as such no test for stationarity[21] is performed. Since these prices have been log-transformed, the price for security j at day t can be expressed as a 'log return':

[21] In addition, consistent utilisation of the same process facilitates neutral cross-sectional comparisons between results of different financial securities.

$$y_t = \ln(P_{j,t}) - \ln(P_{j,t-1}) = \ln\left(\frac{P_{j,t}}{P_{j,t-1}}\right)$$

Hence, the context of the ACP is the rate of change of logged prices. The $\mathrm{ARIMA}(1,1,1)$ process arrives at estimate $\hat{y}_t$ as follows:

$$\hat{y}_t = \Delta + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1}$$

Where ϕ denotes the AR coefficient, θ denotes the MA coefficient, the constant Δ denotes the average change between periods, and the forecasting error ε of day $t-1$ is given as:

$$\varepsilon_{t-1} = y_{t-1} - \hat{y}_{t-1}$$

Through fitting of the model, data is transformed so that the growth-trend, the smoothed local-trend, and the non-stationary structure of the data are removed. The resultant forecasting error ε_t is the transformed data for the remaining models. This relationship is the difference between the expected rate of change and the observed rate of change of the logged prices, and can be expressed:

$$\varepsilon_t = y_t - \hat{y}_t$$

Since:

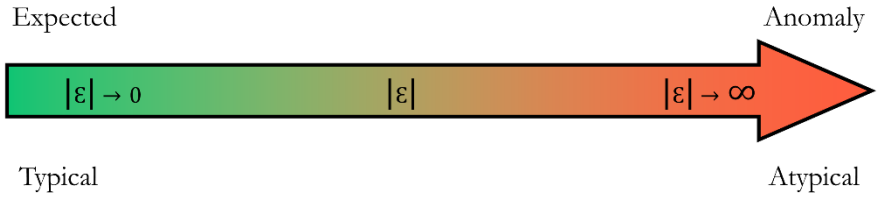
$$\hat{y}_t = y_t + \varepsilon_t$$

And it is assumed that:

$$\varepsilon_t \sim N(0,\ \sigma^2)$$

There are two useful outcomes from this model, (1) the estimated locality μ_j of the price level of security j and (2) its deviation from this trended locality $\varepsilon_{j,t}$. $\varepsilon_{j,t}$ is expected to increase as any given data point of the equity-price time-series deviates from what is typical; thus, the atypical nature of the datapoint is expected to be identified through the magnitude of $|\varepsilon_{j,t}|$. The further away a price-level is from where 'it should be', according to the $\mathrm{ARIMA}(1,1,1)$ model, the more unusual that price-level is in its current time-context. Figure 5 illustrates this description of the characterisation of a price-level in the 'typicality continuum':

Figure 5: Typicality Continuum



A consequence of using this model as a transform is the inducement of additional noise, which can make identification of the atypicality more challenging. However, this illustrates one of the benefits of the Undercomplete Autoencoder for this task as seen in Section 3.2.2.2.

3.2.2.2 Stage 2: Characterisation

Dispersion (σ_ε) Estimator: *GJR-GARCH*$(1,1,1)$

In order to estimate the dispersion on day t of the standardised data $\varepsilon_{j,t}$, the $\mathrm{GJR\text{-}GARCH}(1,1,1)$ volatility process is employed, with the orders of symmetric innovation (p), asymmetric innovation (o) and lagged conditional variance (q) all being 1: $p = 1,\ o = 1,\ q = 1$.

Furthermore, a zero-mean with $\mu = 0$ is assumed. Let σ_t^2 denote the variance of ε_t on day t. The parameter δ_{t-1}^2 represents the square of the estimation error of the model on day $t-1$ and the parameter σ_{t-1}^2 represents the conditional variance on day $t-1$. The model is specified with constant ω and coefficients[22] α, β and λ as follows:

$$\sigma_t^2 = \omega + \alpha\,\delta_{t-1}^2 + \lambda\,I_{[\delta_{t-1}<0]}\delta_{t-1}^2 + \beta\sigma_{t-1}^2$$

Where the indicator function $I_{[\delta_{t-1}<0]} = 1$ if $\delta < 0$ and 0 otherwise, and the parameters are constrained such that $\omega > 0,\ \alpha \geq 0,\ \beta \geq 0$ and $\alpha + \beta < 1$. Long-term variance $\overline{\sigma}^2$ is:

$$\overline{\sigma}^2 = \omega \cdot \left(1 - (\alpha + \beta + 0.5\,\lambda)\right)^{-1}$$

And the one-step-ahead forecast is:

$$\hat{\sigma}_{t+1}^2 = \hat{\omega} + \hat{\alpha}\,\hat{\delta}_t^2 + \hat{\lambda}\,I_{[\delta_t<0]}\,\hat{\delta}_t^2 + \hat{\beta}\hat{\sigma}_t^2$$

Conventionally, when modelling volatility, δ_t (the model's one-step-ahead forecasting error) will typically denote market shocks or unexpected return (Alexander, 2008a). Analogously, δ_t denotes the degree of atypicality according to its absolute magnitude in the ACP. The benefit of using a GARCH process is its ability to capture any anomaly clustering that may be present, as it can capture the volatility clustering of equity prices. Furthermore, the one-step-ahead volatility forecast $\hat{\sigma}_{t+1}$ provides a measure of expected atypicality, since $\sigma_{t+1}^2 = \varepsilon_{t+1}^2$. This will be further examined through plots in chapter 4.

Atypicality (Ω_ε) Estimator: Undercomplete Autoencoder

Unsupervised learning provides the most potential for a study of anomalies in the domain of finance, specifically regarding equity price-series (Zhou and Paffenroth, 2017, Chalapathy et al., 2017, Chalapathy and Chawla, 2019). Employing dense layers, data is characterised and anomalies are detected through the reconstruction error of the autoencoder (AE) reconstruction model (as examined in Sections 2.3.5, 2.3.6 and 2.3.7). In other words, the typicality of a datum is characterised by the magnitude of the AE's reconstruction error. A larger magnitude thus denotes greater atypicality.

For AE with input features x and total number of x equal to the natural number υ, let x_υ be the last feature where there is a series, $1, 2, \ldots, \upsilon$, of features. Since the error ε of the stage 1 $\mathrm{ARIMA}(1,1,1)$ model is used as the source of input features, the υ feature is:

$$x_\upsilon = \varepsilon_{t-\upsilon}$$

The mathematical formalisation of Pang et al. (2021) of the basic formulation of an autoencoder is given as:

$$\{\Theta_e^*, \Theta_d^*\} = \underset{\Theta_e, \Theta_d}{\arg\min} \sum_{x \in \chi} \|x - \phi_d(\phi_e(x; \Theta_e); \Theta_d)\|^2$$

[22] In typical use of the model directly upon log-returns in financial markets, the reactivity of the model to market shocks is captured by α. The lag parameter β measures the persistence in conditional volatility, and the negative leverage effect is captured by λ.

The encoding network, indicated as ϕ_e, performs a dimensionality reduction by mapping the original data onto the abstract representation space Z, also known as the 'latent space' and this map is the 'code'. The code represents a simplified description of the input's ground-truth and must generalise to describing all training data due to the mean-squared error cost-function. From these parameters the decoder ϕ_d then reconstructs the original input features. Thus, Z is a shared layer between the encoder and decoder parts of the model. The parameters of this code in Z are given as $\{\Theta_e^*, \Theta_d^*\}$, where Θ_e^* are the parameters for the encoding layer and Θ_d^* are the parameters for the decoding layer. However, these parameters are ultimately shared between the encoding and decoding layers and thus:

$$\Theta_e^* \stackrel{\text{def}}{=} \Theta_d^*$$

After the decoder has reconstructed the input data from the code layer, the reconstruction error s_x is calculated according to the L2 norm as follows:

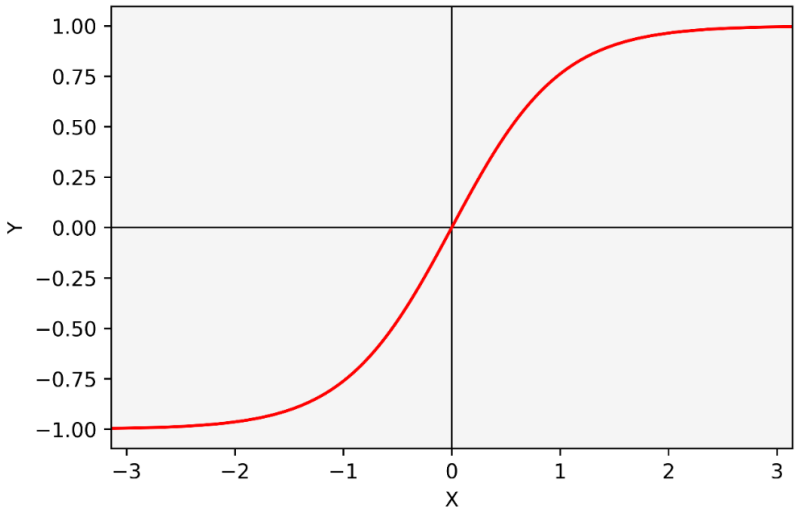
$$s_x = \|x - \phi_d(\phi_e(x; \Theta_e^*); \Theta_d^*)\|^2$$

This is done on the train-data and then tested on the test-data through the backpropagation[23] optimisation process. Due to the imposition of neuron scarcity, $\{\Theta_e^*, \Theta_d^*\}$ will only capture the underlying regularities in the data.

As discussed in Section 3.2.2.1, error ε_t implicitly captures the rate of change between observations and their expected changes. It is reasonable to assume the daily percentage price change of security j to be bounded between -100% and 100%. The minimum and maximums of the data reside within these bounds. Consequently, it is sufficient under the circumstances to use the tanh activation function in the network's neurons. This function aligns with the assumption as it ensures the input and output of the AE is regularised by bounding observations between -1 and 1. Furthermore, the tanh function is most responsive to values nearest to the expected value of 0, resulting from the near linearity in this region of the function's behaviour (see Figure 6). LeCun et al. (2012) observe that tanh enables more efficient backpropagation during training than other options such as the logistic function due to the larger gradients arising from this near linearity. The function graphically represented in Figure 6 is given as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

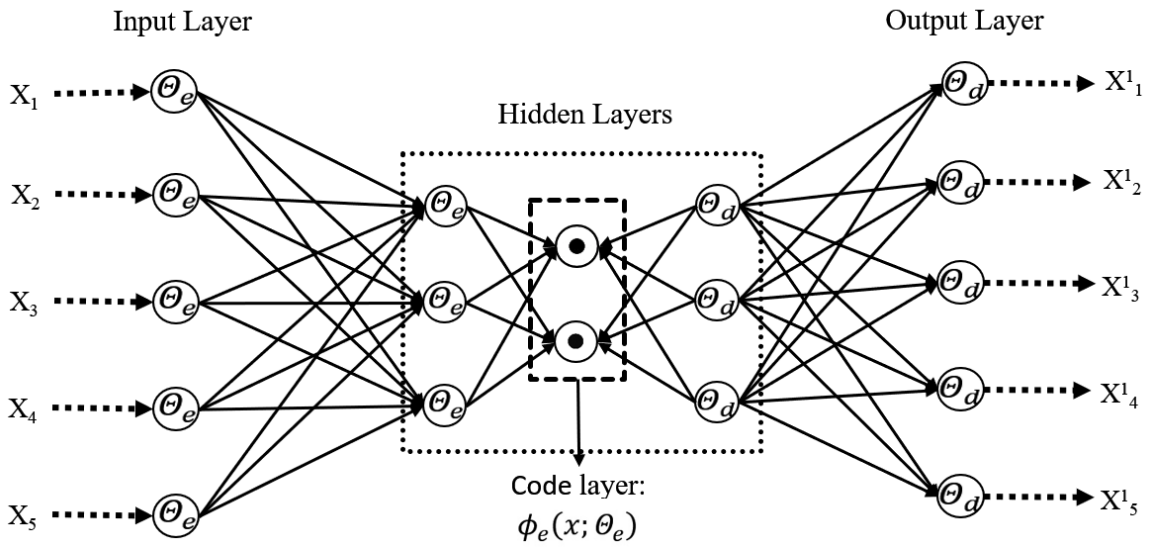
Figure 6: The Tanh Function



[23] Refer to Rumelhart, Hinton and Williams (1986) for detailed analysis of backpropagation.

As indicated by Pang et al. (2021), AEs can be vulnerable to severe overfitting of noisy inputs. We have thus designed the AR's structure to explicitly remedy this potential issue by using an undercomplete autoencoder. The term 'undercomplete' denotes that the code layer is severely constrained to prevent sufficient neurons being available for the model to memorise the noise. This imposed neuron scarcity prevents the model from utilising code-neurons to memorise noise. As anomalies are by definition rare, the code neurons do not learn to accurately reconstruct them. Therefor the magnitude of the reconstruction error indicates the degree of atypicality, arising as a consequence of this neuron scarcity. Consequently, increases in s_x indicate increases in atypicality. The model structure is represented graphically for an AE with $\upsilon = 5$ in Figure 7:

Figure 7: Example of Undercomplete Autoencoder Structure



During preliminary experimentation, it was determined that sufficient width and depth was necessary in order for the model to be suitable for the purpose described[24]. The encoder ϕ_e is given the depth[25] of 3. The width of each layer is given as $64 \rightarrow 46 \rightarrow 28 \rightarrow 10$, where 10 is the width of the code layer and $\rightarrow$ denotes direction of dimensionality reduction. The decoder ϕ_d mirrors the dimensions of the encoder ϕ_e. Table 4 provides the summary model structure as recorded by Keras. The total number of learned parameters is 13 406.

Table 4: Keras Model Structure Summary

Layer (type)	Output Shape	Number of Learned Parameters
flatten (Flatten)	(None, 64)	0
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 46)	2990
dense_2 (Dense)	(None, 28)	1316
dense_3 (Dense)	(None, 10)	290
dense_4 (Dense)	(None, 28)	308
dense_5 (Dense)	(None, 46)	1334
dense_6 (Dense)	(None, 64)	3008

[24] 'Width' and 'depth' refer to the number of input features and the number of hidden layers respectively.

[25] 1 visible layer, 1 hidden layer, and 1 shared code layer.

reshape (Reshape)	(None, 64)	0

There are $\upsilon = 64$ features as the input layer receives 64 lags of ε_t and thus has a width of 64, so that:

$$\{x_\upsilon : \upsilon = 1, \dots, 64\}$$

Characterisation of atypicality is given by the Root Mean Square Error (RMSE) where this error is the reconstruction error and is given as:

$$s'_x = \sqrt{s_x^2}$$

This is a similar metric to that used in the Dispersion Estimator, with similar interpretability. The RMSE is of primary interest to the study and is expected to be more powerful in measuring the atypicality than the Dispersion estimator. As the two models have different strengths, the deviation in magnitude of their atypicality assessments will be of particular interest to the study. Additionally, interest is paid to how additive the atypicality measures are for the shorter-term GARCH and the longer-term AE.

Stage 2 comparative model: *Auto-ARIMA*

In order to provide additional evidence of the efficacy of the ACP's Stage 2, additional ARIMA models are implemented[26] for each equity index, trained on ε_t. The p, d and q parameters for these models are determined through an automated procedure known as automated ARIMA[27]. Numerous versions of automated ARIMA model specification have been successfully utilised for various forecasting applications (Mélard and Pasteels, 2000, Tran and Reed, 2004, Yermal and Balasubramanian, 2017, Jamil and Akbar, 2017, Alghamdi et al., 2019, Alzyout et al., 2019). Following is a brief discussion of the utilised procedure. Determination of the order of d is first undertaken by assessing the time-series' stationarity through the Augmented Dickey-Fuller, Phillips–Perron and Kwiatkowski–Phillips–Schmidt–Shin unit-root tests (Dickey and Fuller, 1979, Phillips and Perron, 1988, Kwiatkowski et al., 1992). Thereafter, the best fitting model is ascertained through a non-stepwise optimisation process whereby model selection is governed by Akaike Information Criterion minimisation (Akaike, 1998). The outcome of this process results in the best fitting orders of the p, d and q parameters being chosen for the specified ARIMA model. For fair comparison, auto-ARIMA's prediction is plotted alongside the prediction of GJR-GARCH model, but auto-ARIMA's absolute error is plotted against AE's RMSE.

3.2.3 Anomaly Categories & Expected Detection

In accordance with the definitions provided in Section 2.3.2, and the ACP system described in Section 3.2.2, the anomalies to be observed are the point and conditional anomaly types. The former concerns a single observation that is anomalous in comparison to the data's consensus; the latter concerns an instance or instances whose atypicality is dependent on the characteristics of the surrounding data. With both types, the typicality of a price movement is conditional on surrounding data and thus we note that point anomalies are inherently a subset of the more general conditional anomalies. Considering the Dispersion Estimator, the conditionality of an observation on day t is predicated on the previous observation $t-1$, suggesting this estimator may be better suited to the characterisation of point anomalies and observation pairs. In contrast, the Atypicality Estimator's estimate on day t is predicated on the sequence of observations from $t-1$ to $t-65$, suggesting it is more appropriate for the detection of conditional anomaly sequences.

[26] Our code is available at our GitHub repository: https://github.com/keegangclarke/acp.git

[27] We utilise a python implementation of the process, provided in the package pmdarima whose documents can be found at https://alkaline-ml.com/pmdarima/index.html

Considering that Black Swan events are by definition[28] unpredictable and arrive without warning, they are expected to present without warning in the ACP derived measures. In contrast Gray Rhino events are, per definition, indicated by an arbitrary length sequence of progressively strengthening weak signals. These are expected to present in a progressive manner in the ACP derived measures. Therefore, we categorise Black Swan events as Category A anomalies and Gray Rhino events as Category B anomalies.

3.2.4 Downside Protection: the ACP as a Potential Trading Signal

In order to test the potential of using systems such as the ACP in portfolio management, a simple rule-trading simulation is proposed. Since many trading systems and risk-metric systems already make use of GARCH models, this test is focussed on deploying the Atypicality Estimator practically. The trading rule is to be long the equity index when the market is behaving 'normally' and to be long risk-free assets when the market is not. 10 basis points per trade are assumed to be lost as fees. The primary objective is to outperform on a risk-adjusted basis. The rule's implications are as follows:

1. Set a threshold R for the AE's root mean squared error recorded (RMSE) based on the AE's training RMSE.
2. Trade portfolio
    a. When $\mathrm{RMSE}_t > \mathrm{R}$ on day t, then on day $t+1$ the portfolio is sold and placed in risk-free interest-bearing assets (long-term treasury bonds) simultaneously.
    b. When $\mathrm{RMSE}_t < \mathrm{R}$ on day t, then on day $t+1$ the portfolio is repurchased, and the risk-free interest-bearing assets are sold simultaneously.

The static threshold is to take the RMSE recorded in the training data and calculate the mean μ and the standard deviation θ, with thresholds μ, 1θ, 2θ and 3θ.

A comprehensive exposition of this study and its findings ensues

[28] See Section 1 for discussion regarding the definition of Black Swan and Gray Rhino events.

4. Analysis and Results

4.1 Event Study

4.1.1 Expected Event Effects

We exhibit our expectation of event effects per the recommendations of McWilliams and Siegel (1997), in order to aid the reader in assessing the efficacy of the results. A priori, the later events are expected to display larger effect magnitudes than the earlier events, with E_4 expected to display the largest effect. Negative (positive) AARs and CAARs are expected in securities that investors perceive to be negatively (positively) affected by the event. As the virus spread globally from its Wuhan epicentre, these are expected to first develop in the more rapidly affected securities - equities situated in the Eastern-Asian and South-eastern Asian regions and in the Travel and Leisure supersector.

4.1.2 Findings

The AAR plots only illustrate the AARs that were at least parametrically significant[29], whilst demonstrating the consensus of these remaining significant AARs through plotting the cross-sectional average. The latter is referred to as the 'average-plot'. The legends of these figures indicate the data's statistical significances according to the typical method of interpretation, such that "NA" is not statistically significant[30], "10%" denotes statistically significant at a 10% level of significance, "5%" denotes statistically significant at a 5% level of significance, and "1%" denotes statistically significant at a 1% level of significance.

Supersectors

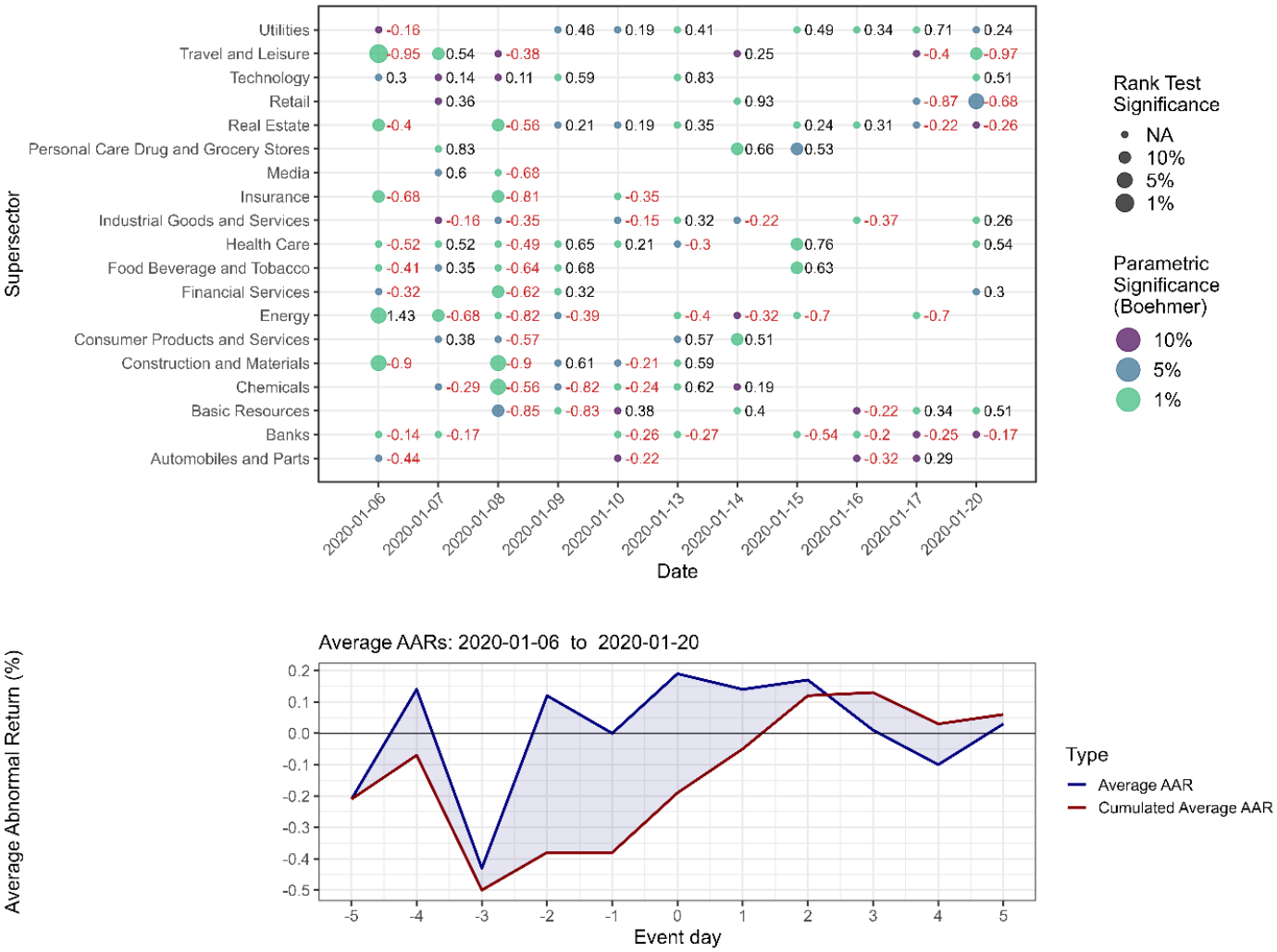
Figures 8 to 11 indicate the progressive impact of COVID on a sample of G-20 indices, organised according to ICB supersectors.

[29] The measure of parametric significance is more sensitive than the non-parametric measure.

[30] As we remove data that is not parametrically significant, "NA" refers to the non-parametric measure.

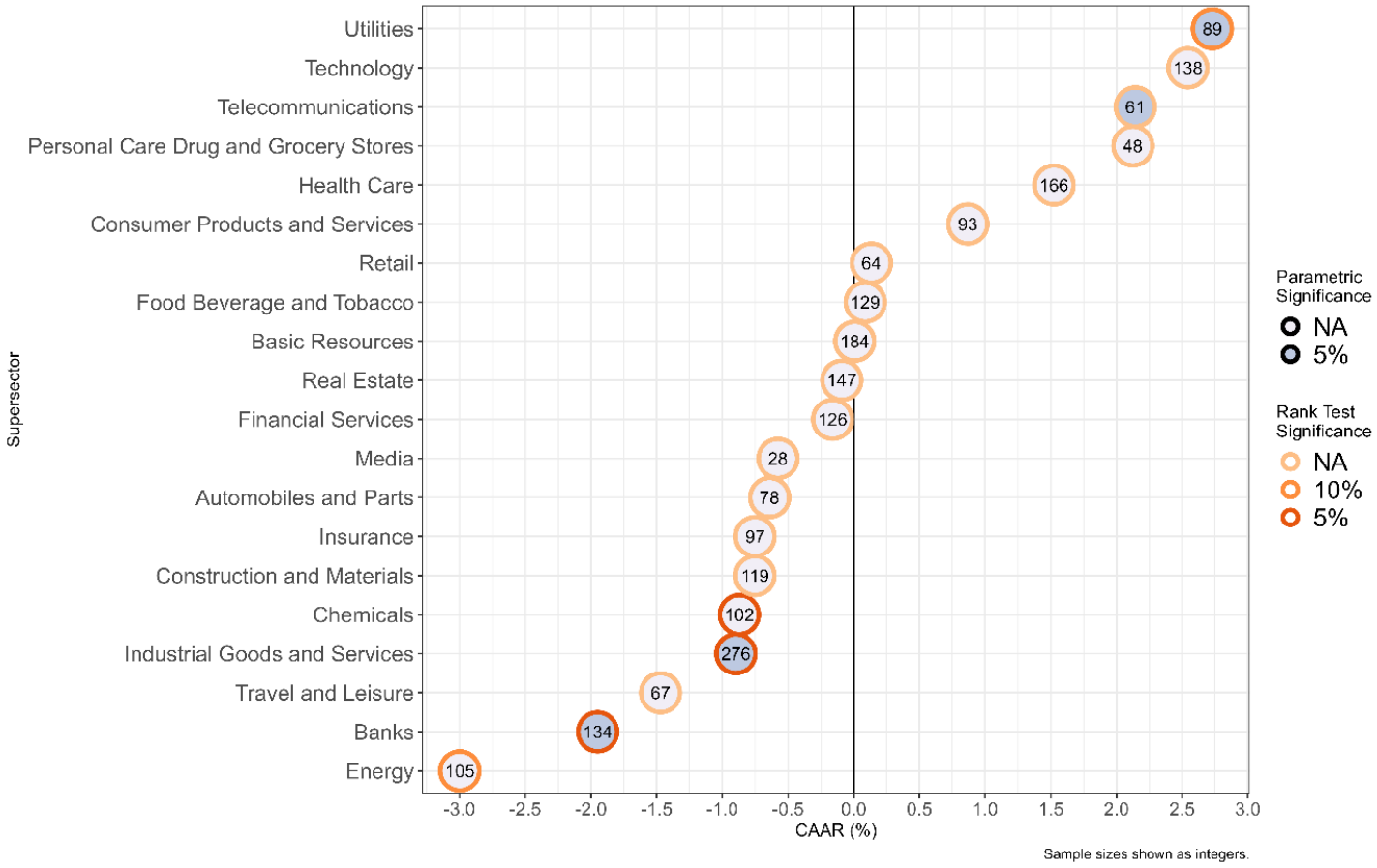
In Figure 8, the effect of E_1 on the 13th of January 2020 is represented by 70% of significant AARs being positive. Conspicuously, the substantial portion of negative AARs prior to day 0 is evident. Of these significant AARs, approximately 62% are negative. Yet after day 0, only 44% are negative. The 8th of January stands out with a nearly uniform negative result. These observations are distinctive in the average-plot. Furthermore, the negative cumulation approaches positivity on day 1, providing support to the notion of a positive response on average. Of the supersector groupings, both Utilities and Technology experienced a majority of significant AARs over the period, however the latter saw only 1 statistically significant AAR after the event day. In contrast, Energy, Banks, and Industrial Goods and Services all experienced a majority of negative AARs over the window, suggesting that these groups were perceived to be negatively affected by the spreading virus. Telecommunications recorded no parametrically significant AARs for the period in question and thus are not present in the Figure. We note that the statistical significance of the AARs prior to day 0 are greater than those succeeding the event day.

Figure 8: Combined AAR plot of Event 1 (grouped by supersector)



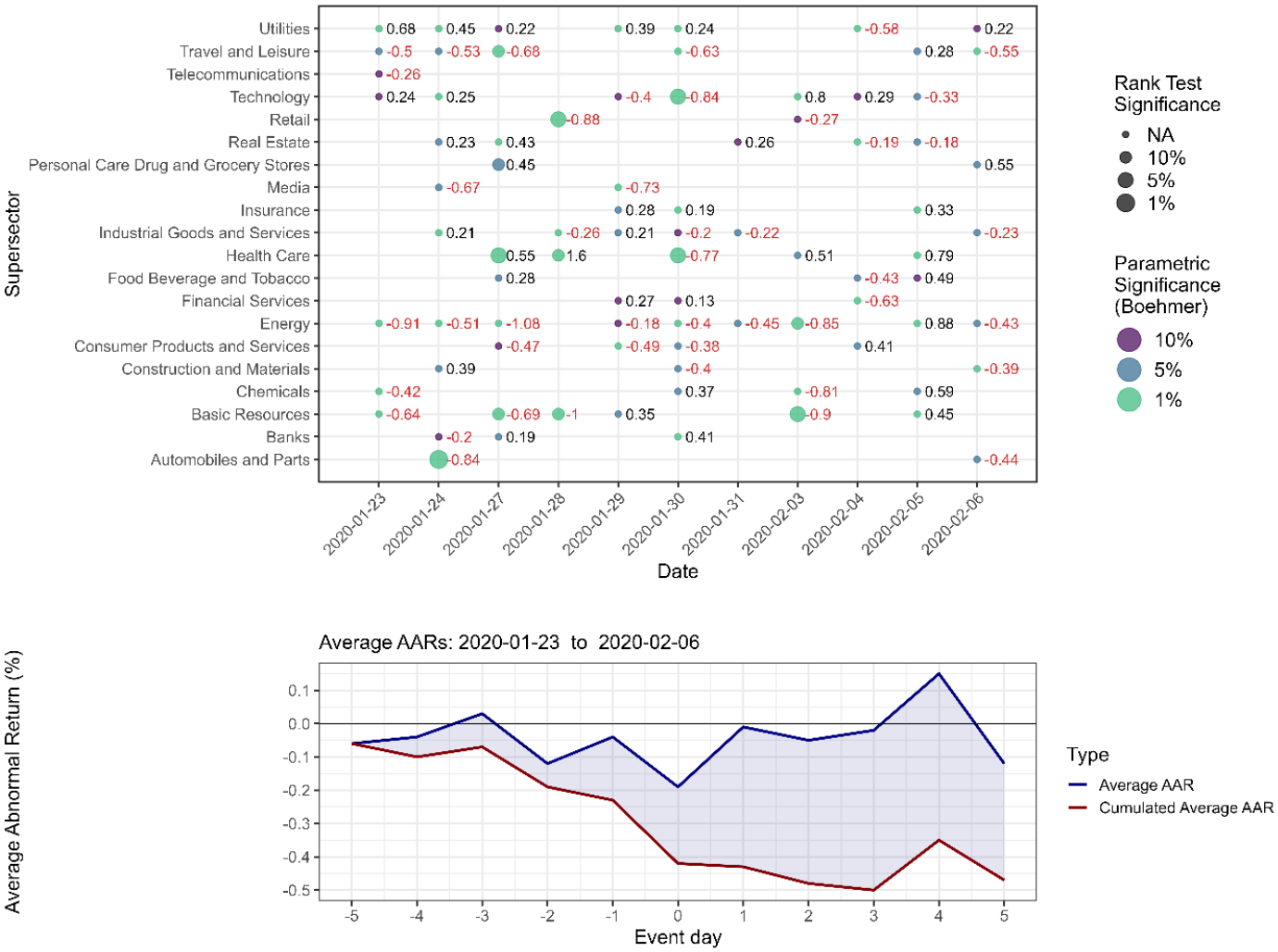
We observe in Figure 9 the cumulative position of each supersector with respect to Event 1. Utilities records a surprisingly strong, positive result. Although Telecommunications observed no statistically significant AARs over the event (hence its absence from Figure 8) we record that its CAAR is positively significant. Similarly, Utilities is cumulatively positive, which stands in contrast to Chemicals, Industrial Goods and Services, Banks, and Energy, all recording significant negative CAARs. The response of Energy appears to be linked to the increased crude oil volatility at the end of 2019, coupled with dropping prices, weak demand for natural gas and tensions in the Middle East (FS Investments, 2020, International Energy Agency, 2020). However, due to the magnitude of the CAAR, we suggest that market participants began to assess the negative implications of the virus on Energy, of which the PRC is the largest consumer.

Figure 9: CAAR plot of Event 1 (grouped by supersector)



In Figure 10, we see this narrative weakening during Event Period 2, with approximately 58% of significant AARs negative on day 0. More significant AARs prior to event day remain, however this approaches numerical equality with the succeeding period. The proportion of negative AARs before and after day 0 has likewise narrowed, being 65% and 55% respectively. Regarding individual supersectors, both Travel and Leisure as well as Energy experienced a majority of negative AARs. The former finds a greater proportion of these prior to and including day 0, whereas the latter records this behaviour on most days. Contrastingly, of Health Care's significant AARs only day 0's is negative. Overall, a larger portion of AARs were not significant for many of the groupings. Returning to the total sample, we observe that the cumulative average of the supersectors shows a negative trend from day -5 to 5, which strengthens around day 0 and reaches its minimum on day 3. Repeatedly, the general significance level of AARs prior to day 0 are of higher significance than those post.

Figure 10: Combined AAR plot of Event 2 (grouped by supersector)



In Figure 11, Health Care's response to E_2 is cumulatively positive. However, Retail's near 0% CAAR is significant according to the Rank test. Although this test is well specified in the presence of unsymmetrical abnormal returns and we observe Retail's unsymmetric response, the near 0% result is inconclusive in its interpretation. Notable is the absence of significance amongst 18 of the 20 supersectors. Considering the AAR-related observations, we suggest this response is evidence of the mixed responses to the event in question.

Figure 11: CAAR plot of Event 2 (grouped by supersector)

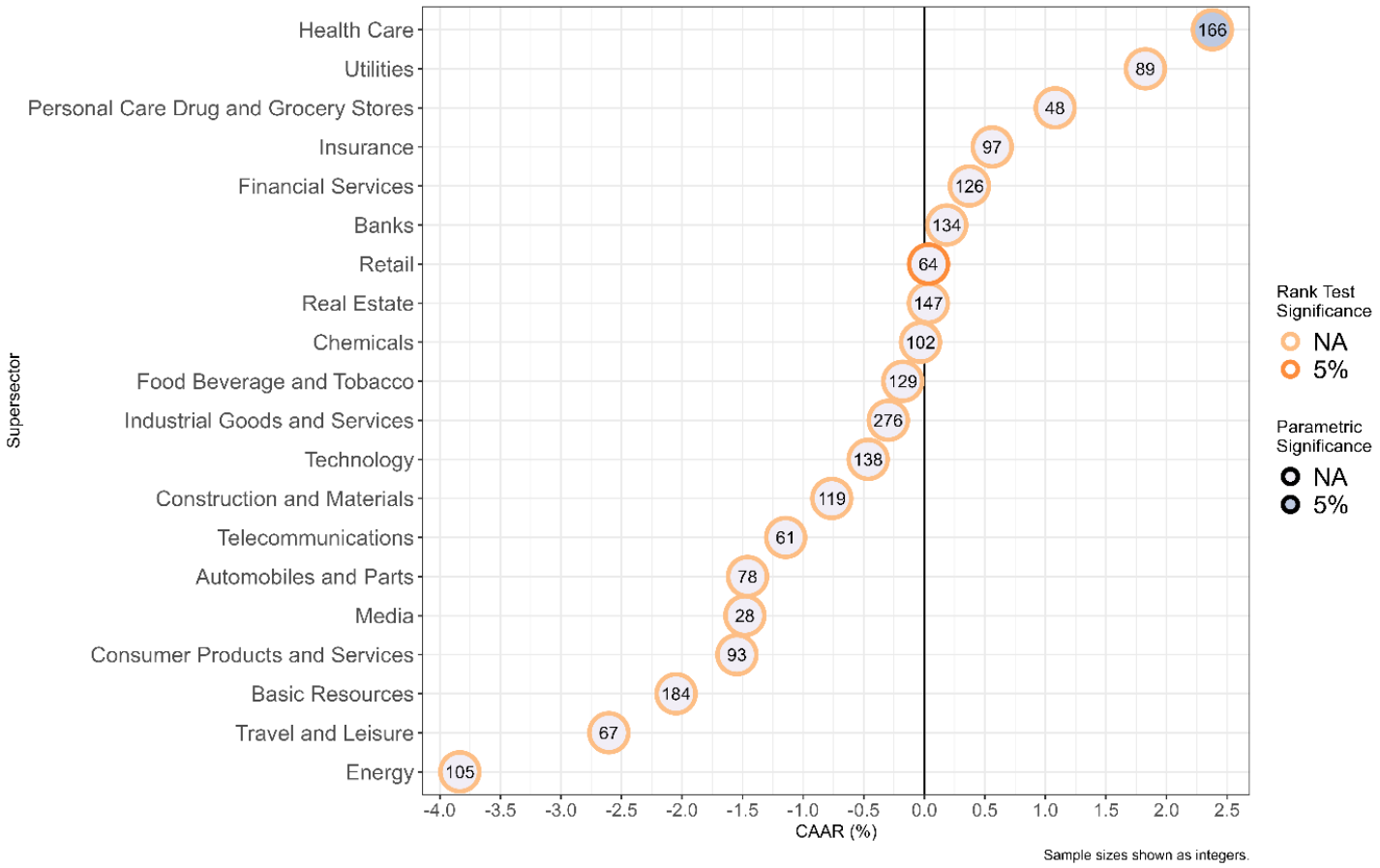
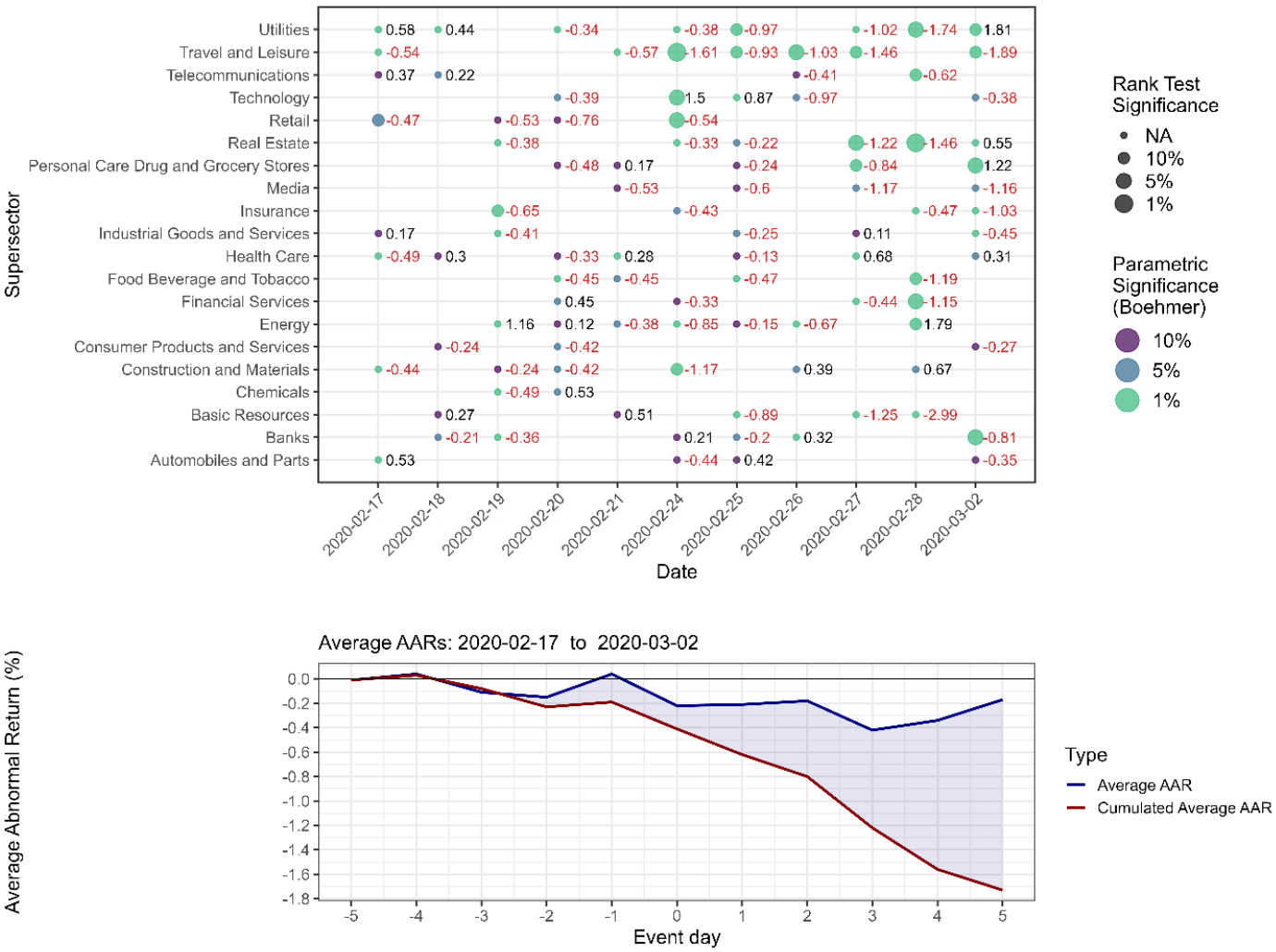


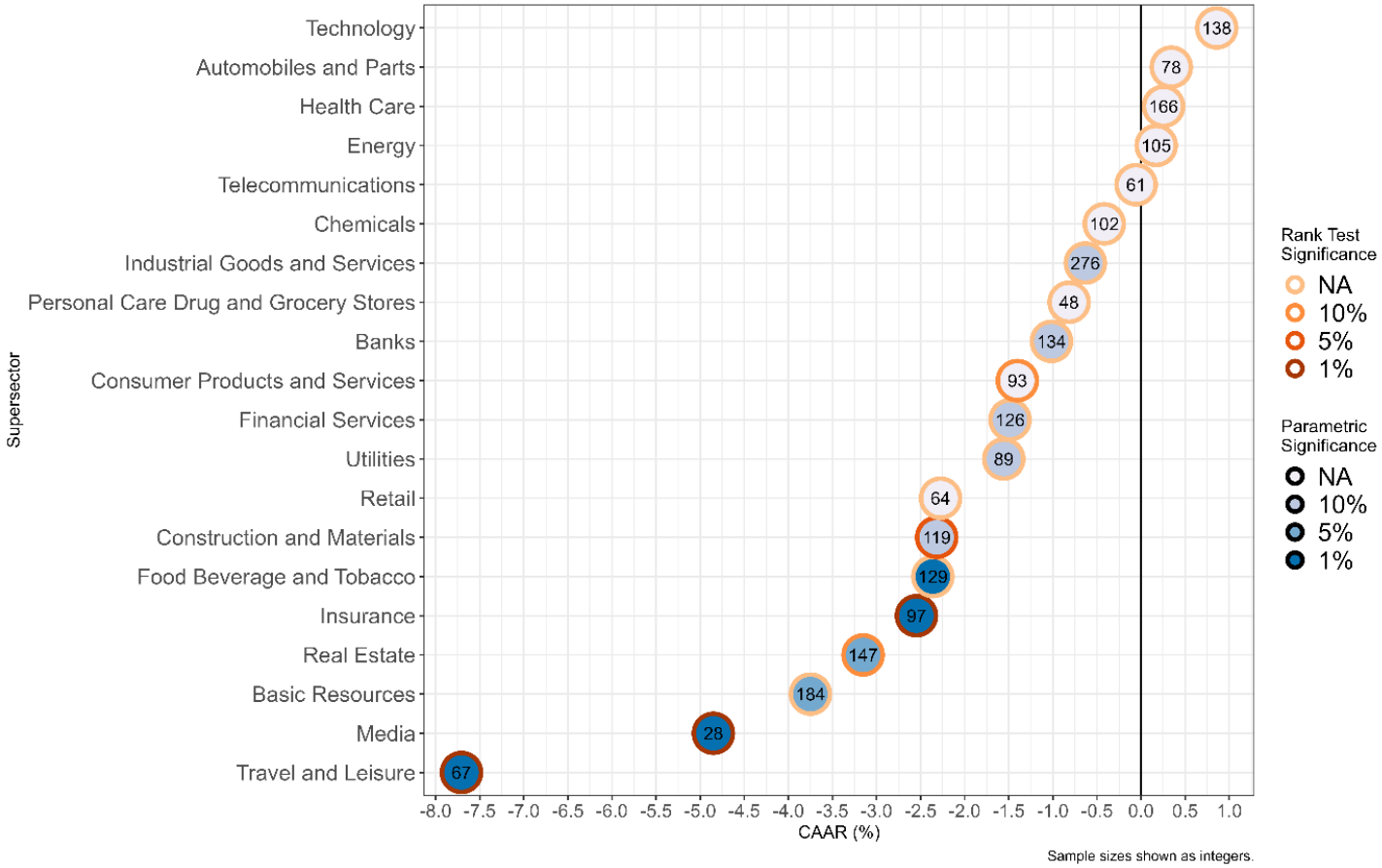
Figure 12 records an accelerated negative trend for the average, with a peak cumulation of the cross-sectional average AAR of approximately -1.7% on day 5. The narrative has now fully reversed with the day 0 proportion of negative returns recorded at approximately 81%. Furthermore, a larger proportion of significant AARs are present after day 0 than is prior, which are respectively 77% and 63%. It is evident that AARs succeeding day 0 are now larger and we are more confident of their statistical significance. Individually, Utilities responded negatively. Travel and Leisure is now negative for all significant AAR recorded and Real Estate is similarly dominated by negative AARs from day 0 onwards. In contrast, we record negative AARs for Retail as well as Construction and Materials prior to day 0, with the former featuring a conspicuous absence of statistically significant AARs thereafter. Construction and Materials features a similarly weak response to the event.

Figure 12: Combined AAR plot of Event 3 (grouped by supersector)



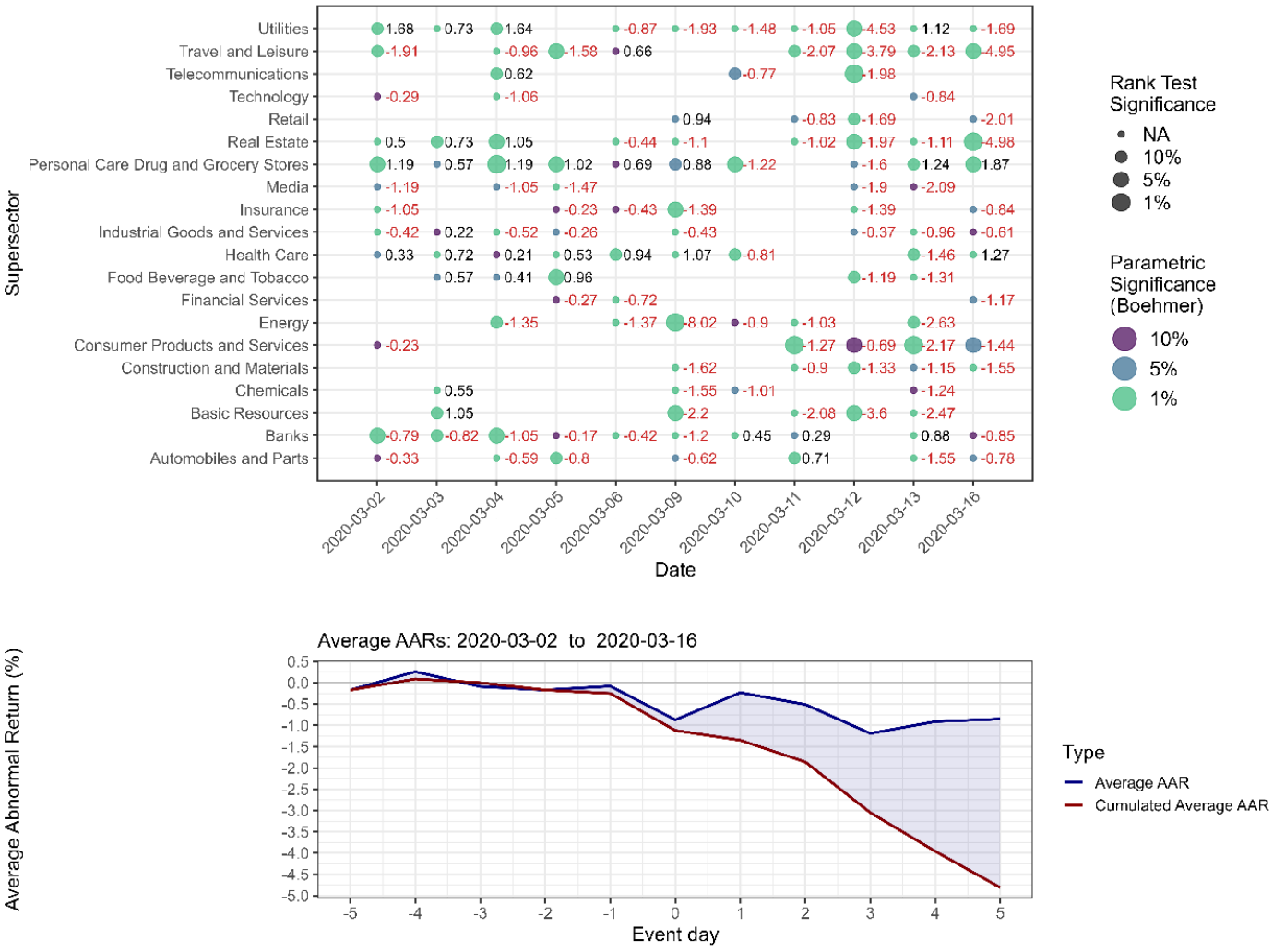
The negative effect of E_3 is apparent in Figure 13, with all 12 significant CAARs negative. Moreover, the cumulative negative effect upon Travel and Leisure substantially exceeds that upon other supersectors. Industry that is related to industrial processes and their supply chains are recorded with cumulatively negative CAARs, such as Industrials (Construction and Materials, Industrial Goods and Services) and the Basic Resources supersector of the Basic Materials industry. Alongside this, the market subsumed the event to exert negative consequences upon Financials (Banks, Insurance, Financial Services). The Food, Beverage and Tobacco supersector displayed a significant negative response to the event, as did the Consumer Products and Services, and Media supersectors.

Figure 13: CAAR plot of Event 3 (grouped by supersector)



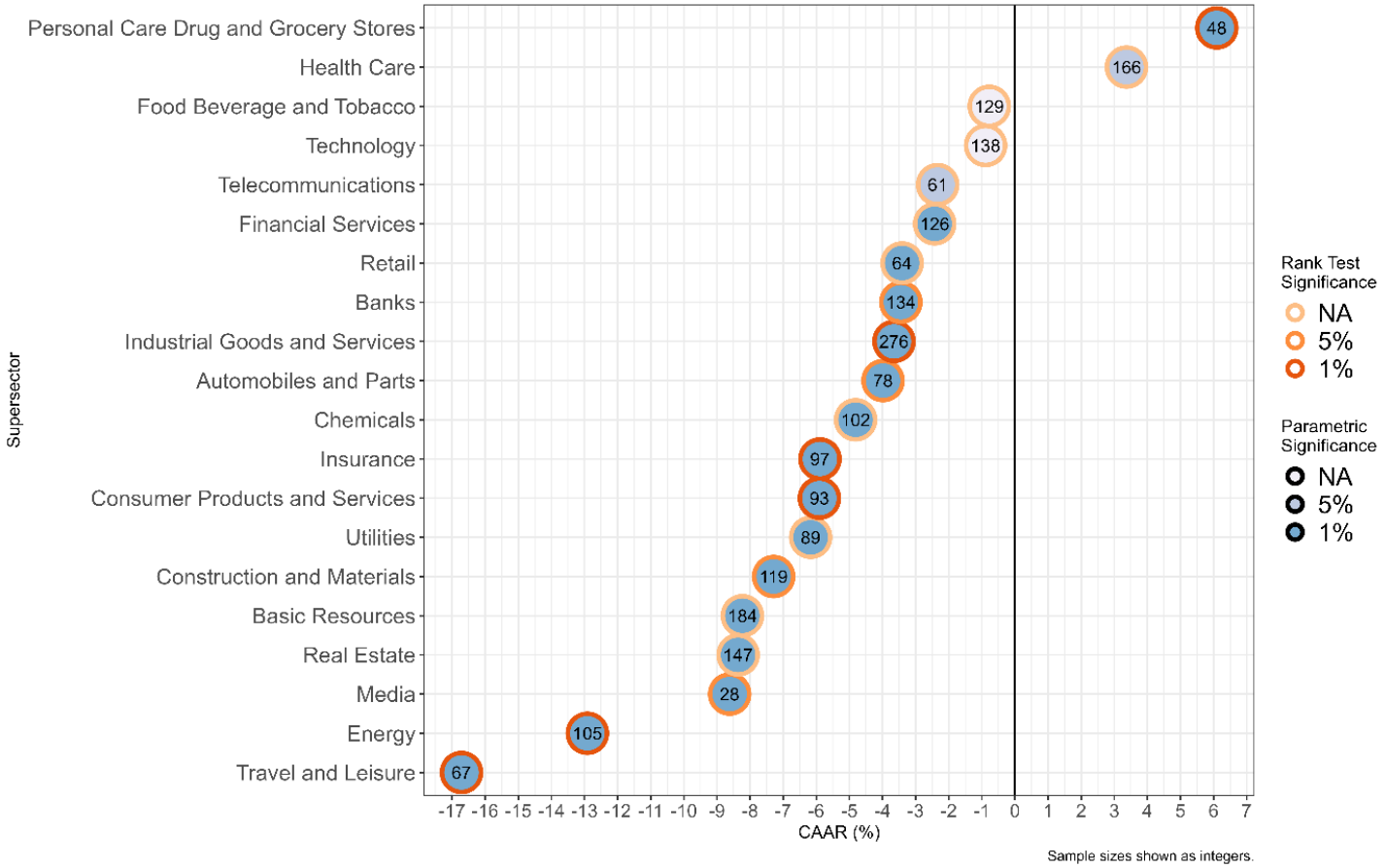
We observe an intensification of this negative trend in Figure 14. It is conspicuous in the average-plot that the effect of E4 is negative, with the cross-sectional averages closely tracking 0% prior to the event yet reaching an approximate -4.75% minimum on day 5. On event day 77% of AARs are recorded to be negative, with the magnitude of Energy's AAR at -8.02%. Furthermore, we record the proportion of negative AARs prior and succeeding day 0 as 55% and 86% respectively. Technology exhibits prominent absence of significant AARs. Similarly, Consumer Products, Construction and Materials, and Basic Resources participate in this behaviour prior to day 2, whereafter they all turn significantly negative. Real Estate experiences positive AARs approaching day 0 and then joins the negative consensus. In contrast, Personal Care, Drug and Grocery Stores as well as Health Care record large positive AARs prior to day 0, and only weaken on days -1 through 3 before returning to positive strength. It is evident investors preferred buying these supersectors as they sold their holdings in the others.

Figure 14: Combined AAR plot of Event 4 (grouped by supersector)



We see this preference further illustrated in Figure 15, with these two supersectors' salient positive response. Of the remaining supersectors, 80% respond negatively to the acceleration of COVID's spread, culminating in its pandemic classification.

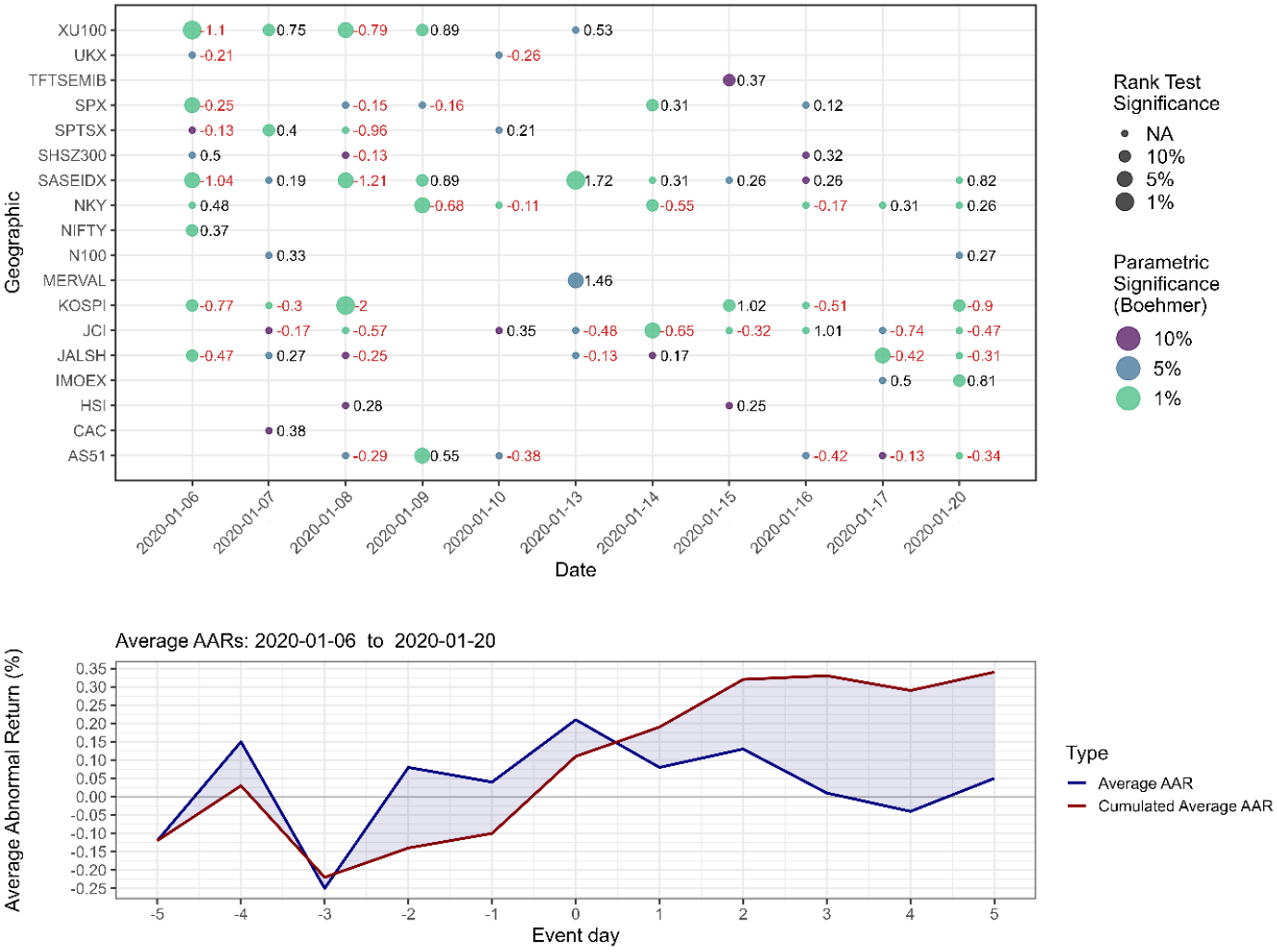
Figure 15: CAAR plot of Event 4 (grouped by supersector)



Geographic Regions

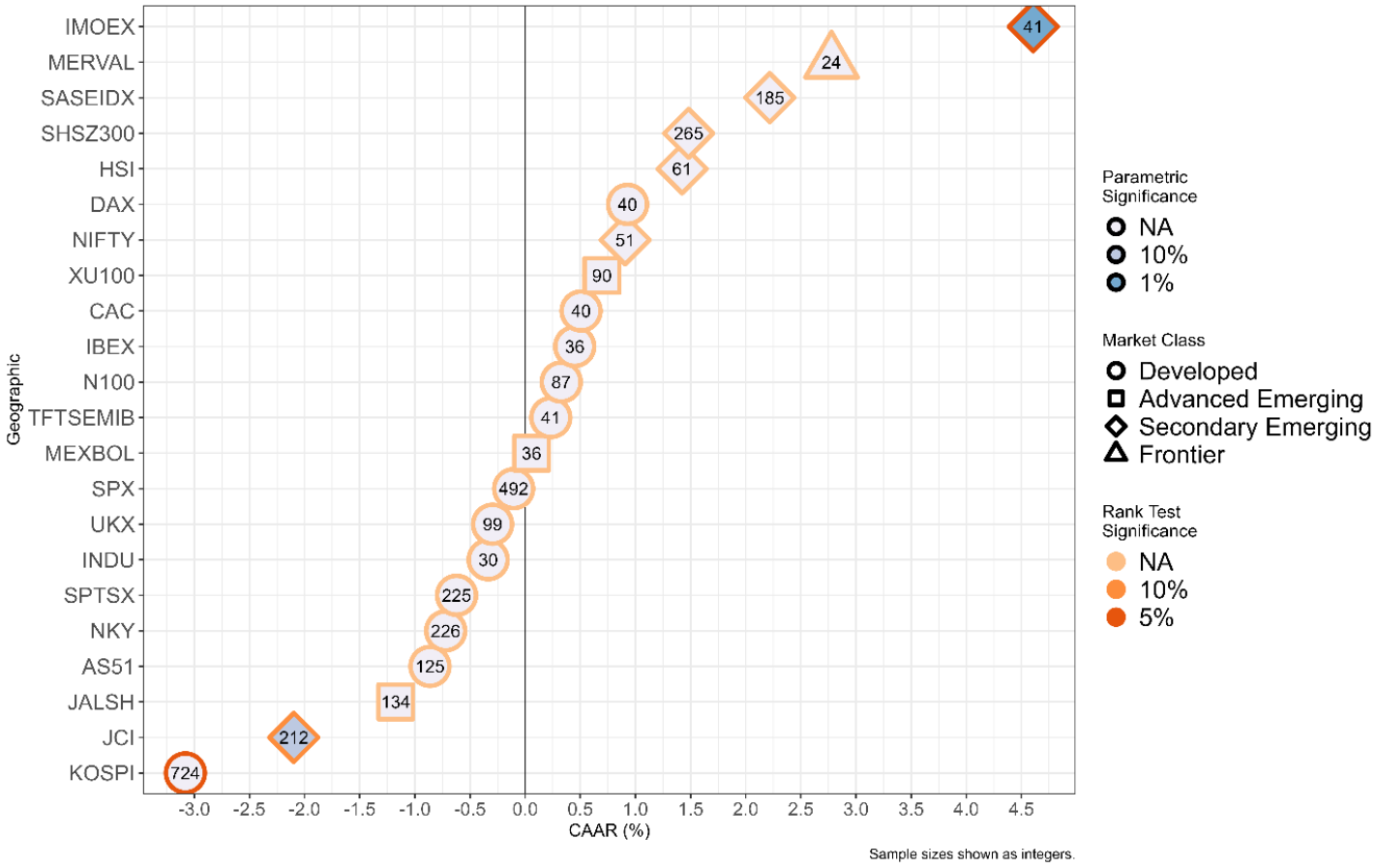
We now consider the event effects upon the geographic regions, referred to as 'markets'. Similar to the supersector groupings, the majority response of the markets to E_1 in Figure 16 is negative prior to day 0, whereafter many recorded positive or insignificant AARs. Proportionally, 61% of significant AARs prior to day 0 reduce to 43% and we have likewise lower confidence of their significance. We also record lower significance AARs after day 0. The low number of AARs upon event day illustrates a weak effect, yet this effect emulates the response of the supersectors through the significant AARs being positive in majority. Most responses appear mixed, however, and the JCI records a negative response. In contrast, the SASEIDX responds positively. On the balance, few markets record definitive responses to the event with the TFTSEMIB, SHSZ300, NIFTY, N100, MERVAL, IMOEX, HSI and the CAC recording few significant AARs. The IBEX, INDU, MEXBOL and the DAX failed to record significant AARs.

Figure 16: Combined AAR plot of Event 1 (grouped by geographic region)



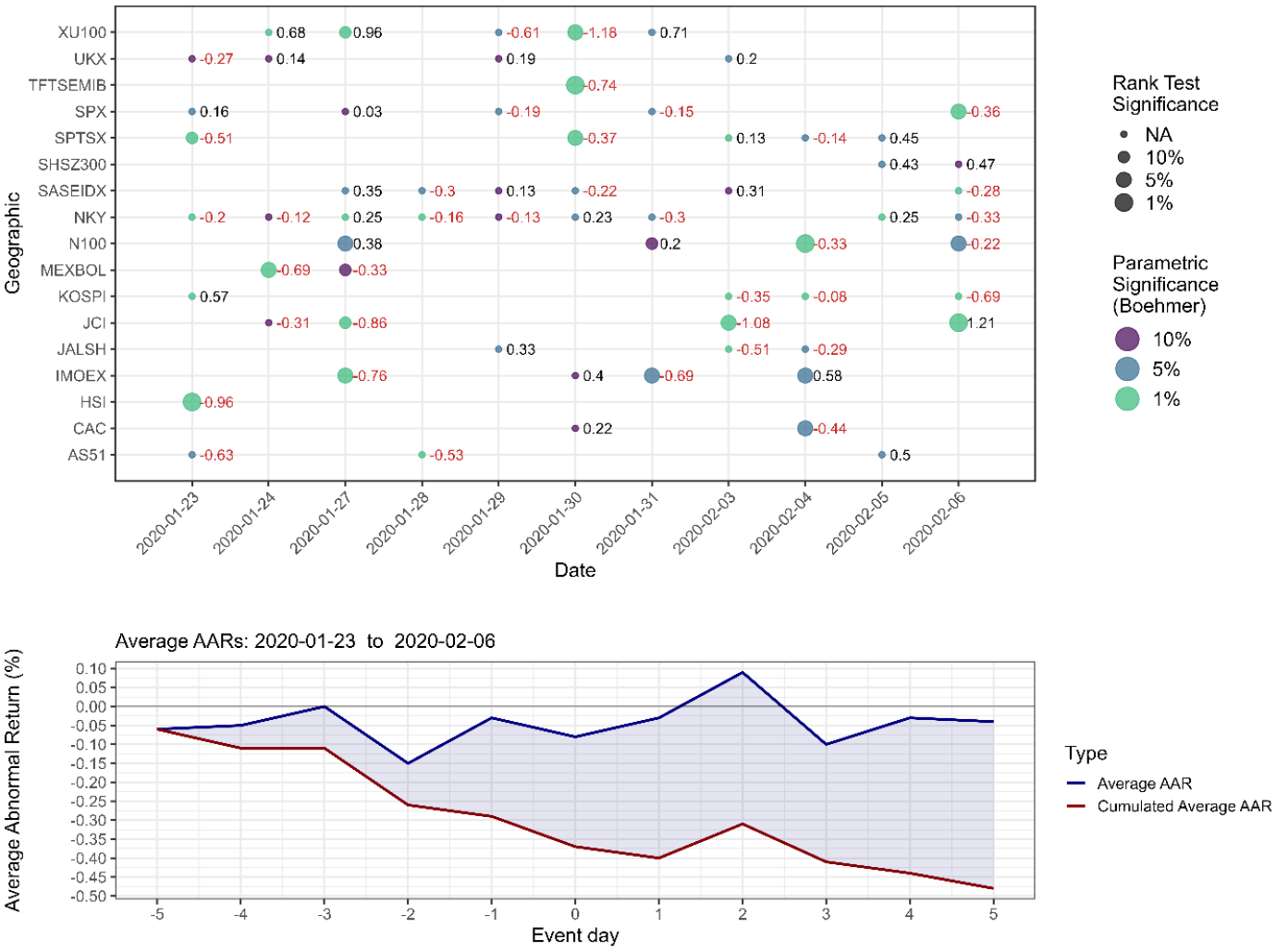
Over the period, it is evident in Figure 17 that the JCI and the KOSPI both recorded negative responses. Unexpectedly, the IMOEX records a large magnitude positive CAAR. We observe that all markets which responded are classified under 'Emerging'. On the balance, the cumulative responses are underwhelming.

Figure 17: CAAR plot of Event 1 (grouped by geographic region)



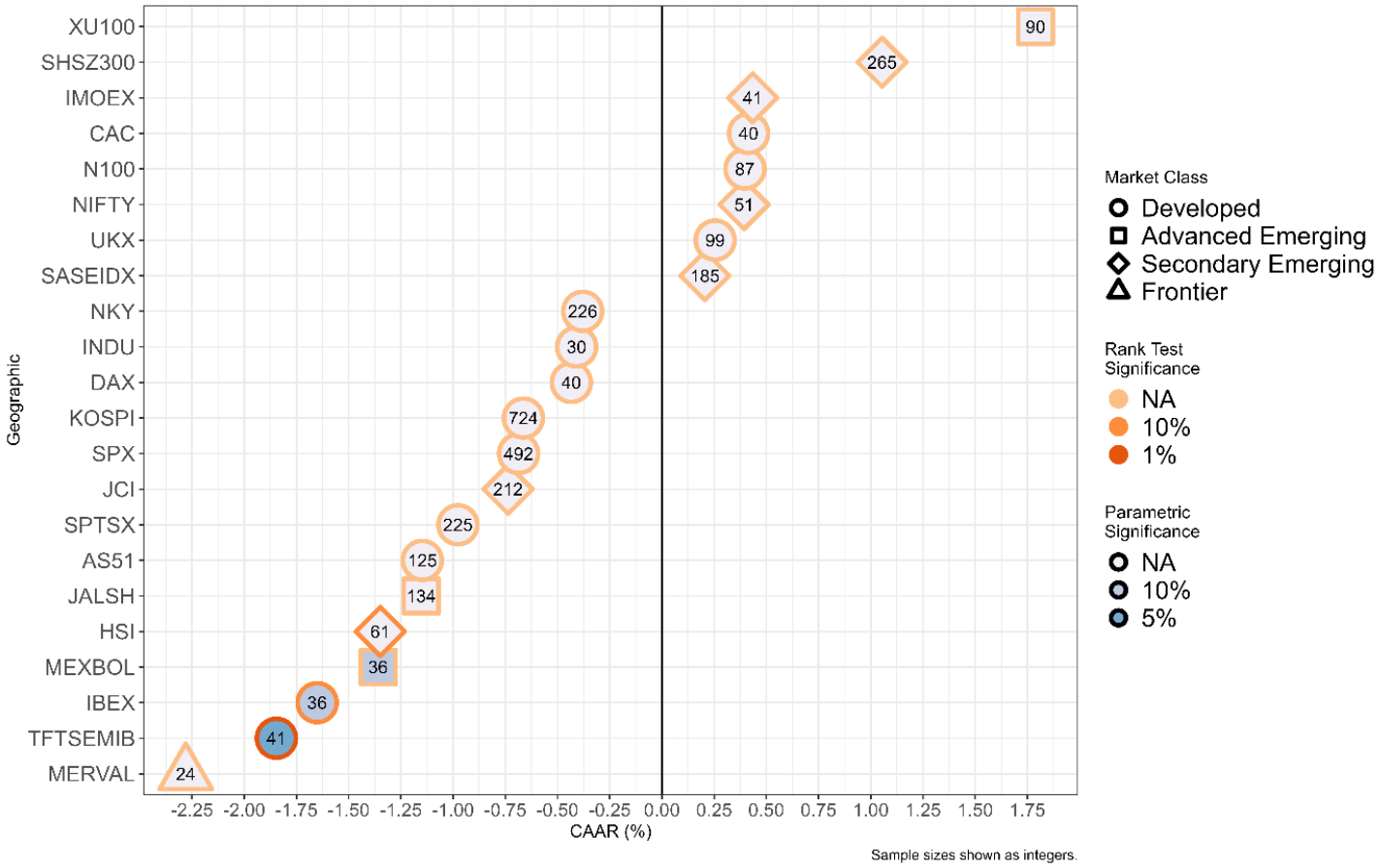
In Figure 18, the response of the markets to E_2 begins to countermand the aforementioned positive response. As can be observed in the average-plot, the average response is negative over the event period. Prior to event day, 59% of significant AARs are negative, 57% on day 0 and 57% thereafter. Conspicuous is the near absence of significant AARs from the PRC's representative markets, the HSI and the SHSZ300. Due to the PRC situation as the epicentre of the virus, this is unexpected. Contrastingly, the day 0 response of the XU100, the UKX and the TFTSEMIB is evident, which is the date COVID was declared a PHEIC. However, the record illustrates generally muted responses to the Wuhan lockdown on the 24th. The NIFTY and the MERVAL now join the IBEX, INDU, and the DAX in failing to record significant AARs.

Figure 18: Combined AAR plot of Event 2 (grouped by geographic region)



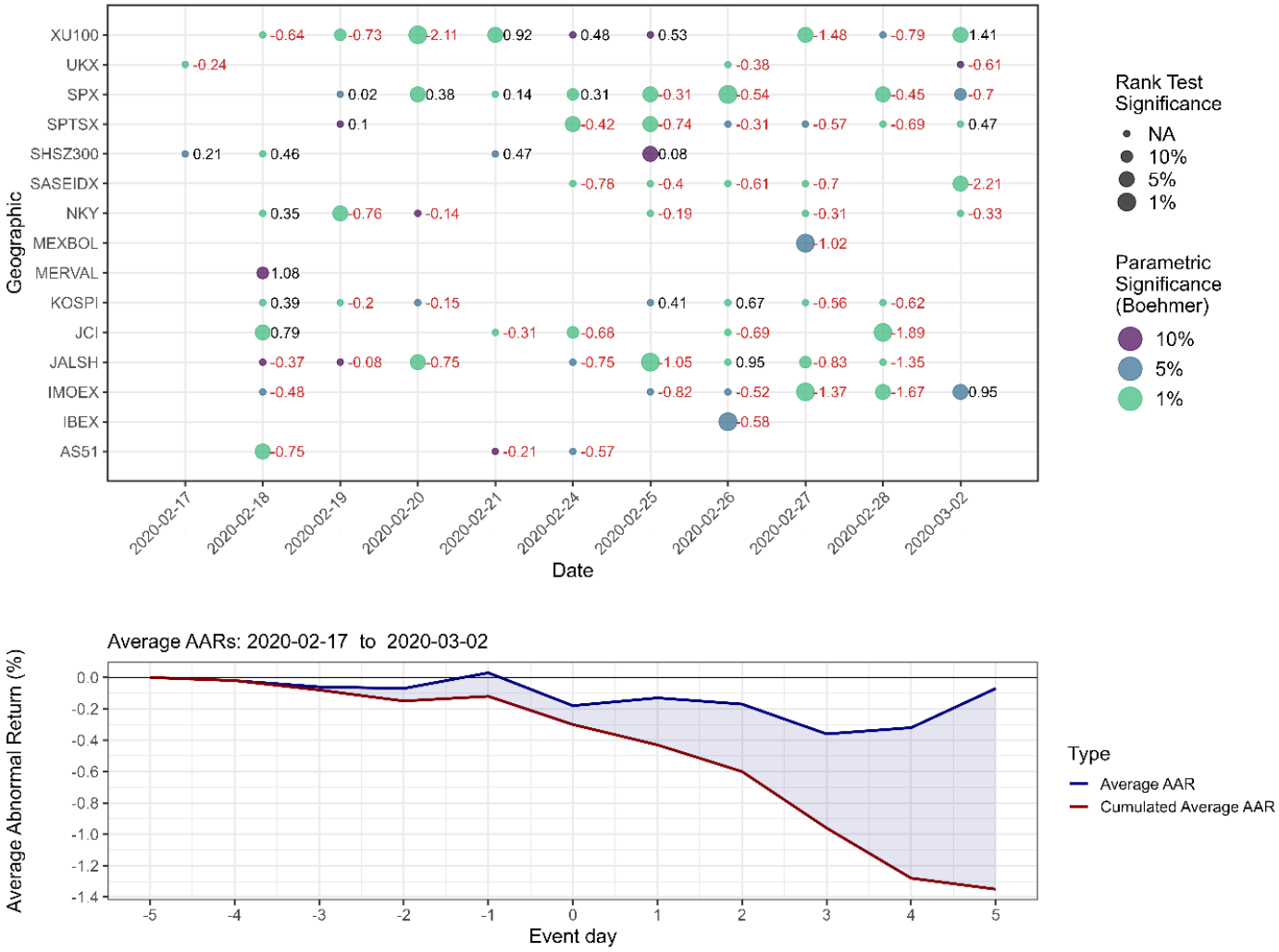
The conspicuous absence of significant AARs in PRC markets is alleviated by HSI's negative cumulative response in Figure 19. Likewise, the IBEX and the MEXBOL, whose AARs have hitherto been insignificant, record significant negative CAARs over E_2's window. On the balance, responses are again underwhelming when considered on a cumulative basis.

Figure 19: CAAR plot of Event 2 (grouped by geographic region)



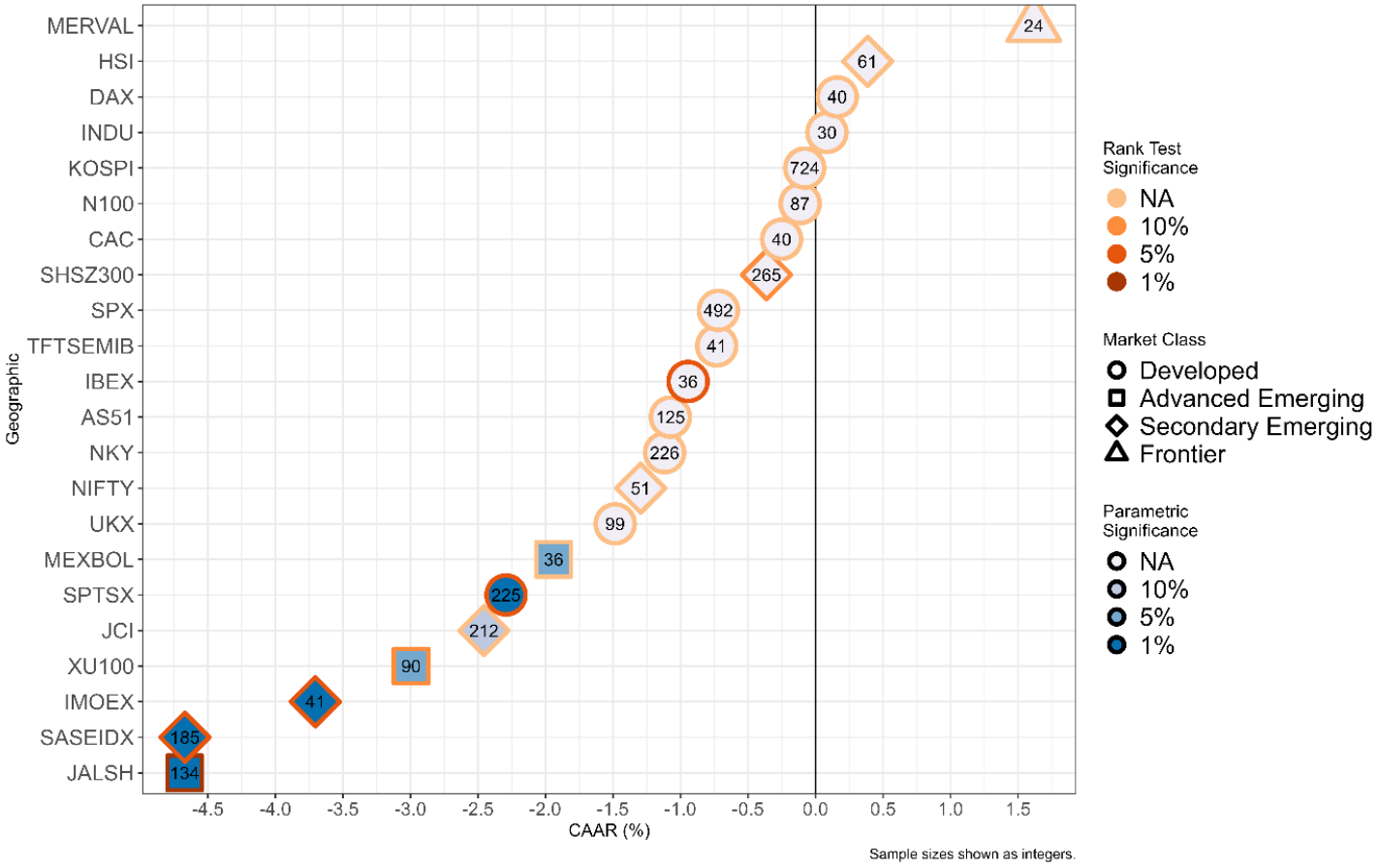
As with the supersectors, the response of markets to E_3 is similarly negative in Figure 20. On event day, 71% of significant AARs were negative, and the proportion of AARs changed from 55% prior to day 0 to 80% thereafter. We observe that the HSI, N100, CAC, and the TFTSEMIB no longer present significant AARs, whilst the DAX, and the INDU continue to be absent in this regard. In contrast, the IBEX and the MEXBOL now present a significant AAR each, respectively on days 2 and 3. The MERVAL and the UKX display a muted response. Nevertheless, it is evident in both the AAR plot and the average-plot that the response to the WHO's recommendation of stringent measures is negative. Furthermore, the Cumulated Average AAR now reaches a trough of approximately -1.3%, approaching 3 times the magnitude of E_2's response. The SPX, SPTSX and the SASEIDX demonstrate this plainly, with positive or no significant AARs prior to day 0 whereafter they all turn negative and with higher levels of confidence.

Figure 20: Combined AAR plot of Event 3 (grouped by geographic region)



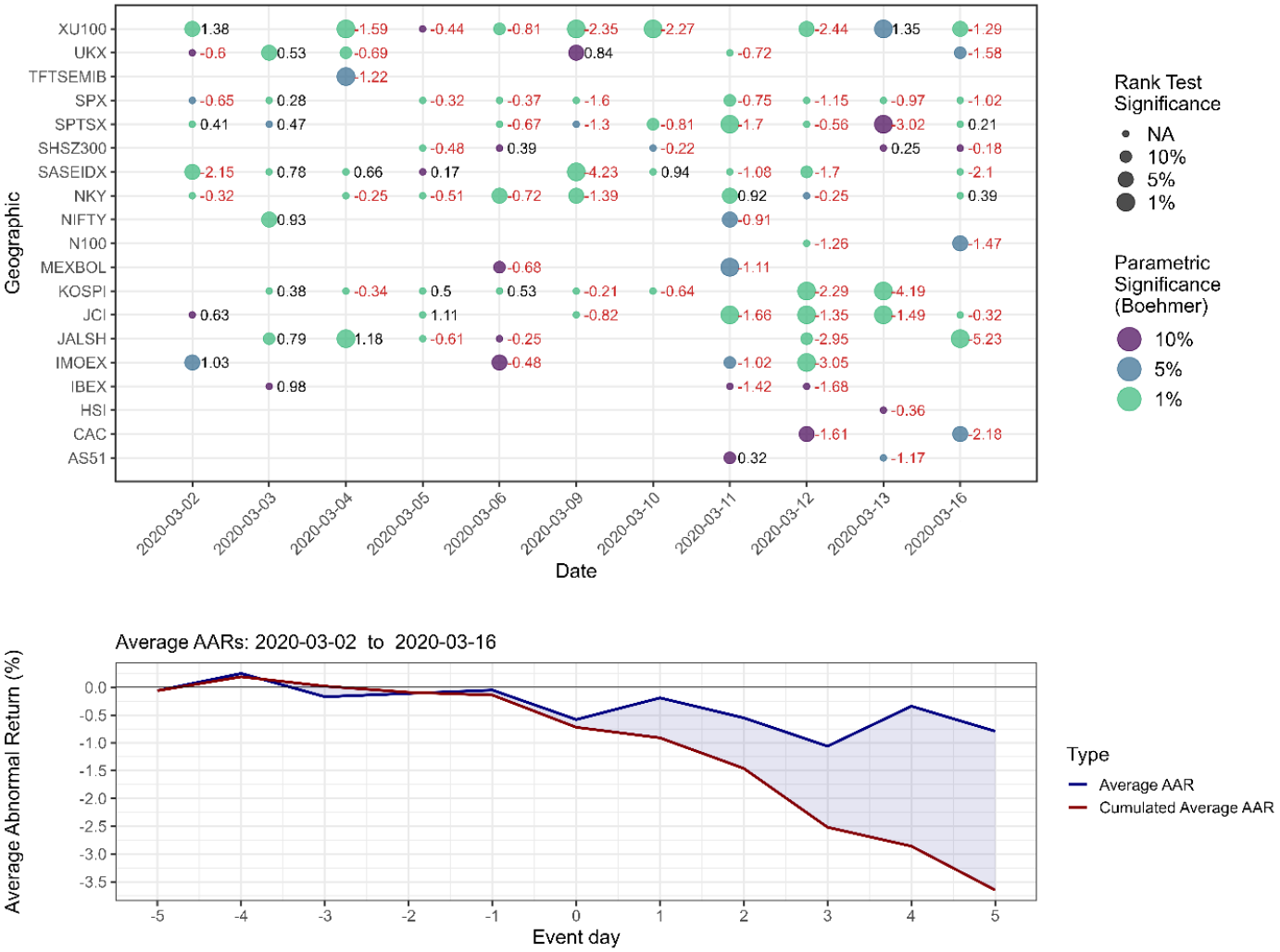
In Figure 21, the negative slant of the sample is evident. Furthermore, the magnitude is greater in Emerging Markets (EMs) than it is Developed, with the number of EMs presenting with significant CAARs exceeding that of Developed.

Figure 21: CAAR plot of Event 3 (grouped by geographic region)



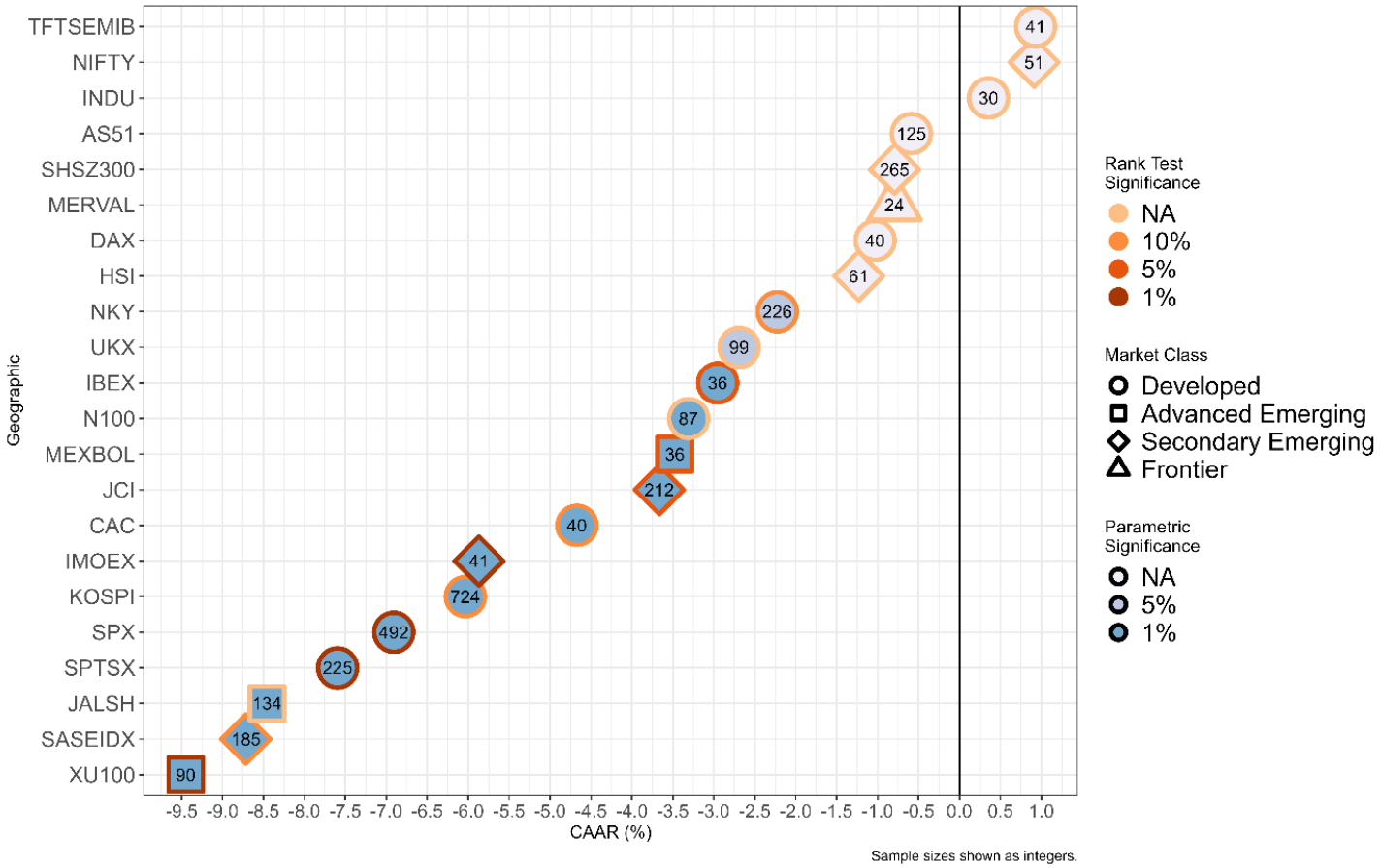
The negative effect of E_4 is most pronounced in Figure 22's average-plot, reaching a cumulated average AAR of approximately -3.5% in magnitude. The event's negative effect is further evidenced by the 52.5% proportion of significant AARs prior to day 0 reaching 87.5% on day 0, and 85% thereafter. Furthermore, the response to the WHO declaration upon the 12th is evident. There is, however, a conspicuous absence of response in the MERVAL, INDU, and the DAX. The INDU and the DAX record neither significant AARs nor CAARs for each of the events studied.

Figure 22: Combined AAR plot of Event 4 (grouped by geographic region)



Apparent in Figure 23 is the increased number of Developed Markets recording significant CAARs, whilst Emerging Markets respond with greater negative magnitudes. This is evidenced by contrasting the approximately -4.6% average CAAR amongst Developed Markets with the approximately -6.7% of the Emerging Markets. The negative slant of the event is now fully developed, however the cumulative response of 8 markets is still insignificant. We contrast this with the supersector groupings of which only 2 recorded insignificant CAARs.

Figure 23: CAAR plot of Event 4 (grouped by geographic region)

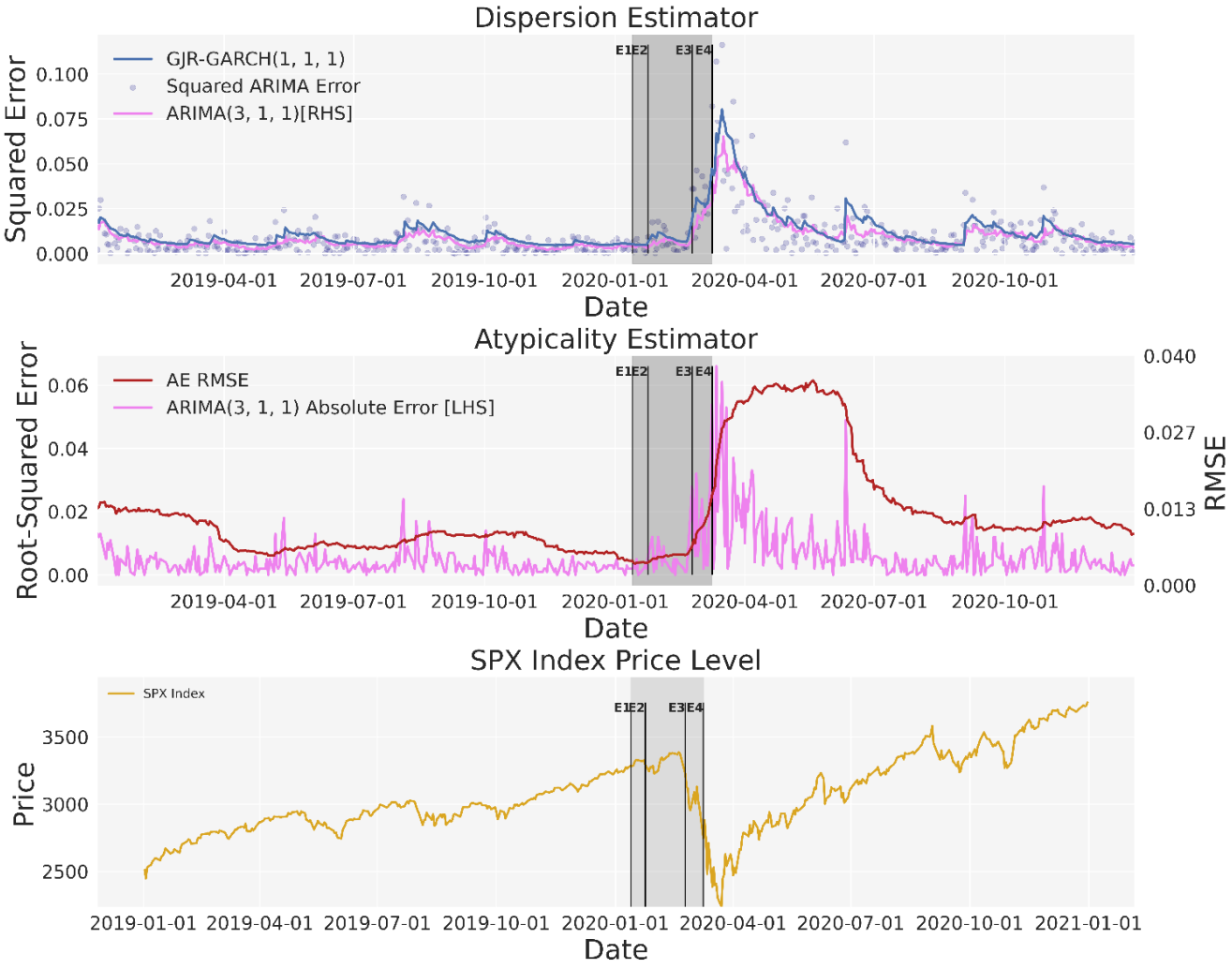


4.2 Anomaly Characterisation Process

4.2.1 Identification of Weak Signals

Subsequent to application of the ACP upon the full samples, the ensuing graphical analysis facilitated identification of weak signals preceding the COVID event. The process's efficacy of characterising anomalies became immediately apparent, resulting in clear distinction between typical and atypical price movements. The event is assumed to begin on E_4's day 0, as described in Section 3.2.1.4, which we denote as "E_4". Greyed-out periods in ACP-related figures represent the period between E_1's and E_4's event day 0. Finally, the Dispersion Estimator and Atypicality Estimator's estimates plotted as line-plots are described as the 'dispersion line' and 'atypicality line' respectively.

Figure 24: Close-Up of COVID Period - SPX Index



When examining plots of the atypicality estimator longitudinally and cross-sectionally for all securities in the sample, commonalities emerge. Anomalies present with steep inclines (stepwise or otherwise) on the atypicality line (red), as perceived in Figure 24. Similarly, the dispersion estimator presents a regular, minor upwards movement followed directly by rapid incline of the dispersion line (blue)[31]. Subsequently we suggest the following ACP rule: that an anomaly is unfolding when

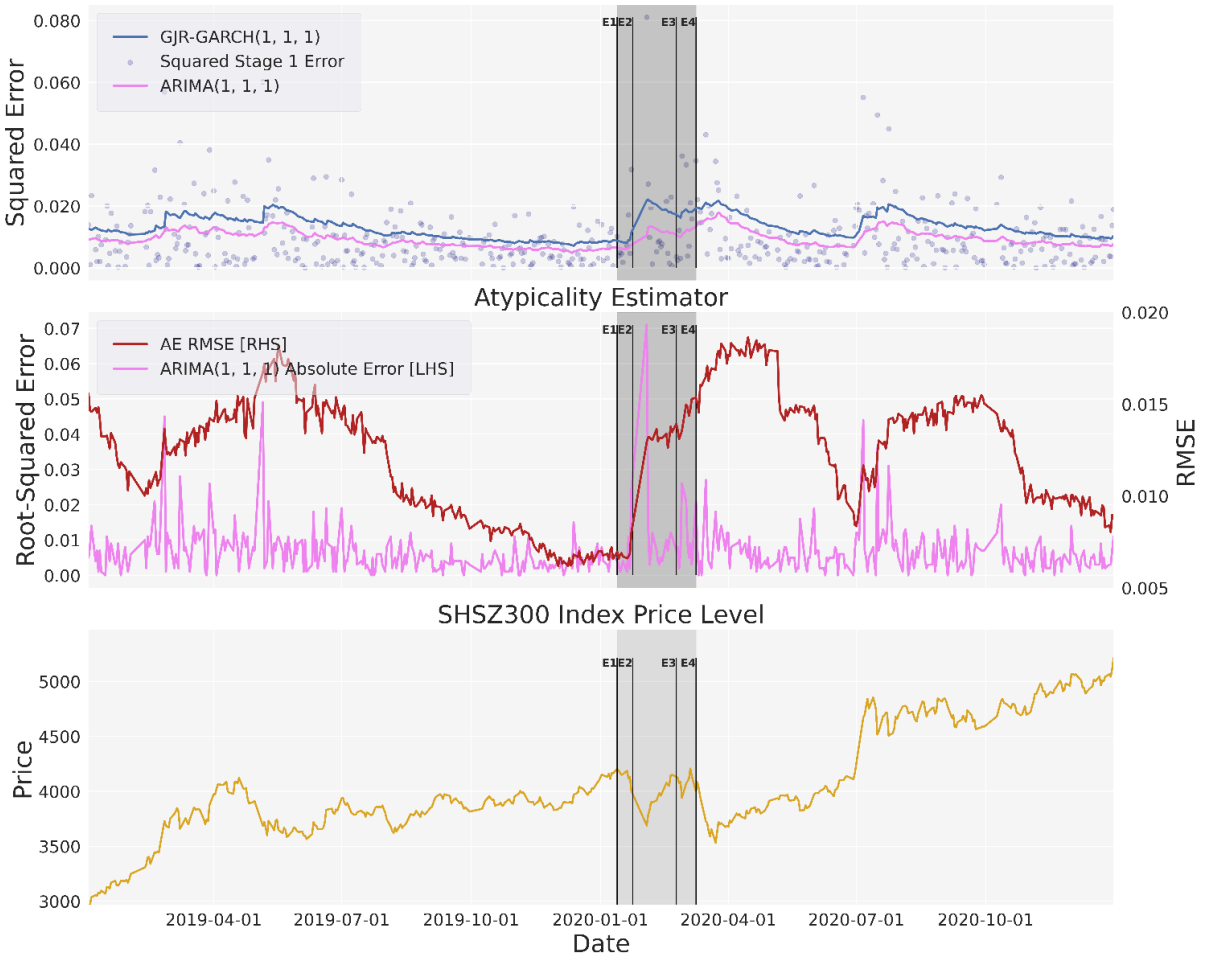
[31] Figures that place the dispersion and atypicality lines on the same x-axis are available in Appendix 4, wherein the co-movement of these lines, and the differences in their rapidity and divergences in trend can be observed.

the atypicality line increases at a gradient greater than recent history and simultaneous upwards jumps leading to new medium-term highs are observed by the dispersion line.

Consequently, and as discernible in Figure 24 that this behaviour began to unfold prior to E_4's event day, we suggest that this prototypical version of the ACP has detected weak signals in our ex-post analysis. We detect that an unfolding anomaly per the atypicality and dispersion lines is concomitant with a divergence between the most recent atypicality/dispersion line observation and its prior trend, a possible point of further study.

In the same way, we graphically analysed all markets contained within the sample. Those most similar to the SPX Index or exhibiting a hybridisation of the SPX and the HSI Indices' characteristics, present this same weak signal following E_3, and between E_3 and E_4, as described above. No such signal presented in markets exhibiting characteristics similar to the HSI Index itself. Consequently, we define three categories: (1) the SPX, (2) the SPX-HSI Hybrid (HYBRID), and (3) the HSI[32]. The ACP's results are discussed as pertaining to these three categories and not per market. The SHSZ300 could not be classified according to these and is thus categorised as being unique and as can be seen in in Figure 25, we discover a significant anomalous signal in the SHSZ300's data preceding all other markets. This market did not react as expected during the event study analysis; failing to react in a statistically significant way for most observations. We suggest this demonstrates the proficiency of the ACP.

Figure 25: Close-Up of COVID Period - SHSZ300 Index



[32] A list characterising each market as one of the categories or a hybrid thereof, can be found in Appendix 5.

In each of the figures provided, we find the models chosen for the ACP superior in characterising the data's typicality. These models indicate perspicuously the typicality of data whilst providing context for this characterisation. Disparately, the auto-ARIMA's estimation of ε_t (comparable to the dispersion line) and the consequent model error (analogous to the AE's RMSE) provide deficient characterisations of the data's typicality. The proficiency of the ACP at characterising equity-price data according to its atypicality is evident, and this can be used to discover point or contextual/conditional anomalies. Furthermore, cross-sectional juxtaposition of the atypicality measures provides the potential observation of collective or group anomalies.

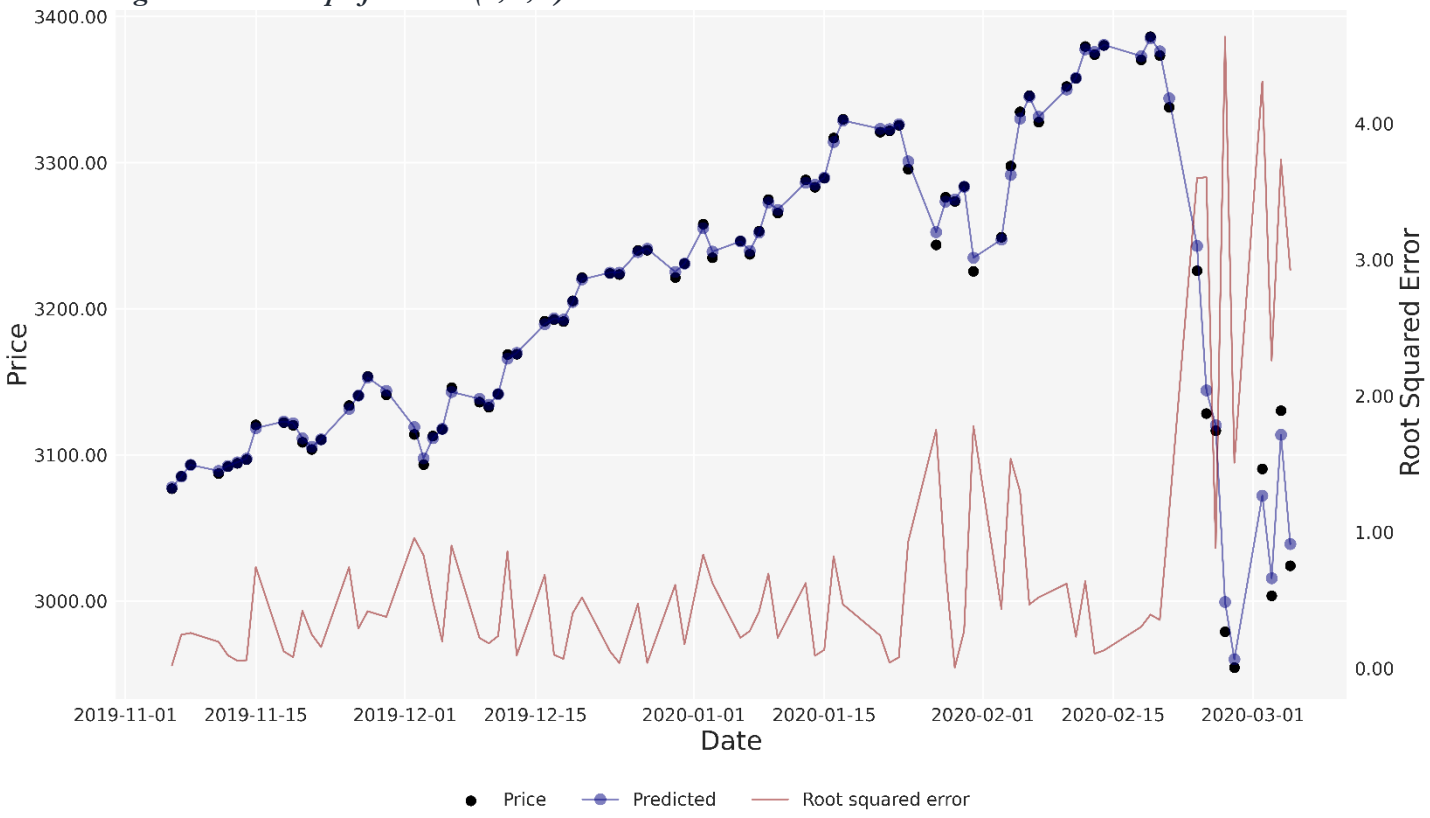
We now analyse the three categories over their entire samples, and demonstrate graphically the purpose of Stage 1 of the ACP.

4.2.2 Characterisation of Markets

Stage 1: Graphical Demonstration

In Figure 26 we illustrate the objective of the Stage 1 $\mathrm{ARIMA}(1,1,1)$ transform. The model is able to fit the underlying structure contained within the SPX Index data resulting in the residual information being useful in characterising typicality[33]. We draw attention to the close proximity of the model's fit, resulting in a low root-squared error (the square root of the squared residual error). However, this closeness diverges during the 2020/03 period, where the model's prediction $\hat{Y}_t$ (represented by 'Predicted') diverges to a greater magnitude than is typical from the actual closing price Y_t (represented by 'Price') – with an increase observed in the residual. We recollect that the $\mathrm{ARIMA}(1,1,1)$ prediction $\hat{Y}_t$ is forecasted in terms of the estimated one-day log return $\hat{y}_t$ according to a function of the prior value and the prior moving average plus some constant. Thereby, the error $Y_t - \hat{Y}_t$ extracts the information useful in characterising the typicality of these datapoints in relation to their prior values, permitting the consequential identification and description of abnormalities.

Figure 26: Close-Up of ARIMA(1, 1, 1) Model's Fit on SPX Index



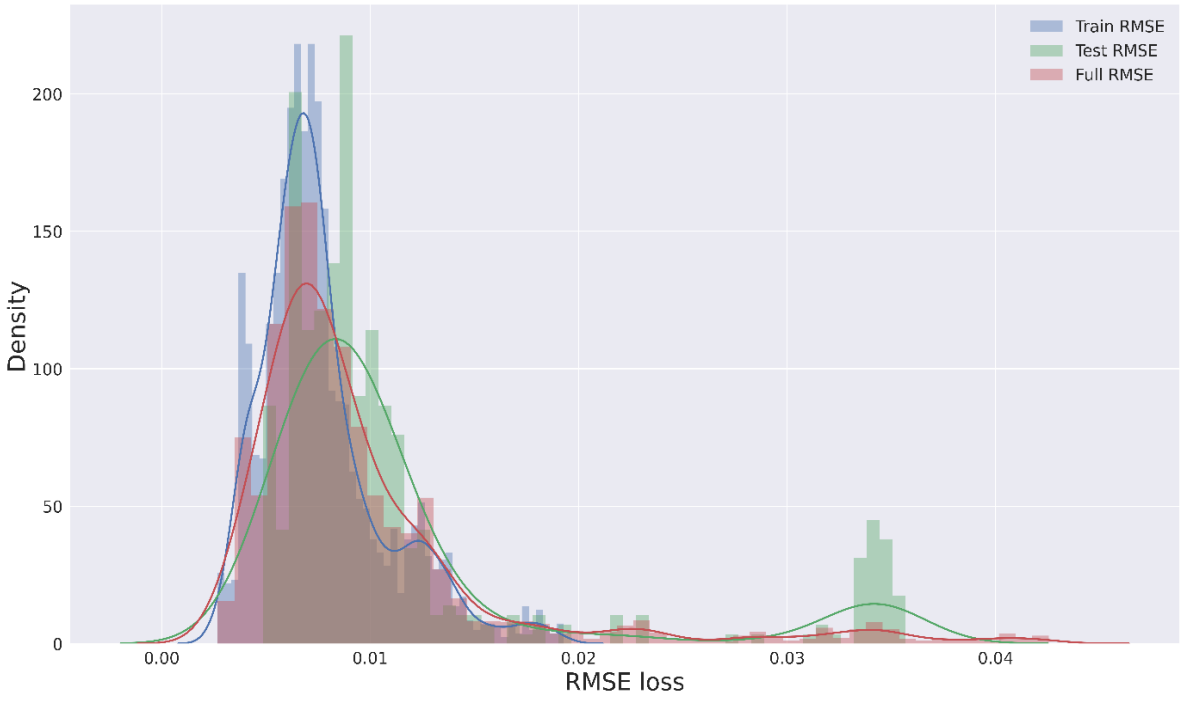
[33] The results of stage one can be found in our GitHub repo https://github.com/keegangclarke/acp.git at the following directory: acp/ACP Results/stage1/covid_close_up_arima_fit

Stage 2: Anomaly Characterisation and Market Categories

The dispersion and atypicality estimators can be used to arrive at similar conclusions through their distinct methods. The dispersion estimator displays greater reactivity and is consequently noisier in its characterisation enabling a perspicuous analysis of a single day's typicality. The atypicality estimator displays lower reactivity, lowering the noise in its typicality estimate. Nevertheless, it remains as responsive as the dispersion estimator to larger anomalous movements. Consequently, it estimates atypicality with greater clarity by considering the previous 64-days in its assessment of a single day's atypicality. Using this combination, we are able to characterise anomalies in the data whilst simultaneously characterising the behaviour of the market to which the data belongs. Moreover, the atypicality estimator's RMSE's distribution assists in the characterisation of a market's price movements. This is as the different shapes help illustrate the market's tendency and characteristics.

Further analysis incorporating characteristic[34] description upon the three dominant categories and the unique market ensues, including the SHSZ300 and the MERVAL. The former is largely dissimilar to the three categories records and the latter records an unusual event. For the density plots of the atypicality lines, the red, blue, and green distributions are the full sample, training sample and the testing sample respectively. Furthermore, this section pertains to ACP results upon the concerned equity securities over the full period of their samples and is not limited to the period around the COVID event.

Figure 27: SPX Density Plot of AE's RMSE (SPX Category)

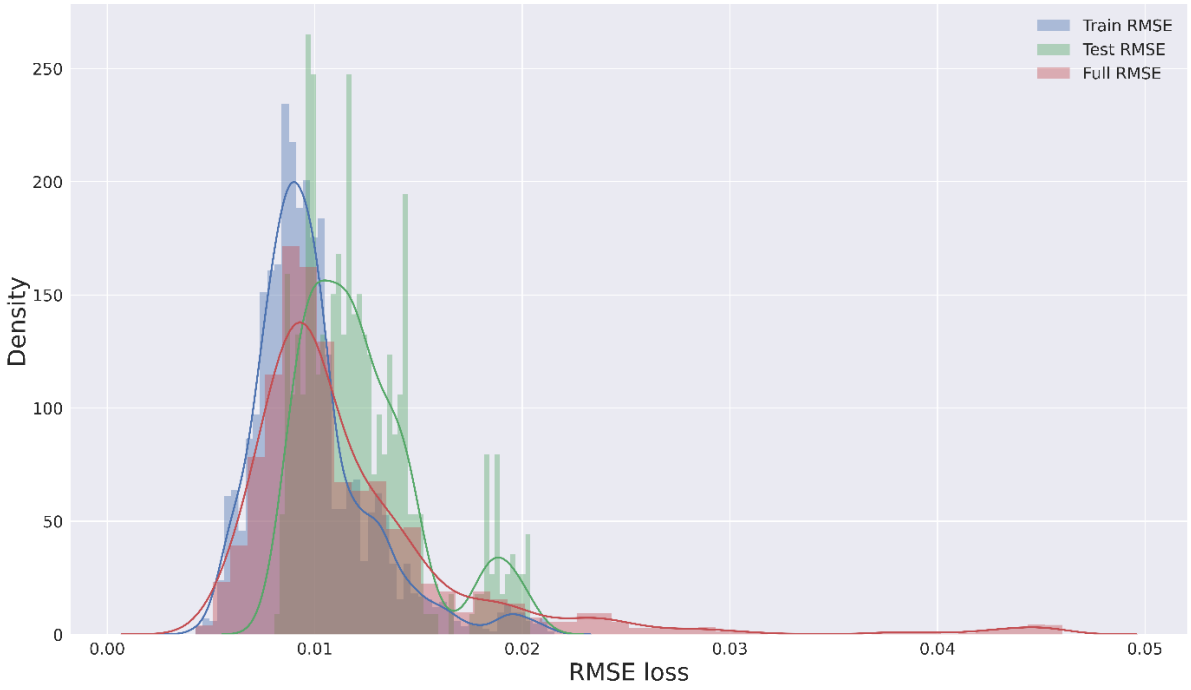


Pictured in Figure 27 is the density distribution of the SPX Category's atypicality line (the AE's RMSE), which appears to be similarly shaped to a generalized extreme value distribution with secondary mode apparent in the Testing RMSE' tail. The SPX Category presents with large and sudden responses to major global events as measured in both the dispersion and atypicality lines.

[34] Figures of ACP results for all equity securities can be found in our GitHub https://github.com/keegangclarke/acp.git in the following directory: acp/ACP Results/stage2/

In Figure 32 this is perspicuous during the GFC and COVID periods, recording its greatest atypicality for both events. Similar magnitudes of atypicality are prominent for these two periods. Examination of the dispersion line discloses a volatile clustering signal yielding a rough dispersion line throughout the sample, including periods of market calm. In terms of typicality, the category displays sensitivity towards price movements divergent from the underlying trend. In other words, changes from the status quo are most often anomalous to some degree. Further, the atypicality line presents with large and sudden peaks, interspersed with periods of relative calm wherein the atypicality line is fairly smooth. Finally, observation of the price plots elucidates an apparent exponential trend.

Figure 28: HSI Density Plot of AE's RMSE (HSI Category)



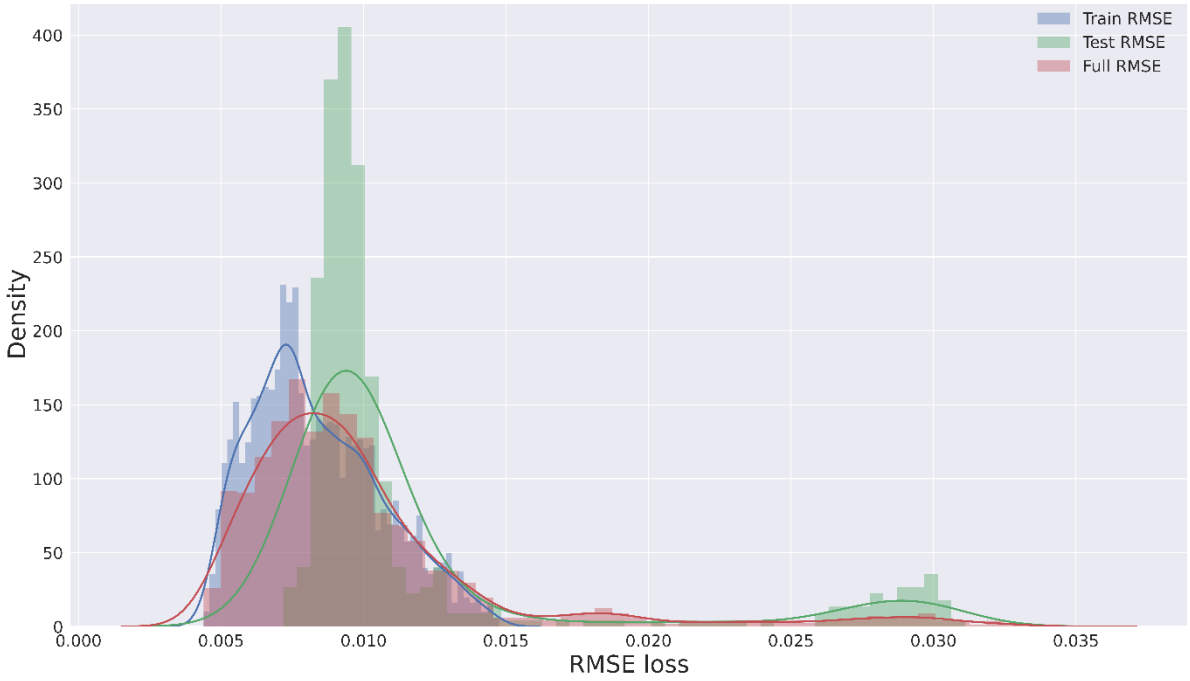
In contrast to the SPX, the HSI Category presents with muted responses to global events according to both the dispersion and atypicality estimators. This is most easily observed during the GFC and COVID periods. This category's typicality density plot is pictured in Figure 28 and is likewise similarly shaped to a generalized extreme value distribution[35], however, it has a longer and thinner right-side tail as well as greater kurtosis. Furthermore, a secondary mode is apparent in the Testing RMSE', however not in the tail as with the SPX Category. Although the HSI Category still records a large magnitude of atypicality during the GFC in Figure 33, it records much lower atypicality during the COVID period (as measured by both estimators). The HSI Category displays a smoother dispersion line, only rising vertically to highly anomalous movements. Additionally, this category does not present with the same clustering behaviour in its dispersion line. Observing the atypicality line, the gradient of peaks is on average lower than the SPX Category. The dispersion line however returned to a typical baseline after the COVID period, whereas the SPX continued to have experienced elevated atypicality of the clustering kind. Finally, the underlying prices follow an apparent mean-reversion trend in their line-plots.

For completeness, an example of the SPX-HSI Hybrid Category (HYBRID) is given in the JALSH Index. Figure 29 illustrates the density plot of the autoencoder's RMSE. This distribution is flatter than the previous two categories and consequently resembles a Gumbel in shape more than GEV.

[35] The well-known distributions are used to *describe* the RMSE distributions. We do not claim the RMSE distributions are following these distributions, the determination of which requires further tests outside of the scope of the study.

Similar to the SPX category, the HYBRID also presents with a secondary mode in the Test RMSE's tail and the tail is not as long as the HSI's. The HYBRID follows the SPX in recording comparable magnitudes of the atypicality and dispersion measures with respect to the GFC and the COVID periods, as evidenced in Figure 34. However, the dispersion line is smoother than the SPX's whilst remaining rougher than the HSI's. In contrast, the HYBRID records a rougher atypicality line than both the SPX and the HSI Categories.

Figure 29: JALSH Density Plot of AE's RMSE (HSI-SPX Hybrid Category)



In all three categories we find that the atypicality measure's peaks are frequently paired with a primary peak as a result of the initial shock, followed by an 'after-shock'. Figure 30 is an example extracted from Figures 32 and 33. We suggest that a major atypical decline in price is then often followed by a strong rebound of which the increase is similarly atypical.

Figure 30: HSI [LHS] and SPX [RHS] Atypicality Line during GFC

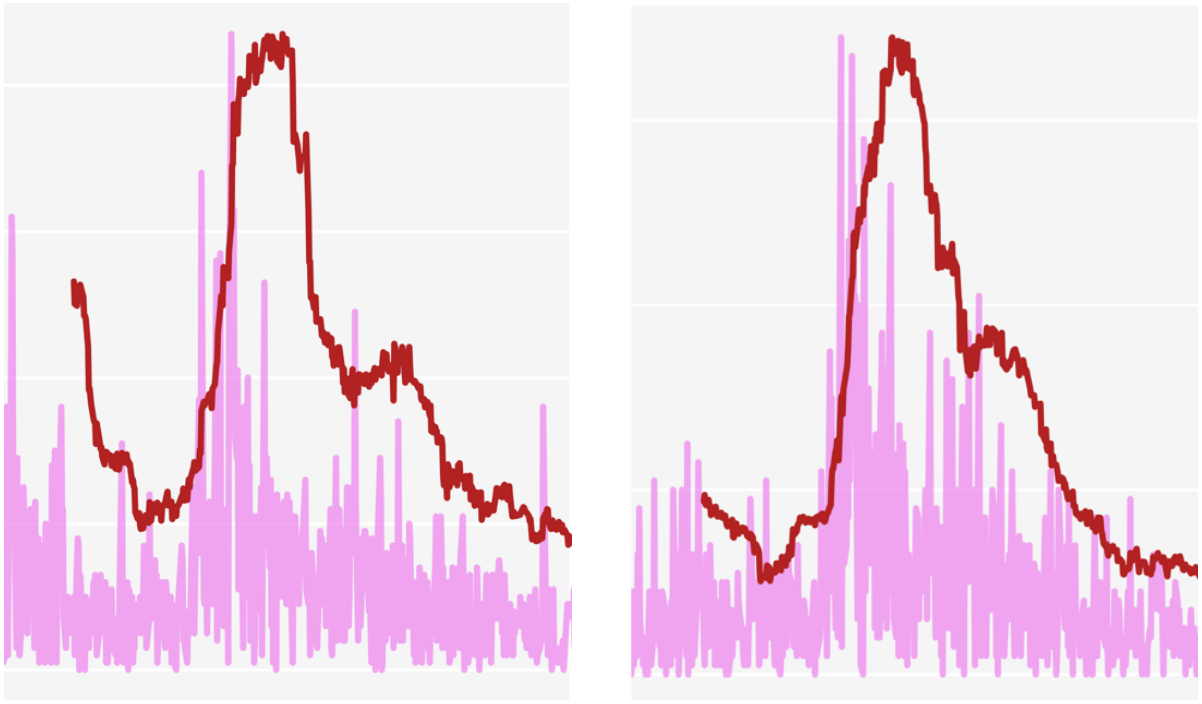
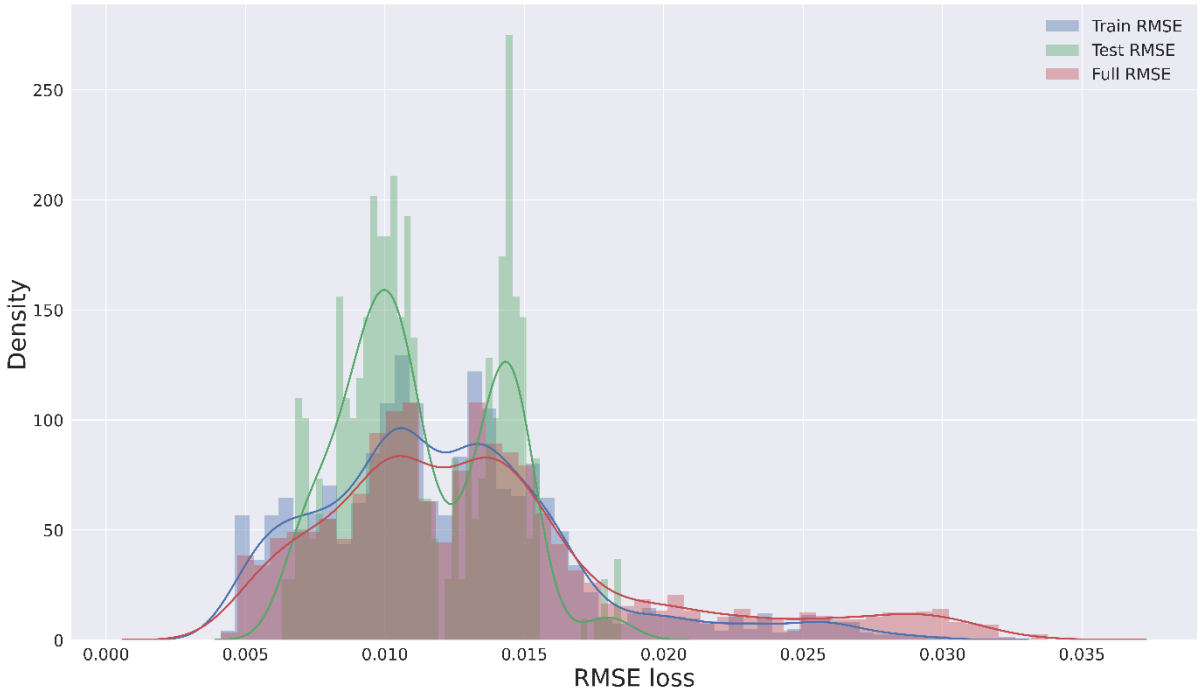


Figure 31: SHSZ300 Density Plot of AE's RMSE (Unique)



The ACP records the SHSZ300's typicality as dissimilar to the three identified categories. Whereas the categories have single distribution RMSE density plots, the SHSZ300's RMSE density plot, pictured in Figure 31, appears to be multi-distributional. Furthermore, the dispersion line, illustrated in Figure 35, displays a roughness similar to the HYBRID category withal the maximum peaks are considerably lower. The dispersion line of both the GFC and the COVID periods record lower atypicality than observed in other markets. For all other markets studied, the GFC period shows a large magnitude of atypicality. Additionally, the COVID period recorded a similar magnitude of atypicality for the SPX Category.

SHSZ300's dispersion estimator displays a unique relationship with the comparative auto-ARIMA model. There is greater divergence between the inferior model's estimate of ε_t and the GJR-GARCH's estimate, where the latter is consistently above the former. In the other markets examined, a large divergence between the two estimates generally occurs when the dispersion estimator characterises a datapoint as highly atypical. This suggests that at a basic level the price movements of this market are unusual. Corroborating this is the atypicality line's profile. For the other markets examined, peaks are often spaced by calm periods. The SHSZ300 stands in contrast, as peaks are more often spaced by troughs prior to reaching a new peak. Consequently, it seems that this market is highly tumultuous.

We now examine the atypical event discovered in the MERVAL in brief. Represented in Figure 36, this equity index experienced a one-day decline of approximately 37% on the 2019/08/12 (according to own calculations). This decline occurred in response to unexpected and undesired election news (Ponczek, 2019). As designed, the dispersion (atypicality) estimator records this as an incredibly unusual one day reaction, approximately more than 2.5x (2x) the atypicality of either the GFC or COVID for this market. Further observation of the atypicality line identifies instantaneous verticality of the line, not seen elsewhere in any of the markets examined. Accordingly, there was no prior indication of the event in the atypicality line until its lapse. Therefore, as this "extreme impact" event was only predictable in hindsight we ascertain that it is accurate to classify it as a 'black swan'.

Figure 32: ACP fitted on SPX Index (SPX Category)

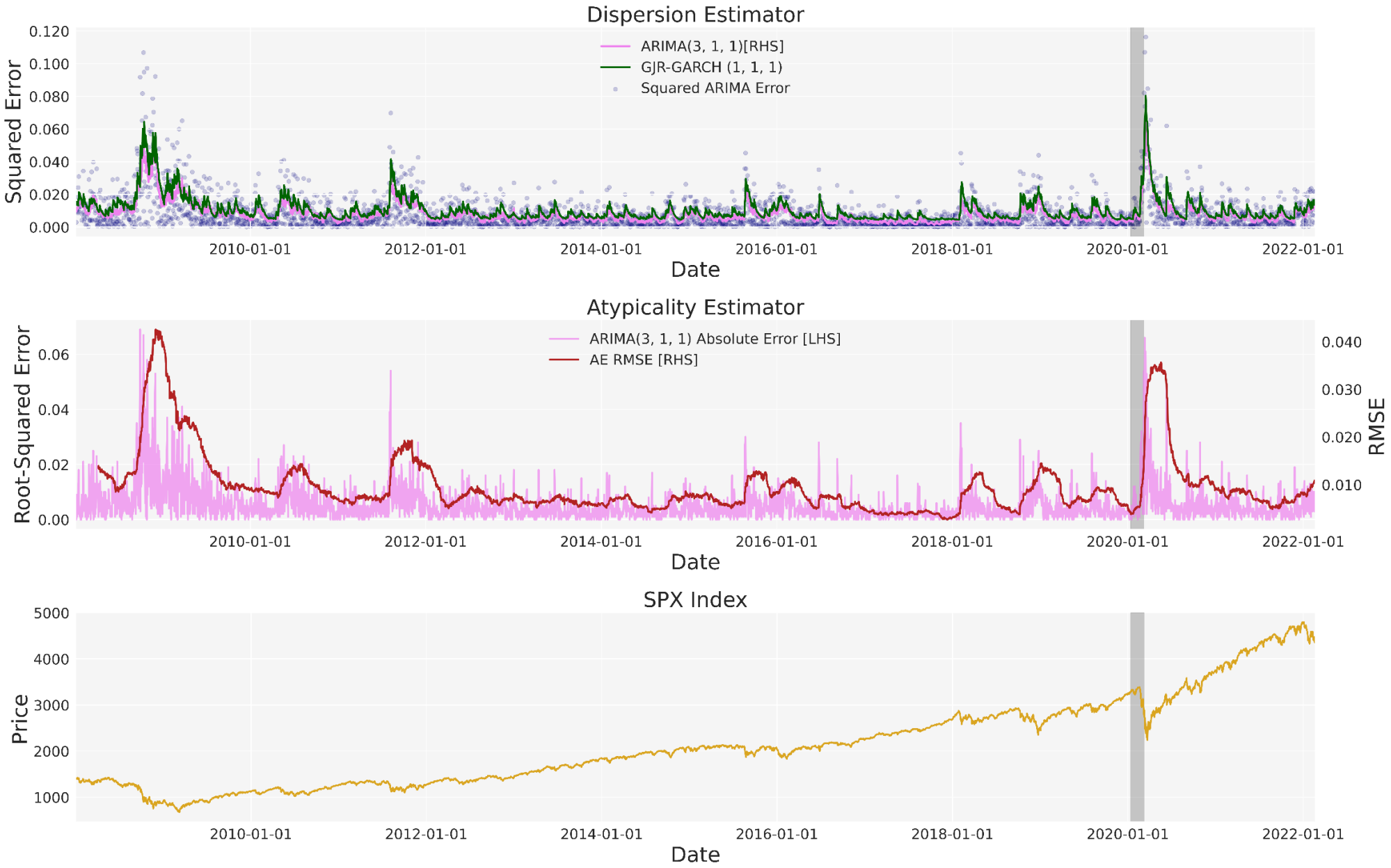


Figure 33: ACP fitted on HSI Index (HSI Category)

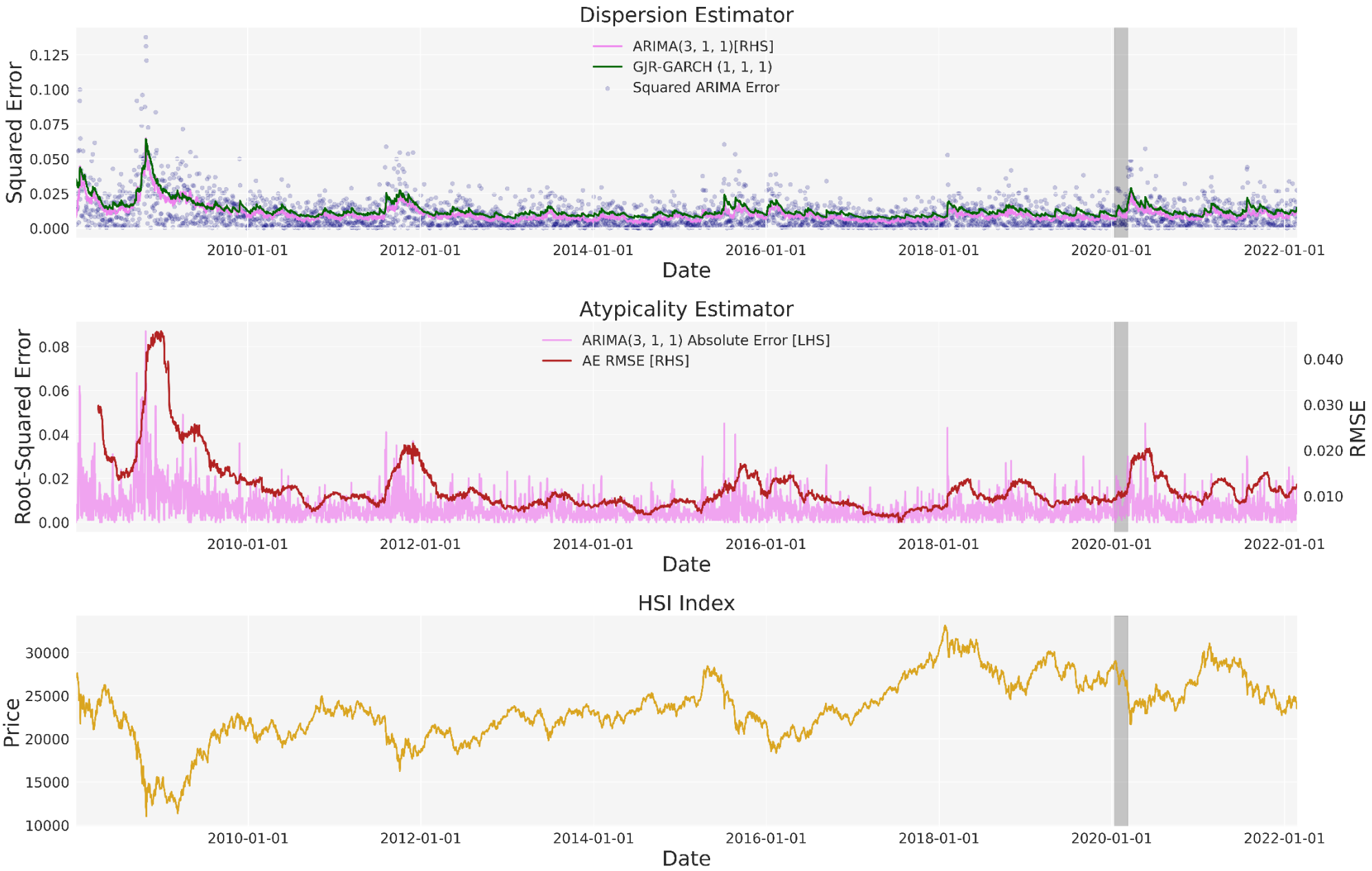


Figure 34: ACP fitted on JALSH Index (SPX-HSI Hybrid Category)

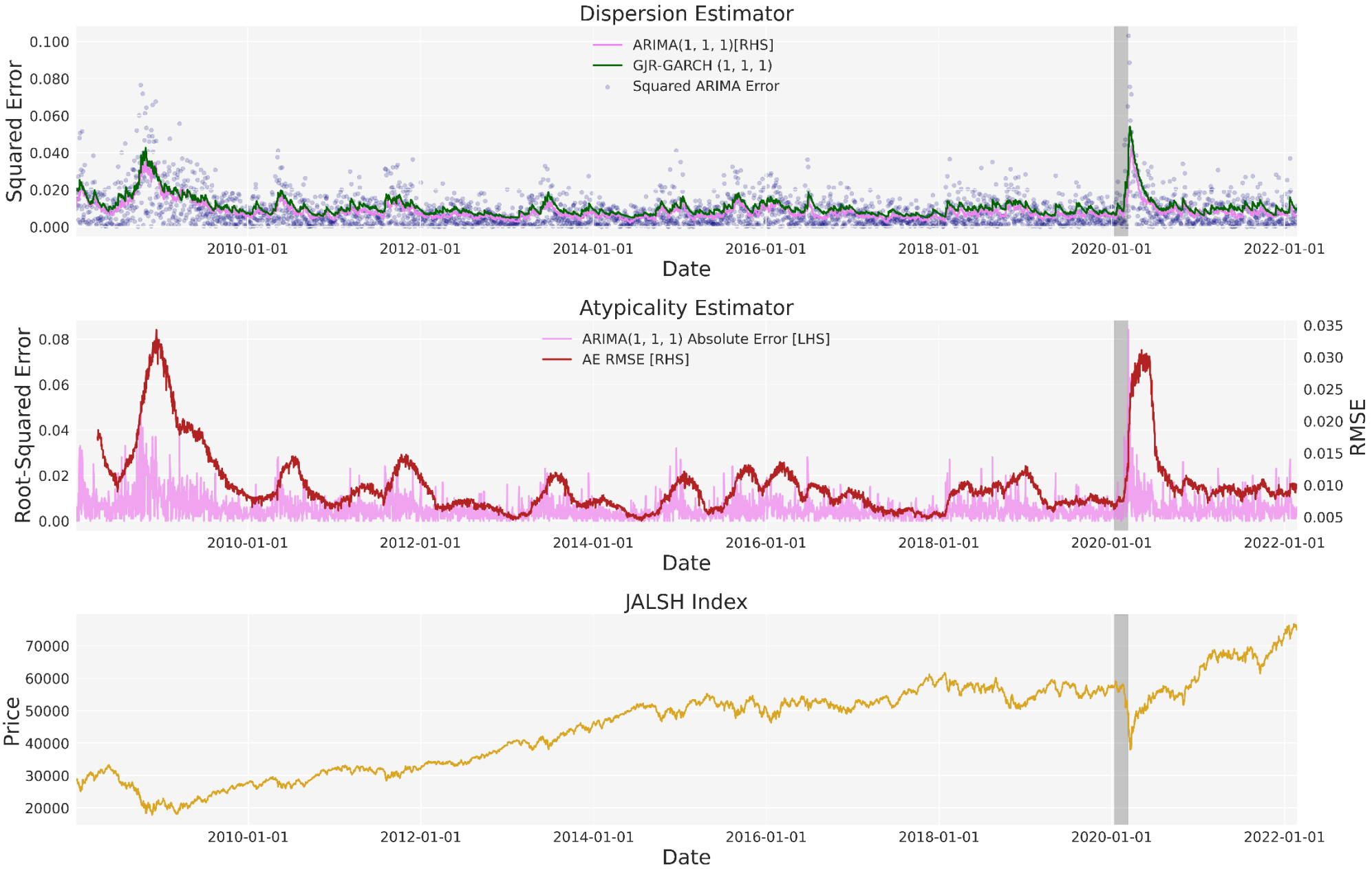


Figure 35: ACP fitted on SHSZ300 Index (Unique)

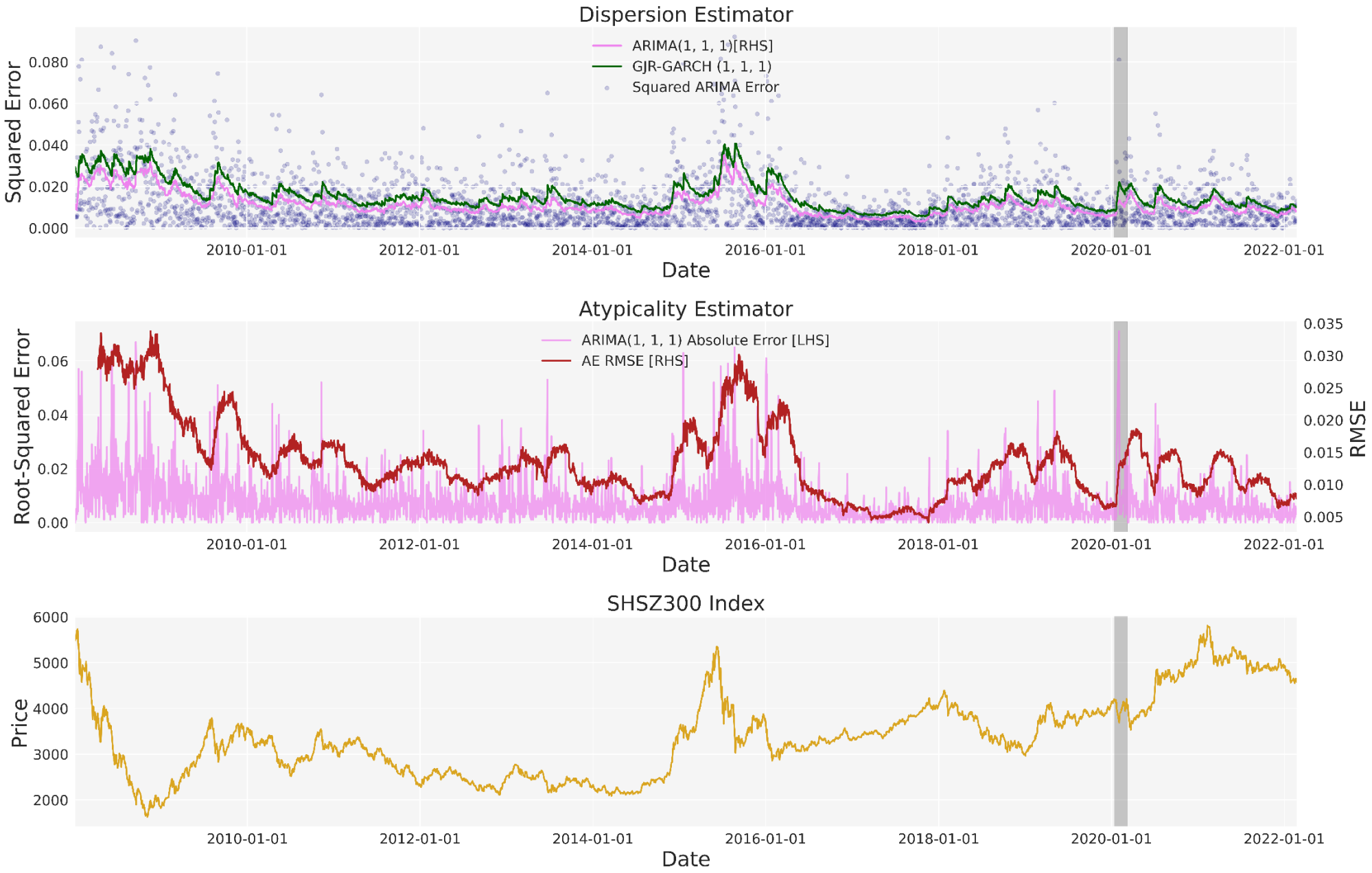
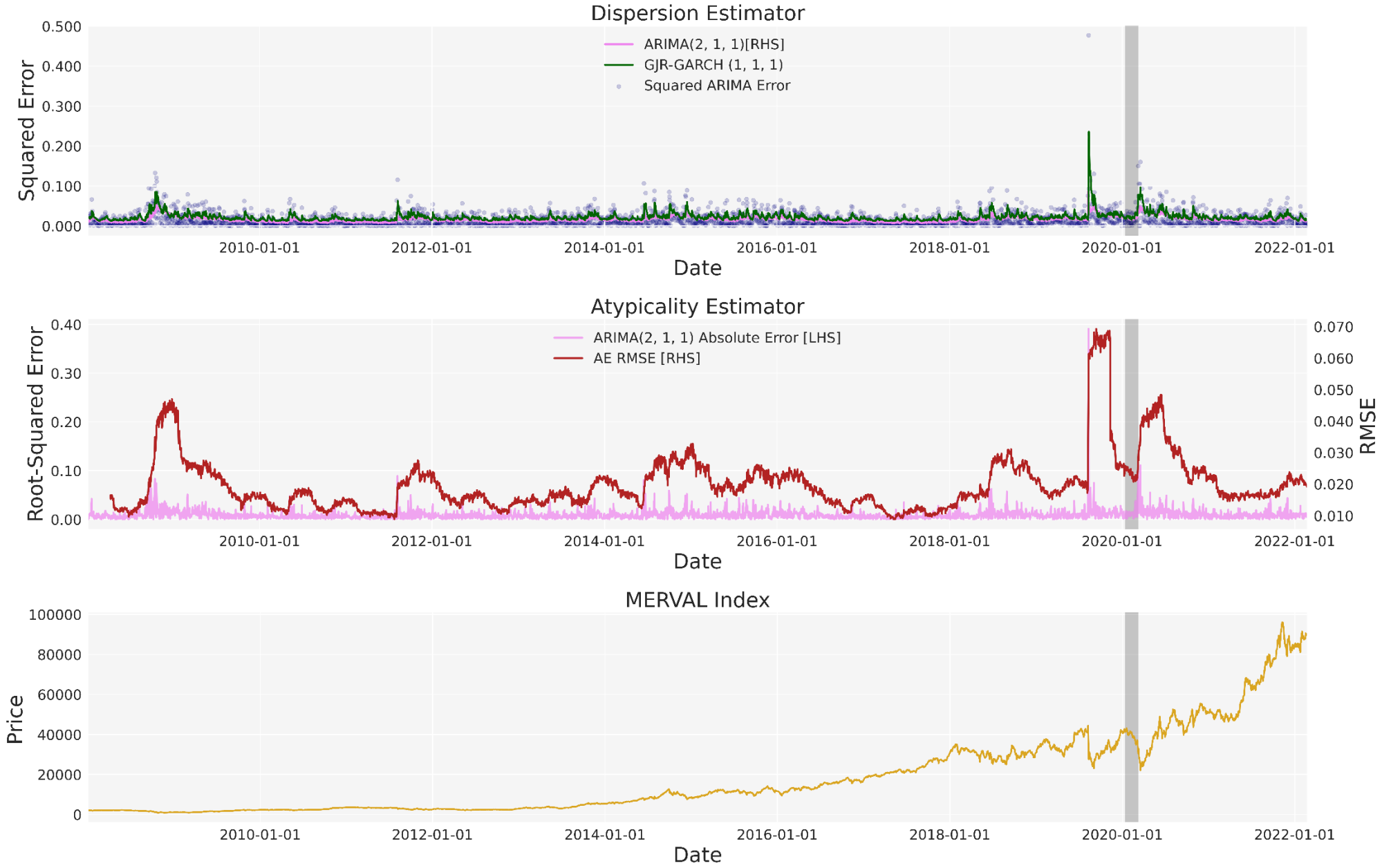


Figure 36: ACP fitted on MERVAL Index (SPX-HSI Hybrid Category)



Further observations of the ACP revealed two apparent kinds of anomalies. Those that follow our definition of Category A, occurring suddenly and reaching their maximum nearly as soon as they began, and Category B types which initially present with weak signals rapidly progressing until reaching its maximum. The majority of anomalies were characterised as the latter, according to the atypicality line, with few - such as MERVAL Index's rapid response in 2019 to undesired election news - characterised as the former. These kinds of distinctions are not clear in the dispersion line, suggesting that estimation of the typicality of an observation on day t on the basis of $t-1$ observations is insufficient to detect weak signals.

The GFC and the COVID market rout are characterised as having comparable atypicality to the SPX and SPX-HSI Hybrid Categories. For most markets, these anomalies presented with behaviour consistent with a substantial magnitude Category B. Interestingly, the atypicality of COVID for the HSI category is not anomalous on the order of magnitude as the GFC, and is in fact similar to the smaller, more commonly sized anomalies - such as the 2012-2013 Cypriot financial crisis.

4.2.3 Downside Protection: the ACP as a Potential Trading Signal

Presented in Table 5 is the static thresholding's risk-adjusted outperformance relative to a buy-and-hold strategy (which buys on the first observation of the sample and holds the position for the entire sample period)[36]. Where a developed market lacked positive rates, the U.S. interest rates were substituted due to the reasonable assumption that those domestic investors would invest in U.S. risk-free interest instruments, determining negative rates to be un-investible. The percentages under 'Outperformance of the Strategy' represent the percentage of the sample for which that threshold outperformed the buy-and-hold strategy. g is the long-term compound annualised growth rate (CAGR, higher is better), σ is the annualised standard deviation of the portfolio (lower is better) and g/σ measures the given growth per unit risk (higher is better). g_d is the annualised loss, (lower is better) and σ_d is the annualised downside deviation of the portfolio (lower is better). The Min and Max values represent the maximum single-day drawdown and upwards surge respectively.

In 77% of the 22 markets an improved risk-adjusted return could be achieved over the sample periods through use of the trading rule (at least one of the four thresholds). In 95% of the sample improved downside characteristics were realised as a result of reducing the total losses and the downside deviation significantly. Which threshold outperformed varied market to market, owing to idiosyncrasies, however the μ threshold and the 3θ tended to perform best. This illustrates some of the potential of using such a system practically in the marketplace. Table 11, an expanded version of Table 5, may be found in Appendix 7, with additional information contained in Appendix 6's Table 10.

[36] This analysis was performed over a sample that had large losses near the beginning (the GFC) and the end (COVID).

Table 5: Summary table of performance

Index	Metric	Index Return	LT Interest Rates	μ Rule	1θ Rule	2θ Rule	3θ Rule	O/P
				Performance				
Average	g	3.56%	3.03%	2.06%	2.24%	2.71%	3.20%	
	σ	22.88%	0.08%	11.84%	16.33%	17.60%	19.30%	
	g/σ	0.146	N/A	0.162	0.131	0.153	0.157	
				Outperformance of Strategy				
Total Sample	g/σ			50%	41%	50%	55%	77%
Best performer	g/σ			41%	9%	14%	18%	N/A
				Downside Performance				
Average Sample Performance	g_d	-51.41%		-28.10%	-39.87%	-44.13%	-47.28%	Yes
	σ_d	13.89%		8.03%	10.72%	11.40%	12.48%	Yes
	Min	-13.24%		-7.02%	-10.31%	-11.07%	-12.11%	Yes
	Max	11.73%		5.91%	8.22%	7.30%	8.85%	
				Outperformance of Strategy				
Total Sample	g_d			95%	91%	91%	91%	95%
	σ_d			95%	91%	91%	91%	95%
	Min			95%	77%	64%	50%	95%
	Max			95%	77%	82%	73%	5%
Best Performer	g_d			95%	5%	5%	0%	
	σ_d			0%	0%	0%	9%	
	Min			100%	27%	18%	14%	
	Max			0%	23%	18%	27%	
Overall				95%	84%	83%	76%	73%

5. Discussion

5.1 Event Study

The effects of COVID leading informational events upon global equity markets is the primary concern of this investigation. Through rigorous application of event study methodology, the progressive nature of the COVID market rout becomes apparent. It is evident that as E_4 was approached, the magnitude and lucidity of abnormal pricing behaviour increased, as did the number of significant responses in the examined sample. Consequently, the question is posed as to whether we discover weak signals of E_4 in the prior E_1, E_2 and E_3 events. Finally, the heterogeneous effects of E_4 are observed.

5.1.1 Supersectors

Through the observation of ICB Supersectors, some evidence of weak signals prior to the COVID market rout of E_4 become apparent. E_1 generated a weakly positive response on average – an unexpected result. Market participants reduced their exposures to Chemicals, Industrial Goods and Services, Banks and Energy whilst increasing exposure to Utilities and Telecommunications – with this perceptible in the significant CAARs in Figure 9. He et al. (2020) find a similar negative effect on Chinese A-shares classified in the analogous Manufacturing industry. We suggest the increased exposure in defensive supersectors suggests caution was displayed by market participants due to uncertainty arising from this information. However, the detection of the first case of coronavirus outside of China (E_1) is ascribed to have the least perceptible implication of the 4 events. As such, we further assessed veracity by attempting to observe global macroeconomic and sectoral events that may have caused uncertainty during this time period. Superficially, the negative results of the Energy supersector appear to be linked to several factors. These factors include increased crude oil volatility at the end of 2019, dropping prices, weak demand for natural gas and a heightening of tensions in the Middle East (FS Investments, 2020, International Energy Agency, 2020, Lubold et al., 2020). We suggest this pattern is more consistent with market participants expressing concern of economic slowdown - to which these supersectors are highly exposed. The defensive positioning of market participants in Utilities and Telecommunications provides further substantiation. This weakly suggests that some market participants endeavoured to protect their equity portfolios from the potential threat observed in the early phases of COVID's spread.

The benefits of our experiment design are highlighted when examining event E_2. In the E_1 evidence was discovered in the CAARs; in contrast, for E_2, evidence of weak signals is identified in the AARs. The (strong) positive event effect upon Health Care suggests that market participants assessed COVID to have a clear impact on this supersector. Conspicuous is the absence of the statistical significance in the remaining 18 supersectors' CAARs in Figure 11. However, observation of the single-day AARs reveals a negative response in Travel and Leisure and in Energy, as corroborated by the results of He et al. (2020). Travel and Leisure's response suggests that market participants had begun to assess the negative impact COVID would have on this supersector. Energy's response provides further evidence of concerns of economic slowdown. However, this event presents reduced specificity on the nature of this slowdown due to the fewer number of AAR responses in the Chemicals, Industrial Goods and Services, and Banks supersectors. The negative cumulation of the cross-sectional average AAR in Figure 10, indicating a negative event effect overall, further supports these notions.

Concerns were further exacerbated in response to E_3, displaying a perspicuously strong negative effect across supersectors (per the cumulatively negative trend of the cross-sectional average AAR). Evidently, investors subsumed the WHO recommendation for large-scale, stringent containment measures as more sizeably negative than they did E_1 and E_2. This is most discernible on and after event day in Figure 12. Furthermore, our results record a negative response versus the hitherto

positive response of the defensive Utilities supersector. In Figure 13, all significant CAARs materialised in the negative showing that market participants no longer increased exposure to defensives, instead decreasing equity exposure (on average). This suggests market participants garnered deeper concerns regarding the pandemic and had begun to determine that COVID would impact the overall global economy. Participants demonstrated uncertainty as to which components might benefit, if any, however it appears that they identified specific supersectors would be more adversely affected by COVID than others. This is evidenced in the stronger negative impacts upon Travel and Leisure, Basic Resources, Real Estate, Insurance, Construction and Materials, Food, Beverage and Tobacco, and Media. Many of these are related to global industry (the Industrials and Basic Resources industries) which by implication suggest exacerbation of economic slowdown concerns. The negative response of Industrial Goods and Services further corroborates this notion. These concerns were not limited to the aforementioned supersectors, as indicated by the negative responses Consumer Products and Services. We propound this displays evidence of market participants' uncertainty of COVID's full implications, with stronger action placed where they held greater conviction of COVID's ramifications.

E_4 records the arrival of the COVID market rout. The event effect is perspicuously negative. Overall, market participants appear to have attained a consensus view on the global repercussions of the virus. Travel and Leisure records a prodigiously negative effect in Figure 15 as market participants forecasted the coming lockdowns to proliferate globally. In addition, the expectation of a global economic slowdown appears evident in Energy's similarly sized negative response. When coupled with 14 of the remaining supersectors experiencing large negative responses, there is evidence of participants forecasting grim ramifications of COVID upon the wider economy. However, the response is disparate in some supersectors. In particular, both the Personal Care, Drug and Grocery Stores, and the Health Care supersectors are on record with large positive responses - albeit a lower magnitude than their negative counterparts. Evidence suggests that market participants ascertained these supersectors stood to benefit from COVID's economic implications. However, as these are of lower magnitude than their negative counterparts, we suggest that comparably fewer participants were confident of potential positive repercussions than they were of negative ones.

5.1.2 Geographic Regions

Resembling the supersector groupings in E_1, the geographic regions experience a correspondingly positive cross-sectional average event effect per Figure 16. However, the number of significant AARs in this event window is indicative of this effect being weak. This is supported by the lack of significant responses recorded in Spain (IBEX), a part of the U.S.A.'s market (INDU), Mexico (MEXBOL) and a part of the German market (DAX). Market participants increased portfolio exposures to Russia (IMOEX) whilst reducing exposure to Indonesia (JCI) and South-Korea (KOSPI). Antithetical to expectations, statistically significant results for the two Chinese equity markets (HSI and SHSZ300) are absent. This is in contrast to the results of He et al. (2020), who found a significant negative impact on Shanghai A-shares, and a minor, yet significant, positive impact on Shenzhen A-shares. However, when the authors grouped together both sets of A-Shares (analogous to the SHSZ300) their statistical significance evanesced. Therefore, it appears the composite nature of the SHSZ300 is obfuscating detection of significant results in these equities. Considering this, results from this study and from He et al. (2020) encapsulate minor quantities of de-risking in the concentrated area of Indonesia (JCI) and of South Korea (KOSPI) and in China (Shanghai A-Shares). Domestic investors also increased exposures to specific Chinese equities (Shenzhen A-Shares). Further supporting this notion are the results of Zhang et al. (2020) who find that subsequent to the WHO announcement, the Eastern Asian markets formed a cluster. This cluster showed an increase in corelation within, but a decrease in corelation with markets excluded from the cluster. We suggest these unusual market tensions signalled weakly of the coming rout.

Subsequently, E_2's newer and more substantial information regarding the severity of COVID's spread produced a different effect in the equity markets. In particular, a coherent change is apparent in the cross-sectional average's trend in Figure 18, from a positive to a negative cumulation with some markets responding strongly on event day. However, this response remains heterogeneous overall, and the consensus view of market participants remains indeterminate. Localised effects in Eastern and South-eastern Asian regions (Indonesia and South Korea respectively) propagated to Western Asia (Turkey , XU100) on event day and to Southern Europe (TFTSEMIB and IBEX – Italy and Spain) along with some parts of Central Europe (MEXBOL - Mexico). Liu et al. (2020) likewise record a rapid negative response in China (HSI), Malaysia (KLSE), Japan (NKY), Thailand (SET50) and Asia ex Japan. He et al. (2020) report that the Shanghai A-shares followed suit, whilst the Shenzhen A-shares diverged with a positive response. Once again, the Shanghai-Shenzhen grouping recorded statistically insignificant results, as did our results for the SHSZ300. Nevertheless, further evidence for Shanghai's negative response is presented by Al-Qudah and Houcine (2021). Furthermore, Al-Qudah and Houcine (2021) find that Western Pacific region reacted more rapidly relative to the remaining regions and consequently endured a greater downfall. The results of Singh et al. (2020) suggest these aforementioned negative responses continued until 2020/02/27 at least. Results from our study and from the literature reveal that global equity markets responded negatively through a sequence of outward propagation from a centralised Eastern Asian region. This market contagion appears to exhibit weak signals preceding the impending COVID market rout, implying this is an example of a Gray Rhino event.

E_3 provides the most cogent signals of an impending Gray Rhino event. Results in Figure 20 reveal that responses were on average negative, particularly those subsequent to event day, whilst also materialising with greater magnitudes than prior events. Emerging Markets appear to experience a materially larger effect than Developed Markets, with Harjoto et al. (2020) observing the same. Moreover, the propagation continues as South Africa (JALSH) and Saudi Arabia (SASEIDX) experience a markedly negative effect, followed by Russia (IMOEX), Turkey (XU100), Indonesia (JCI), Canada (SPTSX), Mexico (MEXBOL) and Spain (IBEX). Incongruously, Al-Qudah and Houcine (2021) suggest that the response of South Africa (JALSH) was delayed in comparison to USA (INDU), China (Shanghai Composite SSEC) and Russia (IMOEX). Nonetheless, Singh et al. (2020) corroborated with our results deviating however by claiming the UK (UKX), Germany (DAX), France (CAC), Australia (AS51), South Korea (KOSPI) and India (NIFTY) all had statistically significant results. Since our methodology is more robust against correlated cross-sectional returns, and makes use of a more dynamic return generating process in estimating ARs, we suggest that our results supersede Singh et al. (2020). Whilst the HSI (China) remains conspicuously absent from these results, He et al. (2020) dispute this purported lack of response through their sample-split between Shanghai A-share and Shenzhen A-share maintaining their respective negative and positive responses. Contrastingly, their Shanghai-Shenzhen composite remained statistically insignificant whereas we record the SHSZ300 to have a significant negative response. Considering the differences in statistical tests utilised in our results and those of He et al. (2020), the Crude Dependence Adjustment Test and the Rank Test which we utilised are documented to possess greater test power. Therefore, we defer to our results for this composite and to their results of the Shanghai and Shenzhen A-Shares sample-split. We suggest that the Shanghai A-shares had now experienced sufficiently large magnitudes of a response so as to overwhelm the positive response recorded in the Shenzhen A-shares. We perceive that the negative effects of COVID upon global equity markets had been markedly exacerbated in response to E_3. Considering the responses of prior events, a progressive accumulation of negative sentiment towards the virus is observed. The escalating nature of Events 1 to 3 indicated that any further increase of the virus would likely lead to consequences of both greater magnitude and of greater intensity. Therefore, it is evident in our results and the literature that weak signals occurred prior to the COVID market rout, implying this can be classified a Gray Rhino event.

E_4 records the beginning of the 2020 COVID equity market rout and not weak signals. This rout began in response to the lockdowns of Italy and Saudi Arabia, which was shortly followed thereafter by the WHO's pandemic classification of the disease. Both the magnitude and intensity of E_4's negative effect is perspicuous in the AAR, average AAR, and CAAR plots (Figures 22 and 23). No market showed immunity against the event effects on a daily basis – every market recorded a significant negative AAR after E_4 in Figure 22. However, 8 of the 22 markets did not record significant CAARs in Figure 23. We suggest this is indicative that the rout had not yet spread in full measure to the remaining equity markets within this event's window. Additionally, we observe that on average Emerging Markets experience larger magnitude negative event effects than Developed Markets, Harjoto et al. (2020) concur. Additionally, it is evident that market participants reached a consensus that COVID would affect the global economy negatively overall, and this notion is further supported by the literature (Al-Awadhi et al., 2020, Harjoto et al., 2020, He et al., 2020, Liu et al., 2020, Singh et al., 2020, Zhang et al., 2020, Al-Qudah and Houcine, 2021). The aforementioned literature pertaining to E_4 arrives at a concurrence that COVID affected equity markets negatively on the whole, leading to the widely reported equity market rout (Westbrook, 2020) which impacted equity markets worldwide.

5.2 Anomaly Characterisation Process

In addition to the aforementioned Event Study, our results concern the description of equity price data according to its typicality. This pursuit contains the observation of quantifiably atypical movements in the price data. We seek to detect the weak signals of a Gray Rhino event in the period preceding it. In addition, we endeavour to detect these anomalous price movements prior to the COVID market rout in particular, and thus classify them cohesively as a Gray Rhino event. In the pursuit of answering our research questions several discoveries are made.

In order to answer our pursuits, the Anomaly Characterisation Process was developed and deployed upon equity price data sampled from market representative indices of the G-20 members. Stage 1 of this process extracts the typicality information of price movements through the use of endogenous modelling (as detailed in Section 3.2.2). It implicitly demonstrates the effectiveness of this model's utilisation in the extraction of typicality information (on an ex-post basis). We can successfully describe equity price movements through a parsimonious model with three parameters – an $\mathrm{ARIMA}(1,1,1)$ model in this case. This simplicity allows for a consistent relationship to be modelled throughout the time-series whilst containing sufficient dynamism to enable efficient extraction of the typicality information via the model's error. This is true even if the modelled relationship does not fully capture all the endogenous mechanics underlying the time-series. This notion of extraction is observable in Figure 26. We infer that it is necessitated that the Stage 1 model's prediction accuracy is not perfect in order for data to be classified according to its typicality. Stage 1 could have utilised a naïve model. The naive model defines prices of security i, with error $\varepsilon'_{i,t}$ on day t calculated as the difference between observed 1st difference $y_{i,t}$ and empirical mean difference of logged prices μ_i such that:

$$\varepsilon'_{i,t} = y_{i,t} - \mu_i$$

However, this error severely limits the measures' usefulness in characterising atypicality, since $\varepsilon'_{i,t}$ indicates how distant an observed price is from the mean and is not conditional on recent market conditions. In contrast to the $\mathrm{ARIMA}(1,1,1)$ model, the naive model fails to remove the majority of the endogenous mechanics underlying the dynamics of the time-series.

Stage 2 of the ACP employs two different methods to sufficiently evince atypicality arising from $\varepsilon_{i,t}$. The dispersion estimator does this indirectly, since the utilised volatility modelling process estimates the variance σ_t^2 of $\varepsilon_{i,t}$, thereby indicating smaller (larger) values of atypicality through smaller (larger) values of σ_t^2. Concomitantly, the atypicality estimator exploits the class imbalance of atypical and typical price movements. Through the undercomplete model architecture, the model is constrained to focus on the more numerous typical price movements during model training. Therein the model learns to reconstruct typical price movement behaviour with vastly greater accuracy than it does atypical behaviour. When translated into the model's RMSE this indicates atypicality as the model is considerably less capable of reconstructing atypical price movements. If Stage 2 were not utilised, different techniques would be required in order to potentially infer atypicality.

Event study methodology deploys additional techniques to ascertain typical or atypical error; however, this process traditionally deploys exogenous models instead of endogenous. As presented in Section 3.2.1, this methodology requires strict experiment design in order for analysis to be valid. In particular, event windows are placed around expertly selected dates which are expected a priori to have a price effect. The model error is then assessed via several statistical tests which is highly effective in ex-post analysis. However, it has the primary disadvantage of requiring the researcher to have prior expert knowledge of the specific event(s) studied. This prohibits timeliness of analysis and any application of insights upon future events necessarily rely on the assumption that this historical information is valid for the future event. Furthermore, the analysis is limited to the short event window, with any supplemental windows requiring additional models trained on dissimilar samples – consequently possessing differing parameters. The latter diminishes the effectiveness of comparative analysis of different periods. For example, comparison of the 2008 Global Financial Crisis with the 2020 COVID market rout. Although they can be compared productively through event study analysis, we cannot compare the two events on an identical basis due to changes in the pricing relationship modelled (the market model).

In contrast, the ACP avoids these problems since length of sample is unconstrained and thus analysis does not require differing models for different periods of the time-series. The same relationship is modelled throughout. Due to this consistency, we can compare different event's effects occurring in different time periods on the same basis. Furthermore, provided there is sufficient data, the timeliness of ACP-based analysis is currently only constrained by frequency of data – we can characterise up to a single lag of data. Consequently, we do not require prior expert knowledge of an event and can potentially discover the event as it unfolds (depending on the rapidity and magnitude of the event's effect). Additionally, we demonstrate that it is possible to effectively describe the typicality of equity price data via the Anomaly Characterisation Process. This method does not require deployment of commonly used traditional statistical techniques, such as modelling a fit to some known statistical distribution in order to detect outliers through empirical deviation from the assumed distribution. This method is also free from the techniques necessitated by the event study methodology. Through the Atypicality Estimator component, the ACP fits the empirical distribution and thus forgoes assumptions of the data's distributional nature. The outliers are described and detected through contextually atypical changes in this description, circumnavigating the need to (potentially erroneously) assume the data comprises specific distributional characteristics. This is, however, different to simple comparison with the empirical distribution. The ACP is not limited to historical data and can effectively characterise unseen data with the same efficacy as it does historical data (on a numerically continuous basis as opposed to numerically discrete one). The AE implicitly ignores outliers through its undercomplete (constrained) model architecture, ergo estimating the 'true' empirical distribution sans outliers. Finally, the ACP places anomalies on the same standing and thereby handles the problem of heterogeneity.

The event study revealed that the approach of the COVID event was marked by weak signals, consequently implying classification as a Gray Rhino event per the definition supplied in Section 2.1. This implied classification stands in contrast to other literature (He et al., 2020, Singh et al., 2020). The weak signals were at first subtle but became detectably unusual and escalated as the time-series proceeded towards the event. These signals commenced 3 months prior to the COVID market rout – at earliest and in their least discernible form. In analysis of ACP results on the respective equity indices, corresponding to the Geographic Regions grouping of the event study, some evidence of these weak signals manifests graphically. This evidence is predominantly in the SPX Category. These signals do not begin to present until the commencement of E_2 and occur in the dispersion line at this point. Thereafter, in both the dispersion and atypicality lines, the signals are clarified in response to E_3 with their culmination shortly after the commencement of E_4. Preceding all other markets, a significant anomalous price movement in the SHSZ300 was discovered - this being most perceptible in the atypicality line. It is therefore suggested that an anomaly characterisation process can detect anomalous price movements in the data prior to the COVID rout. This is particularly the case in the presence of a strong, simultaneous divergence of both the atypicality and dispersion lines from their recent trends. Consequently, it is conceivable that a version of the ACP running on higher frequency live data, would be able to provide practitioners weak and strong signals of events at higher resolution and with more rapidity. Thereby, timeous detection of Gray Rhino events could occur. Such insight would allow practitioners to act faster than less informed competitors, thereby enabling them to protect their equity portfolios pre-emptively through tail risk hedging. Our simple rule-trading simulation supports this notion, whereby risk-adjusted outperformance was achieved 77% of the time through a simple threshold (whose value was dependent on each market). Moreover, for 95% of the total sample, both the downside and the portfolio's volatility were mitigated - the latter with the result of less extreme returns.

Our results demonstrate that an undercomplete deep autoencoder using simple neural networks can be utilised successfully in detection of anomalies within financial time-series (provided this data is first transformed per Stage 1). This is in contrast to more complex variants constructed by other researchers upon various kinds of non-financial data such as: the Variational AEs by Ikeda et al. (2018), the Enhanced Long-Short-Term-Memory-AE by Malekia et al. (2021), and the recurrent neural network based AEs of various other researchers (Assendorp, 2017, Nucci et al., 2018, Hundman et al., 2018, Nolle et al., 2018). We suspect that further improvements may be found in more powerful model designs such as these, particularly in relation to recurrent neural networks with their explicit modelling of time-dependent relationships. Additionally, the literature demonstrates the efficacious adaptation of autoencoder-type deep learning models in adapting to different kinds of data for the purposes of anomaly detection. For example Internet of Things data (Kieu et al., 2019, Kieu et al., 2018), COVID medical data (Homayouni et al., 2021), and various other kinds of non-financial time-series data (Chen et al., 2017). Our results contribute to the evidence of this model structure's adaptation to equity price time-series data.

Additional insights are garnered regarding the ACP through it's deployment on the varied sample of G-20 equity indices. Our model choice for Stage 2 is demonstrably superior to a best-fit ARIMA model whose parameters are selected via an automated process. The GJR-GARCH is more reactive, with an apparent closer fit, and the concomitant AE model provides a clarified signal in its RMSE as opposed to the noisy signal of the auto-ARIMA model's error. Stage 2's utilisation of the two models provides synergistic insight into the typicality of the analysed data, aiding the detection of point and contextual anomalies. This enables the classification of dissimilar anomalous events through identification of differing characteristics.

Through the deployment of the ACP on indices representing the global equity markets, commonalities between the differing markets with respect to their typicality were discovered and then grouped into three different categories through graphical analysis. Our results - contribute to new insight into the relationships between the G-20 indices in the literature. Consequently, global investors may, for example, develop their own ACP systems and deploy them upon their universe of investible assets in order to analyse the similarities and divergences therein. Similarly, researchers may analyse assets' relationships from the perspective of the data's typicality in lieu of more commonly used techniques such as correlation. This study shows that the equity indices in the sample have tended towards three kinds of general behaviours since the beginning of 2008. These are categorised as the SPX Category, the HSI Category and the SPX-HSI Hybrid Category. We discuss the former two categories. The HYBRID is defined as some combination of the distinct characteristics identified in each of the other two. Plots of the dispersion line indicate the SPX category presents with sudden, atypical price movements throughout periods of calm and that these atypical movements often follow clustering behaviour. This category continues to display heightened and clustered atypicality post-COVID up until the remainder of the sample. In contrast, the HSI Category displays smaller magnitude atypical price movements which present in a smoother dispersion line than in the SPX Category. We suggest this is a consequence of the normal returns generating process, as the highly reactive dispersion estimator did not spike upwards as frequently or with the same clustering behaviour as the SPX Category. The HSI then returns to its base level post-COVID, with smoothness similar to known calm periods, again unlike the SPX Category. Atypicality for both categories plainly illustrated larger magnitude atypical events. For the SPX Category, the atypicality of COVID is similar to the atypicality of the GFC, whereas for the HSI category, COVID presents far less atypically than the GFC. By concurrently examining price plots, it is observed that the SPX category presents with an apparent exponential trend whereas the HSI category presents with mean-reversion behaviour. The apparent exponential trend of the SPX is a possible cause for both the large, sudden spikes in the dispersion, as well as the calm periods presented by the atypicality line. The apparent mean-reversion tendency of the HSI category may be responsible for its smooth single day responses recorded by the dispersion line.

The SHSZ300 is an exception to the aforementioned three categories. During the 2015–2016 Chinese stock market turbulence the atypicality line records similar magnitudes of atypicality to the GFC. Similar to the HSI, its atypicality during the COVID period is heightened to a lesser degree than during the GFC (per both measures). Per the dispersion line, this unique market records greater one day atypicality during 2015–2016 Chinese stock market turbulence than it does during GFC. According to He et al. (2020) the apparent mixed-distribution of the SHSZ300's atypicality line is explained by the negatively correlated reactions of Shenzhen and Shanghai A-shares. Based on both our and He et al. (2020) event study analyses, the unusual atypicality of the SHSZ300 appears to be due to dual-market composite nature: equities in the Shenzhen A-share market deviate from the expected behaviour of the global market.

Two kinds of anomalous events are implicitly present in the ACP results. The first are those that appear to follow Taleb's (2004, 2007) definition of Black Swan events: "extreme impact" events that are only predictable in hindsight. The second are those that appear to follow the definition of a Gray Rhino: weak signals that escalate in magnitude up until the event's arrival. Significant in the ACP results is the pervasiveness of the GFC event's impact upon all markets examined. It is present in general public parlance as well as being suggested by Swango (2020) that the GFC is a Black Swan event. We are inclined to agree due to the pervasiveness of its 'extreme impact', however determination of this is outside of our scope. The magnitude of impact of MERVAL's 2019/08/12 event and the lack of prior signals (suggesting the event is only predictable in hindsight) suggest that it is a Black Swan event. In contrast, we suggest the COVID market rout of 2020 is a Gray Rhino. Both our results and the literature provide evidence of weak signals prior

to the event. Furthermore, both our Event Study and ACP results illustrate progressive exacerbation of these signals.

6. Conclusions and Recommendations for Further Research

Our study, by observing the defining presence of antecedent weak signals in equity stock prices, sought to determine whether Gray Rhino events are observable employing an innovative combination of traditional and novel methodologies. We implement event study methodology to this end and discover weak signals prior to the market rout. Concurrent to this we formulate and deploy a novel anomaly characterisation process in order to characterise equity price data according to its typicality facilitating the discovery of atypical price changes. Hereby we detect weak signals whilst enabling the examination and comparison of different events on the same basis of comparison. These research goals are attained on a sample of indices representing the G-20 equity markets.

6.1 Event Study

In order to detect weak signals prior to the COVID market rout, we examined the effects of significant information concerning COVID, prior to (E_1 to E_3) and including the rout (E_4). Moreover, we examine whether carefully specified event study can detect these ex-ante weak signals. The literature identifies its effect to be vastly negative upon global equity markets and suggests that there are effects prior to this event. However, prior studies have not explicitly investigated the potential of prior weak signals. Our results suggest that the COVID market rout is preceded by weak signals. We suggest the first signal we identify (E_1) is related to concerns of economic slowdown linked with defensive positioning of global investors. Concurrently, we add to the evidence that minor reductions of investor's exposures to the South-eastern and Eastern Asia area occurred in response. We suggest the observed signal begins to exacerbate in response to supplemental and comparatively clearer information regarding COVID (E_2) as the market response is likewise clarified. Furthermore, we suggest that additional information resulted in the reversal of the aforementioned defensive positioning recorded (E_1) as a market consensus to COVID began to assemble (E3). Finally, we corroborate the consensus view of the literature of COVID on the representative G-20 equity indices; that it resulted in a large negative response across equity global markets (E4). Concurrently we discover the positive responses certain ICB supersectors experienced and suggest it is due to investors that expected the supersectors to benefit positively from COVID's ramifications. Our results demonstrate prior weak signals and record its effects, and we thereby suggest that COVID is appropriately classified as a Gray Rhino event, opposing classification as a Black Swan event by He et al. (2020).

Possible extensions to the event study include replacing the (exogenous) market model with ARIMA-type endogenous models. This may enable more exacting ARs and may improve the rigor of the statistical analysis, and thus improve detection of effects – clarifying the weak signals of E_1 and E_2. The utilisation of larger sample sizes may increase the depth of understanding and allow finer granularization of industry-related groupings and further clarification of weak signals. Additional research into appropriate test statistics would be required. Alternatively, additional veracity of statistical analysis may be obtained through deployment of more powerful test-statistics such as the Generalised Rank Test by Kolari and Pynnonen (2011). Thereby, future researchers may be able to clarify the apparent discrepancies in the SHSZ300. We suggest that researchers sample the Shenzhen and Shanghai A-shares separately to aid their analysis by avoiding the apparent mixed distribution of the SHSZ300.

6.2 Anomaly Characterisation Process

To describe equity price movements according to their typicality, we developed and deployed an anomaly characterisation process and uncovered its efficacy through graphical analysis. The process is shown to be able to effectively describe equity price movements' typicality on a quantitative basis whilst yielding supplementary insight into the commonalities and disparities between the different equity markets. This enabled us to observe that many market events present

with generic changes in the magnitude of atypicality measured which rapidly deteriorate into the main event's peak atypicality. We observe the COVID rout to follow this pattern. Therefore, we suggest the ACP can be used to identify weak signals, and that these present prior to the COVID market rout. Consequently, this reinforces our view of its Gray Rhino status, since per definition weak signals do not precede Black Swan events but do precede Gray Rhino events.

The efficacious ACP describes the typicality of prices through their deviance from the expected endogenous mechanics and enables us to classify the representative G-20 equity market indices into three groups – one of which is a hybrid of the other two. Through examination we discover that the COVID market rout is similar in atypicality to the GFC for the SPX Category, but this atypicality is divergent for HSI Category markets where they record the COVID rout as being less atypical than the GFC. This stands in contrast to the response of all markets to the GFC, whereby we find similar magnitudes of atypicality across the sample. Since the GFC is commonly viewed as a Black Swan type event, this contrast supplements our view of COVID being a Gray Rhino. In addition to this, we contribute to the literature a novel method for the analysis of equity market relationships by characterising and comparing their atypicality in response to the same global events. Furthermore, we discover the unique atypicality of the SHSZ300 and further support the notion of using alternative groupings for these equities in order to clarify analysis thereof. Finally, we observe what we suggest being a 'true' Black Swan event in the MERVAL index. This event presents with near instantaneous repercussions whilst following the definition of its ex-ante forecast appearing unfeasible. Since we do not observe any antecedent weak signals, this further supports our classification of COVID as Gray Rhino. Black Swan events cannot be predicted, and thus weak signals do not precede them.

In addition to answering our research questions, several supplemental insights are gained. We contrast our novel method with the aforementioned event study methodology and discover the merits of both. Similarly, we observe that the ACP resolves several issues related to anomaly detection including those of heterogeneity and rarity. We also contribute to the literature by demonstrating the autoencoder model structure can be used upon equity price time-series, in contrast to the literature's use of different model structures upon the same kind of data and similar autoencoder models upon different kinds of data (see Sections 2.3.5 and 2.3.6).

We suggest that the utilization of an ACP upon intraday data may increase its practical use through reduction of the system's lag. Additional data-engineering as well as more complex transformation of the data in Stage 1 is anticipated to reach appropriate levels of distinction between observations. Higher frequency sampling presents unique challenges, primarily in the form of the data's tendency towards degenerate distributions which are difficult to model. Furthermore, the use of recurrent neural network (RRN) layers in the autoencoder may provide demonstrable improvements. We expect this to capture additional longitudinal relationships present in the data, with the option to add a concurrent cross-sectional basis. This process may capture panel relationships and potentially generate a clearer characterisation of equity price data whilst potentially being able to uncover group anomalies. Finally, the introduction of a new Stage 3, whereby the Dispersion and Atypicality lines are forecasted, may provide additional practical use to investors in the form of faster risk mitigation and the potential to trade the systems' signals.

References

AGGARWAL, C. C. 2018. *Teaching Deep Learners to Generalize*, Springer International Publishing AG.

AITKEN, A. C. 1936. IV.—On Least Squares and Linear Combination of Observations. *Proceedings of the Royal Society of Edinburgh,* 55**,** 42-48.

AKAIKE, H. 1998. Information Theory and an Extension of the Maximum Likelihood Principle. *In:* PARZEN, E., TANABE, K. & KITAGAWA, G. (eds.) *Selected Papers of Hirotugu Akaike.* New York, NY: Springer New York.

AL-ATHARI, F. M. & AL-AMLEH, M. A. 2016. A Comparison between Least Trimmed of Squares and MMEstimation in Linear Regression Using Simulation Technique. *5th International Arab Conference on Mathematics and Computations 2016.* Zarqa University, Jordan: Zarqa University.

AL-AWADHI, A. M., ALSAIFI, K., AL-AWADHI, A. & ALHAMMADI, S. 2020. Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns. *Journal of behavioral and experimental finance,* 27**,** 100326-100326.

AL-QUDAH, A. A. & HOUCINE, A. 2021. A.Stock markets' reaction to COVID-19: evidence from the six WHO regions. *Journal of Economic Studies,* ahead-of-print.

ALEXANDER, C. 2008a. Asymmetric GARCH Models. *Market Risk Analysis: Practical Financial Econometrics.* The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons Ltd.

ALEXANDER, C. 2008b. *Value-at-Risk Models,* The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, John Wiley & Sons, Ltd.

ALGHAMDI, T., ELGAZZAR, K., BAYOUMI, M., SHARAF, T. & SHAH, S. Forecasting Traffic Congestion Using ARIMA Modeling. 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), 2019/06/28 2019. 1227-1232.

ALZYOUT, M., ALSMIRAT, M. & AL-SALEH, M. I. Automated ARIMA Model Construction for Dynamic Vehicle GPS Location Prediction. 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS), 2019/10/25 2019. 380-386.

ANDERSEN, T. G., BOLLERSLEV, T., CHRISTOFFERSEN, P. F. & DIEBOLD, F. X. 2005. Practical Volatility and Correlation Modeling for Financial Market Risk Management. *National Bureau of Economic Research Working Paper Series,* No. 11069.

ANDREWS, J., TANAY, T., MORTON, E. & GRIFFIN, L. 2016. Transfer representation-learning for anomaly detection. *33rd International Conference on Machine Learning.* New York, NY, USA: UCL Discovery.

ANSOFF, H. I. 1975. Managing strategic surprise by response to weak signals. *California Management Review,* 18**,** 21-23.

ARIAN, H., POORVASEI, H., SHARIF, A. & ZAMANI, S. 2020. The Uncertain Shape of Grey Swans: Extreme Value Theory with Uncertain Threshold. *New Economics Papers***,** 1-24.

ASSENDORP, J. P. 2017. *Deep learning for anomaly detection in multivariate time series data.* Master of Science Informatik Masters, Hochschule für Angewandte Wissenschaften Hamburg.

AU YEUNG, J. F. K., WEI, Z.-K., CHAN, K. Y., LAU, H. Y. K. & YIU, K.-F. C. 2020. Jump detection in financial time series using machine learning algorithms. *Soft computing (Berlin, Germany),* 24**,** 1789-1801.

BERNHARDT, D. & ECKBLAD, M. 2013. *Stock Market Crash of 1987* [Online]. Chicago: Federal Reserve Bank of Chicago. Available:

https://www.federalreservehistory.org/essays/stock-market-crash-of-1987 [Accessed 2021/10/08 2021].
BOEHMER, E., MUSUMECI, J. & POULSEN, A. 1991. Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics,* 30**,** 253-272.
BOLLERSLEV, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics,* 31**,** 307-327.
BOX, G. E., JENKINS, G. M., REINSEL, G. C. & LJUNG, G. M. 2015. *Time series analysis: Forecasting and Control*, John Wiley & Sons.
BREUSCH, T. S. 1978. Testing for Autocorrelation in Dynamic Linear Models *Australian Economic Papers,* 17**,** 334-355.
BROOKS, C. & PERSAND, G. 2003. Volatility forecasting for risk management. *Journal of Forecasting,* 22**,** 1-22.
BROWN, S. J. & WARNER, J. B. 1985. Using daily stock returns: The case of event studies. *Journal of financial economics,* 14**,** 3-31.
BUDA, T. S., CAGLAYAN, B. & ASSEM, H. 2018. DeepAD: A Generic Framework Based on Deep Learning for Time Series Anomaly Detection. *In:* PHUNG, D., TSENG, V. S., WEBB, G. I., HO, B., GANJI, M. & RASHIDI, L. (eds.) *Advances in Knowledge Discovery and Data Mining.* Melbourne, VIC, Australia: Springer International Publishing.
CAO, L. 2019. AI Pioneers in Investment Management. CFA Institute.
CAO, L. 2021. AI in Finance: Challenges, Techniques and Opportunities. *ACM Computing Surveys,* 1**,** 40.
CFI, C. F. I. 2020. Black Swan Event-Guide to Unpredictable Catastrophic Events. *Corporate Finance.Institute* [Online]. Available: https://corporatefinanceinstitute.com/resources/knowledge/finance/black-swan-event/.
CHALAPATHY, R. & CHAWLA, S. 2019. Deep learning for anomaly detection: A survey. *Computer Research Repository CoRR,* abs/1901.03407.
CHALAPATHY, R., KRISHNA, A. M. & CHAWLA, S. 2017. Robust, Deep and Inductive Anomaly Detection. *Machine Learning and Knowledge Discovery in Databases.* Skopje, Macedonia: Springer, Cham.
CHALAPATHY, R., MENON, A. K. & CHAWLA, S. 2019. Anomaly Detection using One-Class Neural Networks. *arXiv pre-print server*.
CHANDOLA, V., BANERJEE, A. & KUMAR, V. 2009. Anomaly detection: A survey. *ACM Computing Surveys,* 41**,** Article 15.
CHEN, J., SATHE, S., AGGARWAL, C. & TURAGA, D. 2017. Outlier Detection with Autoencoder Ensembles. *Proceedings of the 2017 SIAM International Conference on Data Mining (SDM).* The Westin Galleria Houston, Houston, Texas, USA: Society for Industrial and Applied Mathematics SIAM.
CHEN, N.-F., ROLL, R. & ROSS, S. A. 1986. Economic Forces and the Stock Market. *The Journal of business (Chicago, Ill.),* 59**,** 383-403.
CORRADO, C. J. 1989. A nonparametric test for abnormal security-price performance in event studies. *Journal of financial economics,* 23**,** 385-395.
CORRADO, C. J. 2011. Event studies: A methodology review. *Accounting & Finance,* 51**,** 207-234.
CORRADO, C. J. & ZIVNEY, T. L. 1992. The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns. *Journal of financial and quantitative analysis,* 27**,** 465-478.
COWAN, A. R. 1992. Nonparametric event study tests. *Review of Quantitative Finance and Accounting,* 2**,** 343-358.
DICKEY, D. A. & FULLER, W. A. 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association,* 74**,** 427-431.

EBBERTH, L. P., LADEIRA, M., CARVALHO, R. N. & MARZAGÃO, T. 2016. Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering. *15th IEEE International Conference on Machine Learning and Applications (ICMLA).* Anaheim, CA, USA: IEEE.

ENGLE, R. F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica,* 50**,** 987-1007.

ERFANI, S. M., RAJASEGARAR, S., KARUNASEKERA, S. & LECKIE, C. 2016. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition,* 58**,** 121-134.

FAMA, E. F., FISHER, L., JENSEN, M. C. & ROLL, R. 1969. The Adjustment of Stock Prices to New Information. *International Economic Review,* 10**,** 1-21.

FERGUSON, N. 2020. Black Swans, Dragon Kings and Gray Rhinos: The World War of 1914-1918 and the Pandemic of 2020-? *Hoover Institution History Working Paper Series,* 2020-1**,** 1-46.

FS INVESTMENTS. 2020. *Energy market commentary: January 2020* [Online]. FS Investments. [Accessed].

FU, X., LUO, H., ZHONG, S. & LIN, L. 2019. Aircraft engine fault detection based on grouped convolutional denoising autoencoders. *Chinese Journal of Aeronautics,* 32**,** 296-307.

GILES, C. L., LAWRENCE, S. & TSOI, A. C. 2001. Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference. *Machine Learning,* 44**,** 161-183.

GLETTE-IVERSEN, I. & AVEN, T. 2021. On the meaning of and relationship between dragon-kings, black swans and related concepts. *Reliability Engineering & System Safey,* 211**,** 13.

GLOPID-R 2020. COVID-19 Public Health Emergency of International Concern (PHEIC) Global research and innovation forum. Global Research Collaboration for Infections Disease Preparedness.

GLOSTEN, L. R., JAGANNATHAN, R. & RUNKLE, D. E. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance,* 48**,** 1779-1801.

GODFREY, L. G. 1978. Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica,* 46**,** 1293-1301.

GUO, Y., LIAO, W., WANG, Q., YU, L., JI, T. & LI, P. 2018. Multidimensional Time Series Anomaly Detection: A GRU-based Gaussian Mixture Variational Autoencoder Approach. *In:* ZHU, J. & TAKEUCHI, I. (eds.) *ACML 2018.*

HAMMOND, P. J. 2016. Catastrophic Risk, Rare Events, and Black Swans: Could There Be a Countably Additive Synthesis? *The Economics of the Global Environment, Studies in Economic Theory,* 29**,** 17-38.

HARJOTO, M. A., ROSSI, F. & PAGLIA, J. K. 2020. COVID-19: stock market reactions to the shock and the stimulus. *Applied Economics Letters,* 28**,** 795-801.

HAWKINS, S., HE, H., WILLIAMS, G. & BAXTER, R. 2002. Outlier Detection Using Replicator Neural Networks. *In:* KAMBAYASHI, Y., WINIWARTER, W. & ARIKAWA, M. (eds.) *DaWaK: International Conference on Data Warehousing and Knowledge Discovery.* 2002/09/02 ed. Aix-en-Provence, France: Springer Link.

HE, P., SUN, Y., ZHANG, Y. & LI, T. 2020. COVID–19's Impact on Stock Prices Across Different Sectors—An Event Study Based on the Chinese Stock Market. *Emerging Markets Finance and Trade,* 56**,** 2198-2212.

HOMAYOUNI, H., RAY, I., GHOSH, S., GONDALIA, S. & KAHN, M. G. 2021. Anomaly Detection in COVID-19 Time-Series Data. *SN Computer Science,* 2.

HUI, C.-H., WONG, J. & LI, K.-F. 2013. Price Disparities between Mainland China's Onshore and Offshore Financial Markets. *In:* WONG, M. C. S. & CHAN, W. F. C. (eds.)

Investing in Asian Offshore Currency Markets: The Shift from Dollars to Renminbi. London: Palgrave Macmillan UK.

HUNDMAN, K., CONSTANTINOU, V., LAPORTE, C., COLWELL, I. & SODERSTROM, T. 2018. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *International Conference on Knowledge Discovery and Data Mining.* ACM.

IKEDA, Y., TAJIRI, K., NAKANO, Y., WATANABE, K. & ISHIBASHI, K. 2018. Estimation of Dimensions Contributing to Detected Anomalies with Variational Autoencoders.

INTERNATIONAL ENERGY AGENCY 2020. Global Energy Review 2020. Paris: International Energy Agency.

JAMIL, M. S. & AKBAR, S. Taxi passenger hotspot prediction using automatic ARIMA model. 2017 3rd International Conference on Science in Information Technology (ICSITech), 2017/10/26 2017. 23-28.

JI, Z., GONG, J. & FENG, J. 2021. A Novel Deep Learning Approach for Anomaly Detection of Time Series Data. *Scientific programming,* 2021.

KIEU, T., YANG, B., GUO, C. & JENSEN, C. S. 2019. Outlier detection for time series with recurrent autoencoder ensembles. *Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19).* Macao, China: International Joint Conferences on Artificial Intelligence.

KIEU, T., YANG, B. & JENSEN, C. S. 2018. Outlier detection for multidimensional time series using deep neural networks. In 2018 19th IEEE International Conference on Mobile Data Management (MDM), pages 125–134. IEEE, 2018. *19th IEEE International Conference on Mobile Data Management.*

KOENKER, R. 1981. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics,* 17**,** 107-112.

KOLARI, J. W. & PYNNONEN, S. 2011. Nonparametric rank tests for event studies. *Journal of Empirical Finance,* 18**,** 953-971.

KUHN, T. S. 1970. *The Structure of Scientific Revolutions,* Chicago, London, University of Chicago Press.

KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P. & SHIN, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics,* 54**,** 159-178.

LECUN, Y. A., BOTTOU, L., ORR, G. B. & MÜLLER, K.-R. 2012. Efficient BackProp. *In:* MONTAVON, G., ORR, G. B. & MÜLLER, K.-R. (eds.) *Neural Networks: Tricks of the Trade: Second Edition.* Berlin, Heidelberg: Springer Berlin Heidelberg.

LIU, H., MANZOOR, A., WANG, C., ZHANG, L. & MANZOOR, Z. 2020. The COVID-19 Outbreak and Affected Countries Stock Markets Response. *International Journal of Environmental Research and Public Health,* 17.

LUBOLD, G., YOUSSEF, N. A. & COLES, I. 2020. Iran Fires Missiles at U.S. Forces in Iraq. *The Wall Street Journal*, 2020/01/08.

LYON, J. D. & TSAI, C.-L. 1996. A Comparison of Tests for Heteroscedasticity. *Journal of the Royal Statistical Society. Series D (The Statistician),* 45**,** 337-349.

MACKINLAY, A. C. 1997. Event Studies in Economics and Finance. *Journal of Economic Literature,* 35**,** 13-39.

MALEKIA, S., MALEKIB, S. & JENNINGS, N. R. 2021. Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering. *Applied Soft Computing,* 108.

MCWILLIAMS, A. & SIEGEL, D. 1997. Event Studies In Management Research: Theoretical And Empirical Issues. *Academy of Management Journal,* 40**,** 626-657.

MÉLARD, G. & PASTEELS, J. M. 2000. Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting,* 16**,** 497-508.

MOEWS, B., HERRMANN, J. M. & IBIKUNLE, G. 2019. Lagged correlation-based deep learning for directional trend change prediction in financial time series. *Expert Systems with Applications,* 120**,** 197-206.

MOHAPATRA, R. K., PINTILIE, L., KANDI, V., SARANGI, A. K., DAS, D., SAHU, R. & PEREKHODA, L. 2020. The recent challenges of highly contagious COVID-19, causing respiratory infections: Symptoms, diagnosis, transmission, possible vaccines, animal models, and immunotherapy. *Chemical Biology & Drug Design,* 96**,** 1187-1208.

NIU, Z., YU, K. & WU, X. 2020. LSTM-based VAE-GAN for time-series anomaly detection. *Sensors 2020 (Basel, Switzerland),* 20:3738**,** 1-12.

NOLLE, T., SEELIGER, A. & MÜHLHÄUSER, M. 2018. BINet: Multivariate Business Process Anomaly Detection Using Deep Learning. *International Conference on Business Process Management.* Cham: Springer International Publishing.

NUCCI, A., CUI, S., GARRETT, J., SINGH, G. & CROLEY, K. 2018. Real-Time Multivariate-time-scale anomaly detection system for next generation networks. *Technical Disclosure Commons***,** 1-22.

OZBAYOGLU, A. M., GUDELEK, M. U. & SEZER, O. B. 2020. Deep learning for financial applications: A survey. *Applied Soft Computing,* 93.

PACE, R. K. & CALABRESE, R. 2021. Ignoring Spatial and Spatiotemporal Dependence in the Disturbances Can Make Black Swans Appear Grey. *Journal of Real Estate Finance & Economics*.

PANG, G., SHEN, C., CAO, L. & VAN DEN HENGEL, A. 2021. Deep Learning for Anomaly Detection: A Review. *ACM Computing Surveys,* 54**,** Article 38.

PATELL, J. 1976. Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Test. *Journal of Accounting Research,* 14**,** 246-276.

PATTERSON, J. & GIBSON, A. 2017. *Deep Learning A Practitioner's Approach,* 1005 Gravenstein Highway North, Sebastopol, CA 95472, O'Reilly Media, Inc.

PETERSON, P. P. 1989. Event Studies: A Review of Issues and Methodology. *Quarterly Journal of Business and Economics,* 28**,** 36-66.

PHILLIPS, P. C. B. & PERRON, P. 1988. Testing for a unit root in time series regression. *Biometrika,* 75**,** 335-346.

POLEMIS, M. & SOURSOU, S. 2020. Assessing the Impact of the COVID-19 Pandemic on the Greek Energy Firms: An Event Study Analysis. *Energy Research Letters*.

PONCZEK, S. 2019. *Argentina's 48% Stock Rout Second-Biggest in Past 70 Years* [Online]. Bloomberg Markets: Bloomberg. Available: https://www.bloomberg.com/news/articles/2019-08-12/argentina-s-46-stock-rout-second-biggest-in-past-70-years [Accessed 2023].

PUTNIŅŠ, T. J. 2012. Market Manipulation: A Survey. *Journal of Economic Surveys,* 26**,** 952-967.

PWC FINANCIAL SERVICES INSTITUTE 2018. Smart money: AI transitions from fad to future of institutional investing. PWC.

ROUSSEAU, P., CAMARA, D. & KOTZINOS, D. 2021. Weak signal detection and identification in large data sets: a review of methods and applications. *Future* [Online]. Available: https://www.researchgate.net/publication/352006977_Weak_signal_detection_and_identification_in_large_data_sets_a_review_of_methods_and_applications [Accessed May 2021].

RUMELHART, D. E., HINTON, G. E. & WILLIAMS, R. J. 1986. Learning representations by back-propagating errors. *Nature,* 323**,** 533-536.

RYNGAERT, M. & NETTER, J. 1990. Shareholder Wealth Effects of the 1986 Ohio Antitakeover Law Revisited: Its Real Effects. *J.L. Econ & Org.,* 6**,** 253.

SCHMITT, T. A., CHETALOVA, D., SCHÄFER, R. & GUHR, T. 2013. Non-stationarity in financial time series: Generic features and tail behavior. *Europhysics Letters*.

SCHÖLKOPF, B. & SMOLA, A. J. 2002. *Learning with Kernels*.
SEILER, M. J. 2000. The Efficacy of Event-Study Methodologies: Measuring EREIT Abnormal Performance Under Conditions of Induced Variance. *Journal of Financial and Strategic Decisions,* 13**,** 101-112.
SHARPE, W. F. 1964. Capital Assset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance,* 19**,** 425-442.
SHORE, D. A. 2020. Today's Leadership Lesson: Mind the Wildlife and Prepare for Tomorrow's Disruption. *Journal of Health Communication,* 00**,** 1-2.
SINGH, B., DHALL, R., NARANG, S. & RAWAT, S. 2020. The Outbreak of COVID-19 and Stock Market Responses: An Event Study and Panel Data Analysis for G-20 Countries. *Global Business Review***,** 1-26.
SMITH, T. C. & FRANK, E. 2016. Introducing machine learning concepts with WEKA. *Statistical Genomics: Methods and Protocols.* Springer.
SONG, X., WU, M., JERMAINE, C. & RANKA, S. 2007. Conditional Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering,* 19**,** 631-645.
SORNETTE, D. 2009. Dragon-Kings, Black Swans and the Prediction of Crises. *Swiss Finance Institute,* 09-36.
SWANGO, D. L. 2020. Black Swans: When the Impossible Occurs. *The Appraisal journal,* 88**,** 140-148.
TALEB, N. N. 2004. The Black Swan; Why Don't We Learn that We Don't Learn. 1-45.
TALEB, N. N. 2007. *The Black Swan: the impact of the highly improbable,* New York, Random House.
TALEB, N. N. 2020. The Coronavirus Pandemic Was Preventable Says Nassim Taleb. *In:* SCHATZKER, E. (ed.) *Bloomberg Markets: European Close.* Bloomberg Television.
TRAN, N. & REED, D. A. 2004. Automatic ARIMA time series modeling for adaptive I/O prefetching. *IEEE Transactions on Parallel and Distributed Systems,* 15**,** 362-377.
TUOR, A., BAERWOLF, R., KNOWLES, N., HUTCHINSON, B., NICHOLS, N. & JASPER, R. 2017a. Recurrent Neural Network Language Models for Open Vocabulary Event-Level Cyber Anomaly Detection.
TUOR, A., KAPLAN, S., HUTCHINSON, B., NICHOLS, N. & ROBINSON, S. 2017b. Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams.
WESTBROOK, T. 2020. *GLOBAL MARKETS - Stocks crash as pandemic panic sweeps markets* [Online]. Thomson Reuters. Available: https://www.reuters.com/article/global-markets-idCNL4N2B614Y [Accessed 2023 2023/03/21].
WHO 2020. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization.
WORLD BANK DATA. 2022. *GDP (current US$) - China, United States, World, European Union* [Online]. World Bank. Available: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2020&locations=CN-US-1W-EU&start=1960&view=chart [Accessed 2022/02/28 2022].
WUCKER, M. 2016. *The Gray Rhino How to recognize and act on the obvious dangers we ignore*, St Martin's Press.
XIAO, Z. & JIAO, J. 2021. Explainable Fraud Detection for Few Labeled Time Series Data. *Security and communication networks,* 2021**,** 1-9.
YAN, L. & QIAN, Y. 2020. The Impact of COVID-19 on the Chinese Stock Market - An Event Study Based on the Consumer Industry. *Asian Economic Letters,* 1.
YEH, C.-C. M., ZHU, Y., ULANOVA, L., BEGUM, N., DING, Y., DAU, H. A., SILVA, D. F., MUEEN, A. & KEOGH, E. 2016. Matrix Profile I: All Pairs Similarity Joins for Time Series:

A Unifying View that Includes Motifs, Discords and Shapelets. *IEEE 16th International Conference on Data Mining.*

YERMAL, L. & BALASUBRAMANIAN, P. Application of Auto ARIMA Model for Forecasting Returns on Minute Wise Amalgamated Data in NSE. 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017/12/16 2017. 1-5.

ZHANG, C., SONG, D., CHEN, Y., FENG, X., LUMEZANU, C., CHENG, W., NI, J., ZONG, B., CHEN, H. & CHAWLA, N. V. 2018. A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data. *Association for the Advancement of Artificial Intelligence www.aaai.org*, 1-9.

ZHANG, D., HU, M. & JI, Q. 2020. Financial markets under the global pandemic of COVID-19. *Finance Research Letters,* 36.

ZHENG, D., DAI, X., LAN, T., ZHANG, W. & MOU, J. 2021. The Negative Effect of Share Pledging by Controlling Shareholders under COVID-19. *Emerging markets finance & trade,* 57, 2826-2837.

ZHOU, C. & PAFFENROTH, R. C. 2017. Anomaly Detection with Robust Deep Autoencoders. *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Halifax, NS, Canada: Association for Computing Machinery.

Appendices

Appendix 1: Summary of key related-event study specifications

Table 6: Summary of key related-event study specifications

Authors	Markets	Return Model	Return Types	Estimation period	Event Day	Event Windows	Conclusions
Liu et al. (2020)	21 Leading stock indices in major affected countries	Market Model	AR, CAR, AAR, CAAR	[-121, -1] [-151, -1] [-181, -1]	2020/01/20	[0, 6] [7, 13] [14, 20] [21, 27] [28, 34]	Negative ARs in all majorly affected countries, more sig. in Asia
Harjoto et al. (2020)	MSCI indices: MXWO MXWOU MXEF MXUSLC MXUSSC	Market-adjusted return and market model	AR, CAR	[-256, -6] [-261, -11]	2020/03/11 2020/04/20	[-5, 5] [-10, 10]	Greater negative impact on emerging than developed markets, on developed ex-US than US-only, and on Small Cap than large Cap.
He et al. (2020)	2 895 listed companies from the Shanghai and Shenzhen A-share market	Market Model	AR, CAR	[-191, -31]	2020/01/23	[-30, 0] [-25, 0] [-20, 0] [-15, 0] [-10, 0] [-5, 0] [0, 0] [0, 5] [0, 10] [0, 15] [0, 20] [0, 25] [0, 30]	Severe impact on several industries, low response in other industries and positive strong response in still other industries.
Polemis and Soursou (2020)	Greek energy companies	Market Model	AR, CAAR	[-151, -11]	2020/03/23	[-10, 10] [-20, 20] [-50, 50]	Sig. heterogenous negative impact on energy stocks prior to event. Insignificant CAARs.
Singh et al. (2020)	G-20 economies	Constant Mean Return Model	CAR	[-151, -1]	2020/01/20	[0, 9] [10, 20] [21, 30] [31, 40] [41, 50] [51, 57]	Sig. negative impact on G-20 countries followed by major recovery
Yan and Qian (2020)	SSE Consumer 80 Index	Constant Mean Return Model	AR, CAR	[-171, -31]	2020/01/22	[-2, 2] [-5, 5] [-10, 10] [-15, 15] [-30, 30]	Sig. negative impact on consumer stocks at beginning, but sharp recovery
Al-Qudah and Houcine (2021)	FTSE/JSE DJIA TEDPIX MOEX BSE Sensex SSEC	Market Model	AR, CAR, AAR, CAAR	[-91, -1]	2020/01/21	[0, 9] [10, 19] [20, 29] [30, 39] [40, 49]	Negative ARs, with relationship between increase in no. cases and ARs. Western Pacific reacted faster to event, and more neg.

The square brackets indicate the ranges of the windows with day 0 indicating the event day. Negative days are days prior to event day and positive days are days after the event day. Thus, for a 120-day estimation window with [-121, -1] indicates the window begins 121 days prior to event day and ends 1 day prior to event day. For a 11-day window with [-5, 5] indicates the window begins 5 days prior to event day and ends 5 days after the event day (not counting the event day).

Appendix 2: Summary of Events Under Study

Table 7: Summary of Events

No.	Events (Signals)	Event Day	Return Model	Return Types	Estimation period	Event Window
1	First case of COVID-19 outside the PRC. (First reported death of COVID-19).	2020/01/13 (2020/01/11)	Market Model (with GLS estimation)	AAR, CAAR	[-256, -6]	[-5, 5]
2	Declaration of PHEIC. (First Official recognition by SCIO (China))	2020/01/30 (2020/01/23)	Market Model (with GLS estimation)	AAR, CAAR	[-256, -6]	[-5, 5]
3	WHO-China JM recommend large-scale stringent measures.	2020/02/24	Market Model (with GLS estimation)	AAR, CAAR	[-256, -6]	[-5, 5]
4	Italy, Saudi Arabia implement lockdown. WHO declares pandemic.	2020/03/09 2020/03/11	Market Model (with GLS estimation)	AAR, CAAR	[-256, -6]	[-5, 5]

Appendix 3: Software Utilised in Analyses

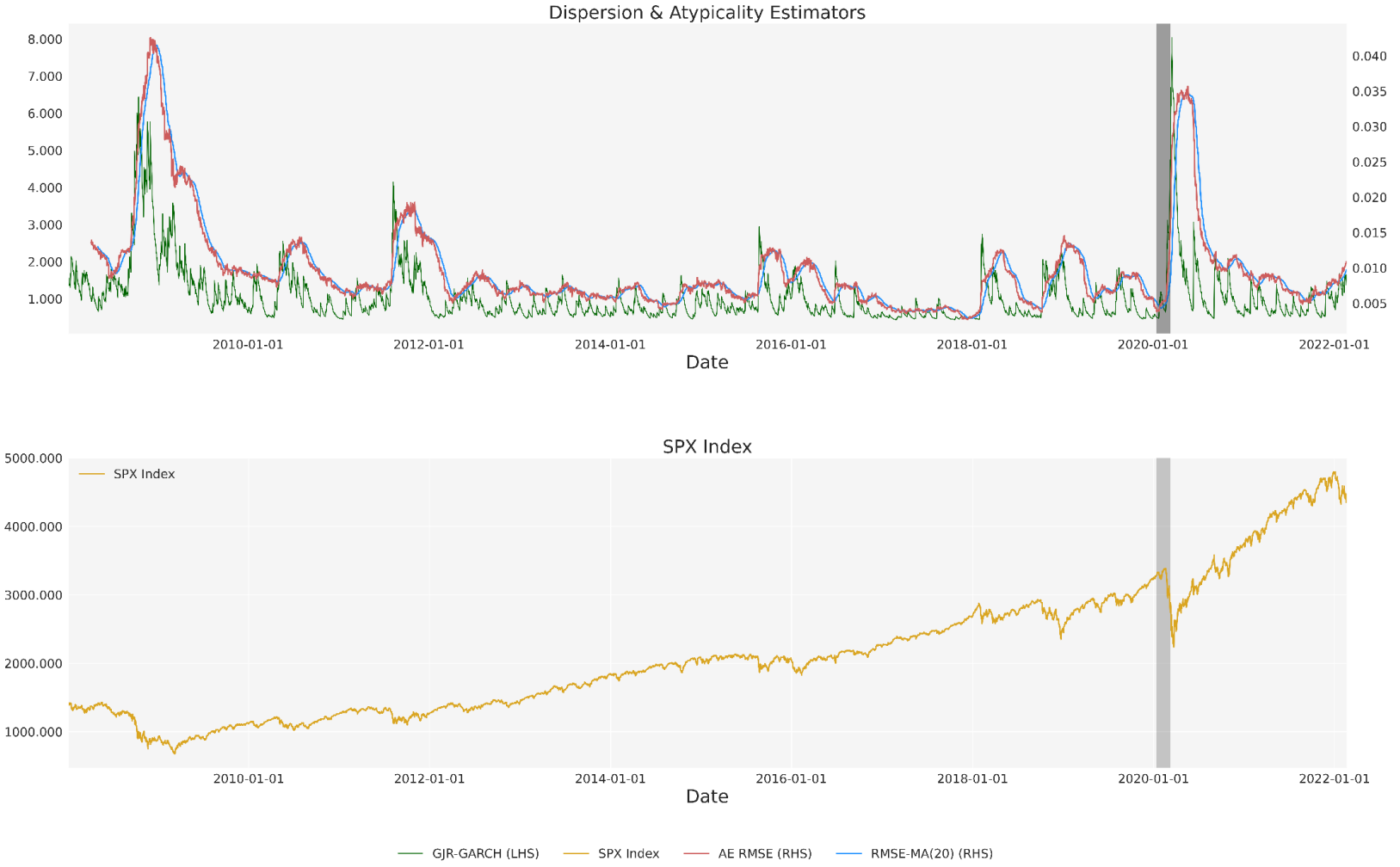
Table 8: Software

Programming Language	Software	Primary Use in Study
Python	pandas	Data manipulation
	numpy	Data manipulation
	matplotlib	Graphs and figures
	seaborn	Graphs and figures
	pmadrima	Arima model
	arch	Garch Model
	tensorflow	Autoencoder
	statsmodels	Autoencoder
R	magrittr	Code pipelining
	ggplot2	Graphs and figures
	bizdays	Date sequences with exclusions
	estudy2	Event study
	lmtest	Heteroscedasticity and Autocorrelation tests
	nlme	GLS linear regression models
	tidyverse	Data manipulation

Appendix 4: Co-movement of stage 2 models

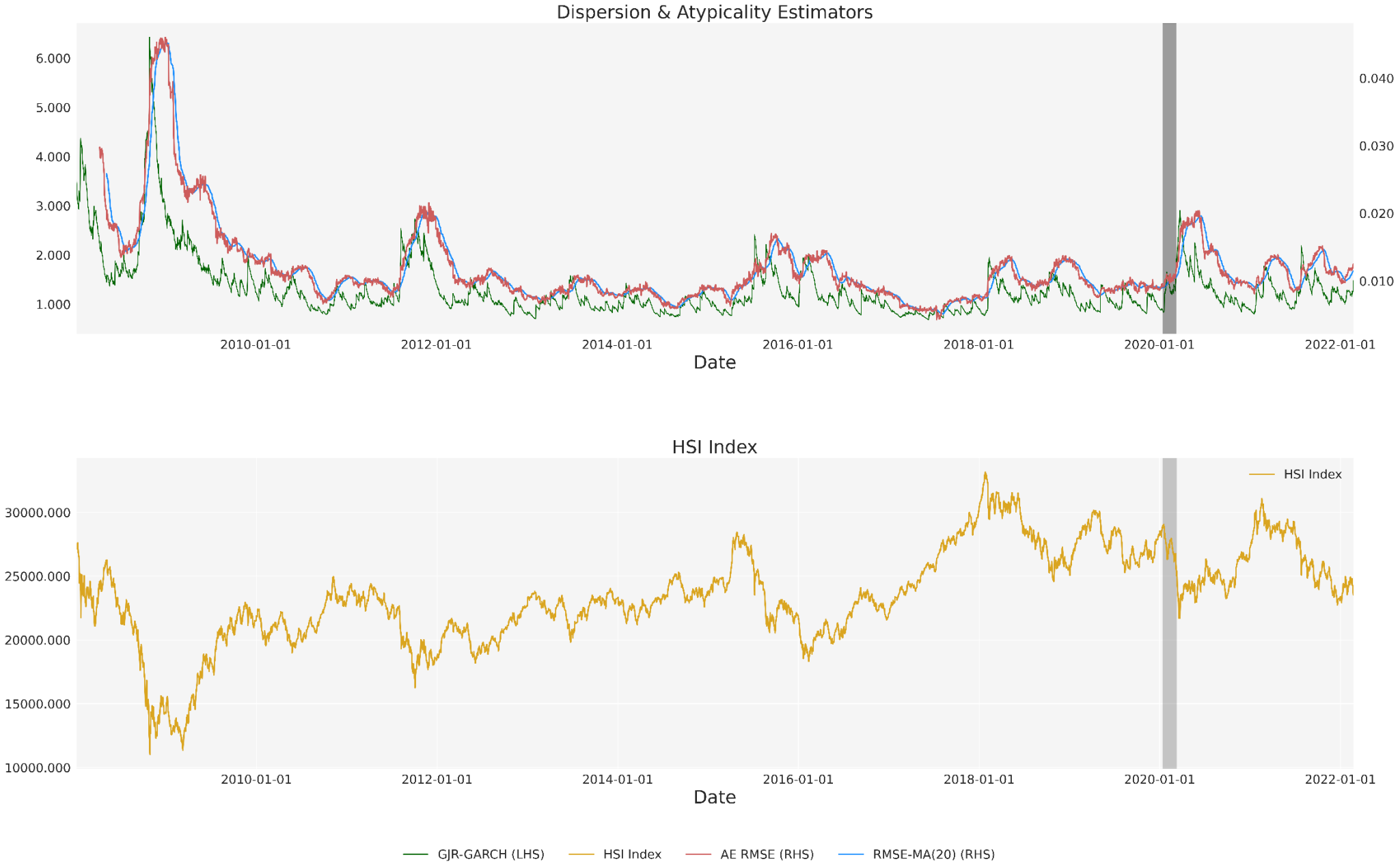
Only the two main categories are presented here[37].

Figure 37: ACP Stage 2 on SPX Index



[37] Plots of the remaining indices are available our GitHub https://github.com/keegangclarke/acp/tree/c20b89e8dc70caed2324f63c347710111c0cbfff/ACP%20Results/stage2/comparison_full_sample_merged

Figure 38: ACP Stage 2 on HSI Index



Appendix 5: Market categories

Table 9: Market categories

Index	Category
AS51	SPX
CAC	SPX
DAX	SPX
HSI	HSI
IBEX	SPX-HSI Hybrid
IMOEX	HSI
INDU	SPX
JALSH	SPX-HSI Hybrid
JCI	SPX-HSI Hybrid
KOSPI	SPX
MERVAL	SPX-HSI Hybrid
MEXBOL	HSI
N100	SPX-HSI Hybrid
NIFTY	SPX
NKY	SPX-HSI Hybrid
SASEIDX	SPX-HSI Hybrid
SHSZ300	SHSZ300
SPTSX	SPX Hybrid
SPX	SPX
TFTSEMIB	SPX-HSI Hybrid
UKX	SPX
XU100	SPX-HSI Hybrid

Appendix 6: Outperformance

Note, the dates under each index identify the start and end of the sample for that equity index. g represents the compound annual growth rate (CAGR), σ represents the annualised standard-deviation, Min and Max represent the smallest and largest returns (representative) recorded in the sample. g/σ represents the risk adjusted return, where greater values represent greater magnitude of returns per unit of risk exposed to.

Table 10: Total Sample Outperformance

Index	Metric	Index Return	LT Interest Rates	μ Rule	1θ Rule	2θ Rule	3θ Rule	O/P
Total Sample	g/σ			***50%***	***41%***	***50%***	***55%***	***77%***
Best performer				***41%***	***9%***	***14%***	***18%***	
	g	3.56%	3.03%	2.06%	2.24%	2.71%	3.20%	
Average	σ	22.88%	0.08%	11.84%	16.33%	17.60%	19.30%	
	g/σ	0.146	N/A	0.162	0.131	0.153	0.157	
AS51	g	0.62%	2.55%	0.69%	1.41%	0.91%	0.54%	***Yes***
2008/01/04	σ	17.41%	0.07%	10.42%	12.48%	14.07%	10.42%	
2022/02/22	g/σ	0.035	N/A	0.066	0.113	0.065	0.052	***Yes***
CAC	g	1.80%	0.86%	1.39%	0.69%	1.23%	1.64%	
2008/01/04	σ	22.60%	0.08%	10.46%	15.89%	17.95%	19.56%	
2022/02/22	g/σ	0.080	N/A	0.133	0.044	0.069	0.084	***Yes***
DAX	g	3.25%	1.73%	-0.33%	-0.39%	1.49%	0.93%	
2008/01/04	σ	22.27%	0.04%	10.84%	15.63%	17.64%	18.73%	
2022/02/22	g/σ	0.146	N/A	-0.030	-0.025	0.085	0.050	
HSI	g	-0.76%	1.26%	-0.70%	-1.53%	-1.04%	-1.31%	***Yes***
2008/01/04	σ	23.69%	0.04%	19.22%	15.85%	17.88%	18.91%	
2022/02/22	g/σ	-0.032	N/A	-0.037	-0.097	-0.058	-0.069	
IBEX	g	-1.58%	1.95%	-3.01%	-3.19%	-1.61%	-1.84%	
2008/01/04	σ	24.20%	0.11%	12.24%	18.11%	20.94%	22.34%	
2022/02/22	g/σ	-0.065	N/A	-0.246	-0.176	-0.077	-0.082	
IMOEX	g	2.03%	5.85%	4.88%	0.93%	4.88%	0.93%	***Yes***
2008/01/09	σ	29.64%	0.11%	12.14%	27.48%	12.14%	27.48%	
2022/02/22	g/σ	0.069	N/A	0.402	0.034	0.402	0.034	***Yes***
INDU	g	4.80%	1.73%	1.43%	3.50%	4.38%	4.56%	
2008/01/04	σ	19.71%	0.04%	8.72%	11.40%	13.28%	14.47%	
2022/02/22	g/σ	0.243	N/A	0.164	0.307	0.330	0.315	***Yes***
JALSH	g	4.70%	5.84%	4.77%	3.81%	3.48%	4.78%	***Yes***
2008/01/04	σ	19.42%	0.05%	8.93%	13.08%	14.85%	15.60%	
2022/02/22	g/σ	0.242	N/A	0.534	0.291	0.234	0.307	***Yes***
JCI	g	7.31%	5.77%	5.47%	6.47%	3.36%	5.66%	
2000/01/04	σ	22.86%	0.14%	12.72%	14.48%	18.24%	20.44%	
2022/02/22	g/σ	0.320	N/A	0.430	0.447	0.184	0.277	***Yes***
KOSPI	g	1.85%	2.15%	-0.82%	-0.58%	0.25%	-0.26%	
2008/01/04	σ	19.53%	0.08%	8.95%	12.59%	14.01%	15.21%	
2022/02/23	g/σ	0.095	N/A	-0.092	-0.046	0.018	-0.017	

MERVAL	g	19.89%	7.69%	14.15%	17.13%	14.27%	18.82%	
2008/01/04	σ	37.30%	0.14%	20.90%	32.37%	34.54%	36.29%	
2022/02/22	g/σ	0.533	N/A	0.677	0.529	0.413	0.519	***Yes***
MEXBOL	g	3.01%	4.62%	0.82%	1.59%	2.35%	2.54%	
2008/01/04	σ	18.78%	0.07%	9.70%	12.23%	13.95%	14.83%	
2022/02/21	g/σ	0.160	N/A	0.085	0.130	0.168	0.171	***Yes***
N100	g	1.31%	1.74%	1.11%	0.74%	1.61%	1.47%	***Yes***
2008/01/04	σ	20.87%	0.05%	10.05%	14.69%	16.72%	17.99%	
2022/02/22	g/σ	0.063	N/A	0.110	0.050	0.096	0.082	***Yes***
NIFTY	g	7.52%	5.33%	4.32%	5.80%	6.51%	7.67%	***Yes***
2000/01/04	σ	22.73%	0.08%	8.43%	12.47%	14.79%	16.36%	
2022/02/22	g/σ	0.331	N/A	0.513	0.465	0.440	0.469	***Yes***
NKY	g	2.89%	0.37%	-0.27%	0.25%	0.77%	2.79%	
2008/01/04	σ	24.10%	0.03%	12.87%	18.17%	20.12%	21.64%	
2022/02/22	g/σ	0.120	N/A	-0.021	0.014	0.038	0.129	***Yes***
SASEIDX	g	0.43%	1.28%	1.41%	-0.67%	-0.28%	-0.29%	***Yes***
2008/01/07	σ	24.20%	0.04%	14.03%	18.19%	19.85%	21.44%	
2022/02/21	g/σ	0.018	N/A	0.100	-0.037	-0.014	-0.013	***Yes***
SHSZ300	g	-0.82%	1.26%	-1.14%	-0.58%	1.10%	2.16%	***Yes***
2008/01/04	σ	26.14%	0.04%	13.86%	19.66%	21.83%	23.36%	
2022/02/23	g/σ	-0.031	N/A	-0.082	-0.030	0.050	0.093	***Yes***
SPTX	g	2.07%	1.49%	1.27%	1.16%	1.71%	1.96%	
2008/01/04	σ	18.49%	0.05%	7.03%	9.91%	11.37%	12.22%	
2022/02/18	g/σ	0.112	N/A	0.181	0.117	0.150	0.161	***Yes***
SPX	g	5.60%	1.73%	1.91%	3.49%	4.41%	5.25%	
2008/01/04	σ	20.56%	0.04%	8.94%	12.22%	14.14%	15.40%	
2022/02/18	g/σ	0.273	N/A	0.214	0.286	0.312	0.341	***Yes***
TFTSEMIB	g	4.83%	1.95%	2.74%	1.27%	2.85%	3.71%	
2011/08/01	σ	24.51%	0.09%	15.93%	20.30%	22.60%	23.49%	
2022/02/22	g/σ	0.197	N/A	0.172	0.063	0.126	0.158	
UKX	g	0.78%	1.52%	-1.14%	0.03%	0.63%	0.56%	
2008/01/04	σ	19.06%	0.07%	9.17%	12.33%	14.30%	15.38%	
2022/02/22	g/σ	0.041	N/A	-0.124	0.003	0.044	0.036	***Yes***
XU100	g	6.71%	7.90%	6.32%	7.85%	6.28%	8.07%	***Yes***
2008/01/04	σ	25.32%	0.18%	14.92%	19.74%	21.90%	23.11%	
2022/02/22	g/σ	0.265	N/A	0.424	0.398	0.287	0.349	***Yes***

Note, the dates under each index identify the start and end of the sample for that equity index.
g_d represents the CAGR of all the negative returns (downside), σ_d represents the annualised downside-deviation, Min and Max represent the smallest and largest returns (representative) recorded in the sample.

Table 11: Downside Performance

Index	Metric	Index Return	μ Rule	1θ Rule	2θ Rule	3θ Rule	O/P
Overall			**95%**	**84%**	**83%**	**76%**	**73%**
Total Sample	g_d		95.45%	90.91%	90.91%	90.91%	**95%**
	σ_d		95.45%	90.91%	90.91%	90.91%	**95%**
	Min		95.45%	77.27%	63.64%	50.00%	**95%**
	Max		95.45%	77.27%	81.82%	72.73%	**5%**
Best Performer	g_d		95.45%	4.55%	4.55%	0.00%	
	σ_d		0.00%	0.00%	0.00%	9.09%	
	Min		100.00%	27.27%	18.18%	13.64%	
	Max		0.00%	22.73%	18.18%	27.27%	
Average Sample Performance	g_d	-51.41%	-28.10%	-39.87%	-44.13%	-47.28%	**Yes**
	σ_d	13.89%	8.03%	10.72%	11.40%	12.48%	**Yes**
	Min	-13.24%	-7.02%	-10.31%	-11.07%	-12.11%	**Yes**
	Max	11.73%	5.91%	8.22%	7.30%	8.85%	
AS51	g_d	-48.78%	-31.56%	-37.60%	-42.49%	-31.64%	**Yes**
2008/01/04	σ_d	11.33%	7.04%	8.22%	9.10%	7.04%	**Yes**
2022/02/22	Min	-9.70%	-7.05%	-7.33%	-7.36%	-7.05%	**Yes**
	Max	7.00%	5.02%	5.02%	5.02%	5.02%	
CAC	g_d	-56.30%	-26.26%	-42.87%	-47.81%	-51.30%	**Yes**
2008/01/04	σ_d	14.29%	7.13%	10.40%	11.69%	12.70%	**Yes**
2022/02/22	Min	-12.28%	-6.83%	-8.39%	-12.28%	-12.28%	**Yes**
	Max	0.00%	6.01%	9.66%	9.66%	9.66%	**Yes**
DAX	g_d	-28.32%	-43.27%	-47.97%	-50.22%	-55.90%	
2008/01/04	σ_d	7.56%	10.41%	11.69%	12.32%	13.98%	
2022/02/22	Min	-12.24%	-7.16%	-7.94%	-12.24%	-12.24%	**Yes**
	Max	11.40%	5.93%	5.93%	5.93%	11.40%	
HSI	g_d	-58.01%	-50.73%	-40.63%	-47.60%	-50.31%	**Yes**
2008/01/04	σ_d	14.72%	12.16%	10.62%	11.52%	12.10%	**Yes**
2022/02/22	Min	-12.70%	-8.65%	-8.65%	-8.65%	-8.65%	**Yes**
	Max	14.35%	10.72%	10.72%	10.72%	10.72%	
IBEX	g_d	-59.24%	-31.57%	-46.17%	-53.23%	-56.00%	**Yes**
2008/01/04	σ_d	15.22%	8.52%	12.14%	13.44%	14.22%	**Yes**
2022/02/22	Min	-14.06%	-7.54%	-14.06%	-14.06%	-14.06%	**Yes**
	Max	14.43%	6.95%	14.43%	14.43%	14.43%	
IMOEX	g_d	-61.79%	-28.42%	-48.59%	-28.42%	-48.59%	**Yes**
2008/01/09	σ_d	18.87%	8.35%	18.23%	8.35%	18.23%	**Yes**

2022/02/22	Min	-18.66%	-10.79%	-18.66%	-10.79%	-18.66%	**Yes**
	Max	28.69%	5.80%	28.69%	5.80%	28.69%	
INDU	g_d	-47.67%	-21.35%	-30.59%	-36.37%	-39.37%	**Yes**
2008/01/04	σ_d	12.79%	6.13%	7.69%	8.65%	9.53%	**Yes**
2022/02/22	Min	-12.93%	-4.60%	-5.55%	-5.55%	-7.79%	**Yes**
	Max	11.37%	3.55%	3.95%	5.09%	5.09%	
JALSH	g_d	-51.19%	-20.82%	-35.78%	-41.99%	-43.84%	**Yes**
2008/01/04	σ_d	12.09%	5.87%	8.37%	9.38%	9.86%	**Yes**
2022/02/22	Min	-9.72%	-4.61%	-4.61%	-6.23%	-9.72%	**Yes**
	Max	7.53%	5.28%	5.28%	5.28%	5.28%	
JCI	g_d	-51.56%	-29.77%	-34.57%	-43.46%	-47.44%	**Yes**
2000/01/04	σ_d	14.57%	8.55%	9.43%	12.37%	13.34%	**Yes**
2022/02/22	Min	-10.38%	-7.13%	-7.03%	-10.36%	-10.36%	**Yes**
	Max	10.19%	4.97%	7.27%	7.27%	7.27%	
KOSPI	g_d	-49.63%	-22.56%	-36.05%	-40.21%	-42.85%	**Yes**
2008/01/04	σ_d	12.62%	6.40%	8.34%	9.19%	10.12%	**Yes**
2022/02/23	Min	-10.57%	-4.44%	-4.69%	-4.69%	-8.39%	**Yes**
	Max	11.95%	4.02%	4.02%	4.02%	5.15%	
MERVAL	g_d	-71.96%	-43.20%	-65.11%	-68.65%	-70.65%	**Yes**
2008/01/04	σ_d	24.06%	13.28%	21.16%	22.89%	23.51%	**Yes**
2022/02/22	Min	-37.93%	-10.73%	-37.93%	-37.93%	-37.93%	**Yes**
	Max	11.00%	10.24%	11.00%	11.00%	11.00%	
MEXBOL	g_d	-48.30%	-23.54%	-32.81%	-38.50%	-40.98%	**Yes**
2008/01/04	σ_d	11.58%	6.71%	8.02%	8.91%	9.40%	**Yes**
2022/02/21	Min	-7.01%	-5.88%	-6.42%	-6.42%	-6.42%	**Yes**
	Max	11.01%	6.36%	6.36%	6.36%	6.36%	
N100	g_d	-53.79%	-24.87%	-40.14%	-45.44%	-48.68%	**Yes**
2008/01/04	σ_d	13.35%	6.91%	9.74%	10.94%	11.76%	**Yes**
2022/02/22	Min	-11.97%	-6.28%	-8.09%	-11.97%	-11.97%	**Yes**
	Max	10.87%	5.98%	8.48%	8.62%	8.62%	
NIFTY	g_d	-55.07%	-17.07%	-30.49%	-37.99%	-42.38%	**Yes**
2000/01/04	σ_d	14.37%	5.90%	8.36%	9.77%	10.57%	**Yes**
2022/02/22	Min	-12.98%	-6.95%	-8.30%	-8.30%	-8.30%	**Yes**
	Max	17.74%	3.69%	5.32%	5.32%	5.32%	
NKY	g_d	-57.92%	-29.40%	-46.63%	-51.15%	-54.49%	**Yes**
2008/01/04	σ_d	15.29%	8.99%	11.98%	13.06%	13.80%	**Yes**
2022/02/22	Min	-11.41%	-6.18%	-10.55%	-10.55%	-10.55%	**Yes**
	Max	14.15%	6.72%	7.71%	8.04%	8.04%	
SASEIDX	g_d	-43.58%	-23.47%	-32.37%	-35.63%	-39.51%	**Yes**
2008/01/07	σ_d	17.03%	10.42%	13.47%	14.38%	15.36%	**Yes**
2022/02/21	Min	-15.43%	-10.30%	-15.43%	-15.43%	-15.43%	**Yes**
	Max	11.79%	5.95%	11.79%	11.79%	11.79%	
SHSZ300	g_d	-61.20%	-29.90%	-48.17%	-52.98%	-55.63%	**Yes**
2008/01/04	σ_d	16.89%	9.65%	12.99%	14.30%	15.21%	**Yes**
2022/02/23	Min	-8.75%	-7.88%	-7.88%	-7.88%	-8.56%	**Yes**
	Max	9.34%	8.27%	8.27%	8.27%	9.24%	
SPTX	g_d	-45.47%	-17.11%	-27.37%	-32.67%	-35.42%	**Yes**

2008/01/04	σ_d	12.54%	5.04%	7.03%	7.78%	8.24%	**Yes**
2022/02/18	Min	-12.34%	-4.75%	-10.27%	-10.27%	-10.27%	**Yes**
	Max	11.96%	4.19%	4.19%	4.19%	4.19%	
SPX	g_d	-21.00%	-17.13%	-32.50%	-38.24%	-40.46%	**Yes**
2008/01/04	σ_d	6.32%	5.58%	8.26%	9.43%	10.37%	**Yes**
2022/02/18	Min	-11.98%	-4.10%	-6.66%	-7.60%	-9.51%	**Yes**
	Max	11.58%	4.24%	4.60%	4.96%	4.96%	
TFTSEMIB	g_d	-60.23%	-39.08%	-51.60%	-56.23%	-58.48%	**Yes**
2011/08/01	σ_d	15.84%	10.71%	13.50%	14.99%	15.34%	**Yes**
2022/02/22	Min	-16.92%	-6.65%	-12.48%	-16.92%	-16.92%	**Yes**
	Max	8.93%	5.49%	5.86%	6.59%	6.59%	
UKX	g_d	-50.03%	-23.55%	-34.63%	-40.87%	-43.16%	**Yes**
2008/01/04	σ_d	12.16%	6.39%	8.14%	9.20%	9.96%	**Yes**
2022/02/22	Min	-10.87%	-5.48%	-5.48%	-7.69%	-10.87%	**Yes**
	Max	9.84%	4.75%	5.16%	5.16%	8.84%	
XU100	g_d	-49.89%	-23.48%	-34.44%	-40.69%	-42.99%	**Yes**
2008/01/04	σ_d	12.19%	6.40%	8.15%	9.21%	9.98%	**Yes**
2022/02/22	Min	-10.47%	-10.47%	-10.47%	-10.47%	-10.47%	
	Max	12.89%	5.81%	7.14%	7.14%	7.14%	