STA Honours Project

An empirical investigation into the modality of the JSE TOP 40 and the S&P 500 price distortions

Yaseed Amod & Luke Barnes
Supervisor: A/Prof. Tim Gebbie

Department of Statistical Science,
University of Cape Town

Abstract

Previous work investigating the price distortion (defined as the log difference between the real price and the fundamental value of the asset) of the S&P 500 has characterised it as having an underlying bimodal distribution. This paper demonstrates that the supposed bimodality occurs only under certain growth rate conditions. Methods are then extended to the TOP 40 index and 7 other major market indices, from which it is seen that the strong bimodal shape of the distribution of the price distortion of an index is a feature unique to the Shiller S&P 500 data set, due to the number of dividend observations and growth rate. It is then shown that standard linear time series, appear to be a good fit for the price distortion of the indices. Nonetheless, a certain class of agent-based financial market models are shown to better characterise and explain some of the dynamics of the price distortions.

Contents

1 Introduction

The efficient market hypothesis proposes that financial markets are exceedingly efficient and thus all information is immediately reflected in the prices of assets, without delay (Fama, 1970; Malkiel, 2003). An asset's price should then reflect its fundamental value, as defined by Shiller (2015), and the distribution of the asset's price distortion (i.e. the log difference between the asset price and its fundamental value) would be unimodal. Indicating that the asset price follows the asset's fundamental value closely, in the long-run. Schmitt and Westerhoff (2017) found that on the contrary, the distribution of the S&P 500's price distortion does not have a unimodal shape, but appears to exhibit bimodal behaviour. This suggests that index price is not in the vicinity of its fundamental value most of the time, as theorised by the efficient market hypothesis. Lux (2021) and Schmitt and Westerhoff (2017) suggest that the large fluctuations of the S&P 500 prices around its fundamental value, as shown by Shiller (2015), are indicative of an undervaluation and overvaluation of the index for extended periods of time. Schmitt and Westerhoff (2017) conclude that the S&P 500's price distortion is bimodal and that the bimodality of the price distortion cannot be explained by any linear models, but can be explained by models that account for non-linear forces, like agent-based models.

A class of heterogeneous agent-based models explain the extended deviations from the fundamental value of an asset through the non-linear interaction of heterogeneous agents and the market itself (Chiarella et al., 2007; Gaunersdorfer and Hommes, 2007; Franke and Westerhoff, 2012; Schmitt and Westerhoff, 2014). These heterogeneous agent-based models are predicated on the existence of two predominant groups of agents, chartists and fundamentalists, who interact with the trading mechanism of the market. Chartists determine trading strategies based on previous price movements, through trend extrapolation. Fundamentalists, on the other hand, trade on the basis of their calculated true fundamental value of the asset (Leal, 2015). When the asset price is close to its supposed fundamental value, chartists (commonly referred to as trend-followers in agent-based modelling literature) push the price away from its fundamental value which in turn causes the fundamentalists to take a more active position in the market to push the asset price back towards its fundamental value (Majewski et al., 2020). Schmitt and Westerhoff (2017) found that the the distribution of the S&P 500 price distortion is similar to the distributions produced by a number of these heterogeneous agent-based models (Gaunersdorfer and Hommes, 2007; Franke and Westerhoff, 2012).

An area of the research that is not explored by Schmitt and Westerhoff (2017) is whether the non-unimodal property of the S&P 500 price distortion is unique to the index, or shared amongst other major international indices. The purpose of this research paper is three-fold. Firstly, an investigation is made into whether the multimodal property of the S&P 500 price distortion is applicable in a South African context, and more specifically to the FTSE/JSE TOP 40 Index, which tracks the 40 largest companies listed on the JSE. Additionally, it is of interest to establish whether this distributional property of the S&P 500 price distortion occurs in 7 other international indices, as strongly. Secondly, are standard linear time series models able to characterise and fit the price distortions of both the S&P 500 and TOP 40. Thirdly, is the non-unimodal property of the price distortion of these indices indicative of bimodality, as described by Schmitt and Westerhoff (2017), and can heterogeneous agent-based financial market models be used to explain their dynamics and stylised facts.

The paper is structured as follows: section 2 outlines the methodology, data and formulae used throughout the paper; section 3 presents the real price, fundamental value and price distortion results obtained for the S&P 500 data sets; similarly, section 4 and section 5 display the results for the TOP 40 index and the 7 other major market indices, respectively; section 6 outlines the statistical methods used to test the modality of the price distortion distributions of the S&P 500 and TOP 40; section 7 shows the results of fitting standard linear time series models to the price distortions of both the S&P 500 and TOP 40; section 8 describes the agent-based models of Gaunersdorfer and Hommes (2007) and Franke and Westerhoff (2012) and section 9 discusses the results of the above sections.

2 Methodology

2.1 Data

The data collection process for the indices is two-fold. First, the unique data set of the S&P 500 by Robert Shiller from January 2015 to July 2022 (updated on a monthly basis) is obtained [1]. The monthly composite index price of the S&P500, the monthly annualised dividend and the monthly CPI, from January 1871 to December 2021, are extracted into a single excel file. There are 3 periods of consideration: January 1871 to June 2014, January 1871 to December 2015 and January 1871 to December 2021. The first period is to replicate the results in Shiller (2001) and apply the methodology used in Schmitt and Westerhoff (2017). The second period is to replicate similar results to that of Schmitt and Westerhoff (2017). The last time period extends the results up to the present and is used to compare the Shiller data to the Bloomberg data [2].

Secondly, the S&P 500 and JSE TOP 40 data, in addition to data from 7 other market indices (FTSE 100, DAX 40, NIKKEI 225, Hang Seng, BOVESPA, NIFTY 50, ASE) is attained from Bloomberg[3]. The data is imported into individual excel files for each market index and contains the monthly closing price, monthly annualised dividends and the monthly CPI. The data for all of the markets[4] is from their earliest recorded closing price date up to December 2021 [5]. The data imported from Bloomberg consists of completely raw data[6] in 2 Excel files, which were then manually transferred into a single excel file to differentiate between each of the market indices.

2.1.1 Data Preparation

Index	Price date	Ann. div. date	CPI date	# monthly obs.
TOP 40	06/1995	10/2002	06/1980	231
S&P 500	12/1927	12/1970	12/1927	613
FTSE 100	12/1983	02/2002	01/1988	293
DAX 40	10/1959	05/1997	01/1991	296
Nikkei 225	02/1970	05/1993	01/1970	344
Hang Seng	07/1964	09/1993	10/1980	340
BOVESPA	03/1993	05/1998	12/1979	284
NIFTY 50	07/1990	04/2000	01/2011	132
ASE	01/1987	11/1995	01/1959	313

Table 2.1: This table shows the starting dates for the closing price, annualised dividends and CPI for each index to 12/2021. The starting date used for each index is the starting date with non-missing data for each variables, up to 12/2021. E.g. For the TOP 40, the starting date used is 10/2002. The resulting number of monthly observations for each index, after missing data is removed is given in the last column. Dates are in the format of [mm/yyyy].

[1] Available on Robert Shiller's website http://www.econ.yale.edu/~shiller/data.htm.

[2] Robert Shillers dataset and the Bloomberg data used can be found here Amod and Barnes (2022).

[3] From the Bloomberg Terminal.

[4] The ASE Index is an exception - a missing price observation at July 2015 was encountered and this was made equal to the average price between June 2015 and August 2015.

[5] Missing data observations resulted in each variable in the data set beginning from the earliest recorded observation at which their no missing data to December 2021.

[6] The raw data, cleaned data, .RData and the R code is made available on our Github repository Amod and Barnes (2022).

Bloomberg Name	Adjusted Name
Date	Date
Index Closing Price	Price
Dividend Per Share 12 Month(Gross)	Annualised dividends
Consumer Price Index	CPI

Table 2.2: Shows the names of the variables as they are on the Bloomberg Terminal in comparison to the naming convention used throughout this paper. Take note that dividends not annualised is referred to as dividends.

2.1.2 Data Discrepancy for the S&P 500 Data

The S&P 500 data from Robert Schiller's unique data set from January 1871 to December 2021 consists of 1800 monthly observations.[7] The S&P 500 data obtained from Bloomberg from December 1970 to December 2021 consists of 613 monthly observations. The price is the closing price at the end of every month. The annualised dividends is defined as the Dividend[8] Per Share 12 Month (Gross). The CPI is the Consumer Price Index for All Urban Consumers. The data in the two data sets is similar for the periods December 1970 to December 2021, as one would expect.[9]

2.2 Robert Shiller's Fundamental Value

Shiller (2015) suggests that a way to calculate the fundamental value of an index is to present value the real dividends, using a constant real discount rate and a constant real average dividend growth rate to account for future unobserved dividends. The discount rate (r) is calculated by taking the geometric average real return rate of the index across the entire time period of the index price. The growth rate (g) is the geometric average growth rate of the real dividends over a 10 year period or longer. Shiller's formula for the Fundamental Value is similar to a two-stage Dividend Discount model, where one stage is comprised of the discounting of dividends up to a point in time, and the second stage accounts for the discounting of the infinite dividends after that point in time. The fundamental value of an asset, is thus given by:

$$P_t^* = \frac{D_t}{(1+r)^1} + \frac{D_{t+1}}{(1+r)^2} + \frac{D_{t+2}}{(1+r)^3} + ... + \frac{D_{t+k}}{(1+r)^n} + \frac{D_\infty}{(1+r)^n}, \tag{1}$$

Here D_{t+k} is the last observed real dividend and $t+k=n$ where n is the number of observations, and the final value, discounting all unseen future dividends, is given by:

$$D_\infty = \frac{D_{t+k}(1+g)}{r-g}. \tag{2}$$

3 S&P 500

3.1 Robert Shiller: 1871 to 2013

In this section, 1716 monthly observations are considered from January 1871 to December 2013[10] from Robert Shiller's S&P 500 data set. The aim is to replicate the plots seen in Shiller (2015) with regards to the fundamental value, the discount rate (r) and the growth rate (g). The real price is

[7] Robert Shiller gives a good overview of his Stock market data and some additional information on his website http://www.econ.yale.edu/~shiller/data.htm.

[8] Take care as the "Dividend" variable within the excel files in Amod and Barnes (2022) is the "Annualised Dividends".

[9] Variable selection (closing price and Dividend Per Share 12 month(Gross)) is chosen in order to get it as close to Robert Shillers data used in Shiller (2015) and in Schmitt and Westerhoff (2017).

[10] The data set goes to June 2014 since all real values should be in terms on June 2014 dollars similar to that in Shiller (2015).

obtained using (25).[11] It is important to note that in this section the ending period CPI is defined to be 06/2014, exactly as in Shiller (2015).[12] Annualised dividends are divided by 12 to get the monthly dividend. (26) is used to obtained the Total Real Return of the S&P 500 on a monthly basis. The real monthly return is then used to calculate the discount rate (r), which is defined as the geometric average monthly real return[13] on the market from January 1871 to December 2013. The geometric average monthly growth rate of real dividends is calculated in a similar manner, but for 3 periods (the last 10, 20 and 30 years) and minimising the objective function[14].

Shiller (2015) uses a r and g value of 7.6% and 5.1% per year respectively. None of the methods used were able to recover these results from the data set. Recovery of these values was attempted using both the first month of every year from 1871 to 2013 and the last month of every year from 1871 to 2013. However, neither of the r and g values obtained resulted in a fundamental value time series (1) in the vicinity of the real price. The geometric average monthly real return on the market from January 1871 to December 2013 was then defined as the r value. The geometric average monthly growth rate of real dividends, for the last 10, 20 and 30 years and one calculated by minimising the objective function, are denoted by g_{10}, g_{20}, g_{30} and $g_{\min}$ respectively. Table 3.1 shows the values calculated for each of these rates from the Shiller data set.

Time period [mm/yyyy]	r
01/1871-12/2013	0.55%
	g
01/2003-12/2012	0.36%
01/1993-12/2012	0.15%
01/1983-12/2012	0.13%
Min obj. function	0.40%

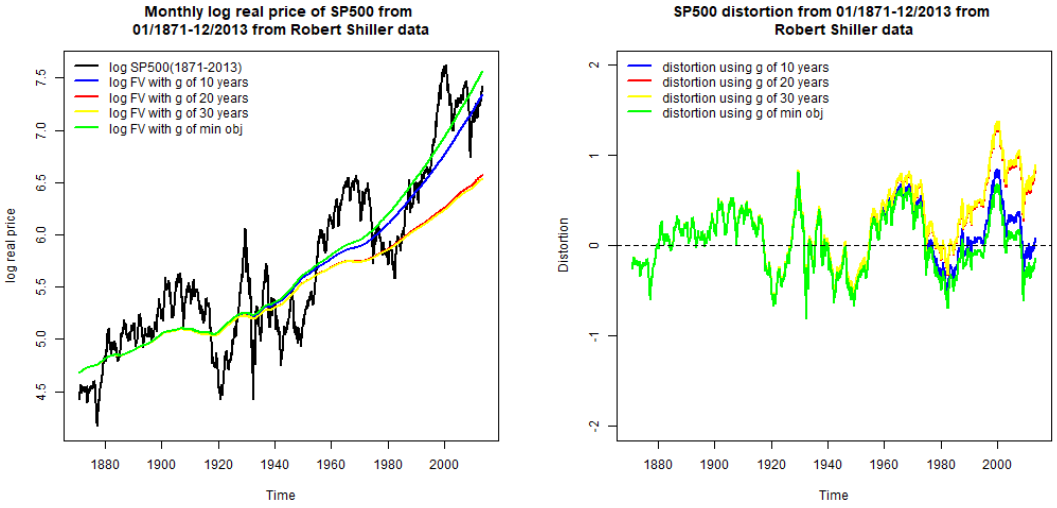
Table 3.1: Table of the monthly discount rate and the four different growth rates calculated from the Shiller data set for the period ending 12/2013. See Figure 3.1 for the log fundamental paths for these growth rates (per month) of the last 10, 20, 30 years of real dividends and minimising the objective function (28) not annualised.

From Figure 3.1a, it is interesting to note that choosing a lower g results in a log fundamental value time series that is more flat for later time periods. This can be seen for g_{20} and g_{30}, whose log fundamental value time series are far below the g_{10} and $g_{\min}$ log fundamental value time series.[15] The log fundamental value time series calculated using the g of the last 20 and 30 years, result in a more positively weighted price distortion, as the log S&P 500 real price is above the fundamental value for such extended lengths of time, more so than when using g_{10} and $g_{\min}$.

See Appendix B for a visual impression of the distribution and smoothed density of the price distortion. These figures only give a visual impression of what the underlying distribution could be. Modality Testing is considered at a later stage. Additionally, the bandwidth b_w[16] used to smooth all distributions is:

$$b_w = n^{-\frac{1}{5}} \times \min\left(\sigma(x) \times \frac{\text{interquartile}(X)}{1.34}\right) \times 4 \times 1.06 \tag{3}$$

[11] Derived from formulas used in the Shillers excel file.
[12] Prices are in terms of June 2014 dollars.
[13] Calculated using (27).
[14] The minimised objective function is defined in (28).
[15] See Table 3.1 for the values of g_{10}, g_{20}, g_{30} and $g_{\min}$.
[16] The **default bandwidth** calculation in the `density()` function in R Core Team (2021).



(a) Log Fundamental Values series

(b) Price Distortion series

Figure 3.1: Figure 3.1a shows the monthly log fundamental value time series calculated using a g_{10}, g_{20}, g_{30} and $g_{\min}$ and the constant discount rate r from Table 3.1, together with the monthly log real price of the S&P 500 from January 1871 to December 2013. Figure 3.1a highlights that smaller growth rates result in a more flat log fundamental value time series. Figure 3.1b shows the monthly price distortion for each of the four fundamental value time series. The price distortion calculated using g_{20} and g_{30} results in more positively weighted price distortions, in comparison to g_{10} and $g_{\min}$.

3.2 Robert Shiller: 1871 to 2015

In this section, 1740 monthly observations are considered from January 1871 to December 2015 from Robert Shiller's data set. The aim is to replicate the results seen in Schmitt and Westerhoff (2017), with regards to the fundamental value the discount rate (r) and growth rate (g). The real price is obtained using (25).[17] In this section, the ending period CPI is December 2015[18]. Annualised dividends are divided by 12 to get the monthly dividend. (26) is used to obtained the Total Real Return of the S&P 500 on a monthly basis. The real monthly return is then used to calculate the discount rate (r), which is defined as the geometric average monthly real return on the market from 01/1871 to 12/2015. The geometric average monthly growth rate of real dividends is calculated in a similar manner, but for 3 periods (the last 10[19], 20 and 30 years) and minimising the objective function[20].

Time period [mm/yyyy]	r
01/1871-12/2015	0.55%
	g
01/2005-12/2014	0.32%
01/1995-12/2014	0.20%
01/1985-12/2014	0.16%
Min obj. function	0.38%

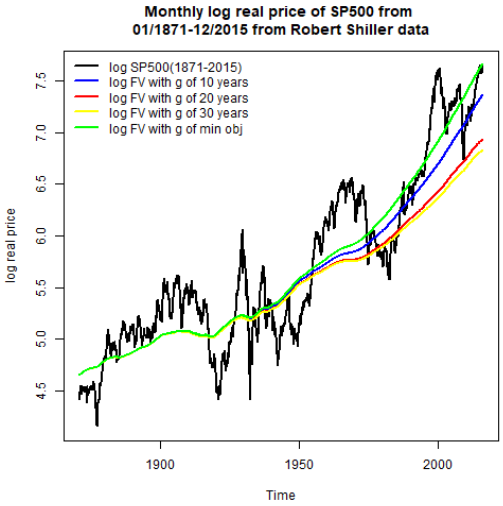
Table 3.2: Table of the monthly discount rate and the four different growth rates calculated from the Shiller data set for the period ending December 2015. See Figure 3.2 for the log fundamental paths for these growth rates (per month) of the last 10, 20, 30 years of real dividends and minimising the objective function not annualised.

[17] Derived from formulas used in the Shiller excel file.

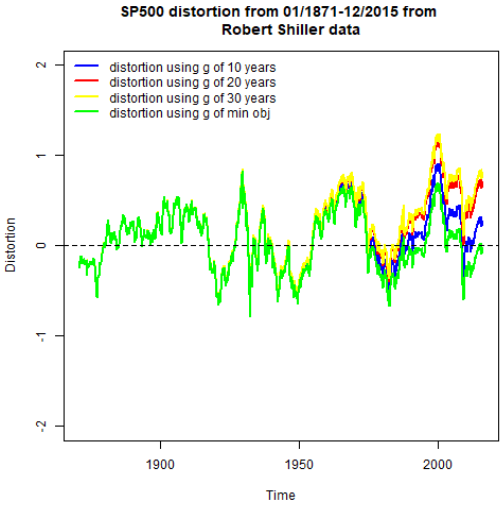
[18] Not mentioned explicitly in Schmitt and Westerhoff (2017)so real prices are assumed in terms of 12/2015 dollarss

[19] Schmitt and Westerhoff (2017) only demonstrates the fundamental value using a an average dividend growth rate of the last 10 years.

[20] The minimised objective function is defined in (28).

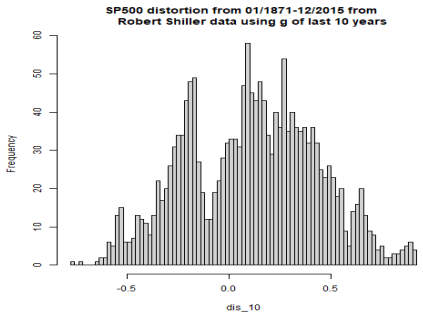


(a) Log Fundamental value series

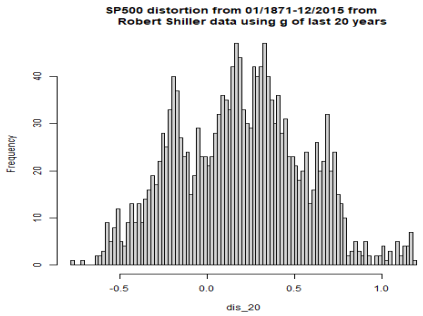


(b) Price Distortion series

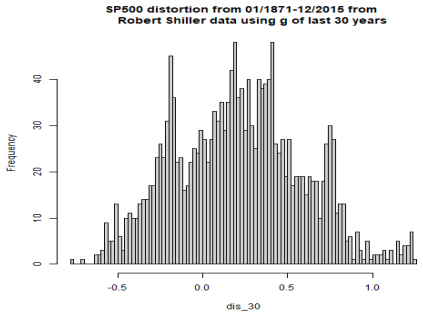
Figure 3.2: Figure 3.2a shows the monthly log fundamental value time series calculated using a g_{10}, g_{20}, g_{30} and $g_{\min}$ and the constant discount rate r from Table 3.2, together with the monthly log real price of the S&P 500 from January 1871 to December 2015. Similarly to Figure 3.1a, it can be noted that smaller growth rates result in a more flat log fundamental value time series. Figure 3.2b shows the monthly price distortion for each of the four fundamental value time series. The price distortion calculated using g_{20} and g_{30} results in more positively weighted price distortions, in comparison to g_{10} and $g_{\min}$.



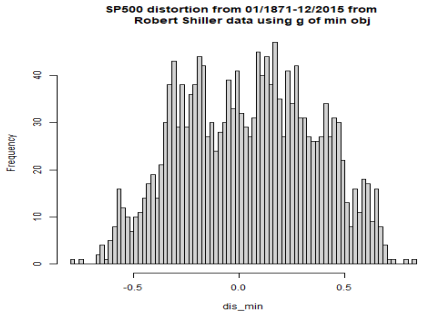
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using g_{20}

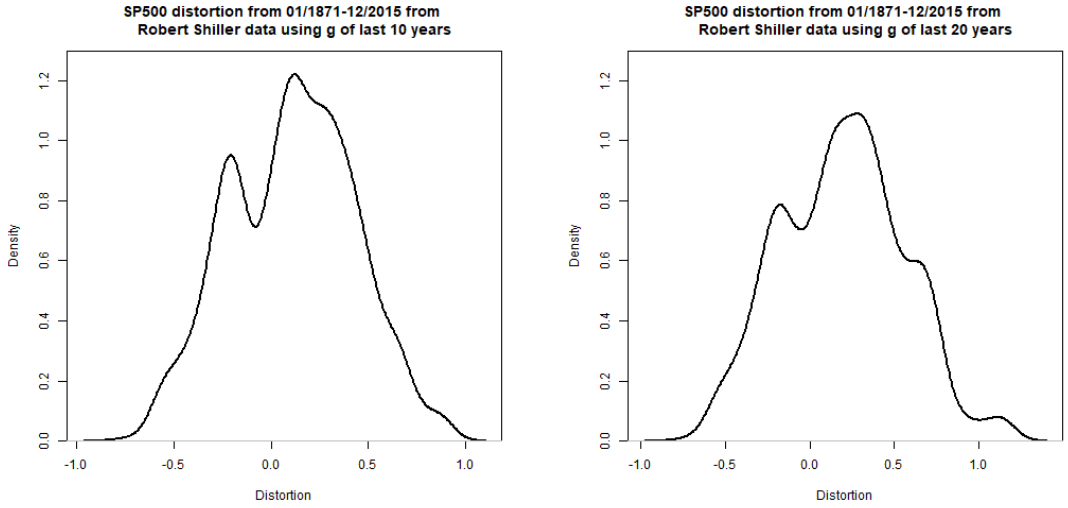


(c) Distribution of distortion using g_{30}



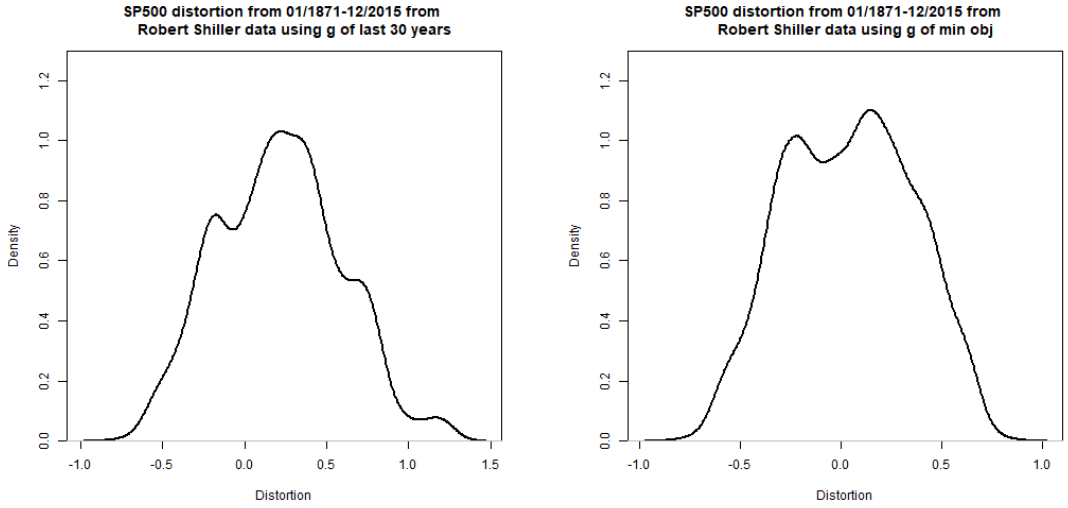
(d) Distribution of distortion using $g_{\min}$

Figure 3.3: Figure 3.3 shows the distribution of the price distortion using a g of the last 10, 20, 30 years and g minimising the objective function. When using a fundamental value with a g_{10} it appears to produce a price distortion with a distinctly bimodal distribution. The modality of distributions for each of the other price distortions is not as visually clear - they do however appear to be multi-modal.



(a) Density of distortion using g_{10}

(b) Density of distortion using g_{20}



(c) Density of distortion using g_{30}

(d) Density of distortion using $g_{\min}$

Figure 3.4: Figure 3.4 shows the smoothed distributions of the price distortions in Figure 3.3.Using g_{10} and $g_{\min}$ produce price distortions with visually bimodal distributions. g_{20} and g_{30} produce price distortions with multimodal distributions of three or more modes.

It is important to note that in Schmitt and Westerhoff (2017), there is no discount and growth rate mentioned using the period January 1871 to December 2015. The discount rate r (7.6% per year) and growth rate g_{10} (5.1% per year) used, are the rates calculated for the period January 1871 to December 2013, by Shiller Shiller (2015). When these rates are used, the fundamental value time series they produce are similar to the results in Schmitt and Westerhoff (2017), however this was not considered in this report. Additionally, the bandwidth used to smoothed all histogram in Figure 3.3 is calculated from (3).[21]

3.3 Robert Shiller: 1871 to 2021

In this section, 1800 monthly observations are considered from January 1871 to December 2021 from Robert Shiller's data set. The ending period CPI is defined as December 2021. Annualised dividends

[21] See subsection 6.1 for formal modality testing.

are divided by 12 to get the monthly dividend. (26) is used to obtained the Total Real Return of the S&P 500 on a monthly basis. The real monthly return is then used to calculate the discount rate (r), which is defined as the geometric average monthly real return on the market from January 1871 to December 2021. The geometric average monthly growth rate of real dividends is calculated in a similar manner, but for 3 periods (the last 10, 20 and 30 years) and minimising the objective function[22]. Table 3.3 shows these values calculated from the Shiller data set.

Time period [mm/yyyy]	r
01/1871-12/2021	0.57%
	g
01/2011-12/2020	0.45%
01/2001-12/2020	0.24%
01/1991-12/2020	0.17%
Min obj. function	0.43%

Table 3.3: Table of the monthly discount rate and the four different growth rates calculated from the Shiller data set for the period ending 12/2021. See Appendix C for the log fundamental time series for these growth rates (per month) of the last 10, 20, 30 years of real dividends and one minimising the objective function not annualised.

3.4 Bloomberg: 1970 to 2021

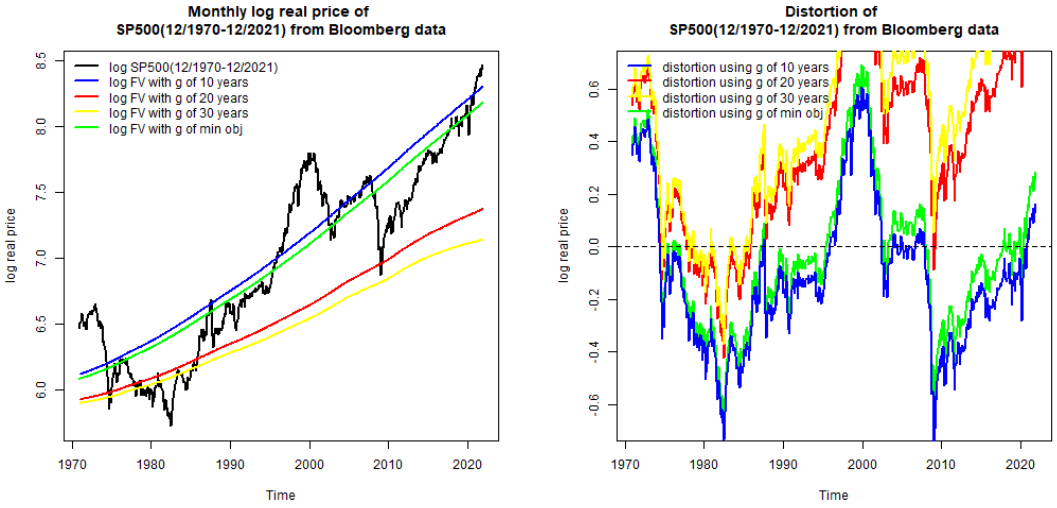
The S&P 500 data set, imported from the Bloomberg Terminal, consists of only 613 monthly observations after removing missing data[23]. This data set is thus considerably small in comparison to the Robert Shiller S&P 500 data from January 1871 to December 2021 which consists of 1800 monthly observations. The Real Price and Real Dividends were calculated the exact same way as described in subsection 3.3. The geometric average monthly real return on the market, from 12/1970 to 12/2021, was calculated as r. Similarly to the Shiller fundamental value time series, the geometric average monthly real growth rate of dividends is computed as g. Table 3.4 shows these values calculated from the Bloomberg data set, and can be compared to Table 3.3. Interestingly enough, these r and g values are relatively similar.

Time period [m/yyyy]	r
12/1970-12/2021	0.56%
	g
01/2011-12/2020	0.44%
01/2001-12/2020	0.25%
01/1991-12/2020	0.17%
Min obj. function	0.42%

Table 3.4: Table of the monthly discount rate and the four different growth rates calculated from the Bloomberg data set for the period ending December 2021. See Figure 3.5 for the log fundamental paths for these growth rates (per month) of the last 10, 20, 30 years of real dividends and minimising the objective function not annualised.

It is crucial to observe that Figure 3.5 is similar, if not identical to Figure C.1, if the price and fundamental value time series in Figure C.1 are considered, from December 1970[24].

[22] The minimised objective function is defined in (28).
[23] Mainly due to the starting date for dividend data.
[24] See Appendix E for additional visual results.



(a) Log Fundamental value series

(b) Price Distortion series

Figure 3.5: Figure 3.5a shows the monthly log fundamental value time series calculated using a g_{10}, g_{20}, g_{30} and $g_{\min}$ and the constant discount rate r from Table 3.4, together with the monthly log real price of the S&P 500 from 12/1970 to 12/2021. It is important to note that smaller growth rates result in more flat log fundamental value time series. The fundamental value calculated using g_{20} and g_{30}, are much lower in comparison to fundamental value time series using g_{10} $g_{\min}$. Additionally, it can be seen that those two fundamental values experience a downward shift - which is attributed to the number of dividend observations. This is not seen in subsection 3.3. Figure 3.5b shows the monthly price distortion for each of the four fundamental value time series. Once again, see that the price distortion calculated using g_{20} and g_{30} result in more positively weighted price distortions, in comparison to g_{10} and $g_{\min}$.

3.5 Additional Information on the discount and growth rates

As discussed above, the discount rate (7.6% per year) and growth rate (5.1% per year), as used by Shiller (2015) and subsequently in Schmitt and Westerhoff (2017), were not recoverable from the Shiller data set - regardless of the methods attempted. Methods included: using the first month of every year from 1871-2013 and calculating the real price and real dividends (in 06/2014 dollars) and then using the geometric average return to get r and g, using the last month of every year from 1871-2013 and calculating the real price and real dividends (in 06/2014 dollars) and then using the geometric average return to get r and g.[25] Further inspection is required of the g used by Shiller (2015). It is defined as the geometric average monthly growth rate of real dividends for the past 10 years for the period 2003-2013. However, this period is 11 years long. It should be for the period starting 2003 and ending 2012 (10 years).[26]

Furthermore, in Shiller (2001) the historical geometric average real monthly return on the market from January 1871 to June 1999 is calculated as r (0.6% per month) and the historical average growth rate of real dividends from January 1871 to December 1999 is calculated as g (0.1% per month).[27] These values are similar to the values recovered for r and g [28].

Additionally, it is easily noticed that the reason Figure 3.2a differs from that of in Schmitt and Westerhoff (2017) is due to the lower discount rate (r) calculated. A lower discount rate shifts the whole fundamental value time series upwards while keeping the overall shape. It is entirely possible to find a discount rate where the log fundamental time series lies very close to the starting point of the log real price.

[25] Geometric average return as defined in (27), compunded monthly returns.
[26] See Figure F.1, Figure F.2 and Figure F.3 for excel snippets.
[27] Real prices are in terms of January 2000 dollars.
[28] See Figure F.4 for excel snippet.

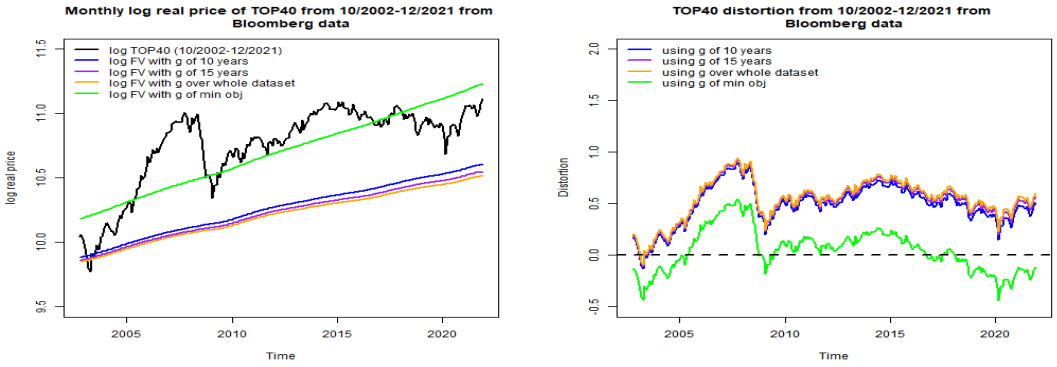
4 TOP 40

4.1 Bloomberg: 2002 to 2021

The TOP 40 data set considers the time period from October 2002 to December 2021 and consists of 231 monthly observations. In comparison to both S&P 500 data sets, it is a small data set with much fewer price and dividend observations. The aim is to implement the methods, used by Shiller (2015) and subsequently Schmitt and Westerhoff (2017), on the S&P 500 and the distributional properties of its price distortion and see if any of these properties are consistent for the JSE TOP 40 index. The real price is obtained using (25)[29]. For the TOP 40, the ending period CPI is 12/2021[30]. Annualised dividends are divided by 12 to get the monthly dividend. (26) is used to obtained the Total Real Return of the TOP 40 on a monthly basis. The real monthly return is then used to calculate the discount rate (r), which is defined as the geometric average monthly real return on the market from October 2002 to December 2021. The geometric average monthly growth rate of real dividends is calculated in a similar manner, but for 3 periods (the last 10, 15 years and over the entire data set) and minimising the objective function [31]. Table 4.1 shows these growth rates calculated for the last 10 (g_{10}), 15 (g_{15}) years, over the whole data set (g_ω) and minimising the objective function ($g_{\min}$).

Time period [mm/yyyy]	r
10/2002-12/2021	0.71%
	g
01/2011-12/2020	0.13%
01/2006-12/2020	0.10%
10/2002-12/2020	0.08%
Min obj. function	0.40%

Table 4.1: Table of the monthly discount rate and the four different growth rates calculated from the TOP 40 data set for the period ending December 2021. See Figure 4.1 for the log fundamental paths for these growth rates (per month) of the last 10 years, 15 years, over the whole dataset of real dividends and minimising the objective function and not annualised.

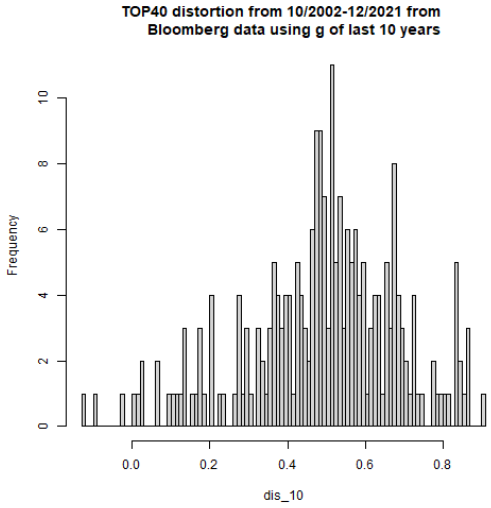


(a) Log Fundamental value series

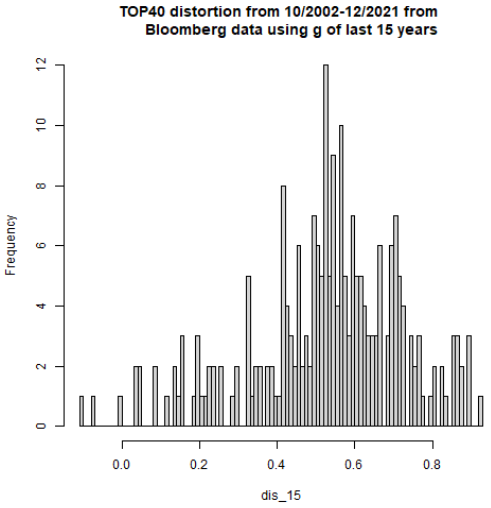
(b) Price Distortion series

Figure 4.1: Figure 4.1a, shows the monthly log fundamental value time series calculated using g_{10}, g_{15}, g_ω and $g_{\min}$ and the constant discount rate r from Table 4.1, together with the monthly log real price of the TOP 40 from October 2002 to December 2021. It is important to note that smaller growth rates result in more flat log fundamental value time series. The fundamental value calculated using g_{10}, g_{15} and g_ω, are much lower in comparison to fundamental value time series using $g_{\min}$. Figure 3.5b shows the monthly price distortion for each of the four fundamental value time series. Once again, see that the price distortion calculated using g_{10}, g_{15} and g_ω result in more positively weighted price distortions, in comparison to $g_{\min}$.

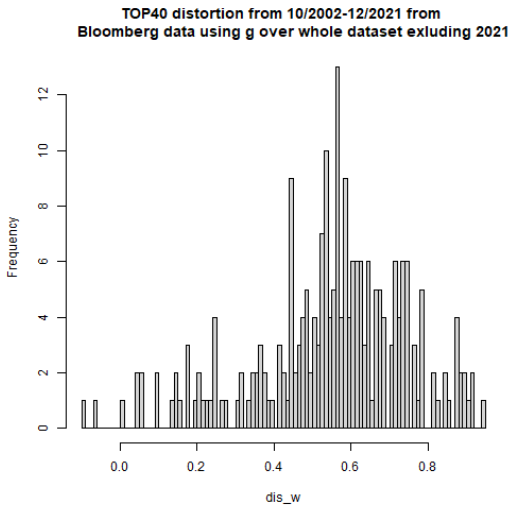
[29] Derived from formulas used in the Shiller excel file.
[30] Prices and dividends are in terms of December 2021 Rands.
[31] The minimised objective function is defined in (28).



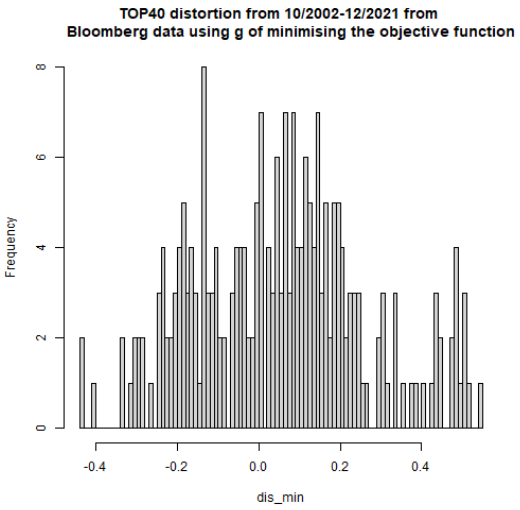
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using g_{15}

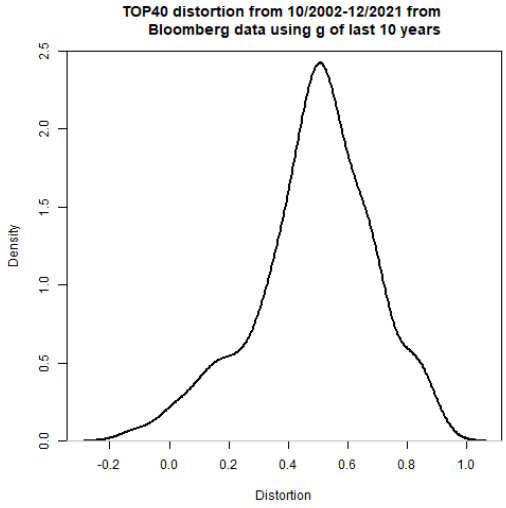


(c) Distribution of distortion using g_{ω}

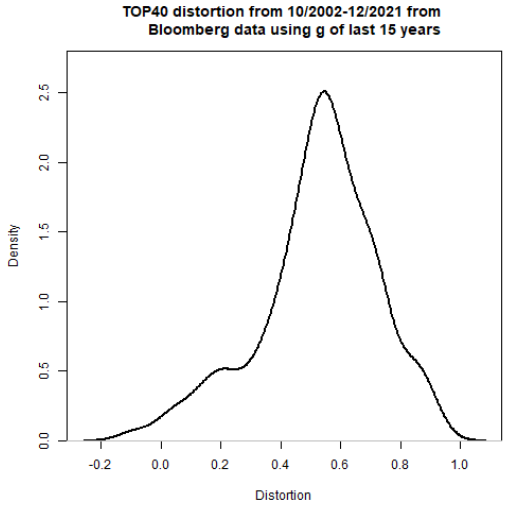


(d) Distribution of distortion using $g_{\min}$

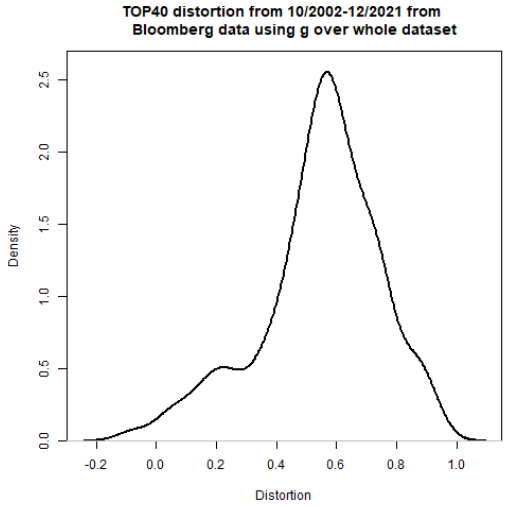
Figure 4.2: Figure 4.2 shows the distribution of the price distortion using a g of the last 10, 15, one over the whole data set and g minimising the objective function. Figure 4.2a (g_{10}) and Figure 4.2d ($g_{\min}$) appear to produce some form of a multimodal impression for the underlying distribution.



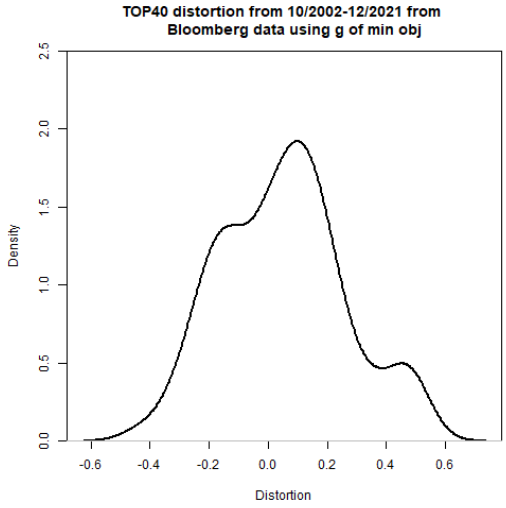
(a) Smooth density distortion using g_{10}



(b) Smooth density distortion using g_{15}



(c) Smooth density distortion using g_{ω}.



(d) Smooth density distortion using $g_{\min}$

Figure 4.3: Figure 4.3 shows the smoothed distributions of the price distortions in Figure 4.2. When smoothing the distributions, there appears to be a less clear multimodal distribution for the price distortion using g_{10} Figure 4.3a. However, looking at Figure 4.3d, note that $g_{\min}$ produces an underlying distribution for the price distortion that appears to be multi-modal.

5 Other Bloomberg Major Market Indices

Above, the methods extrapolated from Schmitt and Westerhoff (2017) are implemented on the two major market indices of interest in this research - the S&P 500 and the TOP 40. These methods have been implemented on 7 other major market (namely the FTSE 100, DAX 40, NIKKEI 225, HANG SENG, BOVESPA, NIFTY 50 and ASE) indices[32]. The only deviation from the implementation above is that only two dividend growth rates g were calculated from the data sets, one being the average real dividend growth rate of the last 10 years g_{10} and the other being the growth rate that

[32]See Table 2.1 for information regarding the starting date for each of the time series and the number of observations in addition to this the real price and real dividends for all indices are in terms of the countries currency at December 2021.

minimises the objective function[33] $g_{\min}$[34] The real price, fundamental value, price distortion and distribution plots for each of the 7 major market indices mentioned above can be found in Appendix J [35].

6 Modality Testing

Statistical methods are used, in this section, to test the modality of the distributions of the various price distortions. Schmitt and Westerhoff (2017) use the Silverman Test, in Silverman (1981), to test for the number of modes the density of their price distortion has. The null hypothesis of the Silverman Test says that the underlying density has at most k modes. The alternative hypothesis, on the other hand, says that it has more than k modes. The test uses 10000 bootstrap replications to calculate the frequency that the critical bandwidth of one bootstrap sample data set is greater than the critical bandwidth of the given data. This event is equivalent to the sample data having more than k modes. The resulting frequency is the p-value. The "silvermantest" package in R [36] is used to test the modality of the price distortions in this section[37].

6.1 Robert Shiller S&P 500: 1871 to 2015

This section focuses on the set of price distortions given by the Robert Shiller price and fundamental value time series from 1871 to 2015, as in subsection 3.2 above. Modality testing is done on the densities for each of the four price distortions found using the different growth rates (g_{10}, g_{20}, g_{30} and $g_{\min}$). In Table 6.1, the results of the computation of the Silverman Test on this data set are seen. Each row in Table 6.1 shows the p-value for up until the null hypothesis is accepted ($p > 0.05$) for each growth rate g.[38]

g	$k=1$	$k=2$	$k=3$	$k=4$
g_{10}	0	0.058	-	-
g_{20}	0	0.007	0.229	-
g_{30}	0	0.051	0.021	0.984
$g_{\min}$	0	0.201	-	-

Table 6.1: Each row shows the computed p-value for the null hypothesis of testing that the density of the distortion using a growth rate of the last 10 (g_{10}), 20 (g_{20}), 30 (g_{30}) years and minimisming the objective function ($g_{\min}$) has at most k modes using `silverman.test()`.

6.1.1 Impact of Growth rates on Modality

From the research conducted and Table 6.1, it was identified that the growth rate used to calculate the fundamental value of the index is influential on the modality of price distortion of that very index. This section attempts to highlight the impact of differing growth rates on the modality of the price distortion. 1000 values are sampled from a uniform sequence between $(-r,r)$ [39]. Each of these values is used to calculate a fundamental value time series. As a result of the simulation, there are 1000 price distortions, each with differing underlying distributions. Below in Table 6.2, are the proportion of these price distortions that can be considered unimodal, bimodal or multimodal with three or more

[33] See (28) for the minimising objective function used.

[34] the growth rate $g_{\min}$ should be smaller than the discount rate r, which is why the bounds searched for $g_{\min}$ are given by (-1, $r - 0.001$).

[35] See Appendix J for the FTSE 100, DAX 40, NIKKEI 225, HANG SENG, BOVESPA, NIFTY 50 and ASE results.

[36] See Schwaiger and Holzmann (2022) for more information on the "silvermantest" package which was used.

[37] When the `silverman.test()` from the "silvermantest" package (Schwaiger and Holzmann, 2022) is ran without setting a seed, it will result in different p-values due to the bootstrapping that takes place and see Amod and Barnes (2022) how the seed was defined in each R script (R Core Team, 2021).

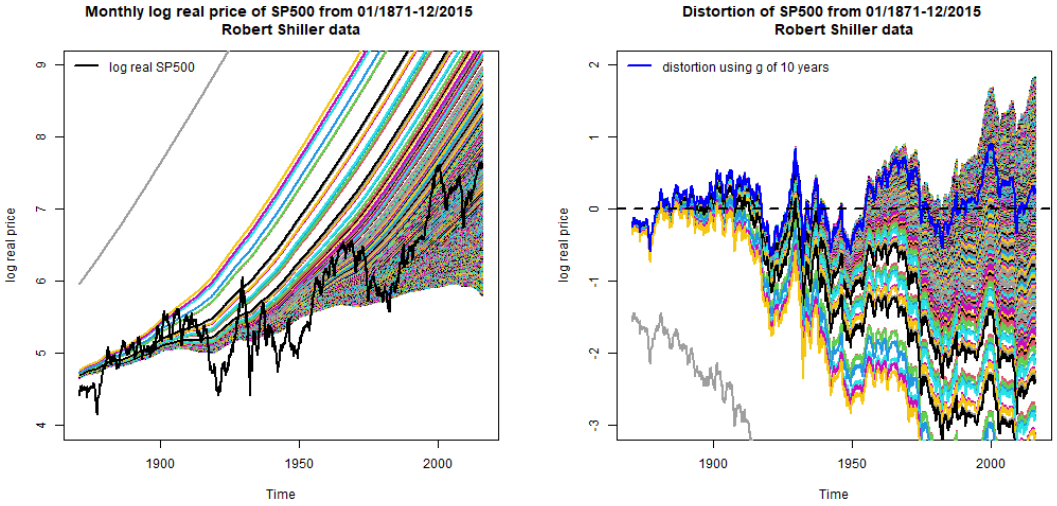
[38] $R = 10000$ bootstrap replications are used to obtain each p-value.

[39] These bounds are chosen as fundamental value cannot be computed for growth rates greater than r.

modes. Figure 6.1 shows each of the fundamental value paths and the corresponding price distortion time series.

Modes	Number of Paths	% of total
$k = 1$	26	2.6%
$k = 2$	218	21.8%
$k \geq 3$	756	75.6%

Table 6.2: Shows the proportion of price distortions, from the 1000 simulated g values, that have 1, 2 or more than 3 modes (k modes). The Silverman Test was used to determine the modality of each price distortion.



(a) Simulated Fundamental Value paths

(b) Simulated Price Distortion paths

Figure 6.1: Figure 6.1a shows the log real S&P 500 price with 1000 log fundamental value paths calculated from the 1000 simulated values of g. It is seen that many paths begin close to log real price initially, but diverge more as time progresses. Note the singular log fundamental path that can be seen extremely undervaluing the log S&P 500 price. On further inspection, this path results from a value of g that is close to fixed discount rate r as in Table 3.2.

6.1.2 Significance and non-significance of p-values

From the testing conducted on this data set, it is seen that when using the Silverman Test to test whether the price distortion using g_{30} has at most 2 modes, significant and non-significant p-value relative to 0.05 are found[40]. This occurred numerous times, which is why 1000 Silverman Tests were ran using 10000 bootstrap replications. The number of p-values > 0.05 were recorded. Surprisingly, the results showed that 73.7% of the tests were not significant relative to 0.05. This is concerning as the modality of the price distortion using g_{30} appears to be visually similar to the modality of the price distortion using g_{20}, but their modality classifications, from (Silverman, 1981; Schwaiger and Holzmann, 2022), are vastly different.

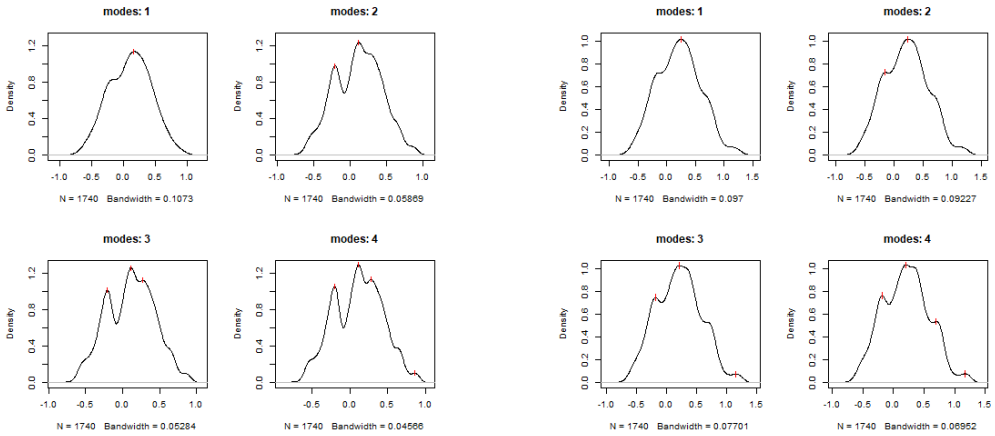
A similar problem was experienced when the number of bootstrap replications were set to the default amount R = 999. Nonetheless, the data from an empirical perspective, does not appear to be strictly bimodal or have weak bimodal characteristics and properties. Therefore, the p-value reporting may change and appear to be significant or non-significant due to the randomness of the bootstrap sampling if no seed is set.

[40] The recording or non-recording of these values are influential (see p-hacking).

6.1.3 Influence of Critical Bandwidth on Modality

bw	k=1	k=2	k=3	k=4
g_{10}	0.107280	0.058692	0.052842	0.045659
g_{30}	0.096997	0.092270	0.077009	0.069517

Table 6.3: Table 6.3 show a decreasing critical bandwidth (computed from the `h.crit()` in R from the "silvermantest" package for smoothing the density of the distortion using a growth rate of the last 10 years and 30 years. As the critical bandwidth decreases the number of modes increase, this is visually seen in Figure 6.2.



(a) Price Distortion densities (g_{10})

(b) Price Distortion densities (g_{30})

Figure 6.2: Figure 6.2a visually shows how the number of modes change, for price distortions using a g of the last 10 years, as the bandwidths change (as shown in Table 6.3 above). Figure 6.2a visually shows how the number of modes change, for price distortions using a g of the last 30 years, as the bandwidths change (as shown in Table 6.3 above).

6.2 Bloomberg TOP 40

This section focuses on the set of price distortion time series calculated using the JSE TOP 40 data set [41]. Modality testing is done on the densities for each of the four price distortions that were found using the different growth rates (g_{10}, g_{15}, g_ω and $g_{\min}$). In Table 6.4, the results of the computation of the Silverman Test on this data set are seen. Similar to Table 6.1, each row in Table 6.4 shows the p-value for up until the null hypothesis is accepted ($p > 0.05$) for each growth rate g.

g	k = 1	k = 2	k = 3	k = 4
g_{10}	0.440	-	-	-
g_{15}	0.446	-	-	-
g_ω	0.479	-	-	-
$g_{\min}$	0	0.023	4e-04	1

Table 6.4: Each row shows the computed p-value for the null hypothesis of testing that the density of the distortion using a growth rate of the last 10 (g_{10}), 15 (g_{15}), whole dataset (g_ω) years and minimisming the objective function ($g_{\min}$) has at most k modes using `silverman.test()`.

[41] As seen in section 4,

g	unimodal	bimodal	tri-modal
g_{10}	TRUE	FALSE	FALSE
g_{15}	FALSE	TRUE	FALSE
g_{ω}	FALSE	TRUE	FALSE
$g_{\min}$	FALSE	TRUE	FALSE

Table 6.5: Shows the number of modes detected using the default density smoothing for each of the price distortions using the `nr.modes()` part of the "silvermantest" package.

`nr.modes()` (Schwaiger and Holzmann, 2022) calculates the number of modes of a density as follows:

```
d1 <- diff(y)        # where y is the density
signs <- diff(d1 / abs(d1))
length(signs[signs == -2])
```

6.2.1 Impact of Growth rates on Modality

Similar to what was done for the S&P 500, this section considers the effect of growth rates on modality of the underlying distribution of the price distortions of the TOP 40 index. Again, 1000 values are sampled from a uniform sequence between $(-r,r)$. Each of these values is used to calculate a fundamental value time series and the corresponding price distortion time series. Below in Table 6.6 and Table 6.7, are the proportion of these price distortions that can be considered unimodal, bimodal, tri-modal or multimodal with four or more modes [42]. Figure 6.3 shows each of the simulated fundamental value paths and the corresponding price distortion time series [43].

Modes	$k=1$	$k=2$	$k=3$	$k\geq 4$
Number of paths	320	403	146	131
% of total	32.0%	40.3%	14.6%	13.1%

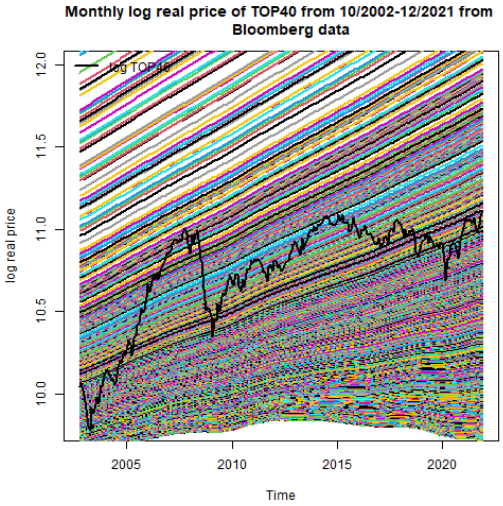
Table 6.6: This table shows the number of distortion paths from $\boldsymbol{g}$ that have k modes from the Silverman Test (Schwaiger and Holzmann, 2022; Silverman, 1981).

Modes	Number of paths	% of total
$k=1$	130	13.0%
$k=2$	734	73.4%
$k=3$	136	13.6%
$k\geq 4$	0	0%

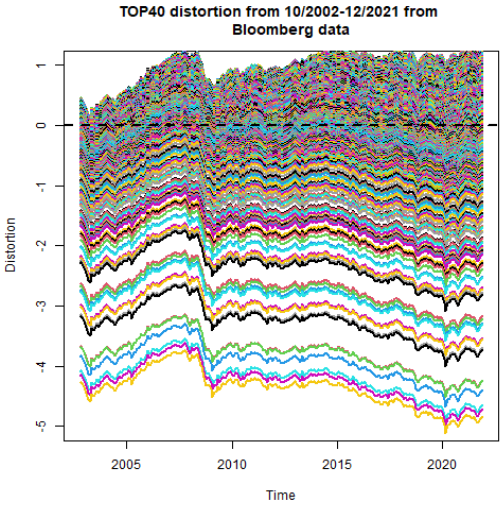
Table 6.7: Shows the number of price distortion paths where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022).

[42] Price distortions that the Silverman Test concludes have at most one mode, are removed when calculating the number of price distortions with at most two modes. Similarly, for at most three modes.

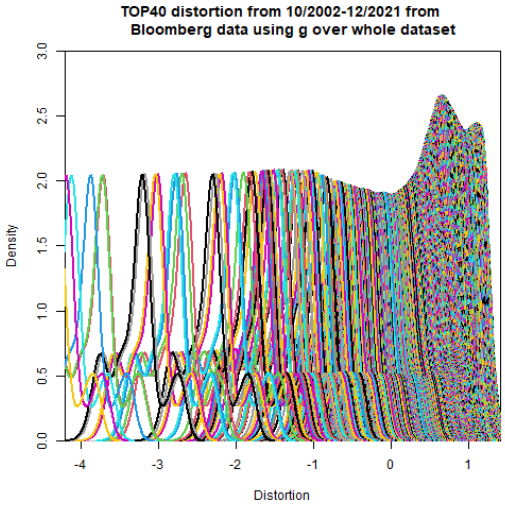
[43] See Appendix I Figure I.1a and Figure I.1b for the contour plots of these simulations.



(a) Simulated Fundamental Value paths



(b) Simulated Price distortion paths



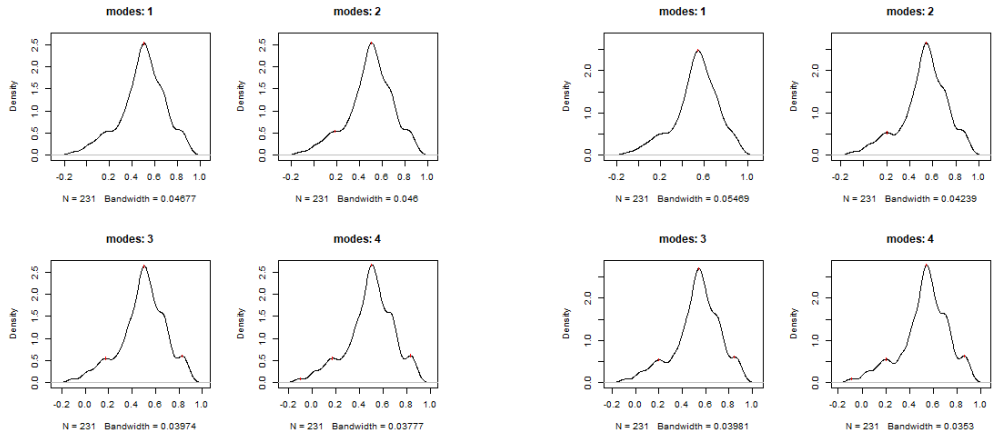
(c) Smoothed Densities of simulated price distortions

Figure 6.3: Figure 6.3a shows the log real TOP40 price with 1000 log fundamental value paths calculated from the 1000 simulated values of g. As g increases, the fundamental value paths appear to shift up - inspectioo suggests that this is due to the number of observations. Figure 6.3b shows the simulated price distortions and Figure 6.3c shows the smoothed densities of these simulated price distortions. For increasing g, it appears that the distribution of the associated price distortion has a visually multimodal impression. Figure I.1 shows the contour of these plots.

6.2.2 Influence of Critical Bandwidth on Modality

bw	k=1	k=2	k=3	k=4
g_{10}	0.046766	0.046000	0.039743	0.037774
g_{15}	0.054687	0.042386	0.039805	0.035301

Table 6.8: Shows the critical bandwidth (computed from the `h.crit()` function in R (R Core Team, 2021) from the "silvermantest" package. The density of the price distortions for two growth rates are considered (g_{10} and g_{15}). As the critical bandwidth decreases the number of modes increase. This is visually seen in Figure 6.4

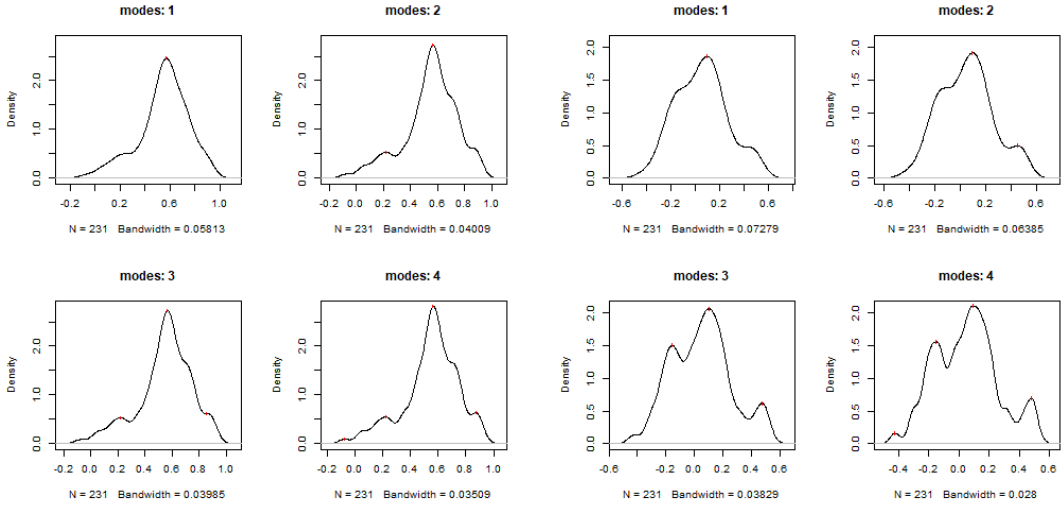


(a) Density of price distortions (g_{10}) (b) Density of price distortions (g_{15})

Figure 6.4: Figure 6.4a visually shows how the number of modes change, for price distortions using a g of the last 10 years, as the bandwidths change (as shown in Table 6.8 above). Figure 6.4b visually shows how the number of modes change, for price distortions using a g of the last 15 years, as the bandwidths change (as shown in Table 6.8 above).

bw	k=1	k=2	k=3	k=4
g_ω	0.058131	0.040086	0.039846	0.035094
$g_{\min}$	0.072790	0.063850	0.038292	0.027998

Table 6.9: Shows the critical bandwidth (computed from the `h.crit()` function in R (R Core Team, 2021) from the "silvermantest" package (Schwaiger and Holzmann, 2022). The density of the price distortions for two growth rates are considered (g_ω and $g_{\min}$). As the critical bandwidth decreases the number of modes increase. This is visually seen in Figure 6.5.



(a) Density of price distortions (g_ω) (b) Density of price distortions ($g_{\min}$)

Figure 6.5: Figure 6.5a visually shows how the number of modes change, for price distortions using a g over the whole data set, as the bandwidths change (as shown in Table 6.9 above). Figure 6.5b visually shows how the number of modes change, for price distortions using a g that minimises the objective function, as the bandwidths change (as shown in Table 6.9 above).

7 Standard Linear Time Series Models

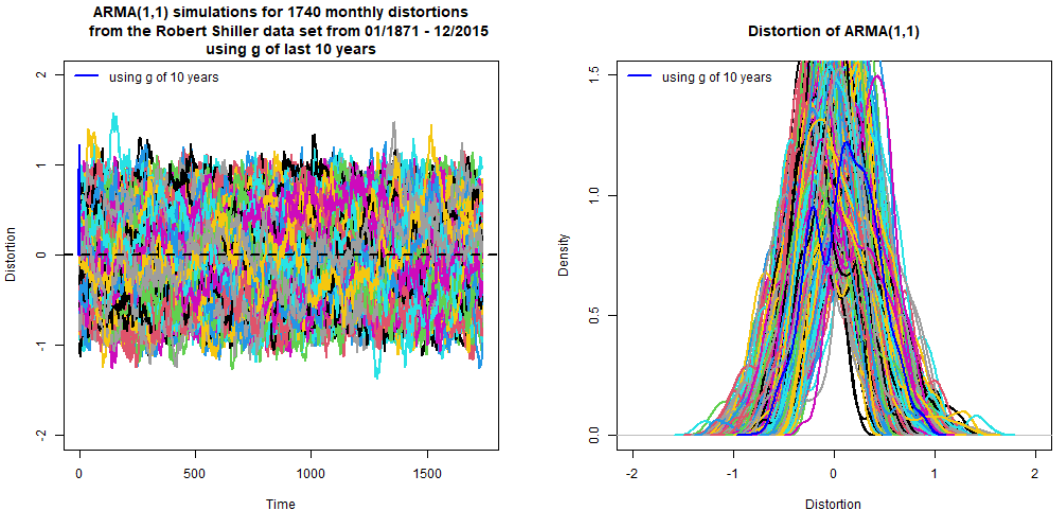
Schmitt and Westerhoff (2017) conclude that a linear time series model is unable to explain the supposed bimodal distribution found for the S&P500 price distortion. Below, simulations are conducted to assess whether the conclusion of Schmitt and Westerhoff (2017) is indeed correct with respect to the distribution of the S&P 500 price distortion. Linear time series models are then fit to the price distortion of the TOP 40 to see if one of them replicates the dynamics of the TOP 40 price distortion as well.

7.1 Robert Shiller S&P 500: 1871 to 2015

Schmitt and Westerhoff (2017) find that the two linear time series models that best fit the S&P 500 price distortion are the ARMA(1, 1) and the ARMA(2, 2). [44] These two linear time series models are considered and fitted to the S&P 500 price distortion [45]. Additionally, an ARIMA(4,1,1) model is fitted to the price distortion [46].

7.1.1 ARMA(1,1)

An ARMA(1, 1) is fit to the price distortion, whose fundamental value is calculated using a growth rate of the last 10 years (g_{10}) [47]. 1000 ARMA(1,1) simulations are then done to determine the number of modes using the Silverman Test (Silverman, 1981; Schwaiger and Holzmann, 2022). Figure 7.1 shows the 1000 simulations using the ARMA(1,1) model and their smoothed densities. Table 7.1 shows the proportion of these 1000 price distortion simulations that have one, two or more than three modes.



(a) 1000 ARMA(1,1) Price Distortions

(b) Smoothed Densities

Figure 7.1: Figure 7.1a shows 1000 distortion simulations from the ARMA(1,1) model and Figure 7.1b shows the smoothed densities of the price distortions.

[44] Schmitt and Westerhoff (2017) use Mathematica's in built function TimeSeriesModelFit to find these models and the configuration.

[45] Using the "forecast" package from Hyndman and Khandakar (2008) in R (R Core Team, 2021).

[46] See Appendix G for all linear time series model parameters.

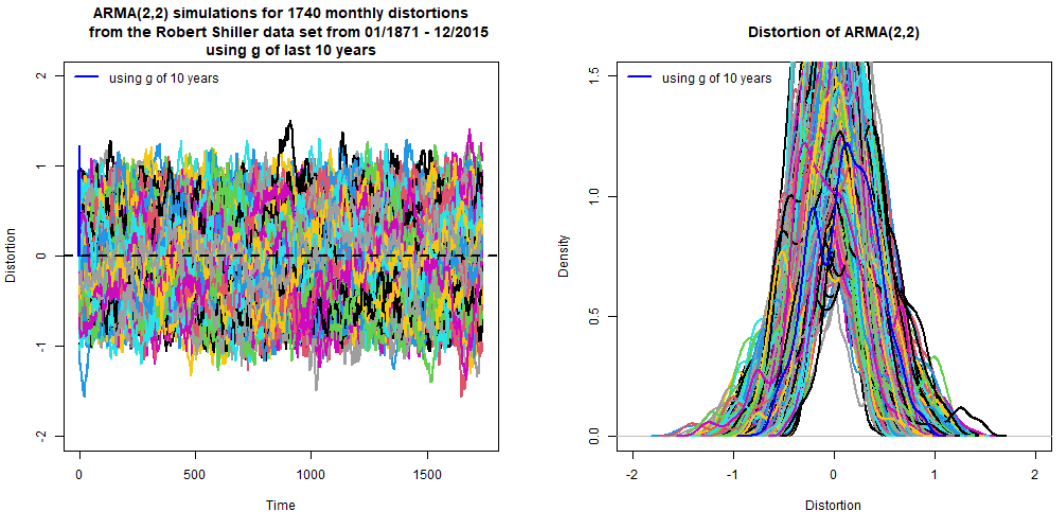
[47] Using the "forecast" package from Schwaiger and Holzmann (2022) in R.

Modes	Number of paths	% of total
$k = 1$	50	5.0%
$k = 2$	309	30.9%
$k \geq 3$	641	64.1%

Table 7.1: Shows the proportion of these 1000 simulations that have one, two or more than 3 modes. The Silverman Test was used to determine the modality of each price distortion.

7.1.2 ARMA(2,2)

An ARMA(2, 2) is fit to the price distortion, whose fundamental value is calculated using a growth rate of the last 10 years (g_{10}) [48]. 1000 ARMA(2,2) simulations are then done to determine the number of modes using the Silverman test (Silverman, 1981; Schwaiger and Holzmann, 2022) [49]. Figure 7.2 shows the 1000 simulations using the ARMA(2,2) model and their smoothed densities. Table 7.2 shows the proportion of these 1000 price distortion simulations that have one, two or more than three modes.



(a) 1000 ARMA(2,2) price distortions

(b) Smoothed density distortions

Figure 7.2: Figure 7.2a shows 1000 distortion simulations from the ARMA(2,2) model and Figure 7.2b shows the smoothed densities of the price distortions.

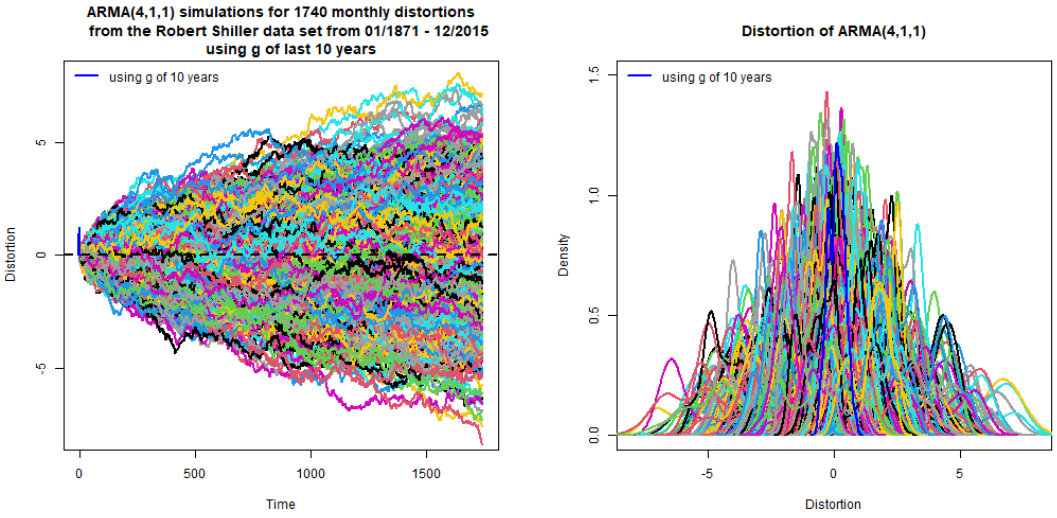
Modes	Number of paths	% of total
$k = 1$	75	7.5%
$k = 2$	506	50.6%
$k \geq 3$	419	41.9%

Table 7.2: shows the proportion of these 1000 simulations that have one, two or more than 3 modes. The Silverman Test was used to determine the modality of each price distortion.

[48] Using the "forecast" package Schwaiger and Holzmann (2022) in R.
[49] Schmitt and Westerhoff (2017) use a dip test to test the modality of the price distortions.

7.1.3 ARIMA(4,1,1)

An ARIMA(4, 1, 1) is fit to the price distortion, whose fundamental value is calculated using a growth rate of the last 10 years (g_{10}) [50] [51]. 1000 ARIMA(4, 1, 1) simulations are then done to determine the number of modes using the Silverman Test (Silverman, 1981; Schwaiger and Holzmann, 2022). Figure 7.3 shows the 1000 simulations using the ARIMA(4, 1, 1) model and their smoothed densities. Table 7.3 shows the proportion of these 1000 price distortion simulations that have one, two or more than three modes. An ARIMA(4,1,1) was found to be a better fit of the S&P500 price distortion, than the ARMA(1, 1) or ARMA(2, 2). Table 7.4 shows the fit statistics for the various models based on the AIC, BIC and AICC. The ARIMA(4,1,1) was found to be the best model based on the AICC minimisation.



(a) 1000 ARMA(4,1,1) price distortions

(b) Smoothed density distortions

Figure 7.3: Figure 7.3a shows 1000 distortion simulations from the ARIMA(4,1,1) model and Figure 7.3b shows the smoothed densities of the price distortions.

Modes	Number of modes	% of total
$k = 1$	75	7.5%
$k = 2$	498	49.8%
$k \geq 3$	427	42.7%

Table 7.3: shows the proportion of these 1000 simulations that have one, two or more than 3 modes. The Silverman Test was used to determine the modality of each price distortion.

[50] Using the the `auto.arima()` function in the "forecast" package (Schwaiger and Holzmann, 2022) in R.

[51] `auto.arima()` uses a combination of unit root tests, minimization of the AIC and MLE to obtain the best fit ARIMA model.

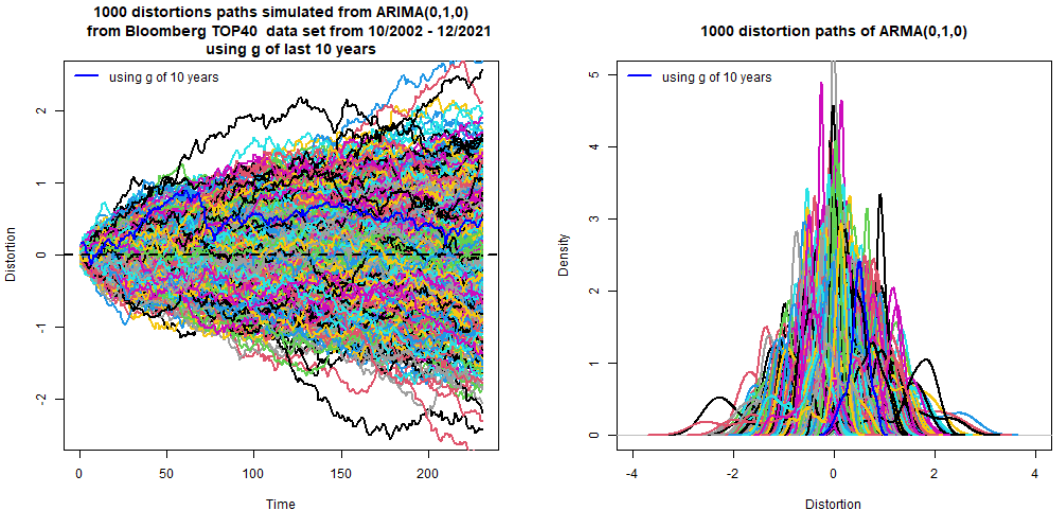
Fit statistics			
	AIC	BIC	AICC
ARIMA(4,1,1)	-6321.932	-6289.165	-6321.883
ARMA(2,2)	-6320.443	-6287.673	-6320.395
ARMA(1,1)	-6319.284	-6297.437	-6319.261

Table 7.4: shows the fit statistics for the above time series models. It appears that the ARIMA(4,1,1) has the best fit based on the AICC. Additionally, this contrasts the claim that the price distortion (using g_{10}) is at least weakly stationary. One should notice that the ARIMA(2,2) and ARIMA(1,1) have AICC values close to that of the ARIMA(4,1,1).

7.2 TOP 40

Linear Time Series models are fit to each of the four price distortions found for the TOP 40 index (using the varying growth rates: g_{10}, g_{15}, g_{ω} and g_{min}) [52]. An ARIMA(0, 1, 0) is found to be the best fit linear model to the price distortion for each of the growth rates (g_{10}, g_{15}, g_{ω} and g_{min}) [53]. 1000 ARIMA(0, 1, 0) simulations are then done for each of the price distortions to determine the modality of the underlying distribution [54]. Note that the behaviour of the residuals for these linear time series models were not considered [55] [56].

7.2.1 Growth rate of the last 10 years (g_{10})



(a) 1000 ARMA(0,1,0) price distortions

(b) Smoothed density distortions

Figure 7.4: Figure 7.4a shows 1000 distortion simulations from the ARIMA(0,1,0) model and Figure 7.4b shows the smoothed densities of the price distortions. See Appendix I Figure I.2a for the contours of these simulations.

[52] Using the the `auto.arima()` function in the "forecast" package (Schwaiger and Holzmann, 2022) in R.
[53] An ARMA(1,1) is fit to the TOP 40 price distortion in Appendix K with the residuals of the the ARMA(1,1) model in Appendix L.
[54] Using the Silverman Test Schwaiger and Holzmann (2022); Silverman (1981).
[55] This is done for other markets in Appendix L.
[56] See Appendix H Table H.1 for noise variance of ARIMA(0,1,0) models below.

Modes	$k = 1$	$k = 2$	$k = 3$	$k \geq 4$
Number of paths	42	365	472	121
% of total	4.2%	36.5%	47.2%	12.1%

Table 7.5: Shows the proportion of these 1000 simulations that have one, two, three or more than 4 modes (k modes). The Silverman Test (Schwaiger and Holzmann, 2022; Silverman, 1981) was used to determine the modality of each price distortion.

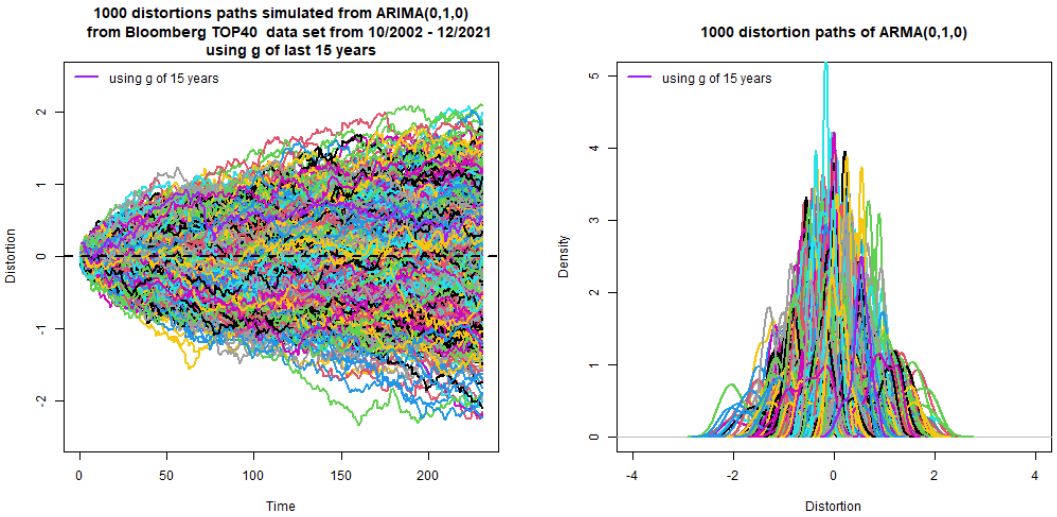
Modes	Number of paths	% of total
$k = 1$	110	11.0%
$k = 2$	642	64.2%
$k = 3$	231	23.1%
$k \geq 4$	17	1.7%

Table 7.6: Shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022).

Fit statistics			
	AIC	BIC	AICC
ARIMA(0,1,0)	-743.0338	-739.5957	-743.0163

Table 7.7: shows the fit statistics (AIC, BIC and AICC) for the above time series models. The ARIMA(0,1,0) model was chosen on the basis of minimising the AICC the best, minimising the AICC which is the default function of `auto.arima()` from the forecast package Hyndman and Khandakar (2008). Note that for large n, the AICC converges to the AIC. Additionally, in this instance, the model would also be selected on the basis of it minimising the AIC and BIC the best.

7.2.2 Growth rate of the last 15 years (g_{15})



(a) 1000 ARMA(0,1,0) price distortions

(b) Smoothed density distortions

Figure 7.5: Figure 7.5a shows 1000 distortion simulations from the ARIMA(0,1,0) model and Figure 7.5b shows the smoothed densities of the price distortions. See Appendix I Figure I.2b for the contours of these simulations.

Modes	k=1	k=2	k=3	$k \geq 4$
Number of paths	42	366	460	132
% of total	4.2%	36.6%	46.0%	13.2%

Table 7.8: Shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022).

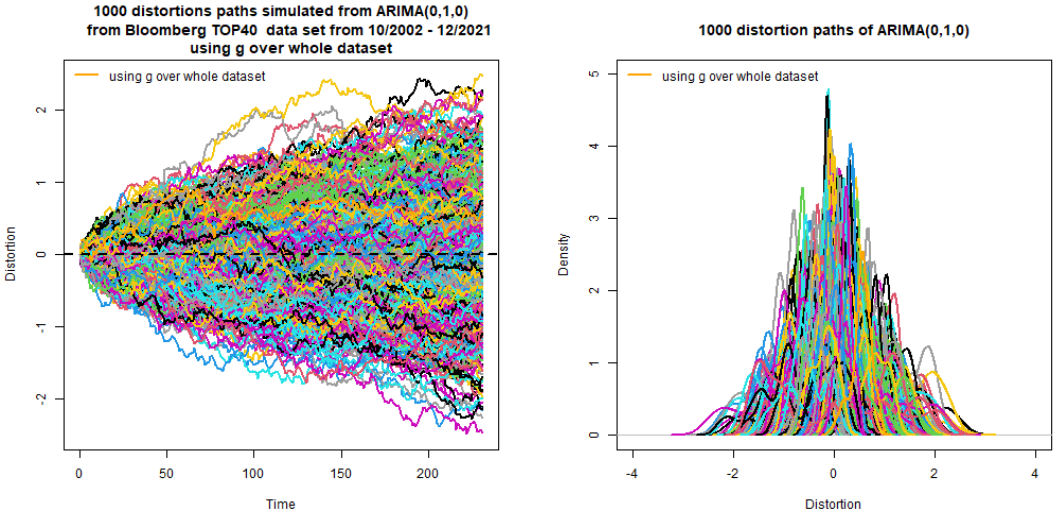
Modes	Number of paths	% of total
$k = 1$	123	12.3%
$k = 2$	633	63.3%
$k = 3$	225	22.5%
$k \geq 4$	19	1.9%

Table 7.9: shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022).

Fit statistics			
	AIC	BIC	AICC
ARIMA(0,1,0)	-743.0245	-739.5864	-743.0069

Table 7.10: shows the fit statistics (AIC, BIC and AICC) for the above time series models. The ARIMA(0,1,0) model was chosen on the basis of minimising the AICC the best, minimising the AICC which is the default function of `auto.arima()` from the forecast package Hyndman and Khandakar (2008). Note that for large n, the AICC converges to the AIC. Additionally, in this instance, the model would also be selected on the basis of it minimising the AIC and BIC the best.

7.2.3 Growth rate over the whole data set (g_ω)



(a) 1000 ARMA(0,1,0) price distortions

(b) Smoothed density distortions

Figure 7.6: Figure 7.6a shows 1000 distortion simulations from the ARIMA(0,1,0) model and Figure 7.6b shows the smoothed densities of the price distortions. See Appendix I Figure I.3a for the contours of these simulations.

Modes	$k = 1$	$k = 2$	$k = 3$	$k \geq 4$
Number of paths	43	380	448	129
% of total	4.3%	38.0%	44.8%	12.9%

Table 7.11: Shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022).

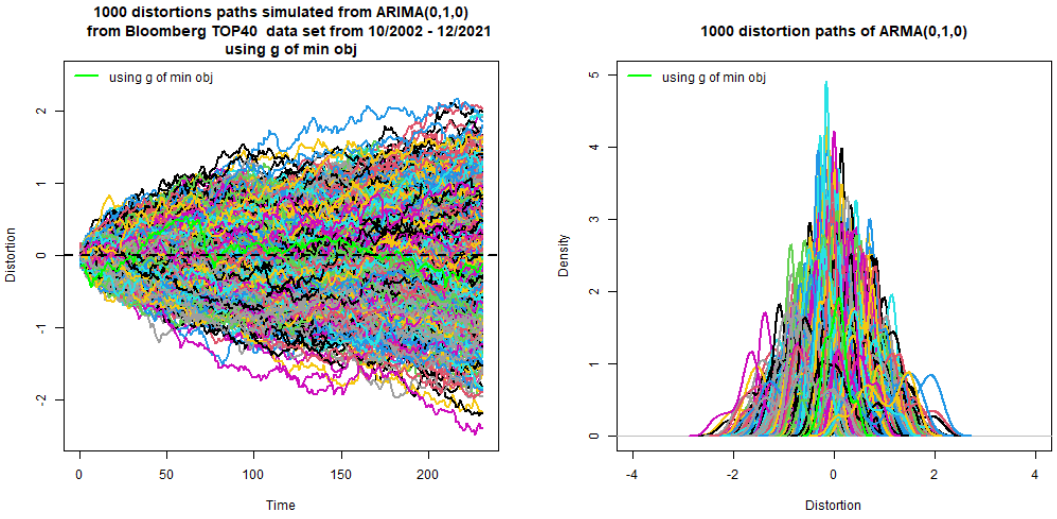
Modes	Number of paths	% of total
$k = 1$	126	12.6%
$k = 2$	634	63.4%
$k = 3$	219	21.9%
$k \geq 4$	21	2.1%

Table 7.12: shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022)

Fit statistics			
	AIC	BIC	AICC
ARIMA(0,1,0)	-742.8911	-739.453	-742.8735

Table 7.13: shows the fit statistics (AIC, BIC and AICC) for the above time series models. THe ARIMA(0,1,0) model was chosen on the basis of minimising the AICC the best, minimising the AICC which is the default function of `auto.arima()` from the forecast package Hyndman and Khandakar (2008). Note that for large n, the AICC converges to the AIC. Additionally, in this instance, the model would also be selected on the basis of it minimising the AIC and BIC the best.

7.2.4 Growth rate that minimises the objective function (g_{min})



(a) 1000 ARMA(0,1,0) price distortions

(b) Smoothed density distortions

Figure 7.7: Figure 7.7a shows 1000 distortion simulations from the ARIMA(0,1,0) model and Figure 7.7a shows the smoothed densities of the price distortions. See Appendix I Figure I.3b for the contours of these simulations.

Modes	k=1	k=2	k=3	$k \geq 4$
Number of paths	45	375	448	132
% of total	4.5%	37.5%	44.8%	13.2%

Table 7.14: Shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022).

Modes	Number of paths	% of total
$k = 1$	109	10.9%
$k = 2$	643	64.3%
$k = 3$	226	22.6%
$k \geq 4$	22	22.0%

Table 7.15: shows the proportion of ARIMA(0,1,0) simulations where the density will have k modes. These calculations are done using the default bandwidth and the `nr.modes()` function in the "silvermantest" package Schwaiger and Holzmann (2022)

Fit statistics			
	AIC	BIC	AICC
ARIMA(0,1,0)	-743.0164	-739.5783	-742.9988

Table 7.16: shows the fit statistics (AIC, BIC and AICC) for the above time series models. THe ARIMA(0,1,0) model was chosen on the basis of minimising the AICC the best, minimising the AICC which is the default function of `auto.arima()` from the forecast package Hyndman and Khandakar (2008). Note that for large n, the AICC converges to the AIC. Additionally, in this instance, the model would also be selected on the basis of it minimising the AIC and BIC the best.

8 Heterogeneous Agent Based Models

Schmitt and Westerhoff (2017); Lux (2021) propose that there exists a class of heterogeneous agent-based models that explain the visual bimodal impression of these index price distortions. These HABMs suggest that the price distortion of the S&P500 and other indices can be explained by the interaction of two types of agents, Chartists and Fundamentalists, and the market trading mechanism.

8.1 Agent-Based Model by Gaunersdorfer and Hommes

The speculators (agents) in the heterogeneous agent-based model (HABM) of Gaunersdorfer and Hommes (2007) draw on similar principles to how Shiller's Fundamental Value is calculated to calculate the fundamental value of the asset - discounting the summation of all expected future dividends by a fixed rate of return r. For the asset, a dividend process y_t that has a constant mean is assumed. The fundamental value of the asset is then given by:

$$p^* = \sum_{i=1}^{\infty} \frac{\bar{y}}{(1+r)^i} = \frac{\bar{y}}{r}, \tag{4}$$

Fundamentalists believe that the price will always shift towards the fundamental value. They calculate the expected price as:

$$E_{F,t}[p_{t+1}] = p^* + \nu(p_{t-1} - p^*), \tag{5}$$

where $0 \leq \nu \leq 1$ is the mean rate at which fundamentalists expect the price to revert to the fundamental value. Chartists, or trend-followers as they are referred to in literature, base their strategy on what they believe the price of the asset will do. If they believe it will increase, they will acquire more and similarly if they believe the price will decrease at the next time period, they would dispose of the asset. They calculate the expected price of the asset at the next time period as:

$$E_{C,t}[p_{t+1}] = p_{t-1} + g(p_{t-1} - p_{t-2}), \tag{6}$$

where $g \geq 0$ is the strength of the chartists' technical analysis. The demand for the risky asset can be derived by using the fact that all speculators are mean-variance maximisers. The demand for the risky asset by fundamentalists is given by:

$$z_{F,t} = \frac{E_{F,t}[p_{t+1} + y_{t+1} - (1+r)p_t]}{a\sigma^2}, \tag{7}$$

And similarly, the demand for the risky asset by chartists is given by:

$$z_{C,t} = \frac{E_{C,t}[p_{t+1} + y_{t+1} - (1+r)p_t]}{a\sigma^2}, \tag{8}$$

where a is the aversion to risk across all speculators. If z^s is defined as the outside supply of the risky asset, then it is known that the market equilibrium for the risky asset is given by:

$$n_{C,t}z_{C,t} + n_{F,t}z_{F,t} = z^s, \tag{9}$$

where $n_{C,t}$ and $n_{F,t}$ are the fraction of all speculators that are chartists or fundamentalists. Assuming simply that the outside supply of the risky asset is 0, the price of the risky asset at the current period t can be derived.

$$(1+r)p_t = n_{C,t}E_{C,t}[p_{t+1} + y_{t+1}] + n_{F,t}E_{F,t}[p_{t+1} + y_{t+1}], \tag{10}$$

Given that the dividend process has a constant mean, thus $E_{C,t}[y_{t+1}] = E_{F,t}[y_{t+1}] = \bar{y}$.

$$(1+r)p_t = n_{C,t}E_{C,t}[p_{t+1}] + n_{F,t}E_{F,t}[p_{t+1}] + \bar{y}, \tag{11}$$

Therefore, the price of the risky asset is given by:

$$p_t = \frac{1}{(1+r)}\left[n_{C,t}E_{C,t}[p_{t+1}] + n_{F,t}E_{F,t}[p_{t+1}] + \bar{y}\right] + \epsilon_t, \tag{12}$$

where ϵ_t is added to account for the additional dynamic noise from noise traders.

$$U_{C,t} = (p_t + y_t - (1+r)p_{t-1})z_{C,t-1} + \eta U_{C,t-1} \tag{13}$$
$$U_{F,t} = (p_t + y_t - (1+r)p_{t-1})z_{F,t-1} + \eta U_{F,t-1}, \tag{14}$$

Let $U_{C,t}$ and $U_{F,t}$ denote the accumulated profits by Chartists and Fundamentalists respectively. For the agent-based model by Gaunersdorfer and Hommes (2007), speculators decide on following the technical or fundamental financial rule based on how profitable the rule has been up to that time period and the size of the mispricing of the asset (i.e. the difference between the price and the fundamental value). The accumulated profits for each financial rule can then be used to calculate the percentage of speculators that are chartists or fundamentalists.

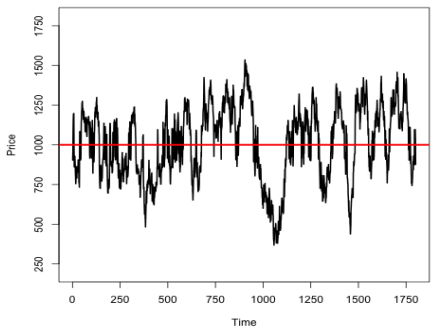
$$n_{C,t} = \frac{\exp[\beta U_{C,t-1}]}{\exp[\beta U_{C,t-1}] + \exp[\beta U_{F,t-1}]} \times \exp\left[-\frac{(p^* - p_{t-1})^2}{\alpha}\right] \tag{15}$$
$$n_{F,t} = 1 - n_{C,t}, \tag{16}$$

where the $\beta \geq 0$ parameter accounts for the rate at which speculators will switch to the financial rule that is the most profitable at that point in time. The $\alpha \geq 0$ parameter in the correction term on the right is a sensitivity measure of the chartists and fundamentalists to the mispricing of the asset in the market. The larger the mispricing the smaller the correction term, and the more speculators move towards the fundamental financial rule.

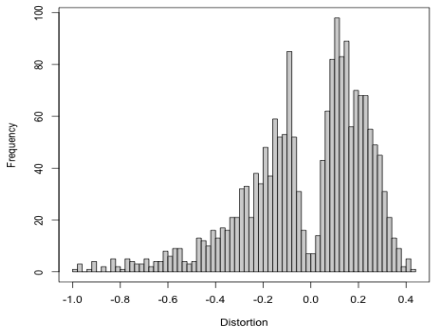
Parameter	Estimate
$\overline{y}$	1
σ_δ	0
r	0.001
ν	1
g	1.89
aσ^2	1
σ_ϵ	10
η	0
α	2000
β	2

Table 8.1: Shows the parameter estimates for the agent-based model, as done by Gaunersdorfer & Hommes **?**, used in the simulation process.

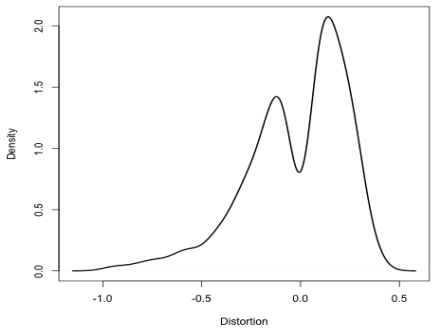
Gaunersdorfer and Hommes (2007) calibrate their model using the parameters found in Table 8.1. For this simulation of their agent-based financial market model, the same parameter calibration will be used. Additionally, to ascertain that the simulation is replicable, a seed value of 2022 is set. 37800 daily observations (252 trading days × 150 years) are simulated. Every 21st observation is then extrapolated to get 1800 monthly observations (the same length as the Shiller S&P500 data set) of the price from the simulation of this agent-based model.



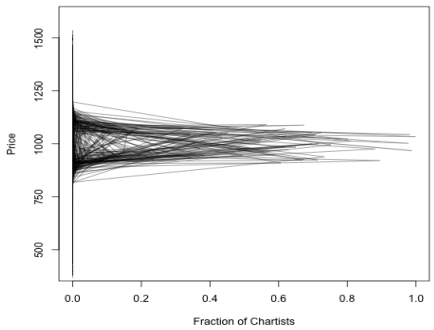
(a) Price and Fundamental Value.



(b) Price Distortion Histogram.



(c) Price Distortion Smoothed Density.



(d) Chartist Share of Market.

Figure 8.1: Figure 8.1a depicts the monthly price of the risky asset and it's fundamental value. Figure 8.1b shows a histogram of the price distortion distribution for the model. Figure 8.1c shows the density of the model's price distortion. Figure 8.1d shows how the percentage of chartists change as the price changes over time.

8.2 Agent-Based Model by Franke and Westerhoff

In comparison to the agent-based model by Gaunersdorfer and Hommes (2007), one of Franke and Westerhoff (2012) agent-based models does not consider past profitability as a determining factor when choosing between chartist and fundamentalist financial rules. Franke and Westerhoff (2012) propose that for this specific agent-based model there are three factors that influence which financial rule speculators will choose to follow. These factors are: predisposition (speculators have an affinity to either chartist or fundamentalist rules, regardless of other factors), herding (the more speculators following a certain financial rule, the more attractive it is) and market misalignments (the larger the mispricing, the more risky it is to choose the chartist financial rule). Finally, the ABM by Franke and Westerhoff (2012) does not utilise market clearing to set the price. Rather, a market-maker sets the price of the risky asset based on the orders from both chartists and fundamentalists. The price set for current period t, by the market-maker, is thus given by:

$$p_t = p_{t-1} + \mu(n_{C,t-1}d_{C,t-1} + n_{F,t-1}d_{F,t-1}), \tag{17}$$

where $n_{C,t-1}$ and $n_{F,t-1}$, as we've already established, are the fraction of all speculators that are chartists or fundamentalists in the previous time period. p_t is the log price of the risky asset. $\mu > 0$ is the adjustment parameter the market-maker uses on the price in relation to the excess demand. $d_{C,t-1}$ and $d_{F,t-1}$ are the average demands for the risky asset by an individual chartist and fundamentalist in the previous period, respectively. The demand for the risky asset by fundamentalists can be formulated as:

$$d_{F,t} = \phi(p^* - p_t) + \epsilon_{F,t},\ \epsilon_{F,t} \sim N(0, \sigma_F^2), \tag{18}$$

where ϕ is non-negative and constant parameter, $\epsilon_{F,t}$ is a random term added to account for the noise of the fundamentalist demand and p^* is the log fundamental value. Similarly, a formulation of the demand for the risky asset by chartists can be given as:

$$d_{C,t} = \chi(p_t - p_{t-1}) + \epsilon_{C,t},\ \epsilon_{C,t} \sim N(0, \sigma_C^2), \tag{19}$$

where χ is also a non-negative and constant parameter and $\epsilon_{C,t}$ is a random term added to account for the noise of the chartist demand. Almost identically to the Gaunersdorfer and Hommes (2007) agent-based model (with the exception of the correction term), the percentage of speculators that are chartists or fundamentalists is calculated as:

$$n_{C,t} = \frac{\exp[\beta U_{C,t-1}]}{\exp[\beta U_{C,t-1}] + \exp[\beta U_{F,t-1}]} \tag{20}$$

$$= \frac{1}{1 + \exp[\beta(U_{F,t-1} - U_{C,t-1})]} \tag{21}$$

$$= \frac{1}{1 + \exp[\beta a_{t-1}]} \tag{22}$$

$$n_{F,t} = 1 - n_{C,t}, \tag{23}$$

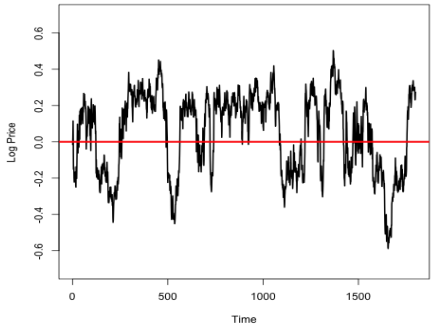
where β is a measure of the strength of the speculator's choice for either financial rule and $U_{C,t-1}$ and $U_{C,t-1}$ are the utilities associated with choosing that financial rule up to the previous time period. It can then be said that the difference between the two utilities is the relative attractiveness of one financial rule over the other. In this instance, $a_{t-1} = U_{F,t-1} - U_{C,t-1}$ is the relative attractiveness of the fundamentalist financial rule over the chartist financial rule at the previous time period. The relative attractiveness of the fundamentalist financial rule over the chartist financial rule can thus be defined as:

$$a_t = \alpha_0 + \alpha_n(n_{F,t} - n_{C,t}) + \alpha_p(p_t - p^*)^2 \tag{24}$$

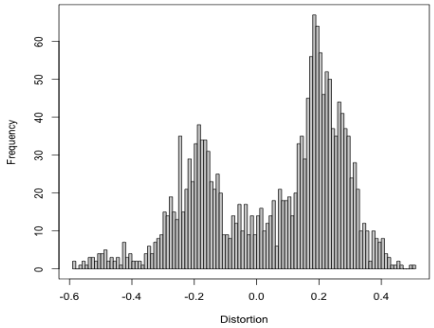
where α_0 is a constant parameter measure of the preference speculators have for either the fundamental or chartist financial rule. $\alpha_0 > 0$ indicates a preference for the fundamental financial rule and $\alpha_0 < 0$ is indicative of a preference for the chartist financial rule. α_n is a measure of the strength of the herding behaviour by speculators and α_p measures the intensity of the effect market misalignments have on speculators.

Parameter	Estimate
μ	0.001
β	1
p^*	0
χ	1.5
ϕ	0.12
σ_C	2.087
σ_F	0.758
α_0	-0.327
α_n	1.79
α_p	18.43

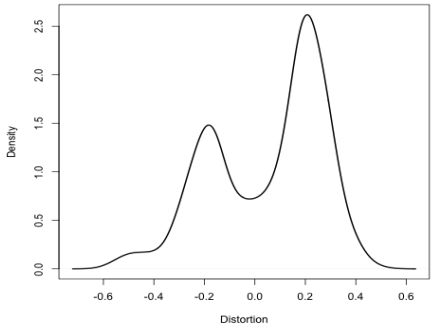
Table 8.2: Table 8.2 shows the parameter estimates for the agent-based model, as estimated by Franke & Westerhoff Franke and Westerhoff (2012), used in the simulation process.



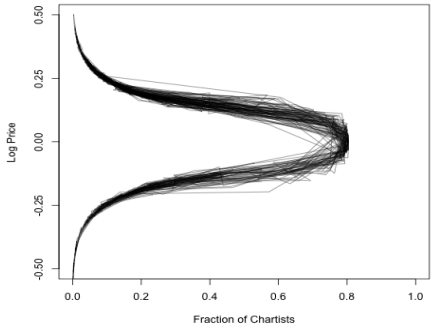
(a) Price and Fundamental Value.



(b) Price Distortion Histogram.



(c) Price Distortion Smoothed Density.



(d) Chartist Share of Market.

Figure 8.2: Figure 8.2a depicts the monthly price of the risky asset and it's fundamental value. Figure 8.2b shows a histogram of the price distortion distribution for the model. Figure 8.2c shows the density of the model's price distortion. Figure 8.2d shows how the percentage of chartists change as the price changes over time.

Franke and Westerhoff (2012) calibrate their model using the parameters found in Table 8.2. These parameters were found using the Method of Simulated Moments. For the purpose of this simulation of their agent-based financial market model, the same parameter calibration will be used. Additionally, to ascertain that the simulation is replicable, a seed value of 2022 is set. 37800 daily observations (252 trading days $\times$ 150 years) are simulated. Every 21st observation is then extrapolated to get 1800 monthly observations (the same length as the Shiller S&P500 data set) of the price from the simulation of this agent-based model.

9 Discussion of Results

9.1 S&P 500

Three different periods of time were considered for the time series constructed using the Shiller S&P 500 data set. January 1871 to December 2013, January 1871 to December 2015 and January 1871 to December 2021. The aim of the first period is to replicate the plots seen in Shiller (2015) with regards to the fundamental value, the discount rate (r) and the growth rate (g). The aim of the second period is to replicate the fundamental value time series seen in Schmitt and Westerhoff (2017), and find the the discount rate (r) and growth rate (g) used from the data set. Lastly, the intended use of the third period is to be a comparison tool used to contrast the distributions of the price distortions found using the Shiller data set to those found using the Bloomberg S&P 500 data set.

Consider the first time period (January 1871 to December 2013). Figure 3.1a shows the price and

fundamental value time series calculated using a growth rate of the last 10 years g_{10}, 20 years g_{20}, 30 years g_{30} and one that minimises the objective function $g_{\min}$ and the constant discount rate r [57]. From Figure 3.1a, it is seen that smaller growth rates result in a more flat log fundamental value time series. The impact of the smaller growth rates is seen in Figure 3.1b, which shows the price distortion for each of the four fundamental value time series. The smaller growth rates, g_{20} and g_{30}, have price distortions that are more positively weighted in favour of the price, in comparison to g_{10} and $g_{\min}$, as their log fundamental value time series are significantly under the log real price. Thus, producing an over valuation of the real price of the S&P 500, in comparison to each of the fundamental values. Looking at the distribution of the price distortion using different growth rates in Figure B.1, it is seen that the price distortion using a growth rate of the last 10 years (g_{10}) and a growth rate minimising the objective function ($g_{\min}$) appear have a bimodal visual impression for their price distortion distributions. The bimodal impression dissipates when assessing the distribution of the price distortion using a growth rate of the last 20 years (g_{20}) in Figure B.1b and 30 years (g_{30}) in Figure B.1c. These two price distortions appear to exhibit much more multimodality. The smoothed density of each of the price distortions in Figure B.2 allows for a clearer view of the modality of each distribution. Figure B.2a and Figure B.2d confirm the visual bimodality suspected in Figure B.1 for g_{10} and $g_{\min}$. The modality of the price distortions using g_{20} and g_{30} are not visually clear in Figure B.2b and Figure B.2c. However, both do appear to be multi-modal. Formal modality testing is considered later on with respect to the second time period.

The second time period (January 1871 to December 2015), is used to replicate the results from Schmitt and Westerhoff (2017). Figure 3.2a shows the price and fundamental value time series calculated using the constant discount rate r and a growth rate of the last 10 years g_{10}, 20 years g_{20}, 30 years g_{30} and one that minimises the objective function $g_{\min}$ [58]. The use of multiple growth rates to calculate the fundamental value of the index is a divergence from Shiller (2015). Shiller (2015) states that the fundamental value is calculated using an average dividend growth rate of the last 10 years. However, there is no clear reason for this. Schmitt and Westerhoff (2017) replicate Shiller's fundamental value time series, as is and without question. A main interest of the research is assessing whether changing the growth rate affects the distributional properties of the price distortion. It was found that when an average dividend growth rate of the last 20 years, the last 30 years and a growth rate found by minimising the objective function [59] were used to determine the fundamental value, the distinct bimodal distributional property of the price distortion using a 10 year growth rate dissipates or is not as visually clear. In Figure 3.4, which shows the smoothed distributions of the price distortions for the different growth rates, this is seen clearly. The fundamental value calculated using g_{10} produces a price distortion whose distribution is visually bimodal, as was found by Schmitt and Westerhoff (2017). Likewise for $g_{\min}$, even though the bimodal impression is more questionable. However, g_{20} and g_{30} produce price distortions with distributions that visually appear to be multimodal with three or more modes. This is a key finding of the research. The bimodal distributional property of the S&P500 price distortion, as discovered by Schmitt and Westerhoff (2017), is dependent on the average dividend growth rate used to calculate the fundamental value.

The third period (January 1871 to December 2021) is used to contrast the two S&P 500 data sets that were collected. Looking at the real price and real dividend comparison between the Shiller data set and the Bloomberg data set (See Appendix D), they are almost identical as one would expect. There are little to no discrepancies between those two data sets, from when the Bloomberg data begins. However, when looking at the distributions of the price distortions for both data sets, as seen in Appendix C and Appendix E, it is seen that for an average dividend growth rate of the last 10 years (which is almost identical across data sets) [60], the Shiller data maintains a visual bimodal impression for the distribution. The Bloomberg data, on the other hand, has a visual tri-modal distribution for the price distortion. This contrast between the two data sets proposes and highlights another key finding of the research. That the strong bimodal impression of the price distortion found using the Shiller data set is a function of the number of price and dividend observations in the data set.

[57] These values are found in Table 3.1.
[58] These values are found in Table 3.2.
[59] See (28) for minimising objective function.
[60] See Table 3.3 and Table 3.4 for exact growth rates.

All of the market data sets collected from Bloomberg are much smaller in comparison to the Shiller data set [61]. Later on it is seen that when the Shiller fundamental value is implemented for all of the other indices, the strong bimodal impression of the distribution of the price distortion is not recoverable, as it is Shiller S&P500 data [62]. There are many potential reasons for this. One of them being that the bimodality of the price distortion is solely a feature of the S&P 500 and is not found for any other major market indices. However, this is difficult to statistically assess when the length of the data set is not ruled out as a key influential factor in this distributional property.

9.2 TOP 40

The JSE TOP 40 index is considered as it is of interest in the research to assess whether the distributional properties of the price distortion, as found by Schmitt and Westerhoff (2017), are apparent and applicable to this index. The geometric average monthly real return on the TOP 40 (r) is 0.71% per month. For the growth rate of the last 10 (g_{10}), 15 years (g_{15}), over the whole data set (g_ω) and minimising the objective function ($g_{\min}$) the values found are 0.13%, 0.10%, 0.08% and 0.40% per month respectively[63]. Notice that g_{10}, g_{15} and g_ω appear to be roughly the same - which would suggest that their fundamental value time series are going to be close to each other. This is visualised in Figure 4.1a, which shows the price and fundamental value time series calculated using these growth rates above and the constant discount rate (r). It is evident that the fundamental value time series with similar values for the growth rate are extremely close to each other. On the other hand, $g_{\min}$, produces a fundamental value time series that lay above the other three. It passes through the real price, as one would expect due to how the growth rate is calculated [64]. A closer look at the shape of all of these fundamental value time series highlights that these series have less curvature and are more straight when compared to the fundamental values produced by the Shiller S&P 500 data set. They are much more similar in shape to the fundamental value time series calculated for the Bloomberg S&P 500 data set, which is smaller than Shiller's data set. This distinct shape factor of the fundamental value is attributed to the number of observations in the data sets. For larger data sets with more price and dividend observations, the fundamental value appears to be a better visual 'fit' of the price.

Figure 4.3 shows the smoothed distributions of each of the four price distortions found for the TOP 40. The price distortions using a g_{10}, g_{15} and g_w appear to be visually unimodal, with no multimodal impression. Conversely, looking at Figure 4.3d, $g_{\min}$ seemingly produces price distortion where visually the modality of the distribution is unclear. The visual modality of these price distortions, when compared to the visual modality of the S&P 500 suggest that the multimodal property of the distribution of the price distortion of the S&P 500 is not applicable to the TOP 40 when using the Shiller fundamental value to intrinsically value the index. The results above propose that the reason for this is related to the number of observations used in the Shiller Fundamental Value time series.

9.3 Other Major Market Indices

Consider briefly the 7 other major market indices (FTSE 100, DAX 40, NIKKEI 225, HANG SENG, BOVESPA, NIFTY 50 and ASE). Each of these data sets, like the TOP 40 data set, have less observations than both the Shiller S&P 500 data set and the Bloomberg S&P 500 data set. Therefore, issues regarding the Shiller Fundamental Value calculation and the number of dividend observations, that are apparent for the TOP 40 are discernible for all of the indices [65]. Note that many of these indices appeared to exhibit unimodal behaviour for the distribution of their price distortions, when a growth rate of the last 10 years (g_{10}) and one minimising the objective function are used to determine

[61] See Table 2.1 for the number of observations in each market data set.
[62] See Appendix J for visual impressions of distributions for the price distortions of the other major market indices.
[63] See section 4 on how these rates were calculated and Table 4.1 for the table of these values.
[64] See (28) for the minimised objective function.
[65] See Appendix J for FTSE 100, DAX 40, NIKKEI 225, HANG SENG, BOVESPA, NIFTY 50 and the ASE results.

their fundamental value time series. However, indices like the HANG SENG, NIKKEI [66] and FTSE 100 do appear have an underlying multimodal distribution under certain conditions.

9.4 Modality Testing

Statistical methods are used to test the modality of the distributions of the S&P 500 price distortions and the TOP 40 price distortions. The Silverman Test tests the null hypothesis that the underlying density has at most k modes against the alternative hypothesis that the density has more than k modes.

Regarding the Shiller S&P 500 data set (for the period 1871 to 2015) it is seen in Table 6.1 that the Silverman Test suggests that the distributions of the price distortions using a g of the last 10 years (g_{10}) and a g that minimises the objective function ($g_{\min}$), have at most two modes. The evidence, especially considering g_{10}, supports the Schmitt and Westerhoff (2017) proposal that the price distortion of the S&P 500 has an underlying bimodal distribution. However, the Silverman Test concludes that with respect to g_{20} and g_{30}, their is little evidence that these growth rates produce price distortions with distributions that have less than 3 modes. As discussed in abundance above, it was identified that the growth rate used to calculate the fundamental value of the index is influential on the modality of price distortion of that very index. Table 6.2 shows the proportion of the distributions of 1000 simulated price distortions, each with a differing g, that can be considered unimodal, bimodal or multimodal with three or more modes. Over 75% of the distributions found had at least 3 modes. The results of the simulations highlight that the g_{10} used by Schmitt and Westerhoff (2017) to support the bimodal distribution claim regarding the price distortion is only in the 21% of growth rates that produce bimodality. Interestingly, an examination of Figure 6.1 shows that the non-bimodal impressions usually occurred from a fundamental value path that spends more time overvaluing or undervaluing the market.

Modality testing is done on the densities for each of the four TOP 40 price distortions that were found using the different growth rates (g_{10}, g_{15}, g_{ω} and $g_{\min}$). As expected, when considering the strong visual unimodal impression of the distributions of their price distortions, a growth rate of the last 10, the last 15 years and one over the whole data set have at most one mode. In contrast, the empirical distribution of the price distortion using a growth rate minimising the objective function is indicated to have at most four modes. Similar to the S&P 500 price distortions, it is seen that the growth rate used to calculate the fundamental value has an impact on the on the modality of the distribution of the price distortion. Table 6.6 and Table 6.7 show the proportion of the distributions of a 1000 simulated price distortions, each with a differing g, that can be considered unimodal, bimodal, tri-modal or multimodal with four or more modes. 32% of the simulated fundamental value paths produced a price distortion with a unimodal distribution, alike to the ones found using g_{10}, g_{15} and g_{ω}. 40% of simulated fundamental value paths produced bimodality with respect to the distribution of the price distortion. While this is the largest proportion of the simulated price distortions, it is not evidence enough that the underlying distribution of the TOP 40 price distortion is bimodal or multimodal at a minimum. Figure I.1a shows the contours of the 1000 simulated log fundamental paths for the TOP 40, using the growth rates generated from a uniform distribution between $(-r,r)$. Majority of the simulated paths begin close to the log real price time series, before remaining under the real price time series for the rest of the period. This is the reason why a high portion of the simulated paths significantly over value the market, as can be seen by the high density of the price distortions in Figure I.1b.

9.5 Standard Linear Time Series Models

It is well known that linear time series model tend towards a unimodal distribution for large n. Schmitt and Westerhoff (2017) suggest that a standard linear time series model is not appropriate for explaining the dynamics of the price distortion of the S&P 500 [67]. Nonetheless, Schmitt and

[66] When calculating the growth of the last 10 years of the NIKKEI, it was double the discount rate - which in does not hold for the Gordan growth model.

[67] when using a growth rate of the last 10 years

Westerhoff (2017) find that the two linear time series models that best fit the S&P 500 price distortion are the ARMA(1, 1) and the ARMA(2, 2). The ARMA(1, 1) and the ARMA(2, 2) are fitted to the price distortion, before simulations are done using each of the linear models. Of the 1000 ARMA(1, 1) simulations fitted to the price distortion, Table 7.1 shows that 30% of the simulated price distortions have a bimodal distribution. In comparison, 4% of the simulated price distortions have an underlying multimodal distribution with more than two modes [68]. The ARMA(2, 2) simulations fitted to the S&P 500 price distortion provided a stronger case for the supposed bimodality of the underlying distribution. Table 7.2 shows that 50% of those simulated price distortions have a bimodal distribution and 41% have a multimodal distribution with more than two modes. These linear models however were not found to be the best fit of the S&P 500 price distortion. Table G.5 shows the AIC, AICC and BIC for the ARMA(1,1), ARMA(2,2) and another linear time series model, the ARIMA(4,1,1). Using the AICC and AIC as fit statistics[69], it is found that the linear time series model that best characterises the price distortion of the S&P500, when using a growth rate of the last 10 years, is the ARIMA(4,1,1). Of the 1000 ARIMA(4,1,1) simulations fitted to the price distortion, Table 7.3 shows that 49% of the simulated price distortions have a bimodal distribution and 42% of the simulated price distortions have an underlying multimodal distribution with three modes or more modes. Table G.1 shows the auto-regressive term parameters of the ARIMA(4,1,1), ARMA(1,1) and ARMA(2,2) together with the standard 95% confidence interval ($\theta \pm 1.96 s.e(\theta)$). The θ_3 parameter of the ARIMA(4,1,1) appears to have an associated 95% confidence interval (-0.080,0.051) that contains 0 (non-significant). The implication of this being that that there may be no association between the price distortion and the θ_3 parameter. Consideration was not put towards removing this parameter, but should be in further research. Similarly, further analysis hinted that the ARIMA(4,1,1) may be over-fitting the model, but analysis of the residuals was not considered in the research. Lastly, note from Table G.5 that the AIC, AICC and BIC values for all three linear models are close together, which provides support that the ARMA(1,1) and the ARMA(2,2) models are marginally worse fits of the S&P 500 price distortion when compared to the ARIMA(4, 1, 1).

Linear Time Series models are fit to each of the four price distortions found for the TOP 40 index (using the varying growth rates: g_{10}, g_{15}, g_ω and $g_{\min}$). An ARIMA(0, 1, 0), equivalent to a random walk process, is found to be the best fit linear model to the price distortion for each of the growth rates, based on the AIC, AICC and BIC fit statistics. Of the 1000 ARIMA(0, 1, 0) simulations fitted to each of the TOP 40 price distortions, the Silverman Test concludes that over 90% of each of the 1000 simulated price distortions produce multimodal distributions (see Table 7.5, Table 7.8, Table 7.11 and Table 7.14). The noise variance of the ARIMA(0,1,0) models for each of the price distortions found using the different growth rates can be found in Appendix H Table H.1. Additionally, an ARMA(1,1) model is fitted to each of the TOP 40 price distortions (using a growth rate of the last 10 (g_{10}), 15 years (g_{15}), over the whole data set (g_ω) and one minimising the objective function ($g_{\min}$)) [70]. When the fit statistics of the ARMA(1,1) models (see Table K.1) are compared to those of the ARIMA(0, 1, 0) [71], it is seen that there is a minimal difference between all three fit statistics (AIC, AICC and BIC) for each of the four TOP 40 price distortions (using the varying growth rates: g_{10}, g_{15}, g_ω and $g_{\min}$). These results propose that the ARIMA(0, 1, 0) is a marginally better fit of the TOP 40 price distortion than the ARMA(1,1) model.

The ARMA(1, 1) appears to be a consistent good fit of the price distortions for both the S&P 500 and the TOP 40. As such, an ARMA(1,1) is fit to the price distortions [72] of all of the other major market indices (FTSE 100, DAX 40, NIKKEI 225, HANG SENG, BOVESPA, NIFTY 50 and ASE) [73]. The

[68] Silverman Test (Schwaiger and Holzmann, 2022) is used to determine the modality of the simulated price distortion paths

[69] We were concerned that the ARIMA(4,1,1) may be overfitting the distortion but we do not explore any of the residuals of the model and do not consider forecasting however we did consider fitting a model based on minimising the BIC and obviously this lead to a more parsimonious model which is a ARIMA(0,1,1).

[70] See the residuals of the the ARMA(1,1) model fitted to the TOP 40 price distortion using a growth rate of the last 10 years and one minimising the objective function in Figure L.1 which appear to well behaved have being centered around 0 and appear to have low correlation between each other.

[71] See Table 7.7, Table 7.10, Table 7.13, Table 7.16.

[72] Using a growth rate of the last 10 years (g_{10} and one that minimises the objective function ($g_{\min}$).

[73] See results in Appendix K.

value fit statistics do not hold as much information when only fitting a singular linear model to the data sets. Rather, assessment is made of the residuals of each of the ARMA(1,1) models fit to the price distortions of the 7 indices (see Appendix L). The residuals appear to be centred at 0 for each of the models - which is indicative of them being unbiased. Additionally, there appears to be very low correlation among the residuals. Note that the residual distribution of some indices (the NIFTY 50 and the BOVESPA) appears to be skewed to some extent. These are not interpreted further, but caution should be advised as they could lead to unreliable predictions if the ARMA(1,1) is used to fit their price distortions. Alike to the S&P 500 and the TOP 40, the ARMA(1,1) is seemingly good fit of the price distortions of each of the other market indices.

9.6 Heterogeneous Agent Based Models

Schmitt and Westerhoff (2017) and Lux (2021) propose that there exists a class of heterogeneous agent-based models that explain the visual bimodal impression for the distribution of the S&P 500 price distortion. Figure 8.1 shows the results of the simulation of the Gaunersdorfer and Hommes (2007) agent-based model. The top left plot, Figure 8.1a, depicts the monthly price of the risky asset and it's fundamental value. The price seemingly oscillates around the fundamental value as it does for the S&P500. The model ostensibly depicts and recovers many of the empirical statistical facts of the S&P500 like it's volatility clustering. Figure 8.1a highlights these bubbles and bubble bursts. The top right plot, Figure 8.1b, shows a histogram of the price distortion distribution for the model. Similarly, the bottom left plot, Figure 8.1c, shows the density of the model's price distortion. As is displayed in these two plots, the Gaunersdorfer and Hommes (2007) agent-based model's price distortion has a strong bimodal impression when looking at its distribution. Finally, the bottom right plot, Figure 8.1d, shows how the percentage of chartists change as the price evolves over time. When the index price is in the vicinity of the fundamental value - a high percentage of speculators choose the chartist financial rule - however, as the price moves further away from the fundamental value, more speculators choose the fundamental financial rule in fear of the market misalignment.

The results of the simulation of one of Franke and Westerhoff (2012) agent-based models, are presented in Figure 8.2. The top left plot, Figure 8.2a, shows the monthly log of the price of the risky asset and the log of it's fundamental value. Similarly to the Gaunersdorfer and Hommes (2007) agent-based model, the log price fluctuates around the fundamental value as it does for the S&P500 price distortion. The agent-based model is able to produce the bubbles (extended periods of the asset price above the fundamental value) and bubble bursts (the sudden decrease of the asset price below the fundamental value after a bubble) that is seen for the S&P500. The top right plot, Figure 8.2b, and the bottom left plot, Figure 8.2c, present the distribution of the price distortion of the agent-based model. From these two plots, there is a clear visual bimodal distribution for the price distortion, using the model configuration of Franke and Westerhoff (2012). The last plot, Figure 8.2d, in the bottom right, shows how the percentage of chartists change as the price changes over time. Similarly, and more evidently than the Gaunersdorfer and Hommes (2007) agent-based model, it is seen that when the index price is close to the fundamental value - a high percentage of speculators choose to follow the chartist financial rule - however, as the price moves further away from the fundamental value, more speculators opt for the fundamental financial rule due to the strong influence of the market misalignments factor on the speculators in the agent-based model.

10 Conclusion

Schmitt and Westerhoff (2017) propose that the price distortion of the S&P 500 is bimodal. They further conclude that standard linear time series models are unable to explain this distributional property of the price distortion, but suggest that it can be explained through the dynamics of non-linear agent-based financial market models. The results of this paper propose that the S&P 500 price distortion only has an underlying bimodal distribution under certain average dividend growth rate conditions. For growth rates outside of these conditions, the distribution of the price distortion changes and exhibits multimodality (with three or more modes) more often than it does not. This

is a key finding of the above research. The modality of the underlying distribution of the price distortion is dependent on the growth rate used to calculate the fundamental value. Additionally, it is concluded that the bimodality of the S&P 500 price distortion (under certain conditions) is a feature of the number of price and dividend observations in the Shiller data set. For the S&P 500 data sets with a smaller number of observations, the strong bimodal impression dissipates into multimodality. This is further confirmed when the fundamental value is implemented for all of the other market indices, including the JSE TOP 40 index. The strong bimodal shape of the distribution of the price distortion is not recoverable for smaller data sets. Modality testing on TOP 40 index showed that regardless of the length over which the average dividend growth rate is calculated, the distribution of its price distortion cannot be considered multimodal at a minimum. The linear time series model that best characterises the price distortion of the S&P500, when using a growth rate of the last 10 years, is the ARIMA(4,1,1). However, the ARIMA(4,1,1) is only a marginally better fit, according to the fit statistics, than the ARMA(1,1) and the ARMA(2,2) models. An ARIMA(0,1,0) model, equivalent to a Random Walk process, is found to be the best linear model that most appropriately fits the price distortion of the JSE TOP 40. Still, when compared to an ARMA(1,1), the ARIMA(0,1,0), like the ARIMA(4,1,1) for the S&P 500 distortion, is only a marginally better fit of the TOP 40 Price Distortion. The final result of this research paper shows that some of the dynamics and properties of the S&P 500 price distortion, like the price bubbles and bubble bursts and its supposed bimodality, can be explained by heterogeneous agent-based financial market models and the interaction of the agents in the model, financial rules and the market trading mechanism. The results of the paper provide a valuable insight into the modality of the price distortion distribution of indices and the possible reasons that explain the given modality.

It is important to note that while this research paper has conducted a thorough investigation into the modality of the distributions of both the S&P 500 price distortion and the JSE TOP 40 price distortion, there are areas that should be explored for further research in this field. These areas include: considering an alternative method to determining the number of modes, applications of the price distortion using Shiller's fundamental value to assets who have a similar length of price history to that of the S&P 500; use of a different intrinsic calculation to the one proposed by Shiller to determine the fundamental value time series of an asset; calibration of the parameters of an agent-based model to mimic the dynamics and behaviour of the TOP 40 price distortion and those of other major market indices; and lastly further investigations into other statistical models, like Hidden Markov Models that could better explain the non-linear behaviour of the price distortions of the indices mentioned in this paper.

References

Amod, Y., Barnes, L., 2022. Honours Thesis Code. URL: https://github.com/yam0d/honours-project.git.

Benedictow, A., Boug, P., 2017. Calculating the real return on a sovereign wealth fund. Canadian Journal of Economics/Revue canadienne d'économique 50, 571–594.

Carmona, R., 2014. Statistical analysis of financial data in R. volume 2. Springer.

Chiarella, C., Dieci, R., He, X.Z., 2007. Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework. Journal of Economic Behavior & Organization 62, 408–427.

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. The journal of Finance 25, 383–417.

Franke, R., Westerhoff, F., 2012. Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. Journal of Economic Dynamics and Control 36, 1193–1211.

Gaunersdorfer, A., Hommes, C., 2007. A nonlinear structural model for volatility clustering, in: Long memory in economics. Springer, pp. 265–288.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 26, 1–22. URL: https://www.jstatsoft.org/article/view/v027i03.

Leal, S.J., 2015. Fundamentalists, chartists and asset pricing anomalies. Quantitative Finance 15, 1837–1850.

Lux, T., 2021. Can heterogeneous agent models explain the alleged mispricing of the s&p 500? Quantitative Finance 21, 1413–1433.

Majewski, A.A., Ciliberti, S., Bouchaud, J.P., 2020. Co-existence of trend and value in financial markets: Estimating an extended chiarella model. Journal of Economic Dynamics and Control 112, 103791.

Malkiel, B.G., 2003. The efficient market hypothesis and its critics. Journal of economic perspectives 17, 59–82.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Schmitt, N., Westerhoff, F., 2014. Speculative behavior and the dynamics of interacting stock markets. Journal of Economic Dynamics and Control 45, 262–288.

Schmitt, N., Westerhoff, F., 2017. On the bimodality of the distribution of the s&p 500's distortion: Empirical evidence and theoretical explanations. Journal of Economic Dynamics and Control 80, 34–53.

Schwaiger, F., Holzmann, H., 2022. silvermantest: Investigate the Number of Modes using Kernel Density Estimates. URL: https://github.com/jenzopr/silvermantest. r package version 0.1.0.

Shiller, R., 1980. Do stock prices move too much to be justified by subsequent changes in dividends? .

Shiller, R., 1987. The volatility of stock market prices. Science 235, 33–37.

Shiller, R., 2001. Irrational exuberance: Broadway .

Shiller, R., 2015. Irrational exuberance, in: Irrational exuberance. Princeton university press.

Silverman, B.W., 1981. Using kernel density estimates to investigate multimodality. Journal of the Royal Statistical Society: Series B (Methodological) 43, 97–99.

Wickham, H., Bryan, J., 2019. readxl: Read Excel Files. URL: https://CRAN.R-project.org/package=readxl. r package version 1.3.1.

Wickham, H., Hester, J., Chang, W., 2021. devtools: Tools to Make Developing R Packages Easier. URL: https://CRAN.R-project.org/package=devtools. r package version 2.4.2.

A Formulae

The Real Price and Real Dividend calculation:

$$\text{Real value} = \text{Nominal value} \times \frac{\text{Ending period CPI value}}{\text{Current period CPI value}} \tag{25}$$

Total Real Return (Real Return):

$$\text{TR}_\text{i} = \text{TR}_{\text{i}-1}(\frac{\text{RP}_\text{i} + \text{RD}_\text{i}/12}{\text{RP}_{\text{i}-1}}), \tag{26}$$

for $i = 2$,.....n where $\text{TR}_1 = \text{RP}_1$ and RP_i is the i^{th} monthly real price and $\text{RD}_\text{i}/12$ is the i^{th} monthly real dividends (not annualised).

The Discount (r) and Growth rate (g) calculation:

Geometric Average Return:

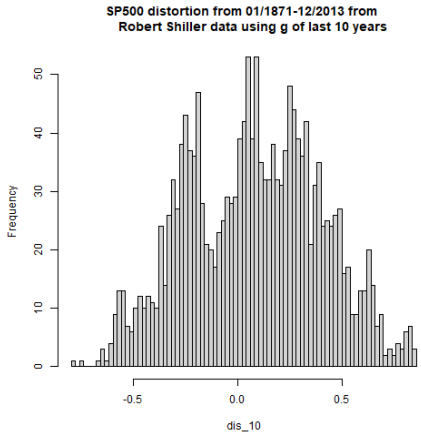
$$\text{GAR} = (\prod_{\text{i}=1}^{\text{n}}(1 + \text{r}_\text{i}))^{1/\text{n}} - 1, \tag{27}$$

where r_i is the i^{th} return (%) for total real return or real dividends and n is the number of returns.

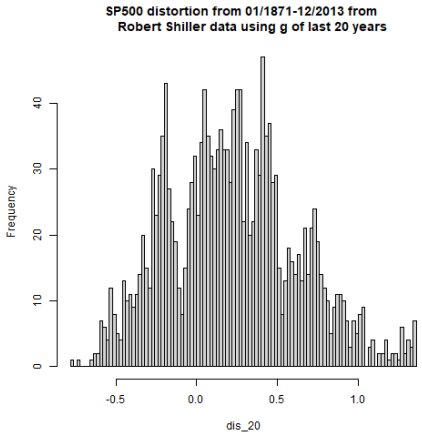
Objective Function minimised to obtain g_{min}:

$$\text{min}_\text{g} \sum_{\text{i}=1}^{\text{n}}(|\text{P}_\text{i} - \text{FV}_\text{i}|)^2 \tag{28}$$

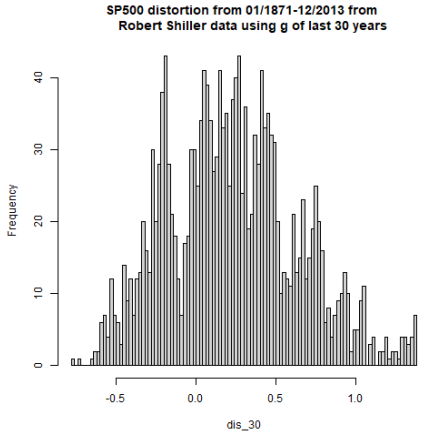
B Robert Shiller S&P 500 1871 to 2013



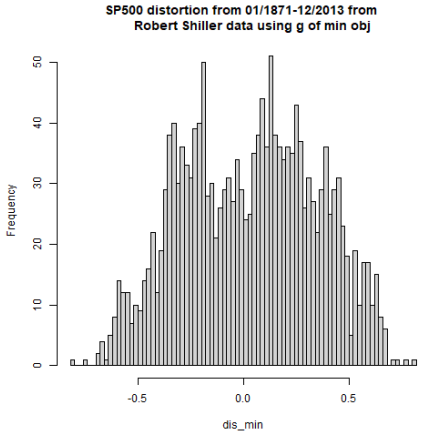
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using g_{20}

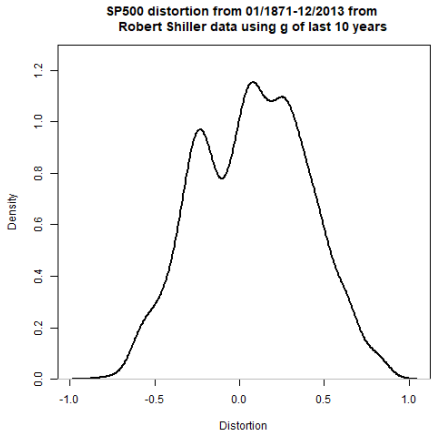


(c) Distribution of distortion using g_{30}

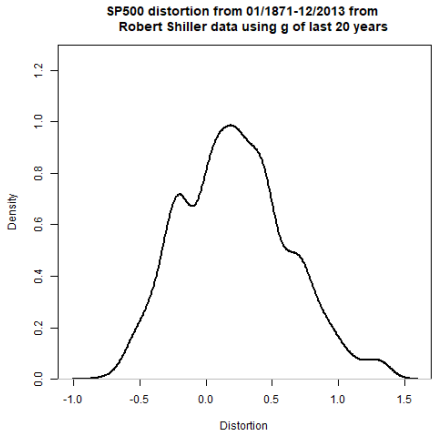


(d) Distribution of distortion using $g_{\min}$

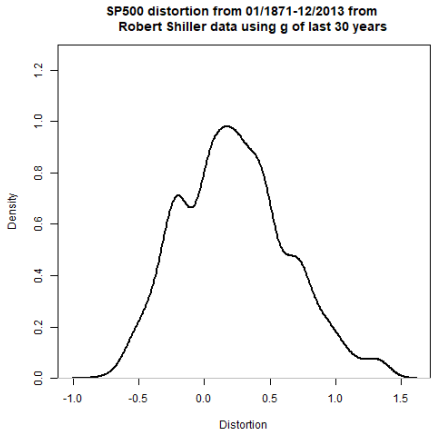
Figure B.1: In Figure B.1 above displayed are the frequency of the distortion using a g of the last 10, 20, 30 years and g minimising the objective function. Some figures appear to give a stronger bimodal impression such as when using g of the last 10 years in Figure B.1a and g when minimising the objective function in Figure B.1d while when using a g of the last 20 years in Figure B.1b and 30 years in Figure B.1c one can see some form of it being trimodal.



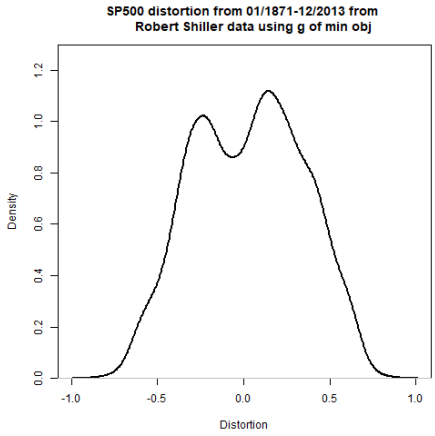
(a) Smoothed density distortion using g_{10}



(b) Smoothed density distortion using g_{20}



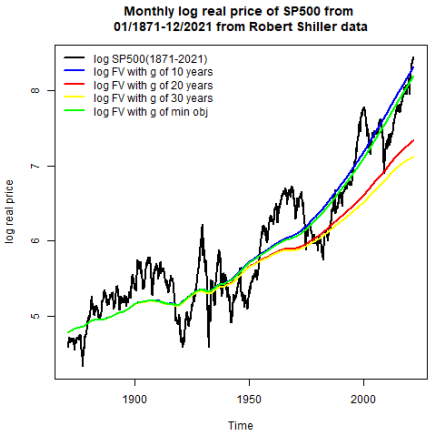
(c) Smoothed density distortion using g_{30}



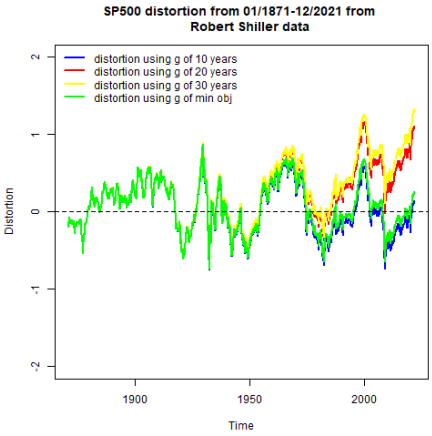
(d) Smoothed density distortion using $g_{\min}$

Figure B.2: Figure C.1 displays the smoothed densities for Figure B.1. When smoothing the distribution there appears to be a less clear bimodal distribution when using g of the last 10 years in Figure B.2a but a clear bimodal distribution when using a g minimising the objective function in Figure B.2d while when using a g of the last 20 years in Figure B.2b and 30 years in Figure B.2c it appears to have 3 or even 4 modes?.

C Robert Shiller S&P 500 1871 to 2021

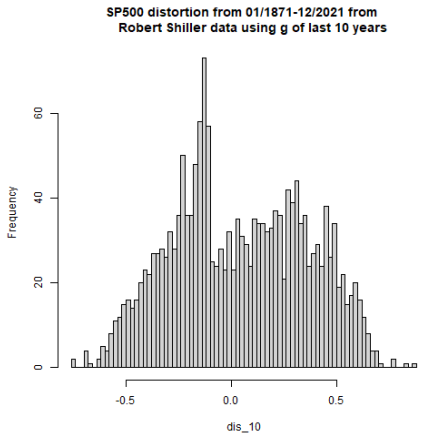


(a) Log fundamental values paths

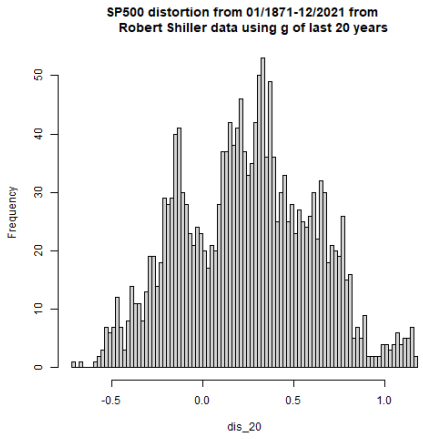


(b) Price distortions paths

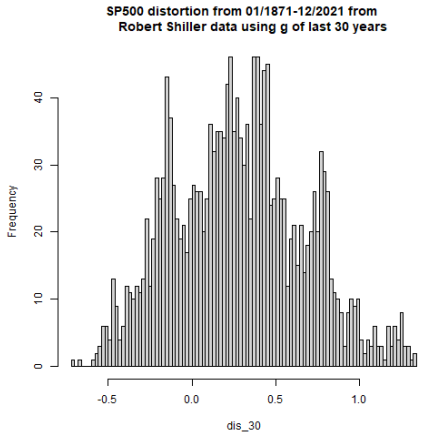
Figure C.1: In Figure C.1a, the monthly log fundamental values for using a g of the last 10, 20, 30 years and when minimising the objective function (1) is displayed together with the monthly log real price of the S&P 500 from 01/1871 to 12/2021 while keeping the discount rate r fixed as calculated in Table 3.3. From Figure C.1a it can noted that by choosing a lower g this results in the monthly log fundamental value becoming more flat as time goes on, the g of the last 20 and 30 years are much lower than that of the last 10 and the g obtained when minimising the objective function defined in equation (28), thus resulting in a more flatter line (see Table 3.3 for the values). In Figure C.1b the distortion paths which is defined as the log difference between the monthly real price of the real SP 500 and the monthly fundamental value calculated using a g of the last 10. 20, 30 years and minimising the objective function. The fundamental values calculated from g of the last 20 and 30 years result in more positive distortion due to their log fundamental lines lying significantly under the log real price and thus gives a under valuation of the log real S&P 500.



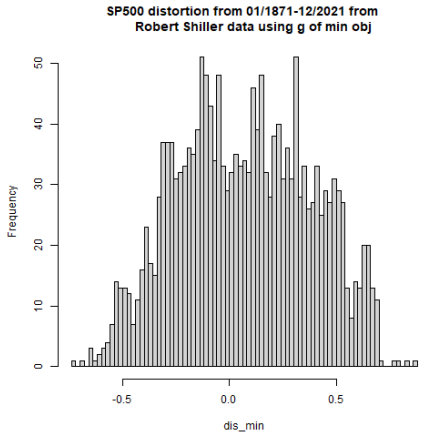
(a) Price distortion using g_{10}.



(b) Price distortion using g_{20}.

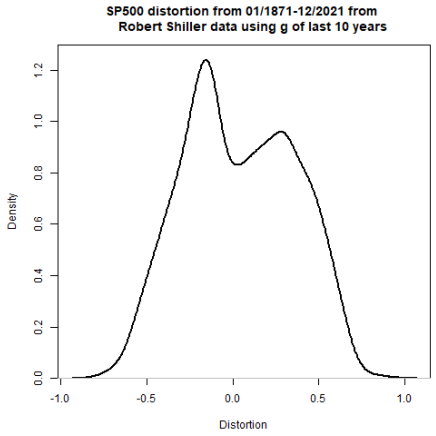


(c) Price distortion using g_{30}.

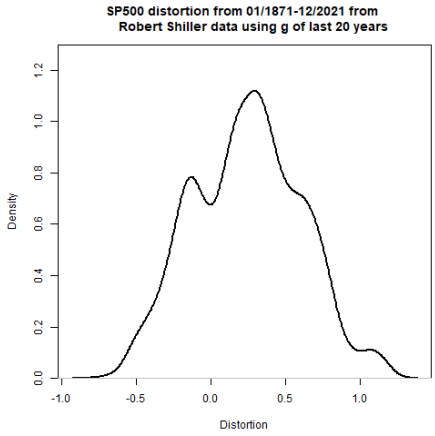


(d) Price distortion using $g_{\min}$.

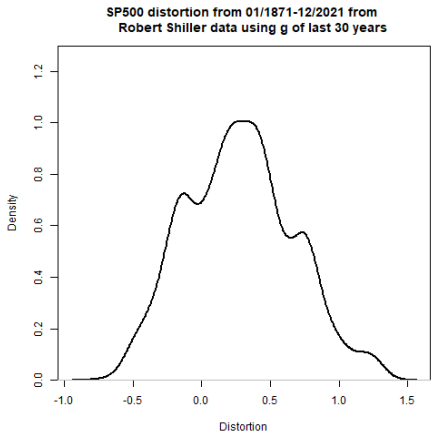
Figure C.2: In Figure C.2 displayed are the frequency of the distortion using a g of the last 10, 20, 30 years and g minimising the objective function. Some Figures appear to give a stronger bimodal impression such as when using g of the last 10 years Figure C.2a and a less clear modality when using g minimising the objective function Figure C.2d while when using a g of the last 20 years Figure C.2b and 30 years Figure C.2c one can see some form of it being trimodal.



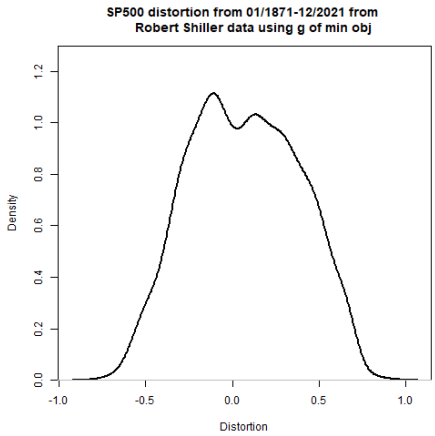
(a) Smoothed price distortion using g_{10}.



(b) Smoothed price distortion using g_{20}.



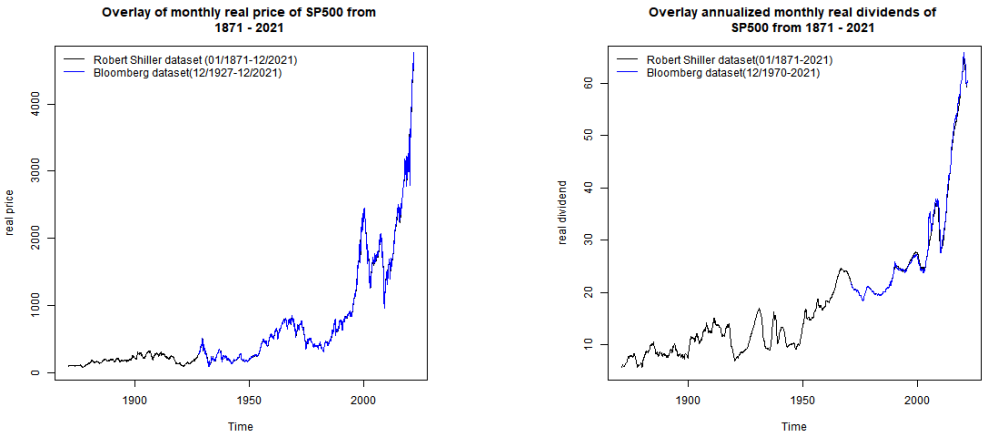
(c) Smoothed price distortion using g_{30}.



(d) Smoothed price distortion using $g_{\min}$.

Figure C.3: In Figure C.3 the smoothed densities for Figure C.2 can be seen. When smoothing the data there appears that the plots seem very similar to that of Figure B.2 but some features are more distinct such as when using g of the last 10 years appears to be bimodal now.

D S&P 500 data visualisation comparison

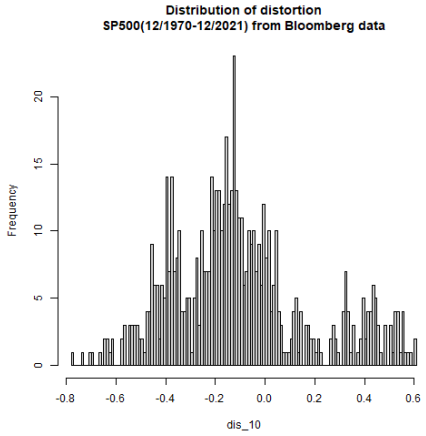


(a) Price comparison.

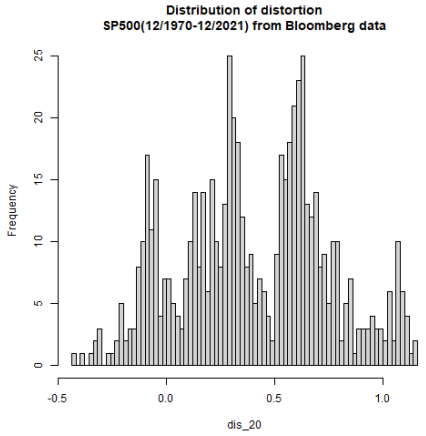
(b) Annualised dividend comparison.

Figure D.1: Figure D.1a compares the real Composite price index of obtained from Shillers data set (black line) from 01/1871 to 12/2021 to the real closing price of each month obtained from the Bloomberg data set (blue line) from 12/1927 to 12/2021. The data appears to be very similar to each other despite it being from 2 different sources and different number of observations. The same goes for the real dividends annualised in Figure D.1b. This visualisation was motivation for using the data.

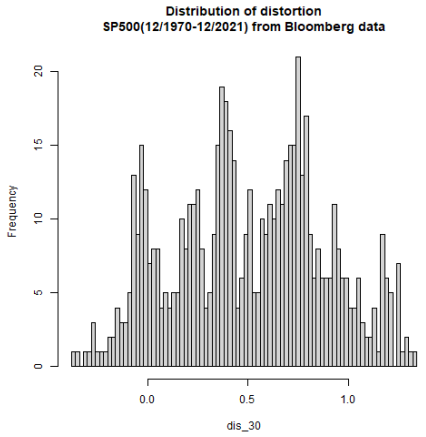
E Bloomberg S&P 500 1970 to 2021



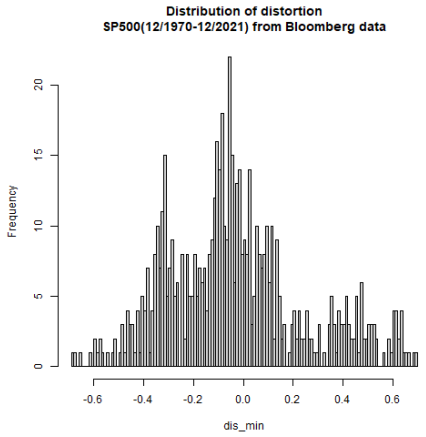
(a) Price distortion using g_{10}.



(b) Price distortion using g_{20}.

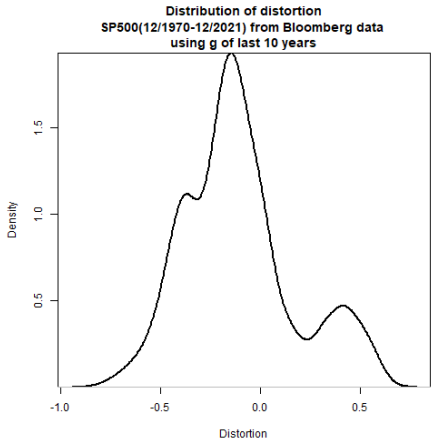


(c) Price distortion using g_{30}.

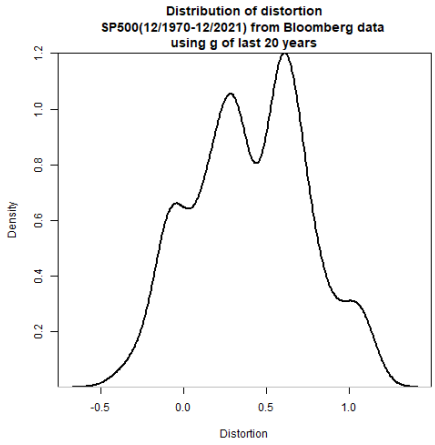


(d) Price distortion using $g_{\min}$.

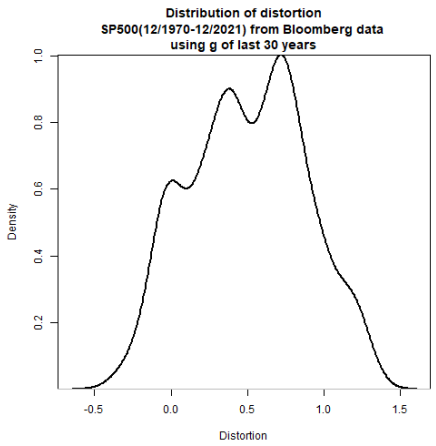
Figure E.1: In Figure E.1 are the frequency of the distortion using a g of the last 10, 20, 30 years and g minimising the objective function. Some the plots appear to give a multimodal impression such as when using g of the last 20 years Figure E.1b and g of the last 30 years Figure E.1c while when using a g of the last 10 and minimising the objective function one can see some form of it being bimodal or trimodal. This results are very different compared to that of the results obtained when consider the Robert Shiller data which was expected due to the number of observations Figure C.2



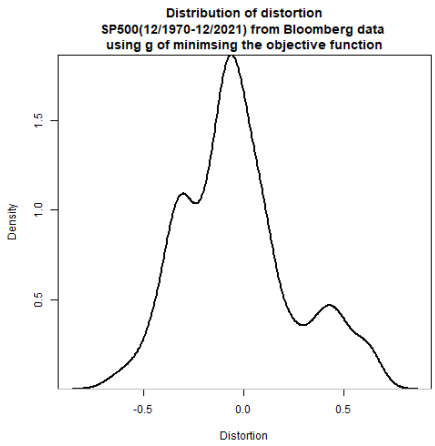
(a) Smoothed price distortion using g_{10}.



(b) Smoothed price distortion using g_{20}.



(c) Smoothed price distortion using g_{30}.



(d) Smoothed price distortion using $g_{\min}$.

Figure E.2: When smoothing E.1 the data the plots are completely different to Figure C.3 which uses roughly triple the amount of observations.

F Discount and growth rate

Taking January of every year from Sheet 2014 From 1871 - 2013													
Date	Comp. Price	Annualised Dividend	CPI	real price	Annualised real div	Total real return	return	return + 1	GAR (01/1871 to 01/2013)	real div not annualised	return	return + 1	GAR (01/2003 to 01/2012)
1871,01	4,44	0,26	12,46406	84,90354	4,971829	89,27698				4,971829			
1872,01	4,86	0,2633	12,65439	91,53715	4,959204	101,4669	0,136541	1,136541	0,065394	4,959204	-0,00254	0,997461	0,028733
1873,01	5,11	0,3025	12,93981	94,12295	5,571857	110,5095	0,089118	1,089118		5,571857	0,123539	1,123539	
1874,01	4,66	0,33	12,3689	89,79608	6,35895	112,8954	0,02159	1,02159		6,35895	0,141262	1,141262	
1875,01	4,54	0,3275	11,51265	93,99027	6,780135	126,6928	0,122214	1,122214		6,780135	0,066235	1,066235	
1876,01	4,46	0,3	10,84658	98,00419	6,59221	140,9891	0,112843	1,112843		6,59221	-0,02772	0,972283	

Figure F.1: Calculation snippet of using the first month of every year to calculate the geometric average monthly real return - r and the geometric average monthly real growth rate of dividend - g for the last 10 years from 01/2003 to 01/2012.

Taking December of every year from Sheet 2014 From 1871 - 2013													
Date	Comp. Price	Annualised Dividend	CPI	real price	Annualised real div	Total real return	return	return + 1	GAR (12/1871 to 12/2013)	real div not annualised	return	return + 1	GAR (12/2003 to 12/2012)
1871,12	4,74	0,26	12,65439	89,27698	4,897049	89,27698				4,897049			
1872,12	5,07	0,3	12,93981	93,38617	5,525809	98,91198	0,107923	1,107923	0,066773	5,525809	0,128396	1,128396	0,04359
1873,12	4,42	0,33	12,17865	86,5019	6,458287	98,46079	-0,00456	0,995438		6,458287	0,16875	1,16875	
1874,12	4,54	0,33	11,51265	93,99027	6,831892	114,7608	0,165549	1,165549		6,831892	0,057849	1,057849	
1875,12	4,37	0,3	10,94174	95,19134	6,534874	124,2063	0,082306	1,082306		6,534874	-0,04348	0,956525	
1876,12	3,58	0,3	10,75149	79,36276	6,65051	112,2307	-0,09642	0,903583		6,65051	0,017695	1,017695	

Figure F.2: Calculation snippet of using the last month of every year to calculate the geometric average monthly real return - r and the geometric average monthly real growth rate of dividend - g for the last 10 years from 12/2003 to 12/2012.

Date	Comp. Price	Annualised Dividend	CPI	real price	Annualised real div	Total real return	return	return + 1	GAR (01/1871 to 12/2013)	real div not annualised	return	return + 1	GAR (01/2003 to 12/2012)
1871,01	4,44	0,26	12,46	84,90354	4,971829	84,90354				0,414319			
1871,02	4,50	0,26	12,84	83,50124	4,824516	83,90328	-0,01178	0,988219	0,005491	0,402043	-0,02963	0,970371	0,003562
1871,03	4,61	0,26	13,03	84,29333	4,754071	85,09727	0,01423	1,01423		0,396173	-0,0146	0,985398	
1871,04	4,74	0,26	12,56	89,95346	4,934156	91,22648	0,072026	1,072026		0,41118	0,03788	1,03788	
1871,05	4,86	0,26	12,27	94,37549	5,048895	96,13779	0,053836	1,053836		0,420741	0,023254	1,023254	
1871,06	4,82	0,26	12,08	95,07304	5,128421	97,28371	0,01192	1,01192		0,427368	0,015751	1,015751	
1871,07	4,73	0,26	12,08	93,29782	5,128421	95,90452	-0,01418	0,985823		0,427368	0	1	
1871,08	4,79	0,26	11,89	95,99266	5,210458	99,12099	0,033538	1,033538		0,434205	0,015996	1,015996	
1871,09	4,84	0,26	12,18	94,72154	5,088347	98,24629	-0,00882	0,991175		0,424029	-0,02344	0,976564	

Figure F.3: Calculation snippet of using the monthly values to calculate the geometric average monthly real return - r and the geometric average monthly real growth rate of dividend - g for the last 10 years from 01/2003 to 12/2012.

Date	Comp. Price	Annualised Dividend	CPI		real price	Annualised real div		Total real return	return	return + 1	GAR (01/1871 to 06/1999)		real div not annualised	return	return + 1	GAR (01/1871 to 12/1999)
1871,01	4,44	0,26	12,46		60,13064	3,521164		60,13064					0,2934303			
1871,02	4,50	0,26	12,84		59,1375	3,416833		59,42	-0,01178	0,988219	0,005957		0,2847361	-0,02963	0,970371	0,001008
1871,03	4,61	0,26	13,03		59,69848	3,366942		60,27	0,01423	1,01423			0,2805785	-0,0146	0,985398	
1871,04	4,74	0,26	12,56		63,70711	3,494483		64,61	0,072026	1,072026			0,2912069	0,03788	1,03788	
1871,05	4,86	0,26	12,27		66,83889	3,575743		68,09	0,053836	1,053836			0,2979786	0,023254	1,023254	
1871,06	4,82	0,26	12,08		67,33292	3,632066		68,90	0,01192	1,01192			0,3026722	0,015751	1,015751	
1871,07	4,73	0,26	12,08		66,07566	3,632066		67,92	-0,01418	0,985823			0,3026722	0	1	
1871,08	4,79	0,26	11,89		67,98421	3,690166		70,20	0,033538	1,033538			0,3075138	0,015996	1,015996	
1871,09	4,84	0,26	12,18		67,08397	3,603685		69,58	-0,00882	0,991175			0,300307	-0,02344	0,976564	

real price	Annualised real div		Total real return	return	return + 1	GAR (01/1871 to 06/1999)		real div not annualised	return	return + 1	GAR (01/1871 to 12/1999)
=B3*$D$1551/D3	=C3*$D$1551/D3		60,1306420418752					=H3/12			
=B4*$D$1551/D4	=C4*$D$1551/D4		=J3*((G4 + H4/12)/(=J4/J3 -1		=K4+1	=GEOMEAN(I4:I154		=H4/12	=O4/O3 - 1	=P4+1	=GEOMEAN(Q4:Q15
=B5*$D$1551/D5	=C5*$D$1551/D5		=J4*((G5 + H5/12)/(=J5/J4 -1		=K5+1			=H5/12	=O5/O4 - 1	=P5+1	
=B6*$D$1551/D6	=C6*$D$1551/D6		=J5*((G6 + H6/12)/(=J6/J5 -1		=K6+1			=H6/12	=O6/O5 - 1	=P6+1	
=B7*$D$1551/D7	=C7*$D$1551/D7		=J6*((G7 + H7/12)/(=J7/J6 -1		=K7+1			=H7/12	=O7/O6 - 1	=P7+1	
=B8*$D$1551/D8	=C8*$D$1551/D8		=J7*((G8 + H8/12)/(=J8/J7 -1		=K8+1			=H8/12	=O8/O7 - 1	=P8+1	
=B9*$D$1551/D9	=C9*$D$1551/D9		=J8*((G9 + H9/12)/(=J9/J8 -1		=K9+1			=H9/12	=O9/O8 - 1	=P9+1	
=B10*$D$1551/D10	=C10*$D$1551/D10		=J9*((G10 + H10/12	=J10/J9 -1	=K10+1			=H10/12	=O10/O9 - 1	=P10+1	
=B11*$D$1551/D11	=C11*$D$1551/D11		=J10*((G11 + H11/1	=J11/J10 -1	=K11+1			=H11/12	=O11/O10 - 1	=P11+1	
=B12*$D$1551/D12	=C12*$D$1551/D12		=J11*((G12 + H12/1	=J12/J11 -1	=K12+1			=H12/12	=O12/O11 - 1	=P12+1	
=B13*$D$1551/D13	=C13*$D$1551/D13		=J12*((G13 + H13/1	=J13/J12 -1	=K13+1			=H13/12	=O13/O12 - 1	=P13+1	
=B14*$D$1551/D14	=C14*$D$1551/D14		=J13*((G14 + H14/1	=J14/J13 -1	=K14+1			=H14/12	=O14/O13 - 1	=P14+1	
=B15*$D$1551/D15	=C15*$D$1551/D15		=J14*((G15 + H15/1	=J15/J14 -1	=K15+1			=H15/12	=O15/O14 - 1	=P15+1	
=B16*$D$1551/D16	=C16*$D$1551/D16		=J15*((G16 + H16/1	=J16/J15 -1	=K16+1			=H16/12	=O16/O15 - 1	=P16+1	

Figure F.4: The top image shows a calculation snippet of using the monthly values to calculate the geometric average monthly real return from 01/1871 to 06/1999 - r and the geometric average monthly real growth rate of dividend from 01/1871 to 12/1999- g. The bottom image shows the formula behind the top image.

G Robert Shiller S&P 500 2015 time series model fits

Time Series model fits using g_{10}						
	θ_1	$\theta_1^{95\%}$	θ_2	$\theta_2^{95\%}$	θ_3	$\theta_3^{95\%}$
ARIMA(4,1,1)	0.903	(0.667,1.139)	-0.230	(-0.321,-0.137)	-0.015	(-0.080,0.051)
ARMA(1,1)	0.988	(0.980,0.995)	-	(-)	-	(-)
ARMA(2,2)	0.678	(0.215,1.140)	0.305	(-0.151,0.762)	-	(-)

Table G.1: Table G.1 shows the autoregressive terms parameters of the ARIMA(4,1,1), ARMA(1,1) and ARMA(2,2) together with the standard 95% confidence interval ($\theta \pm 1.96 s.e(\theta)$). Looking at θ_3 of the ARIMA(4,1,1) it appears to that the associated 95% confidence interval (-0.080,0.051) contains 0 (non-significant), which implies that there may be no association between the distortion and the θ_3 although if we consider removing this term then surely the ARIMA(3,1,1) will have a higher AICC than the ARIMA(4,1,1) which explains the distortion using a growth rate of the last years the best although the ARIMA(3,1,1) may be good enough but this area was not explored. (continue to Table G.2)

Time Series model fits for g_{10}		
	θ_4	$\theta_4^{95\%}$
ARIMA(4,1,1)	0.084	(0.036,0.131)
ARMA(1,1)	-	-
ARMA(2,2)	-	-

Table G.2: Remaining autoregressive terms. (continue to Table G.3)

Time Series model fits using g_{10}				
	ϕ_1	$\phi_1^{95\%}$	ϕ_2	$\phi_2^{95\%}$
ARIMA(4,1,1)	-0.622	(-0.855,-0.389)	-	(-)
ARMA(1,1)	0.276	(0.233,0.319)	-	(-)
ARMA(2,2)	0.603	(0.144,1.061)	0.141	(0.011,0.270)

Table G.3: Moving average terms.

Time Series model fits using g_{10}			
	μ	$\mu^{95\%}$	σ^2
ARIMA(4,1,1)	-	(-)	0.002
ARMA(1,1)	0.111	(-0.071,0.292)	0.002
ARMA(2,2)	0.113	(-0.063,0.288)	0.002

Table G.4: Mean and noise variance of time series models.

Time Series model fits using g_{10}			
	AIC	AICC	BIC
ARIMA(4,1,1)	-6321.93	-6321.88	-6289.17
ARMA(1,1)	-6319.284	-6319.261	-6297.437
ARMA(2,2)	-6320.44	-6320.39	-6287.67

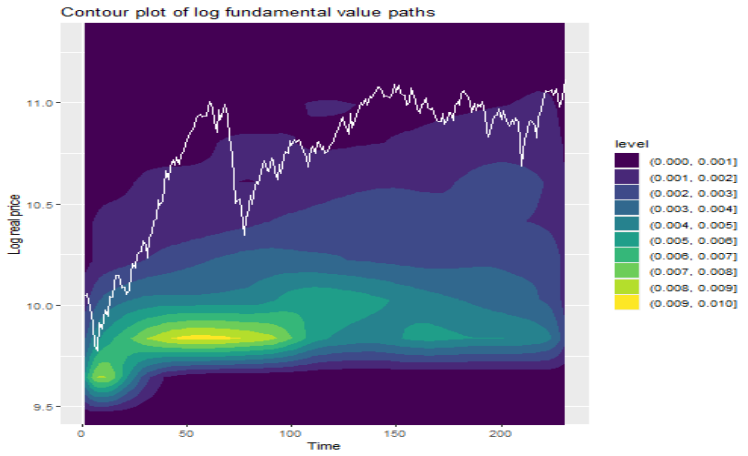
Table G.5: In Table G.5 shows the AIC, AICC and BIC for the ARIMA(4,1,1), ARMA(1,1) and ARMA(2,2). The ARIMA(4,1,1) were selected on the basis of the AICC since it is well known that the AICC converges to the AIC for large observations and the ARMA(1,1) and ARMA(2,2) via the maximum likelihood. We were concerned that the ARIMA(4,1,1) may be overfitting the model but we do not explore any of the residuals of the model and do not consider forecasting however we did consider fitting a model based on minimising the BIC and obviously this lead to a more parsimonious model which is a ARIMA(0,1,1). Nonetheless it appears from Table G.5 that the ARIMA(4,1,1) best explains the distortion using a growth rate of the last 10 years although its plausible for the ARMA(1,1) and the ARMA(2,2) to similarly model the distortion well looking at the proximity of the AICC and AICC values.

H Noise variance of ARIMA(0,1,0) for TOP40

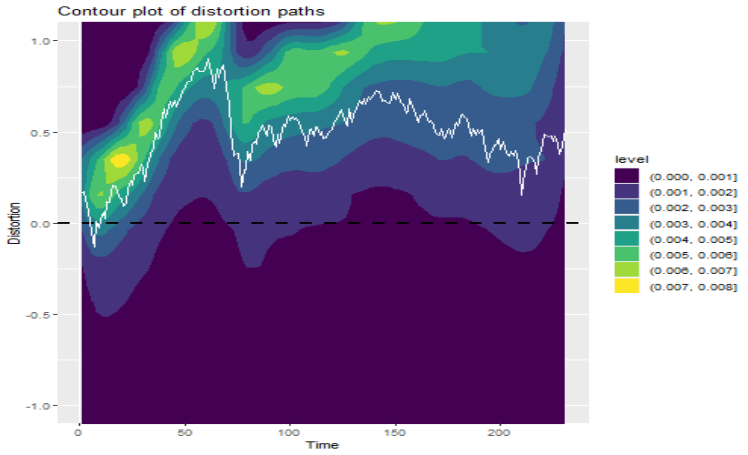
ARIMA(0,1,0)	
	σ^2
g_{10}	0.002295
g_{15}	0.002295
g_{ω}	0.002296
$g_{\min}$	0.002295

Table H.1: Noise variance of ARIMA(0,1,0) models in subsection 7.2 for the TOP 40 distortion when using a growth rate of the last 10, 15 years, over the whole date set and minimising the objective function.

I TOP 40 contour plots

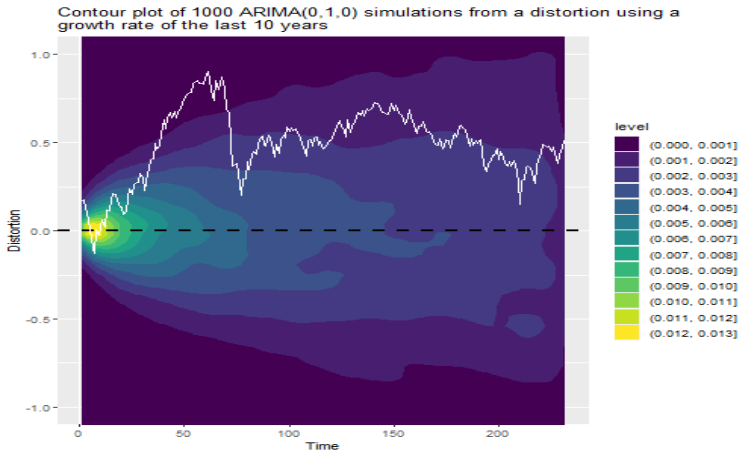


(a) Log fundamental value paths

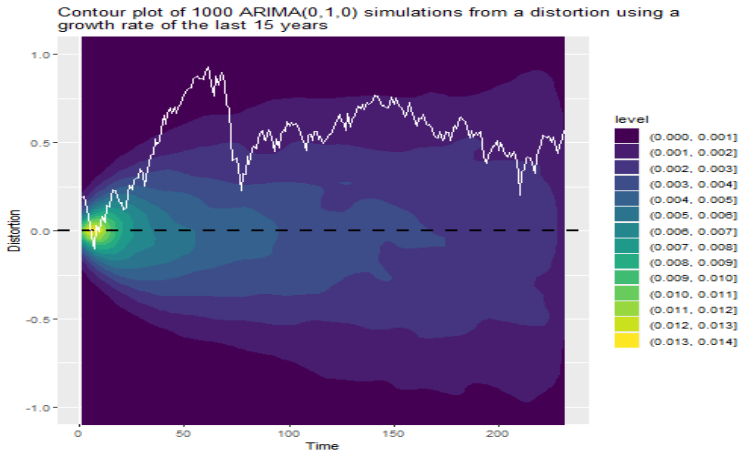


(b) Price distortion paths

Figure I.1: Figure I.1a shows the contours of the 1000 log fundamental paths for the TOP40 generated from a uniform distribution between $(-r,r)$. It can be seen that majority are situated close to the log real price. Figure I.1b shows the contours of 1000 different distortion paths obtained from the above. There appears to be a high number of distortions that give a similar shape to that of the log real price. Most of the distortions overvalue the market seen by the high density in Figure I.1b, due to those low growth rates being produced which led to most log fundamental value paths lying beneath the log real price. These are related to growth rates that undervalue or overvalue the market somewhat significantly. The white line in Figure I.1a shows the log fundamental time series using a growth rate of the last 10 years (g_{10}) and the white line in Figure I.1b shows the distortion path of using a g_{10}.

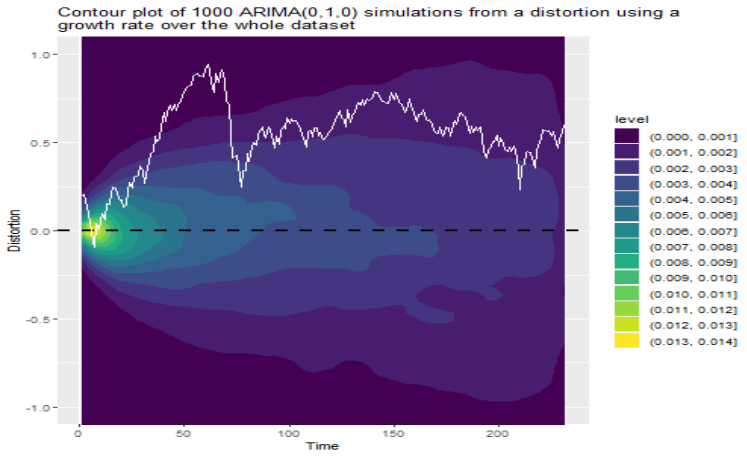


(a) Contour plot of 1000 simulations from a random walk fitted to a distortion using a growth rate of the last 10 years.

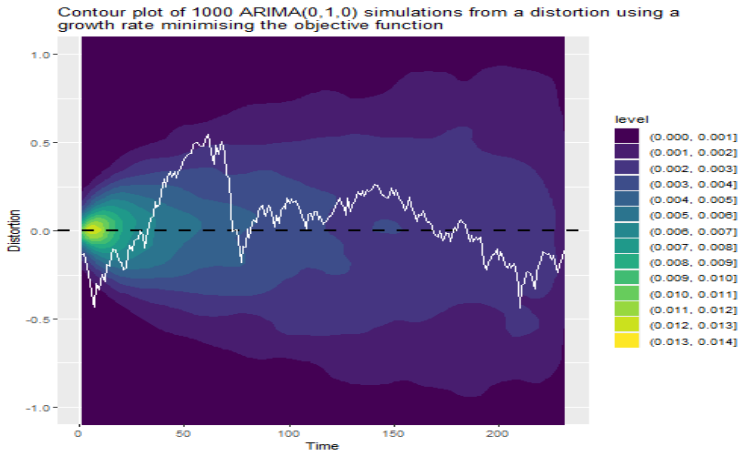


(b) Contour plot of 1000 simulations from a random walk fitted to a distortion using a growth rate of the last 15 years.

Figure I.2: Figure I.2a and Figure I.2b shows the contours of the 1000 distortions paths generated from their respected ARIMA(0,1,0) model. Both look fairly similar and the more time goes on some distortions appear to deviate more further from 0 than others. The white line in Figure I.2a shows the price distortion using a g_{10} and the white line in Figure I.2b shows the price distortion using a g_{15}.



(a) Contour plot of 1000 simulations from a random walk fitted to a distortion using a growth rate over the whole data set.



(b) Contour plot of 1000 simulations from a random walk fitted to a distortion using a growth rate minimising the objective function.

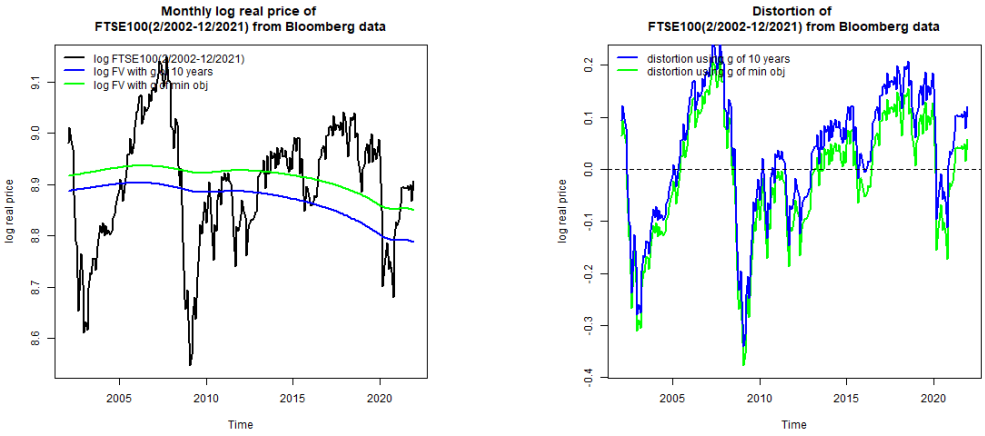
Figure I.3: Figure I.3a and Figure I.3b shows the contours of the 1000 distortions paths generated from their respected ARIMA(0,1,0) model. Both look fairly similar and the more time goes on the some distortion appears to deviate from 0 more further than others. These are related to growth rates that undervalue or overvalue the market somewhat significantly. The white line in Figure I.3a shows the price distortion using a g_ω and the white line in Figure I.3b shows the price distortion using a $g_{\min}$.

J Bloomberg markets

J.1 FTSE 100

Time period	r
02/2002-12/2020	0.28%
	g
01/2011-12/2020	-0.07%
Min obj. function	-0.05%

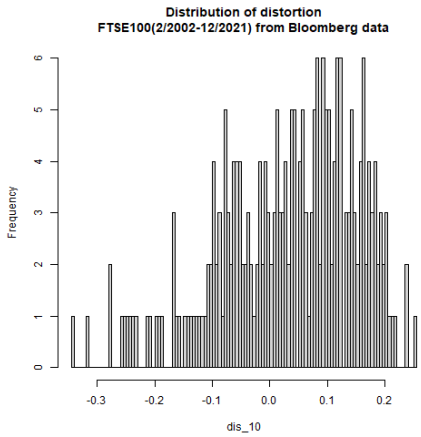
Table J.1: Table of the monthly discount and growth rate per month.



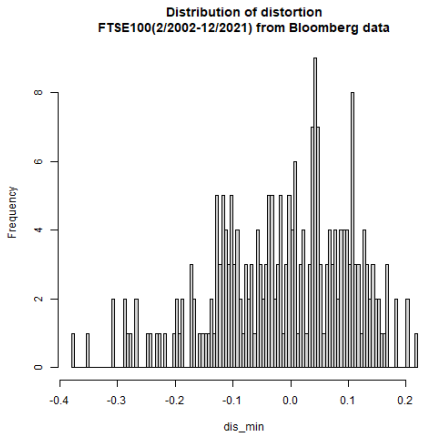
(a) Log fundamental value paths

(b) Price distortion paths

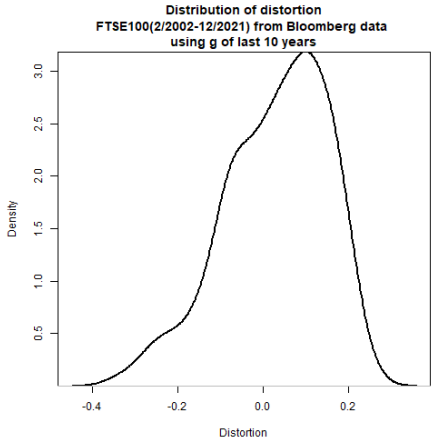
Figure J.1: In Figure J.1a, the monthly log fundamental value paths for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the FTSE 100 from 02/2002 to 12/2021 while keeping the discount rate r fixed as calculated in Table J.1. From Figure J.1a the log real price appears to cross the log fundamental value line sometimes and it appears to give that impression that it may be bimodal as it looks similar to the S&P 500 in Figure 3.2a but with no trend. In plot Figure J.1b shows the distortion of the FTSE 100 when using g_{10} and $g_{\min}$.



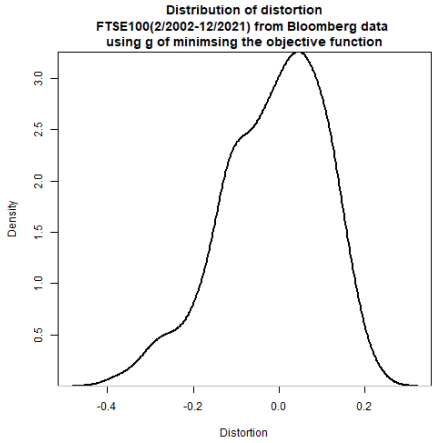
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using $g_{\min}$



(c) Smoothed density distortion using g_{10}



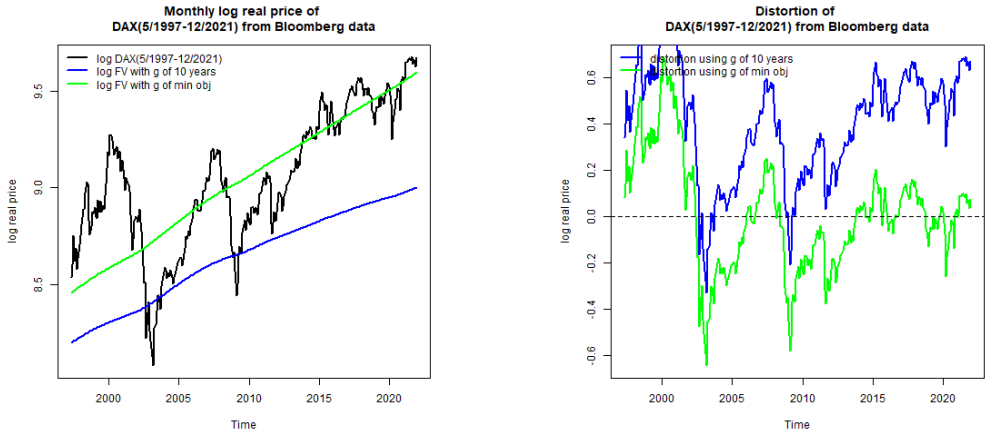
(d) Smoothed density distortion using $g_{\min}$

Figure J.2: From Figure J.2 shows that the FTSE 100 distortion when using a growth rate of the last 10 years and minimsing the objective function gives a more unimodal impression than bimodal which we previously though in Figure J.1.

J.2 DAX 40

Time period	r
05/1997-12/2020	0.61%
	g
01/2011-12/2020	0.28%
Min obj. function	0.43%

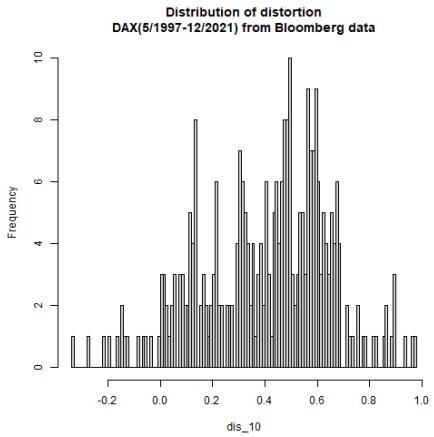
Table J.2: Table of the monthly discount and growth rate per month.



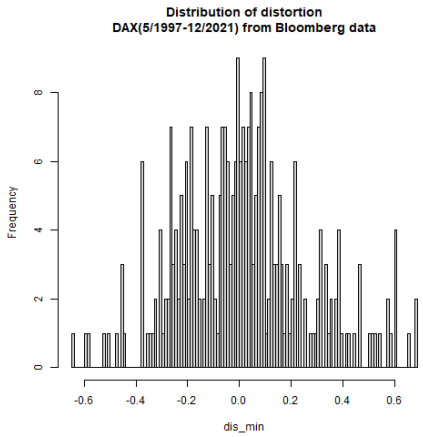
(a) Log fundamental value paths

(b) Price distortion paths

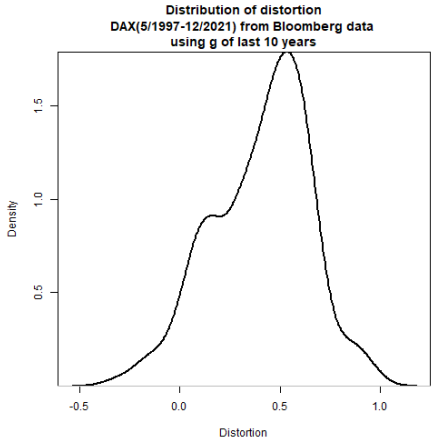
Figure J.3: In Figure J.3a, the monthly log fundamental value paths for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the DAX 40 05/1997 to 12/2021 while keeping the discount rate r fixed as calculated in Table J.2. In Figure J.3b shows the distortion of the DAX 40 when using g_{10} and $g_{\min}$.



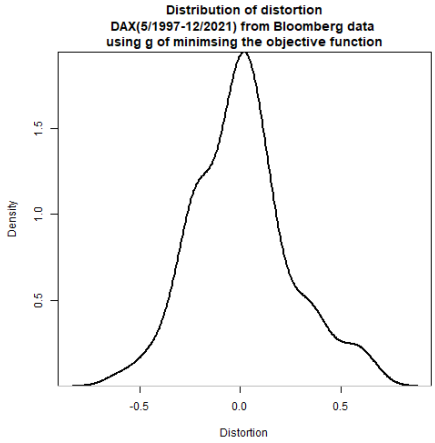
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using $g_{\min}$



(c) Smoothed density distortion using g_{10}



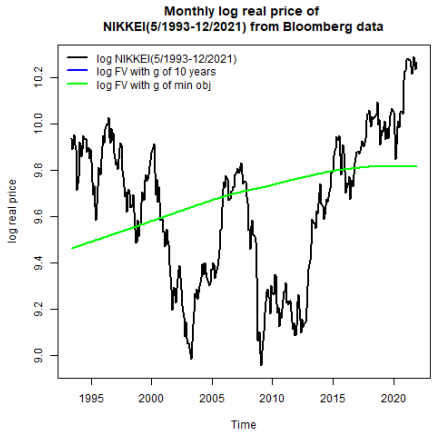
(d) Smoothed density distortion using $g_{\min}$

Figure J.4: From Figure J.4 shows that the density of the DAX 40 distortion when using a growth rate of the last 10 years and minimising the objective function gives a questionable bimodal impression.

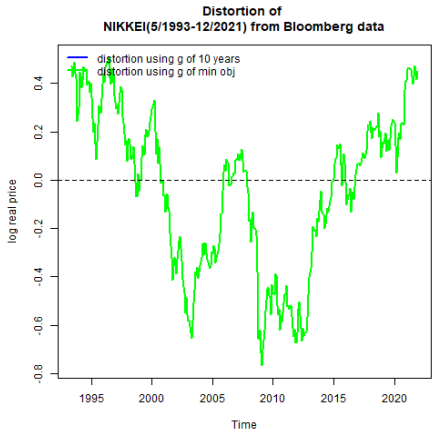
J.3 NIKKEI

Time period	r
05/1993-12/2020	0.20%
	g
01/2011-12/2020	0.45%
Min obj. function	-0.02%

Table J.3: Table of the monthly discount and growth rate per month.

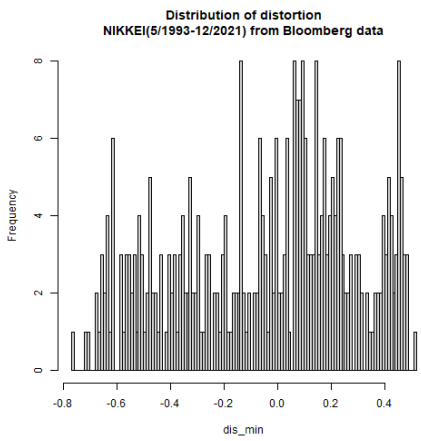


(a) Log fundamental value paths

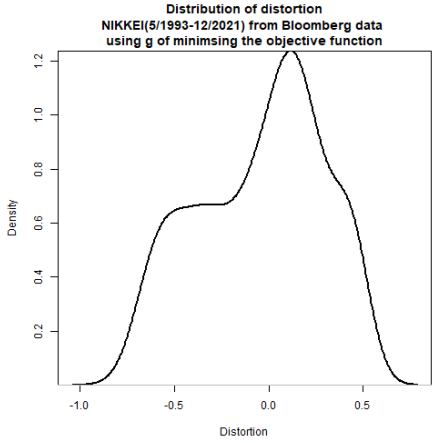


(b) Price distortion paths

Figure J.5: In Figure J.5a, the monthly log fundamental values for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the NIKKEI 05/1993 to 12/2021 while keeping the discount rate r fixed as calculated in Table J.3. In Figure J.5b shows the distortion of the NIKKEI when using g_{10} and $g_{\min}$. The growth rate of the last 10 years give a growth rate that is greater than the discount rate which makes using the Gordan growth model invalid since a growth rate less than the discount rate is required.



(a) Distribution of distortion using $g_{\min}$.



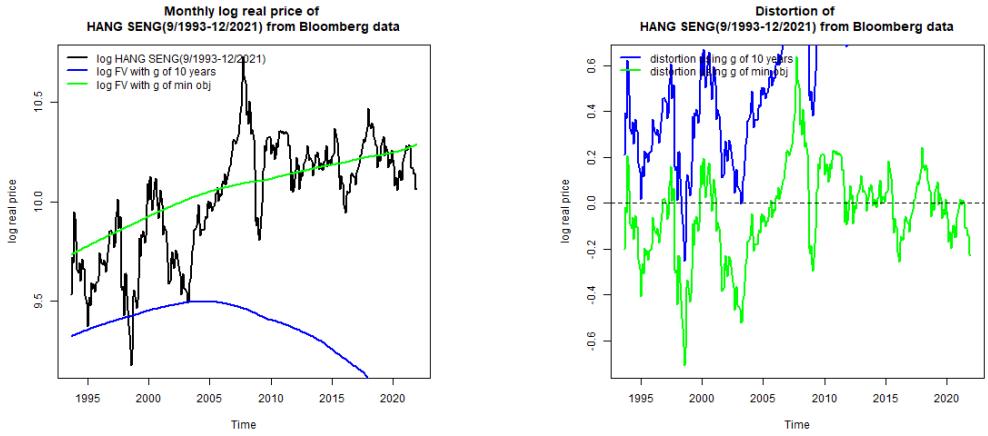
(b) Smoothed density distortion using $g_{\min}$.

Figure J.6: From Figure J.4 shows that the density of the NIKKEI distortion when using a growth rate of the last 10 years and minimIsing the objective function gives a questionable bimodal impression as well.

J.4 HANG SENG

Time period	r
09/1993-12/2020	0.40%
	g
01/2011-12/2020	-0.31%
Min obj. function	0.24%

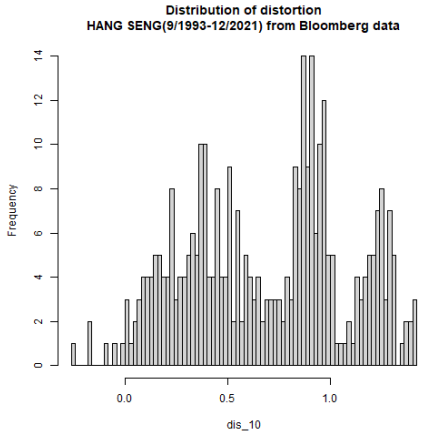
Table J.4: Table of the monthly discount and growth rate per month.



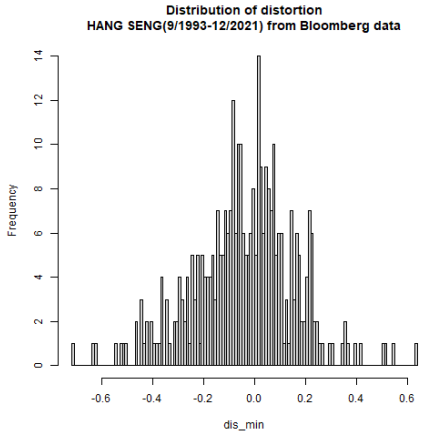
(a) Log fundamental value paths

(b) Price distortion paths

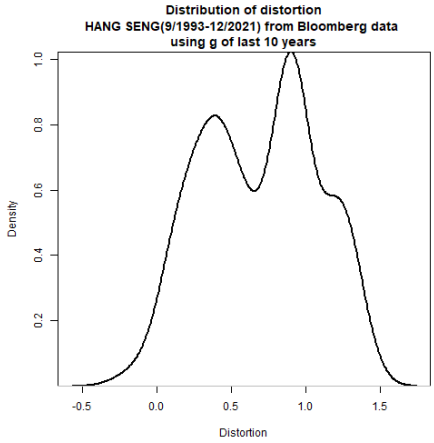
Figure J.7: In Figure J.7a, the monthly log fundamental values for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the HANG SENG from 09/1993 to 12/2021 while keeping the discount rate r fixed as calculated in Table J.4. In Figure J.7b shows the distortion of the HANG SENG when using g_{10} and $g_{\min}$. We see that the growth rate of the last 10 years just takes a dip somewhere around 2005 this leads to the market being extremely over valued.



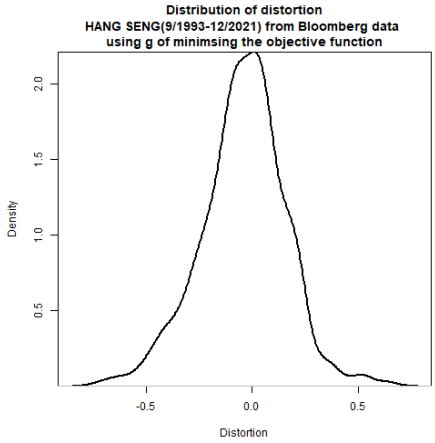
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using $g_{\min}$



(c) Smoothed density distortion using g_{10}



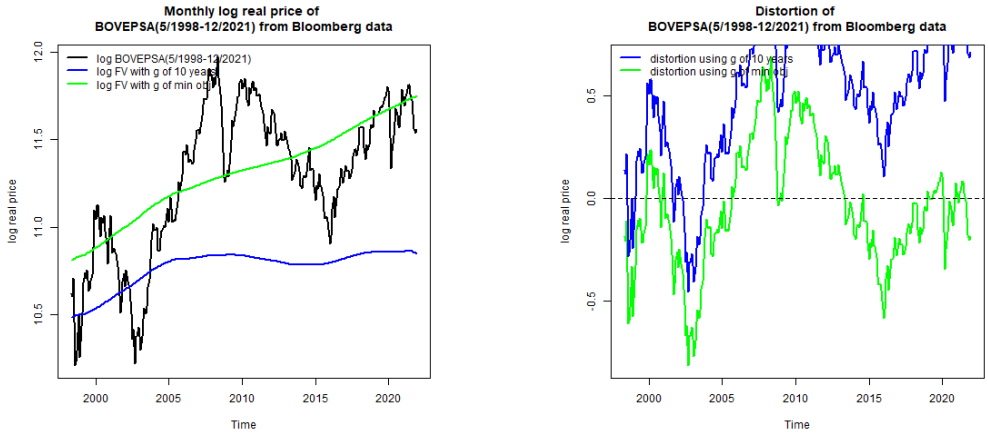
(d) Smoothed density distortion using $g_{\min}$

Figure J.8: From Figure J.8a shows that the density of the HANG SENG distortion when using a growth rate of the last 10 years and minimising the objective function gives an bimodal impression and a unimodal impression respectively.

J.5 BOVEPSA

Time period	r
05/1998-12/2020	0.60%
	g
01/2011-12/2020	-0.81%
Min obj. function	0.02%

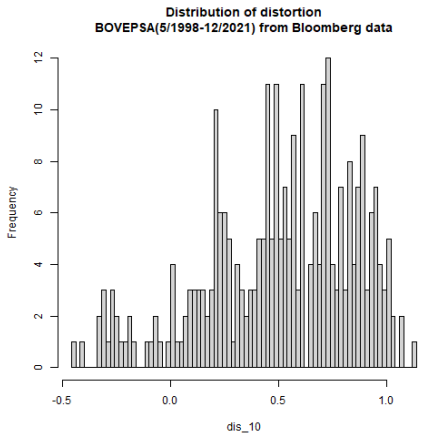
Table J.5: Table of the monthly discount and growth rate per month.



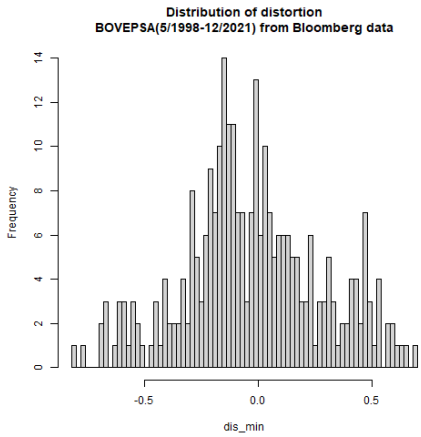
(a) Log fundamental value paths

(b) Price distortion paths

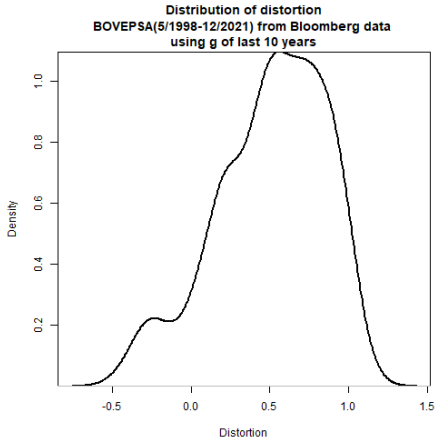
Figure J.9: In Figure J.9a, the monthly log fundamental values for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the BOVEPSA from 05/1998 to 12/2020 while keeping the discount rate r fixed as calculated in Table J.5. From Figure J.9a, we see that when using the growth rate of the last 10 years, the log price is consistently (From 2005) above the log fundamental value line. But when using $g_{\min}$, it fluctuates around the fundamental value. Figure J.9b shows the price distortion of the BOVEPSA when using both g_{10} and $g_{\min}$.



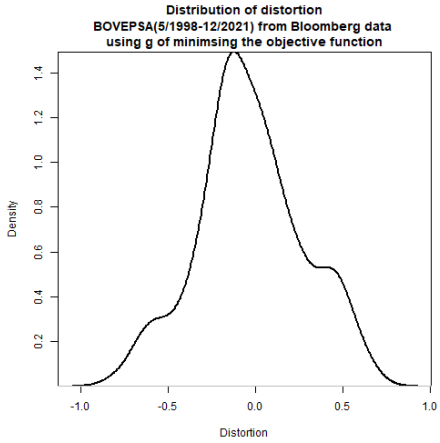
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using $g_{\min}$



(c) Smoothed density distortion using g_{10}



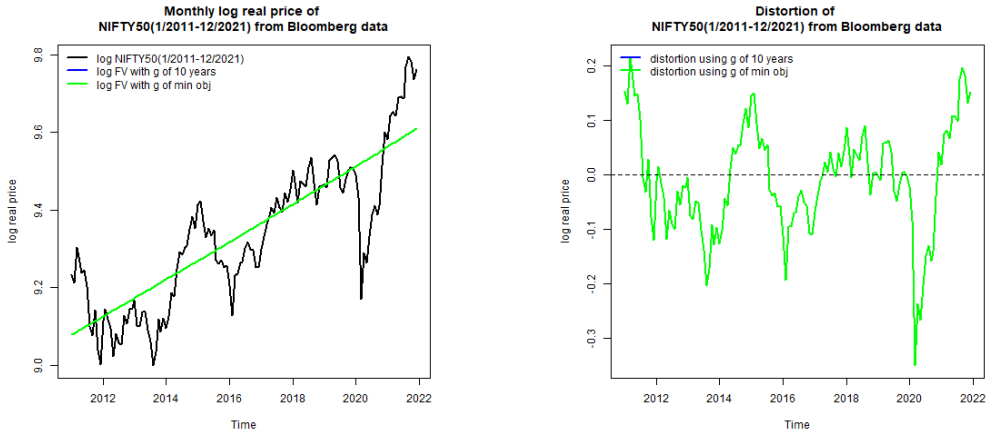
(d) Smoothed density distortion using $g_{\min}$

Figure J.10: From Figure J.10 shows that the density of the BOVEPSA price distortion when using a growth rate of the last 10 years and one minimising the objective function. Neither growth rates give a visual strong bimodal impression.

J.6 NIFTY 50

Time period	r
04/2000-12/2020	0.52%
	g
01/2011-12/2020	-0.02%
Min obj. function	0.40%

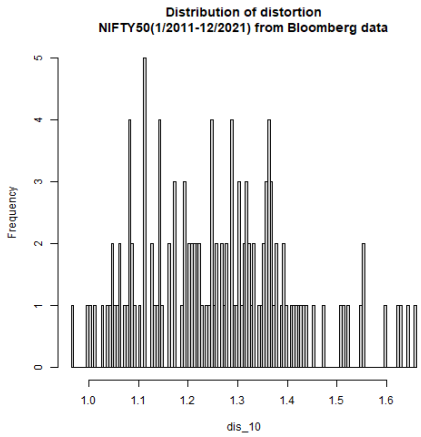
Table J.6: Table of the monthly discount and growth rate per month.



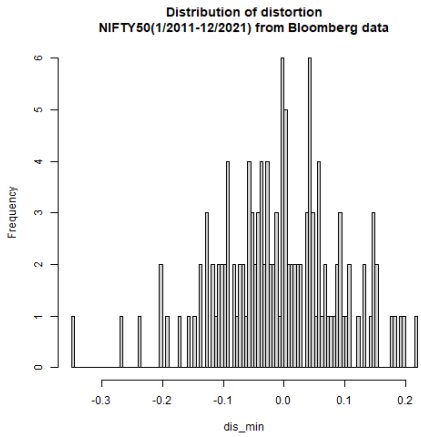
(a) Log fundamental values.

(b) Price distortion.

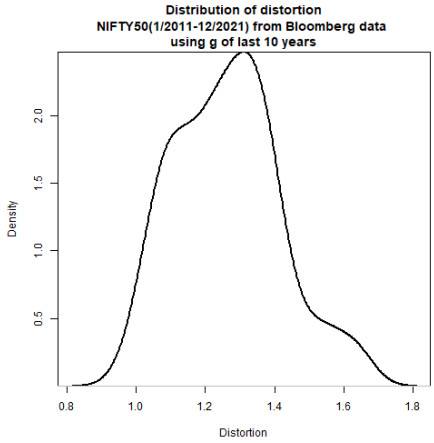
Figure J.11: In Figure J.11a, the monthly log fundamental values for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the NIFTY 50 from 04/2000 to 12/2020 while keeping the discount rate r fixed as calculated in Table J.7. From Figure J.11a, we see that when using the growth rate of the last 10 years, the log price is consistently above the log fundamental value path. But when using $g_{\min}$, the log price has extended periods where it is above and below the fundamental value. Plot Figure J.9b shows the price distortion of the NIFTY 50 when using both g_{10} and $g_{\min}$. The NIFTY 50 distortion when using g_{10} does not appear due to being cut off because it over values the market extremely but one can expect it to have a high positive distortion.



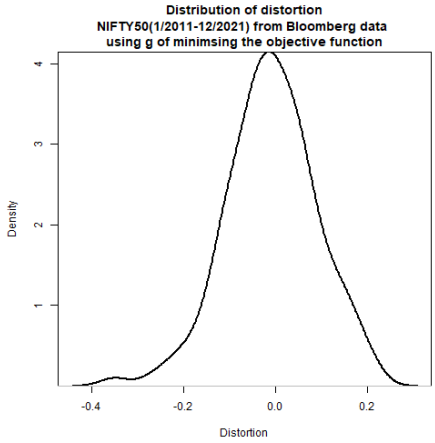
(a) Distribution of distortion using g_{10}.



(b) Distribution of distortion using $g_{\min}$.



(c) Smoothed density distortion using g_{10}.



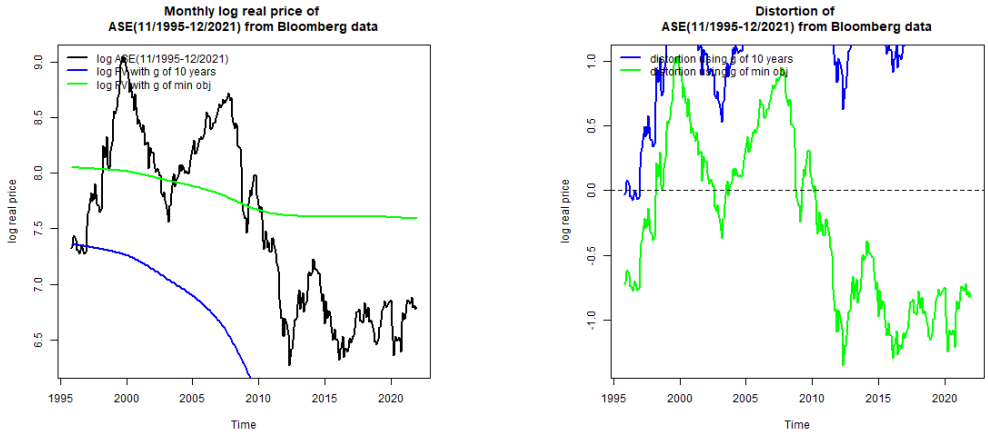
(d) Smoothed density distortion using $g_{\min}$.

Figure J.12: From Figure J.12 shows that the density of the NIFTY 50 price distortion when using a growth rate of the last 10 years and one minimising the objective function. Neither appear to give a visual bimodal distribution.

J.7 ASE

Time period	r
11/1995-12/2020	0.06%
	g
01/2011-12/2020	-1.39%
Min obj. function	-0.03%

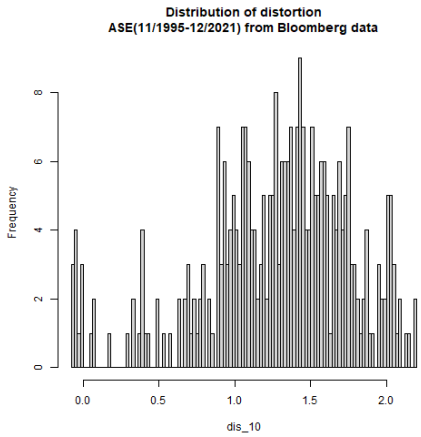
Table J.7: Table of the monthly discount and growth rate per month.



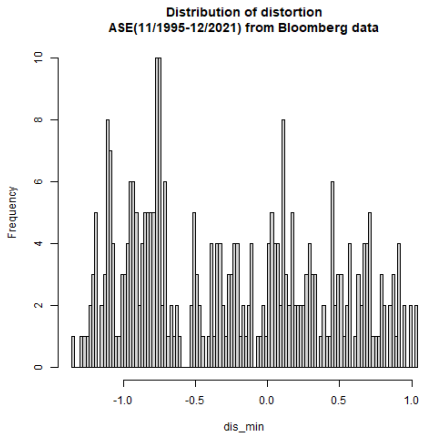
(a) Log fundamental value paths

(b) Price distortion paths

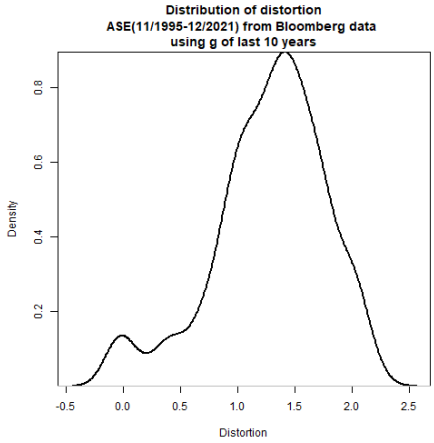
Figure J.13: In Figure J.13a, the monthly log fundamental values for using a g of the last 10 years and when minimising the objective function (1) is displayed together with the monthly log real price of the ASE from 11/1995 to 12/2020 while keeping the discount rate r fixed as calculated in Table J.7. From Figure J.13a, we see that when using the growth rate of the last 10 years, the log price is consistently above the log fundamental value line. But when using $g_{\min}$, the log price has extended periods where it is above and below the fundamental value. Figure J.13b shows the price distortion of the ASE when using both g_{10} and $g_{\min}$.



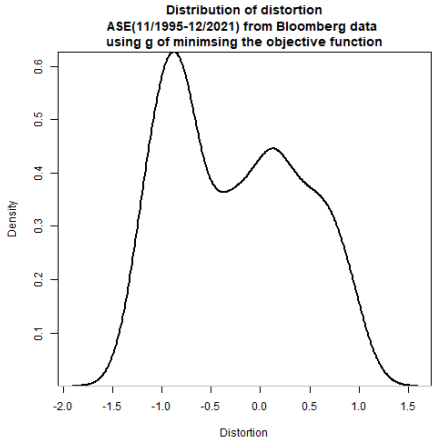
(a) Distribution of distortion using g_{10}



(b) Distribution of distortion using $g_{\min}$



(c) Smoothed density distortion using g_{10}



(d) Smoothed density distortion using $g_{\min}$

Figure J.14: From Figure J.14 shows that the density of the ASE price distortion when using a growth rate of the last 10 years and one minimising the objective function. $g_{\min}$ appears to give a questionable visual bimodal distribution, where as g_{10} does not.

K ARMA(1,1) models for Bloomberg markets

K.1 TOP 40 ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu_1^{95\%}$	σ^2
g_{10}	0.973	(0.945,1.002)	-0.026	(-0.148,0.097)	0.455	(0.257,0.653)	0.002
g_{15}	0.974	(0.946,1.002)	-0.026	(-0.149,0.097)	0.491	(0.289,0.694)	0.002
g_{ω}	0.974	(0.947,1.001)	-0.026	(-0.148,0.097)	0.007	(-0.197,0.211)	0.002
$g_{\min}$	0.975	(0.946,1.003)	-0.026	(-0.149,0.097)	0.513	(0.307,0.719)	0.002

Fit stat	AIC	AICC	BIC
g_{10}	-740.7	-740.59	-727
g_{15}	-740.6	-740.42	-726.83
g_{ω}	-740.81	-740.64	-727.04
$g_{\min}$	-740.48	-740.3	-726.71

Table K.1: Table K.1 shows the ARMA(1,1) model parameter fits of the Bloomberg TOP 40 distortion using a growth rate of the last 10, 15 years, over the whole data set and minimising the objective function together with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Focusing more on their AIC, AICC and BIC fit statistics rather than the model parameters, the ARMA(1,1) appears to model the distortion using g_{10}, g_{15}, g_{ω} and $g_{\min}$ all fairly the same and when compared to Table 7.7, Table 7.10, Table 7.13, Table 7.16 they differ slightly but could possibly model the distortion. See Figure L.1 for residual plots of ARMA(1,1) fit to the distortion when using a growth rate of the last 10 years and minimising the objective function.

K.2 S&P 500 ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.987	(0.975,1)	0.038	(-0.046,0.123)	-0.044	(-0.301,0.213)	0.002
$g_{\min}$	0.987	(0.975,1)	0.038	(-0.046,0.122)	0.032	(-0.223,0.287)	0.002

Fit stat	AIC	AICC	BIC
g_{10}	-2074.1	-2074	-2056.38
$g_{\min}$	-2074.1	-2074.1	-2056.44

Table K.2: We consider fitting a ARMA(1,1) model to the Bloomberg S&P 500 distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). It does not appear to model the data as best as in Table 7.3 although with the Robert Shiller data from 1871 to 2015 there is a increased number of observations nearly triple the observations itself are not exactly the same but good enough but still this should not be compared to Table 7.3. In Figure L.2 residuals appear to be have well with low correlation among each other and appear to be centered around 0. Although taking a look at the distribution both residuals plots appear to be somewhat left skewed but may follow a normal distribution.

K.3 FTSE 100 ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.938	(0.894,0.983)	0.050	(-0.089,0.189)	0.038	(-0.042,0.118)	0.002
$g_{\min}$	0.941	(0.895,0.986)	0.055	(-0.084,0.194)	-0.443	(-0.526,-0.360)	0.002

Fit stat	AIC	AICC	BIC
g_{10}	-860.58	-860.41	-846.67
$g_{\min}$	-859.27	-859.09	-845.36

Table K.3: We consider fitting a ARMA(1,1) model to the Bloomberg FTSE 100 distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the FTSE 100 using a growth rate of the last 10 years and minimising the objective function are at Figure L.3.

K.4 DAX 40 ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.963	(0.932,0.993)	0.054	(-0.064,0.173)	0.423	(0.239,0.606)	0.004
$g_{\min}$	0.015	(0.931,0.991)	0.060	(-0.064,0.174)	0.423	(-0.160,0.194)	0.004

Fit stat	AIC	AICC	BIC
g_{10}	-796.9	-796.76	-782.14
$g_{\min}$	-797.43	-797.3	-782.67

Table K.4: We consider fitting a ARMA(1,1) model to the Bloomberg DAX 40 distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the DAX 40 using a growth rate of the last 10 years and minimising the objective function are at Figure L.4.

K.5 NIKKEI ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	-	(-)	-	(-)	-	(-)	-
$g_{\min}$	0.985	(0.968,1.003)	0.010	(-0.010,0.209)	0.092	(-0.296,0.481)	0.003

Fit stat	AIC	AICC	BIC
g_{10}	-	-	-
$g_{\min}$	-993.51	-993.4	-978.15

Table K.5: We consider fitting a ARMA(1,1) model to the Bloomberg NIKKEI distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the NIKKEI using a growth rate of the last 10 years and minimising the objective function are at Figure L.5.

K.6 HANG SENG ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.984	(0.965,1.003)	0.048	(-0.057,0.152)	0.703	(0.288,1.118)	0.005
$g_{\min}$	0.930	(0.890,0.970)	0.070	(-0.037,0.176)	-0.054	(-0.161,0.053)	0.005

Fit stat	AIC	AICC	BIC
g_{10}	-835.67	-835.55	-820.36
$g_{\min}$	-846.95	-846.83	-831.63

Table K.6: We consider fitting a ARMA(1,1) model to the Bloomberg HANG SENG distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the HANG SENG using a growth rate of the last 10 years and minimising the objective function are at Figure L.6.

K.7 BOVEPSA ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.967	(0.939,0.996)	0.060	(-0.061,0.181)	0.496	(0.217,0.775)	0.007
$g_{\min}$	0.956	(0.923,0.989)	0.065	(-0.056,0.186)	-0.046	(-0.260,0.165)	0.007

Fit stat	AIC	AICC	BIC
g_{10}	-608.57	-608.42	-593.97
$g_{\min}$	-610.6	-610.45	-596

Table K.7: We consider fitting a ARMA(1,1) model to the Bloomberg BOVEPSA distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the BOVEPSA using a growth rate of the last 10 years and minimising the objective function are at Figure L.7.

K.8 NIFTY 50 ARMA(1,1) fit

ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.960	(0.908,1.012)	-0.039	(-0.231,0.153)	1.309	(1.130,1.490)	0.002
$g_{\min}$	0.881	(0.784,0.977)	0.002	(-0.202,0.207)	0.008	(-0.059,0.075)	0.002

Fit stat	AIC	AICC	BIC
g_{10}	-409.23	-408.91	-397.7
$g_{\min}$	-415.61	-415.29	-404.08

Table K.8: We consider fitting a ARMA(1,1) model to the Bloomberg NIFTY 50 distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the NIFTY 50 using a growth rate of the last 10 years and minimising the objective function are at Figure L.9.

K.9 ASE ARMA(1,1) fit

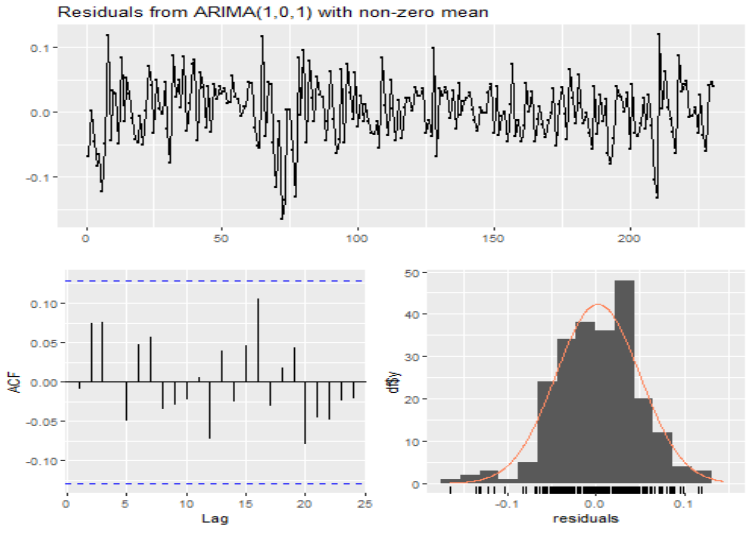
ARMA(1,1)							
	θ_1	$\theta_1^{95\%}$	ϕ_1	$\phi_1^{95\%}$	μ	$\mu^{95\%}$	σ^2
g_{10}	0.989	(0.970,1.007)	0.087	(-0.031,0.206)	1.204	(0.425,1.982)	0.008
$g_{\min}$	0.987	(0.972,1.002)	0.091	(-0.207,0.208)	-0.414	(-1.124,0.296)	0.008

Fit stat	AIC	AICC	BIC
g_{10}	-599.83	-599.7	-584.83
$g_{\min}$	-599.43	-599.3	-584.44

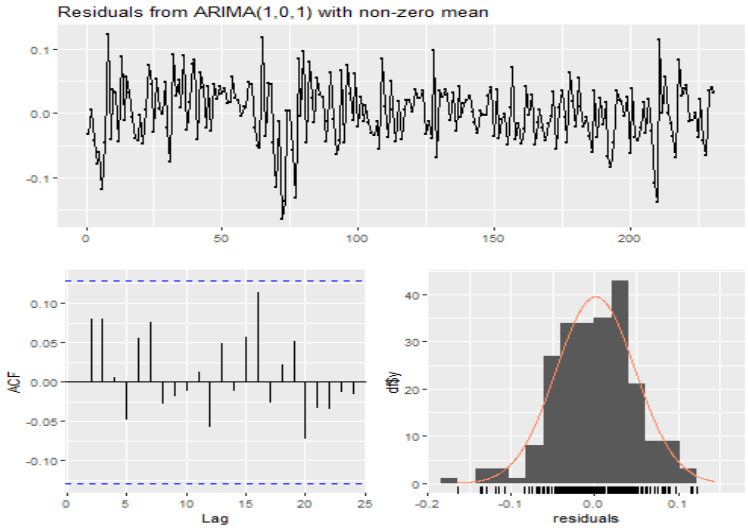
Table K.9: We consider fitting a ARMA(1,1) model to the Bloomberg ASE distortions using a growth rate of the last 10 years and minimising the objective function with the standard 95% confidence interval (i.e. $\theta_1^{95\%} = \theta_1 \pm 1.96 s.e(\theta_1)$). Residuals for the distortion of the ASE using a growth rate of the last 10 years and minimising the objective function are at Figure L.9.

L ARMA(1,1) residuals for Bloomberg markets

L.1 TOP 40



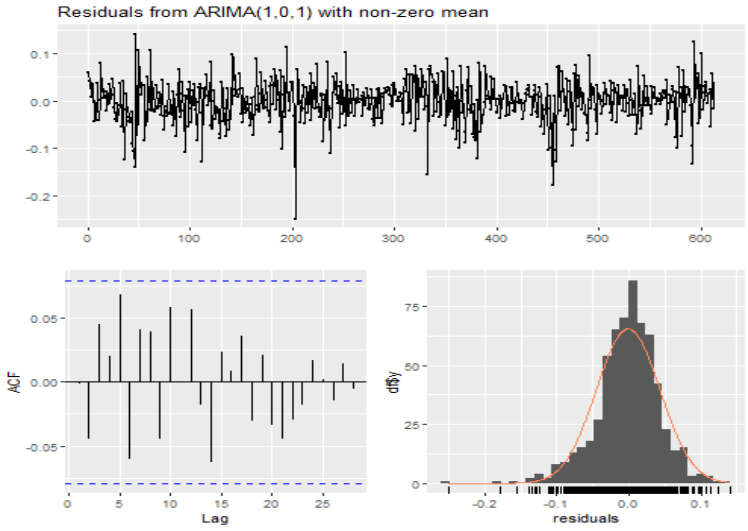
(a) using g_{10}



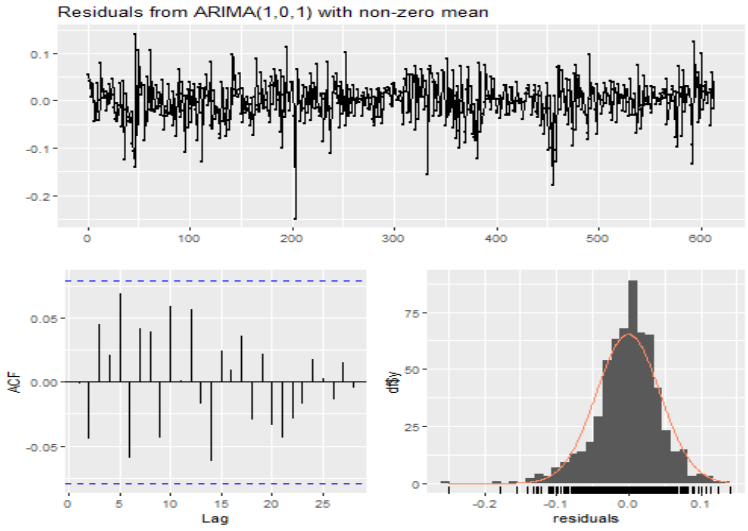
(b) using $g_{\min}$

Figure L.1: Residuals of the ARIMA(1,0,1) fitted to the TOP 40 distortion using a growth rate of the last 10 years and minimising the objective function.

L.2 S&P 500



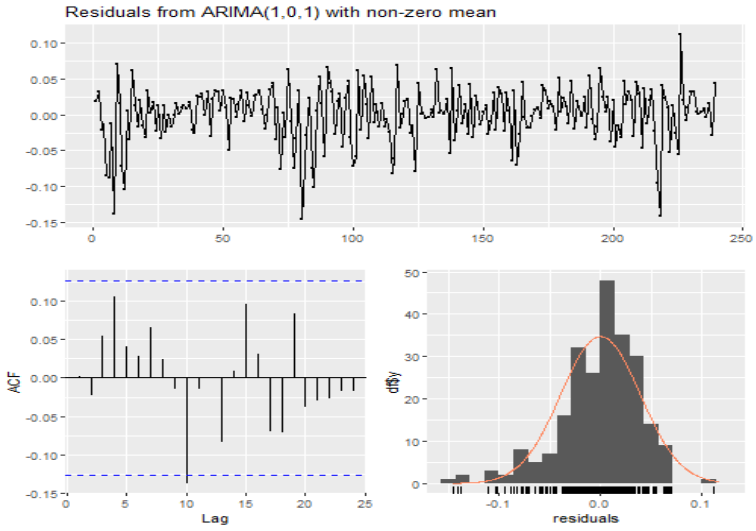
(a) using g_{10}



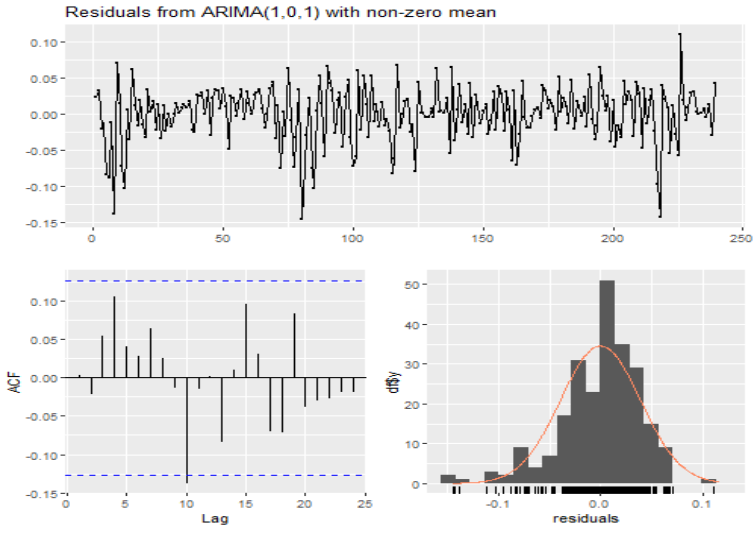
(b) using $g_{\min}$

Figure L.2: Residuals of the ARIMA(1,0,1) fitted to the S&P 500 distortion using a growth rate of the last 10 years and minimising the objective function.

L.3 FTSE 100



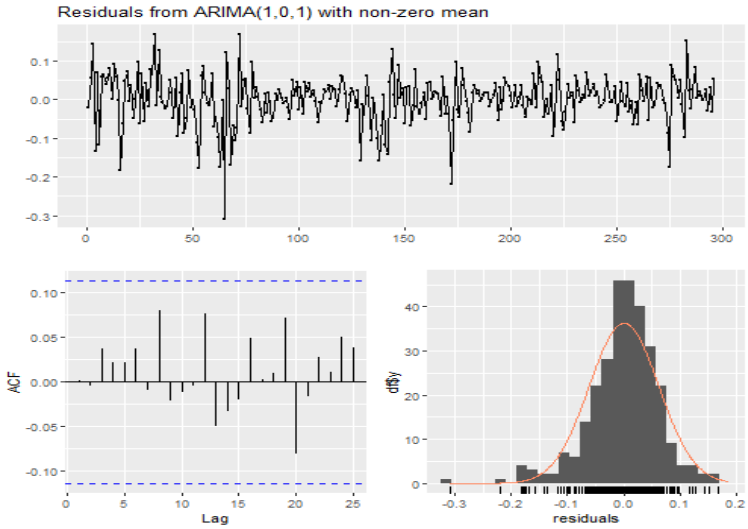
(a) using g_{10}



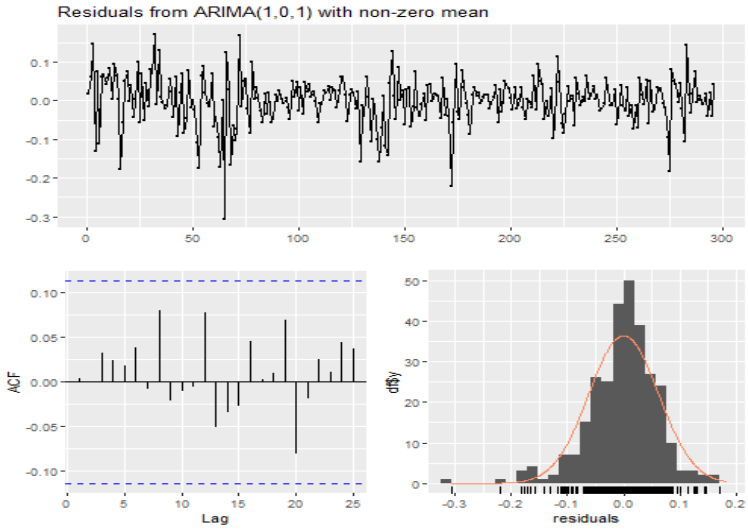
(b) using $g_{\min}$

Figure L.3: Residuals of the ARIMA(1,0,1) fitted to the FTSE 100 distortion using a growth rate of the last 10 years and minimising the objective function.

L.4 DAX 40



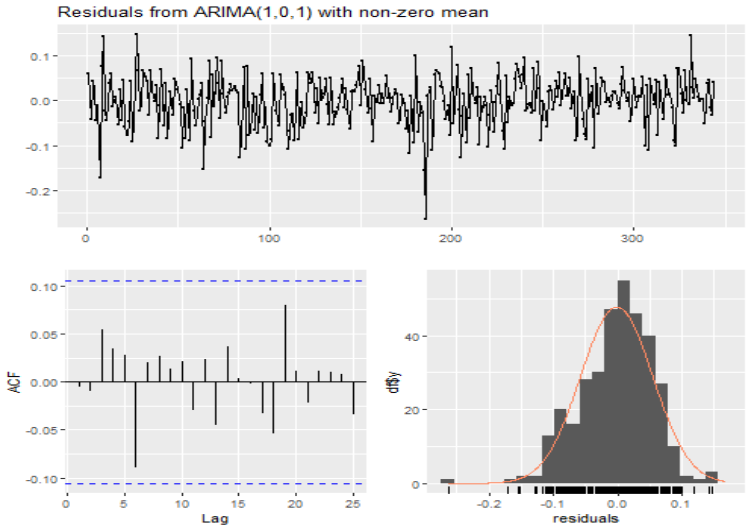
(a) using g_{10}



(b) using $g_{\min}$

Figure L.4: Residuals of the ARIMA(1,0,1) fitted to the DAX distortion using a growth rate of the last 10 years and minimising the objective function.

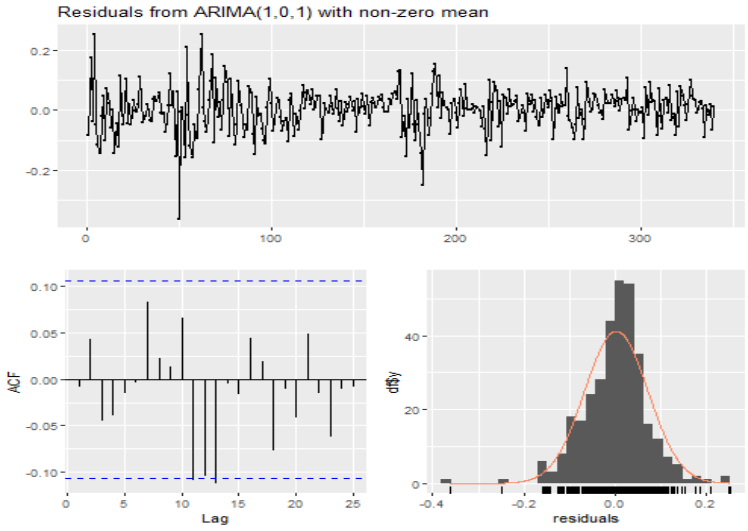
L.5 NIKKEI



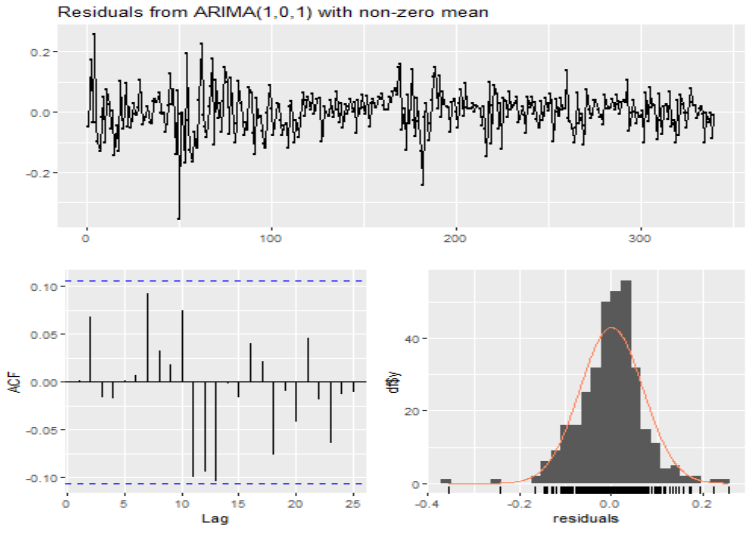
(a) using $g_{\min}$

Figure L.5: Residuals of the ARIMA(1,0,1) fitted to the NIKKEI distortion using a growth rate minimising the objective function.

L.6 HANG SENG



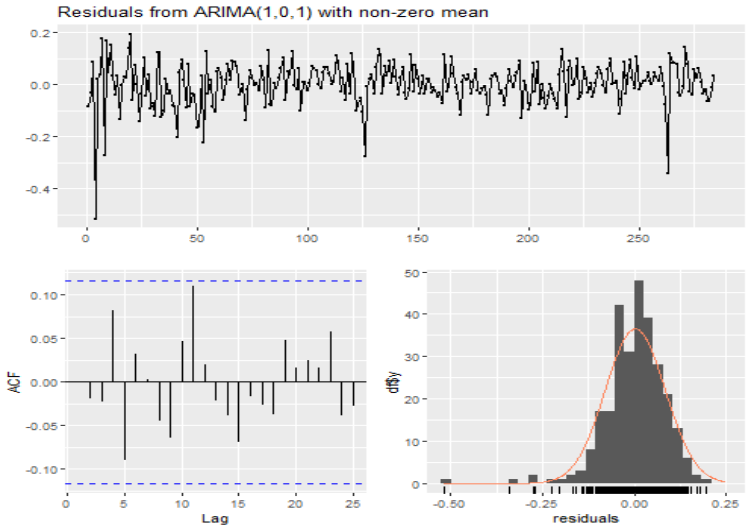
(a) using g_{10}



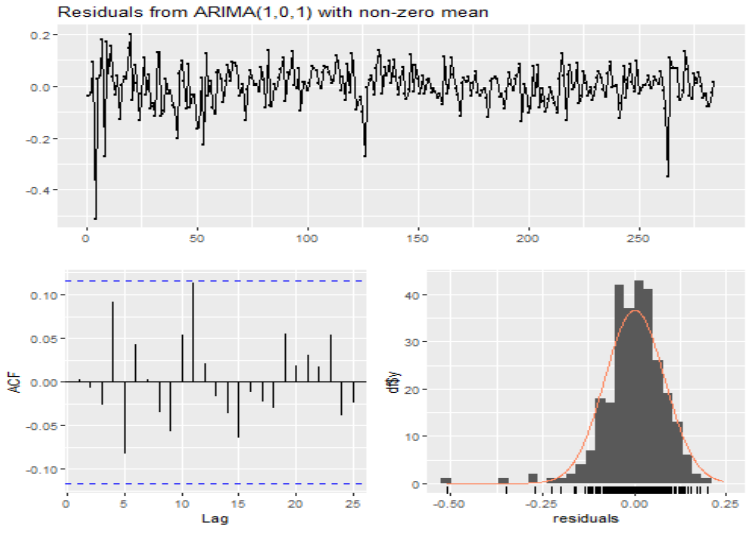
(b) using $g_{\min}$

Figure L.6: Residuals of the ARIMA(1,0,1) fitted to the HANG SENG distortion using a growth rate of the last 10 years and minimising the objective function.

L.7 BOVEPSA



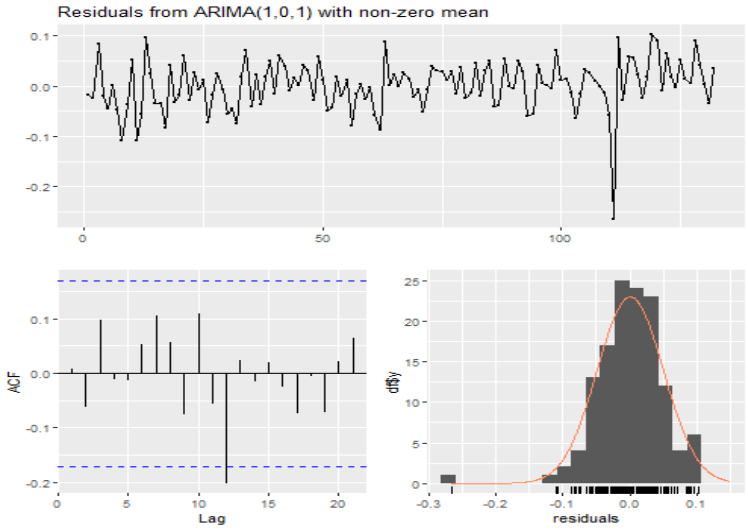
(a) using g_{10}



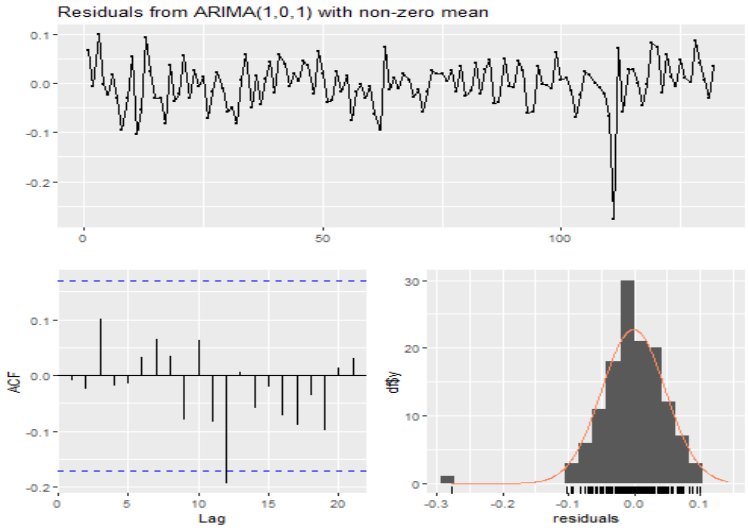
(b) using $g_{\min}$

Figure L.7: Residuals of the ARIMA(1,0,1) fitted to the BOVEPSA distortion using a growth rate of the last 10 years and minimising the objective function.

L.8 NIFTY 50



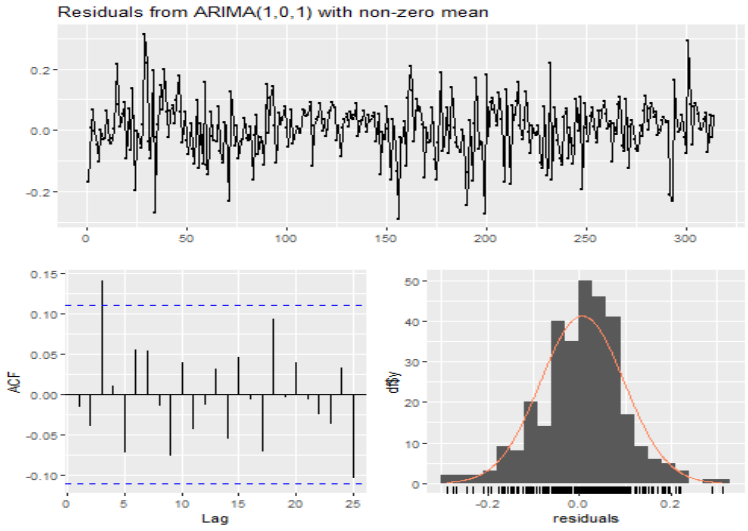
(a) using g_{10}



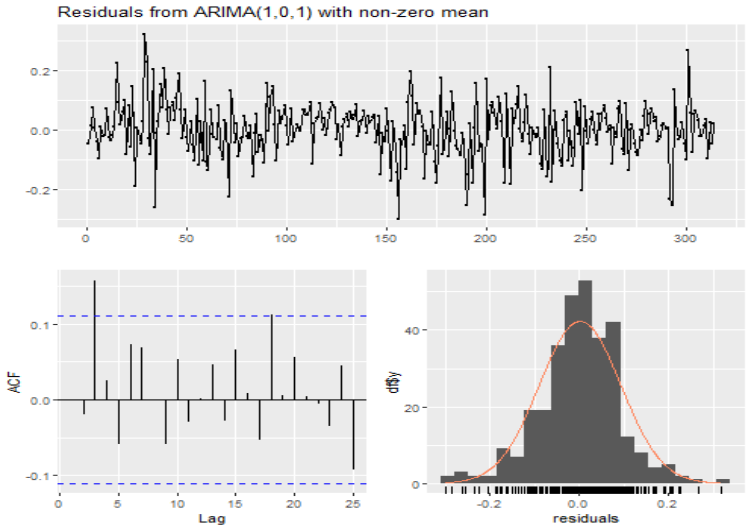
(b) using $g_{\min}$

Figure L.8: Residuals of the ARIMA(1,0,1) fitted to the NIFTY 50 distortion using a growth rate of the last 10 years and minimising the objective function.

L.9 ASE



(a) using g_{10}



(b) using $g_{\min}$

Figure L.9: Residuals of the ARIMA(1,0,1) fitted to the ASE distortion using a growth rate of the last 10 years and minimising the objective function.

M Agent-Based Models Algorithms

Algorithm 1 Gaunersdorfer and Hommes Agent-Based Model Simulation.

$N = 38000$
$t \leftarrow 3$
$y_t = \overline{y}$
$p_1, p_2 = 1000$
$E_{C,1}, E_{C,2} = 1000$
$E_{F,1}, E_{F,2} = 1000$
$z_{C,1}, z_{C,2} = 0$
$z_{F,1}, z_{F,2} = 0$
$U_{C,1}, U_{C,2} = 0$
$U_{F,1}, U_{F,2} = 0$

$\overline{y} = 1$
$\sigma_\delta = 0$
$r = 0.001$
$\nu = 1$
$g = 1.89$
$a\sigma^2 = 1$
$\sigma_\epsilon = 10$
$\eta = 0$
$\alpha = 2000$
$\beta = 2$

for $t \leq N$ **do**

$$E_{C,t}[p_{t+1}] = p_{t-1} + g(p_{t-1} - p_{t-2})$$

$$E_{F,t}[p_{t+1}] = p^* + \nu(p_{t-1} - p^*)$$

$$n_{C,t} = \frac{\exp[\beta U_{C,t-1}]}{\exp[\beta U_{C,t-1}] + \exp[\beta U_{F,t-1}]} \times \exp\left[- \frac{(p^* - p_{t-1})^2}{\alpha} \right]$$

$$n_{F,t} = 1 - n_{C,t}$$

$$\epsilon_t \sim N(0, \sigma_\epsilon)$$

$$p_t = \frac{1}{(1+r)} \left[n_{C,t} E_{C,t}[p_{t+1}] + n_{F,t} E_{F,t}[p_{t+1}] + \overline{y} \right] + \epsilon_t$$

$$z_{C,t} = \frac{E_{C,t}[p_{t+1}] + \overline{y} - (1+r)p_t}{a\sigma^2}$$

$$z_{F,t} = \frac{E_{F,t}[p_{t+1}] + \overline{y} - (1+r)p_t}{a\sigma^2}$$

$$U_{C,t} = (p_t + \overline{y} - (1+r)p_{t-1})z_{C,t-1} + \eta U_{C,t-1}$$

$$U_{F,t} = (p_t + \overline{y} - (1+r)p_{t-1})z_{F,t-1} + \eta U_{F,t-1}$$

end for

Algorithm 2 Franke and Westerhoff Agent-Based Model Simulation.

$N = 38000$
$t \leftarrow 2$
$p_1 = 0$
$d_{C,1} = 0$
$d_{F,1} = 0$
$n_{C,1} = 0$
$n_{F,1} = 0$
$a_1 = 0$

$\mu = 0.01$
$\beta = 1$
$p^* = 0$
$\chi = 1.5$
$\phi = 0.12$
$\sigma_C = 2.087$
$\sigma_F = 0.758$
$\alpha_0 = -0.327$
$\alpha_N = 1.79$
$\alpha_P = 18.43$

for $t \leq N$ **do**

$p_t = p_{t-1} + \mu(n_{C,t-1}d_{C,t-1} + n_{F,t-1}d_{F,t-1})$

$\epsilon_C \sim N(0, \sigma_C)$

$d_{C,t} = \chi(p_t - p_{t-1}) + \epsilon_{C,t}$

$\epsilon_F \sim N(0, \sigma_F)$

$d_{F,t} = \phi(p^* - p_t) + \epsilon_{F,t}$

$n_{C,t} = \frac{1}{1+\exp[\beta a_{t-1}]}$

$n_{F,t} = 1 - n_{C,t}$

$a_t = \alpha_0 + \alpha_n(n_{F,t} - n_{C,t}) + \alpha_p(p_t - p^*)^2$

end for

N R execution time

Table N.1: Below shows the how long the code takes to run in order to produce the results mentioned in the report. In our Github repository Amod and Barnes (2022) there is a ".RData" file saved for each market if running the code from scratch is not of interest however it is advised to look at the "README.md" of a specific folder in the Github repository Amod and Barnes (2022)

Software	Task	Runtime
R	Data cleaning and processing of all markets	$<$ 5 minutes
R	Bloomberg TOP 40	$\approx$ 5.50 hours
R	Robert Shiller S&P 500 1871 - 2015	$\approx$ 7.46 hours
R	Bloomberg other markets (excluding TOP 40) Robert Shiller S&P 500 1871 - 2013 Robert Shiller S&P 500 1871 - 2021	$<$ 5 minutes