Doctoral Programme in Economics and Business

Universitat Jaume I Doctoral School

Empirical analysis of market behaviour: a mesoscopic approach

Report submitted by David Vidal Tomás in order to be eligible for
a doctoral degree awarded by the Universitat Jaume I

Author:
David Vidal Tomás

Supervisors:
Dr. Simone Alfarano
Dr. Gabriele Tedeschi

Castellón de la Plana, June 2020

Funding

- This doctoral dissertation was funded by the Spanish Ministry of Education (grant number FPU2015/01434).

Publications

- Vidal-Tomás, D., & Alfarano, S. (2020). An agent-based early warning indicator for financial market instability. Journal of Economic Interaction and Coordination, 15(1), 49-87.

“This thesis has been accepted by the co-authors of the publications listed above that have waved the right to present them as a part of another PhD thesis”

Cuando emprendas tu viaje a Ítaca
pide que el camino sea largo,
lleno de aventuras, lleno de experiencias.
No temas a los lestrigones ni a los cíclopes
ni al colérico Poseidón,
seres tales jamás hallarás en tu camino,
si tu pensar es elevado, si selecta
es la emoción que toca tu espíritu y tu cuerpo.
Ni a los lestrigones ni a los cíclopes
ni al salvaje Poseidón encontrarás,
si no los llevas dentro de tu alma,
si no los yergue tu alma ante ti.

Pide que el camino sea largo.
Que muchas sean las mañanas de verano
en que llegues -¡con qué placer y alegría!-
a puertos nunca vistos antes.
Detente en los emporios de Fenicia
y hazte con hermosas mercancías,
nácar y coral, ámbar y ébano
y toda suerte de perfumes sensuales,
cuantos más abundantes perfumes sensuales puedas.
Ve a muchas ciudades egipcias
a aprender, a aprender de sus sabios.

Ten siempre a Ítaca en tu mente.
Llegar allí es tu destino.
Mas no apresures nunca el viaje.
Mejor que dure muchos años
y atracar, viejo ya, en la isla,
enriquecido de cuanto ganaste en el camino
sin aguantar a que Ítaca te enriquezca.

Ítaca te brindó tan hermoso viaje.
Sin ella no habrías emprendido el camino.
Pero no tiene ya nada que darte.
Aunque la halles pobre, Ítaca no te ha engañado.
Así, sabio como te has vuelto, con tanta experiencia,
entenderás ya qué significan las Ítacas.

As you set out for Ithaka
hope your road is a long one,
full of adventure, full of discovery.
Laistrygonians, Cyclops,
angry Poseidon—don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians, Cyclops,
wild Poseidon—you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

Hope your road is a long one.
May there be many summer mornings when,
with what pleasure, what joy,
you enter harbors you're seeing for the first time;
may you stop at Phoenician trading stations
to buy fine things,
mother of pearl and coral, amber and ebony,
sensual perfume of every kind—
as many sensual perfumes as you can;
and may you visit many Egyptian cities
to learn and go on learning from their scholars.

Keep Ithaka always in your mind.
Arriving there is what you're destined for.
But don't hurry the journey at all.
Better if it lasts for years,
so you're old by the time you reach the island,
wealthy with all you've gained on the way,
not expecting Ithaka to make you rich.

Ithaka gave you the marvelous journey.
Without her you wouldn't have set out.
She has nothing left to give you now.
And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you'll have understood by then what these Ithakas mean.

Ithaka, Konstantinos P. Kavafis

Acknowledgements

Abraham Lincoln decía que "al final, lo que cuenta no son los años de tu vida, sino la vida de tus años". Esta simple frase tiene mucho significado para mí, ya que la tesis doctoral ha llenado estos años de momentos que jamás olvidaré. Momentos buenos y momentos malos que me han hecho crecer no sólo intelectualmente, sino sobre todo, personalmente. No puedo jurar que tomara siempre las decisiones correctas durante estos años. Sin embargo, hay una cosa de la que sí estoy seguro: estos años no estarían tan llenos de vida sino hubiera sido por todas las personas que han estado conmigo a lo largo de este camino. Un camino que acaba para volver a comenzar.

Me gustaría aprovechar estas páginas para agradecer, de la mejor forma que me es posible, a todos aquellos que habéis hecho que hoy esté dónde estoy... que hoy sea quién soy. Obviamente he aprendido muchísimo técnicamente durante el doctorado gracias a mis directores: modelos, programación, tratamiento de datos y análisis de todo tipo. Pero de ellos me llevo algo más que "simples" conceptos de libro. Por tanto, quiero agradecerles por haberme enseñado "más" de lo técnicamente esperado. En primer lugar, quiero darle las gracias a mi director, Simone Alfarano, por su ayuda y consejo durante estos años. Él fue quién me enseñó dónde comenzaba el camino de *baldosas amarillas* que hoy sigo recorriendo. Gracias a él aprendí que lo importante de la universidad es aprender a aprender. Y por encima de todo, gracias a él aprendí lo que William Golding escribió en el Señor de las moscas: "tenemos que darnos cuenta de que el miedo no nos puede hacer más daño que un sueño". Gracias Simone por haberme guiado hasta aquí. También quiero agradecer a mi director Gabriele Tedeschi por todo su tiempo y dedicación. Él siempre hizo más de lo que debería para ayudarme durante la tesis. Ha sido un alivio saber que siempre podía contar contigo. Pero sobre todo, quiero darle las gracias por enseñarme que la felicidad de recorrer el camino que queremos es mayor a la riqueza que éste nos puede dar. Me alegra saber que mi monedero nunca sabrá contar los tesoros que este pobre camino nos da. Por último, en lo que a la guía se refiere, non posso dimenticare l'aiuto e il lavoro di Maria Cristina Recchioni durante il mio soggiorno ad Ancona. Lavorare al suo fianco mi ha aiutato a crescere come ricercatore e a raggiungere mete a cui non pensavo di poter arrivare. Maria Cristina per me è stata, ed è, un esempio in termini di perseveranza e duro lavoro. Non ho abbastanza parole per ringraziarla di tutta l'attenzione (personale e accademica) che mi ha dato mentre ho visuto in Italia.

Por otro lado, me gustaría dar las gracias a todos aquellos que a lo largo de estos años han sido mis compañeros en la Universitat Jaume I: Alejandro Barrachina, Omar Blanco, Emanuele Ciola, Annarita Colasante, Anabel Doñate, Jaume García, Mariya Ivanova, Diego de Mingo, Sergi Moliner, Isabel Narbón, Jesús Peiró, Marko Petrovic, Giulia Provenzano, Jordi Ripollés, Alba Ruiz, Enrique

Salvador, Andrea Teglio y Lidia Vidal. Con todos vosotros he vivido mil historias que siempre estarán conmigo. Gracias por haberme dado tanto, a cambio de tan poco. Estoy seguro de que el sombrerero loco tendría envídia de nuestras fiestas del té. Agradecer también a Ana María Ibañez y Jose Emilio Farinós por todo su apoyo y trabajo desde que obtuve el Master en la Universidad de Valencia. Por último, en esta parte, y no menos, quiero darle las gracias a Silvia Giménez por su paciencia y amabilidad ante cualquier petición estúpida, y a Eva Camacho, Miguel Ginés y Aurora García por haberme ayudado siempre que lo necesitaba.

En el capítulo personal, me gustaría agradecer a mis amigos todo su tiempo y comprensión en estos años. Gracias a Vicente Martí por haberme aguantado con todas las conversaciones, habidas y por haber, sobre todos los temas posibles. Gracias a Paula Payán por siempre estar ahí a pesar del tiempo separados. Gracias a Adrián García, Alba Lucerga, Blue y Edgar Arribas, por darme una segunda casa igual de cálida y acogedora que la primera. Gracias a Paul Tarcea, Maikel Tena y Ricardo Martinez por las cenas, las conversaciones y risas sin sentido. Gracias a Natalia Rubert, Laura Kuret, Juan Linares, Sandra Rojas, Laura Reyes, Gema Ribes y Samuel Algorta por hacer que este camino haya sido más fácil. Y gracias a todos aquellos que hoy os encontráis en el planeta del olvido. Sin vosotros tampoco hubiera llegado hasta aquí.

Para concluir los agradecimientos, quiero darle las gracias a mi familia por haberme apoyado incondicionalmente en las buenas y en las malas. Darle las gracias a mi sobrina, Vera, por hacernos sonreír cada día llenándonos de ilusión por el futuro. Gracias a mi hermana, Inma, por haber sido mi segunda madre durante tantos años, habiéndomelo dado todo a cambio de nada. Y para acabar, quiero que las últimas palabras que escribo en esta tesis sean para mis padres, Jose Manuel y Paquita. Gracias por haber hecho del niño que fui, el hombre que hoy escribe estas líneas.

David Vidal Tomás

Contents

List of Figures xv

List of Tables xxiii

Abstract xxvii

Resumen xxix

1 Motivation 1

2 An agent-based early warning indicator for financial market instability 5

2.1 Introduction 5

2.2 Data 8

2.3 Sentiment Index 10

2.4 Kirman Herding Model 13

2.4.1 Transition probabilities 14

2.4.2 The aggregate dynamics and the Langevin equation 15

2.5 Validation of the model and estimation of its parameters 18

2.5.1 Estimation of the parameters of the Kirman model 19

2.5.2 Choice of the parameter L of the EMA 22

2.6 Empirical application: Early warning indicator 24

2.7 Robustness analysis 29

2.7.1 US stock markets 29

2.7.2 Worldwide stock markets 30

2.7.3 Global financial village 32

2.8 Conclusion 36

3 Should I herd or should I not? A stochastic herding model for Bitcoin dynamics 39

3.1 Introduction 39

3.2 The model 42

3.2.1 The Switching mechanism 45

3.3 Data description 49

3.4 Estimation and forecasting procedures 50

3.4.1 The indirect inference method 50

3.4.2 Choice of the time window N_w 51

3.4.3 Forecasting procedure 51

3.5 Empirical analysis 52

3.5.1 Model consistency 52

3.5.2 Forecasting analysis 54

3.5.3 Early warning alert system 59

3.6 Concluding remarks 61

4 A cross-sectional analysis of growth and profit rate distribution : the Spanish case 63

4.1 Introduction 63

4.2 Data 66

4.3 Methodology 68

4.4 Empirical results 70

4.4.1 Symmetric case 70

Distributional properties conditional on size and business cycle phase 73

4.4.2 Asymmetric case 75

Distributional properties conditional on size and business cycle phase 78

4.5 Conclusion 82

5 Concluding remarks 85

Appendices 87

A An agent-based early warning indicator for financial market instability 89

A.1 Derivation of the Fokker-Planck equation 89

A.2 Derivation of the equilibrium distribution 90

A.3 Robustness analysis of the determination of the thresholds 91

A.4 Early warning indicators 92

A.5 Worldwide stock markets 95

A.5.1 Sentiment index 95

A.5.2 Tests of volatility clustering 98

A.5.3 Box plots of the moments of simulated data 98

A.5.4 Choice of the parameter L of the EMA 101

A.5.5 Autocorrelation function 105

B A cross-sectional analysis of growth and profit rate distribution : the Spanish case 109
B.1 Firms 109
B.2 Likelihood ratio test for the 200 largest long-lived firms 109
B.3 Probability density function of growth rates of sales 111
B.4 Likelihood ratio test for all the samples 113

Bibliography 117

List of Figures

2.1 Sentiment index and index changes of the S&P 500 (L=100). 13

2.2 Probability density function of the sentiment index compared to the theoretical distribution, given the estimates of each method. 19

2.3 Probability density function of the sentiment index compared to the theoretical distribution, given the estimates of each method and excluding extreme events ($z_t > 0.95$). 20

2.4 Empirical index (continuous line) and simulated sentiment index (dashed line). The Langevin equation was used with the parameters of Table 2.3 to simulate the time series. 21

2.5 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. 22

2.6 Correlogram of the empirical sentiment index z_t as compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 2.3. . . 22

2.7 Estimated parameters ε_1, ε_2, and b according to different values of L. The scaling between the variation of $E[\Delta z^2]$ as a function of L and the parameter b is reported (dotted black line). 24

2.8 Scatter plot of $b(\varepsilon_1 + \varepsilon_2)$ and β_L using different values of L (from 10 to 200 with an increment of 5). We can appreciate a deviation from the 45-degree line just for very low values of L, which corresponds to large values of β_L. 24

2.9 Three possible scenarios for the stock index based on ε_1 and ε_2. In the first one ($\varepsilon_1 < \varepsilon_2$), there is optimism in the market. In the second one ($\varepsilon_1 \approx \varepsilon_2$), there is no dominant mood. In the last one ($\varepsilon_1 > \varepsilon_2$), there is pessimism in the market. 26

2.10 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. S&P 500 index. 27

2.11 Probability density function of the US sentiment indices compared to the theoretical distribution. The Langevin equation was used to obtain the estimates, excluding extreme events ($z_t > 0.95$). 30

2.12 Probability density function of the worldwide sentiment indices compared to the theoretical distribution. The Langevin equation was used to obtain the estimates excluding extreme events ($z_t > 0.95$). 32
2.13 Standardized early warning indicators of the S&P 500, Nasdaq 100, Nikkei 225, and Euro Stoxx 600. Light areas represent the US Business Cycle contractions according to the National Bureau of Economic Research. 34
2.14 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. Nasdaq 100 index. 35
2.15 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. Euro Stoxx 600 index. 35
2.16 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. Nikkei 225 index. 36

3.1 Ants model. 45
3.2 Bitcoin daily prices and daily log-prices, along with the daily returns, from 16/07/2010 to 01/01/2019. 50
3.3 Kolmogorov Smirnov test based on Pedersen approach for each N_w. 51
3.4 Dynamics (left panel) and density (right panel) of $4(z_t - 1/2)$ (black solid line) and the log-return x_t (blue dashed line). 53
3.5 One period ahead forecast (blue dashed line) and observed prices (black solid line). In the left panel prices and forecast are represented one out of every fifty ones. In the right panel the worst fifty one-period ahead forecasts. 54
3.6 Observed (black solid line) and forecast (blue dashed line) realized variance over time. 58
3.7 Bullish/bearish traders ratio computed as $\varepsilon_{1,t}/\varepsilon_{2,t}$ (blue dashed line) and the logarithm of the price (black solid line). Gray area highlights bubble periods detected by Phillips et al. (2015) methodology (GSADF). $N_w = 450$. 61
3.8 Bullish/bearish traders ratio computed as $\varepsilon_{1,t}/\varepsilon_{2,t}$ (blue dashed line) and the logarithm of the price (black solid line). Gray area highlights bubble periods detected by Phillips et al. (2015) methodology (GSADF). $N_w = 130$. 61

4.1 The evolution of the cross-sectional median and standard deviation of growth ($\tilde{g}$) and profit rates for the 200 largest long-lived firms and the entire sample at our disposal (base year 2016). 67

4.2 Estimates of the shape and scale parameter of the Subbotin distribution for profit rates. Error bars show two standard errors. The results refer to the 200 largest long-lived firms according to their sales in 2016. The dashed line in the scale parameter figure represents the median of the estimates. 71

4.3 Estimates of the shape and scale parameter of the Subbotin distribution for growth rates of total assets and sales. Error bars show two standard errors. The results refer to the 200 largest long-lived firms according to their sales in 2016. The dotted line (sales) and dashed line with dots (total assets) in the scale parameter figure represent the median of the estimates. 72

4.4 Estimates of the shape parameter of the Subbotin distribution of profit rates, growth of total assets and sales. Error bars show two standard errors. Results refer to the largest long-lived firms of our sample according to their sales in 2016. 74

4.5 Estimates of the scale parameter of the Subbotin distribution for profit rates, growth of total assets and sales. Error bars show two standard errors. Results refer to the largest long-lived firms of our sample according to their sales in 2016. The dashed line (profit), dotted line (sales) and dashed line with dots (total assets) in the scale parameter figure represent the median of the estimates. 74

4.6 Probability density function (PDF) of profit rates along with the AEP (dotted line) and Laplace (dashed line) distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016. 76

4.7 Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) for the profit rates distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016, removing the two extreme values in each year. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively. . . 77

4.8 Probability density function (PDF) of growth rates of total assets along with the AEP (dotted line) and Laplace (dashed line) distributions. The results refer to the 200 largest long-lived firms according to their sales in 2016. 77

4.9 Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) for growth rates of total assets and sales. Results refer to the 200 largest long-lived firms according to their sales in 2016. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively. 78

4.10 Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) of the AEP distribution for profit rates conditional on size. Error bars show two standard errors. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively. 79

4.11 Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) of the AEP distribution for growth rates of total assets conditional on size. Error bars show two standard errors. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively. 80

4.12 Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) of the AEP distribution for growth rates of sales conditional on size. Error bars show two standard errors. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively. 81

A.1 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. S&P 500 index. 91

A.2 Standardized early warning indicators of all the international markets. Light and areas represent the US Business Cycle contractions according to the National Bureau of Economic Research. 92

A.3 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. S&P 400 midcap index. 92

A.4 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. ASX 200 index. 93

A.5 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. TSX index. 93

A.6 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. JSE All-Share index. 94
A.7 Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. FTSE 100 index. 94
A.8 Sentiment index and index changes of the S&P 400 midcap. 95
A.9 Sentiment index and index changes of the Nasdaq 100. 95
A.10 Sentiment index and index changes of the ASX 200. 96
A.11 Sentiment index and index changes of the TSX. 96
A.12 Sentiment index and index changes of the Euro Stoxx 600. 96
A.13 Sentiment index and index changes of the Nikkei 225. 97
A.14 Sentiment index and index changes of the JSE All-Share. 97
A.15 Sentiment index and index changes of the FTSE 100. 97
A.16 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. S&P 400 midcap. 99
A.17 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. Nasdaq 100. 99
A.18 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. ASX 200. 99
A.19 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. TSX. 100
A.20 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. Euro Stoxx 600. 100
A.21 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. Nikkei 225. 100

A.22 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. JSE All-Share. 101

A.23 Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. FTSE 100. 101

A.24 Estimated parameters ε_1, ε_2 and b according to different values of L. The scaling between the variation of $E[\Delta z^2]$ as a function of L and the parameter b is reported (dotted black line). 102

A.25 Scatter plot of $b(\varepsilon_1 + \varepsilon_2)$ and β_L using different values of L (from 10 to 200 with an increment of 5). Worldwide stock markets. 104

A.26 Scatter plot of $b(\varepsilon_1 + \varepsilon_2)$ and β_L using different values of L (from 10 to 200 with an increment of 5). S&P 500 index including 40, 80, 150 and 208 stocks. 105

A.27 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 4 of the paper. S&P 400 midcap index. 106

A.28 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 4 of the paper. Nasdaq 100 index. 106

A.29 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. ASX 200 index. 106

A.30 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. TSX index. 107

A.31 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. Nikkei 225 index. 107

A.32 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. Euro Stoxx 600 index. 107

A.33 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. JSE All-Share index. 108

A.34 Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. FTSE 100 index. 108

B.1 Sales as a function of GDP for the largest long-lived firms in our sample. 109
B.2 Probability density function (PDF) of growth rates of sales along with the AEP (dotted line) and Laplace (dashed line) distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016. 111

List of Tables

2.1 Details of the stock indices used in the empirical analysis. 9

2.2 Estimated parameters, ε_1, ε_2 and b for the sentiment index. 20

2.3 Estimated parameters, ε_1, ε_2 and b for the sentiment index, excluding extreme events ($z_t > 0.95$). 21

2.4 Estimated parameters, ε_1, ε_2 and b for the US stock indices. The Langevin equation was used to obtain the estimates excluding extreme events ($z_t > 0.95$). 30

2.5 Estimated parameters, ε_1, ε_2 and b for worldwide stock indices. The Langevin equation was used to obtain the estimates excluding extreme events when $z_t > 0.95$. 32

3.1 Descriptive statistics. 49

3.2 Testing model 1: $b_t \Delta t(\varepsilon_{1,t} + \varepsilon_{2,t}) = \alpha_0 + noise$, $\alpha_0 = 1$ (first two lines); Testing model II: $\widehat{\mu}_t = \gamma_0 + \gamma_1 \left(\varepsilon_{1,t}/(\varepsilon_{1,t} + \varepsilon_{2,t}) - 1/2\right) + noise$, $\gamma_0 = 0$, $\gamma_1 = 4$ (last two lines). 54

3.3 Mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and percentage of accuracy of our model for the corresponding periods. Forecast log-prices. 55

3.4 Comparison of our model when forecasting prices in relation to the models proposed in the literature by means of the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and % of accuracy of our model for the corresponding out-sample periods. 57

3.5 Mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and % of accuracy of our model of the corresponding periods. Volatility. 59

3.6 Comparison of our model when forecasting volatility in relation to the models proposed in the literature by means of the mean square error (MSE) for the corresponding out-sample periods. 59

4.1 P-values of the likelihood ratio test for profit and growth rates of total assets and sales. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the Subbotin distribution. The results refer to the 200 largest long-lived firms, according to their sales in 2016, when deleting the extreme positive and negative value. 72
4.2 Median of the estimates of the shape parameter reported in Fig. (4.4). 73
4.3 Median of the estimates of the scale parameter reported in Fig. (4.5). 73
4.4 Pearson correlation coefficient between the time series of the estimates of m, α and σ with the time series of GDP growth rates. 75
4.5 Median of the estimates α_l and α_r reported in Figs. (4.10), (4.11) and (4.12). 81
4.6 Median of the estimates σ_l and σ_r reported in Figs. (4.10), (4.11) and (4.12). 82
4.7 Pearson correlation coefficient between the time series of the estimates of α_l and α_r with the time series of GDP growth rates. 82
4.8 Pearson correlation coefficient between the time series of the estimates of σ_l and σ_r with the time series of GDP growth rates. 82

A.1 *P*-values of the McLeod-Li and Engle's ARCH test for empirical data, and median of the *p*-values obtained from the 500 Monte Carlo simulations (Langevin equation) for simulated data. 98

B.1 P-values of the likelihood ratio test for profit and growth rates of total assets and sales. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the Subbotin distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016. 110
B.2 P-values of the likelihood ratio test for profit rates and growth rates of total assets and sales. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the AEP distribution. Results refer to the 200 largest long-lived firms, according to their sales in 2016. 110
B.3 P-values of the likelihood ratio test for profit rates and growth rates of total assets and sales. The null hypothesis is the asymmetric Laplace distribution, while the alternative hypothesis is the AEP distribution. Results refer to the 200 largest long-lived firms, according to their sales in 2016. 111
B.4 P-values of the likelihood ratio test for profit and growth rates of total assets and sales conditional on size. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the Subbotin distribution. 113

B.5 P-values of the likelihood ratio test for profit and growth rates of total assets and sales conditional on size. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the AEP distribution. 114

B.6 P-values of the likelihood ratio test for profit rates and growth rates of total assets and sales. The null hypothesis is the asymmetric Laplace distribution, while the alternative hypothesis is the AEP distribution. Results refer to the 200 largest long-lived firms, according to their sales in 2016. 115

Abstract

In recent decades, most developed countries have suffered several financial crises, which highlights the need of creating methods and techniques that assess the stability of their stock markets. Specifically, the literature on financial contagion has shown that models based on the interaction between heterogeneous agents shed some light on the fragility of financial markets. Many scholars proposed models with different settings (like coordination mechanism in traders' behaviour, competition among different investors' strategies or behavioural switching among different trading strategies) in order to describe the stock market dynamics by means of the connection between price fluctuations and social interaction. In the same line, firm dynamics literature also demonstrates that simple agent based models give rise to the statistical regularities identified in the empirical data at the macro level. The best example is the Laplace distribution of cross-sectional firms growth rates, which can be thought as the macroscopic evidence of the existence of complex interactions among firms.

With this doctoral dissertation, we aim to contribute to the literature in two different ways. On the one hand, using an agent based model, we provide policymakers and investors with new methods that can be used in financial markets to assess their stability, describe their dynamics, and forecast their future performance. On the other hand, we analyse the evolution of the cross-sectional distribution of activity in the real economy to better understand the firms' destiny. The main results of our two first chapters show that models based on the social interactions provide investors with useful information to set trading strategies, while the last chapter of this thesis demonstrates that the firms' destiny can be tracked by means of the analysis of their profit and growth rate cross-sectional distribution over time.

Resumen

En décadas recientes, la mayoría de los países desarrollados han sufrido diversas crisis financieras, lo que subraya la necesidad de crear métodos y técnicas que evalúen la estabilidad de los mercados financieros. Específicamente, la literatura enfocada en el contagio financiero ha demostrado que modelos basados en la interacción entre agentes heterogéneos arroja luz sobre la fragilidad de estos mercados. De esta forma, muchos investigadores proponen modelos con distintas especificaciones (cómo mecanismos de coordinación entre el comportamiento de los inversores o competición entre distintas estrategias de inversión) para describir la dinámica de los mercados financieros mediante la conexión entre las fluctuaciones de los precios y la interacción social de los inversores. En la misma línea, la literatura enfocada en la dinámica de las empresas también demuestra que simples modelos basados en agentes dan lugar a regularidades estadísticas identificadas en los datos empíricos a nivel macro. El mejor ejemplo es la distribución Laplace de la tasa de crecimiento de las empresas, lo cual puede ser entendido como una evidencia macroscopica de la existencia de interacciones complejas entre empresas.

Con esta tesis doctoral, contribuimos en la literatura de dos formas distintas. Por un lado, mediante un modelo basado en agentes, proporcionamos a los inversores e instituciones gubernamentales nuevos métodos que pueden ser usados en los mercados financieros para evaluar su estabilidad, describir su dinámica, y predecir su comportamiento. Por otro lado, analizamos la evolución de la distribución transversal de la actividad económica para entender apropiadamente el futuro de las empresas. Los principales resultados de los dos primeros capítulos muestran que los modelos basados en las interacciones sociales proporcionan a los inversores información útil para establecer estrategias de inversión, mientras que el último capítulo de la tesis demuestra que el destino de las empresas puede ser evaluado mediante el análisis de la distribución de la rentabilidad económica y los ratios de crecimiento.

Chapter 1

Motivation

The increasing complexity of modern economies, composed of many individuals that mutually interact, obliges scholars to consider in their models the agents' interactions in order to understand the dynamics of interactive complex systems and the corresponding emergent properties (Tesfatsion and Judd, 2006). The mainstream economic literature, in fact, shows that existing theories do not include interactions between economic agents (Tedeschi et al., 2012).

One of the recent methodologies that can take into account those characteristics among a large number of individuals is the Agent-Based model (ABM hereafter), which is based on the "bottom up" approach (De Grauwe, 2010). Compared to the Stochastic Dynamics General Equilibrium approach, characterized by a representative agent and an aggregate system outcome, ABMs are able to explain different phenomena at the macroeconomic level through the interaction of heterogeneous agents at the micro level (Bargigli and Tedeschi, 2013). More specifically, this new approach has been applied to different fields like Industrial Dynamics (Delli Gatti et al., 2005), Macroeconomics (Teglio et al., 2012) and Behavioural Finance (Lux and Alfarano, 2016), among others. Thus, the "bottom up" approach is able to better describe real world phenomena, such as herding in financial markets, banking crises, financial bubbles or competitive processes among firms. Moreover, it also provides an explanation of some empirical regularities, such as the power law probability density function for firms' size (Delli Gatti et al., 2005), or the fatness of the tail of return distribution in financial markets (Alfarano et al., 2005).

In this thesis, we are mainly focused on two fields: Behavioural Finance and Industrial Dynamics. The field of Behavioral Finance aims at examining stock markets by taking into account the behaviour of the traders as a relevant factor in describing the dynamics of financial markets. The Efficient Market Hypothesis, the paradigm to describe the behaviour of financial markets, is essentially based only on the incorporation of exogenous flow of information. Nevertheless, the price in a financial market is also the result of the investors' interactions. Most scholars of the ABM approach consider generally two main categories of traders in the market: fundamentalists and chartists. The former behaves according to the deviation of prices from the fundamental value, while the latter deviates from the rational behaviour, with different approaches, like the imitative behaviour. However, it is possible to find different group of traders depending on the objective of the research, like informed/non-informed, smart/naive, pessimist/optimist or rational/noise traders among other examples (Lux and Alfarano, 2016). Within this framework, the heterogeneity of these groups is always considered as a crucial feature, thus it allows to shed more light on the effect of the traders' social interaction on financial markets (Alfarano, 2006).

In fact, most of these models (Lux, 1995, 1996, 1998; Aoki and Yoshikawa, 2002; Wagner, 2003; Alfarano et al., 2005, 2008; Chiarella et al., 2009) show that social dynamics gives rise to the main empirical regularities in financial markets, namely heteroskedasticity in the fluctuations of financial returns, their unpredictability, the fat tails of their unconditional distribution, and the long-term dependence in their volatility. In other words, these contributions show that in the financial markets do not exist a simple one-to-one relationship between price changes and the new incoming information hitting the market, given that the statistical regularities can be considered as an emergent property of the traders' behaviour. Therefore, although these studies model differently the way that traders interact with each other, they can replicate these stylised facts observed in the empirical data by using social dynamics as the main ingredient of their models. Nevertheless, the direct estimation of the parameters represents one of the main drawbacks of the literature given the complexity of the underlying social dynamics. To address this issue, some scholars introduce indirect estimation methods using the time series of returns to estimate some parameters related to the behaviour of traders (Gilli and Winker, 2003; Franke and Westerhoff, 2011).

Considering this context, in the first chapter of this thesis we propose a market breadth indicator that describes the agents' behaviour by means of the collective movement of the stocks. This proxy of the optimistic/pessimistic mood of traders can be considered as a kind of "mesoscopic device" to analyse the social interaction of traders. We are able to estimate directly from the price time series some parameters connected with the internal dynamics of the market by using the modification of the Kirman model proposed by Alfarano et al. (2005). These parameters can be empirically used to detect bull and bear market phases due to the optimism/pessimism of the investors. Hence, investors can examine the state of the market allowing them to set different strategies to beat the market. In particular, we show that the methodology proposed is useful for nine different international stock markets, which provides enough robustness to the empirical findings of this chapter. However, one of the main limitations of the market breadth indicator proposed in the first chapter is the need of having a considerable number of assets in a given market to represent the sentiment of investors by means of the collective movement of the stocks.

We propose in the second chapter of this thesis a behavioural model that allows us to describe the internal dynamics of a *given* asset in terms of the interaction between two different groups of traders.

More specifically, we define a simple stochastic model that describes the ratio of two consecutive prices as the odds ratio of bullish to bearish traders. By means of this approach, we assess the price changes as the historical competition between bullish traders, who believe in an increase in price, and bearish traders, who believe in a decrease in price. Related to the first chapter of this thesis, the odds ratio is modelled by a stochastic dynamics that is derived from the modification of the Kirman model proposed by Alfarano et al. (2005). However, unlike the first chapter, we employ the behavioural model to analyse Bitcoin.

Compared to other financial assets, Bitcoin is mainly driven by the interaction of traders (Baek and Elbeck, 2015) since there is not a explicit fundamental value, which most of the times is related to particular features like hashrate, blockchain position and liquidity (Corbet et al., 2018). Due to this fact, Bitcoin is characterised by an explosive behaviour with continuous bubbles (Cheah and Fry, 2015) generated by the enthusiasm and fear of the traders that buy and sell according to the news in social media. Even though most of the proposed models in the literature try to describe Bitcoin dynamics by means of machine learning and neuronal networks (Jang and Lee, 2018; McNally et al., 2018), we believe that a behavioral model based on the Kirman model is more suitable to describe this kind of asset whose value is given by the interaction of agents. With this model, we show in the second chapter that it is possible not only to describe the state of the market, but also to forecast Bitcoin price and volatility with satisfactory results compared to other recent models (Jang and Lee, 2018; McNally et al., 2018; Catania et al., 2019; Atsalakis et al., 2019).

In the third chapter, we are focused on the field of Industrial Dynamics. Gibrat (1931) was the first scholar to examine the dynamics and growth of individual firms whose influential idea was to contend that the growth rate of each firm is independent of its size. This fact would imply normally distributed growth rates, and a log-normal distribution of firm sizes. However, many authors demonstrate that (i) the distribution of firms' size follows a power law rather than a log-normal distribution (Axtell, 2001; Gaffeo et al., 2003) and (ii) growth rate of firms follows a Laplace distribution rather than a Normal distribution (Amaral et al., 1997; Bottazzi et al., 2001; Bottazzi and Secchi, 2003, 2006; Bottazzi and Secchi, 2011; Buldyrev et al., 2007; Alfarano et al., 2008; Riccaboni et al., 2011). The non-normality of the growth rate distribution could be considered as an imprint of the complex interactions among firms in a competitive process. In fact, Delli Gatti et al. (2005) show that a simple financial fragility agent based model based on complex interaction of heterogeneous agents (firms and banks) gives rise to (i) the power law distribution of firms'size and (ii) a Laplace distribution for growth rates of firms' output and countries' GDP. The approach used in this study is based on the assumption that the interactions among the single units of a given system may give rise to some statistical regularities at the macro level that are not related to the behavior of the sub-units. Bottazzi and Secchi (2006) also underline the relevance of agents' interaction given that the Laplace distribution stems from a competitive context in which firms are able to seize new growth opportunities proportional to opportunities already taken. Therefore, Industrial dynamics relates the interaction of agents to some statistical regularities like the cross-sectional distribution of activity in the real economy. In fact, Aoki and Yoshikawa (2011) underlined the relevance of studying probability distributions since "equilibrium in the macroeconomy is better described by a probability distribution than by a 'point' in some space or set." This view is in line with Haltiwanger (1997), who stated that "it is becoming increasingly apparent that changes in the key macroaggregates at cyclical and secular frequencies are best understood by tracking the evolution of the cross-sectional distribution of activity and changes at the micro level." Given the relevance of the probability distribution in macroeconomics, it is important to choose those variables that are able to characterise properly the complexity of the economic system. The existing literature (like Bottazzi and Secchi, 2003, 2006) considers that growth rates is the best measure to describe the dynamics of individual firms. However, some authors (Alfarano et al., 2012; Erlingsson et al., 2013; Mundt et al., 2016) have changed the attention towards profit rates instead of growth rates. In fact, Mundt et al. (2016) affirm that profit rates are more convenient to find a general principle for the dynamics of corporations since volatility is independent of size, they exhibit significant autocorrelation and are more stable regardless of the country and period considered. In this context, we analyse in the third chapter whether profit rates are more convenient to analyse firm dynamics, compared to growth rates, in the Spanish economy from 1998 to 2016. This unstable period, characterised by the remarkable impact of the housing bubble on the economy, allows us to observe which variable shows more robust statistical regularities. The main result of our study underlines the stability of profit rates over time for the largest long-lived firms, even during the downturn.

All in all, this thesis consists of three chapters that, to a greater or lesser extent, underline the relevance of the agent interactions at different levels. Despite the fact that these chapters do not show a complete common thread, the thesis presents those topics that have attracted the attention of the author over these years with the social dynamics as the cornerstone. The readers of my dissertation will hopefully excuse the absence of a unique scientific topic. This fact has allowed me to broaden my mind in different ways, which would not have happened otherwise.

Chapter 2

An agent-based early warning indicator for financial market instability

“I define a speculative bubble as a situation in which news of price increases spurs investor enthusiasm, which spreads by psychological contagion from person to person, in the process amplifying stories that might justify the price increases and bringing in a larger and larger class of investors, who, despite doubts about the real value of an investment, are drawn to it partly through envy of others’ successes and partly through a gambler‘s excitement." (Robert J Shiller; Irrational Exuberance, 2015)

2.1 Introduction

Traditional economic literature, based on the efficient market hypothesis (Fama, 1965, 1991; Malkiel and Fama, 1970), highlights the irrelevant role of irrational traders in the long run, since they would disappear from the stock market due to their poor performance (Friedman, 1953). A popular term for investors acting irrationally is "noise traders" (Kyle, 1985; Black, 1986), whose trading might generate crashes and bubbles due to social contagion. This phenomenon is emphasized in Shiller’s quote (from above) that underlines the importance of social interactions in the stock market as a source of price deviations from their fundamental value. Thus, the presence of noise traders creates situations in which rational investors with risk aversion cannot maintain their position given that "price divergence can become worse before it gets better" (Shleifer, 2000). In other words, noise traders are able to create a particular place in the stock market in which to survive (Lux, 2011) to the detriment of rational traders (De Long et al., 1990).

Agent-based models (ABMs) in finance shed more light on the connection between price fluctuations and social interactions, since they are able to include more features like herding and contagion phenomena as main features of their dynamics (Lux and Alfarano, 2016). The herding model of Kirman (1991, 1993) represents one of the first references in this field with the introduction of a stochastic model of information transmission that was initially designed to explain the behavior of an ant colony in the presence of two sources of food. His seminal work was adapted into a financial perspective in which foreign exchange dealers choose their strategy (chartist or fundamentalist) under the influence of social interactions. Most of these models (Lux, 1995, 1996, 1998; Aoki and Yoshikawa, 2002; Wagner, 2003; Alfarano et al., 2005, 2008) have been proposed to provide an explanation of empirical regularities in financial markets, namely, heteroskedasticity in the fluctuations of financial returns, their unpredictability, the fat tails of their unconditional distribution, and the long-term dependence in their volatility. These studies show mainly that those statistical regularities can be considered as an emergent property of internal dynamics governed by the interaction of different groups of traders and are not just a mere reflection of the new incoming information hitting the market.

One of the main problems in the empirical application of agent-based models to financial data lies in the hidden nature of some of the quantities responsible for the internal dynamics of the market. In particular, the fraction of chartists/fundamentalists or pessimist/optimist investors is an unobservable quantity, which should be estimated indirectly from the time series of price returns. Gilli and Winker (2003) were the first to estimate some parameters of the Kirman model by means of a global optimization heuristic, whose main outcome underlined that the DM/USD (German Mark/U S dollar) foreign exchange market was characterized by the changing moods of the agents. Franke and Westerhoff (2011) employ the method of simulated moments (SMM), along with bootstrap and Monte Carlo techniques, to estimate a structural stochastic volatility model of asset pricing. Their main empirical finding underlines the different behavior of agents in the S&P 500 index and USD/DM exchange rate. Chen and Lux (2015) use the SMM to estimate the model of Alfarano et al. (2008), which is applied to three stock market indices (DAX30, S&P 500, Nikkei225), three foreign exchange rates (USD/EUR [Euro], JPY [Japanese Yen]/USD, CHF [Swiss Franc]/EUR) and the gold price, showing the different behavior of traders in those markets.

The aim of this paper is to shed more light on the empirical research of agent-based models that have been previously mentioned. In particular, we employed the herding model introduced by Kirman (1991, 1993) to describe the evolution of a sentiment index that was inspired by the Bank of America Merrill Lynch (BofAML) global breadth rule (Hartnett et al., 2015). This rule is defined as a contrarian indicator of the equity market according to the nomenclature of financial technical analysis. A market breadth indicator measures the number of companies advancing relative to those declining during a given time period. Positive market breadth occurs when more assets in a market are advancing than declining, suggesting that the market is bullish. Conversely, a disproportionate number of declining stocks are interpreted as a bearish market phase.

Instead of estimating indirectly the fraction of pessimists/optimists from a financial time series of returns, we consider the evolution of the breadth of the entire market as a direct proxy of the optimistic/pessimistic mood of traders. Such a dichotomic classification of the sentiment of traders, very often employed in the behavioral finance literature, is obviously an oversimplification of a much more complex psychological description of the investment attitude of the traders. The use of this simple classification allows us to apply an extensive battery of analytical tools already available in the literature. We implicitly assume that the collective behavior of the stocks in a given market is a reflection of the collective mood of traders and the idiosyncratic shocks to the individual firms. We conjecture that the collective movement of the stocks can be decomposed into idiosyncratic shocks to each stock and a global coupling among them. Such a decomposition is the core of the Kirman herding model, typically employed to describe the behavior of a population of investors. In other words, we consider all the stocks in a given market as sorts of "mesoscopic devices" to detect some aspect of the micro behavior of the investors. We do not explicitly model the connection between the micro motives of the traders and the dynamics of the stock prices and the market index. Such complex interaction has been exhaustively modeled in an extensive part of agent-based finance literature. The main advantage of a shift from the investors' sentiment to the market breadth is that the latter can be directly computed from the price time series, while the former should be indirectly estimated after assuming a particular agent-based model.

The aim of the paper is twofold. The first contribution of the paper is showing that this model can accurately reproduce the main statistical properties of the empirical sentiment index, constructed using the S&P 500 financial index. This was accomplished after noticing that the evolution of the market breadth visually resembled a time series generated by the herding model of Kirman. In particular, we focused our attention on the unconditional distribution, the heteroskedasticity of the fluctuations of the index changes, and the exponential decay of the autocorrelation function. Moreover, we could easily estimate the three parameters of the model with a straightforward application of the maximum likelihood method.

Our second contribution focuses on an empirical application of our model to the sentiment index in order to detect the turning points from bull markets to bear markets. We used a measure of asymmetry of the attitude of the investors based on the sentiment index z_t as an early warning indicator. By means of this tool, we can identify bull markets in which most of the stocks are simultaneously moving together due to the optimism of investors. Hence, this indicator allows investors and policy-makers to anticipate the emergence of pessimistic phases that stem from persistent bull markets. Investors are able to reduce their financial exposure through the progressive sale of stocks or purchasing put-options, while policy-makers can control the mood of investors and apply efficient policies to minimize the contagion from financial markets to the real economy. Finally, the robustness of our results is demonstrated using two different stock indices in the USA (S&P 400 midcap and Nasdaq 100) and six different worldwide stock indices: ASX 200 (Australia), TSX (Canada), Euro Stoxx 600 (Europe), Nikkei 225 (Japan), JSE (South Africa), and FTSE 100 (UK). We observed generally similar features in all the stock indices, thus we can apply our sentiment index and early warning indicator in any market regardless of the country-specific characteristics. It is remarkable that, despite quite a large degree of heterogeneity in the sample of assets and markets considered, the early warning indicator as well as the underlying herding model proved to be fairly robust in predicting the optimistic and pessimistic phases and in describing the dynamics of the market breadth.

The rest of the paper is organized as follows. We describe the data in Section (2.2) along with the introduction of the sentiment index, z_t, in Section (2.3). The herding mechanism is described in Section (2.4), and the estimation of the parameters is shown in Section (2.5) along with the validation of the model. The definition of the early warning indicator and its empirical application are described in Section (2.6). The estimations for the USA and worldwide stock markets are shown in Section (2.7). Finally, we summarize the main results of this paper with the conclusion in Section (4.5).

2.2 Data

Our data sample was constituted of daily stock prices that were sourced from the Thomson Reuters Datastream. More specifically, we employed all the S&P 500 stocks that had been trading from 01/01/1981 to 01/06/2018. The sample included 208 stocks.

To ensure the robustness of our results, we employed two other stock market groups. On the one hand, we studied the S&P 400 midcap and Nasdaq 100 to observe whether the firm size and sector characteristics gave rise to different results. On the other hand, we analyzed six stock markets from different countries to examine whether investors' sentiment can be described by the same model regardless of the different market structures, trading mechanisms or country-specific characteristics (Australia, Canada, Europe[1], Japan, South Africa, and the UK). We examined these indices from 01/01/1993 to 01/06/2018 due to the availability of data.[2] Table 2.1 shows all the stock markets that were used in this study, including the corresponding index, sample period, country, and number of stocks.

---

[1] Given the similarities in relation to the Euro Stoxx 600, we did not report results related to countries like Germany (DAX), France (CAC), and Spain (IBEX). Hence, we focused on the Euro Stoxx 600 to include representation of the European countries.

[2] The only exception was the JSE All-Share index. In this case, our initial year was 2000.

Table 2.1: Details of the stock indices used in the empirical analysis.

Number	Stock index	Sample period	Country	Stocks
1	S&P 500	01/01/1981 - 01/06/2018	USA	208
2	S&P 400 midcap	01/01/1993 - 01/06/2018	USA	188
3	Nasdaq 100	01/01/1993 - 01/06/2018	USA	48
4	ASX 200	01/01/1993 - 01/06/2018	Australia	53
5	TSX	01/01/1993 - 01/06/2018	Canada	81
6	Euro Stoxx 600	01/01/1993 - 01/06/2018	Europe	282
7	Nikkei 225	01/01/1993 - 01/06/2018	Japan	183
8	JSE All-Share	01/01/2000 - 01/06/2018	South Africa	86
9	FTSE 100	01/01/1993 - 01/06/2018	UK	57

Given that we were analyzing stock indices for a time span of 37 years (in the case of S&P 500) and 25 years (for the rest of stock markets with the exception of the JSE index), we detrended the stock prices in order to focus on the behavior of fluctuations around the long-term trends. The basic idea is to detrend the exponential long-term trend of the stock prices using a simple linear regression on the cumulative log-returns for each stock index. The long-term trend of the stock market is mainly affected by the general evolution of the economy and not by the short-term fluctuations in the sentiment of the investors, which was the focus of our study. In order to do so, the first step was to compute the cumulative returns of each stock index calculated as:

$$r_{j,\tau} = \ln\left(\frac{S_{j,\tau}}{S_{j,1}}\right), \quad j = 1, ..., 9; \quad \tau \geq 1, \tag{2.1}$$

where $S_{j,\tau}$ is the stock index from Table 2.1. Afterwards, we employed an ordinary least squares regression to obtain the corresponding average daily return for each index, defined as β_j.[3] We then computed the detrended prices for each one of the stocks in the corresponding market, defined as:

$$_jp_{i,\tau} = \frac{_jP_{i,\tau}}{e^{\beta_j \tau}}, \tag{2.3}$$

where $_jP_{i,\tau}$ are the raw prices while $_jp_{i,\tau}$ denotes the detrended prices of the stock i in market j.[4]

---

[3] The equation that has been estimated is

$$r_{j,\tau} = \alpha_j + \beta_j \tau \tag{2.2}$$

where τ denotes an incremental index for each day of the time series and, consequently, β denotes the average daily return for each index. We checked for the impact of the starting value $S_{j,1}$ on the estimation of β_j, varying it in a time window of approximately 200 days. The estimation was not significantly affected.

[4] If we consider the original time series of prices, our results were not significantly affected, since for the sentiment index, we used a moving average to extract the traders' mood. The detrending technique has been employed to eliminate the small systematic optimistic bias in the sentiment index due to the presence of the exponential long-term growth.

2.3 Sentiment Index

Our paper was inspired by the GBR from the Bank of America Merrill Lynch, which indicates extreme pessimistic scenarios when most of the stock indices around the world are oversold, thus triggering a buy signal to take advantage of possible rebounds. The GBR is based on the market breadth technical trading analysis, which employs a comparison of moving averages as a classic chartist technique. The GBR triggers a buy signal when 88% of stock indices included in the MSCI All Country World Index[5] are simultaneously below their 200-day moving average and 50-day moving average. Such collective movement of all international markets in the same direction is considered an imprint of a pessimistic market mood affecting all the global financial markets. The origin of this collective movement can be found in a pessimistic persistent sentiment of investors, possibly overreacting to some exogenous factors. The main idea of this trading rule is to identify a period of overall pessimistic sentiment in the world financial markets, and consequently, to set the proper timing to enter into the market.

Given that the GBR and other similar indices of market breadth rely inherently on the sentiment of agents in the worldwide stock market, we similarly created sentiment indices based on individual stocks belonging to the stock indices reported in Table 2.1. The technical trading rules based on market breadth indicators are widespread among financial traders to estimate the changes of sentiment of financial investors. Our starting research question was whether we could employ concepts and techniques developed in the agent-based literature to conveniently model such indicators. To do so, we constructed an index similar to GBR that is based on the stocks of a single financial market instead of the financial indices included in the MSCI All Country World Index.

In order to construct our sentiment index, we categorized each one of the N stocks of a given market (N is the fixed number of stocks that we considered) according to two possible states: n_t (State 1) and $N - n_t$ (State 2), where n_t is defined as the number of stocks advancing and $N - n_t$ as the number of stocks declining.

As a proxy for the impact of optimism and pessimism on a given stock, we computed the relative price in a given day with respect to its exponential moving average (EMA, hereafter). In contrast to the GBR index, we only used the price and the moving average, instead of two moving averages, so that we had a binary and mutually exclusive characterization of the state of each stock.[6] The EMA was calculated as a linear combination of the day's closing price and the previous day's EMA value with a given weight W. We denote the EMA of stock i at time t as $\bar{p}_{i,t}$, recursively defined as:

$$\bar{p}_{i,t} = W \cdot p_{i,t} + (1 - W) \cdot \bar{p}_{i,t-1} \quad \text{for } t \geq 1, \tag{2.4}$$

---

[5] The MSCI All Country World Index aggregates 46 indices, 23 developed and 23 emerging markets.

[6] The GBR allows for three possible states of a stock, given the relative level of the price with respect to the two moving averages.

where the first term denotes the percentage of the day's price that is added to the previous day's exponential moving average.[7] We define the weight W to compute the EMA as:

$$W = \frac{2}{L+1}, \tag{2.5}$$

where L is the "equivalent" length of the time window measured in days.[8] We computed the starting value of $\bar{p}_{i,0}$ as the simple arithmetic average over a window of length H, i.e.,

$$\bar{p}_{i,0} = \frac{1}{H} \sum_{\tau=1}^{H} p_{i,\tau}. \tag{2.6}$$

This value was essentially irrelevant for our results, since after a few iterations of Eq. (2.4) (roughly L) the starting value did not have an impact on the value of $\bar{p}_{i,t}$, so that a possible constraint is $H \leq L$.[9]

Having defined $p_{i,t}$ and $\bar{p}_{i,t}$, we consider that the stock i at time t is in a "pessimistic state" when the price of the stock, $p_{i,t}$, is below $\bar{p}_{i,t}$, while we consider that the stock is in an "optimistic state" when the price of the stock is above $\bar{p}_{i,t}$. We denote the state of each stock i as $n_{i,t}$, which takes the values:

$$n_{i,t} = \begin{cases} 1 \text{ if} & p_{i,t} < \bar{p}_{i,t-1}, \\ 0 \text{ if} & p_{i,t} \geq \bar{p}_{i,t-1}. \end{cases} \tag{2.7}$$

This binary characterization of the state of a given stock is just a rough approximation of the mood of the investors, since the state of a single stock might change abruptly due to small fluctuations of the current price with respect to its exponential moving average. It is the aggregate index, in fact, which carries the information on the collective dynamics of the sentiment of the investors in the market. We define n_t as the sum of the individual state of each stock at time t, i.e.,

$$n_t = \sum_{i=1}^{N} n_{i,t}. \tag{2.8}$$

We introduce the normalized sentiment index z_t, which represents the collective behavior of the market at time t in relation to the size of the market N:

$$z_t = \frac{\sum_{i=1}^{N} n_{i,t}}{N} = \frac{n_t}{N}. \tag{2.9}$$

We are implicitly assuming that our estimator of a sentiment index is a reflection of the behavior of the investors. We have several layers involved in the dynamics of the system, some of them not explicitly modeled: i) the micro level where investors, based on their particular trading strategies and expectations, trade the stocks; ii) the price dynamics of the single stock, which reflect to some extent the behavior of the agents and their sentiment; iii) the index of Eq. (2.9), which captures at an aggregate level the sentiment of the investors and its dynamics. A connection between (i) and (ii) is assumed to exist; the price dynamics of a single stock partially reflect the overall sentiment of the investors in the market. Many models in the literature on agent-based finance have indirectly shown the existence of a connection between the trading strategies employed by the financial investors and the properties of the financial time series of prices and returns, like the model proposed by Lux and Marchesi (1999a) or the nonlinear dynamical system introduced by Brock and Hommes (1997) (for references to other models in the literature, see Lux and Alfarano, 2016). In this paper, we proxied the evolution of the sentiment of financial investors with our index constructed from the collective movement of the stocks, which had the great advantage to be an observable quantity.

---

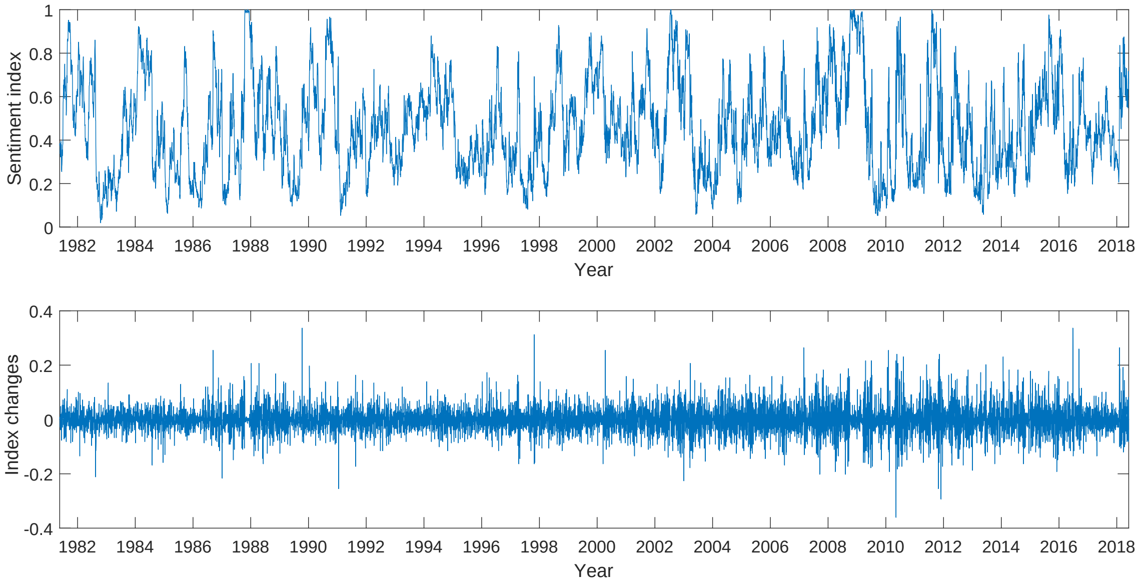
[7] In this and the following sections, we omit an index indicating the particular market to which a stock belongs in order to avoid unwieldy notation. We are implicitly assuming that all the stocks refer to the corresponding market.

[8] Such value for the weight factor allows for an intuitive comparison between an EMA and a simple arithmetic average of the rolling window L. For $W = 2/(L+1)$, both moving averages have the same center of mass. Using Eq. (2.5), we can characterize an EMA with a timescale (see Murphy, 1999 and Hansun, 2013).

[9] Without a loss of generality, we set $H = L$. The choice did not influence our results as soon as $H << T$.

A bear market is characterized by a value of the variable z_t that is close to one; conversely, a bull market is associated to a value of z_t close to zero. So, when the vast majority of stocks have a price lower than their corresponding moving average, we are in the presence of a widespread pessimistic sentiment of the market or possibly of a negative exogenous event, affecting all the stocks at the same time. Fig. (2.1) shows the resulting sentiment index z_t of the S&P 500 along with the time series of daily changes, computed as the difference between z_t and z_{t-1}. We observed swings of sentiment of the market between the two extreme values, $z = 0$ and $z = 1$. Moreover, a visual inspection of the time series of index changes showed the presence of volatility clusters, a phenomenon shared with the time series of price returns (Cont, 2001). Interestingly, periods of high volatility of the index changes coincided with the value of z_t around a balanced configuration ($z_t \approx 1/2$). Conversely, when z_t takes extreme values ($z_t \approx 1$ or $z_t \approx 0$) the index changes exhibit lower volatility. Our simple filtering technique might be useful to characterize the overall behavior of investors in the market. Fig. (2.1) shows a remarkable visual similarity (swings between the two extremes, a heteroskedastic time series of index changes, a specific correlation of high/low value volatility of the index changes with z_t) to the time series generated by the herding model proposed by Kirman (Kirman, 1991; Kirman, 1993), which was inspired by an entomological experiment on the behavior of ants. In this paper, we analyzed to what extent such visual similarity can be translated into a quantitative correspondence. In addition to the visual correspondence, we take the position that the collective movement of the stocks at the macro level can be an imprint of the behavior of financial investors at the micro level. The model of Kirman has been extensively used in the literature to successfully model the behavior of financial agents. Based on this premise, we conjectured that this model could be useful to quantitatively describe not only the behavior of traders, but also the collective movement of stocks in an asset market.

Figure. 2.1: Sentiment index and index changes of the S&P 500 (L=100).



2.4 Kirman Herding Model

Kirman (1991, 1993) has popularized an entomological experiment on the behavior of an ant colony where the insects could choose between two identical sources of food near their nest. Surprisingly, one can observe that most of the ants rest in the proximity of one source of food instead of observing an approximately equal number of ants around the two sources, a scenario which one might intuitively think at first sight. Even more surprisingly, such an asymmetric distribution is not stable over time, but it switches between the two sources. Those experimental observations have been explained through a combination of recruitment interactions among ants and autonomous switchings of individual ants as a result of their stochastic search for food. In particular, the pairwise recruiting interaction of two ants, due to the exchange of pheromones to communicate with each other, can be interpreted as a herding behavior in that one ant decides to follow another ant regardless of its own information on the location of the source of food.

Kirman has introduced a simple stochastic model in order to explain the aggregate asymmetric behavior of the colony together with the switching of the majority of ants over time. His model has been successfully applied to financial markets in order to describe the switching mechanism among different strategies used by agents to trade an asset.[10] The main idea is that investors can be categorized into a few groups based on their trading strategies: fundamentalists/chartists, optimists/pessimists, followers/contrarians, etc. They can switch category depending on some sort of peer pressure or herding interaction, which might create a prevalent optimistic or pessimistic sentiment among investors in a given time period. They can also change investment strategies following some idiosyncratic influences. We assume that the herding coupling among traders at a market level retains its effects on the dynamics of the individual stock price and ultimately allows us to estimate the prevailing sentiment in the market. In line with a large number of contributions in agent-based finance and taking into account a qualitative similarity of the time-series properties, we applied the ant model of Kirman to formally describe the dynamics of the observable sentiment index defined in Eq. (2.9). This allowed us to have an analytical model for the evolution of the sentiment of investors in a financial market whose parameters could be easily estimated.

---

[10] Two main examples of this literature are Lux and Marchesi (2000) and Alfarano et al. (2005). The former relates volatility clustering, fat tails, and the unit root property of assets to the interaction of chartists, who can be optimistic and pessimistic, and fundamentalists. In fact, chartists change their mood not only due to the price trend but also because of the majority opinion. The latter also shows that fat tails and volatility clustering can be considered as emergent properties of the interaction of traders, whose main contribution lies in the direct estimation of the underlying parameters of their herding model given a closed-form solution for the distribution of returns.

2.4.1 Transition probabilities

Here, we illustrate the main properties of the Kirman herding model, and how we can adopt it to describe the evolution of the sentiment index introduced in Section (2.3). The formalization of the model assumes that the stochastic population dynamics, determined by the aggregate variable n_t from Eq. (2.8), evolves according to the Poisson probabilities governing the state of one stock. The aggregate indicator changes from n_t at time t to some $n'_{t+\Delta t_0}$ at time t+Δt_0 according to the given conditional probabilities denoted by $\rho(n'_{t+\Delta t_0} \mid n_t)$. For sufficiently small time increments Δt_0, the probabilities are linear in the time interval and are defined as:

$$\begin{aligned}\rho(n'_{t+\Delta t_0} = n_t + 1 \mid n_t) = &\ (N - n_t)(a_1 + bn_t) \cdot \Delta t_0, \\ \rho(n'_{t+\Delta t_0} = n_t - 1 \mid n_t) = &\ n_t(a_2 + b(N - n_t)) \cdot \Delta t_0,\end{aligned} \tag{2.10}$$

with the further constraint that:

$$\rho(n'_{t+\Delta t_0} = n_t \mid n_t) = 1 - \rho(n'_{t+\Delta t_0} = n + 1 \mid n_t) - \rho(n'_{t+\Delta t_0} = n - 1 \mid n_t). \tag{2.11}$$

Moreover, given that the conditional probabilities $\rho(n'_{t+\Delta t_0} \mid n_t)$ must be bounded to 1 and must be positive, an upper limit can be computed for the elementary time step, Δt_0:

$$\Delta t_0 \leq \frac{2}{bN(N + \frac{a_1}{b} + \frac{a_2}{b})}. \tag{2.12}$$

The probabilities from Eqs. (2.10) and (2.11) define a Markov chain that can be more precisely classified within the class of nonlinear one-step processes (Van Kampen, 1992).[11] The constants a_1 and a_2 account for the average intensity of random switching of state of a stock. Those changes are originated by idiosyncratic exogenous shocks affecting a particular asset. The quadratic term instead accounts for the global coupling among all assets responsible for their collective movements in the market. In this regard, the constant parameter b accounts for the relative intensity of such global coupling with respect to the idiosyncratic propensity to switch.

An individual stock might change its state due to an idiosyncratic external event, for instance, new information affecting the prospects of future cash flow of a given firm. New information affects the trading attitude of investors, triggering their buy or sell signals on that particular stock, which, in turn, might change the current stock price, possibly modifying its relative position with respect to the EMA and eventually its state. The terms a_1 and a_2 trigger random switches among states, regardless of the state of the other stocks. Notice, however, that this sensitivity to change the state depends on the direction of the transition, given the asymmetry of the two coefficients (a_1 and a_2). The state of a stock might also change under the market pressure modeled by the dependence of the probabilities on the overall number of stocks in the opposite state, governed by the constant parameter b. In other words, with this term we account for the global coupling among all stocks in a given market, which is captured within the capital asset pricing model (CAPM) by the dependence of the individual stock return on the return of the index. Collective changes of the mood of investors due to social interactions based on herd behavior are reflected in the corresponding changes of the states of the stocks. Our sentiment index can be considered, therefore, a device to detect indirectly the unobservable movement of the sentiment of investors. The nonlinear term in Eq. (2.10) accounts for the impact that the mutual influence in the behavior of traders has on the collective movements of the stocks.

2.4.2 The aggregate dynamics and the Langevin equation

In Alfarano (2006) and Garibaldi and Scalas (2010), the finitary equilibrium distribution of the Markov chain of Eq. (2.10) is derived, which turns out to be a Polya distribution. In addition to the unconditional distribution, it is possible to compute the conditional distribution of the variable n_t, $\rho(n_{t+\Delta t} \mid n_0)$ for any value of Δt, as an infinite polynomial expansion in the leading eigenvalue of the transition matrix defined in Eq. (2.10) and (2.11) (see Garibaldi and Scalas, 2010). Such a solution can be handled just with numeric techniques, and it does not allow for a clear comprehension of the dynamical structure of the process. Even though the discrete process can be computed numerically, we decided to introduce the continuous approximation which allows for a more intuitive understanding of the dynamics of the aggregate index n_t than the description in terms of the discrete variable. In order to approximate the discrete stochastic process by a continuous diffusion process (Alfarano et al., 2005, 2008, 2013; Alfarano and Milaković, 2009), we define the collective behavior of the whole market with respect to the intensive variable z_t. Alfarano et al. (2005) shows that the Markov chain of Eq. (2.10) can be approximated by the dynamics of the variable z_t within the framework of the Fokker-Planck equation (FPE),

---

[11] The Markov chain described in Eq. (2.10) and (2.11) is characterized by being homogeneous, irreducible, aperiodic, ergodic, with a unique equilibrium distribution (Feller, 1968). In particular, the transition probabilities are homogeneous since they are time independent. Each state can be reached from the rest of the states in a finite number of states, thus the chain is irreducible. Given that the probabilities $\rho(n'_{t+\Delta t_0} \mid n_t) \neq 0$, the chain is aperiodic. Hence, the Markov chain is ergodic, since it is aperiodic and irreducible. Finally, the equilibrium distribution exists and is unique. Garibaldi and Scalas (2010) give a detailed account of the properties of such process.

$$\frac{\partial p(z,t)}{\partial t} = -\frac{\partial}{\partial z}[A(z)p(z,t)] + \frac{1}{2}\frac{\partial^2}{\partial z^2}[D(z)p(z,t)], \tag{2.13}$$

where $A(z)$ represents the drift term,

$$A(z) = a_1 - (a_1 + a_2)z, \tag{2.14}$$

and the diffusion term $D(z)$ is given by

$$D(z) = 2b(1-z)z. \tag{2.15}$$

The resulting equilibrium distribution, obtained by Alfarano et al. (2005) is

$$p_0(z) = \frac{1}{B(\varepsilon_1, \varepsilon_2)} z^{\varepsilon_1 - 1}(1-z)^{\varepsilon_2 - 1}, \tag{2.16}$$

where

$$B(\varepsilon_1, \varepsilon_2) = \frac{\Gamma(\varepsilon_1)\Gamma(\varepsilon_2)}{\Gamma(\varepsilon_1 + \varepsilon_2)}, \tag{2.17}$$

being $\Gamma(\cdot)$ the gamma function.[12] Interestingly, it turns out that it depends only on the ratios $\varepsilon_1 = a_1/b$ and $\varepsilon_2 = a_2/b$ but not on the size of the constants a_1, a_2, and b. The resulting distribution of Eq. (2.16) is known as the beta distribution which is characterized by being a flexible distribution in a bounded domain.

---

[12] Despite the fact that our sentiment index z_t is not a continuous variable given the finite number N of stocks in the system, for coherence of the estimation procedure, we examined the beta as an equilibrium distribution instead of the Polya distribution (Alfarano et al., 2005), since we employed the continuous conditional likelihood to estimate the underlying parameters of the model.

The time evolution of the stochastic variable z_t, from Eq. (2.13), can be alternatively described by the stochastic differential equation in continuous time,

$$dz_t = A(z_t)dt + \sqrt{D(z_t)}\, dW_t, \tag{2.18}$$

where dz_t is the infinitesimal increment of z_t, and dW_t is the Wiener increment.[13] The Eq. (2.18) can be approximated by the discrete time stochastic difference equation, known in physics terminology as a Langevin equation, using the Euler–Maruyama approximation (see Kloeden and Platen, 2013). Therefore, instead of simulating the stochastic process of Eq. (2.10) at the microscopic level with a single transition at a time or solving the FPE of Eq. (2.13) (see Alfarano, 2006), we can also describe the herding mechanism at the macroscopic level by means of the Langevin equation. This approach allows an approximation of the conditional distribution, $\rho(n_{t+\Delta t} \mid n_0)$ of the discrete process of Eq. (2.10) with a Gaussian distribution. In other words, instead of following the herding dynamics at the microscopic timescale Δt_0, when we observed at most a switch of a single asset, we considered a macroscopic time scale Δt, during which we aggregated several variations of the variable z_t in order to obtain a simpler description of its dynamics. The main intuition is that the aggregation of many "micro increments" of the variable z_t leads to a Gaussian conditional distribution because of the central limit theorem. Based on this intuition, Alfarano et al. (2005) derive the following approximation for the stochastic process of Eq. (2.10):

$$\begin{aligned} z_{t+\Delta t} &= z_t + (\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z_t)b\Delta t + \sqrt{2b\Delta t(1 - z_t)z_t} \cdot \lambda_t, \\ &= z_t + (\varepsilon_1 + \varepsilon_2)(\bar{z} - z_t)b\Delta t + \sqrt{2b\Delta t(1 - z_t)z_t} \cdot \lambda_t, \end{aligned} \tag{2.19}$$

where λ_t is an iid normally distributed random variable, and $\bar{z}_t$ is defined as $\varepsilon_1/(\varepsilon_1 + \varepsilon_2)$, which is the mean of the process itself. The macroscopic time scale Δt is proportional to N^2, i.e., $\Delta t \sim N^2 \Delta t_0$.[14] The process of Eq. (2.19) is characterized by a linear mean reverting component and a heteroskedastic random term, conditional to the value of z_t. Given the parabolic dependence of the diffusion function, values of z_t close to 0.5 generate higher fluctuations than when z_t is close to the boundaries of its range. This dependence generates heteroskedastic fluctuations in the time series of Δz_t, which may resemble those illustrated in Fig. (2.1). The main justification in using the continuous approximation of Eq. (2.18) and its associated discrete version of Eq. (2.19) lies in the possibility of clearly understanding the connection between the dynamical properties of the generated time series and the structure of the stochastic process. It is evident that such a connection is not trivial to understand from the original Markov chain of Eq. (2.10). Obviously, Eq. (2.19) is an approximation for a large N of the process in Eq. (2.10) with the further restriction that the variable z_t cannot be close to the boundaries $z_t = 0$ and $z_t = 1$. In those regions, the continuous approximation is no longer valid since it may violate the boundary conditions. Hence, the natural boundaries implemented in Eq. (2.10) must be exogenously added to Eq. (2.19). Consequently, in order to simulate the process of Eq. (2.19), we have to add reflecting boundaries at $z_t = 0$ and $z_t = 1$ by hand:

---

[13] The classic book written by Gardiner et al. (1985) accounts for the connection between the Fokker-Planck equation and the associated stochastic differential equation.

[14] See Alfarano et al. (2008) for the details of the derivation.

$$\begin{aligned} &\text{if} \quad z_t > 1 \quad \text{then} \quad \frac{z_{t+\Delta t} + z_t}{2} = 1, \\ &\text{if} \quad z_t < 0 \quad \text{then} \quad \frac{z_{t+\Delta t} + z_t}{2} = 0, \end{aligned} \tag{2.20}$$

which are equivalent to a reflection around the edges of the domain of z_t, $z_t = 1$, and $z_t = 0$, respectively. One of the advantages of the Langevin equation is that it is relatively simple to estimate via maximum likelihood, since the conditional probability density function of $z_{t+\Delta t}$ given z_t is a Gaussian with mean $z_t + (\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z_t)b\Delta t$ and standard deviation $\sqrt{2b\Delta t(1 - z_t)z_t}$. However, note that using the results from the theory of stochastic processes, it can be shown that the stochastic discrete Eq. (2.19) can be characterized by an equilibrium distribution from Eq. (2.16).

Finally, Alfarano et al., 2005 derive the autocorrelation function (ACF) through the Langevin equation by using a recursive method[15], leading to an exponential autocorrelation function:

$$C_z(t) = e^{-b(\varepsilon_1+\varepsilon_2)t}. \tag{2.21}$$

Summarizing, the model of Eq. (2.19) is characterized by a beta equilibrium distribution, a mean reverting property with heteroskedastic persistent fluctuations and an exponential autocorrelation function. Furthermore, an additional advantage in using Eq. (2.19) derived from the continuous approximation of Eq. (2.10) lies in the possibility of employing the maximum likelihood technique in order to estimate the underlying parameters.

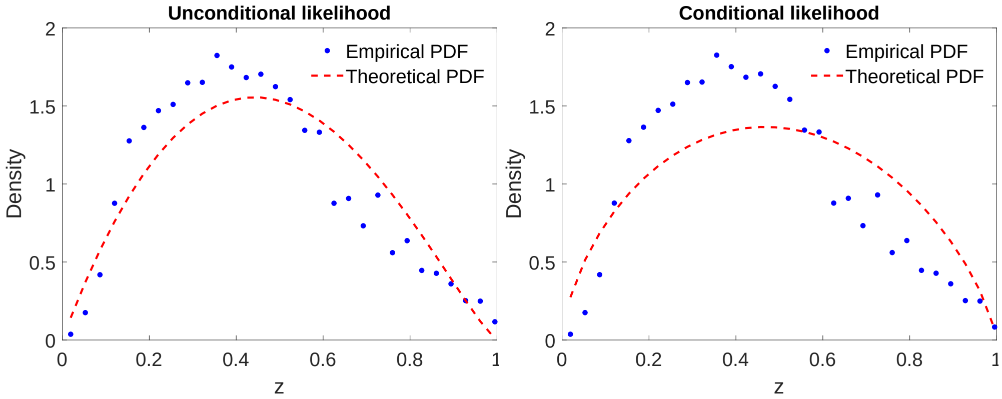
2.5 Validation of the model and estimation of its parameters

In this section, we validated the Kirman model by comparing its theoretical predictions to the empirical properties of the sentiment index defined in Eq. (2.9). We used the time series of the sentiment index based on the 208 stocks of the S&P 500 that have been listed in this stock market from 01/01/1981 to 01/06/2018. To compute the sentiment index we fixed the parameter of the EMA L=100. We show in Section (2.5.2) the sensitivity of the parameter values of the model to the choice of L.

---

[15] Eq. (2.21) can be also computed using the FPE or the Markov chain of Eq. (2.10).

Figure. 2.2: Probability density function of the sentiment index compared to the theoretical distribution, given the estimates of each method.



2.5.1 Estimation of the parameters of the Kirman model

We estimated the three parameters (ε_1, ε_2, and b) that characterize the conditional dynamics of z_t based on Eq. (2.19) using the maximum likelihood method (conditional likelihood).[16] Alternatively, we estimated the parameters ε_1 and ε_2 using the maximum likelihood method based only on the unconditional distribution (unconditional likelihood) from Eq. (2.17) (for the details of this methodology, see Alfarano et al., 2005). The parameter b, which governs the timescale of the process, does not enter in the determination of the unconditional distribution and, therefore, cannot be estimated with the last method. The comparison of the two methods allowed us to evaluate the goodness of the Langevin equation in describing the evolution of the empirical time series. Asymptotically, the two methods should provide the same estimates for the parameters ε_1 and ε_2. A significant discrepancy between the two methods would indicate a limitation of Eq. (2.19) in describing the conditional properties of the sentiment index.

In Table 2.2 and Fig. (2.2), we show the estimations of ε_1, ε_2 and b, and the corresponding fit of the unconditional distribution. Comparing the values of the Chi-squared statistics, we observed that the unconditional likelihood gave a better fit of the empirical probability density of the sentiment index than the corresponding parameters estimated using Eq. (2.19). We observed, in fact, the presence of a clear asymmetry in the empirical distribution, while the values of ε_1 and ε_2 estimated with the conditional likelihood point to a symmetrical distribution. Such a poor performance of the conditional likelihood, compared to the estimation based solely on the unconditional distribution, revealed a certain degree of incoherence between the process of Eq. (2.19) and the evolution of the empirical time series.

---

[16] The value of the parameter b is expressed in units of Δt. We set $\Delta t = 1$ day.

Table 2.2: Estimated parameters, ε_1, ε_2 and b for the sentiment index.

Method	ε_1	ε_2	b	Chi-squared
Unconditional likelihood	1.99 ± 0.07	2.25 ± 0.06	-	618
Conditional likelihood	1.64 ± 0.16	1.74 ± 0.14	0.0055 ± 0.0001	1004

In order to identify the origin of such a discrepancy, we eliminated from the time series the large values of z_t, i.e., $z_t > 0.95$. In Table 2.3 and Fig. (2.3), we show the results of the estimation procedure excluding those events. Interestingly, considering the standard errors of both estimates, the estimated values from the conditional and the unconditional likelihood were not statistically different. From our exercise, we can conclude that the empirical realizations of z_t at the upper boundary were much more persistent than what was expected under the Kirman model.

Fig. (2.4) shows the empirical and simulated indices with their corresponding daily changes for the first 2000 observations for better visibility. The synthetic series was simulated using the Langevin equation with the parameters from Table 2.3. Both time series exhibit swings of pessimism and optimism with similar characteristics. The time series of the empirical changes of the index and the corresponding simulated time series show heteroskedastic fluctuations. We detected volatility clustering in the empirical data by means of the ARCH (autoregressive conditional heteroskedastic) test (Engle, 1982) and McLeod-Li test (McLeod and Li, 1983) with p-values extremely close to zero in both cases. In the case of the simulated data, we have used the parameters of the Langevin equation $\varepsilon_1 = 2.22$, $\varepsilon_2 = 2.86$, and $b = 0.0054$ to compute the 500 Monte Carlo simulations. The medians of the p-values for the two tests were 0.0023 and 0.0032, respectively. Thus, the model captured the volatility clustering phenomenon in the time series of the sentiment index.

Figure. 2.3: Probability density function of the sentiment index compared to the theoretical distribution, given the estimates of each method and excluding extreme events ($z_t > 0.95$).

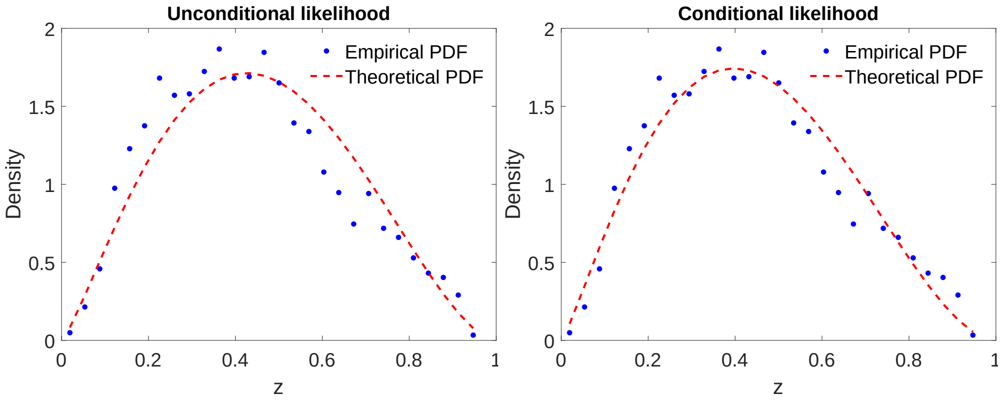
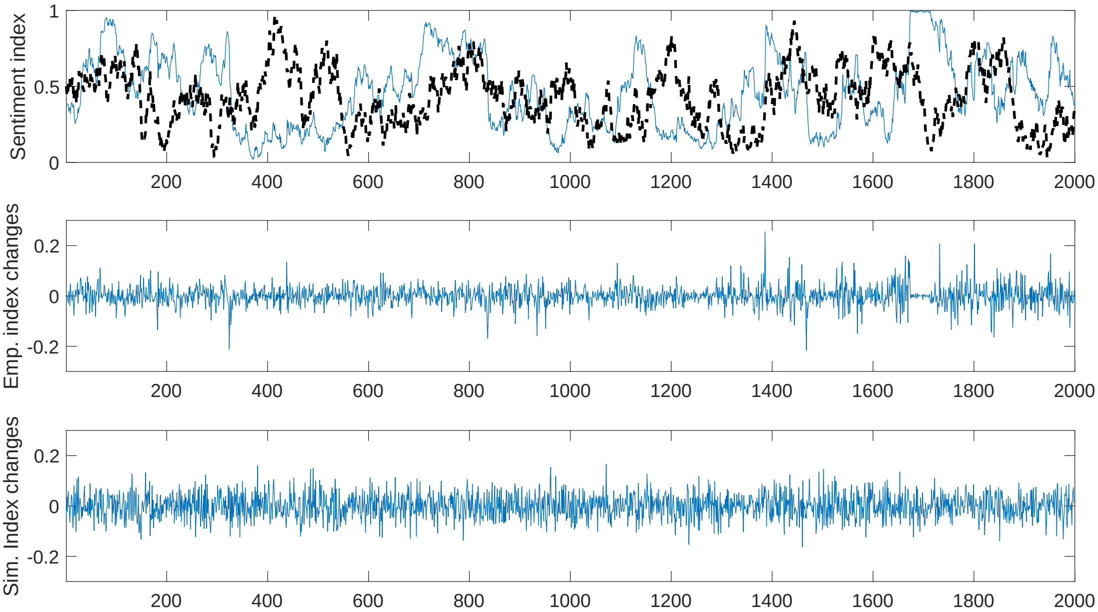


Table 2.3: Estimated parameters, ε_1, ε_2 and b for the sentiment index, excluding extreme events ($z_t > 0.95$).

Method	ε_1	ε_2	b	Chi-squared
Unconditional likelihood	2.28 ± 0.09	2.72 ± 0.10	-	336
Conditional likelihood	2.22 ± 0.18	2.86 ± 0.22	0.0054 ± 0.0001	279

Figure. 2.4: Empirical index (continuous line) and simulated sentiment index (dashed line). The Langevin equation was used with the parameters of Table 2.3 to simulate the time series.



To better capture the similarities of the empirical index and the simulated series, in Fig. (2.5), we show box plots that report some moments (mean, median, and standard deviation) obtained from the 500 Monte Carlo simulations of Eq. (2.19) with the parameters estimated with the conditional likelihood without extreme events. As can be observed, the moments of the empirical time series are in line with those of the simulated time series at the 95% confidence level.

Figure. 2.5: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added.

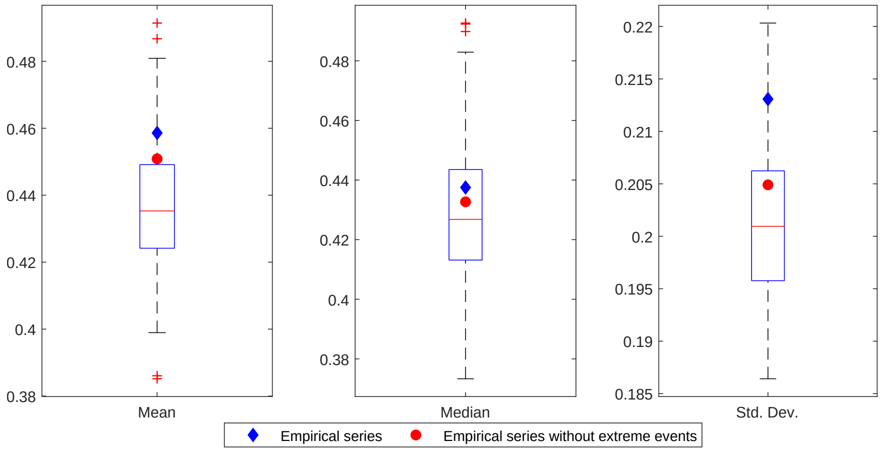
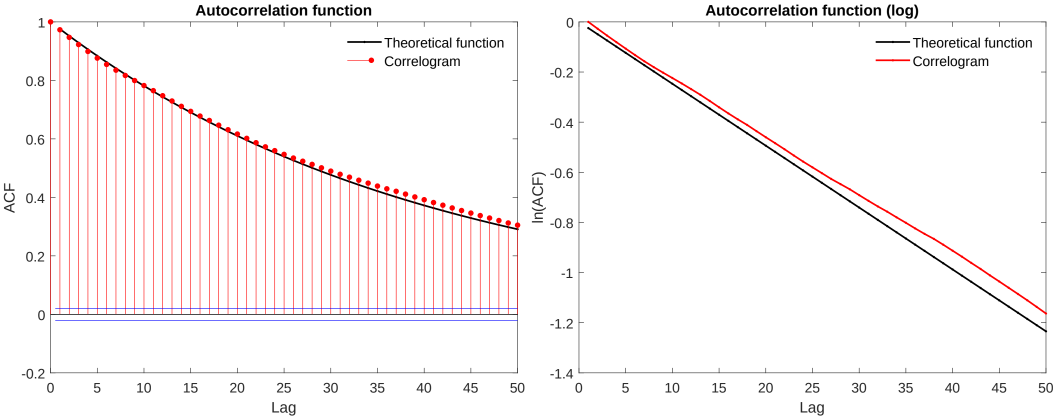


Figure. 2.6: Correlogram of the empirical sentiment index z_t as compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 2.3.



In order to analyze to what extent the model captured well the time evolution of the empirical data, we considered the autocorrelation function. Fig. (A.34) shows the fit of the correlogram of the empirical time series of the index z_t as compared to the theoretical autocorrelation function from Eq. (2.21), which turned out to be particularly accurate. We concluded that the herding model could satisfactorily describe the degree of persistence of the time series of the sentiment index z_t.

2.5.2 Choice of the parameter L of the EMA

In this section, we examine the sensitivity of the estimation of the main parameters of the model to different values for L. We estimated the parameters using the conditional likelihood, changing the value of L to compute the sentiment index from Eq. (2.9). In particular, we considered a plausible range of values for L from 10 to 200 days, based on the moving average chartist techniques typically applied in the analysis of financial data. As can be observed in Fig. (2.7), both ε_1 and ε_2 did not exhibit large deviations by increasing L, thus the unconditional distribution was robust to the choice of L. On the contrary, the herding parameter b decreased significantly the higher the value of L. Interestingly, the choice of L affected the level of fluctuations of the increments of the variable z_t, as shown in Fig. (2.7). Employing Eq. (2.19), we show that a direct connection exists between the magnitude of the fluctuations of z_t, measured as their variance $E[\Delta z^2]$, and the parameter b:

$$E[\Delta z^2] = (\varepsilon_1 + \varepsilon_2)^2 \cdot \mathrm{Var}[z] b^2 \Delta t^2 + 2b\Delta t E[z(1-z)] \approx \frac{b\Delta t}{2}. \tag{2.22}$$

To obtain Eq. (2.22), we have neglected the term proportional to $b^2\Delta t^2$, and we have considered the approximation $\varepsilon_1 \approx \varepsilon_2$, so that $E[z(1-z)] = 1/4$. The dependence of the parameter b from L stems directly from an equivalent dependence of the influence in the level of fluctuations of Δz as a function of L.[17] Fig. (2.7) shows that the scaling of Eq. (2.22), which is a direct consequence of the Langevin equation, accurately describes the empirical relation between b and the fluctuations of z_t. Given that the estimated value of the herding parameter b varies with L, the decay rate of the autocorrelation function will also change, see Eq. (2.21). We wanted to verify if the good fit of the correlogram from Fig. (A.34) was just related to the particular choice L=100. In order to study the robustness of the fit of the correlogram for different choices of L, we analyzed whether the correlogram of the sentiment index was well approximated by the theoretical ACF, described in Eq. (2.21), for any value of L. To do so, we computed the decay rate of the ACF from Eq. (2.21), i.e., $b(\varepsilon_1 + \varepsilon_2)$, using the estimated values of the parameters for a given L. Thus, this decay rate is compared to the slope (β_L) of the correlogram, which is computed by means of an OLS regression:

$$\ln(ACF_L(t)) = \alpha_L - \beta_L \, t, \tag{2.23}$$

where t denotes the lags, and β_L is the empirical decay given L. [18] Fig. (2.8) shows that the decay rate computed using the estimated values and the empirical rate β_L were approximately equal for the different choices of L. We can infer that our model provided a good fit of the autocorrelation of the sentiment index independently of the choice of L.

Given these results, we can contend that the parameter L did not alter the parameters of the unconditional distribution. Moreover, despite the fact that b changes with L, the correlogram was in line with the theoretical prediction, thus the ACF was also robust to the choice of the parameter L. We have decided to fix L=100, since it was a sufficiently long timescale compared to a single day and not too long with respect to the duration of the different phases of the market.

---

[17] More rigorous calculation of the term $E[z(1-z)]$ does not change this line of reasoning.

[18] We employ a maximum of 20 lags in the regression. There are not significant differences if we use a number of lags equal to 30, 40 or 50.

Figure. 2.7: Estimated parameters ε_1, ε_2, and b according to different values of L. The scaling between the variation of $E[\Delta z^2]$ as a function of L and the parameter b is reported (dotted black line).

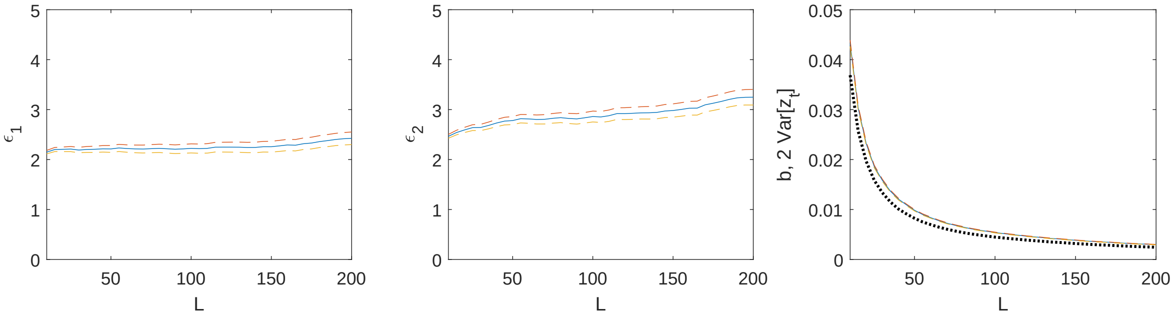
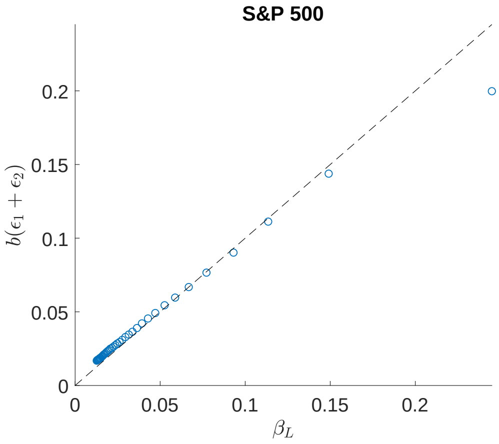


Figure. 2.8: Scatter plot of $b(\varepsilon_1 + \varepsilon_2)$ and β_L using different values of L (from 10 to 200 with an increment of 5). We can appreciate a deviation from the 45-degree line just for very low values of L, which corresponds to large values of β_L.



2.6 Empirical application: Early warning indicator

The sentiment of investors in financial markets has been analyzed from many different approaches in a sizeable empirical literature often with the objective of detecting bubbles with their subsequent financial crises. Some examples are the index cohesive force (Kenett et al., 2011; Kenett et al., 2012), power law distributions (Kaizoji, 2006; Mizuno et al., 2016), the leverage of the banking sector (Adrian and Shin, 2009), cross-correlations (Podobnik et al., 2009), the CAPE (cyclically-adjusted price-to-earnings) ratio (Shiller, 2015) and regime-switching approaches (Preis et al., 2011), among other studies. We aim now to employ the sentiment index developed in the previous sections to detect potential optimistic phases in the market, with the objective of providing an early warning indicator of possible downturn periods. A direct use of the indicator z_t to detect possible downturn periods would be based on the simplistic premise that the time evolution of the stock market index is a "perfect mirror" of the waves of optimism and pessimism of investors. The connection between the sentiment of the investors, as measured by our proxy, and the prices of stocks is much more complex than a mere one-to-one relationship. As a consequence of that, we can observe visually that the dynamics of the variable z_t turns out to be too volatile to be employed as a direct measure of the turning points of the market. Relying on the value of z_t to detect an early warning indicator for the market phase would in fact suggest too many false positive turning points. Therefore, even though the Kirman model had proved to be a valuable tool to reasonably describe the dynamics of the sentiment index, we had to develop a filtering technique in order to extract a more readable signal from the raw time series of the sentiment index z_t.

To extract the information from the raw index, we rely on the battery of analytical tools introduced in Section 2.4. An extreme value of z_t (either close to 0 or 1) can be, in principle, an outcome of a symmetric transition probabilities of the stocks between the two states, i.e., when $\varepsilon_1 = \varepsilon_2$ in Eq. (2.10). Extreme events of z_t can then be observed without being an indicator of a systematic bias in the underlying sentiment of the traders, i.e., when $\varepsilon_1 \neq \varepsilon_2$. Instead, a persistence asymmetry of the two coefficients would systematically favor extreme values of z_t. For instance, the case $\varepsilon_1 > \varepsilon_2$ would favor values of z_t close to 1, contemporaneously reducing the probability of observing the opposite extreme case. We conjecture that a persistent optimistic or pessimistic phase in the sentiment of the traders can anticipate critical movements in the market prices. An imbalance between the two values of ε_1 and ε_2 can be generally employed as a more reliable measure for a systematic predisposition of the traders to pessimism or optimism and, therefore, used as the basis of an early warning indicator.

In order to aggregate the information of the sentiment index during a given time interval and identify the phase of the market, we estimated the parameters $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ through a rolling window of 750 days (three years).[19] We implemented a recursive estimation for $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ using the observations of z_t from $t - 750$ to t. Considering that Eq. (2.19) is the law of motion of the sentiment index, we estimated the parameters $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and b_t using the conditional maximum likelihood method for each rolling window.[20] We can now employ as a summary indicator for the phase of the market the difference between the value of the two estimated parameters[21]:

$$\Lambda_t = \varepsilon_{2,t} - \varepsilon_{1,t}. \tag{2.24}$$

We can have three different scenarios: $\varepsilon_{2,t} \approx \varepsilon_{1,t}$, so that $\Lambda_t \approx 0$, which represents a "symmetric" market with approximately the same number of stocks in a pessimistic and optimistic phase; $\varepsilon_{2,t} > \varepsilon_{1,t}$ and $\Lambda_t > 0$ represent a strong bull market, when most of the stock prices are, on average, raising

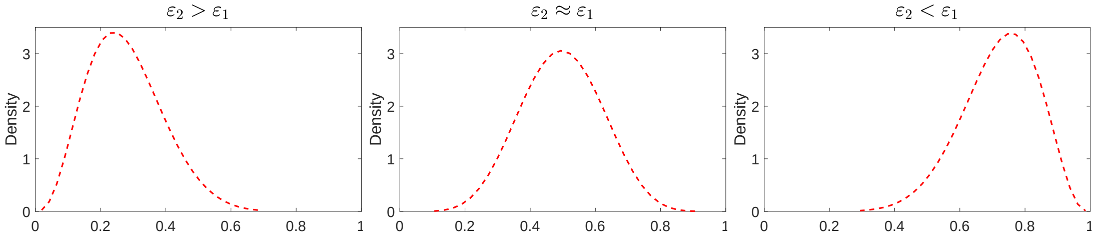
---

[19] The choice of 750 days for the rolling window was based on the length of the standard business cycle, which is around 7-11 years. More specifically, given that we wanted to observe the transitions of the different market phases, we used a shorter time window equal to 3 years, i.e., 750 days. On the other hand, shortening the length of the rolling window would affect the precision of the estimates of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$.

[20] We use all data in each window without removing extreme values. In this case, using short running windows, we did not observe that extreme events affected the fit of the distribution. In fact, we observed similar results even excluding those values of $z_t > 0.95$.

[21] We have employed the alternative measure $\tilde{\Lambda}_t = \varepsilon_{2,t}/\varepsilon_{1,t}$. *Mutatis mutandis*, the results of the early warning indicator based on the ratio of the two parameters ($\tilde{\Lambda}_t$) were consistent with respect to the results of the early warning indicator Λ_t.

Figure. 2.9: Three possible scenarios for the stock index based on ε_1 and ε_2. In the first one ($\varepsilon_1 < \varepsilon_2$), there is optimism in the market. In the second one ($\varepsilon_1 \approx \varepsilon_2$), there is no dominant mood. In the last one ($\varepsilon_1 > \varepsilon_2$), there is pessimism in the market.



at the same time; finally, $\varepsilon_{2,t} < \varepsilon_{1,t}$ and $\Lambda_t < 0$ underline the existence of a bear market, when most of the stock prices are, on average, falling simultaneously, as shown in Fig. (2.9).

Fig. (2.10) shows the evolution of the indicator Λ_t using a 100-day EMA and a time interval of 750 days rolling window for the estimation of parameters. To obtain a sharper characterization of the phase of the market, we considered certain levels of "excess asymmetry", establishing a threshold equal to the 90^{th} percentile, which are represented by dark gray areas, for bull phases, and a threshold equal to the 10^{th} percentile, which are represented by light gray areas, for bear markets.[22] We show the sequence of the S&P 500 index and its detrended series.[23] Moreover, we plot in Fig. (2.10) the sequence of the 30 most negative weeks in S&P 500 history.

---

[22] We computed the two thresholds using the entire data set at our disposal. However, as can be observed in the Appendix (A.3), we obtained very similar results using past data to compute recursively the value of the two thresholds.

[23] Given that we were not analyzing exactly the S&P 500, since we only had 208 stocks at our disposal, we show the corresponding index of the available stocks ($\bar{S}_t$), defined as the exponential cumulative sum of market returns at each time t,

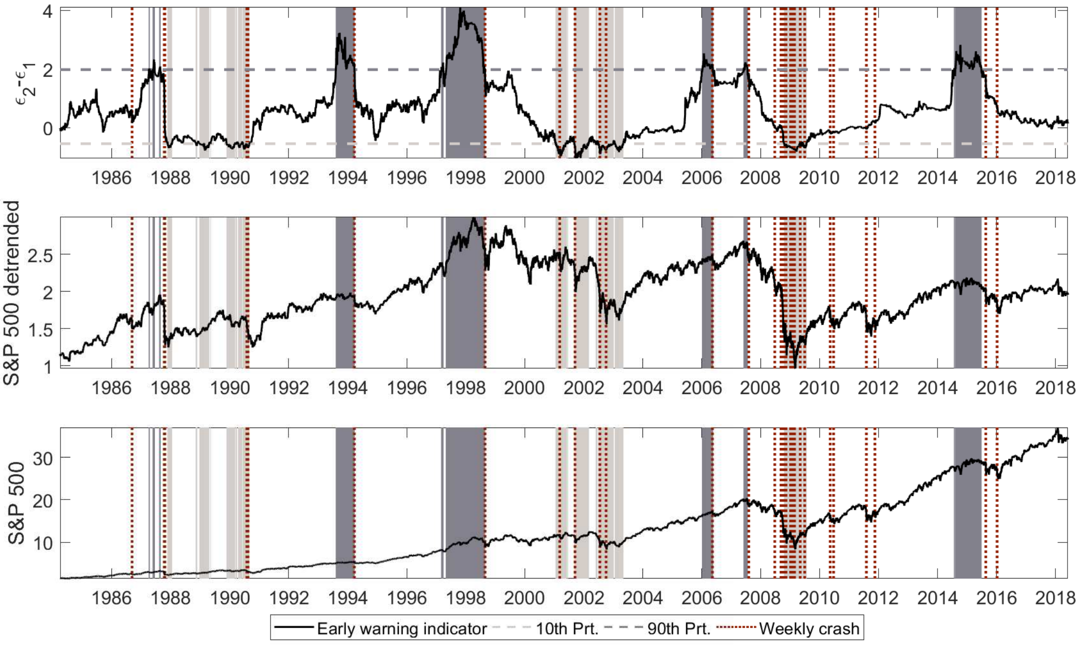
$$\bar{S}_t = e^{mr_t \cdot t}, \tag{2.25}$$

where mr_t denotes the market return at t defined as:

$$mr_t = \frac{1}{N}\sum_{i=1}^{N} r_{i,t}, \quad N = 208, \tag{2.26}$$

and $r_{i,t}$ denotes the return of each stock i. We also report the S&P 500 detrended index, which was calculated following the same procedure using the returns calculated by means of the detrended prices ($p_{i,t}$).

Figure. 2.10: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. S&P 500 index.



Note. Bull market phases and the subsequent negative events: 1987 (Black Monday), 1994 (tightening monetary policy), 1997-1998 (Russian financial crisis), 2006 (tightening monetary policy), 2007 (sub-prime mortgage crisis) and 2015-2016 (weakness of the Chinese economy). Bear market phases: 1987-1990 (Black Monday and the Gulf War), 2001-2003 (burst of the dot-com bubble), 2008-2009 (burst of the housing bubble).

Fig. (2.10) shows the different periods in which we observed the sequence of optimistic and pessimistic phases of the market, along with extreme negative weekly events. The objective of this figure was to evaluate the performance of the early warning indicator with empirical data, that is:

- We expected to observe at least a downtrend period (underlined by negative weeks in the stock index) shortly after the early warning indicator had detected a bull market phase (around its 90^{th} percentile).
- We expected higher prices after bear market phases (around the 10^{th} percentile) due to the low prices of stocks during extreme pessimistic phases.

The detection of the bull and the bear market phases can be exploited by investors to set a long-term trading strategy based on the prospect of future prices. In particular, a possible trading strategy detecting a bull market phase would be based on the progressive sale of stocks, or the purchase of put-options, in order to possibly avoid or smooth the downtrends or persistent downturns. In the opposite scenario, a bear market would give the opportunity to accumulate stocks due to the low stock prices as a consequence of the pessimism in the market. The early warning indicator can be also considered as a valid instrument for policy-makers to set more efficient policies in order to avoid the contagion from the financial to the real side of the economy.

In Fig. (2.10), we identify the following optimistic market phases: 1987, 1994, 1997-1998, 2006, 2007, and 2014-2015. In favor of the proper performance of the indicator, all these optimistic phases of the market preceded at least an extreme negative weekly return, which were triggered by well-identified events. In chronological order, the first optimistic period appeared some months before the Black Monday (October, 1987), from April to August 1987. In fact, we observed in our sample the most negative event with a weekly return equal to -18.92% due to the errors of the computerized trading system, which triggered sell orders of an enormous block of stocks as prices fell (Waldrop, 1987). The second optimistic period was detected from March 1997 to August 1998, whose end was found on the last week of August 1998 as a result of the Russian financial crisis (Buchs, 1999) with a negative weekly return of -7%. Another extreme optimistic period in the S&P 500 could be identified, from January to May 2006, in which a downtrend stopped the increase in prices due to a new prospect of further tightening of the monetary policy in USA, Europe, and Japan (IMF, 2006). Despite this small decrease in prices, the optimism remained in the market with another peak of optimism from June to July 2007. Interestingly, our indicator included the previous maximum price on the US bear market (2007-2009), triggered by the subprime mortgage crisis in October 2007 (Demyanyk and Van Hemert, 2009). The last optimistic period observed in our sample was detected during 2014-2015, which arose from the unconventional expansionary monetary policy implemented by the central banks, known as quantitative easing, generating an increase in prices without precedence that currently still remains. This surprising scenario has only been interrupted by the stock market sell-off in 2015, as a result of fears about the Chinese economy and investors' concerns over the end of the quantitative easing (Bendini, 2015; Sornette et al., 2015). Our empirical analysis stressed the validity of the early warning indicator since the six extreme bull market phases identified by the early warning indicator systematically preceded a downtrend of the market.

Focusing on the second hypothesis, Fig. (2.10) shows generally three extreme pessimistic periods (levels around the 10^{th} percentile): after the Black Monday (1987-1990), the burst of the dot-com bubble (2001-2003), and the burst of the housing bubble (2008-2009). As can be observed from the S&P 500 index, there was a considerable increase in prices after the detection of the pessimistic phase. In particular, we noted three increases in prices: after the Gulf War (1990), after the burst of the dot-com bubble (2003), and after the burst of the housing bubble (2005). Thus, persistent downturns can be used to progressively accumulate call-options or stocks since they are expected at the end of the pessimistic phase of the market.

2.7 Robustness analysis

So far, we have examined the statistical properties of our proxy for the mood of investors estimated using the collective movement of S&P 500 stocks. However, any stock market can be potentially driven in some measure by the sentiment of agents. Different statistical properties of the evolution of the sentiment index could characterize diverse stock markets due to factors like country's features (Anderson et al., 2011, Karlsson and Nordén, 2007) proportion of institutional investors, firm size (Ferreira and Matos, 2008) or quality of reports (Biddle et al., 2009) among other aspects. Given these differences among markets, it is interesting to study whether the sentiment index and the early warning indicator could be meaningfully applied to other financial markets regardless of the apparent differences among countries and indices.

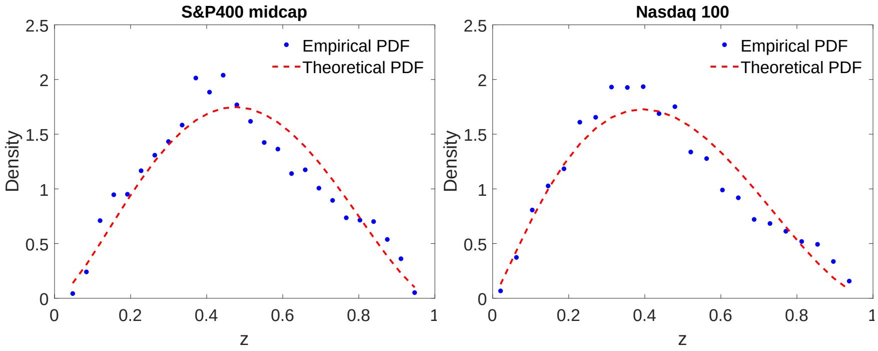
In order to do so, we repeated the procedure to compute the sentiment index of Eq. (2.9) for two alternative data sets: US stock markets and worldwide stock markets. On the one hand, by examining different stock markets in the same country, i.e., assuming that the country features were invariant, we could study whether aspects like firm size, proportion of institutional investors, liquidity, or the specificity of a sector could affect the behavior of the sentiment index. Therefore, we analyzed the S&P 400 midcap index whose companies are smaller than S&P 500 firms with a lower level of liquidity and a clearly different proportion of institutional investors given the absence of analyst recommendations for these companies. We also analyzed the Nasdaq index, whose core business of the companies is focused on the information technology sector.

On the other hand, we studied the indices of six different countries in order to observe whether a different result arose, not only due to stock market characteristics, but also due to country features like cultural or economic conditions. Thus, we examined the following countries and indices: ASX 200 (Australia), TSX (Canada), Euro Stoxx 600 (Europe), Nikkei 225 (Japan), JSE (South Africa), and FTSE 100 (UK).

2.7.1 US stock markets

Investor preference is one of the main features to take into account in order to know who is investing in each type of market. In fact, Ferreira and Matos (2008) show that institutional investors prefer firms that are characterized by being large, well-governed, and with high levels of stock-trading liquidity. In the same line, Aggarwal et al. (2005) find that US funds are more interested in investing in large firms with transparent accounting policies. The relevance of financial reports is highlighted by Biddle et al. (2009), since those firms with higher financial reporting quality suffer less from macroeconomic conditions and deviate less from predicted investment levels. Financial aspects are also crucial, since domestic managers prefer companies with large dividends, low financial distress, and

Figure. 2.11: Probability density function of the US sentiment indices compared to the theoretical distribution. The Langevin equation was used to obtain the estimates, excluding extreme events ($z_t > 0.95$).



low-return variability (Covrig et al., 2006), which are features of firms with high-market capitalization. Considering these characteristics, institutional investors would prefer to invest in companies from the S&P 500 rather than those from the S&P 400 midcap. If we consider the median total market cap, the S&P 500 is five times larger than the S&P 400 midcap with the following market capitalization: 20493.91 and 4178.83 USD millions for the S&P 500 and the S&P 400, respectively.[24]

Despite differences among these market indices, it is possible to observe in Fig. (2.11) and Table 2.4 that the estimates were in line with those obtained for the S&P 500. Moreover, when we used stock markets like the Nasdaq, focused on firms from the information technology sector, we kept observing very similar estimates. Therefore, even with a different type of investor, liquidity, sector, or market cap, our proxy for the sentiment of investors showed very similar statistical properties: volatility clustering in the time series of sentiment index increments and the exponential autocorrelation function of the sentiment index (a detailed statistical analysis is shown in the Appendix A.5).

Table 2.4: Estimated parameters, ε_1, ε_2 and b for the US stock indices. The Langevin equation was used to obtain the estimates excluding extreme events ($z_t > 0.95$).

Stock market	Date	ε_1	ε_2	b	Chi-squared
S&P 400 midcap	1993-2018	2.55 ± 0.23	2.72 ± 0.23	0.0060 ± 0.0002	220
Nasdaq	1993-2018	2.18 ± 0.17	2.82 ± 0.21	0.0088 ± 0.0002	208

2.7.2 Worldwide stock markets

A remarkable feature of worldwide stock markets is the high level of portfolio concentration in domestic markets. This phenomenon, known as "home bias", goes not only against the advantages of international diversification but also many standard asset-pricing models. This traditional feature, which was highlighted in the 1990s by French and Poterba (1991), Cooper and Kaplanis (1994), and

---

[24] The median total market capitalization of the S&P 500 and S&P 400 midcap refers to June 2018.

Tesar and Werner (1995), is still present nowadays despite the many improvements in terms of information channels.[25] Anderson et al. (2011) demonstrate that home bias exists in institutionally managed portfolios from more than 60 countries. In the same vein, Grinblatt and Keloharju (2001) with a database focused on Finland, contend that investors prefer those firms that are nearby with a common language and culture.

Home bias underlines the fact that local agents are investing in those companies and stocks that are present in the corresponding local or domestic financial market. Therefore, we can assume that when analyzing a worldwide stock market, the main proportion of capital is given by local investors. This aspect is of paramount importance for this study since the fact of observing a proper fit of the unconditional distribution of the sentiment index, regardless of the country, implies that we can use our model to describe the optimism and pessimism of agents regardless of the country-specific characteristics. Table 2.5 and Fig. (2.12) show how the Langevin process and its equilibrium distribution provided us with a fairly good description of financial markets from different countries. We obviously observed different values of the estimates given the diverse dynamics of each sentiment index z_t, which was reasonably explained by the different features of each index and, above all, the investors' moods. More specifically, Euro Stoxx 600, JSE All-Share, and FTSE 100 are characterized by a generalized optimism of the investors ($\varepsilon_2 > \varepsilon_1$), which was higher than the optimism observed in the ASX 200 and TSX given the difference between ε_2 and ε_1. A special case was found in the Nikkei 225 given a "quasi" uniform distribution representing a market in which there was not a prevailing (optimistic/pessimistic) mood.

Excluding the latter index, the sentiment index of all the other financial markets seemed to be generally well characterized by an asymmetric uni-modal beta distribution, with a non-monotonic probability density. Nevertheless, Euro Stoxx 600 and TSX showed a worse performance both visually and quantitatively (Chi-squared), compared to the rest of the market, due to an excess optimism that was not completely described by the herding process and its estimates. Despite the fact that the fit of these markets was not always satisfactory, the optimistic mood of investors was roughly captured in both cases.

Finally, also in the case of international markets, we observed heteroskedastic fluctuations of the index increments and an exponential decay of the autocorrelation.[26]

---

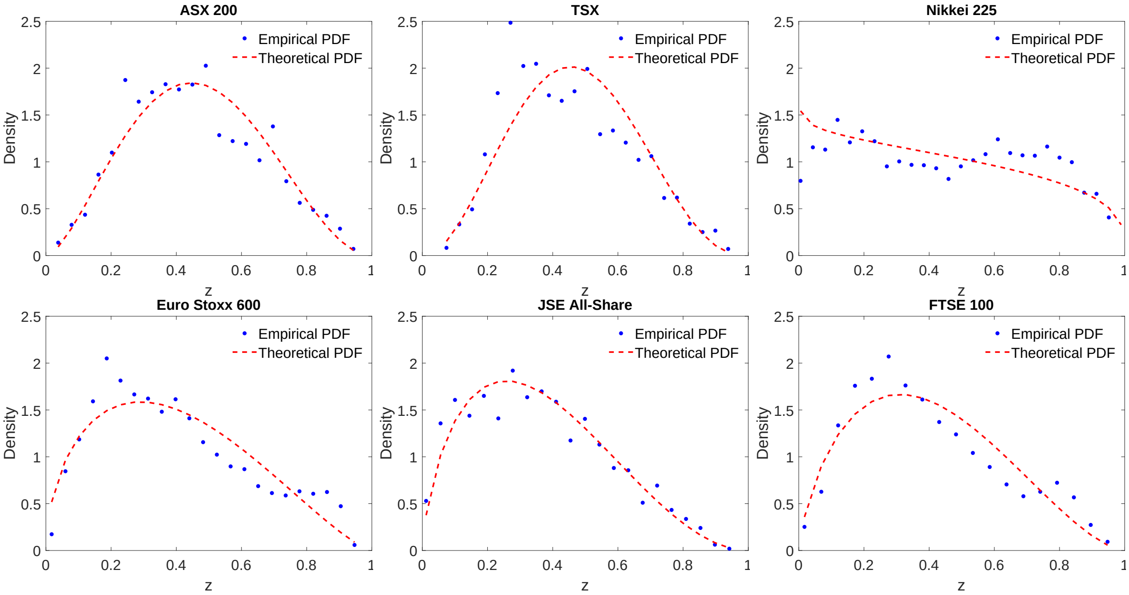
[25] Although Gehrig (1993), Ivković and Weisbenner (2005), and Massa and Simonov (2006) suggest that investors can take advantage of exploiting local information, Seasholes and Zhu (2010) state that investors do not have value-relevant information on local companies.

[26] As with the S&P 500, we used the ARCH test (Engle, 1982) and McLeod-Li test (McLeod and Li, 1983) to study volatility clustering in the empirical and simulated indices. The results showed that most of the indices were characterized by volatility clustering. The only exception seemed to be the TSX index. We also analyzed the sensitivity of the estimated parameters to the choice of L, as done for the S&P 500. The results are included in the Appendix A.5.

Table 2.5: Estimated parameters, ε_1, ε_2 and b for worldwide stock indices. The Langevin equation was used to obtain the estimates excluding extreme events when $z_t > 0.95$.

Index	Country	Date	ε_1	ε_2	b	Chi-squared
ASX 200	Australia	1993-2018	2.68 ± 0.23	3.09 ± 0.26	0.0060 ± 0.0001	261
TSX	Canada	1993-2018	3.18 ± 0.33	3.62 ± 0.36	0.0036 ± 0.0001	530
Nikkei 225	Japan	1993-2018	0.96 ± 0.10	1.29 ± 0.13	0.0086 ± 0.0002	354
Euro Stoxx 600	Europe	1993-2018	1.56 ± 0.18	2.37 ± 0.25	0.0050 ± 0.0001	454
JSE	South Africa	2000-2018	1.70 ± 0.21	3.00 ± 0.40	0.0081 ± 0.0001	118
FTSE 100	UK	1993-2018	1.74 ± 0.15	2.63 ± 0.20	0.0081 ± 0.0002	365

Figure. 2.12: Probability density function of the worldwide sentiment indices compared to the theoretical distribution. The Langevin equation was used to obtain the estimates excluding extreme events ($z_t > 0.95$).



2.7.3 Global financial village

The fact that financial financial markets have become a coupled complex system is not surprising given the growing correlation between stock market indices (Mantegna and Stanley, 1996; Forbes and Rigobon, 2002), as it is underlined by Kenett et al. (2012) when identifying that US, UK, German and Japanese indices are highly interconnected. Consequently, the "global financial village is highly prone to systemic collapses which can sweep the entire village" (Kenett et al., 2012). In the same line, by means of our early warning indicator, we observe how all the stock indices were behaving in a similar manner regardless of the country-specific characteristics of each index. The effect of the Russian financial crisis in 1998, the dot-com crash in 2001-2003, the burst of the housing bubble in 2008-2009 and the downtrend in prices during 2015-2016 (due to the weakness of China economy) are the best example of the consequences of the "global financial village", since to a greater or lesser extent, it was possible to observe those events regardless of the market [see in this section: Nasdaq

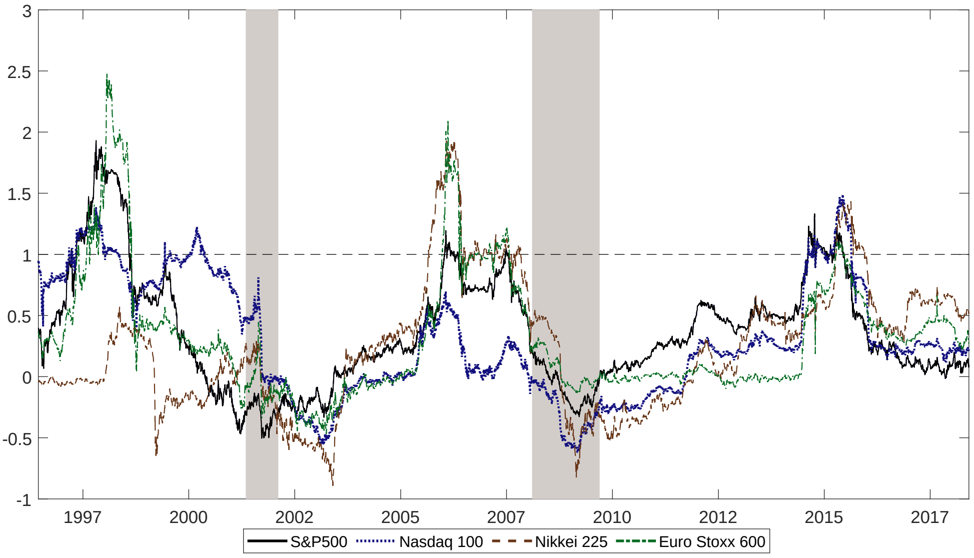
100 (USA), Fig. (2.14); Eurostoxx 600 (Europe), Fig. (2.15) and Nikkei 225 (Japan), Fig. (2.16)].[27] As a result, in favor of the two hypotheses of our early warning indicator, we could identify negative weekly returns after an optimistic market phase had been detected, while a pessimistic market phase was characterized by low prices compared to the future price of the indices. We identified a high concentration of negative weekly returns in all the stock indices after the burst of the housing bubble (2008-2009), along with a lower concentration during the burst of the dot-com bubble (2001-2003), and the downtrend caused by the Chinese economy (2015-2016). To show better the co-movements of the financial markets, we plotted standardized early warning indicators of the S&P 500, Nasdaq 100, Nikkei 225, and Euro Stoxx 600 in Fig. (2.13) along with shaded areas that represent the US Business cycle contractions according to the National Bureau of Economic Research.[28] The standardization has been computed by dividing $j\Lambda_t$ by the market-specific 90% quantile threshold. From Fig. (2.13), we can identify business cycle contractions after worldwide waves of optimism, like the dot-com bubble and the housing bubble, which supports the inter-connectedness of stock markets through the investor's moods. In this regard, it was interesting to observe that a bubble in one specific market was generating financial downturns in other markets that were not involved in such an optimistic scenario. To explain this point, we underline two main periods of recent financial history (the dot-com bubble and housing bubble) using the main stock indices [see Fig. (2.13) and S&P 500 (USA), Fig. (2.10); Nasdaq 100 (USA), Fig. (2.14); EuroStoxx 600 (Europe), Fig. (2.15) and the Nikkei 225 (Japan), Fig. (2.16)]. As can be easily observed, despite the Russian financial crisis in 1998, investors trading in Nasdaq stocks maintained their positive mood giving rise to a bull market from February to May 2000. The burst of the bubble was in March 2000, as we can note due to the continuous negative weeks in Fig. (2.14).

---

[27] In the Appendix (A.4): S&P 400 midcap (USA), Fig. (A.3); ASX 200 (Australia), Fig. (A.4); TSX (Canada), Fig. (A.5); JSE All-Share (South Africa), Fig. (A.6), and FTSE 100 (UK), Fig. (A.7)

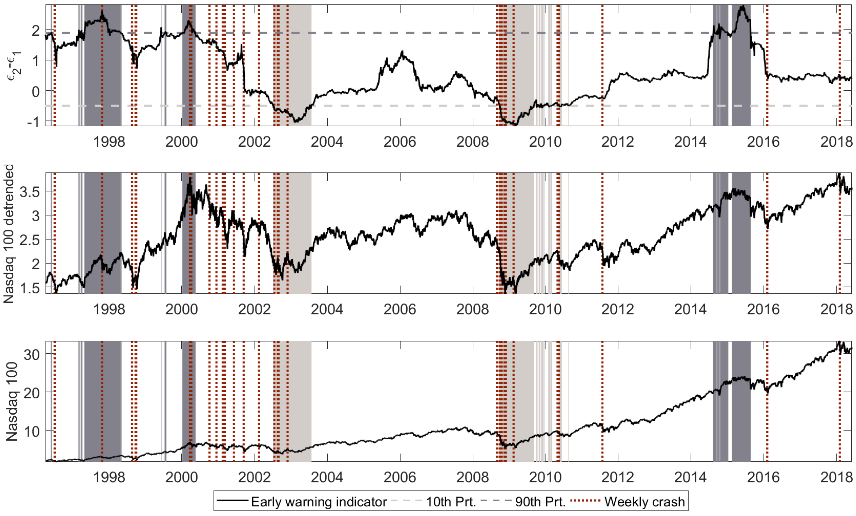
[28] We also show in the Appendix (A.4) all the the early warning indicators ($j\Lambda_t$) computed for all the financial markets considered in the paper, Fig. (A.2).

Figure. 2.13: Standardized early warning indicators of the S&P 500, Nasdaq 100, Nikkei 225, and Euro Stoxx 600. Light areas represent the US Business Cycle contractions according to the National Bureau of Economic Research.



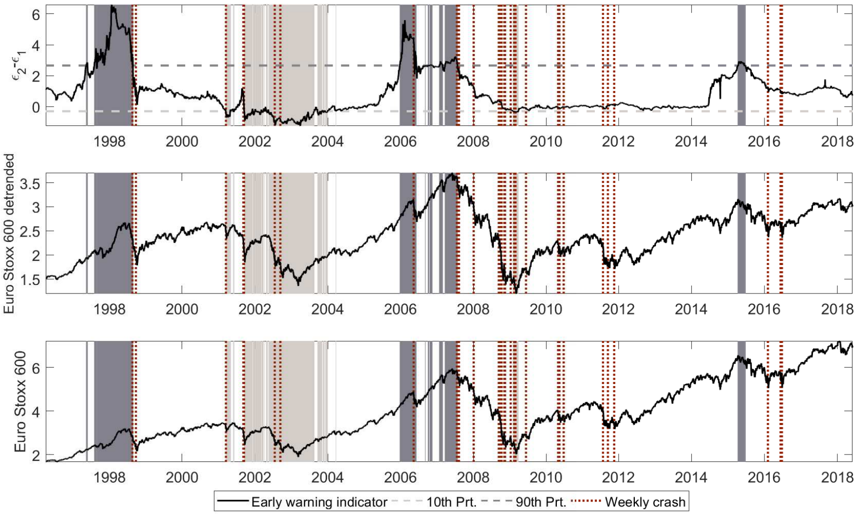
Surprisingly, we did not observe the previous optimism to the burst in the rest of indices, like the S&P 500 or Euro Stoxx 600. In fact, the Nikkei 225 did not even have an optimistic period before the Russian financial crisis. However all the markets suffered from the burst of the dot-com bubble with a wave of pessimism in the ensuing years, mainly from 2001. In other words, the bubble originated in a particular market (Nasdaq) due to the optimism of their traders, and the resulting herding phenomena gave rise to a pessimistic scenario in very different markets [S&P 500, Euro Stoxx 600, Nikkei 225, and the rest of the markets in Appendix (A.4)] whose traders were not so optimistic. Interestingly, we could observe during the housing bubble the opposite scenario. A wave of optimism dominated most of the stock markets like the Euro Stoxx 600 and Nikkei 225 indices, which were affected by the same negative events in May 2006 and October 2008 as in the S&P 500 index. Nevertheless, the Nasdaq 100 did not have an optimistic period nor was it affected by the same negative events during this period. At any rate, and in the same line as the dot-com bubble, Nasdaq and all the markets suffered from the housing crash, even though there was not an optimistic phase in the Nasdaq index, i.e., despite the fact that there was not a bubble in a specific market, it will be still affected by the mood of other markets due to financial contagion. Therefore, the early warning indicator also allowed us to describe the different behavior of each stock index with more detail, even though all of them are connected, as can be observed due to the effect of the financial downturns.

Figure. 2.14: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. Nasdaq 100 index.



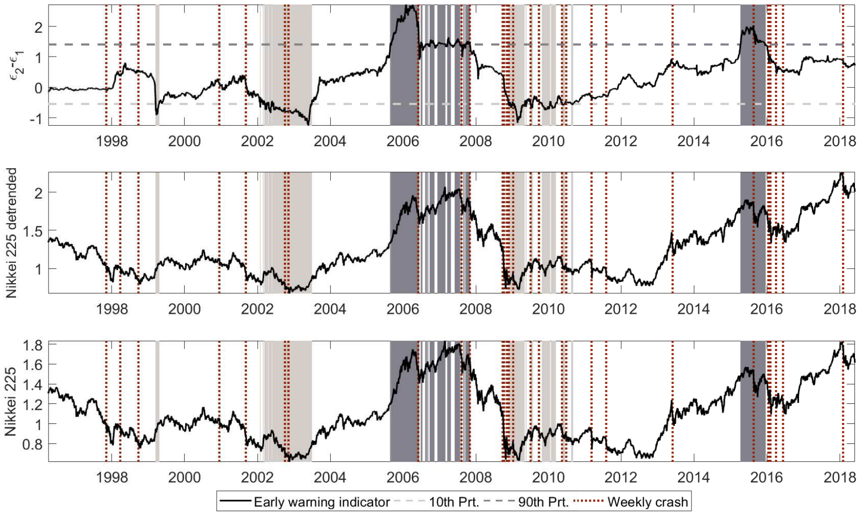
Note. Bull market phases and the subsequent negative events: 1997-1998 (Russian financial crisis), 2000 (burst of the dot-com bubble) and 2015-2016 (weakness of the Chinese economy). Bear market phases: 2002-2003 (burst of the dot-com bubble) and 2008-2010 (burst of the housing bubble).

Figure. 2.15: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. Euro Stoxx 600 index.



Note. Bull market phases and the subsequent negative events: 1997-1998 (Russian financial crisis), 2006 (tightening monetary policy), 2007 (subprime mortgage crisis) and 2015 (weakness of the Chinese economy). Bear market phases: 2001-2003 (burst of the dot-com bubble) and 2009 (burst of the housing bubble).

Figure. 2.16: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. Nikkei 225 index.



Note. Bull market phases and the subsequent negative events: 2006 (tightening monetary policy), 2007 (subprime mortgage crisis) and 2015-2016 (weakness of the Chinese economy). Bear market phases: 2002-2003 (burst of the dot-com bubble) and 2008-2010 (burst of the housing bubble).

2.8 Conclusion

Inspired by the Bank of America Merrill Lynch global breath rule, we have introduced an index of financial investor sentiment based on the collective movements of the stocks in a given financial market. The underlying hypothesis was that such an index reflects and captures the collective behavior of the investors influencing each other in waves of optimism and pessimism transmitted by the social interactions. The index z_t aggregates the state of each single stock, which depends on the relative position of its price with respect to its EMA. The time evolution of the index can be successfully described by the herding model introduced by Kirman (1991, 1993). In particular, the unconditional distribution of the sentiment index, the heteroskedastic fluctuations of the time series of its increments and the autocorrelation function match satisfactorily the analytical properties of the herding model. Based on the herding model and the sentiment index, we introduced an early warning indicator, using the distributional asymmetry of the sentiment index computed in a rolling window. Our early warning indicator can clearly identify the optimistic and pessimistic phases of the market. Our results were robust when applying the early warning indicator to other indices of the US financial markets and the financial market indices of other countries like Japan, Australia, and Canada, among others.

Moreover, our research points out that one of the most celebrated models in the agent-based literature demonstrates once again to be very flexible in describing different aspects of financial data. The Kirman model can successfully describe the behavior of traders and the collective movement of the stocks. The fact that the same model can be employed to account for the dynamics of different entities (traders' expectations and the collective movement of stock in a market) can be considered as a further indirect indication of a connection between the complex behavior of the traders and the market movements. The strength of our approach lies in its analytical apparatus, which allows for a (partial) identification of the connection of the time-series properties of the sentiment index with the structural parameters of the model, i.e., the nature of the interactions among the model constituents and the fluctuations generated at the macro level. However, as a main weakness, we cannot satisfactorily take into account the extreme events in the evolution of the sentiment index, which is partially due to the homogeneity assumption in the influence of the global coupling among stocks instead of the fully connected network implicit in the formalization of Eq. 2.10 (see Alfarano and Milaković, 2009). Ongoing research is focused on the introduction of a more realistic network of influence among the different stocks. The model can also be applied to alternative definitions of market breadth in order to design efficient estimators for the sentiment of the investors and to develop more timing early warning indicators.

As a final comment, note that we do not propose any model that links the sentiment index to the evolution of the price of a stock or the market index. This connection is more complex than a one-to-one relationship. The agent-based financial literature has identified additional important ingredients to account for such a connection. In particular, pointing out the crucial role played by the fundamentalists. Future research will be devoted to incorporating the fundamental information in the dynamics of the early warning indicator.

Chapter 3

Should I herd or should I not? A stochastic herding model for Bitcoin dynamics

3.1 Introduction

In 2008, inspired by the growing number of currency crises, Nakamoto created a means of payment disconnected from central banks and governments: Bitcoin. As many economists, journalists and practitioners claimed, at the beginning, this cryptocurrency was motivated by philosophical reasons: the desire to separate money from the state (see, Feuer, 2013, Krugman, 2018). Nakamoto himself wrote in his white paper: "The root problem with conventional currencies is all the trust that's required to make it work. The central bank must be trusted not to debase the currency, but the history of fiat currencies is full of breaches of that trust" (see, Nakamoto, 2009 and Feuer, 2013). In the wake of Bitcoin's great success, more than 1000 other digital currencies were created. However just a small subgroup of them, made up of 50 currencies, remained functional for more than one year (see, Wei, 2018). The great impact of these currencies, seen both as a means of payment and speculative stocks, did not go unnoticed by the scientific community and investors, who began to carefully analyze them. It was easy to realize that, those objects created to be free from the possibility of interference from malicious governments or banks were, on the other hand, highly manipulable and volatile (see, Dodd, 2018). This appeared especially true for Bitcoin, which proved to be particularly volatile in the face of external shocks driving traders to buy or sell large amounts of the cryptocurrency in short time periods. The initial utopian desire to free this currency from external influences failed, and it was soon understood that the Bitcoin dynamics was determined by the behavior of a few investors. Specifically, the economic literature showed how this cryptocurrency moved due to the effect of herding phenomena normally trigged by few traders. For instance, Gandal et al. (2018)

demonstrated that, during the Bitcoin bubble of 2013, the cryptocurrency was manipulated by two investors who, with their trading activity, caused a significant increase in the Bitcoin price in the Mt.Gox, the Bitsamp, the Bifinex and the BTC-e trading platforms. In this regard, the authors wrote: "It is likely their activity sent a signal to the market and encouraged others to enter and purchase Bitcoins. This may be one of the reasons why their activity could have such a large effect on the Bitcoin price".

The fact that Bitcoin is driven by herding phenomena is now well documented by empirical literature.[1] In fact, many studies have shown the existence of a co-movement between the cryptocurrency price trend and the information published on social media. Many are the mechanisms which have been proven able to influence / manipulate the Bitcoin price dynamics. Part of the literature has focused on the existence of a bidirectional relationship between searched queries and the Bitcoin price, volatility, volume and returns at different time lags (see, Kristoufek, 2013 and Urquhart, 2018). All these studies have highlighted the presence of a "positive feedback", meaning that on-line searches on Bitcoin influence their dynamics (i.e., price, volatility, volume and returns) which, in turn, influence the number of searches. Clearly, positive or negative news also play an important role in the cryptocurrency movement. Specifically, a recent study has shown how positive and negative news have a strong positive impact on currency returns (see, Vidal-Tomás and Ibañez, 2018). Therefore, all these studies have confirmed how the presence of animal spirits can generate strong fluctuations in this currency dynamics *via* waves of optimism and pessimism. More generally, this "mass-uniform" behavior has been proven to be the key ingredient in affecting the currency price oscillation, which seems completely separated from any economic fundamental values, such as unemployment, production or consumption (Baek and Elbeck, 2015).

Some insights into coordination of traders' strategies and large aggregate fluctuations have been provided by theoretical models which have taken into account the interaction among heterogeneous agents (or groups of agents). For example, Lux (1995, 1996, 1998), Brock and Hommes (1998), Licalzi and Pellizzari (2003), LeBaron and Yamamoto (2008) and Chiarella et al. (2009) have studied how coordination phenomena in traders' behavior (for example by following chartist trading rules) or mechanisms of behavioral switching lead to speculative bubbles. Specifically, these studies have introduced some kind of competition among different investors' strategies and a switching mechanism re-addressing traders from the least profitable tactic to the most profitable one. When, at the aggregate level, one strategy dominates the others in popularity, strong prices fluctuations emerge due to the coordination in agents trading positions.

A similar mechanism capable of triggering large price fluctuations is the phenomenon known as herding which occurs in presence of information externalities, when agents' private information is swamped by the information derived from directly observing others' actions (see, Banerjee, 1992, 1993; Orléan, 1995; Cont and Bouchaud, 2000; Tedeschi et al., 2009, 2012).

---

[1] Herding phenomenon has been also underlined as a relevant feature in the rest of the cryptocurrency market (see, for instance, Vidal-Tomás et al., 2019 and da Gama Silva et al., 2019).

Inspired by the above mentioned empirical and theoretical studies, in this paper we present a model where a population of traders with heterogeneous beliefs can change their expectations based on a switching herding mechanism. We show how this model is able to describe the specifications of traders into a simple dicotomic identification and forecast prices and volatility of this cryptocurrency. Specifically, we model two groups: bullish and bearish traders. Our investors can switch between the two investment attitudes when they observe that one of the two strategies becomes more popular. The switching mechanism is modeled following the modified version of the Kirman's ant colony model (see, Kirman, 1993) proposed by Alfarano et al. (2005). Moreover, we model the ratio of two Bitcoin consecutive prices as the odds ratio of bullish to bearish traders. The reason behind this important assumption derives from the empirical evidence. In fact, several studies show that markets driven by the bullish strategy are characterized by waves of optimism which inflate prices. Conversely, the bearish behavior acts as a thermostat and deflates speculative bubbles (see, for instance, Recchioni et al., 2015; Tedeschi et al., 2019). Overall, it is exactly this alternation of optimism and pessimism waves that allows us to describe price fluctuations and some other stylized facts of the returns time series (see, for instance, Cont, 2001). As far as we know, Cocco et al. (2017) are the only ones to use a behavioral switching model to reproduce the stylized facts of the cryptocurrency returns time series.

It is worthy of note that, while the ability of the odds ratio to model prices' volatility is well known in the economic literature (see, for instance, Kirman, 1993; LeBaron and Yamamoto, 2008; Lux and Marchesi, 2000; Chiarella et al., 2009; Raberto et al., 2001; Tedeschi et al., 2009), the idea of linking this ratio with that of two consecutive prices is unexplored and has the advantage to provide a price model able to explain several empirical findings of the recent literature on Bitcoin. Moreover, our price model has two other important properties we have to mention. First, since the dynamics of the percentage of traders following each strategy (see, z_t in Eq. (3.3)) belongs to the Pearson diffusions, we can statistically deal with the diffusion process (see, Forman and Sørensen, 2008). Second, our price model just involves three parameters (i.e. the propensities to change strategy and the herding parameter) and this allows us to easily estimate it via the indirect inference method (see, Gouriéroux and Valéry, 2004). Precisely these two model characteristics allow us to contribute to the financial literature in two important aspects. On the one hand, we enrich the still too scarce literature on the estimation of behavioral financial models (see, for instance, Gilli and Winker, 2003; Alfarano et al., 2005, 2006; Boswijk et al., 2007; Kukacka and Barunik, 2017). On the other hand, the mathematical properties of the model and its estimation allow us both to describe the agents' behavior trading in Bitcoin and to predict prices. To our knowledge, indeed, this is the only model simultaneously able to produce a good input validation (i.e. to describe traders' behaviour) and to generate an efficient output prediction[2] capable of competing with other models focused on Bitcoin price prediction (see, McNally et al., 2018; Jang and Lee, 2018; Catania et al., 2019; Atsalakis et al., 2019).

The rest of the paper is organized as follows. In Sec. (3.2), we explain the model. In Sec. (3.3), we describe Bitcoin data used for the model estimation. The estimation procedure and forecasting are defined in Sec. (3.4). In Sec. (3.5), we prove the model ability in successfully describing the investors' behavior trading in the cryptocurrency market and in efficiently generating predictive output validation. Finally, in Sec. (3.6) we draw conclusions.

3.2 The model

Inspired by a rich economic literature, which explains price fluctuations starting from the interaction of heterogeneous agents, our model describes Bitcoin dynamics by combining a switching with a herding mechanism. In fact, the co-ordination of traders' strategies by market mediated interactions (for example by following chartist trading rules) or mechanisms of behavioural switching have been shown to lead to large aggregate fluctuations (see, Lux, 1995, 1996, 1998, Lux and Marchesi, 1999b and Brock and Hommes, 1998, among the first studies)[3]. Therefore, inspired by these studies, and following the modified version of the Kirman's ant colony model (see Kirman, 1993) proposed by Alfarano et al. (2005), we show that a combination of these two forces well represents Bitcoin dynamics.

Given the influence of different groups of traders at the micro-level to explain several phenomena at the macro-level, we focus our analysis on the behavior of two groups of traders to describe Bitcoin dynamics, namely, bullish and bearish traders. We assume that in the Bitcoin market the absence of a clear definition of fundamentals renders the use of fundamental strategies rather marginal. Specifically, we employ the odds ratio of bullish traders to bearish traders, $z_t/(1-z_t)$, to quantify the relative impact of one group of agents on the other one at any time in the market. We connect the odds ratio with Bitcoin prices and we assume that bullish traders expect prices to rise, while the other group expects them to decrease. Consequently,

$$\frac{\mathrm{P}_t}{\mathrm{P}_{t-\Delta t}} = \frac{z_t}{1-z_t}, \tag{3.1}$$

---

[2] As Leigh Tesfatsion points out in her website on the validation, there are three different ways of validating computational models: (a) Descriptive output validation, which matches computationally generated output against already-acquired system data; (b) Input validation, which ensures that the structural conditions, institutional arrangements and behavioral dispositions incorporated into the model capture the salient aspects of the actual system; (c) Predictive output validation, which matches computationally generated output against yet-to-be-acquired system data.

[3] A good survey of this type of work is Hommes (2006), where these models have been extensively discussed.

where P_t denotes the price of Bitcoin at time t and z_t is a discrete time stochastic difference equation, known in physics as a Langevin equation[4] (Alfarano et al., 2005):

$$z_{t+\Delta t} = z_t + (\varepsilon_1 - (\varepsilon_1 + \varepsilon_2) z_t) b \Delta t + \sqrt{2b\Delta t\,(1 - z_t) z_t}\,\lambda_t\,. \tag{3.3}$$

In Eq. (3.3) variables ε_1 and ε_2 represent the propensity to change strategy from bullish to bearish, b is the herding parameter, and λ_t, $t > 0$, a random variable drawn from a standard normal distribution (i.e., $\lambda_t \sim N(0,1)$). In Sec. (3.2.1), we derive Eq. (3.3) starting from a Markov Chain, which represents a stylised herding model.

The model (3.1)-(3.3) has several advantages. The first one is the combination of a switching and a herding mechanism driving the dynamics of z_t (see Sec. (3.2.1) for more details). This composite mechanism, that operates at the micro level, is approximated at the aggregate level by a Langevin equation which drives the percentage of bullish traders, z_t. Interestingly, the continuous counterpart of the z-dynamics in Eq. (3.3) belongs to the family of Wright-Fisher diffusion processes, widely used for evolutionary models (see, for example, Ewens, 2004; Larsen and Sorensen, 2007):

$$dz_t = b(\epsilon_1 + \epsilon_2)\left(\frac{\varepsilon_1}{(\varepsilon_1 + \varepsilon_2)} - z_t\right) dt + \sqrt{2b\,(1 - z_t) z_t}\, dB_t, \tag{3.4}$$

where B_t is a Wiener process with $B_0 = 0$. Originally, this Itô stochastic differential equation is used in population dynamics to model the frequency of genes or alleles within a population Ewens (2004). The continuous version of the z-dynamics highlights that the process z_t, $t > 0$ is trapped in the interval $[0,1]$ (i.e., $0 \leq z_t \leq 1$, $t > 0$) given that $0 < z_0 < 1$ and $\epsilon_1 > 1$, $\epsilon_2 > 1$ and the dynamics of the conditional expected value of $z_{t+\Delta t}$ given z_t is (see, Larsen and Sorensen, 2007):

$$E\left(z_{t+\Delta t} \mid z_t\right) = \frac{\epsilon_1}{\epsilon_1 + \epsilon_2} + \left(z_t - \frac{\epsilon_1}{\epsilon_1 + \epsilon_2}\right) e^{-(\epsilon_1 + \epsilon_2) b \Delta t}. \tag{3.5}$$

Thus, the process (3.4) shows a mean reverting behavior in that when $\Delta t \to \infty$ the long term mean of the process is $\epsilon_1 / (\epsilon_1 + \epsilon_2)$ while the speed of mean reversion is $b(\epsilon_1 + \epsilon_2)$.

The second advantage is the low number of model parameters. In fact, model (3.1)-(3.3) allows for a fast estimation procedure in that only three parameters have to be estimated. Going into details, an indirect inference method (see, Gouriéroux and Valéry, 2004) can be successfully used to estimate the model parameters ε_1, ε_2 and b (see Sec. (3.4.1)).

---

[4] In Eq. (3.3) we have to add the reflecting boundaries conditions at $z_t = 0$ and $z_t = 1$ by hand:

$$\begin{aligned} &\text{if} \quad z_t > 1 \quad \text{then} \quad \frac{z_{t+\Delta t} + z_t}{2} = 1, \\ &\text{if} \quad z_t < 0 \quad \text{then} \quad \frac{z_{t+\Delta t} + z_t}{2} = 0, \end{aligned} \tag{3.2}$$

which are equivalent to a reflection around the edges of the domain of z_t, $z_t = 1$, and $z_t = 0$, respectively.

The third benefit is the relationship between the model (3.1)–(3.3) and the normal model for the log-return. Specifically, model (3.1)–(3.3) allows us to explain the connection between the price drift and the fraction of bullish traders. In order to show it, we need to compare the odds model dynamics (3.1) and the following dynamics, which we refer to as "standard dynamics", namely the standard log-normal price model:

$$\frac{P_t}{P_{t-\Delta t}} = e^{\mu_t + \xi_t}, \tag{3.6}$$

where μ_t is a deterministic function of time while ξ_t is a random number drawn from a normal distribution with zero mean and variance $\sigma^2 \Delta t$, where σ is a constant. Taking the logarithms of both sides of Eq. (3.3) and Eq. (3.6) we obtain:

$$r_t = \log\left(\frac{P_t}{P_{t-\Delta t}}\right) = \log\left(\frac{z_t}{1-z_t}\right) = \mu_t + \xi_t. \tag{3.7}$$

We note that z_t is $1/2$ when the fraction of bullish traders equals that of bearish one, and the price does not change with respect to the previous period. Hence, we use the Taylor expansion of the odds ratio logarithm with base point $z^* = 1/2$ to investigate the relationship between the drift μ_t and the fraction of bullish traders. It is easy to see that the Taylor series of the logarithm of the odds ratio is given by:

$$\log\left(\frac{z}{1-z}\right) = \sum_{j=1}^{+\infty} \frac{4^j}{(2j-1)} \left(z - \frac{1}{2}\right)^{2j-1}, \; z \in (0,1). \tag{3.8}$$

Thus, using the first order of the expansion, we obtain the following relationship:

$$r_t = \mu_t + \xi_t \approx 4\left(z_t - \frac{1}{2}\right) + o\left(\left(z_t - \frac{1}{2}\right)^2\right), \; z_t \to \frac{1}{2}, \tag{3.9}$$

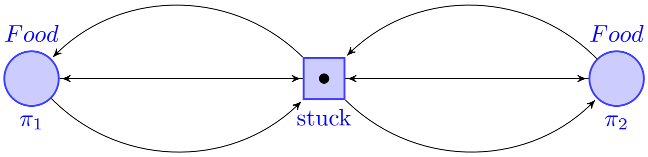
where $o(\cdot)$ is the Landau's symbol. Neglecting the terms with order higher than 2 and taking the unconditional expected value of both sides in Eq. (3.9) we obtain:

$$E[r_t] = E[\mu_t] = \widehat{\mu}_t = E\left[\log\left(\frac{P_t}{P_{t-\Delta t}}\right)\right] \approx 4E\left[z_t - \frac{1}{2}\right] = 4\left(\frac{\varepsilon_1}{(\varepsilon_1+\varepsilon_2)} - \frac{1}{2}\right). \tag{3.10}$$

Eq. (3.10) shows that the estimated drift depends on the deviation of the long term mean, $\frac{\varepsilon_1}{(\varepsilon_1+\varepsilon_2)}$, from $1/2$.

The fourth advantage is a formula for the variance of the price. Model (3.1) and the stochastic dynamics of the fraction of bullish traders, Eq. (3.4), allow us to determine the variance of the logarithm of the odds ratio, which coincides with returns. This formula is obtained by simply differentiating the logarithm of the ratio of two consecutive prices (see, Eq. (3.1)) and applying Ito's lemma. This

Figure. 3.1: Ants model.



very standard computation provides the following expression for the integrated variance, namely the quadratic variation, of the Bitcoin log-price:

$$\sigma_t^2 = \int_{t-\Delta t}^{t} \frac{2b}{z_\tau(1-z_\tau)} d\tau \,. \tag{3.11}$$

Eq. (3.11) tells us that the higher the herding parameter, the higher the volatility of Bitcoin. Interestingly, the volatility increases also in the presence of an excess of bullish or bearish sentiment (i.e., $z_t \rightarrow 1$ or $z_t \rightarrow 0$). As shown in the empirical analysis the fraction of bullish traders z_t oscillates around its average $1/2$, thus the price volatility mainly depends on the herding parameter.

Finally, the fifth advantage is the model ability in detecting bubble periods in the Bitcoin market. Given that the parameters allow us to identify those periods in which bullish traders are dominating the market ($\epsilon_1/\epsilon_2 > 1$), at the expense of bearish traders ($\epsilon_1/\epsilon_2 < 1$), we are able to provide investors with an early warning system alert that detect periods with an excess of optimism in the market.

3.2.1 The Switching mechanism

The behavioural model previously described works under the assumption that the price increments are given by the interaction of two groups of agents: bullish and bearish traders. This traders' dynamics is denoted by the variable z_t representing the percentage of bullish traders, while $1 - z_t$ represents the percentage of bearish traders. In this subsection, we illustrate the switching mechanism underlying the price odds ratio, which is based on the modification of Kirman's ant colony model (see Kirman, 1993) proposed by Alfarano et al. (2005).

Let us start from the seminal Kirman's ants model. We have a population of N ants and two sources of food with "a priori" probability of attracting the ants given by π_1, π_2 (see Fig. (3.1)).

The dynamic evolution of the process is given by:

$$p_1 = P(n, n+1) = \left(\frac{N-n}{N}\right)\left(\pi_1 + (1-\delta)\frac{n}{N-1}\right), \tag{3.12}$$

$$p_2 = P(n, n-1) = \frac{n}{N}\left(\pi_2 + (1-\delta)\frac{N-n}{N-1}\right), \tag{3.13}$$

$$p_3 = P(n, n) = 1 - p_1 - p_2, \tag{3.14}$$

where $\pi_1 = \pi_2 = \epsilon$, $\delta < 1$ and $p_1 + p_2 \leq 1$. The quantity $(1-\delta)$ accounts for the herding effect in that the "a priori" probability increases with the number of ants already present in the source. In absence of herding (i.e. $\delta = 1$) π_1 and π_2 are "a priori" probabilities associated with the two urns (sources of food) and consequently, $\pi_1 + \pi_2 \leq 1$ since, at each step of the Markov chain, we have a trinomial random variable. When these probabilities are equal $\pi_1 = \pi_2 = \epsilon$ and no herding is present (i.e., $\delta = 1$) we have the Ehrenfest urn model and $\epsilon = 1/2$.

Let us now come back to Alfarano et al. (2005) model. The model is populated by N agents, which can be in the state 1 or 2. Traders in state 1 (2) are denoted by n $(N-n)$. The stochastic population dynamics evolves according to the probability of changing from state n at time t, n_t, to some $n'_{t+\Delta t_0}$ at time t+Δt_0 according to the given conditional probabilities denoted by $\rho(n'_{t+\Delta t_0} \mid n_t)$. For sufficiently small time increments Δt_0, the probabilities are linear in the time interval and are defined as:

$$\begin{aligned} \rho(n'_{t+\Delta t_0} = n_t + 1 \mid n_t) &= \ (N - n_t)(a_1 + b n_t) \cdot \Delta t_0, \\ \rho(n'_{t+\Delta t_0} = n_t - 1 \mid n_t) &= \ n_t(a_2 + b(N - n_t)) \cdot \Delta t_0, \end{aligned} \tag{3.15}$$

with the further constraint that:

$$\rho(n'_{t+\Delta t_0} = n_t \mid n_t) = 1 - \rho(n'_{t+\Delta t_0} = n + 1 \mid n_t) - \rho(n'_{t+\Delta t_0} = n - 1 \mid n_t). \tag{3.16}$$

We rewrite Eqs. (3.15) to make a comparison with the Kirman model. To this end, we drop t and we write:

$$\begin{aligned} \rho(n'_{t+\Delta t_0} = n_t + 1 \mid n_t) &= (\Delta t_0 N)\ \left(\frac{N-n}{N}\right)\left(a_1 + (b(N-1))\frac{n}{N-1}\right) = \\ &= \left(\frac{N-n}{N}\right)\left(\underbrace{((\Delta t_0 N)a_1)}_{\pi_1} + \underbrace{(\Delta t_0 N)(b(N-1))}_{1-\delta}\frac{n}{N-1}\right), \end{aligned} \tag{3.17}$$

$$\begin{aligned}\rho(n'_{t+\Delta t_0} = n_t - 1 \mid n_t) &= (\Delta t_0 N)\left(\frac{n}{N}\right)\left(a_2 + b(N-1)\frac{(N-n)}{N-1}\right) = \\ &= \frac{n}{N}\left(\underbrace{(\Delta t_0 N)a_2}_{\pi_2} + \underbrace{(\Delta t_0 N)(b(N-1))}_{1-\delta}\frac{(N-n)}{N-1}\right).\end{aligned} \tag{3.18}$$

Comparing Eqs. (3.12)-(3.13) with Eqs. (3.17)-(3.18) we have:

$$\pi_1 = a_1(\Delta t_0 N), \tag{3.19}$$

$$\pi_2 = a_2(\Delta t_0 N), \tag{3.20}$$

$$1 - \delta = b(N-1)(\Delta t_0 N). \tag{3.21}$$

Setting $n_t = z_t N$ which implies $N - n_t = N - N z_t = N(1 - z_t)$, $t > 0$ and $\Delta t = N \Delta t_0$, from Eqs. (3.17)-(3.18) we have:

$$\rho(n'_{t+\Delta t_0} = n_t + 1 \mid n_t) = (1 - z_t)\left(\underbrace{a_1 \Delta t}_{\pi_1} + \underbrace{b(N-1)\Delta t}_{1-\delta} z_t \frac{N}{N-1}\right), \tag{3.22}$$

$$\rho(n'_{t+\Delta t_0} = n_t - 1 \mid n_t) = z_t\left(\underbrace{a_2 \Delta t}_{\pi_2} + \underbrace{b(N-1)\Delta t}_{1-\delta}(1 - z_t)\frac{N}{N-1}\right). \tag{3.23}$$

We have now to define the step from z_t to $z_{t+\Delta t}$. We know that $\Delta t = N \Delta t_0$ so we set

$$z_{t+\Delta t} = z_t + \sum_{i=1}^{N} \eta_{t+i\Delta t_0}, \tag{3.24}$$

where $\eta_{t+i\Delta t_0}$ are independent and identically distributed random variable defined by:

$$\eta_{t+i\Delta t_0} = \begin{cases} \frac{1}{N}, & p_i^+ = (1 - z_t)\left(\underbrace{a_1 \Delta t}_{\pi_1} + \underbrace{b(N-1)\Delta t}_{1-\delta} z_t \frac{N}{N-1}\right) \\ -\frac{1}{N}, & p_i^- = z_t\left(\underbrace{a_2 \Delta t}_{\pi_2} + \underbrace{b(N-1)\Delta t}_{1-\delta}(1 - z_t)\frac{N}{N-1}\right) \\ 0, & 1 - p_i^+ - p_i^- \end{cases} \tag{3.25}$$

with expected value

$$\mu_i = E(\eta_{t+i\Delta t_0}) = \frac{1}{N}\Delta t\,(a_1(1 - z_t) - a_2 z_t) = \frac{1}{N}\Delta t\,(a_1 - (a_1 + a_2) z_t), \tag{3.26}$$

and variance

$$\begin{aligned}\sigma_i^2 &= V(\eta_{t+i\Delta t_0}) = E\left(\eta^2_{t+i\Delta t_0}\right) - \left(E(\eta_{t+i\Delta t_0})\right)^2 = \\ &= \frac{\Delta t}{N^2}\left[(1-z_t)a_1 + z_t a_2 + 2bz_t(1-z_t)(N-1)\frac{N}{N-1}\right] - \frac{\Delta t^2}{N^2}\left(a_1 - (a_1+a_2)z_t\right)^2.\end{aligned} \tag{3.27}$$

Keeping only the dominant term of the variance we have

$$\sigma_i^2 = V(\eta_{t+i\Delta t_0}) \approx \frac{\Delta t}{N^2}\left[2bz_t(1-z_t)(N-1)\frac{N}{N-1}\right] = \frac{\Delta t}{N}(2bz_t(1-z_t)). \tag{3.28}$$

For sufficiently large N, for the central limit theorem, we have that

$$\frac{\frac{1}{N}\sum_{i=1}^N \eta_{t+i\Delta t_0} - \mu_i}{\sigma_i/\sqrt{N}} \sim N(0,1), \tag{3.29}$$

so the variable $\xi_t = \sum_{i=1}^N \eta_{t+i\Delta t_0}$ is a normal variable with mean

$$N\mu_i = NE(\eta_{t+i\Delta t_0}) = \Delta t\left(a_1 - (a_1+a_2)z_t\right), \tag{3.30}$$

and variance

$$N^2\frac{\sigma_i^2}{N} = N^2\frac{V(\eta_{t+i\Delta t_0})}{N} = N^2\frac{\Delta t}{N^2}(2bz_t(1-z_t)) = 2b\Delta t z_t(1-z_t). \tag{3.31}$$

Consequently, we obtain:

$$z_{t+\Delta t} = z_t + \Delta t\left(a_1 - (a_1+a_2)z_t\right) + \sqrt{2b\Delta t z_t(1-z_t)}\lambda_t, \tag{3.32}$$

where λ_t is a random number drawn from a normal distribution. The CLT and Eq. (3.32) hold true far from the boundaries.

In continuous time we can write

$$dz_t = \left(a_1 - (a_1+a_2)z_t\right)dt + \sqrt{2bz_t(1-z_t)}dB_t, \tag{3.33}$$

where dB_t is the differential of a Wiener process $B_t \sim N(0,t)$ and $B_0 = 0$.

The choice of rewriting a_1 and a_2 as follows:

$$\begin{aligned} a_1 &= b\varepsilon_1, \\ a_2 &= b\varepsilon_2, \end{aligned} \tag{3.34}$$

is only a way to parameterize the dynamics. Concluding the restrictions to new parameters that guarantee the correct linkage with the original Markov chain are:

$$\begin{aligned}\delta &= 1 - b(N-1)\Delta t < 1,\\ \pi_1 &= \epsilon_1 b \Delta t \leq 1,\\ \pi_2 &= \epsilon_2 b \Delta t \leq 1,\\ \pi_1 + \pi_2 &\leq 1.\end{aligned} \tag{3.35}$$

Note that in Eq. (3.33), the boundaries conditions are natural and, therefore, it is not necessary to impose any further exogenous restriction like in the discrete version in Eq. (3.32).

3.3 Data description

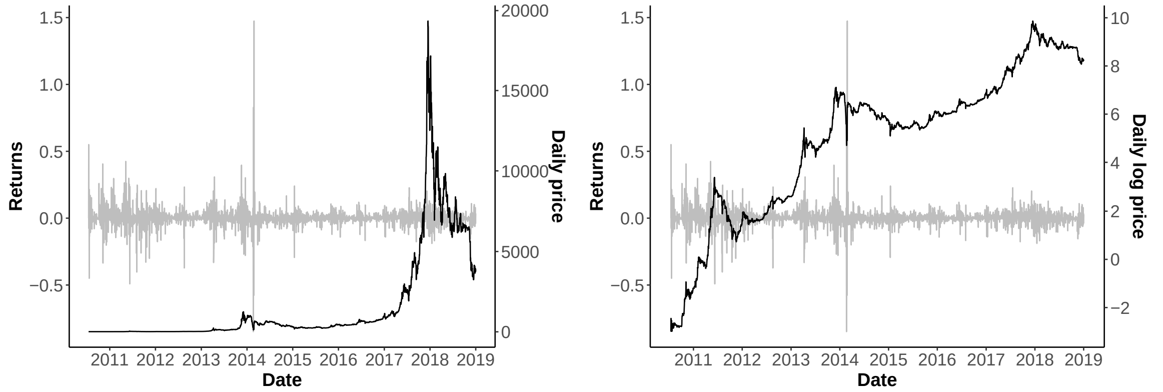
We use the Bitcoin close price from July 16, 2010 to January 1, 2019. Data are freely downloadable from Yahoo Finance website in daily frequency.

Fig. (3.2) shows the evolution of Bitcoin daily price over time. Although the highest value of Bitcoin is 19.345$ on 16/12/2018, using log prices it is possible to observe that the most profitable period of Bitcoin is from 2010 to 2014. The turning point seems to be the bankruptcy of Mt. Gox market in 2014, as shown by returns at the beginning of this year. Fig. (3.2) and Table 3.1 show that Bitcoin returns are characterized by an extremely high standard deviation. Moreover, the daily mean and median of returns are equal to 0.36% and 0.19%, and this highlights the profitability of trading in Bitcoin during the last 8 years.

Table 3.1: Descriptive statistics.

Bitcoin	Mean	Median	Std.Dev	Skewness	Kurtosis	Min.	Max.
Return	0.0036	0.0019	0.0684	2.9370	94.2214	-0.8488	1.4744

Figure. 3.2: Bitcoin daily prices and daily log-prices, along with the daily returns, from 16/07/2010 to 01/01/2019.



3.4 Estimation and forecasting procedures

3.4.1 The indirect inference method

In this subsection, we present the procedure used to estimate the parameters, ε_1, ε_2 and $b\,\Delta t$. Moreover, since these parameters are estimated over time, via a rolling window approach, we explain how the calibration window, N_w, is selected in Sec. (3.4.2).

Let us consider a sample $z_{t_1}, z_{t_2}, \ldots, z_{t_n}$ of fractions of bullish traders. Following Gouriéroux and Valéry (2004), the indirect inference method estimates the parameters appearing in Eq. (3.3) using the autoregression with conditional heteroskedasticity as the instrumental model:

$$\frac{z_{t_{i+1}} - z_{t_i}}{\sqrt{2z_{t_i}(1-z_{t_i})}} = \frac{\beta_0}{\sqrt{2z_{t_i}(1-z_{t_i})}} + \frac{\beta_1 z_{t_i}}{\sqrt{2z_{t_i}(1-z_{t_i})}} + \nu\,\lambda_{t_{i+1}}\,,\ i = 1,2,\ldots,n, \tag{3.36}$$

where $\lambda_{t_{i+1}}$, $i = 1, 2, \ldots, n$ are standard normally distributed errors, while the estimated values of β_0 and β_1 are related to the estimates of the model parameters $\widehat{\varepsilon}_1$, $\widehat{\varepsilon}_2$ and $\widehat{b\,\Delta t}$ from Eq. (3.3) as follows:

$$\widehat{\beta_0} = \widehat{\varepsilon}_1\,\widehat{b\,\Delta t}, \tag{3.37}$$

$$\widehat{\beta}_1 = -(\widehat{\varepsilon}_1 + \widehat{\varepsilon}_2)\,\widehat{b\,\Delta t}, \tag{3.38}$$

and

$$\widehat{\nu^2} = \widehat{b\,\Delta t}\,. \tag{3.39}$$

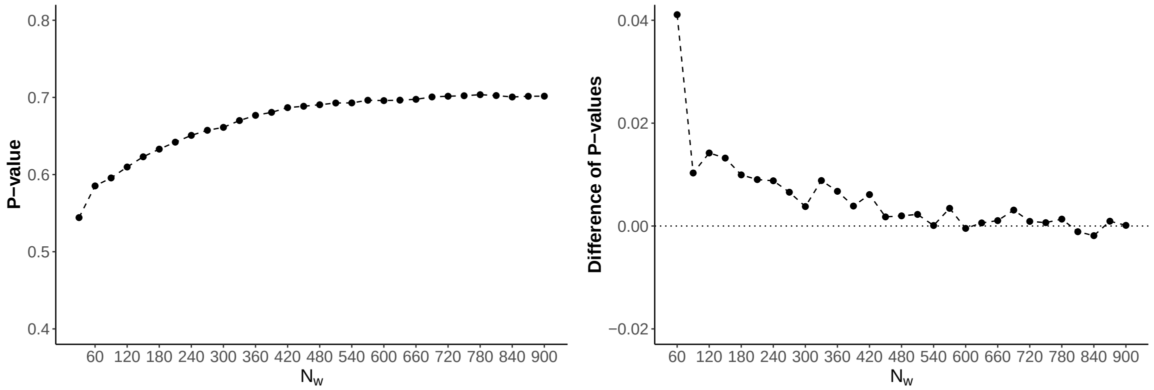
The parameters $(\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\nu})$ are estimated by the ordinary least squares approach.

3.4.2 Choice of the time window N_w

The estimation procedure uses a rolling window of a fixed sample size N_w, which moves along the time series discarding the oldest data observation and inserting the newest one. The first step of the estimation procedure consists in determining the time window size, i.e. the time period over which is reliable to assume that model parameters are constant. We consider a plausible range for N_w, from 30 to 900 days. Following Larsen and Sorensen (2007), to select the proper time window we use the simulated uniform residuals (Pedersen, 1994). Let $F(y \mid x; \varepsilon_1, \varepsilon_2, b\Delta t)$ denotes the conditional distribution function associated with the model (3.3) and $U_i = F(z_{t_i} \mid z_{t_{i-1}}; \widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \widehat{b\Delta t})$, $i = 1, 2, \ldots, n$. If the data z_{t_i}, $i = 1, 2, \ldots, n$, have been generated by Eq. (3.3) with model parameters $\widehat{\varepsilon}_1$, $\widehat{\varepsilon}_2$, $\widehat{b\Delta t}$, then U_i, $i = 1, 2, \ldots, n$, are independent and uniformly distributed in the unit interval. The Kolmogorov-Smirnov test statistics is used to assess the reliability of the assumption of constant values of model parameters over the time window considered. Thus, we test the null hypothesis that the residuals are uniformly distributed in each possible interval, N_w. This approach to select the value of N_w is feasible in that it does not require the knowledge of the model parameters.

The p-values obtained with Kolmogorov Smirnov test statistics are shown in Fig. (3.3), where we can observe their value to be higher than 0.5. To better investigate the p-value dynamics, we calculate the difference of the p-values corresponding to two consecutive values of the window size N_w. We can observe that the time series of the p-value variations approaches zero when N_w is longer than 450. Therefore, we use $N_w = 450$ to estimate the parameters during the forecasting analysis.

Figure. 3.3: Kolmogorov Smirnov test based on Pedersen approach for each N_w.



3.4.3 Forecasting procedure

In this section, we define a price forecast using Eq. (3.40):

$$ln\left(\frac{\mathrm{P}_t}{\mathrm{P}_{t-\Delta t}}\right) = ln\left(\frac{z_t}{1-z_t}\right) = f_{t+\Delta t}, \tag{3.40}$$

where $f_{t+\Delta t}$ is defined using Eq. (3.8) up to the 8th term:

$$f_{t+\Delta t} = \sum_{j=1}^{8} \frac{4^j}{(2j-1)} E\left[\left(z_t - \frac{1}{2}\right)^{2j-1}\right]. \tag{3.41}$$

Then we estimate $f_{t+\Delta t}$ using the sample moments as follows:

$$\widehat{f}_{t+\Delta t} = \sum_{j=1}^{8} \frac{4^j}{2j-1} \frac{1}{N_w} \left[\left(\widehat{z}_{t+\Delta t} - \frac{1}{2}\right)^{2j-1} + \sum_{m=1}^{N_w-1} \left(z_{t-(m-1)\Delta t} - \frac{1}{2}\right)^{2j-1}\right]. \tag{3.42}$$

where, $\widehat{z}_{t+\Delta t} = E\left(z_{t+\Delta t} \mid z_t\right)$ from Eq. (3.5)

Here, N_w is the sample size[5] while $\widehat{z}_{t+\Delta t}$ is the forecast of the fraction of bullish traders at $t + \Delta t$. Thanks to the estimate of $f_{t+\Delta t}$ the one-period ahead price forecast reads:

$$\widehat{P}_{t+\Delta t} = P_t e^{\widehat{f}_{t+\Delta t}}. \tag{3.43}$$

Formula (3.43) can be generalized to define the k-period-ahead forecasts as follows:

$$\widehat{P}_{t+k\Delta t} = P_t \prod_{m=1}^{k} e^{\widehat{f}_{t+m\Delta t}}, k = 1, 2, \ldots, \tag{3.44}$$

where $\widehat{f}_{t+m\Delta t}$ is defined as

$$\widehat{f}_{t+m\Delta t} = \sum_{j=1}^{8} \frac{4^j}{2j-1} \frac{1}{N_w} \left[\sum_{l=1}^{m} \left(\widehat{z}_{t+l\Delta t} - \frac{1}{2}\right)^{2j-1} + \sum_{l=1}^{N_w-1} \left(z_{t-(l-1)\Delta t} - \frac{1}{2}\right)^{2j-1}\right], \tag{3.45}$$

while $\widehat{z}_{t+m\Delta t}$, $m = 2, 3, \ldots$,

$$\widehat{z}(t + m\Delta t) = \frac{\epsilon_1}{\epsilon_1 + \epsilon_2} + \left(\widehat{z}_{t+(m-1)\Delta t} - \frac{\epsilon_1}{\epsilon_1 + \epsilon_2}\right) e^{-(\epsilon_1 + \epsilon_2) b\Delta t}. \tag{3.46}$$

3.5 Empirical analysis

3.5.1 Model consistency

Firstly, we explore the relation between the odds model dynamics (see Eq. (3.1)) and the "standard dynamics" (see Eq. (3.6)). Let us denote with $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ and b_t the daily time series of the estimated model parameters[6], with $\widehat{\mu}_t$ the drift of the log-return estimated over the rolling windows and with x_t the log-return of prices (i.e., $x_t = \log\left(P_t / P_{t-\Delta t}\right)$).

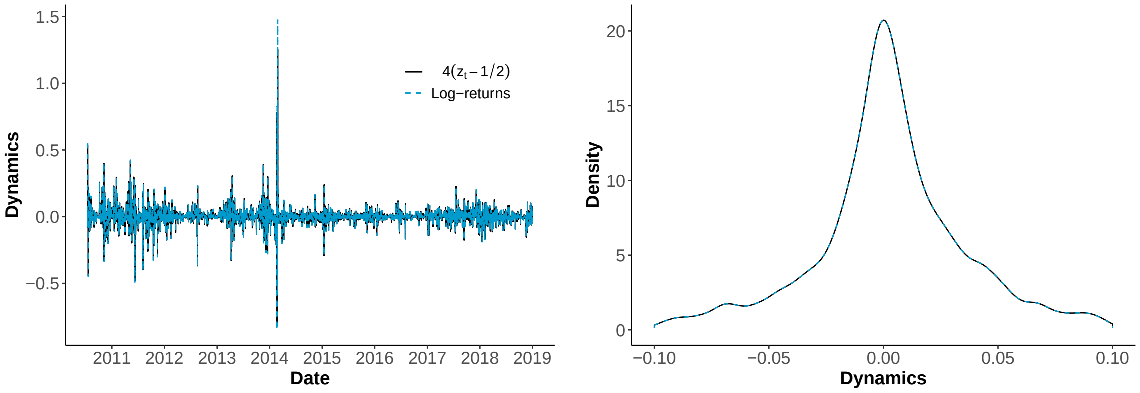
---

[5] Specifically, N_w refers to the window of in-sample data needed to estimate the model's parameters.

[6] We omit the hevaier notation $\widehat{\varepsilon}_{1,t}$, $\widehat{\varepsilon}_{2,t}$, $\widehat{b\Delta t}$.

On the one hand, the left panel of Fig. 3.4 shows the dynamics of $4(z_t - 1/2)$ (black solid line) and the log-return, i.e., x_t (blue dashed line). On the other hand, the right panel shows the density of the log-return x_t (blue dashed line) and the $4(z_t - 1/2)$ (black solid line). As expected, the two lines overlap due to the theoretical link between the two dynamics (see Eq. 3.9 and 3.10).

Figure. 3.4: Dynamics (left panel) and density (right panel) of $4(z_t - 1/2)$ (black solid line) and the log-return x_t (blue dashed line).



We now study the parameters' dynamics by testing two characteristics of the theoretical model shown in Sec. (3.2):

i) the "normalization" condition, i.e., $(a_{1,t} + a_{2,t})\Delta t \leq 1$, or equivalently, $(\varepsilon_{1,t} + \varepsilon_{2,t})b_t\Delta t \leq 1$ (see Eqs. 3.35);

ii) the relationship between the estimated drift and the long term mean given in Eq. (3.10).

Thus, we proceed testing two models. The first one regards the "normalization condition":

$$(\varepsilon_{1,t} + \varepsilon_{2,t})b_t\Delta t = \alpha_0 + \xi_t, \tag{3.47}$$

the second one the behavioral explanation of the drift dynamics:

$$\widehat{\mu}_t = \gamma_0 + \gamma_1 \left(\frac{\varepsilon_{1,t}}{\varepsilon_{1,t} + \varepsilon_{2,t}} - \frac{1}{2} \right) + \xi_t\,. \tag{3.48}$$

Here, ξ_t is a random number normally distributed with zero mean and constant variance, while α_0, γ_0 and γ_1 are constant. The expected results of the two linear regressions are $\alpha_0 \leq 1$, $\gamma_0 = 0$ and $\gamma_1 = 4$. Table 3.2 shows the results of the linear regressions, which confirm our expectations.

Table 3.2: Testing model 1: $b_t \Delta t(\varepsilon_{1,t} + \varepsilon_{2,t}) = \alpha_0 + noise$, $\alpha_0 = 1$ (first two lines); Testing model II: $\widehat{\mu}_t = \gamma_0 + \gamma_1 \left(\varepsilon_{1,t}/(\varepsilon_{1,t} + \varepsilon_{2,t}) - 1/2\right) + noise$, $\gamma_0 = 0$, $\gamma_1 = 4$ (last two lines).

Testing normalization - $b_t \Delta t(\varepsilon_{1,t} + \varepsilon_{2,t}) = \alpha_0 + \xi_t$

Asset	α_0	–	R^2	p-value (test F)	res. std. error
Bitcoin	0.977(***)	-	-	2.0e-16	0.05156

Testing drift relationship - $\widehat{\mu}_t = \gamma_0 + \gamma_1 \left(\varepsilon_{1,t}/(\varepsilon_{1,t} + \varepsilon_{2,t}) - 1/2\right) + \xi_t$

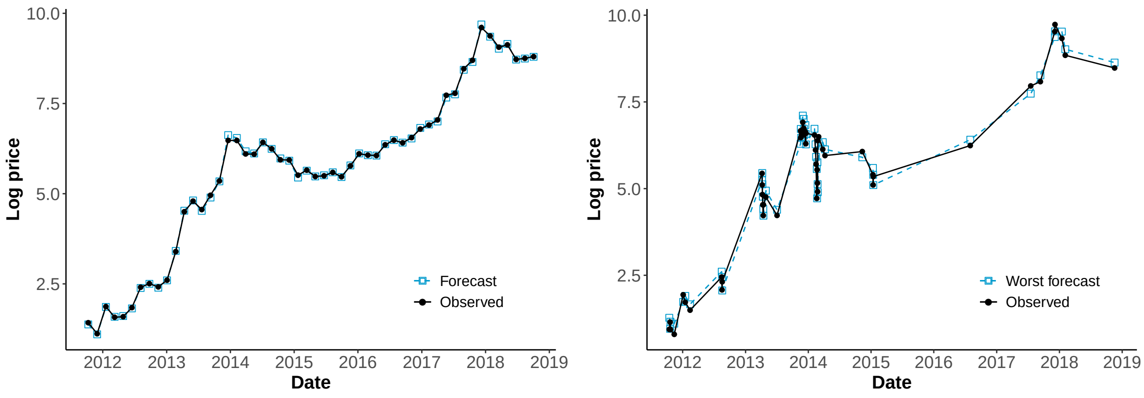
Asset	γ_0	γ_1	R^2	p-value (test F)	res. std. error
Bitcoin	2.27e-5(***)	3.965(***)	0.9946	2.2e-16	2.346e-4

3.5.2 Forecasting analysis

Having shown the approach to select the optimal time window size (N_w) to estimate model parameters, we now investigate the potential of the model in forecasting.

We report in Fig. (3.5) the price forecast one-day ahead using Eq. (3.43). On the left we plot one point out of every fifty points while on the right we plot the worst fifty forecasts. As can be observed, the worst forecasts of our model are mainly found before 2015. In particular, during the bankruptcy of Mt. Gox exchange in February 2014, and during the DDos (distributed denial of service) attack in April 2013. In addition to these particular events, the model is not able to properly forecast prices during the final phase of the bubbles, i.e. the crashes.

Figure. 3.5: One period ahead forecast (blue dashed line) and observed prices (black solid line). In the left panel prices and forecast are represented one out of every fifty ones. In the right panel the worst fifty one-period ahead forecasts.



Following the Bitcoin literature, we employ five measures to examine the predictive power of our model: Root Mean Square Error (RMSE) (Jang and Lee, 2018; McNally et al., 2018; Atsalakis et al., 2019), Mean Square Error (MSE) (Han et al., 2019), Mean Absolute Percentage Error (MAPE) (Jang and Lee, 2018), Mean Absolute error (MAE) (Atsalakis et al., 2019), and the accuracy of our model to predict properly the direction in the change of the future Bitcoin prices (McNally et al., 2018; Catania et al., 2019; Atsalakis et al., 2019). We define these measures in the following way:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2, \tag{3.49}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}, \tag{3.50}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|(x_t - \hat{x}_t)\right|, \tag{3.51}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{(x_t - \hat{x}_t)}{x_t}\right|, \tag{3.52}$$

where T is the number of forecasts of the out-sample period, x_t the $t-th$ observed Bitcoin log-price and $\hat{x}_t$ the t-th log-price forecast. Finally, the accuracy to predict the direction is defined as the ratio of the correct forecasts between the total number of forecasts T. The results are reported in Table 3.3 where three data samples are investigated: i) the complete dataset, ii) a sub-sample reporting the data before Mt. Gox bankruptcy, iii) a sub-sample reporting the data after Mt. Gox bankruptcy. As the reader can notice our model results improve after Mt. Gox bankruptcy and this supports the visual results showed in Fig. (3.5). This fact can be attributed to the increase in liquidity and market capitalization of Bitcoin, which reduce the impact of extreme events in asset prices.

Table 3.3: Mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and percentage of accuracy of our model for the corresponding periods. Forecast log-prices.

Our model	MSE	RMSE	MAE	MAPE	%Accuracy
All sample (11/10/2011 - 01/01/2019)	0.0039	0.0625	0.031	0.00741	54.22%
Before Mt. Gox Bankruptcy (01-02-2014)	0.00385	0.06211	0.03641	0.01433	57.76%
After Mt. Gox Bankruptcy (01-04-2014)	0.001552	0.03939	0.02535	0.00359	53.11%

At this point a question arises: how does our model perform in forecasting prices compared with other studies? To answer this question, we compare the results of our price forecasts with those of other recent studies shown in Table 3.4 (the first column).

Before going into details on the comparison results some explanations on the Bitcoin prices forecasting mechanisms used by other models are necessary. Most of these studies use neural networks for modeling the price prediction. For instance, McNally et al. (2018) employ a Bayesian optimized recurrent neural network (RNN) and a long short term memory (LSTM) network to forecast Bitcoin prices. Similarly, Jang and Lee (2018) compute a Bayesian neural network (BNN) to forecast Bitcoin via 16 explanatory variables. Finally, Atsalakis et al. (2019) employ an artificial neural network (ANN), an adaptive neuro-fuzzy inference system (ANFIS) and a hybrid neuro-fuzzy controller system, namely

PATSOS, which is a combination of two adaptive neuro-fuzzy inference system. Differently from these studies, Catania et al. (2019) study the performance of several alternative univariate and multivariate models with 13 explanatory variables. In particular, their best models are: the dynamic model averaging across all model (DMA), the dynamics model selection (DMS), and their restricted versions, DMA-NR and DMS-NR, where only the lagged values of Bitcoin prices are used. On the one hand, the DMA averages the forecasts across all the different model combinations[7] based on the predictive likelihood that measured the ability of a model to predict Bitcoin prices. On the other hand, the DMS refers to the model with the highest probability among all the possible model combinations.

Let us now compare results on price forecasts of these different methodologies. First of all, two clarifications are needed: i) given that these studies analyze different time periods, we homogenize the out-of-sample data used for our forecast with those of each one of these studies (see the third column of the Table 3.4); ii) the comparison is based on error measures, i.e. the mean square error (MSE), the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) and on the accuracy to provide price trend predictions (see the last five columns of Table 3.4). As the reader can observe there is a strong competition among the different models in terms of forecasting errors. In fact, in some cases our model shows better performances (as demonstrated with the RNN, the LSTM and the ANN) while in others worse results (as with the PATOS). Finally, with some models (see, for instance, the BNN and the ANFIS) the forecast errors are the same and, therefore, a supremacy is not recognizable. Similar findings are observed analyzing the accuracy on price trend predictions. Here, for example, we observe that our model underperforms the DMA, the DMA-NR, the DMS and all the Atsalakis et al. (2019) model specifications[8], but outperforms the RNN, the LSTM and the DMS-NR.

---

[7] Considering the 13 explanatory variables and the additional factors there are 2,621,440 possible combinations.

[8] In this regard, a clarification is necessary: in this comparison analysis we are using only 60 observations out-of-sample, which certainly weakens our performances.

Table 3.4: Comparison of our model when forecasting prices in relation to the models proposed in the literature by means of the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and % of accuracy of our model for the corresponding out-sample periods.

Model	Specification	Out-sample	MSE	RMSE	MAE	MAPE	%Accuracy
Our model		Dec 15-Jul 16	0.00088	0.0286	0.0173	0.0028	59.34%
McNally et al. (2018)	RNN			0.0545			50.25%
	LSTM			0.0687			52.78%
Our model		Dec 16 -Aug 2017	0.00196	0.044	0.0301	0.0041	64.15%
Jang and Lee (2018)	BNN			0.0069		0.018	
Our model		Jan 16 - Dec 17	0.00151	0.03886	0.0244	0.0031	59.64%
Catania et al. (2019)	DMA-NR						62.00%
	DMS-NR						59.00%
	DMA						63.00%
	DMS						60.00%
Our model		Aug 17 - Oct 17	0.00239	0.04893	0.0341	0.0040	51.66%
Atsalakis et al. (2019)	ANN			0.0847	0.0476		55.10%
	ANFIS			0.0489	0.0402		57.70%
	PATSOS			0.0376	0.0307		63.22%

Given the above reported results, we can focus on the main advantages / disadvantages of our model forecast in comparison with the others. Three are the main advantages of our procedure. Firstly, we just use Bitcoin price to describe future dynamics, while the other studies incorporate many other explanatory variables (see, for instance, Jang and Lee, 2018 and Catania et al., 2019). Secondly, we do not need considerable in-sample data to calibrate the model and this implies an easy data achievement for forecasts. Many of the works above presented, in fact, require long time windows of in-sample data (see, McNally et al. (2018), Jang and Lee (2018), and Atsalakis et al. (2019)). Finally, as shown in the next sections, the estimated model parameters $(\varepsilon_1, \varepsilon_2, b)$ allow us to capture market sentiments and, also, to forecast the realized volatility. Regarding the disadvantages, as already stressed, our estimation procedure is particularly sensitive to some episodes that have strongly influenced Bitcoin, such as the Mt. Gox bankruptcy. However it is interesting to note that these episodes, which generated fluctuations in the price time series of up to ±20%, are typical of the first years of this currency creation. This means that from 2015 onwards our procedure produces good results in price forecasts even in the presence of bubbles (see Cheung et al., 2015; Corbet et al., 2018).

Model ability to forecast the realized volatility

Having successfully proven our estimation procedure for prices forecasting, we also test its efficiency in predicting the realized volatility. The realized variance typically used in the swap contract is defined as:

$$\sigma^2_{RV,t} = \frac{A}{N_w} \sum_{i=t-N_w}^{t} (\log(P_i/P_{i-1}))^2 \tag{3.53}$$

where A is the annualisation factor normally chosen to be approximately the number of sampled points in a year (i.e. A=252). Here we choose $A = 1$, that is we consider a daily volatility since the Bitcoin volatility is very high.

We forecast the realized volatility at t by using Eq. (3.11) where the process z_t is replaced by its conditional expectation (3.5) and the integral is approximated by the rectangular rule:

$$\widehat{\sigma}^2_t = \frac{A}{N_w} \left[\frac{2b\Delta t}{\widehat{z}_t(1-\widehat{z}_t)} + \sum_{\tau=t-N_w}^{t-1} \frac{2b\Delta t}{z_\tau(1-z_\tau)} \right], \tag{3.54}$$

where $\Delta t = 1\, day$ and, as already mentioned, $A = 1$.

Fig. (3.6) shows the realized variance $\sigma^2_{RV,t}$ and the model variance $\widehat{\sigma}^2_t$ as function of time. As can be observed, our model seems to forecast properly the volatility with the exception of the period between 2014 - July 2015. In the same line as forecasting prices, the model is sensitive to the bankruptcy of Mt. Gox exchange. However, compared to the price forecast, the volatility is more affected by this event. The impact of this outlier disappears in July 2015 when the event is excluded from the rolling window, i.e. after N_w observations (450) the rolling window excludes the effect of this event. This fact is supported by Table 3.5 where we observe a better fit of the volatility before Mt. Gox bankruptcy and after July 2015.

Figure. 3.6: Observed (black solid line) and forecast (blue dashed line) realized variance over time.

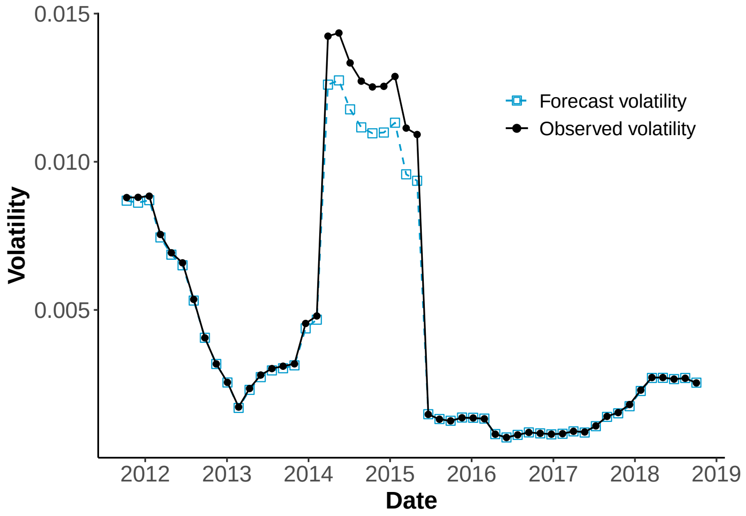


Table 3.5: Mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and % of accuracy of our model of the corresponding periods. Volatility.

Our model	MSE	RMSE	MAE	MAPE
All sample (11/10/2011 - 01/01/2019)	4.402E-07	0.00066	0.000305	0.0000685
Before Mt. Gox Bankruptcy (01/02/2014)	1.146E-08	0.000107	0.0000834	0.0000159
After Mt. Gox Bankruptcy (01/04/2014)	5.867E-07	0.000766	0.000381	0.0000867
After July 2015	3.027E-10	0.0000174	0.0000107	0.00000168

We now compare our model ability in forecasting the realized volatility compared to that of different procedures reported by Trucíos (2019). Specifically, the author examines the performances of several GARCH models (i.e GARCH, IGARCH, EGARCH, GJR, APARCH, CGARCH, TGARCH, AVGARCH, NGARCH, NAGARCH, FGARCH and robust GARCH), alternative GARCH models (i.e betaRGARCH, betaSkewEGARCH and t-GAS) and other techniques using high-frequency data (i.e ARFIMA, ARFIMAL, HAR, HARL, HEAVY and Realised GARCH). In Table 3.6 we just report the GARCH models with lower MSE from each group, that is Robust GARCH Model, t-GAS model and realized GARCH model. As can be observed, our model compares favorably with the robust GARCH model with a lower MSE. However, the t-GAS model and realized GARCH models show slightly better results, even though we observe the same order of magnitude. This result is not surprising since, as Trucíos (2019) contends, t-GAS model is able to down-weight extreme observations while the realized GARCH model is specifically designed to predict the realized volatility using high frequency data. Given that one of the disadvantage of our model is its extreme sensitivity during extreme events, we also report in Table 3.6 the MSE after the Mt. Gox bankruptcy, that is June 2015. As the reader can observe in this subsample we observe a remarkable improve of our forecast which provide empirical evidence of its good performance in the absence of this kind of events.

Table 3.6: Comparison of our model when forecasting volatility in relation to the models proposed in the literature by means of the mean square error (MSE) for the corresponding outsample periods.

	Specification	Date	MSE
Our model		June 2014 - December 2017	6.53E-07
		June 2015 - December 2017	2.79E-10
Trucíos (2019)	Robust GARCH Model	June 2014 - December 2017	1.03E-06
	t-GAS model	June 2014 - December 2017	3.37E-07
	Realised GARCH model	June 2014 - December 2017	2.05E-07

3.5.3 Early warning alert system

As already mentioned, an advantage of our model is its ability to describe the evolution of the market sentiment by estimating the parameters ϵ_1 and ϵ_2, which identify the trader strategy (i.e bullish vs bearish) (see, also Alfarano et al., 2005).

In this regard, we compute the ratio $\varepsilon_1/\varepsilon_2$ using a rolling window of $N_w = 450$. This ratio summarizes three possible behaviors in the currency market: i) $\varepsilon_1/\varepsilon_2 = 1$ means that there is no dominance of one strategy over the other; ii) $\varepsilon_1/\varepsilon_2 > 1$ describes a market where the bullish behaviour controls the market; iii) $\varepsilon_1/\varepsilon_2 < 1$ shows a market influenced by the bearish attitude.

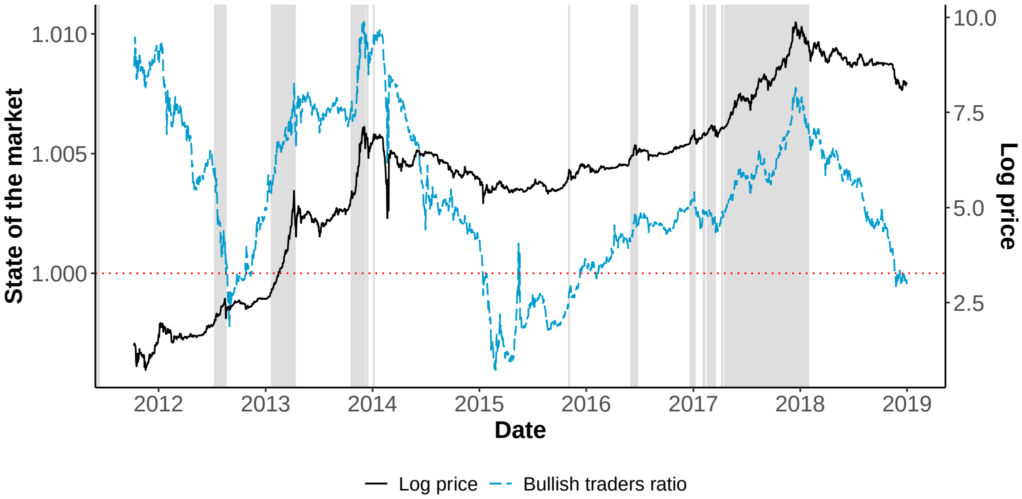
As is known (Alfarano et al., 2005), the predominance of the bullish strategy generates financial bubbles. To verify the impact of this strategy on the dynamics of Bitcoin prices, we compare the values of our ratio with the bubbles identified with the early warning alert system proposed by Phillips et al. (2015).[9] Results are reported in Fig. (3.7) where our ratio (blue dashed line), the bubbles identified by GSADF test (Gray area) and the Bitcoin log prices (black solid line) are shown.

As the reader can observe, Fig. (3.7) shows that the Bitcoin market is mainly dominated by bullish strategy as indicated by the value of the ratio higher than one. This result is certainly in line with the presence of strong fluctuations in Bitcoin prices caused by this strategy. It is worth noting that, when sharp falls in the price series emerge, as in 2012 and in 2014, the bearish behavior takes over. This result is in line with other studies where the emergence of switching behaviors during pre/post periods of financial instability is detected (see Tedeschi et al., 2019 and Grilli et al., 2020). In this sense, in our model the bullish strategy generates fluctuations and volatility, while the bearish one works as a thermostat of the society by realigning prices. The Fig. (3.7) shows another interesting feature of our estimation procedure. By comparing the detection of bubbles obtained with our ratio with respect to those obtained with the GSADF, it seems that this second technique can also identify short-term bubbles. The reason for this important difference lies in the size of the calibration windows. In this regard, Phillips et al. (2015) recommend to use a minimum size window to detect bubbles, $r_0 \cdot T$, where $r_0 = (0.01 + 1.8/\sqrt{T})$, and T is equal to the length of the sample. The resulting window includes 130 observations. However in Fig. (3.7), our parameters ε_1 and ε_2 are estimated using a time window of 450 days.

---

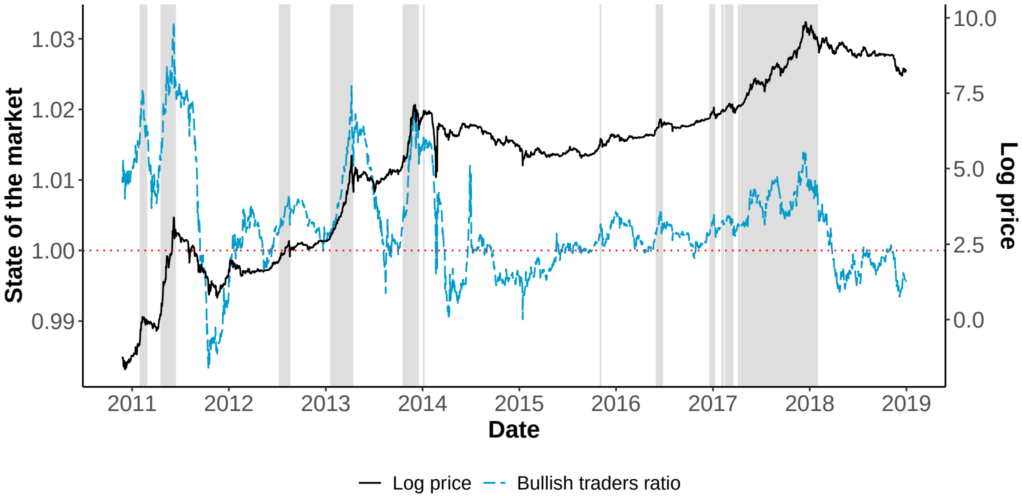
[9] Phillips et al. (2015) developed a generalized version of the sup augmented Dickey-Fuller (GSADF) test (Phillips et al., 2011) able to identify multiple bubbles in the S&P 500 stock market, from January 1871 to December 2010. The new recursive testing procedure is defined as a rolling window right-sided ADF unit root test with a double-sup window selection criteria (Phillips et al., 2015). Compared to their previous work (Phillips et al., 2011), which was just able to identify the existence of one single bubble, Phillips et al. (2015) provides a new framework for testing and dating bubble phenomena in presence of multiple bubbles.

Figure. 3.7: Bullish/bearish traders ratio computed as $\varepsilon_{1,t}/\varepsilon_{2,t}$ (blue dashed line) and the logarithm of the price (black solid line). Gray area highlights bubble periods detected by Phillips et al. (2015) methodology (GSADF). $N_w = 450$.



To verify the impact of the calibration window in detecting short-term bubbles, we estimate the model with $N_w = 130$. As the reader can observe in Fig. (3.8), this simple modification allows us to capture all bubbles and to re-align our results with those of Phillips et al. (2015).

Figure. 3.8: Bullish/bearish traders ratio computed as $\varepsilon_{1,t}/\varepsilon_{2,t}$ (blue dashed line) and the logarithm of the price (black solid line). Gray area highlights bubble periods detected by Phillips et al. (2015) methodology (GSADF). $N_w = 130$.



3.6 Concluding remarks

In this paper we have proposed a simple stochastic model where the Bitcoin price dynamic is driven by the interaction between the bullish and bearish strategy. Specifically, we have described the ratio of two consecutive Bitcoin prices as the odds ratio of bullish and bearish traders. Moreover, the evolution of these two strategic groups has been modeled following the modified version of the Kirman's ant colony model (see Kirman, 1993) proposed by Alfarano et al. (2005).

Our results have shown that the proposed model, estimated via the indirect inference method, well performs in forecasting price trends, dynamics and realized volatilities of Bitcoin daily data. The efficiency of our procedure in generating predictive output validation has been also compared with those of several other studies used to forecast this cryptocurrency. The comparison with some of these models have showed a close competition. Although we have not always been the winners, a consideration with those models against which we underperform must be made: all these models need multiple explanatory variables or/and long time windows of in-sample data to obtain price predictions better than ours. Clearly, the collection of this additional data has time and information costs that our model does not have to face.

In addition, we have shown that the odds ratio of bullish and bearish traders explains market sentiments. In this regard, we have found out that Bitcoin is predominantly driven by the bullish strategy, which generates wider price swings. Instead, the bearish behavior emerges during pre/post periods of financial instability such as the Mt. Gox bankruptcy. This evolutionary analysis on the market sentiments has shown important characteristics in traders' behavior. On the one hand, we have noted that the bullish strategy has a destabilizing power on prices. This behavior, in fact, generates waves of optimism able to create bubbles. On the other hand, the bearish behavior works as a thermostat causing bubble implosion.

In conclusion, our contribution does not want to compete with the other studies, but to support them by trying to describe various aspects of the Bitcoin price dynamics by using a single simple model.

Chapter 4

A cross-sectional analysis of growth and profit rate distribution : the Spanish case

4.1 Introduction

Historically, Gibrat (1931) was the first scholar to propose a stochastic process in order to model the growth of firms based solely on general probabilistic concepts. His basic hypothesis states that the logarithmic growth rate of a firm is independent of its size and it is Normally distributed. The Normal distribution assumption can be justified on the premise of the Central Limit Theorem (CLT hereafter). The logarithmic growth rate of a firm in a given time period (one year for instance) can be decomposed as a sum of a large number of shocks hitting the firm at a higher frequency (e.g. daily). Within this time decomposition, the emergence of the Normal distribution of growth rates is a natural consequence of the CLT, assuming that the shocks are independent and identically distributed. Under these assumptions, the distribution of firms' size is a lognormal distribution. From an economic perspective, the Gibrat's hypotheses are compatible with an ensemble of *independent firms*, experiencing a common trend and idiosyncratic destinies. The Gibrat's statistical approach has been generalised in order to account for other economic phenomena, such as the entry and exit of firms in a market, the turbulence and the learning of firms, leading Sutton to call for the existence of a Gibrat's legacy (Sutton, 1997).

Challenging the Gibrat's hypothesis of Normality, many authors (Amaral et al., 1997; Bottazzi et al., 2001; Bottazzi and Secchi, 2003, 2006; Bottazzi and Secchi, 2011; Buldyrev et al., 2007; Alfarano and Milakovic, 2008; Riccaboni et al., 2011) have empirically shown that firms' growth rates follow a Laplace distribution rather than a Normal distribution.[1] Starting from the basic assumption of iid shocks leading to a Gaussian distribution, the empirical identification of the Laplace distribution can be alternatively interpreted as the imprint of a systemic dependence among the shocks hitting all firms. In order to account for the "Laplacian" deviations from the Gaussian hypothesis, one must replace the assumption of iid shocks by perturbations characterised by some degree of system-wide correlation due to specific economic interactions among firms. The Laplace distribution of cross-sectional firms growth rates, thus, can be thought as the macroscopic evidence of the existence of complex interactions among firms. Some models have been proposed in order to account for the emergence of the Laplace distribution. Bottazzi and Secchi (2006) show that the Laplace distribution stems from a competitive context in which firms are able to seize new growth opportunities proportional to opportunities already taken. Under the resource-based view of the firm (Penrose and Penrose, 2009), Coad and Planck (2012) consider a mechanism of employment growth in a hierarchy, leading to an exponential distributed firm size and a Laplace distribution of growth rates.

Recently, some authors (Alfarano et al., 2012; Mundt et al., 2016) proposed a new focus to analyse firms dynamics from the Gibrat's perspective beyond the growth rates of firm size. They claim that a more informative quantity to account for the dynamics of the ensemble of firms in an economy is to consider firm profit rates instead of growth rates as the key measure of firm performance. This change of focus allows to rely on the general principle of tendency for equalisation of profit rates based on the idea of classical competition. In this respect, Alfarano and Milakovic (2008) introduced a theoretical framework for the profit rate distribution by considering as the intellectual base Adam Smith's notion of classical competition (Smith, 1776), which describes a negative feedback mechanism: capital seeks out those sectors in which profit rates are higher than the economy-wide average, essentially attracting labour, raising output, reducing prices and eventually profit rates. Capital, thus, leaves the sector giving rise to an increase of price and profit rates for those firms that remain in the industry. The entire process tends to equalise profit rates across sectors and firms. The idea of classical competition can be framed in terms of a statistical equilibrium model for the profit rate distribution, which leads to an Exponential Power or Subbotin distribution (Subbotin, 1923). Such theoretical framework has been empirically tested in several contributions (Alfarano et al., 2012; Erlingsson et al., 2013; Mundt et al., 2016), showing that the profit rate distribution can be described by a Laplace distribution, whose first and second moment are very stable over time, much more than the corresponding moments of the growth rate distribution. In particular, it has been shown that such stability is linked to restrict the empirical analysis to large and firms surviving for sufficient time. In this regard, Mundt and Oh (2019) show that the Laplace distribution is not flexible enough to describe the profit rate distribution when entry and exit dynamics of firms is included.

---

[1] Moreover, some authors (Axtell, 2001; Gaffeo et al., 2003) show that the distribution of firms' size follows a power law rather than a lognormal distribution.

They observe an empirical profit rate distribution that exhibits a higher degree of leptokurtosis and a significant asymmetry when compared to a Laplace distribution. Hence, Mundt and Oh (2019) generalise the model proposed by Alfarano et al. (2012) in order to include changes in the nature of the competitive environment and the strength of competitive pressure between entering/existing and incumbent firms. Their model shows that these features can be accounted by the Asymmetric Exponential Power (AEP hereafter) distribution (Bottazzi and Secchi, 2011). The AEP generalises the Subottin distribution in order to include a given degree of asymmetry. Economically, this feature captures the difference between the behaviour of the growth of the typical firm from the average one (Bottazzi et al., 2019), showing its importance to account for aggregate fluctuations.

To shed more light on this strand of literature, we study a large dataset of 35.910 Spanish long-lived firms, analysing the last financial crisis with all the business cycle phases: the period of the real state bubble (1998-2007), the subsequent crisis (2008-2013) and the period of economic recovery (2014-2016). The much larger dataset at our disposal allows for a more extensive analysis of the Laplacian hypothesis of profit rate distribution and its stability over time. In particular, the contribution of our paper is threefold. First, following Alfarano et al. (2012) and Mundt and Oh (2019), we examine whether the empirical profit rate and growth rate distribution of Spanish firms is described by the Laplace, Subbotin or AEP distribution. Compared Mundt and Oh (2019), our analysis is not limited to profit rates but also include the comparison to growth rates. Second, we analyse how the empirical distribution changes according to the different size of the companies. Finally, we also study how the distributions of the profit and growth rates are affected by the different phases of the business cycle. Understanding cross-sectional distribution of growth and profit rates during the different phases of the business cycle can help us to shed more light on macroeconomic fluctuations (Higson et al., 2002, 2004; Gabaix, 2011). In fact, macroeconomic literature has been extended in order to find more sound microfoundations in relation to business cycles models (Holly et al., 2013; Bachmann and Bayer, 2014). Traditionally, most of these analysis used comovements between key macroeconomic aggregates, e.g. consumption and output among other measures (Fidrmuc and Korhonen, 2006). However, the availability of micro-data allows scholars to study how the microeconomic adjustment behaviour of firms affects to the aggregate dynamics of a given economy. For example, Higson et al. (2002) show that fastest growers and declining firms seem to be indifferent to recessions, in the same line as Geroski and Gregg (1997). De Veirman and Levin (2011) analyse trends and cycles in the volatility of U.S. companies observing that firm-specific volatility is not an important driver of the business cycle. Holly et al. (2013) underline that changes in the density of firm growth are a relevant factor to analyse the evolution of the business cycle. Bachmann and Bayer (2014) propose a heterogeneous-firm business cycle model that is able to replicate the procyclical behaviour of the empirical cross-sectional dispersion of firm-level investment rates. Therefore, such as Haltiwanger (1997) states, "it is becoming increasingly apparent that changes in the key macroaggregates at cyclical and secular frequencies are best understood by tracking the evolution of the cross-sectional distribution of activity and changes at the micro level."

The paper is structured as follows. After providing a summary in the introduction, we give a description of our data in Sec. (4.2). The employed methodology for the empirical analysis is described in Sec. (4.3). The results of the empirical analysis are shown in Sec. (4.4), distinguishing between symmetric and asymmetric distributions, namely the Subbotin and the AEP. Finally, Sec. (4.5) summarise the main findings of the paper.

4.2 Data

The dataset is sourced from the System of Analysis of Iberian Balance Sheets and it offers information over the balance sheet of 2.000.000 Spanish firms from 1985 to 2016. Thus, we can examine the evolution of the distribution of growth and profit rates during different phases of the business cycle. In line with most of the industrial dynamics literature mentioned in the introduction, our empirical analysis focuses on long-lived firms. We filter a total of 35.910 firms that have been present in the market for the whole period.[2] Our dataset allows to generalise the previous findings on the distributional properties of the profit rates of large long-lived companies, since we extend the the number of firms of more than two orders of magnitude, from few hundreds to several thousands, whose sizes span five orders of magnitude. In order to compare to previous literature we consider four groups of firms according to their sales in 2016. These groups include the 200, 1.000, 10.000 largest firms and the entire sample.[3]

As starting point, we consider the 200 largest firms due to two main reasons. First, we take as intellectual base the Gabaix's granular hypothesis (Gabaix, 2011). His seminal paper rests on the idea that the idiosyncratic shocks to the largest firms account for a significant fraction of the GDP fluctuations. Following Gabaix (2011), one third of aggregate fluctuations in US GDP growth can be explained by the idiosyncratic shocks of the 100 largest firms. Blanco-Arroyo et al. (2018) and Blanco-Arroyo et al. (2019) show that the Spanish economy is also characterised by granular fluctuations, since the granular residual of the 100 largest firms accounts approximately for 45% of GDP variations. The second reason is related to the fact that we employ the AEP distribution, proposed by Bottazzi and Secchi (2011), to characterise the profit and growth rate distribution. By means of numerical simulations, they state that "the bias of the maximum likelihood estimators, being very small, can be safely ignored at least for samples with more than 100 observations". Therefore, we start the empirical analysis considering the largest 200 firms to ensure the reliability of the parameters estimated.

---

[2] All firms from the financial sector (Standard Industrial Classification (SIC) codes 6000-6799) have been excluded since their total assets are on average about one order of magnitude larger than firms included in all the other sectors. This is due to the different nature of the banking/financial sector, where total assets can be increased due to the financial intermediation activity. The average ROA for banks turns out to be one order of magnitude larger than for firms in other sectors.

[3] Fig. (B.1), in the Appendix, shows the sales as a function of GDP for each group of firms.

As the first step, we compute the logarithmic growth rate for each firm i defined as:

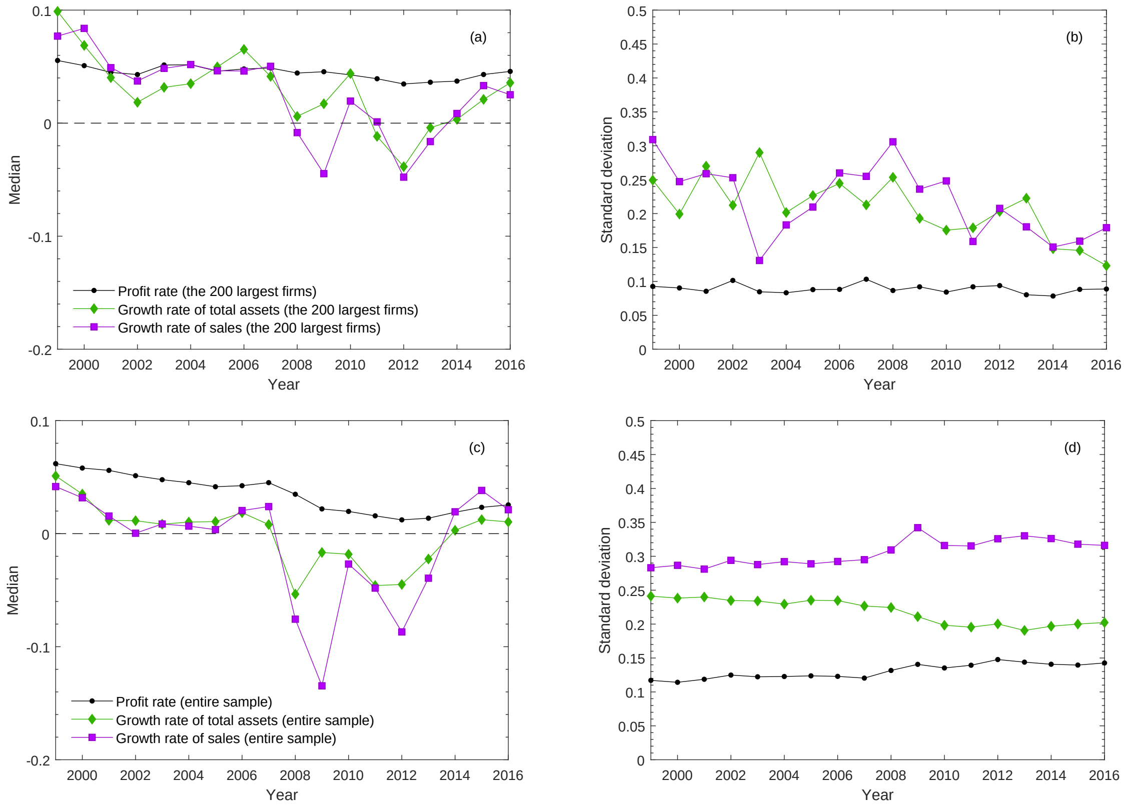
$$\tilde{g}_i(t) = \ln\left(S_i(t)\right) - \ln\left(S_i(t-1)\right), \tag{4.1}$$

where t denotes the year and $S_i(t)$ the firm size, whose proxy is the value of total assets or sales (Axtell, 2001; Stanley et al., 1996).

The variable chosen as a proxy for profit rate is the return on assets (ROA), which is defined as earnings before interest and taxes ($EBIT$) divided by total assets (TA) of firm i at time t,

$$ROA_i(t) = \frac{EBIT_i(t)}{TA_i(t)}. \tag{4.2}$$

Figure. 4.1: The evolution of the cross-sectional median and standard deviation of growth ($\tilde{g}$) and profit rates for the 200 largest long-lived firms and the entire sample at our disposal (base year 2016).



A visual inspection to Fig. (4.1) (a and b) shows that the median of profit rates for the largest 200 long-lived firms exhibits a considerable stability over time compared to the the median growth rates of total assets and sales, which instead exhibits a much higher volatility. The time evolution of the median of profit and growth rates is also reported by Alfarano et al. (2012) using a sample of publicly traded US companies, observing similar results. Our results are also in line with Mundt et al. (2014), who find that the median of profit rates is much more stable than the median of growth rates in more than 40 countries using a dataset of publicly traded companies. Moreover, we confirm the results reported by Coad et al. (2013), who observe a much higher stability of the cross sectional average of profit rates when companies survive more than 11 years. When considering the entire sample of long-lived firms, we observe that the median of profit rates exhibits a higher stability compared to the median of growth rates. However, it shows higher fluctuations with respect to the sample composed by the 200 largest firms, due to the impact of the smaller firms. In both cases reported in Fig. (4.1), the first two moments of the profit rate distribution are more stable than those of growth rates.

The analysis of the first two moments of Fig. (4.1) would be sufficient under a Gaussian hypothesis for the distribution of profit and growth rates. However, an extensive literature in industrial dynamics (see e.g. Sutton, 1997) shows that the empirical distribution of relevant measures of firm performance exhibits significant deviations from the Normality assumption. We, therefore, focus our attention to the characterisation of the entire distribution of profit and growth rates. Following the literature (Bottazzi and Secchi, 2006), we delete possible trends by using the normalized logarithmic size:

$$s_i(t) = \ln(S_i(t)) - N^{-1} \sum_{i=1}^{N} \ln(S_i(t)). \tag{4.3}$$

where N is the number of considered firms in the sample, namely 200, 1.000, 10.000 and the entire sample. We define the annual growth rate of a firm i as:

$$g_i(t) = s_i(t+1) - s_i(t), \tag{4.4}$$

where t denotes time and s_i denotes normalised logarithm of firm size. Profit rates are not normalized in any way and simply remain in the raw form.

4.3 Methodology

Alfarano and Milakovic (2008) introduce a theoretical framework to analyse the distribution of profit rates by considering as an intellectual base the Adam Smith's notion of classical competition (Smith, 1776). It describes a negative feedback mechanism in the reallocation of capital in perpetual search for profitability, leading to a *tendency* for the equalisation of profit rates among competitive economic activities. In the empirical data, however, the complete elimination of profit rates differentials is never achieved.[4] Alfarano et al. (2012), thus, express the outcome of classical competition in terms of a statistical equilibrium model, considering that the complexity of the competitive interactions among firms leads to a non-degenerate distribution of profit rates. In particular, firms disperse their profit rates, denoted as x, around a measure of central tendency, denoted as m, which represents the economy-wide profit rate. The tendency for equalization of profit rates can be encoded as a moment constrain on the dispersion of their distribution measured by the standardized α-th moment:

$$\sigma^{\alpha} = E\left[|x - m|^{\alpha}\right]. \tag{4.5}$$

In order to obtain the profit rate distribution, Alfarano and Milakovic (2008) employ the Maximum Entropy Principle (MEP), which establishes a unique connection between a set of given moment constraints and a probability distribution. The MEP yields the combinatorially most likely distribution maximising the multiplicity of feasible assignments given the moment constrains (see Jaynes, 1979). The result of MEP for the moment constraint in Eq. (4.5) is an Exponential Power or Subbotin distribution, defined as

$$f(x; m, \sigma, \alpha) = \frac{1}{2\sigma\alpha^{\frac{1}{\alpha}}\Gamma(1 + \frac{1}{\alpha})} \exp\left(-\frac{1}{\alpha}\left|\frac{x - m}{\sigma}\right|^{\alpha}\right). \tag{4.6}$$

This symmetric distribution is characterized by three parameters: a location parameter m, a scale parameter $\sigma > 0$ and a shape parameter $\alpha > 0$. Depending on the value of the shape parameter, we have three different cases: (i) a platykurtic distribution for $\alpha > 2$, (ii) a leptokurtic distribution for $\alpha < 2$, and (iii) a Gaussian distribution for the edge case $\alpha = 2$. In particular, the Subbotin distribution reduces to the Laplace distribution when $\alpha = 1$. The distribution in Eq. (4.6) has been widely employed in the literature of industrial dynamics (Bottazzi and Secchi, 2003; Bottazzi and Secchi, 2006; Coad and Planck, 2012; Alfarano et al., 2012; Erlingsson et al., 2013; Mundt et al., 2016) to characterise the empirical distribution of profit and growth rates of firm size, essentially because it interpolates between the Gaussian and the Laplace distribution. Following the growth rate literature (Stanley et al. (1996), Bottazzi and Secchi, 2006 and Coad and Planck, 2012), we consider the Laplace distribution as the benchmark to compare the estimation results.

In this paper, we complement the distributional analysis based on the symmetric distribution of Eq. (4.6) by using the AEP distribution. Mundt and Oh (2019), generalising the result given by Alfarano and Milakovic (2008), provide an economic foundation for the AEP distribution within a statistical equilibrium approach that includes structural differences between the right and left part of the distribution. In particular, they show that the former reflects the activity of incumbent firms while the latter represents the activity of entering/existing companies characterised by low/negative profit rates. Instead of a symmetric behaviour around the measure of central tendency, defined by the Eq. (4.5), which implies the emergence of a symmetric distribution, they define two different conditional measures of dispersion around m: $\sigma^{\alpha_l} = [E|x-m|^{\alpha_l}]$ for $x < m$ and $\sigma^{\alpha_r} = [E|x-m|^{\alpha_r}]$ for $x > m$, where l and r refer to the left and right part of the distribution, respectively. Using the MEP, the probability distribution for the variable x based on the two moment constraints is the following:

---

[4] A perfect elimination of profit rate differentials would lead to a Dirac's delta distribution.

$$f_{AEP}(x;\mathbf{p}) = \frac{1}{C}\exp\left[-\left(\frac{1}{\alpha_l}\left|\frac{x-m}{\sigma_l}\right|^{\alpha_l}\theta(m-x)+\frac{1}{\alpha_r}\left|\frac{x-m}{\sigma_r}\right|^{\alpha_r}\theta(x-m)\right)\right], \qquad (4.7)$$

where $\mathbf{p} = (\alpha_l, \alpha_r, \sigma_l, \sigma_r, m)$, $\theta(x)$ is the Heaviside function[5] and $C = \sigma_l\alpha_l^{1/\alpha_l}\Gamma(1+1/\alpha_l)+\sigma_r\alpha_r^{1/\alpha_r}\Gamma(1+1/\alpha_r)$ is the normalization constant with $\Gamma(\cdot)$ the Gamma function. Eq. (4.7) is a five-parameters family of distributions that is characterized by the location parameter, m, which is the mode of the distribution, two shape parameters, α_l and α_r, describing the density in the lower and upper tail respectively, and two scale parameters, σ_l and σ_r, connected with the distribution width below and above m. The Laplace distribution is nested in the AEP when $\alpha_l = \alpha_r = 1$ and $\sigma_l = \sigma_r = \sigma$. Note that the parameter m in Eq. (4.6) represents the mean, the median and the mode of the distribution. Those three measures of central tendency might not coincide in the AEP distribution.

4.4 Empirical results

In this section, we report the main results of our empirical analysis. In Sec. (4.4.1), we analyse the empirical probability density of profit and growth rates by testing the goodness of fit of the Laplace distribution against the Subbotin distribution. In Sec. (4.4.2), we examine the distributional properties of profit and growth rates testing the Laplace distribution against the AEP distribution.

4.4.1 Symmetric case

We estimate the main parameters of the Subbotin distribution for the largest 200 long-lived firms, using the the maximum likelihood estimation method.[6] We observe that the Laplace distribution provides a relatively poor fit for the profit rate distribution, since, at the 5% significance level, we reject the null hypothesis of $\alpha = 1$ in 11 out of 19 years (see Fig. (4.2)). However, as it has been underlined by Bottazzi and Secchi (2006), Bottazzi et al. (2014) and Mundt et al. (2016), the presence of outliers can significantly affect the estimation of the shape parameter α. Fig. (4.6) shows the presence of some large negative and positive values in several years.[7] Therefore, to avoid the effect of the outliers in the estimation of the parameters α and σ, we delete in each year the most positive and negative observation.[8] We show in the inset of Fig. (4.2) that the Laplace distribution cannot be rejected at the 5% significance level with the exception of 2009. Growth rates of total assets and sales, instead, show a more leptokurtic distribution compared to profit rates distribution with a shape parameter significantly smaller than unity for all years. If we delete the most positive and negative value of sales growth rates, the shape parameter α remains significantly smaller than 1. In the case of growth rates of total assets, the Laplace distribution can be rejected in the majority of cases. Looking at estimators of the scale parameter, we confirm the astonishing stability in the magnitude of profit rate fluctuations. Interestingly, we cannot reject the hypothesis that the scale parameter is constant along the entire period regardless of the phase of the business cycle. This is not the case for the scale parameter of the distributions of growth rates of sales and total assets whose time evolution shows swings, with periods significantly above or below the long term median value (see Sec. (4.4.1)).

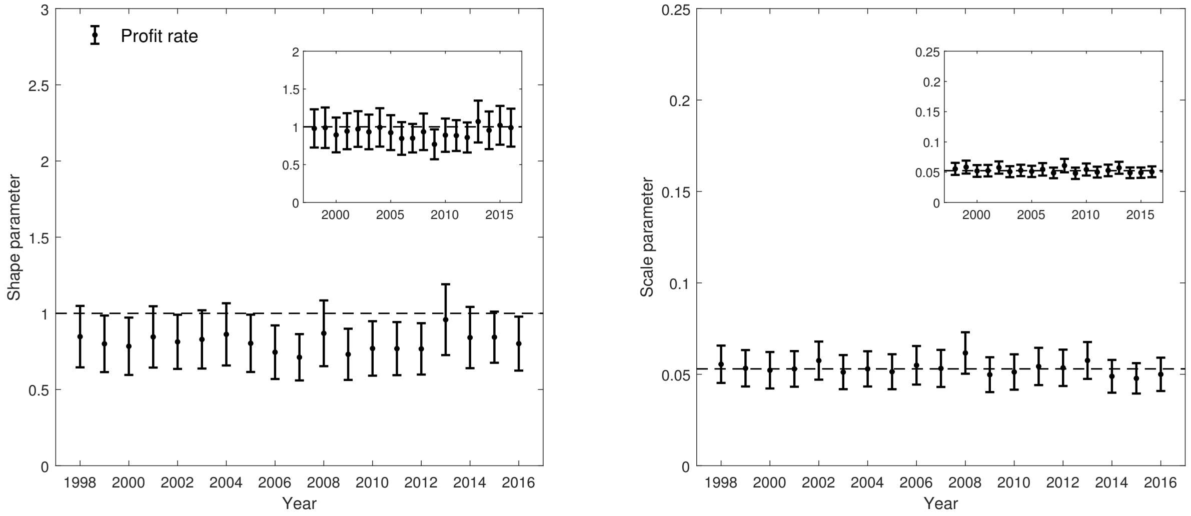
---

[5] The function $\theta(x)$ is equal to 1 for $x > 0$, and 0 for $x < 0$.

[6] For the parameter m the likelihood function is non-analytical and, therefore, the ML estimator does not have the typical asymptotic properties. To avoid convergence and consistency problems, we estimate m as the mode of the distribution and then, we apply the ML to estimate the parameters α and σ conditionally on the value of m (Bottazzi and Secchi, 2006). Using the mean or the median the results are not significantly different when we estimate the parameters of the AEP distribution (material upon request). We opt for the mode for consistency.

[7] The presence of outliers is also observed for growth rates of total assets (Fig. (4.8)) and sales (Fig. (B.2))

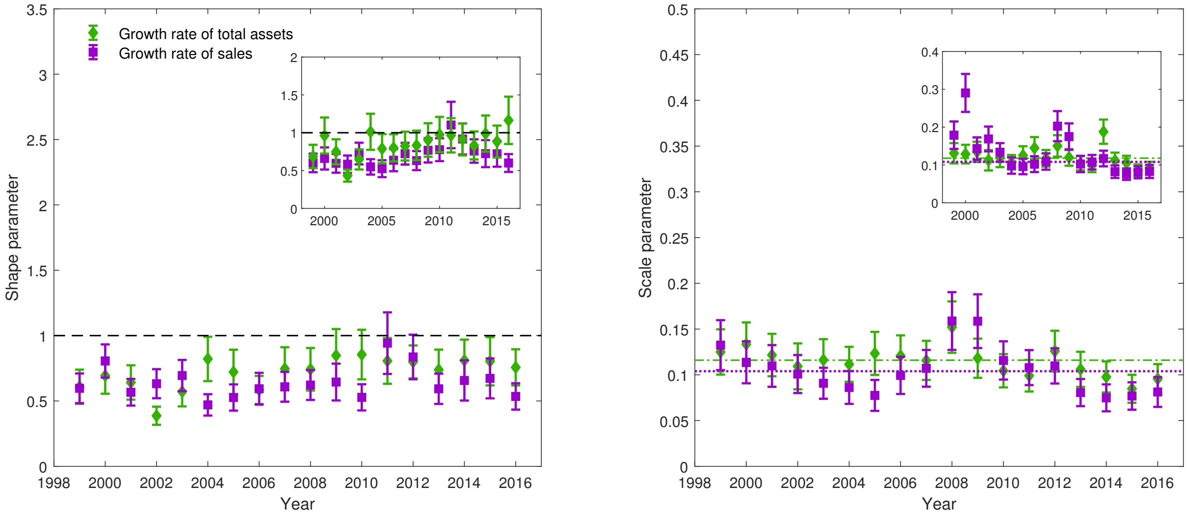
Figure. 4.2: Estimates of the shape and scale parameter of the Subbotin distribution for profit rates. Error bars show two standard errors. The results refer to the 200 largest long-lived firms according to their sales in 2016. The dashed line in the scale parameter figure represents the median of the estimates.



---

[8] From now on, we always delete the extreme positive and negative observations in each year when estimating the parameters of the Subbotin as well as the AEP distribution.

Figure. 4.3: Estimates of the shape and scale parameter of the Subbotin distribution for growth rates of total assets and sales. Error bars show two standard errors. The results refer to the 200 largest long-lived firms according to their sales in 2016. The dotted line (sales) and dashed line with dots (total assets) in the scale parameter figure represent the median of the estimates.



To go beyond a visual inspection, we employ the likelihood ratio test (LRT hereafter) to assess the performance of the Laplace distribution in describing the data. We obtain similar results (see Table B.1 in the Appendix) as compared to the simpler inspection of the estimates of α and σ in Figs. (4.2) and (4.3). The Laplace distribution does not provide a good performance in describing the probability distribution of profit rates, unless deleting the highest and lowest values in each year. In this case, the results of the LRT, reported in Table 4.1, support the previous findings, since we can only reject the null hypothesis for the profit rate distribution in 2009 ($p - value = 0.04$). When comparing the results of the LRT to Fig. (4.3), we observe virtually identical results for the distribution of growth rates of total assets and sales.

Table 4.1: P-values of the likelihood ratio test for profit and growth rates of total assets and sales. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the Subbotin distribution. The results refer to the 200 largest long-lived firms, according to their sales in 2016, when deleting the extreme positive and negative value.

LRT	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Profit rate	0.87	0.93	0.39	0.64	0.81	0.58	0.95	0.53	0.20	0.14
Total assets	-	**0.00**	0.77	**0.01**	**0.00**	**0.00**	0.57	**0.05**	**0.05**	0.09
Sales	-	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

LRT	2008	2009	2010	2011	2012	2013	2014	2015	2016
Profit rate	0.61	**0.04**	0.35	0.29	0.19	0.61	0.72	0.88	0.93
Total assets	0.11	0.42	0.84	0.77	0.42	0.08	0.86	0.29	0.15
Sales	**0.00**	**0.01**	**0.01**	0.48	0.35	**0.01**	**0.01**	**0.01**	**0.00**

Distributional properties conditional on size and business cycle phase

When increasing the sample, Fig. (4.4) shows that the distribution of profit and growth rates exhibits a shape parameter α significantly smaller than 1 most of the years. The data indicate that the distribution of growth rates of firm size roughly retains its shape parameter across the different samples (see Table 4.2), excluding the sample with the largest firms (see Table 4.2). The profit rate distribution, instead, exhibit a clear tendency to become more leptokurtic, while the scale parameter is virtually independent of the size of the considered firms, showing an astonishing stability. For the growth rate distribution, we observe a slight increase in the estimate of the scale parameter with size (see Table 4.3), which is compatible with the power law scaling of the volatility of growth rates as a function of firm size (Bottazzi et al., 2019).

Table 4.2: Median of the estimates of the shape parameter reported in Fig. (4.4).

N° firms	Profit rate	TA	Sales
200	0.94	0.88	0.70
1000	0.83	0.79	0.67
10000	0.74	0.79	0.66
Entire sample	0.66	0.72	0.58

Table 4.3: Median of the estimates of the scale parameter reported in Fig. (4.5).

N° firms	Profit rate	TA	Sales
200	0.055	0.123	0.105
1000	0.054	0.122	0.112
10000	0.053	0.119	0.117
Entire sample	0.055	0.130	0.145

Overall, our results shows that the distribution of profit rates is well described by the Laplace distribution, when we limit the analysis to the case of large long-lived firms. We observe, instead, systematic deviations from the Laplace benchmark when we include smaller firms in the sample, i.e. the smaller the firm we include the fatter the tails of the distribution of profit rate.[9]

---

[9] The LRT shows virtually identical results as Fig. (4.4) and Fig. (4.5), rejecting the Laplace distribution with p-values equal to 0 for most of the cases (see Appendix (B)).

Figure. 4.4: Estimates of the shape parameter of the Subbotin distribution of profit rates, growth of total assets and sales. Error bars show two standard errors. Results refer to the largest long-lived firms of our sample according to their sales in 2016.

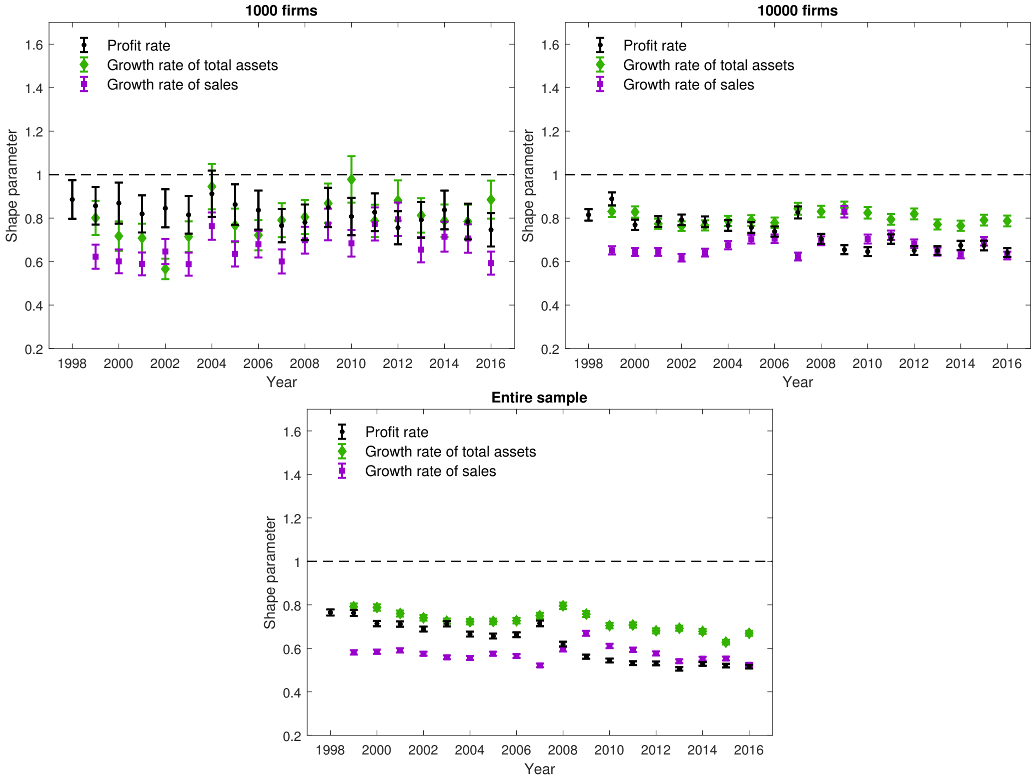
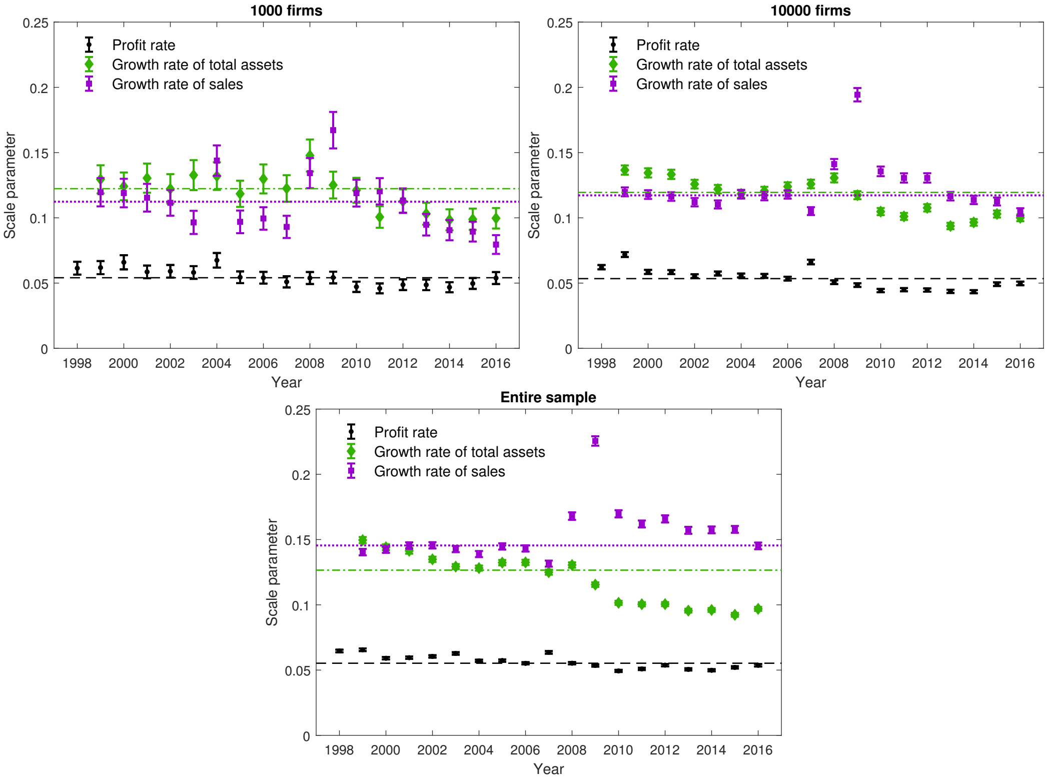


Figure. 4.5: Estimates of the scale parameter of the Subbotin distribution for profit rates, growth of total assets and sales. Error bars show two standard errors. Results refer to the largest long-lived firms of our sample according to their sales in 2016. The dashed line (profit), dotted line (sales) and dashed line with dots (total assets) in the scale parameter figure represent the median of the estimates.



Following Holly et al. (2013), in order to analyse the relation of the estimates for the profit rate distribution with the business cycle, we report in Table 4.4 the Pearson correlation coefficients between the time series of GDP growth rates with those of the estimates of m, α and σ. Regarding the parameter m, we observe a general tendency in which the correlation increases as we include smaller firms in the sample. Interestingly, the parameter σ and α of the profit rate distribution for large firms are essentially independent of the phase of the business cycle, where the only dependence is through m, which confirm the stability of the parameters over time. Such independence is instead lost as soon as we include small firms in the sample. Large firms, then, show more resilience to the business cycle, while small firms are much more dependent on the phase of the economy.

Results of the dependence for σ and α could give rise to misleading findings since we do not know which part of the distribution (right or left) is affected by the business cycle. Using the AEP helps us to understand the dynamics of the firms activity in terms of the GDP.

Table 4.4: Pearson correlation coefficient between the time series of the estimates of m, α and σ with the time series of GDP growth rates.

	m & GDP			α & GDP			σ & GDP		
N° of Firms	Profit rates	TA	Sales	Profit rates	TA	Sales	Profit rates	TA	Sales
200	0.51**	0.27	0.48**	0.07	-0.38	-0.76***	0.00	0.38	-0.14
1000	0.36	0.64***	0.58**	0.37	-0.51**	-0.68***	0.69***	0.46**	-0.25
10000	0.59***	0.81***	0.59***	0.79***	-0.21	-0.50**	0.78***	0.70***	-0.59***
Entire sample	0.63***	0.93***	0.92***	0.83***	0.40	-0.36	0.74***	0.76***	-0.76***

Our results are in line with the literature since, in the case of growth rates, Dosi and Nelson (2010), Bottazzi and Secchi (2011), Erlingsson et al. (2013) and Mundt et al. (2016) show that the growth rate distribution is more leptokurtic than the Laplace distribution. We clearly show that the Laplace distribution nicely accounts for the profit rate distribution just in the case of large and long-lived firms, with the scale parameter almost invariant over time.

4.4.2 Asymmetric case

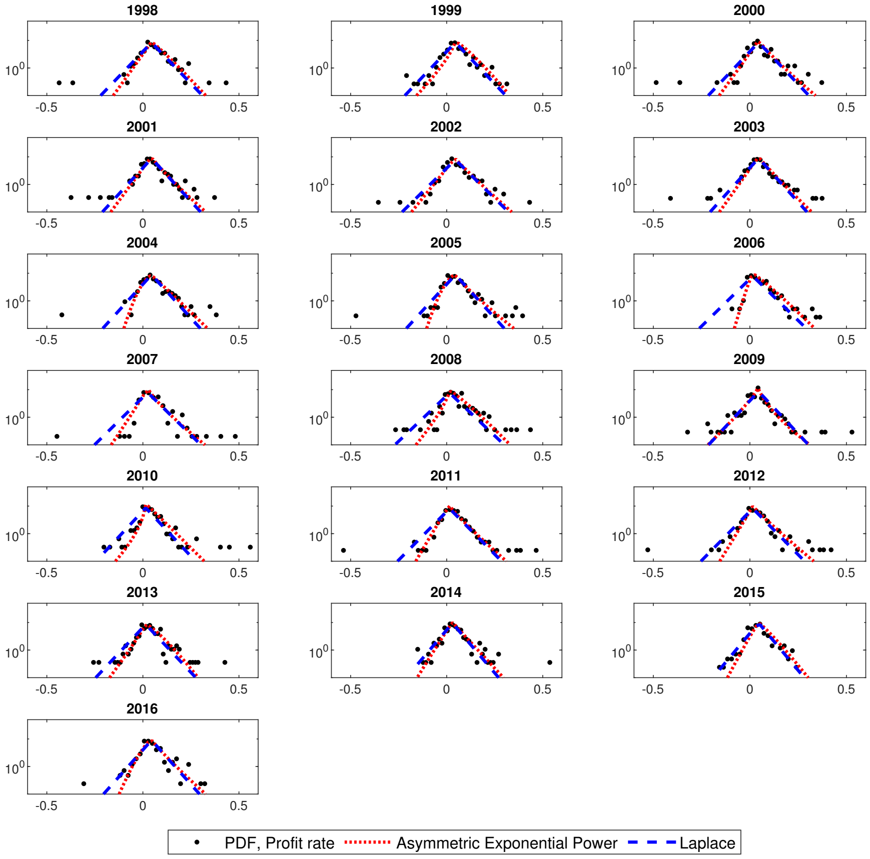
The parameters $\mathbf{p}$=(α_l, α_r, σ_l, σ_r) of the AEP distribution of profit and growth rates are estimated with the maximum likelihood method using the software SUBBOTOOLS created by Bottazzi (2004), conditional on the value of m estimated with the mode of the distribution.[10] The estimation of the slope and scale parameters are shown in Fig. (4.7) and Fig. (4.9) for the 200 largest long-lived firms.

Recall that a given AEP distribution turns out to be a symmetric Laplace as long as $\alpha_l = \alpha_r = 1$ and $\sigma_l = \sigma_r$. Considering the sample of large firms, the shape parameters of the distribution of profit rates fluctuate around the condition $\alpha_l = \alpha_r = 1$ without any systematic pattern, confirmed also by the absence of significant correlations with the growth rate of GDP (see Tables 4.7 and 4.8). The scale parameters, instead, show a significant difference most of the years, favouring the right scale parameter, i.e. $\sigma_r > \sigma_l$. Such gap widens during the housing bubble and the subsequent banking crisis, while it seems to close during the years of the economic recovery. The use of the AEP distribution makes apparent the bad fit of the symmetric Laplace benchmark for large and long-lived companies. A more appropriate model for the profit rate distribution of large firms is an asymmetric Laplace distribution with the mode correlated to the business cycle.[11]

---

[10] For the symmetric distribution the estimates are essentially independent of the chosen estimator for m, namely median, mean or mode. For the AEP distribution, using the mean or the median bias significantly the results (see Bottazzi and Secchi (2011)).

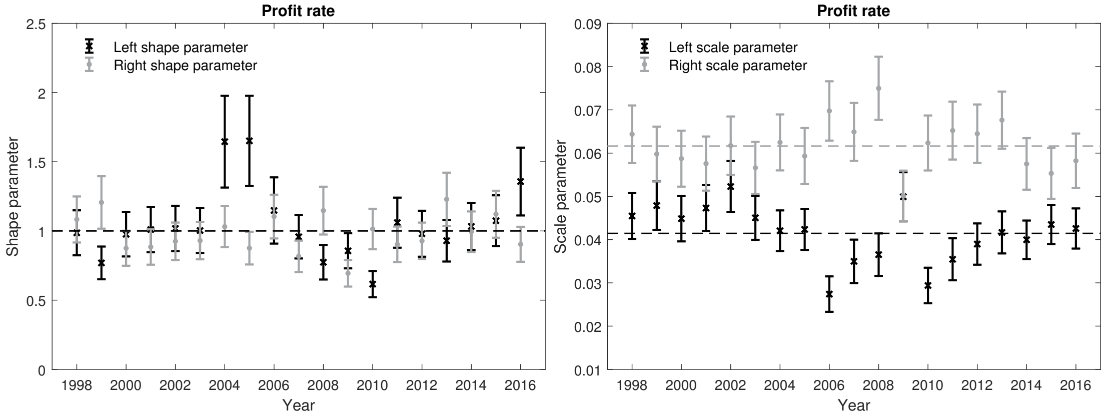
Figure. 4.6: Probability density function (PDF) of profit rates along with the AEP (dotted line) and Laplace (dashed line) distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016.



---

[11] We support these results with the LRT in Sec. (B.2) of the Appendix, in which (i) the symmetric Laplace distribution is rejected most of the years while (ii) the asymmetric Laplace is not rejected in 12 out of 19 years, compared to the AEP distribution.

Figure. 4.7: Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) for the profit rates distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016, removing the two extreme values in each year. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively.



The distribution of growth rates of total assets and sales of the largest long-lived firms (see Fig. (4.9)) are characterised by a strong deviation from the Laplace distribution and a high level of volatility, which is in line with the literature (see for instance,Fu et al., 2005, Bottazzi and Secchi, 2006 and Dosi and Nelson, 2010)

Figure. 4.8: Probability density function (PDF) of growth rates of total assets along with the AEP (dotted line) and Laplace (dashed line) distributions. The results refer to the 200 largest long-lived firms according to their sales in 2016.

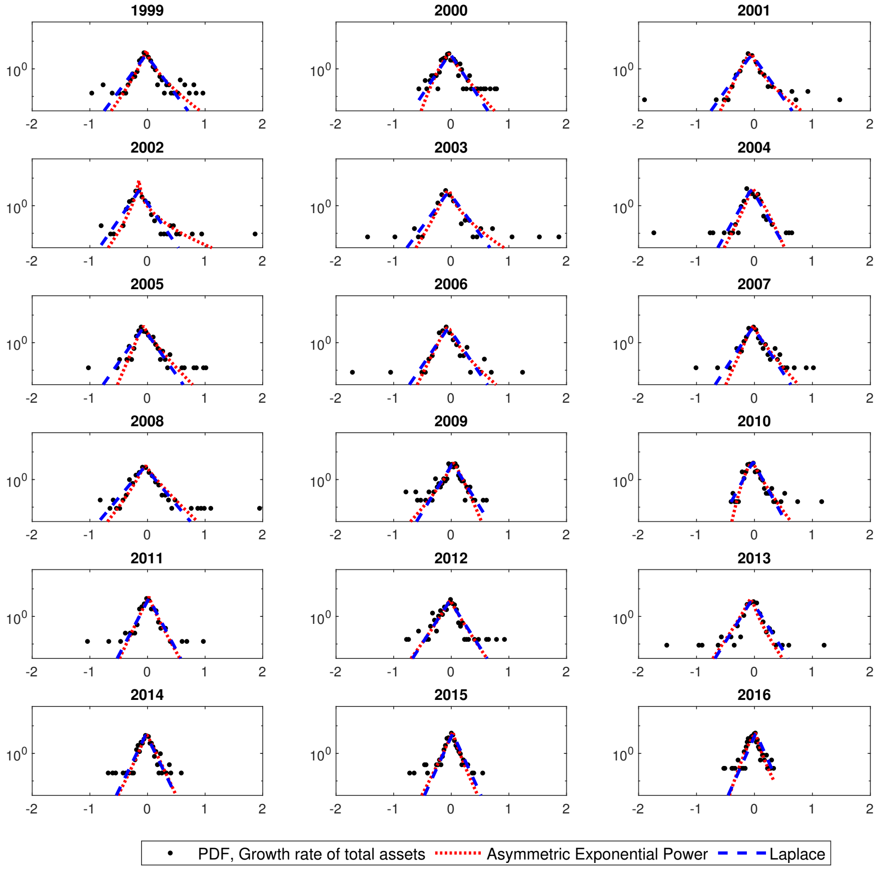
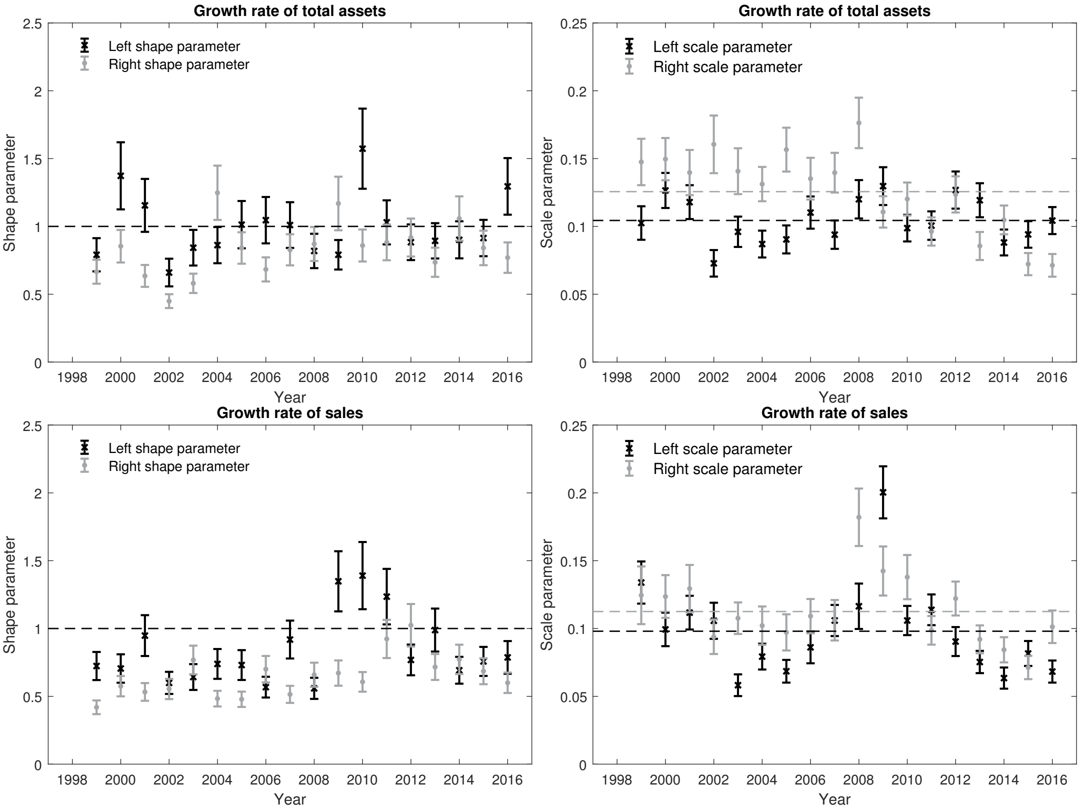


Figure. 4.9: Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) for growth rates of total assets and sales. Results refer to the 200 largest long-lived firms according to their sales in 2016. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively.



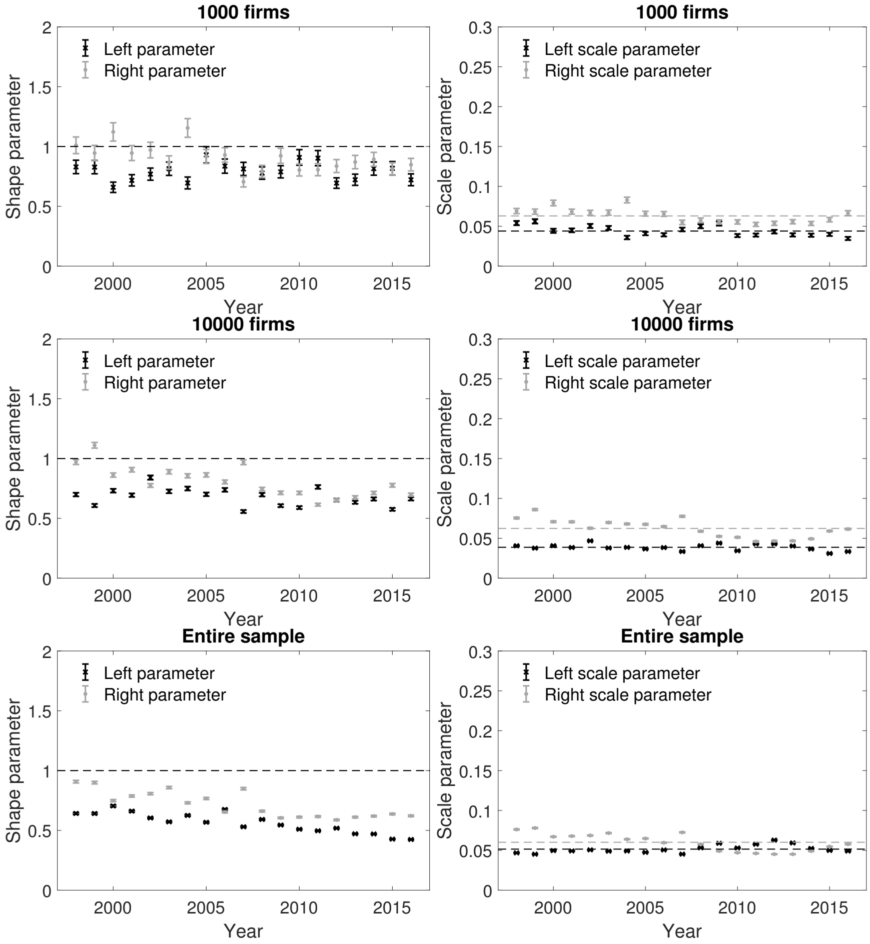
Distributional properties conditional on size and business cycle phase

In Figs. (4.10), (4.11) and (4.12), we report the estimates of the shape and scale parameters of the AEP computed for profit and growth rates of total assets and sales, conditional on size.

Regarding profit rates, we observe that α_l and α_r are significantly smaller than 1 most of the years. The shape of the distribution of profit rates depends on the size of the firms becoming fatter the smaller are the firms included in the sample. The scale parameter, instead, shows a remarkable stability as a function of the size, with the systematic tendency $\sigma_r > \sigma_l$.

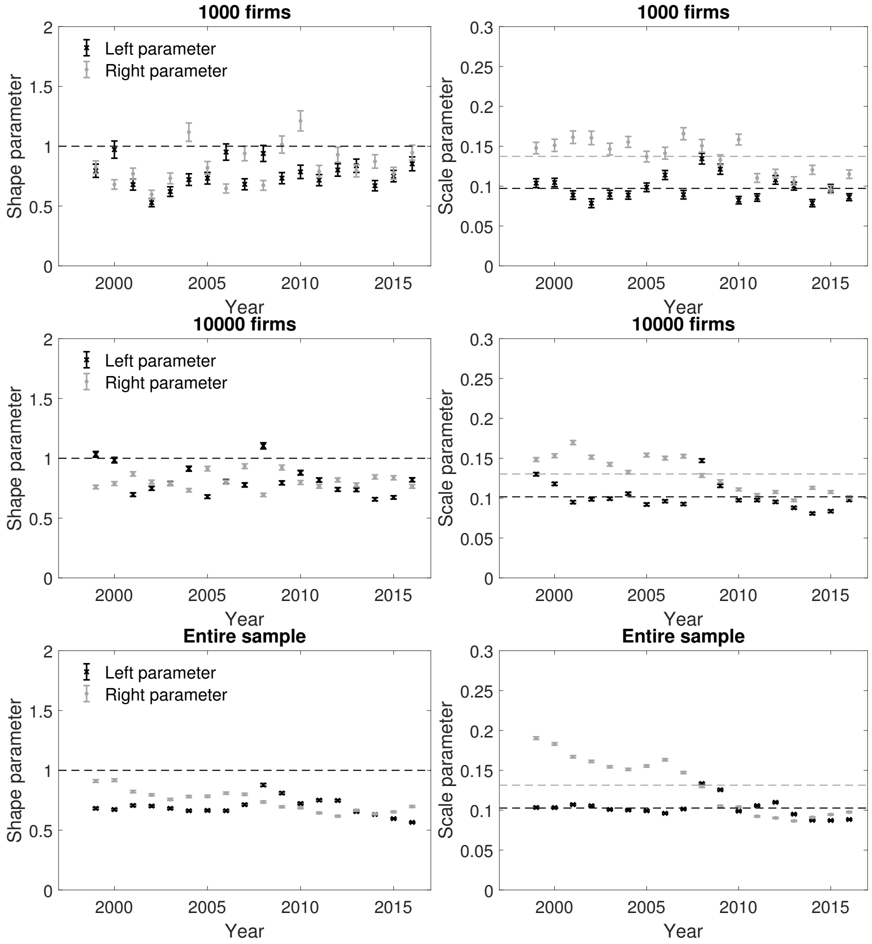
This condition changes when we consider the entire sample. The dispersion on the left side, measured by σ_l, is higher than the σ_r during the phase of the crisis. This change can be attributed to the effect of the business cycle on the profitability of small firms. As can be observed in Tables 4.7 and 4.8, the correlation between the AEP estimated and the GDP growth rates is stronger when including small firms. This result underlines the robustness of the profitability of large firms to the business cycle phase, while the small firms seem to be more affected by the adverse phase of the cycle.

Figure. 4.10: Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) of the AEP distribution for profit rates conditional on size. Error bars show two standard errors. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively.



Focusing on growth rates of total assets, we always reject the Laplace distribution hypothesis, due to the differences in the scale parameters and shape parameters, i.e. $\alpha_l \neq \alpha_r$ and $\sigma_l \neq \sigma_r$. Moreover, the estimates of the shape parameters are different from 1 most of the years. Interestingly, the scale parameters show a similar behaviour to profit rates since the cross-sectional volatility is higher on the right side (σ_r) for the large firms but, when analysing the entire sample, we identify a remarkable decrease/increase of the cross sectional volatility on the right/left side during the crisis period.

Figure. 4.11: Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) of the AEP distribution for growth rates of total assets conditional on size. Error bars show two standard errors. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively.



Finally, in relation to growth rates of sales, the Laplace distribution is also rejected since $\alpha_l \neq \alpha_r \neq 1$ and $\sigma_l \neq \sigma_r$. When including the smallest firms in the analysis, we observe a higher volatility on the left part of the distribution compared to the right one during the downturn, which is consistent with the results reported for profit rates and growth rates of total assets.[12] Thus, with Figs. (4.10), (4.11) and (4.12), we are able to underline the effect of the crisis on small firms by means of the scale parameters of profit and growth rates. On the other hand, in relation to the shape parameters, we observe a different dynamics between profit and growth rates. More specifically, profit rates tend to be more leptokurtic on both parts of the distribution when including smaller firms on the sample. However, the shape parameters of growth rates on the left part of the distribution become more platikurtic (i.e. we observe a slimming down of the left tail) during the crisis period, compared to the right tail. This particular behaviour has been already reported with the 200 largest long-lived firms (see Figs. (4.9) and (B.2)) in which we observe that, during the downturn, growth rates show a higher dispersion on the left part of the distribution with a slimming down of the left tail.[13]

---

[12] The main difference from profit rates and growth rates of total assets is found on the dispersion of the right part of the distribution of growth rates of sales given that it is quite constant regardless of the crisis, as can be observed in Fig. (4.12) and Table 4.8. This feature can be attributed to the natural volatility of sales in both parts of the distribution (Bottazzi et al., 2019).

Figure. 4.12: Estimates of the two shape parameters (α_l and α_r) and two scale parameters (σ_l and σ_r) of the AEP distribution for growth rates of sales conditional on size. Error bars show two standard errors. Gray and black dashed lines refer to the mean of the estimates of σ_l and σ_r, respectively.

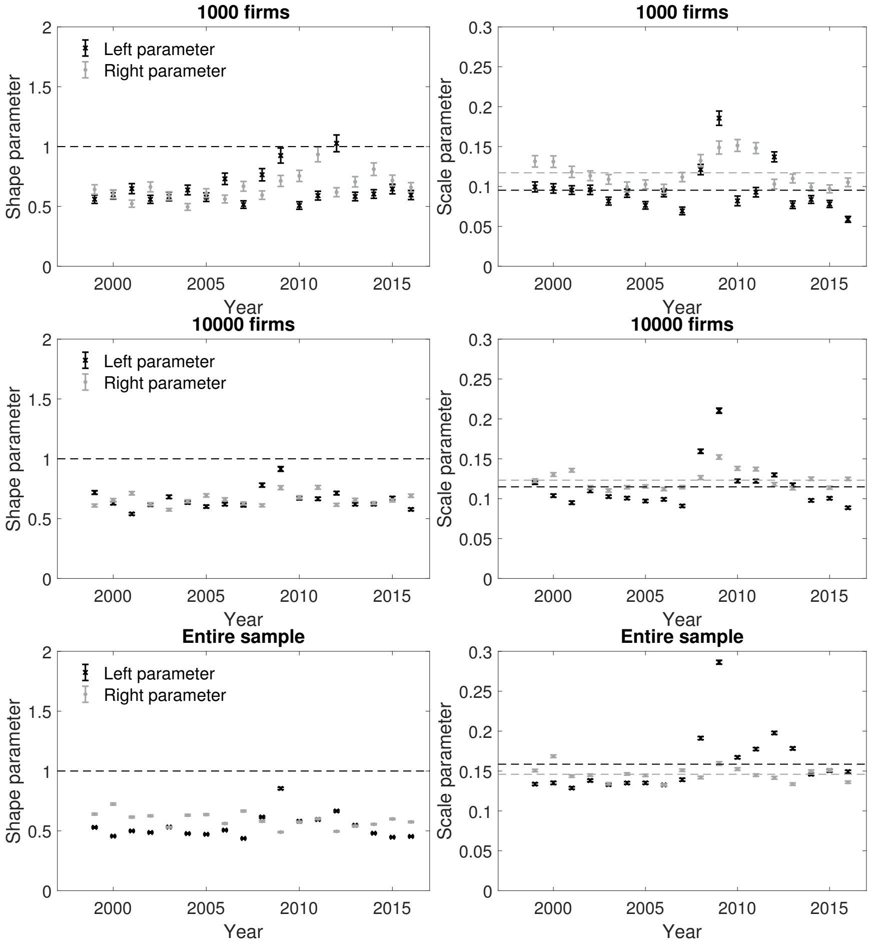


Table 4.5: Median of the estimates α_l and α_r reported in Figs. (4.10), (4.11) and (4.12).

	α_l			α_r		
N° of firms	Profit rates	TA	Sales	Profit rates	TA	Sales
200	1.00	0.91	0.75	0.93	0.84	0.63
1000	0.81	0.74	0.60	0.89	0.81	0.65
10000	0.69	0.79	0.63	0.78	0.80	0.65
Entire sample	0.57	0.68	0.50	0.66	0.75	0.59

---

[13] This feature can be clearly observed in Fig. (B.2) in 2009 and 2010.

Table 4.6: Median of the estimates σ_l and σ_r reported in Figs. (4.10), (4.11) and (4.12).

	σ_l			σ_r		
N° of firms	Profit rates	TA	Sales	Profit rates	TA	Sales
200	0.04	0.10	0.09	0.06	0.13	0.11
1000	0.04	0.09	0.09	0.07	0.14	0.11
10000	0.04	0.10	0.10	0.06	0.13	0.12
Entire sample	0.05	0.10	0.14	0.06	0.14	0.14

Table 4.7: Pearson correlation coefficient between the time series of the estimates of α_l and α_r with the time series of GDP growth rates.

	α_l & GDP			α_r & GDP		
N° of Firms	Profit rates	TA	Sales	Profit rates	TA	Sales
200	0.35	0.05	-0.61***	0.07	-0.43	-0.68***
1000	-0.04	-0.04	-0.49**	0.4	-0.4	-0.61***
10000	0.33	0.16	-0.52**	0.77***	-0.02	-0.35
Entire sample	0.63***	-0.33	-0.72***	0.78***	0.83***	0.71***

Table 4.8: Pearson correlation coefficient between the time series of the estimates of σ_l and σ_r with the time series of GDP growth rates.

	σ_l & GDP			σ_r & GDP		
N° of Firms	Profit rates	TA	Sales	Profit rates	TA	Sales
200	0.07	-0.38	-0.30	0.21	0.54**	-0.11
1000	0.12	-0.14	-0.46*	0.73***	0.58**	-0.35
10000	-0.26	0.11	-0.61***	0.86***	0.83***	-0.45*
Entire sample	-0.88****	-0.21	-0.81***	0.89***	0.87***	0.00

4.5 Conclusion

In this paper, we shed some light on the firm dynamics literature by analysing on what extent the Laplace distribution describes the Spanish long-lived firms distribution of profit and growth rate, against its alternative more general distributions, namely Subbotin and AEP. Moreover, compared to recent literature, we analyse the effect of the different phases of the business cycle and the firm size on the distributional characteristics of profit and growth rates.

We find evidence of systematic deviations of the profit rate distribution from the Laplace benchmark when small firms are included in the analysis. The empirical distribution becomes more leptokurtic without changing the scale parameters. Therefore, the Laplace benchmark turns out to be a reasonable approximation if we limit the sample to large firms. Relaxing the symmetric constraint, the use of the AEP distribution shows that, instead of a Laplace, the better approximation for firm profit rate distribution seems to be an asymmetric Laplace. Interestingly, except for the location parameter, the shape and scale parameters do not depend on the business cycle phase. Small firms, instead, show a much higher dependence of their profit rates on the business cycle phase, signalling again a marked difference with large firms. Taking into account these results, we underline the robustness of the large firms during the financial crisis in terms of profitability given (i) the significant larger dispersion of the right part of the distribution, compared to the left one, and (ii) the absence of relation between the time series of GDP growth rates and the time series of the estimates of σ_l, σ_r, α_l and α_r for the largest 200 long-lived firms' profit rates. However, this robustness is lost when including small firms in the sample since (i) we observe that the dispersion of the left part of the distribution is significantly larger than the right one during the years of the downturn, and (ii) the estimates of the entire sample show a remarkable relation with the GDP growth rates.

Finally, focusing on growth rates, we observe a similar tendency compared to profit rates given the effect of the crisis on small firms growth distribution ($\sigma_r < \sigma_l$). This result is supported by the stronger correlation between the time series of the estimate parameters and GDP growth rates when including small firms in the sample. Interestingly, we observe that profit and growth rates of total assets show a similar dynamics in terms of dispersion, while growth rates of total assets and sales are more similar regarding the shape of the distribution.

Overall, given our outcome, this study underlines the relevance of studying the business cycle and firm size given that the small firms behave differently to the large firms during the turmoil.

Chapter 5

Concluding remarks

To conclude this thesis, we summarise the main results, underlining the future lines of research.

In the second chapter of the dissertation, we have proposed an investor sentiment index based on the collective movement of stock prices in a given market, inspired by the Bank of America Merrill Lynch global breath rule. We show that the time evolution of the sentiment index can be reasonably described by the herding model proposed by Kirman in his seminal paper "Ants, rationality and recruitment" (Kirman, 1993). The correspondence between the index and the model allowed us to easily estimate its parameters. Based on the model and the empirical evolution of the sentiment index, we propose an early warning indicator able to identify optimistic and pessimistic phases of the market. As a result, investors and policy-makers can set different strategies anticipating financial market instability. Investors can reduce the risk of their portfolio while policy-makers can set more efficient policies to avoid the effects of financial instability on the real economy. The validity of our results is supported by means of a robustness analysis showing the application of the early warning indicator in eight different worldwide stock markets. In our future research, we will study how the results presented in this thesis change when including different network topologies in the model. Given that the current chapter considers the existence of a fully connected network, we will be able to observe how the market behaves with more realistic network topologies like the core-periphery network among others.

In the third chapter, we have proposed a behavioural model able to describe the internal dynamics of Bitcoin in terms of the interaction of two contrasting groups of agents. In particular, this model describes the ratio of two consecutive Bitcoin prices as the odds ratio of bullish to bearish traders, which is modeled by a stochastic dynamics and supported by a switching behavioral model. Using the model proposed, we contribute to the literature in different ways. First, we provide an explicit formula for the price drift dynamics as a function of the behavioral parameters. Second, we obtain satisfactory results when forecasting Bitcoin price and volatility. Third, we can use our model as an early warning indicator system to detect bullish periods. In the next studies related to this chapter, we will apply our model to different financial markets and frequencies, in order to prove its robustness and proper performance.

Finally, in the fourth chapter of this thesis, we analyse the evolution of the empirical cross-correlation distribution of profit and growth rates comparing its fitness to the Laplace distribution, considered as a benchmark, additionally to the Subbotin and the Asymmetric Exponential Power distributions. Unlike recent literature, we analyse how the evolution of the distribution depends on the size of the firms and business cycle phase. Our results show that the profit rates of large firms seem to be characterised by an asymmetric Laplace with parameters largely independent of the business cycle. Small firms, instead, are characterised by the AEP distribution, which accounts for the dependence of profit and growth rates on firm size and phase of the business cycle. We also observe that the largest firms are more robust to downturns, in terms of profitability, compared to the small firms given their different distributional characteristics during crisis periods. This result is extended to growth rates, which confirms the dependence of small firms on the business cycle phase. In our future research, we will study whether differences exit in different countries in terms of their distributional properties.

Appendices

Appendix A

An agent-based early warning indicator for financial market instability

A.1 Derivation of the Fokker-Planck equation

Following Alfarano et al. (2005) and Alfarano (2006), the Fokker-Planck equation can be derived as a second-order Taylor approximation for the continuum of our population dynamics considering Eq. (2.10). The transition probabilities in Eq. (2.10) have to be rewritten as a function of the intensive variable z, instead of the extensive variable n. The relation between the transition rates (2.10) for n and z is given by the following formula:

$$\pi_n^{\pm} = \rho(n'_{t+\Delta t_0} = n_t \pm 1 \mid n_t) = N^2 \pi_z^{\pm} \tag{A.1}$$

with:

$$\begin{aligned} \pi(z \to z + 1/N) = \pi_z^{+} = (1-z)\left(\tfrac{a_1}{N} + bz\right), \\ \pi(z \to z - 1/N) = \pi_z^{-} = z\left(\tfrac{a_2}{N} + b(1-z)\right). \end{aligned} \tag{A.2}$$

Following the notation used by Van Kampen (1992), we introduce the "step" operators $\mathbf{E}$ and $\mathbf{E}^{-1}$. Their effect on an arbitrary function $f(n)$ are respectively to add to or to drop off one unit to their integer argument n. Formally:

$$\mathbf{E}[f(n)] = f(n+1) \quad \text{and} \quad \mathbf{E}^{-1}[f(n)] = f(n-1). \tag{A.3}$$

Using these operators $\mathbf{E}$ and $\mathbf{E}^{-1}$, Alfarano (2006) rewrites the Chapman-Kolmogorov equation for the one-step process:

$$\frac{\partial}{\partial t} P_{n,t} = (\mathbf{E} - 1)\left[\pi_n^- P_{n,t}\right] + \left(\mathbf{E}^{-1} - 1\right)\left[\pi_n^+ P_{n,t}\right], \tag{A.4}$$

where $P_{n,t}$ is the probability to have n agents at time t in State 1. The time derivative in the left-hand side of Eq. (A.4) stands for the continuous time approximation for the evolution of the probability $P_{n,t}$. This is called, in the pertinent literature, Master equation.

Using the variable z and excluding the obvious time dependence of $P_{z,t}$, the Master equation, Eq. (A.4), can be approximated with the Fokker-Plank equation obtaining the form:

$$\frac{\partial}{\partial t} P_z = (\mathbf{E} - 1)\left[\pi_z^- N^2 P_z\right] + \left(\mathbf{E}^{-1} - 1\right)\left[\pi_z^+ N^2 P_z\right], \tag{A.5}$$

where the transition rates are given by Eq. (A.2). Note that the probabilities are invariant under this transformation, therefore $P_n = P_z$. The probability density of the variable z is defined according to the following limit:

$$p_{z,t} = \lim_{N\to\infty} \frac{P_{z,t}}{\Delta z} = \lim_{N\to\infty} N P_{z,t}, \tag{A.6}$$

where p_z is a continuous function of z. Eq. (A.6) can be rewritten as:

$$\frac{\partial}{\partial t} p_z = N^2 \left\{ (\mathbf{E} - 1)\left[\pi_z^- p_z\right] + \left(\mathbf{E}^{-1} - 1\right)\left[\pi_z^+ p_z\right] \right\}. \tag{A.7}$$

Using the Taylor expansions for $\mathbf{E}$ and $\mathbf{E}^{-1}$ up to the second order, Alfarano (2006) ends up with:

$$\frac{\partial}{\partial t} p_z = N^2 \left\{ -\Delta z \frac{\partial}{\partial z}\left[\left(\pi_z^+ - \pi_z^-\right) p_z\right] + \frac{1}{2}\Delta z^2 \frac{\partial^2}{\partial z^2}\left[\left(\pi_z^+ + \pi_z^-\right) . p_z\right] \right\} \tag{A.8}$$

The N^2 factor in front of the equation disappears, given the independence of the overall dynamics of the number of agents N. The resulting expression is the Fokker-Planck equation in Eq. (2.13).

A.2 Derivation of the equilibrium distribution

To compute the equilibrium distribution $p_0(z)$, we use the standard formula (Van Kampen, 1992):

$$p_0(z) = \frac{K(\varepsilon_1, \varepsilon_2)}{D(z)} \exp\left(\int^z \frac{2A(y)}{D(y)} dy \right) \tag{A.9}$$

obtained with the condition $\frac{\partial}{\partial t} p(z,t) = 0$ and natural boundary conditions.

Calculating the integral, we obtain the following formula for $p_0(z)$:

$$p_0(z) = K\left(\varepsilon_1, \varepsilon_2\right) z^{\varepsilon_1 - 1} \cdot (1 - z)^{\varepsilon_2 - 1} \tag{A.10}$$

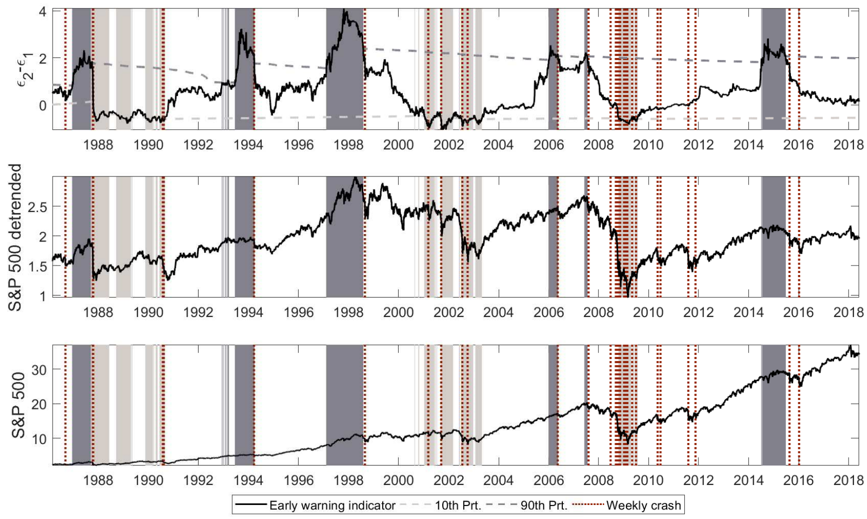
The normalization constant $K(\cdot,\cdot)$ follows from $\int p_0(z)dz = 1$, which turns out to be:

$$\frac{1}{K(\varepsilon_1,\varepsilon_2)} = \int_0^1 z^{\xi_1-1}(1-z)^{\varepsilon_2-1}dz = \frac{\Gamma(\varepsilon_1)\Gamma(\varepsilon_2)}{\Gamma(\varepsilon_1+\varepsilon_2)} = B(\varepsilon_1,\varepsilon_2)\,. \tag{A.11}$$

A.3 Robustness analysis of the determination of the thresholds

In order to study the robustness of the results of Fig. (2.10), we computed the two thresholds at the 10^{th} and 90^{th} percentile using past data instead of the entire sample. So, we computed the percentiles considering (i) 500 data points, from Day 1 to Day 500; (ii) adding the other values of Λ_t, from Day 501 to the end of the time series and computing the new values of the thresholds. As we can see from Fig. (A.1), the results were essentially unchanged with respect to Fig. (2.10).

Figure. A.1: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. S&P 500 index.



Note. Bull market phases and the subsequent negative events: 1987 (Black Monday), 1994 (tightening monetary policy), 1997-1998 (Russian financial crisis), 2006 (tightening monetary policy), 2007 (subprime mortgage crisis) and 2015-2016 (weakness of the Chinese economy). Bear market phases: 1987-1990 (Black Monday and Gulf War), 2001-2003 (burst of the dot-com bubble) and 2008-2009 (burst of the housing bubble).

A.4 Early warning indicators

Figure. A.2: Standardized early warning indicators of all the international markets. Light and areas represent the US Business Cycle contractions according to the National Bureau of Economic Research.

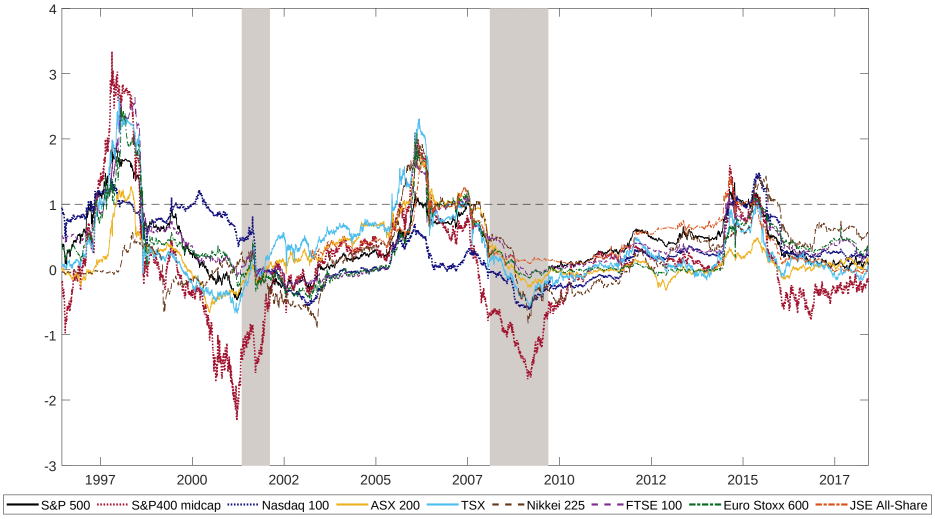
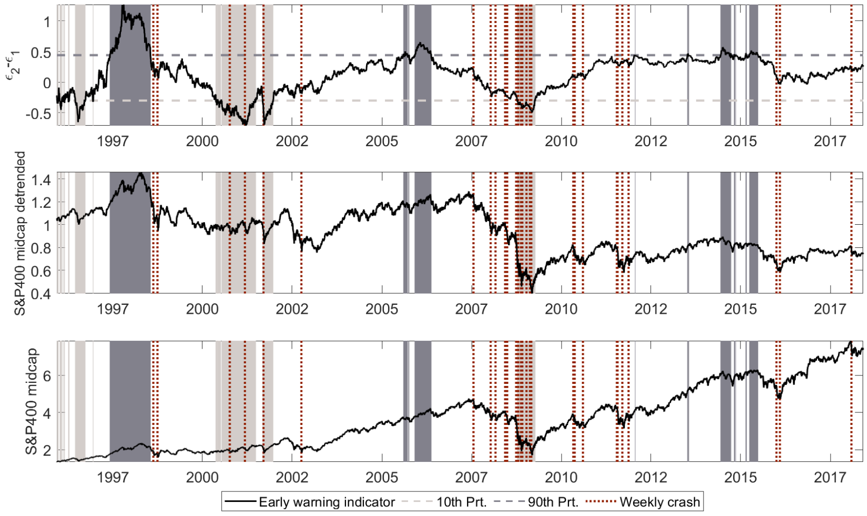
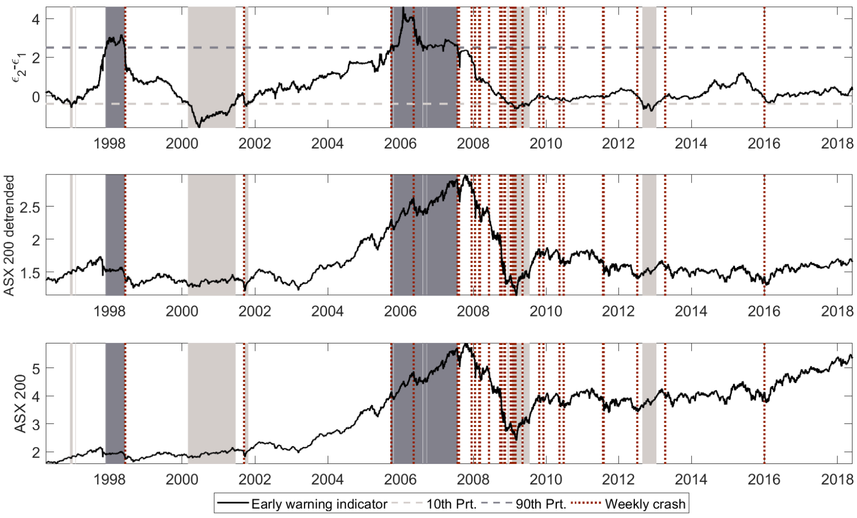


Figure. A.3: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. S&P 400 midcap index.



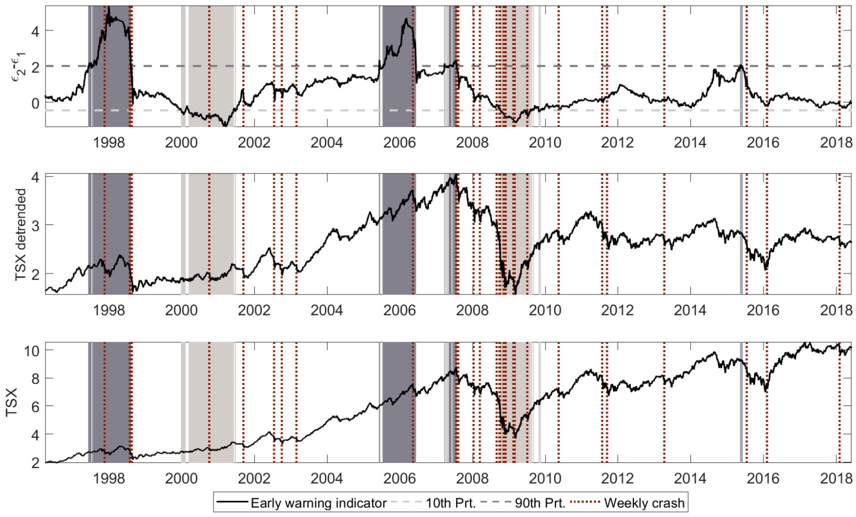
Note. Bull market phases and the subsequent negative events: 1997-1998 (Russian financial crisis), 2006 (subprime mortgage crisis) and 2014-2015 (weakness of the Chinese economy). Bear market phases: 2000-2001 (burst of the dot-com bubble) and 2008-2009 (burst of the housing bubble).

Figure. A.4: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. ASX 200 index.



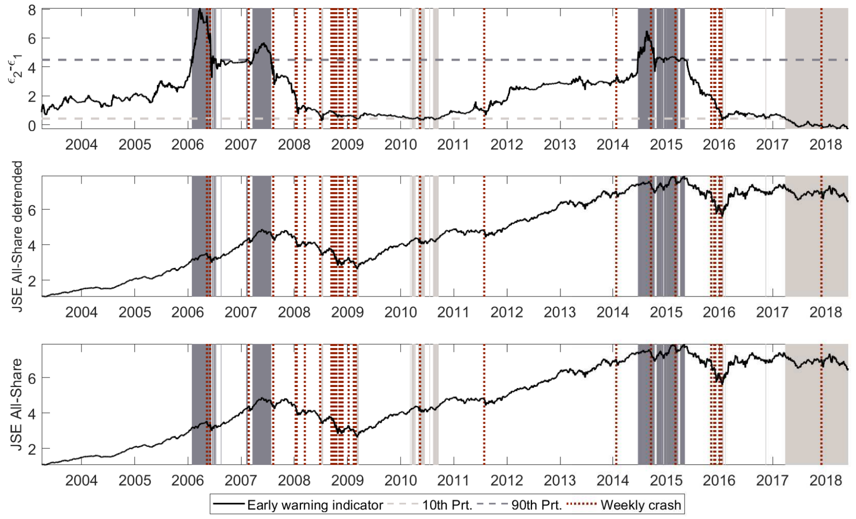
Note. Bull market phases and the subsequent negative events: 1997-1998 (Russian financial crisis), 2006 (tightening monetary policy) and 2007 (subprime mortgage crisis). Bear market phases: 2000-2001 (burst of the dot-com bubble) and 2009 (burst of the housing bubble).

Figure. A.5: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. TSX index.



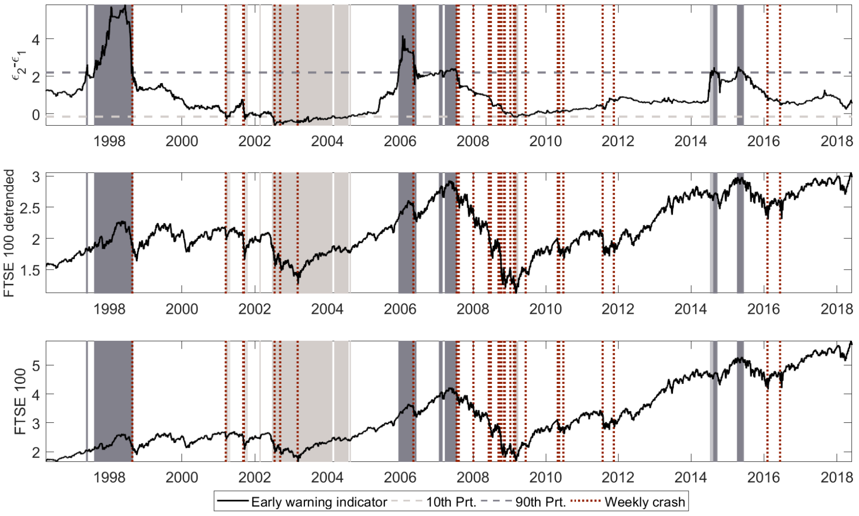
Note. Bull market phases and the subsequent negative events: 1997-1998 (Russian financial crisis), 2006 (tightening monetary policy), 2007 (subprime mortgage crisis) and 2015 (weakness of the Chinese economy). Bear market phases: 2000-2001 (burst of the dot-com bubble) and 2008-2009 (burst of the housing bubble).

Figure. A.6: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. JSE All-Share index.



Note. Bull market phases and the subsequent negative events: 2006 (tightening monetary policy), 2007 (subprime mortgage crisis) and 2014-2015 (weakness of the Chinese economy). Bear market phases: 2008-2009 (burst of the housing bubble) and 2016 (Chinese financial crash).

Figure. A.7: Early warning indicator using a 100-day EMA and a time interval of 750 days for the estimation of parameters. Light and dark gray areas represent the 10^{th} (bear market) and 90^{th} (bull market) percentiles. The dotted red line denotes the 30 most negative weekly returns. FTSE 100 index.



Note. Bull market phases and the subsequent negative events: 1997-1998 (Russian financial crisis), 2006 (tightening monetary policy), 2007 (subprime mortgage crisis) and 2014-2015 (weakness of the Chinese economy). Bear market phases: 2002-2004 (burst of the dot-com bubble) and 2009 (burst of the housing bubble).

A.5 Worldwide stock markets

In this section we show the analysis of the international stock markets. More specifically, we report (i) the computation of the sentiment index, (ii) tests of volatility clustering, (iii) box plots that show the moments obtained from the 500 Monte Carlo simulations for each market, (iv) the effect of the parameter L on ε_1, ε_2, b and the ACF, and (v) the autocorrelation function.

A.5.1 Sentiment index

As with the S&P500 in Fig. (1) of the paper, we compute the sentiment index for each market by using L=100. In all the financial markets it is possible to observe remarkable swings of the index.

Figure. A.8: Sentiment index and index changes of the S&P 400 midcap.

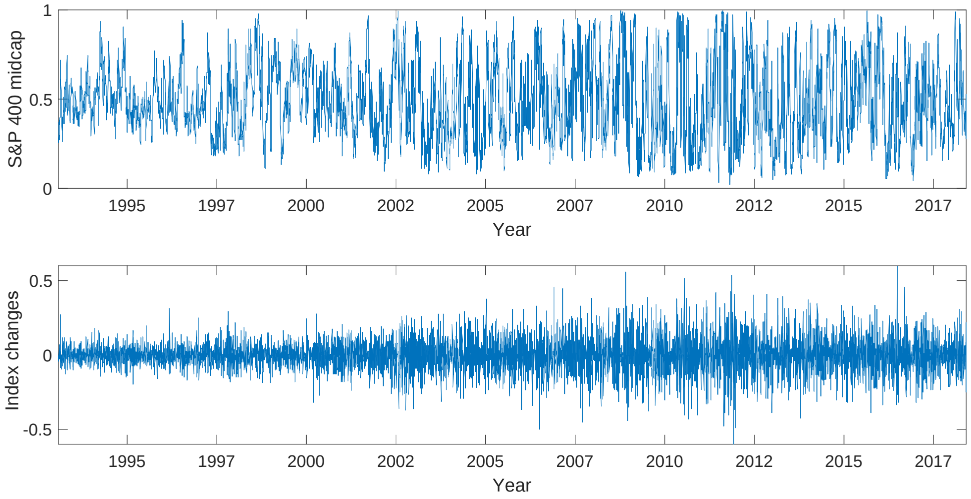


Figure. A.9: Sentiment index and index changes of the Nasdaq 100.

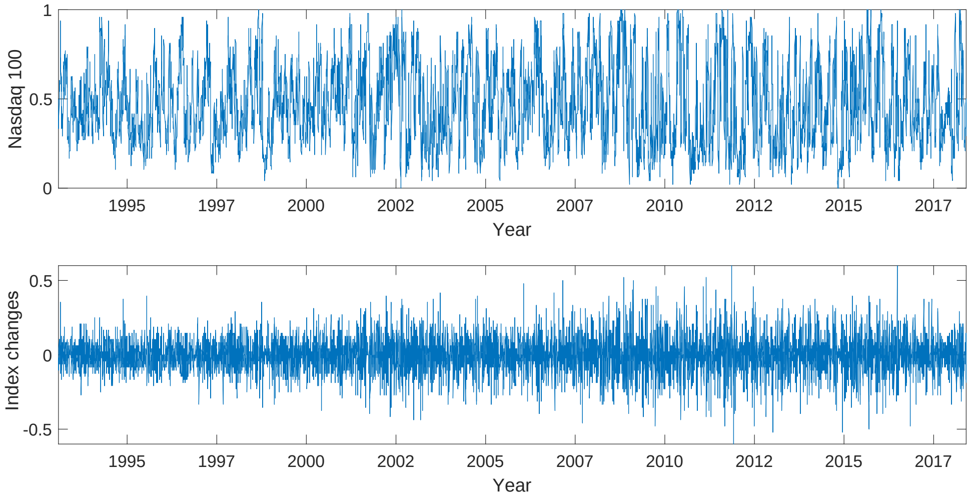


Figure. A.10: Sentiment index and index changes of the ASX 200.

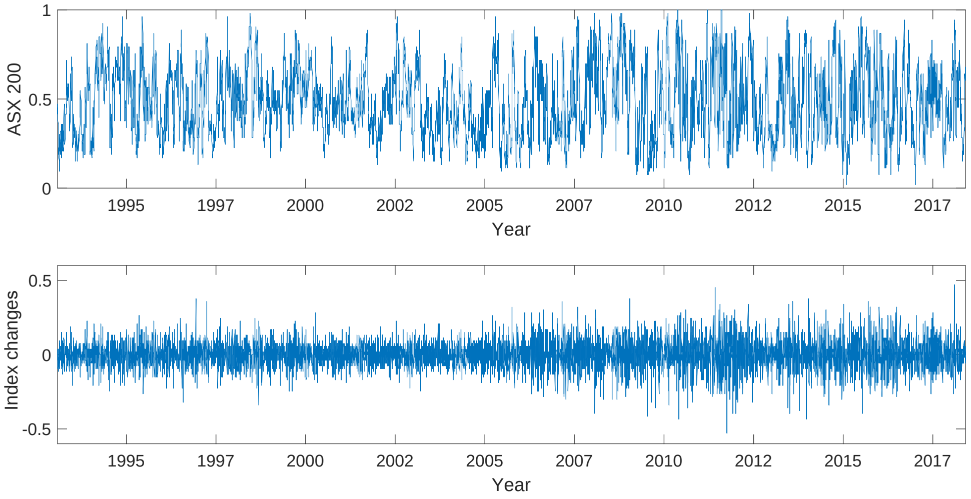


Figure. A.11: Sentiment index and index changes of the TSX.

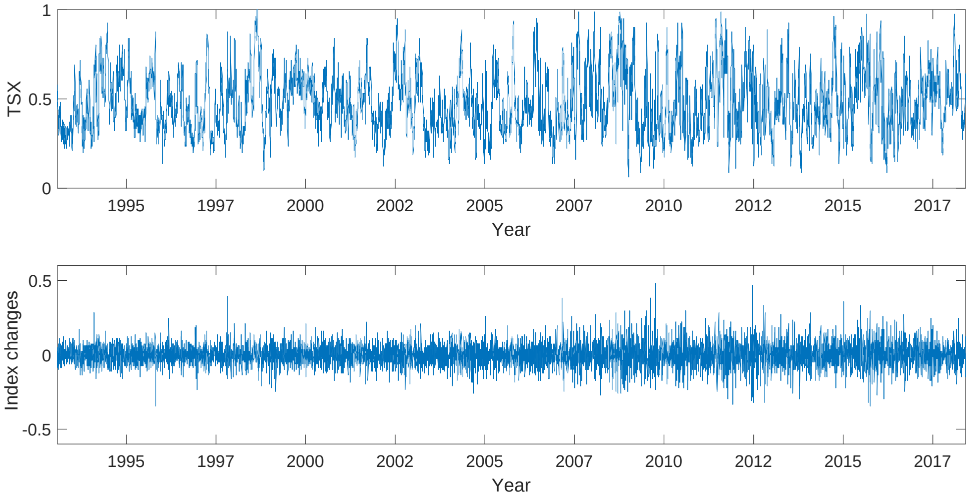


Figure. A.12: Sentiment index and index changes of the Euro Stoxx 600.

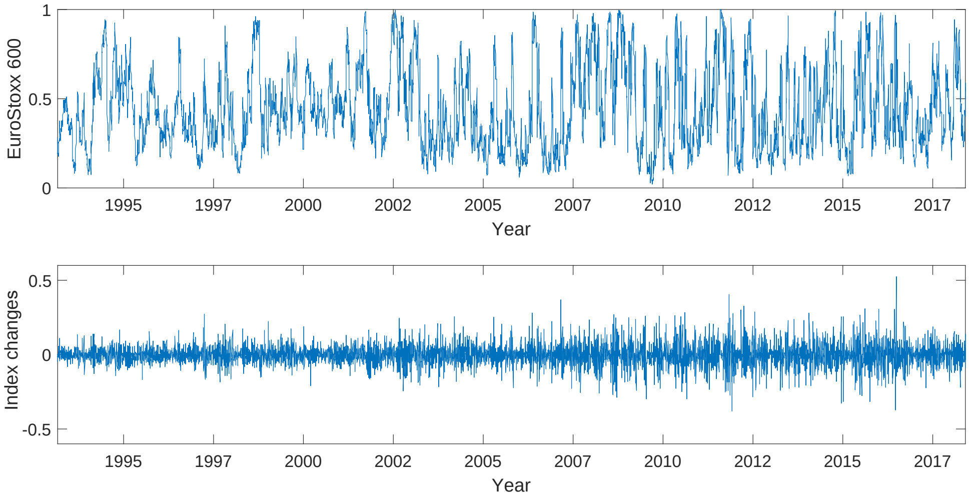


Figure. A.13: Sentiment index and index changes of the Nikkei 225.

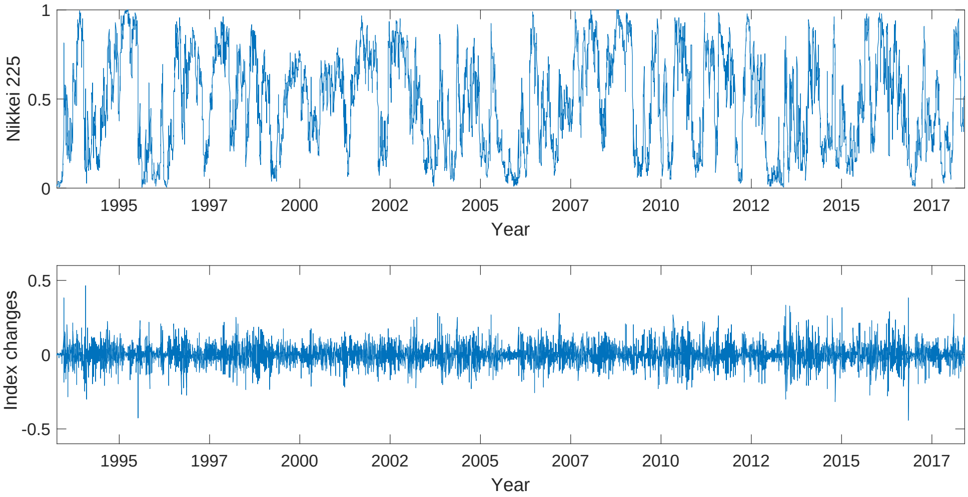


Figure. A.14: Sentiment index and index changes of the JSE All-Share.

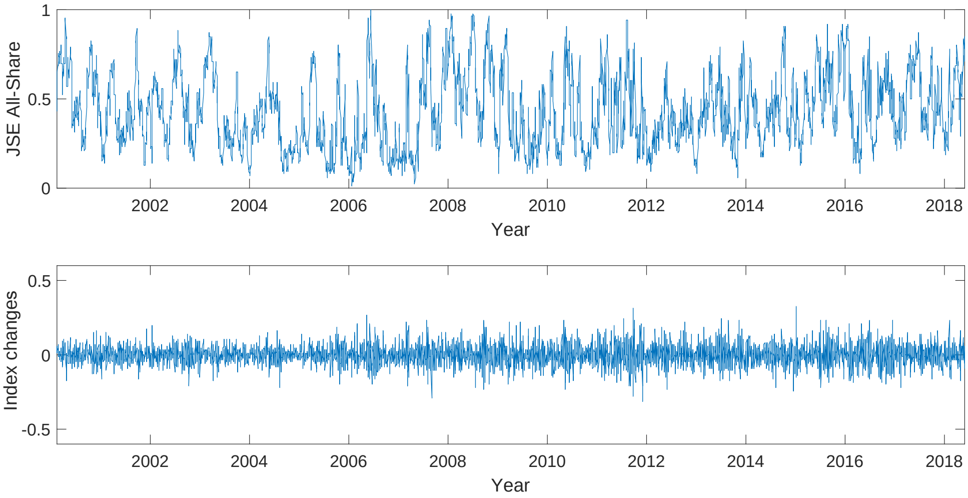
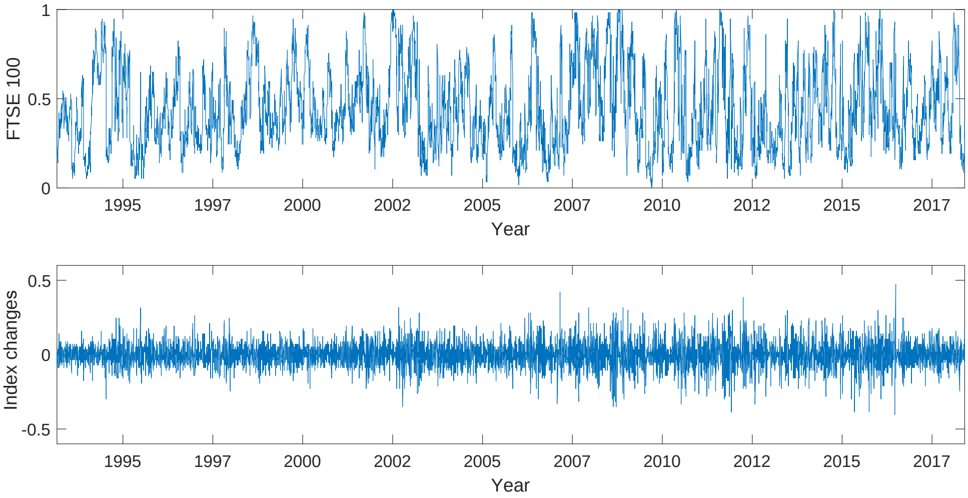


Figure. A.15: Sentiment index and index changes of the FTSE 100.



A.5.2 Tests of volatility clustering

In order to test the existence of volatility clustering we employ two traditional tests of the financial literature: McLeod-Li test (McLeod and Li, 1983) and Engle's ARCH test (Engle, 1982).

On the one hand, for the empirical data, we simply apply the corresponding tests to the sentiment indices. However, for the simulated data, we compute 500 Monte Carlo simulations to generate sentiment indices by using the conditional likelihood with the parameters reported in Table 4 and Table 5 of the paper. Consequently, for each stock market we calculate the median of the *p*-values obtained by the tests from the 500 Monte Carlo simulations. As can be observed in Table A.1, the sentiment indices computed with empirical data are characterised by volatility clustering without any exception. In the case of simulated data we generally observe the existence of conditional heteroskedasticity in all the markets. The only exception is found in the TSX index given that the median of the *p*-values is slightly higher than the 10% significance level. This fact can be explained due to the low herding parameter reported for this market, equal to 0.0036, which signals a low level of fluctuations in the changes of the sentiment index.

Table A.1: *P*-values of the McLeod-Li and Engle's ARCH test for empirical data, and median of the *p*-values obtained from the 500 Monte Carlo simulations (Langevin equation) for simulated data.

	Simulated data		Empirical data	
Stock market	McLeod-Li test	Engle's ARCH test	McLeod-Li test	Engle's ARCH test
S&P 400	0,0420	0,0341	0,0000	0,0000
Nasdaq	0,0219	0,0172	0,0000	0,0000
ASX 200	0,0695	0,0639	0,0000	0,0000
TSX	0,1189	0,1068	0,0000	0,0000
Nikkei	0,0000	0,0000	0,0000	0,0045
FTSE 100	0,0000	0,0000	0,0000	0,0000
Euro Stoxx 600	0,0000	0,0000	0,0000	0,0000
JSE All-Share index	0,0000	0,0000	0,0000	0,0027

A.5.3 Box plots of the moments of simulated data

We also analyse the similarities of the empirical index and our simulated series by means of some moments (mean, median and standard deviation). We compute 500 Monte Carlo simulations with the parameters estimated from the conditional likelihood reported in Table 4 and Table 5 of the paper. The results show that the moments of the empirical data are in line with those of the sentiment indices generated by the Langevin equation, since the mean, median and standard deviation are inside the 95% confidence level. The only exceptions are the Nikkei 225 and FTSE 100 in which the moments are inside the 99% confidence level.

Figure. A.16: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. S&P 400 midcap.

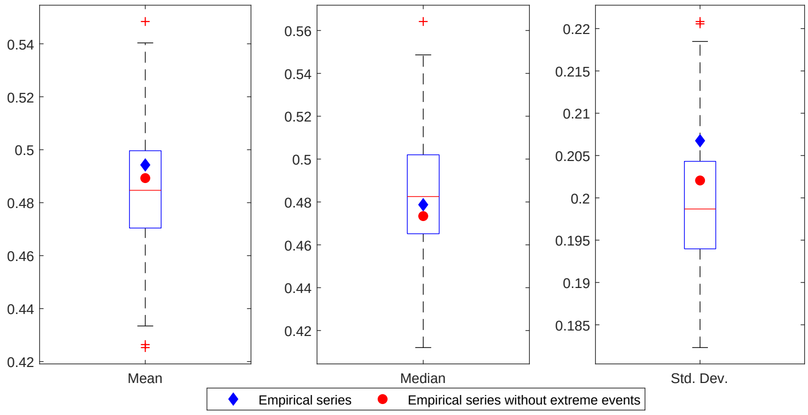


Figure. A.17: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. Nasdaq 100.

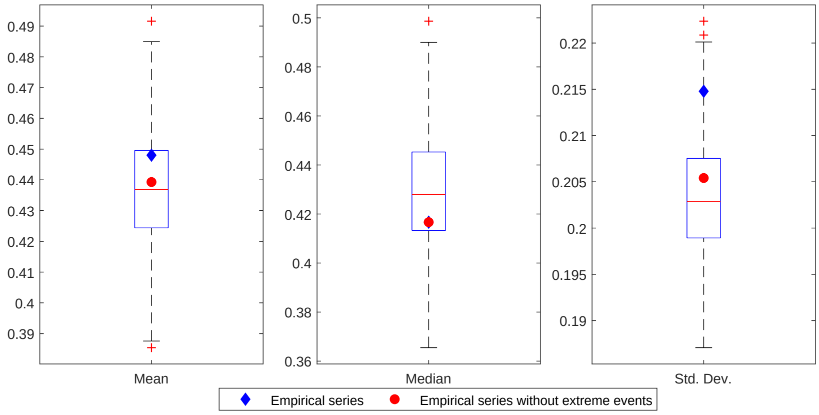


Figure. A.18: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. ASX 200.

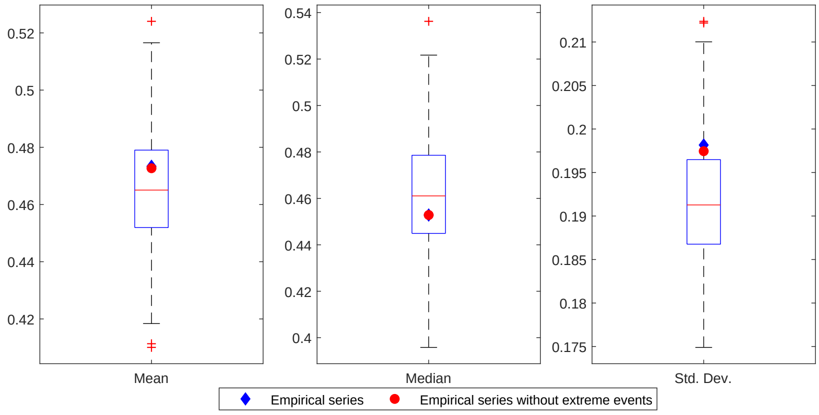


Figure. A.19: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. TSX.

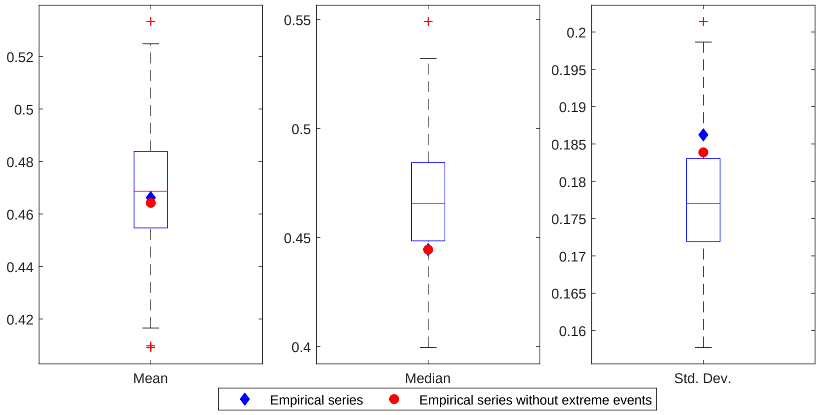


Figure. A.20: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. Euro Stoxx 600.

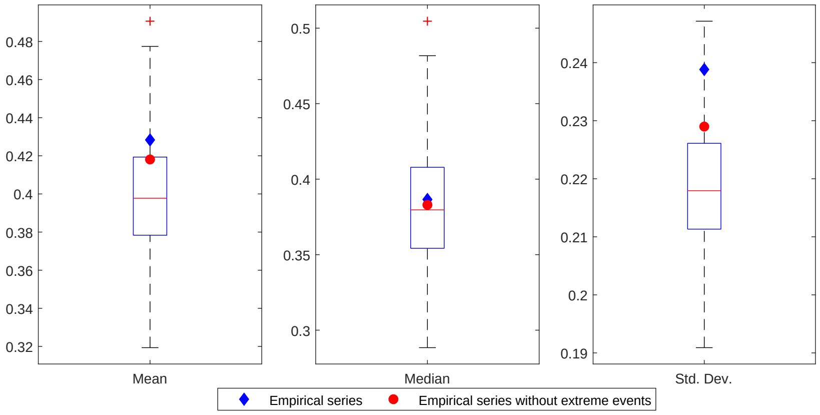


Figure. A.21: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. Nikkei 225.

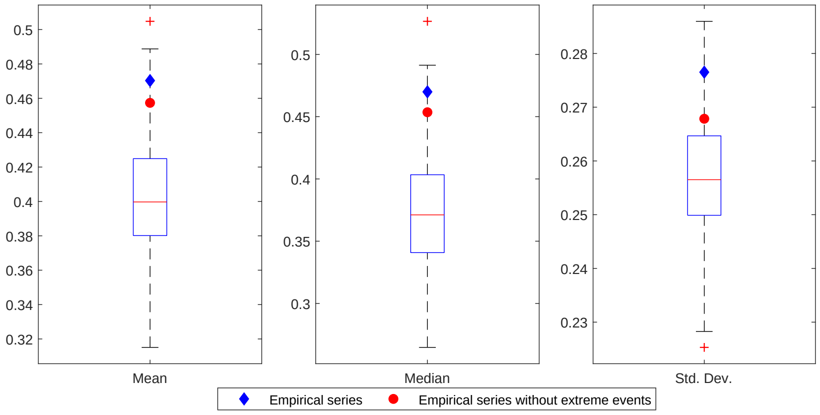


Figure. A.22: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. JSE All-Share.

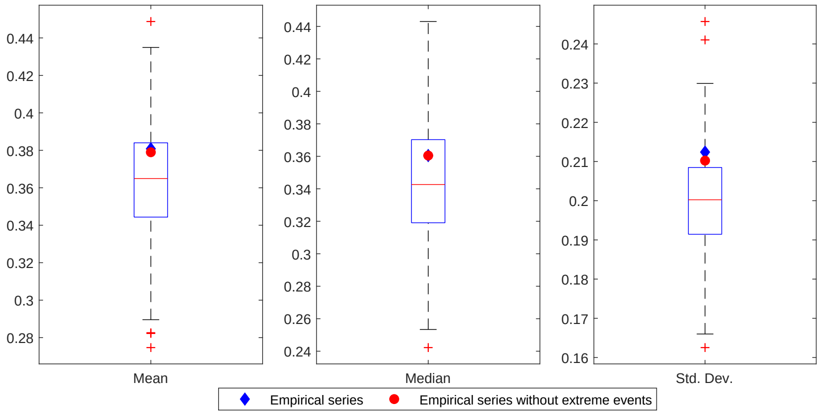
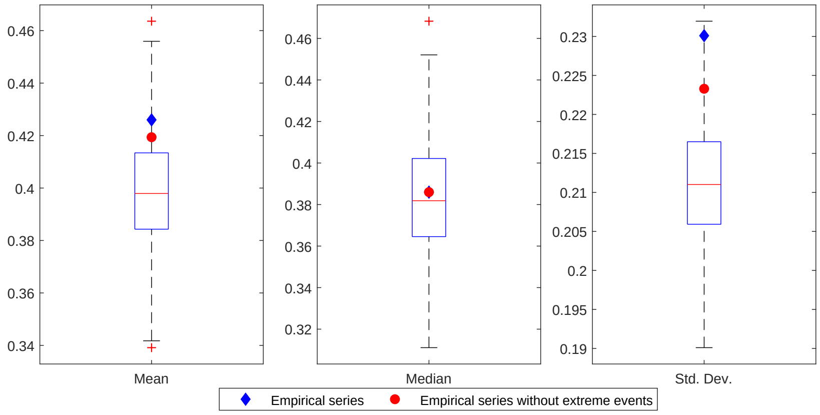


Figure. A.23: Box plots that report the moments of the 500 Monte Carlo simulations. The moments of the empirical time series (blue diamond), and the empirical time series without extreme events ($z_t > 0.95$) (red circle) were added. FTSE 100.

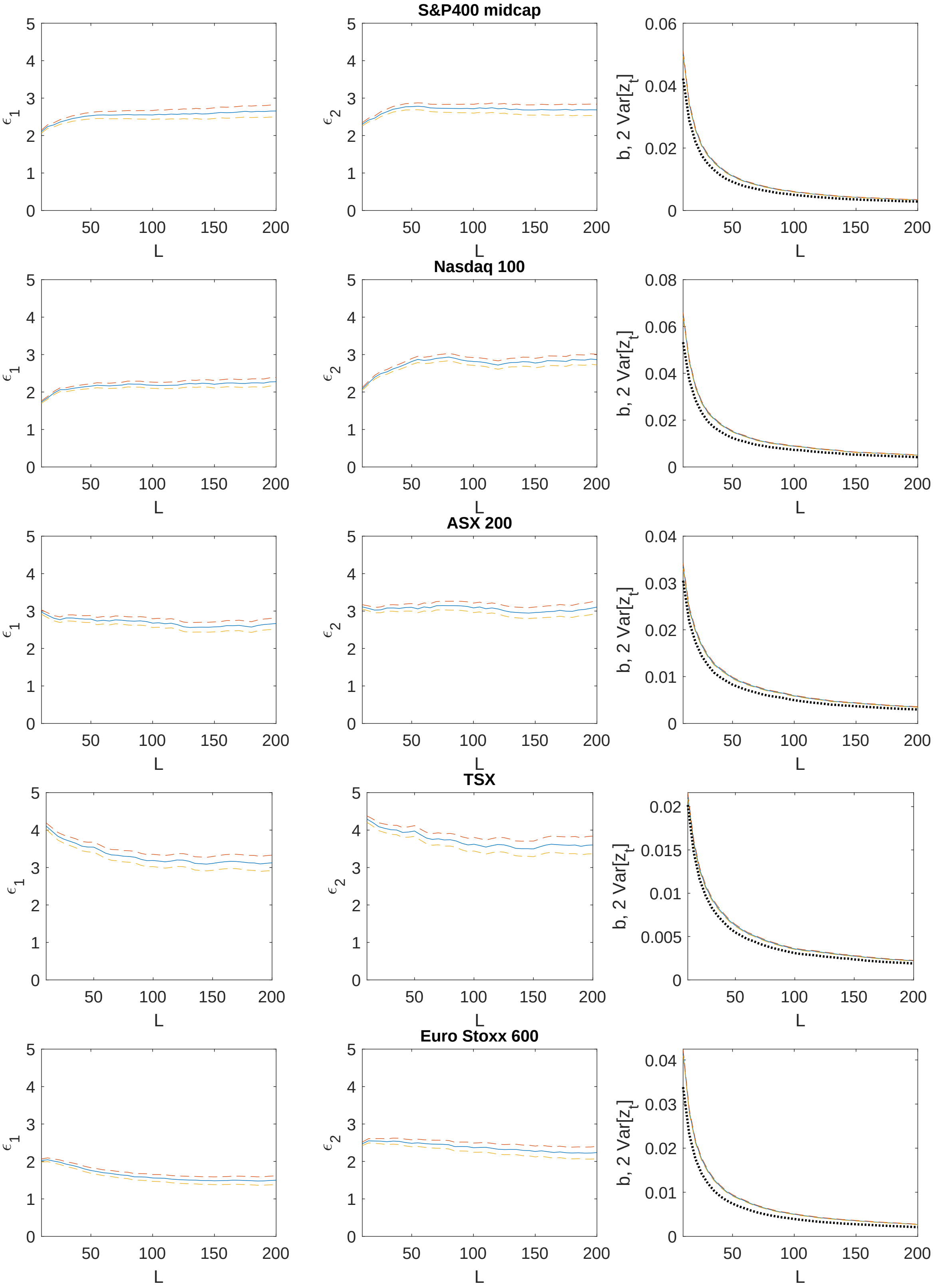


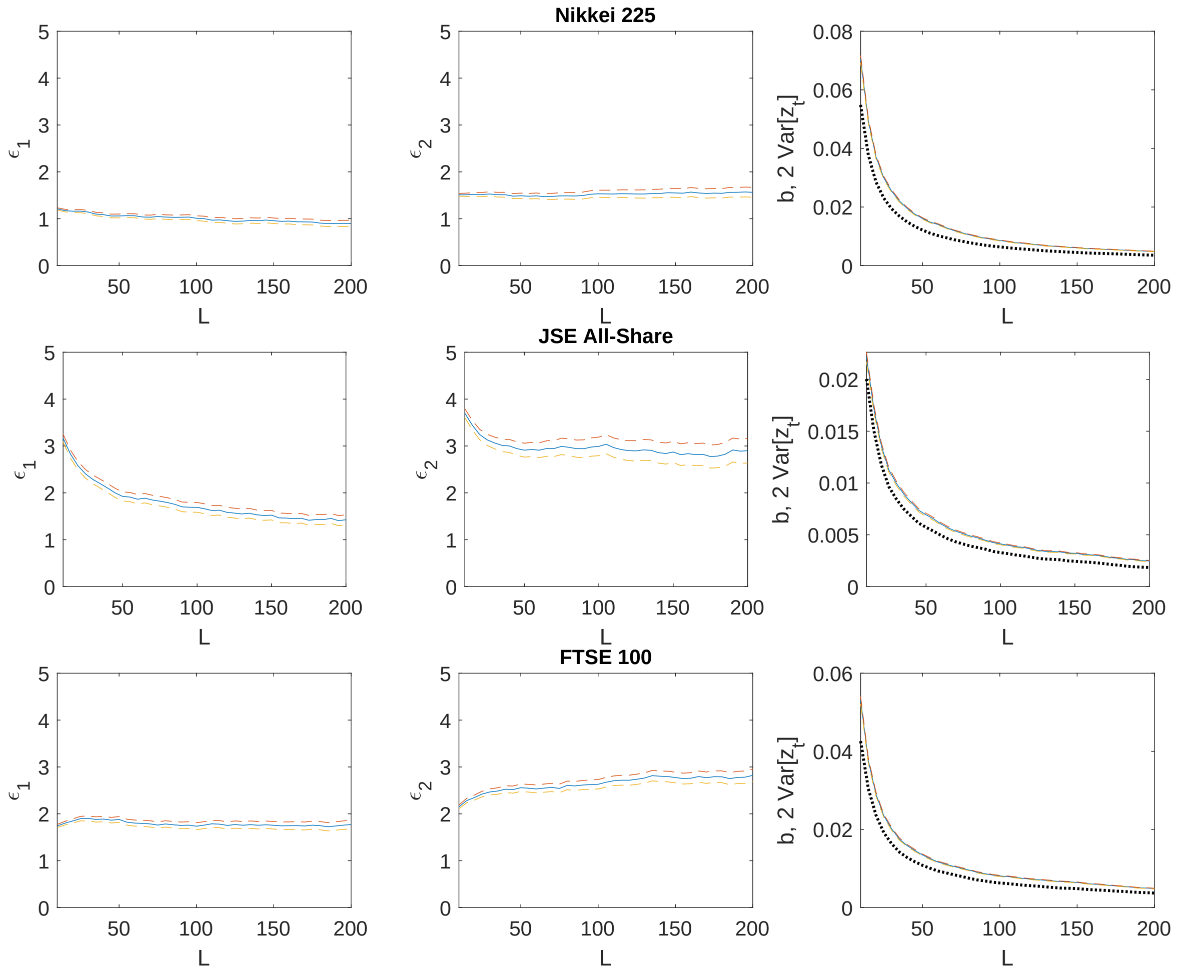
A.5.4 Choice of the parameter L of the EMA

In this section, we analyse for all the international stock markets the sensitivity of the estimation of the main parameters to different values of L.

As reported in the paper, the parameters are estimated by using the conditional likelihood, changing the value of L to compute the corresponding sentiment index. We observe that the evolution of the parameters is line with the S&P 500, i.e. ε_1 and ε_2 do not exhibit large deviations by increasing L, thus the unconditional distribution is robust to the choice of L. Some particular cases could be highlighted like the JSE All-Share or TSX due to their deviations with low values of L, however, the parameters tend to be stable from L=100. Conversely, the herding parameter b also decreases significantly the higher the value of L, regardless of the stock market that we analyse. Therefore, the dynamics of ε_1, ε_2 and b is similar for all the stock markets: ε_1, ε_2 are quite stable while b varies according to the parameter L.

Figure. A.24: Estimated parameters ε_1, ε_2 and b according to different values of L. The scaling between the variation of $E[\Delta z^2]$ as a function of L and the parameter b is reported (dotted black line).

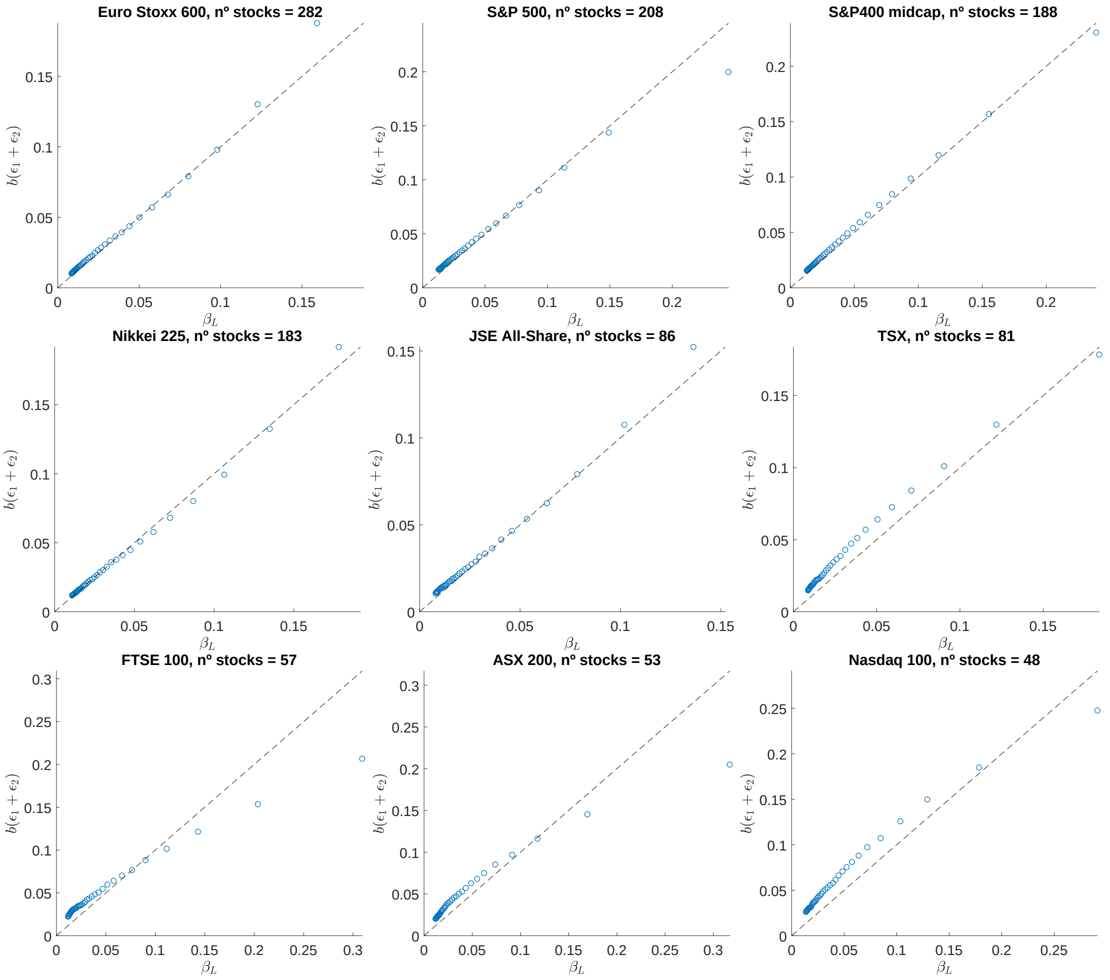




We examine now whether the variations of the herding parameter b affects the decay rate of the autocorrelation function for the worldwide stock markets. In other words, we want to analyse if the parameter L of the EMA is connected with the fit of the correlogram. To do so, we compare the decay rate of the ACF, $b(\varepsilon_1 + \varepsilon_2)$, which is computed using the estimated values of the parameters for a given L, to the slope (β_L) of the correlogram, as it is explained in the Sec. 2.5.2.

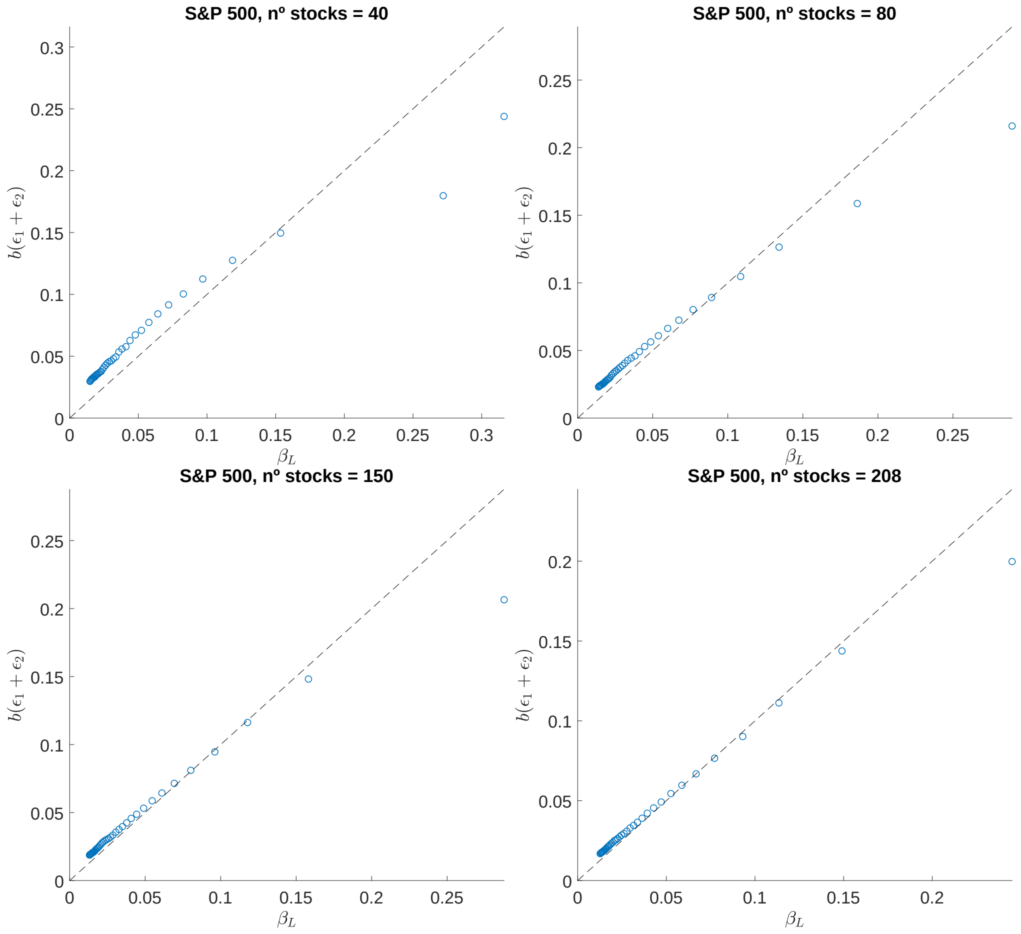
In Fig. (A.25) we show the relation between β_L and $b(\varepsilon_1 + \varepsilon_2)$ for all the stock markets at our disposal, which are ordered according to the number of stocks included in each financial market. As can be observed, for the stock markets with a sufficient number of stocks, β_L and $b(\varepsilon_1 + \varepsilon_2)$ are practically equal regardless of L. This is the case of the Euro Stoxx 600 (282 stocks), S&P 500 (208 stocks), S&P 400 (188 stocks), Nikkei 225 (183 stocks) and JSE All-Share (86 stocks). However, for the rest of the stock markets, we do not observe a proper fit of β_L and $b(\varepsilon_1 + \varepsilon_2)$.

Figure. A.25: Scatter plot of $b(\varepsilon_1 + \varepsilon_2)$ and β_L using different values of L (from 10 to 200 with an increment of 5). Worldwide stock markets.



In order to study this point, we analyse whether it exists a relation between the number of stocks and the difference between β_L and $b(\varepsilon_1 + \varepsilon_2)$. To do so, we examine the S&P 500 by computing the decay rate with different number of stocks. In Fig. (A.26), we show the corresponding results by using 40, 80, 150 and 208 stocks. By analysing the same stock market, but including a different number of stocks in each case, we can observe that the relation of β_L and $b(\varepsilon_1 + \varepsilon_2)$ is affected by the number of stocks used in the computation of each sentiment index z_t. More specifically, in the case of the S&P 500 with 40 stocks, we observe a similar outcome compared to the Nasdaq, FTSE 100, TSX and ASX in Fig. (A.25), however, the fit of the correlogram improves when including more stocks.

Figure. A.26: Scatter plot of $b(\varepsilon_1 + \varepsilon_2)$ and β_L using different values of L (from 10 to 200 with an increment of 5). S&P 500 index including 40, 80, 150 and 208 stocks.



A.5.5 Autocorrelation function

Finally, we analyse the autocorrelation function and the correlogram of each worldwide stock market. The results reported in the following figures support what we have explained previously. Those stock markets with a sufficient number of stocks show a proper fit of the theoretical function with the correlogram. However, the fit of the correlogram and the theoretical ACF is not acceptable when analysing stock markets with few listed stocks.

Figure. A.27: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 4 of the paper. S&P 400 midcap index.

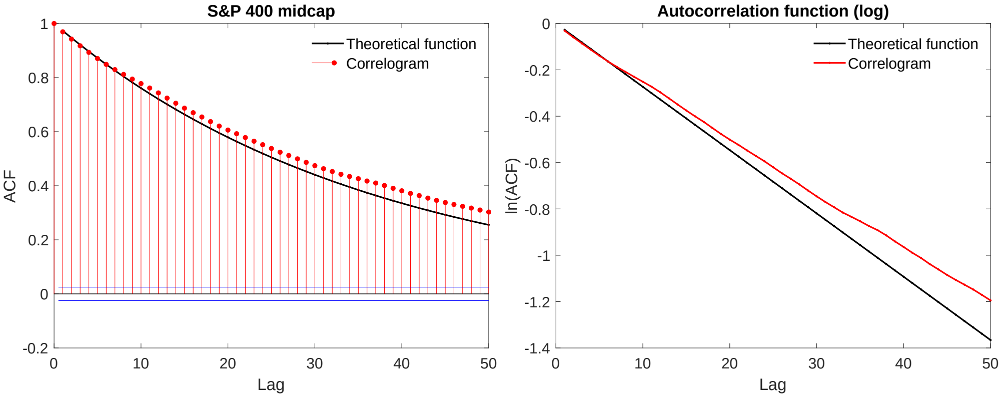


Figure. A.28: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 4 of the paper. Nasdaq 100 index.

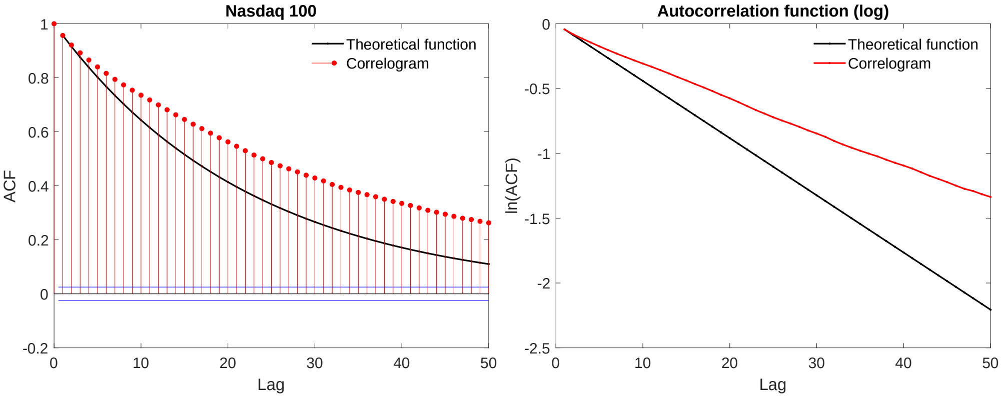


Figure. A.29: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. ASX 200 index.

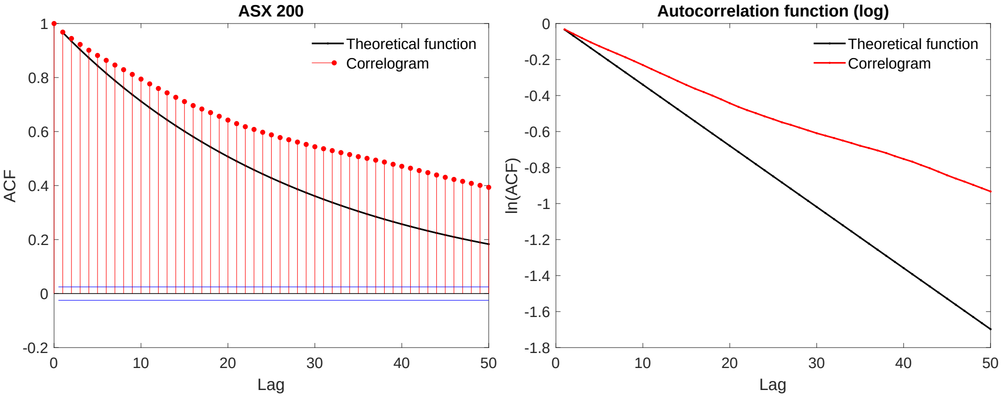


Figure. A.30: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. TSX index.

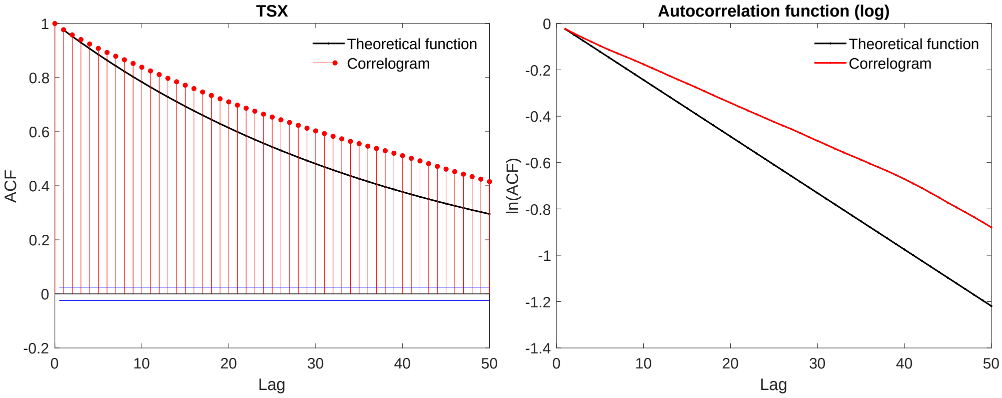


Figure. A.31: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. Nikkei 225 index.

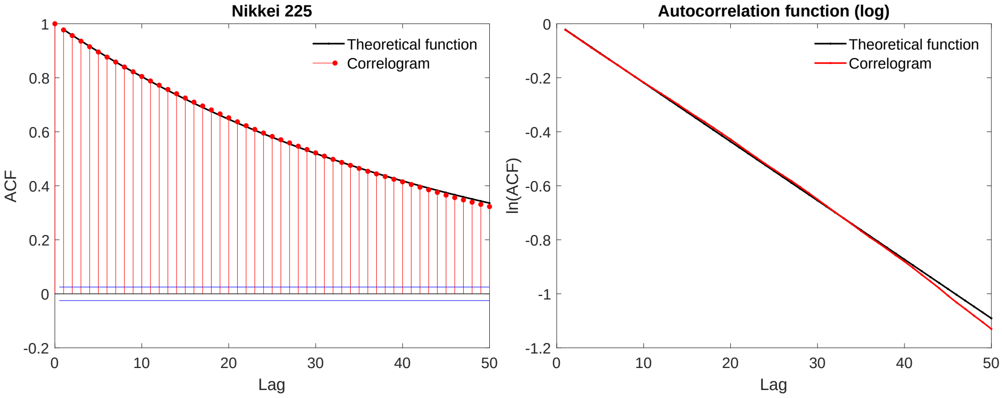


Figure. A.32: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. Euro Stoxx 600 index.

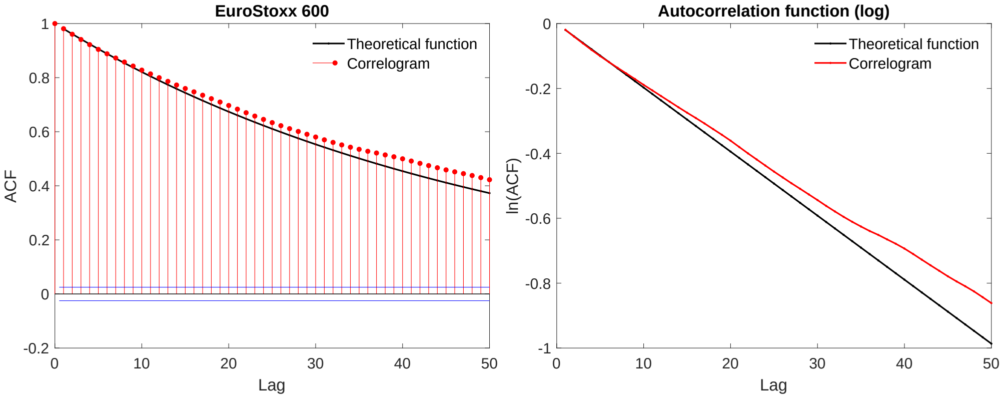


Figure. A.33: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. JSE All-Share index.

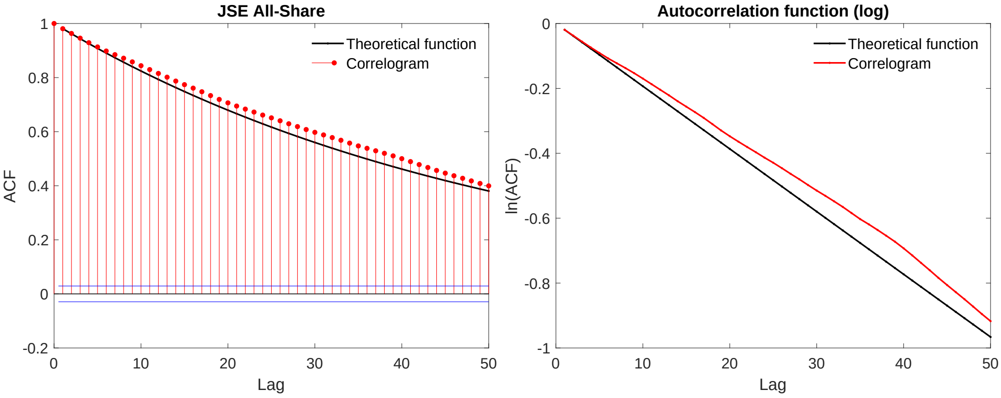
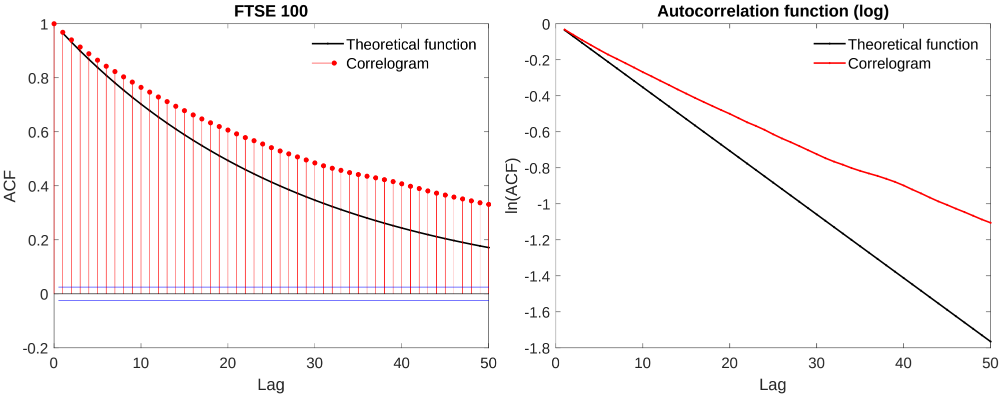


Figure. A.34: Autocorrelation function of the sentiment index compared to the theoretical autocorrelation function. The latter was calculated with the parameters from Table 5 of the paper. FTSE 100 index.

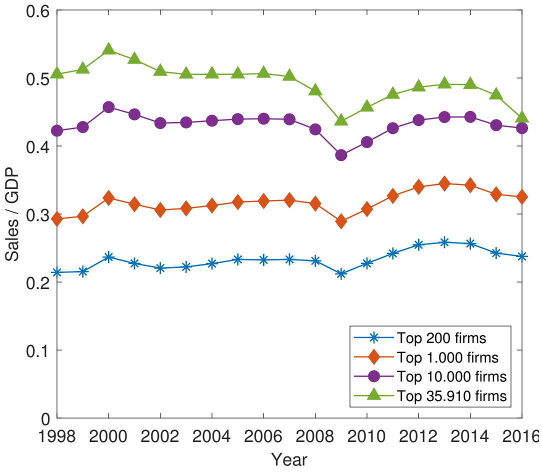


Appendix B

A cross-sectional analysis of growth and profit rate distribution : the Spanish case

B.1 Firms

Figure. B.1: Sales as a function of GDP for the largest long-lived firms in our sample.



B.2 Likelihood ratio test for the 200 largest long-lived firms

Table B.1 and Table B.2 show the LRT in which we test the Laplace distribution compared to the AEP as alternative hypothesis. As can be observed, the null hypothesis of the Laplace distribution is rejected most of the years for profit rates and growth rates of total assets and sales. This result supports the outcome observed by Mundt and Oh (2019) since the AEP seems to characterise better the empirical density of profit rates.

On the other hand, in Table B.3, we report the LRT testing the asymmetric Laplace distribution compared to the AEP distribution. In this case, we observe that the asymmetric Laplace is not rejected in 12 out of 19 years, for profit rates, and in 7 out of 18 years for growth rates of total assets. Compared to Table B.2, the asymmetric Laplace describes better the probability density distribution of profit rates and growth rates of total assets compared to the symmetric Laplace distribution.

Table B.1: P-values of the likelihood ratio test for profit and growth rates of total assets and sales. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the Subbotin distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016.

LRT	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Profit rate	0.17	0.05	**0.04**	0.16	0.06	0.11	0.21	0.06	**0.01**	**0.00**
Total assets	-	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.05	**0.00**	**0.00**	**0.01**
Sales	-	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

LRT	2008	2009	2010	2011	2012	2013	2014	2015	2016
Profit rate	0.27	**0.01**	**0.02**	**0.02**	**0.01**	0.73	0.15	0.08	**0.04**
Total assets	**0.01**	0.17	0.16	**0.04**	**0.00**	**0.00**	**0.02**	0.06	**0.00**
Sales	**0.00**	**0.00**	**0.00**	0.62	0.05	**0.00**	**0.00**	**0.00**	**0.00**

Table B.2: P-values of the likelihood ratio test for profit rates and growth rates of total assets and sales. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the AEP distribution. Results refer to the 200 largest long-lived firms, according to their sales in 2016.

LRT	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Profit rate	**0.00**	**0.00**	**0.00**	**0.03**	**0.04**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**
Total assets	-	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Sales		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

LRT	2008	2009	2010	2011	2012	2013	2014	2015	2016
Profit rate	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Total assets	**0.00**	**0.00**	**0.00**	0.22	0.39	**0.00**	**0.03**	**0.00**	**0.00**
Sales	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

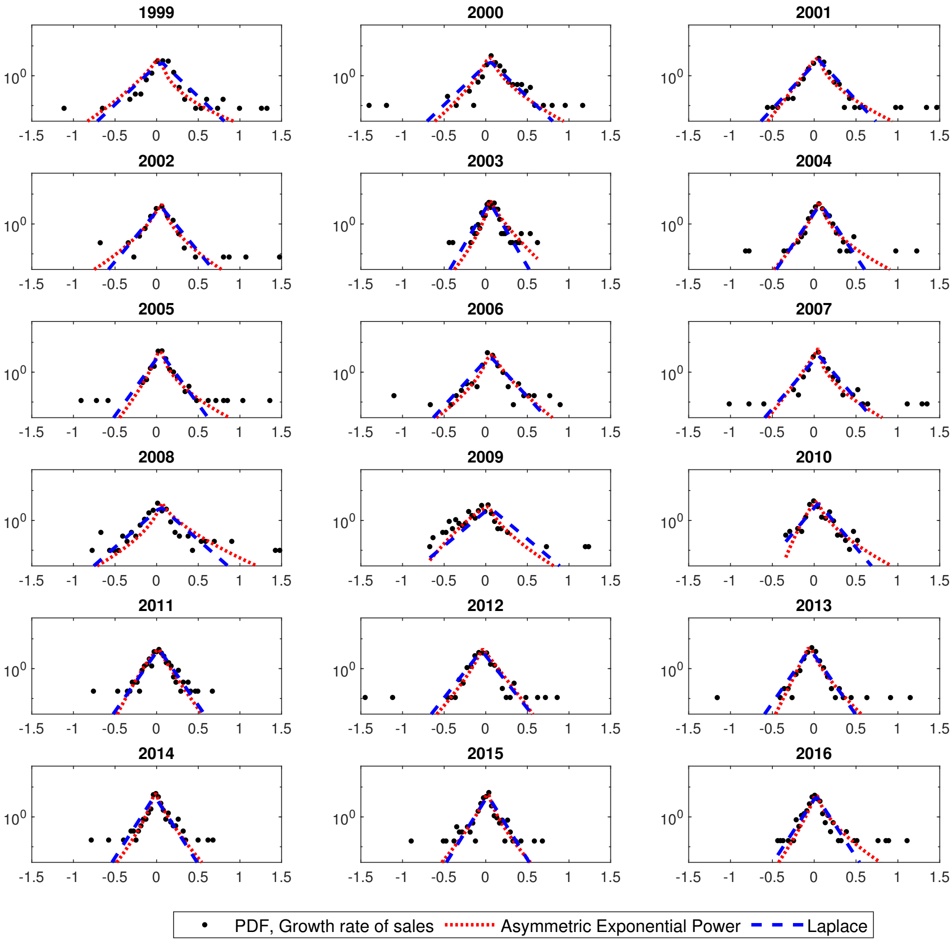
Table B.3: P-values of the likelihood ratio test for profit rates and growth rates of total assets and sales. The null hypothesis is the asymmetric Laplace distribution, while the alternative hypothesis is the AEP distribution. Results refer to the 200 largest long-lived firms, according to their sales in 2016.

LRT	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Profit rates	0.57	**0.01**	0.40	0.39	0.56	0.58	**0.01**	**0.00**	0.39	0.15
Total assets		**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.02**	0.14	**0.00**	0.10
Sales		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

LRT	2008	2009	2010	2011	2012	2013	2014	2015	2016
Profit rates	**0.01**	**0.01**	**0.00**	0.29	0.58	0.07	0.81	0.43	**0.01**
Total assets	0.18	**0.01**	**0.00**	0.28	0.41	**0.03**	0.29	0.29	**0.00**
Sales	**0.00**	**0.00**	**0.00**	0.08	**0.03**	**0.00**	**0.01**	**0.00**	**0.00**

B.3 Probability density function of growth rates of sales

Figure. B.2: Probability density function (PDF) of growth rates of sales along with the AEP (dotted line) and Laplace (dashed line) distribution. The results refer to the 200 largest long-lived firms according to their sales in 2016.



B.4 Likelihood ratio test for all the samples

Table B.4: P-values of the likelihood ratio test for profit and growth rates of total assets and sales conditional on size. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the Subbotin distribution.

Profit rates	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Total assets	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Sales	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Table B.5: P-values of the likelihood ratio test for profit and growth rates of total assets and sales conditional on size. The null hypothesis is the Laplace distribution, while the alternative hypothesis is the AEP distribution.

Profit rates	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Total Assets	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.16	**0**	**0**	**0**	**0**	**0**	**0**
10000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Sales	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Table B.6: P-values of the likelihood ratio test for profit rates and growth rates of total assets and sales. The null hypothesis is the asymmetric Laplace distribution, while the alternative hypothesis is the AEP distribution. Results refer to the 200 largest long-lived firms, according to their sales in 2016.

Profit rates	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.14	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Total assets	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Sales	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10000		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Entire sample		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Bibliography

Adrian, T. and Shin, H. S. (2009). Money, liquidity, and monetary policy. *American Economic Review*, 99(2):600–605.

Aggarwal, R., Klapper, L., and Wysocki, P. D. (2005). Portfolio preferences of foreign institutional investors. *Journal of Banking & Finance*, 29(12):2919–2946.

Alfarano, S. (2006). *An agent-based stochastic volatility model*. PhD thesis, Christian-Albrechts Universität Kiel.

Alfarano, S., Lux, T., and Wagner, F. (2005). Estimation of agent-based models: the case of an asymmetric herding model. *Computational Economics*, 26(1):19–49.

Alfarano, S., Lux, T., and Wagner, F. (2006). Estimation of a simple agent-based model of financial markets: An application to australian stock and foreign exchange data. *Physica A: Statistical Mechanics and its Applications*, 370(1):38 – 42.

Alfarano, S., Lux, T., and Wagner, F. (2008). Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach. *Journal of Economic Dynamics and Control*, 32(1):101–136.

Alfarano, S., Milaković, M., and Raddant, M. (2013). A note on institutional hierarchy and volatility in financial markets. *The European Journal of Finance*, 19(6):449–465.

Alfarano, S. and Milakovic, M. (2008). Does classical competition explain the statistical features of firm growth? *Economics Letters*, 101(3):272–274.

Alfarano, S. and Milaković, M. (2009). Network structure and n-dependence in agent-based herding models. *Journal of Economic Dynamics and Control*, 33(1):78–92.

Alfarano, S., Milakovic, M., Albrecht, I., and Kauschke, J. (2012). A statistical equilibrium model of competitive firms. *Journal of Economics Dynamics and Control*, 36(1):136–149.

Amaral, N., L.A. Buldyrev, S., Havlin, S., Maass, P., Salinger, M., Stanley, H., and Stanley, M. (1997). Scaling behavior in economics: the problem of quantifying company growth. *Physica A*, 244(1):1–24.

Anderson, C. W., Fedenia, M., Hirschey, M., and Skiba, H. (2011). Cultural influences on home bias and international diversification by institutional investors. *Journal of Banking & Finance*, 35(4):916–934.

Aoki, M. and Yoshikawa, H. (2002). Demand saturation creation and economic growth. *Journal of Economic Behavior & Organization*, 48(2):127–154.

Aoki, M. and Yoshikawa, H. (2011). *Reconstructing macroeconomics: a perspective from statistical physics and combinatorial stochastic processes*. Cambridge University Press.

Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F., and Zopounidis, C. (2019). Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research*, 276(2):770–780.

Axtell, R. L. (2001). Zipf distribution of us firm sizes. *Science*, 293(5536):1818–1820.

Bachmann, R. and Bayer, C. (2014). Investment dispersion and the business cycle. *The American Economic Review*, 104(4):1392–1416.

Baek, C. and Elbeck, M. (2015). Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters*, 22(1):30–34.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Banerjee, A. V. (1993). The economics of rumours. *The Review of Economic Studies*, 60(2):309–327.

Bargigli, L. and Tedeschi, G. (2013). Major trends in agent-based economics.

Bendini, R. (2015). Exceptional measures: The shanghai stock market crash and the future of the chinese economy. Technical report, Policy Department, Directorate General for External Policies, European Parliament'.

Biddle, G. C., Hilary, G., and Verdi, R. S. (2009). How does financial reporting quality relate to investment efficiency? *Journal of accounting and economics*, 48(2):112–131.

Black, F. (1986). Noise. *The journal of finance*, 41(3):528–543.

Blanco-Arroyo, O., Ruiz-Buforn, A., Vidal-Tomás, D., and Alfarano, S. (2018). On the determination of the granular size of the economy. *Economics Letters*, 173:35–38.

Blanco-Arroyo, O., Ruiz-Buforn, A., Vidal-Tomás, D., and Alfarano, S. (2019). Empresas granulares y desagregación regional: un análisis del caso español.

Boswijk, H. P., Hommes, C. H., and Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31(6):1938 – 1970.

Bottazzi, G. (2004). Subbotools user's manual: For version 0.9. 7.1, 8 september 2004. Technical report, LEM Working Paper Series.

Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F., and Riccaboni, M. (2001). Innovation and corporate growth in the evolution of the drug industry. *International Journal of Industrial Organization*, 19(7):1161–1187.

Bottazzi, G., Li, L., and Secchi, A. (2019). Aggregate fluctuations and the distribution of firm growth rates. *Industrial and Corporate Change*, 28(3):635–656.

Bottazzi, G. and Secchi, A. (2003). Why are distributions of firm growth rates tent-shaped? *Economics Letters*, 80(3):415–420.

Bottazzi, G. and Secchi, A. (2006). Explaining the distribution of firm growth rates. *The RAND Journal of Economics*, 37(2):235–256.

Bottazzi, G. and Secchi, A. (2011). A new class of asymmetric exponential power densities with applications to economics and finance. *Industrial and Corporate Change*, 20(4):991–1030.

Bottazzi, G., Secchi, A., and Tamagni, F. (2014). Financial constraints and firm dynamics. *Small Business Economics*, 42(1):99–116.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica: Journal of the Econometric Society*, pages 1059–1095.

Brock, W. A. and Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22(8-9):1235–1274.

Buchs, T. D. (1999). Financial crisis in the russian federation: Are the russians learning to tango? *Economics of transition*, 7(3):687–715.

Buldyrev, S. V., Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2007). The growth of business firms: facts and theory. *The Journal of the European Economic Association*, 5(2-3):574–584.

Catania, L., Grassi, S., and Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting*, 35(2):485–501.

Cheah, E.-T. and Fry, J. (2015). Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. *Economics Letters*, 130:32–36.

Chen, Z. and Lux, T. (2015). Estimation of sentiment effects in financial markets: A simulated method of moments approach. *Computational Economics*, pages 1–34.

Cheung, A., Roca, E., and Su, J.-J. (2015). Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on mt. gox bitcoin prices. *Applied Economics*, 47(23):2348–2358.

Chiarella, C., Iori, G., and Perelló, J. (2009). The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, 33(3):525 – 537.

Coad, A. and Planck, M. (2012). Firms as bundles of discrete resources–towards an explanation of the exponential distribution of firm growth rates. *Eastern Economic Journal*, 38(2):189–209.

Coad, A., Segarra, A., and Teruel, M. (2013). Like milk or wine: Does firm performance improve with age? *Structural Change and Economic Dynamics*, 24:173–189.

Cocco, L., Concas, G., and Marchesi, M. (2017). Using an artificial financial market for studying a cryptocurrency market. *Journal of Economic Interaction and Coordination*, 12(2):345–365.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Cont, R. and Bouchaud, J.-P. (2000). Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic Dynamics*, 4(2):170–196.

Cooper, I. and Kaplanis, E. (1994). Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. *The Review of Financial Studies*, 7(1):45–60.

Corbet, S., Lucey, B., and Yarovaya, L. (2018). Datestamping the bitcoin and ethereum bubbles. *Finance Research Letters*, 26:81–88.

Covrig, V., Lau, S. T., and Ng, L. (2006). Do domestic and foreign fund managers have similar preferences for stock characteristics? a cross-country analysis. *Journal of International Business Studies*, 37(3):407–429.

da Gama Silva, P. V. J., Klotzle, M. C., Pinto, A. C. F., and Gomes, L. L. (2019). Herding behavior and contagion in the cryptocurrency market. *Journal of Behavioral and Experimental Finance*, 22:41–50.

De Grauwe, P. (2010). Top-down versus bottom-up macroeconomics. *CESifo Economic Studies*, 56(4):465–497.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.

De Veirman, E. and Levin, A. (2011). Cyclical changes in firm volatility. Technical report, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., and Palestrini, A. (2005). A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic behavior & organization*, 56(4):489–512.

Demyanyk, Y. and Van Hemert, O. (2009). Understanding the subprime mortgage crisis. *The Review of Financial Studies*, 24(6):1848–1880.

Dodd, N. (2018). The social life of bitcoin. *Theory, Culture & Society*, 35(3):35–56.

Dosi, G. and Nelson, R. R. (2010). Technical change and industrial dynamics as evolutionary processes. *Handbook of the Economics of Innovation*, 1:51–127.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Erlingsson, E. J., Alfarano, S., Raberto, M., and Stefánsson, H. (2013). On the distributional properties of size, profit and growth of icelandic firms. *Journal of Economic Interaction and Coordination*, 8(1):57–74.

Ewens, W. (2004). *Mathematical Population Genetics*. Springer.

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, 38(1):34–105.

Fama, E. F. (1991). Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617.

Feller, W. (1968). *An introduction to probability theory and its applications*, volume 1. Wiley, New York.

Ferreira, M. A. and Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics*, 88(3):499–533.

Feuer, A. (2013). The bitcoin ideology. *New York Times*. 14 December.

Fidrmuc, J. and Korhonen, I. (2006). Meta-analysis of the business cycle correlation between the euro area and the ceecs. *Journal of Comparative Economics*, 34(3):518–537.

Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: measuring stock market comovements. *The journal of Finance*, 57(5):2223–2261.

Forman, J. L. and Sørensen, M. (2008). The pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35(3):438–465.

Franke, R. and Westerhoff, F. (2011). Estimation of a structural stochastic volatility model of asset pricing. *Computational Economics*, 38(1):53–83.

French, K. R. and Poterba, J. M. (1991). Investor diversification and international equity markets. Technical report, National Bureau of Economic Research.

Friedman, M. (1953). *Essays in positive economics*. University of Chicago Press.

Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K., and Stanley, H. E. (2005). The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52):18801–18806.

Gabaix, X. (2011). The granular origins of aggregate fluctuation. *Econometrica*, 79(3):733–772.

Gaffeo, E., Gallegati, M., and Palestrini, A. (2003). On the size distribution of firms. additional evidence from the g7 countries. *Physica A*, 324:117–123.

Gandal, N., Hamrick, J., Moore, T., and Oberman, T. (2018). Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics*, 95:86–96.

Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 3. springer Berlin.

Garibaldi, U. and Scalas, E. (2010). *Finitary probabilistic methods in econophysics*. Cambridge University Press.

Gehrig, T. (1993). An information based explanation of the domestic bias in international equity investment. *The Scandinavian Journal of Economics*, pages 97–109.

Geroski, P. A. and Gregg, P. (1997). *Coping with recession: UK company performance in adversity*, volume 38. Cambridge University Press.

Gibrat, R. (1931). *Les inégalités économiques*. Recueil Sirey.

Gilli, M. and Winker, P. (2003). A global optimization heuristic for estimating agent based models. *Computational Statistics & Data Analysis*, 42(3):299 – 312. Computational Ecomometrics.

Gouriéroux, C. and Valéry, P. (2004). Estimation of a jacobi process. `http://www.scse.ca/scse/congres2004/articles/Valery_Gourieroux.pdf`.

Grilli, R., Tedeschi, G., and Gallegati, M. (2020). Business fluctuations in a behavioral switching model: gridlock effects and credit crunch phenomena in financial networks. *Journal of Economic Dynamics and Control*, 114:103863.

Grinblatt, M. and Keloharju, M. (2001). How distance, language, and culture influence stockholdings and trades. *The Journal of Finance*, 56(3):1053–1073.

Haltiwanger, J. C. (1997). Measuring and analyzing aggregate fluctuations: the importance of building from microeconomic evidence. *Review*, 79.

Han, J.-B., Kim, S.-H., Jang, M.-H., and Ri, K.-S. (2019). Using genetic algorithm and narx neural network to forecast daily bitcoin price. *Computational Economics*, pages 1–17.

Hansun, S. (2013). A new approach of moving average method in time series analysis. In *2013 Conference on New Media Studies (CoNMedia)*, pages 1–4. IEEE.

Hartnett, M., Leung, B., and Roche, G. (2015). Rules & tools: Three buy signals and a funeral. Technical report, Bank of America Merrill Lynch.

Higson, C., Holly, S., and Kattuman, P. (2002). The cross-sectional dynamics of the us business cycle: 1950–1999. *Journal of Economic Dynamics and Control*, 26(9):1539–1555.

Higson, C., Holly, S., Kattuman, P., and Platis, S. (2004). The business cycle, macroeconomic shocks and the cross-section: The growth of uk quoted companies. *Economica*, 71(282):299–318.

Holly, S., Petrella, I., and Santoro, E. (2013). Aggregate fluctuations and the cross-sectional dynamics of firm growth. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2):459–479.

Hommes, C. H. (2006). Heterogeneous agent models in economics and finance. *Handbook of Computational Economics*, 2:1109–1186.

IMF (2006). Global markets analysis division: Financial market update. Technical report, International Monetary Fund.

Ivković, Z. and Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, 60(1):267–306.

Jang, H. and Lee, J. (2018). An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, 6:5427–5437.

Jaynes, E. T. (1979). Where do we stand on maximum entropy? *The maximum entropy formalism*, 15.

Kaizoji, T. (2006). A precursor of market crashes: Empirical laws of japan's internet bubble. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1-2):123–127.

Karlsson, A. and Nordén, L. (2007). Home sweet home: Home bias and international diversification among individual investors. *Journal of Banking & Finance*, 31(2):317–333.

Kenett, D. Y., Raddant, M., Lux, T., and Ben-Jacob, E. (2012). Evolvement of uniformity and volatility in the stressed global financial village. *PloS one*, 7(2):e31144.

Kenett, D. Y., Shapira, Y., Madi, A., Bransburg-Zabary, S., Gur-Gershgoren, G., and Ben-Jacob, E. (2011). Index cohesive force analysis reveals that the us market became prone to systemic collapses since 2002. *PLoS one*, 6(4):e19378.

Kirman, A. (1991). Epidemics of opinion and speculative bubbles in financial markets. *Money and financial markets*, 3:54–368.

Kirman, A. (1993). Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, pages 137–156.

Kloeden, P. E. and Platen, E. (2013). *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media.

Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, 3:3415.

Krugman, P. (2018). Bubble, bubble, fraud and trouble. *New York Times*. 29 January.

Kukacka, J. and Barunik, J. (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control*, 85:21 – 45.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Larsen, K. S. and Sorensen, M. (2007). Diffusion models for exchange rates in a target zone. *Mathematical Finance*, 17(2):285–306.

LeBaron, B. and Yamamoto, R. (2008). The impact of imitation on long memory in an order-driven markets. *Eastern Economic Journal*, 34:504–517.

Licalzi, M. and Pellizzari, P. (2003). Fundamentalists clashing over the book: a study of order-driven stock markets. *Quantitative Finance*, 3(6):470–480.

Lux, T. (1995). Herd behaviour, bubbles and crashes. *The Economic Journal*, pages 881–896.

Lux, T. (1996). The stable paretian hypothesis and the frequency of large returns: an examination of major german stocks. *Applied Financial Economics*, 6(6):463–475.

Lux, T. (1998). The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. *Journal of Economic Behavior & Organization*, 33(2):143–165.

Lux, T. (2011). Sentiment dynamics and stock returns: the case of the german stock market. *Empirical economics*, 41(3):663–679.

Lux, T. and Alfarano, S. (2016). Financial power laws: Empirical evidence, models, and mechanisms. *Chaos, Solitons & Fractals*, 88:3–18.

Lux, T. and Marchesi, M. (1999a). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500.

Lux, T. and Marchesi, M. (1999b). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500.

Lux, T. and Marchesi, M. (2000). Volatility clustering in financial markets: a microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3(4):675–702.

Malkiel, B. G. and Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417.

Mantegna, R. N. and Stanley, H. E. (1996). An introduction to econophysics: correlations and complexity in finance.

Massa, M. and Simonov, A. (2006). Hedging, familiarity and portfolio choice. *The Review of Financial Studies*, 19(2):633–685.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of time series analysis*, 4(4):269–273.

McNally, S., Roche, J., and Caton, S. (2018). Predicting the price of bitcoin using machine learning. In *Parallel, Distributed and Network-based Processing (PDP), 2018 26th Euromicro International Conference on*, pages 339–343. IEEE.

Mizuno, T., Ohnishi, T., and Watanabe, T. (2016). Power laws in market capitalization during the dot-com and shanghai bubble periods. *Evolutionary and Institutional Economics Review*, 13(2):445–454.

Mundt, P., Alfarano, S., and Milaković, M. (2016). Gibrat's law redux: think profitability instead of growth. *Industrial and Corporate Change*, 25(4):549–571.

Mundt, P., Förster, N., Alfarano, S., and Milakovic, M. (2014). The real versus the financial economy: A global tale of stability versus volatility. *Economics: The Open-Access, Open-Assessment E-Journal*, 8(2014-17):1–26.

Mundt, P. and Oh, I. (2019). Asymmetric competition, risk, and return distribution. *Economics Letters*, 179:29–32.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Nakamoto, S. (2009). Bitcoin open source implementation of p2p currency. https://p2pfoundation.ning.com/forum/topics/bitcoin-open-source.

Orléan, A. (1995). Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *Journal of Economic Behavior & Organization*, 28(2):257–274.

Pedersen, A. R. (1994). Uniform residuals for discretely observed diffusion processes. *Research Report*, 292.

Penrose, E. and Penrose, E. T. (2009). *The Theory of the Growth of the Firm*. Oxford university press.

Phillips, P. C., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4):1043–1078.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52(1):201–226.

Podobnik, B., Horvatic, D., Petersen, A. M., and Stanley, H. E. (2009). Cross-correlations between volume change and price change. *Proceedings of the National Academy of Sciences*, 106(52):22079–22084.

Preis, T., Schneider, J. J., and Stanley, H. E. (2011). Switching processes in financial markets. *Proceedings of the National Academy of Sciences*.

Raberto, M., Cincotti, S., Focardi, S. M., and Marchesi, M. (2001). Agent-based simulation of a financial market. *Physica A: Statistical Mechanics and its Applications*, 299(1-2):319–327.

Recchioni, M. C., Tedeschi, G., and Gallegati, M. (2015). A calibration procedure for analyzing stock price dynamics in an agent-based framework. *Journal of Economic Dynamics and Control*, 60:1–25.

Riccaboni, M., Growiec, J., and Pammolli, F. (2011). Innovation and corporate dynamics: a theoretical framework. University Library of Munich.

Seasholes, M. S. and Zhu, N. (2010). Individual investors and local bias. *The Journal of Finance*, 65(5):1987–2010.

Shiller, R. J. (2015). *Irrational exuberance*. Princeton university press.

Shleifer, A. (2000). *Clarendon Lectures: Inefficient Markets*. Oxford University Press. Japanese translation, Toyo Keisai, Tokyo, 2001. Chinese translation, 2003.

Smith, A. (1776). *An Inquiry into the Nature and Causes of The Wealth of Nations*. Petersfield: Harriman House.

Sornette, D., Demos, G., Zhang, Q., Cauwels, P., Filimonov, V., and Zhang, Q. (2015). Real-time prediction and post-mortem analysis of the shanghai 2015 stock market bubble and crash. *Swiss Finance Institute Research Paper*, 15-31.

Stanley, M., Amaral, L., Buldyrev, S., Havlin, S., Leschhorn, H., P., M., Salinger, M., and Stanley, H. (1996). Scaling behaviour in the growth of companies. *Nature*, 379:804–806.

Subbotin, M. (1923). On the law of frequency of errors. *Matematicheskii Sbornik*, 31:296–301.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, 35:40–59.

Tedeschi, G., Iori, G., and Gallegati, M. (2009). The role of communication and imitation in limit order markets. *The European Physical Journal B*, 71(4):489.

Tedeschi, G., Iori, G., and Gallegati, M. (2012). Herding effects in order driven markets: The rise and fall of gurus. *Journal of Economic Behavior & Organization*, 81(1):82–96.

Tedeschi, G., Recchioni, M. C., and Berardi, S. (2019). An approach to identifying micro behavior: How banks' strategies influence financial cycles. *Journal of Economic Behavior & Organization*, 162:329–346.

Teglio, A., Raberto, M., and Cincotti, S. (2012). The impact of banks' capital adequacy regulation on the economic system: an agent-based approach. *Advances in Complex Systems*, 15(supp02):1250040.

Tesar, L. L. and Werner, I. M. (1995). Home bias and high turnover. *Journal of international Money and Finance*, 14(4):467–492.

Tesfatsion, L. and Judd, K. L. (2006). *Handbook of computational economics: agent-based computational economics*, volume 2. Elsevier.

Trucíos, C. (2019). Forecasting bitcoin risk measures: A robust approach. *International Journal of Forecasting*, 35(3):836–847.

Urquhart, A. (2018). What causes the attention of bitcoin? *Economics Letters*, 166:40–44.

Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*, volume 1. Elsevier.

Vidal-Tomás, D. and Ibañez, A. (2018). Semi-strong efficiency of bitcoin. *Finance Research Letters*, 27:259–265.

Vidal-Tomás, D., Ibáñez, A. M., and Farinós, J. E. (2019). Herding in the cryptocurrency market: Cssd and csad approaches. *Finance Research Letters*, 30:181–186.

Wagner, F. (2003). Volatility cluster and herding. *Physica A: Statistical Mechanics and its Applications*, 322:607–619.

Waldrop, M. M. (1987). Computers amplify black monday. *Science*, Vol. 238(4827):602–604.

Wei, W. C. (2018). Liquidity and market efficiency in cryptocurrencies. *Economics Letters*, 168:21–24.