DISS. ETH NO. 24574

Information Geometry and the Dynamic Detection of Financial Bubbles and Crashes

A thesis submitted to attain the degree of

Doctor of Sciences of ETH Zürich

(Dr. sc. ETH Zürich)

presented by

Guilherme Do Livramento Demos

M.Sc. in Economics, Federal University of Santa Catarina (UFSC)

born on March 10^{th}, 1989

Citizen of Brazil

Citizen of Italy

accepted on the recommendation of

Prof. Dr. Didier Sornette, examiner

Prof. Dr. Thorsten Hens, co-examiner

Zürich, Switzerland

September 2017

Thesis advisor	Author
Didier Sornette	**Guilherme Do Livramento Demos**

Information Geometry and the Dynamic Detection of Financial Bubbles and Crashes

Abstract

This thesis is a collection of six self-contained research papers of which three are already published in peer-reviewed journals, two are in the process of being published and one working paper. The major focus of the present thesis consists in creating novel statistical metrics for the ongoing diagnostic of financial bubbles across multiple scales and to reduce the uncertainty of inferences about a future crash date using sophisticated Likelihood estimators and Information Geometry.

Chapter 1 is an introductory chapter and provides a literature review about financial bubbles. It also defines the Log-Periodic Power-Law Singularity (LPPLS) model [56] and enumerates some of its' bottlenecks whose the present thesis propose a solution.

In Chapter 2 we ask the following question: it is easier to spot the birth or the burst of financial bubbles? We argue that it is the lack of knowledge regarding when bubbles will burst that allows them to grow even in the presence of rational agents. Using the LPPLS model, we provide a methodology for quantifying the uncertainty of each one of these parameters ($t_1 :=$ birth of the bubble and $t_c :=$ burst) based on the eigenvalues and eigenvectors of the Fisher Information matrix approximated at the best-fit parameters.

Chapter 3 provies a novel statistical metric which allows one to compare goodness of fit of a model using unbalanced sample sizes. When applied to the LPPLS model cost function, the methodology allows one to automatise the process of diagnosing the beginning of financial bubbles without requiring further exogenous information.

Given the occurrence of bubbles across multiple time-scales, Chapter 4 provides a novel multi-scale bubble-diagnostic technology which allows one to automatically perform an ongoing assessment of bubbles maturing across the short, medium and long-term scales.

Chapter 5 focuses on diminishing the uncertainty surrounding the prediction of financial bubbles' burst. Based on eigenvectors and eigenvalues of the Fisher Information Matrix, we sequentially calibrate the LPPLS model parameters using the Modified Profile Likelihood estimator. Using the sloppiest parameter t_c as the focal parameter, the proposed methodology drastically reduces estimation uncertainty of $\hat{t}_c$, specially on small samples and allows one to construct robust confidence intervals around parameter estimates.

Chapter 6 questions if the findings of the previous Chapter can be extended to different classes of models. Using a simple GARCH(p,q) model, we make use of a number of Monte Carlo simulation studies in order to show that calibrating models sequentially based on the eigenvalue hierarchy structure of the Fisher Information Matrix enhances parametric estimation precision and reduce variance. Our results show that an order of 8 reduction on estimation uncertainty can be obtained specially for small sample-sizes (N$\leq$200).

Finally, Chapter 7 reports a successful ongoing diagnosis and post-mortem analysis of the SSEC 2015 bubble made by the Financial Crises Observatory (FCO) while Chapter 8 concludes.

Thesis berater

Autor

Didier Sornette

Guilherme Do Livramento Demos

Information Geometrie und die dynamische Erkennung von finanziellen Blasen und Abstürze

Abstrakt

Diese These ist eine Sammlung von sechs eigenständigen Forschungsarbeiten, von denen drei bereits in Fachzeitschriften veröffentlicht werden, zwei sind in den Prozess der Veröffentlichung und ein Arbeitspapier. Der Schwerpunkt der vorliegenden Arbeit besteht darin, neue statistische Metriken für die laufende Diagnostik von finanziellen Blasen über mehrere Skalen zu schaffen und die Ungewissheit der Schlussfolgerungen über einen zukünftigen Crash-Datum mit anspruchsvollen Likelihood-Schätzern und Information Geometry zu reduzieren.

Kapitel 1 ist ein einleitendes Kapitel und liefert eine Literaturrecherche über finanzielle Blasen. Es definiert auch das Log-Periodische Power-Law Singularity (LPPLS) Modell [56] und zählt einige seiner Engpässe, deren die vorliegende Arbeit eine Lösung vorschlägt.

In Kapitel 2 stellen wir die folgende Frage: Es ist einfacher, die Geburt oder das Platzen von finanziellen Blasen zu erkennen? Wir argumentieren, dass es der Mangel an Wissen darüber ist, wann Blasen platzen werden, die es ihnen erlauben, auch in Gegenwart von rationalen Agenten zu wachsen. Mit dem LPPLS-Modell stellen wir eine Methodik zur Quantifizierung der Unsicherheit eines jeden dieser Parameter (t_1 := Geburt der Blase und t_c := Burst) auf der Grundlage der Eigenwerte und Eigenvektoren der Fisher Information Matrix an Die Best-Fit-Parameter.

Kapitel 3 beweist eine neuartige statistische Metrik, die es erlaubt, die Güte der Passform eines Modells mit unausgewogenen Stichprobengrössen zu vergleichen. Bei der Anwendung auf die LPPLS-Modellkostenfunktion erlaubt die Methodik, den Prozess der

Diagnose des Beginns von Finanzblasen zu automatisieren, ohne weitere exogene Informationen zu erfordern.

Angesichts des Auftretens von Blasen über mehrere Zeitskalen bietet Chapter 4 eine neuartige, mehrstufige Blasendiagnostik-Technologie, die es ermöglicht, automatisch eine laufende Bewertung von Blasen durchzuführen, die kurz-, mittel- und langfristig fällig sind Waage.

Kapitel 5 konzentriert sich auf die Verringerung der Unsicherheit um die Vorhersage der finanziellen Blasen Burst. Basierend auf Eigenvektoren und Eigenwerten der Fisher Information Matrix kalibrieren wir sequentiell die LPPLS-Modellparameter unter Verwendung des Modified Profile Likelihood Schätzers. Mit dem schlimmsten Parameter t_c als focal-Parameter verringert die vorgeschlagene Methodik drastisch die Schätzunsicherheit von $\hat{t}_c$, speziell auf kleine Samples und erlaubt es, robuste Konfidenzintervalle um Parameterschtzungen zu konstruieren.

Kapitel 6 Fragen, wenn die Ergebnisse des vorherigen Kapitels auf verschiedene Klassen von Modellen erweitert werden können. Mit einem einfachen GARCH (p, q) Modell nutzen wir eine Reihe von Monte-Carlo Simulationsstudien, um zu zeigen, dass die auf der Eigenwert-Hierarchiestruktur der Fisher Information Matrix sequentiell basierenden Kalibriermodelle die parametrische Schätzprzision erhöhen und die Varianz reduzieren. Unsere Ergebnisse zeigen, dass eine Ordnung von 8 Reduktion auf Schätzunsicherheit speziell für kleine Stichprobengrössen ($N \leq 200$) erhalten werden kann.

Schliesslich berichtet Chapter 7 eine erfolgreiche laufende Diagnose und Postmortem-Analyse der SSEC 2015 Blase, die von der Financial Crises Observatory (FCO) gemacht wird, während Chapter 8 abschliesst.

Contents

Citations to Previously Published Work

Large portions of Chapter 2 was published in the following paper:

> "Guilherme Demos and Didier Sornette, Birth or burst of financial bubbles: which one is easier to diagnose? Quantitative Finance 17 (5), 657-675 (2017), (online DOI: 10.1080/14697688.2016.1231417, 2016) (http://dx.doi.org/10.1080/14697688.2016.1231417)".

A novel technique for diagnosing the inception of financial bubbles conditional on an end point is presented in Chapter 3. The chapter was submitted to The Journal of Computational Statistics and appears almost entirely in:

> "Demos, Guilherme and Sornette, Didier, Lagrange Regularisation Approach to Compare Nested Data Sets and Determine Objectively Financial Bubbles' Inceptions (July 22, 2017). ETH-Zürich Working Paper Series. Available at SSRN: https://ssrn.com/abstract=3007070".

Chapter 4 provides a novel technique for monitoring the maturation of bubble across multiple scales and was partially submitted to Quantitative Finance,

> "Demirer, R., Demos, G., Gupta, R. and Sornette, D. (2017). On the Predictability of Stock Market Bubbles: Evidence from LPPLS ConfidenceTM Multi-scale Indicators (No. 201752)."

Chapter 5 focuses on diminishing the uncertainty around $\hat{t}_c$ and was published entirely as

> "Vladimir Filimonov, Guilherme Demos and Didier Sornette, Modified Profile Likelihood Inference and Interval Forecast of the Burst of Financial Bubbles", Quantitative Finance, DOI: 10.1080/14697688.2016.1276298, pp. 1-20 (2017) (http://arxiv.org/abs/1602.08258 and http://ssrn.com/abstract=2739832).

Chapter 6 is still a working paper and tries to understand the relationship between the hierarchical calibration of models and precision gains on parameters estimates. The paper can be cited as:

> "Guilherme Demos and Didier Sornette, Hierarchical calibration of sloppy statistical models based on information geometry. ETH-Zürich Working Paper Series".

Finally, Chapter 7 has been published in The Journal of Investment Strategies as

> "Didier Sornette, Guilherme Demos, Qun Zhang, Peter Cauwels, Vladimir Filimonov and Qunzhi Zhang", Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash, Journal of Investment Strategies 4 (4), 77-95 (2015) (Swiss Finance Institute Research Paper No. 15-32. Available at (http://ssrn.com/abstract=2647354)

Acknowledgments

Being able to do research at the Chair of Entrepreneurial Risks here at the ETH zürich was by far one the most incredible experiences of my life. The amount of exchanged information with fellow researchers, the new ideas and the fast pace that discovery takes place are incredible. I have to thank many people whose help was extremely precious during these three years.

First of all, I must thank Didier Sornette for everything; for believing in me and for being there for me both in science and in life. Not too many people have the chance to work with their idols; well, I did and I am extremely thankful for that chance. Your thirst to live is something that never ceases to amaze me. Moreover, thank you for always making sure that we were healthy and fit and for being this inexhaustible source of inspiration and wisdom.

Another person who contributed plenty for my formation and for this thesis is Vladimir Filimonov. Thank you for teaching me how to proper code in Python, how to do statistics and for being this outstanding friend who is always eager to help and to provide valuable advice.

I must also thank Professors Eraldo Sérgio Barbosa da Silva and Guilherme Moura for their support throughout the M.Sc. program at the Federal University of Santa Catarina and for sparkling the flame of curiosity in me which led me to pursue the Ph.D degree outside of Mainstream Economics. Thank you very much.

It wouldn't be fair not giving thanks to all my friends and colleagues from the Chair of Entrepreneurial Risks who definitely made my life and work way more pleasant; Isabella Bieri, Jan C. Gerlach, Diego Ardilla, Zhuli He, Zalan Forro, Qunzhi Zhang, Dimitry Chernov, Qun Zhang, Dorsa Sanadgol, Yavor Kramer, Tobias Huber, Michael Schatz, Lucas Fievet, Ahmed Ahmed, Tatyana Kovalenko, Richard Senner, Rebecca Westphal, Ke Wu and Spencer Wheatley.

I also wish to thank my family for the amazing support and for believing in my dreams. I would not be where I am today were it not for you. Thank you Antonio, Marlene, Gabriel and Ana.

Last but not least, I wish to thank Prof. Thorsten Hens for accepting to be my co-referee and Prof. Antoine Bommier for being the president of the thesis committee.

In loving memory of my brother, Gabriel Demos.

Donec iterum conveniant.

Chapter 1

Introduction and summary

1.1 Introduction

"The history of financial markets has been dotted with episodes of bubbles, during which market values of assets vastly exceeded reasonable assessments of their fundamental value." [125]. Defining financial bubbles or understanding why they occur, persist and when they shall burst it is by no means a straightforward task. Besides problems with the definition of the true fundamental value of a security [48], not all bubbles are the same [36]. There are those fuelled by low interest rates (i.e. Economic Policy driven [76]), agents optimism of ever growing prices (Behavioural driven [68]), technological breakthroughs (Innovation driven [95]) among others.

Due to differences in the underlying generator process, bubbles tend occur over different time-scales. As the reader shall see on Chapter 4, one of our research findings suggests that the magnitude of crashes suffered by bubbly securities are a direct function of the time-scale that they have been inflating. For example, the low interest rate environment which started in the late 1990's with the approval of the Gramm-Leach-Bliley Act. and enhanced post 2001/09/11 terrorist attacks on U.S. territory, culminated in the 2008 liquidity crisis; a long-term bubble maturing throughout the 2000's. Meanwhile, the cryptocurrency Bitcoin (BTC) characterised by its wild fluctuations and lack of institutional coordination in terms of policies, suffered from a number of bubbles due to herding and short-term speculative behaviour of agents [6, 43].

Why do bubbles start and why they persist are subjects of much controversy [103, 125]. I quote Mr. Graham [48] on the problem of defining fundamental value: "*We must recognise, however, that intrinsic value is an elusive concept. In general terms it is understood to be that value which is justified by the facts, e.g., the assets, earnings, dividends, definite prospects, as distinct, let us say, from market quotations established by artificial manipulation or distorted by psychological excesses. But it is a great mistake to imagine that intrinsic value is as definite and as determinable as is the market price*". According to the author, the inadequacy or incorrectness of the data, uncertainties regarding the future, and irrational behaviour of the market are major obstacles to the success of security analysis. Indeed, the task of correctly pricing the fundamental value of a given firm or asset involves much uncertainty, due to inherent sensitivity of the method of discounting future dividends or earnings through time [49].

Neo-Classical Economic stresses that prices should reflect all available information at a given time [87] and consequently any mispricing should be promptly arbitraged by market makers. In reality, many anomalies seem to persist, the bubble phenomenon being arguably the prominent one. In fact, financial bubbles were shown to be persistent even in the presence of rational expectation agents [18]. The behavorial feedback-loop theory for example, stress that higher present observable prices leads to higher expected price appreciation [10, 94]. This resembles the resale option theory [50] where agents are willing to pay more than their own expectation of the asset fundamental value because they expect to re-sell the asset in the future to a more optimistic agent, for a higher price. This argument was explored in a more quantitative way by [54].

In short, Neo-Classical Economic theory is based on the assumptions of rationality of economic agents and of their homogeneous beliefs, or equivalently that their aggregate behaviours can be represented by a representative agent embodying their effective collective preferences. There is growing evidence that equilibrium models and the efficient market hypothesis (EMH) cannot provide a fully reliable framework for explaining the stylized facts of price formation [34] and alternative behavioural theories are being proposed [72] We believe that a complex systems approach to research is crucial to capture the inter-dependent and out-of-equilibrium nature of financial markets, whose total size amounts to at least 300% of the world GDP and of the cumulative wealth of nations [103].

In this thesis, we make use of cutting-edge technology and allow data to do the talk when it comes to explaining bubbles and crashes. We rely on tools and models from statistical-physics and provide an exhaustive analysis of this extreme phenomena (using both synthetic and real-world time-series over multiple scales, markets and securities) within a complex systems framework where financial markets are viewed as an adaptive dynamical system [110]. Our major goal here is not only to understand why bubbles occur and provide confidence intervals about when a change of regime will occur but rather to create novel data-driven techniques for diagnosing in real-time the growing of financial bubbles which allows one to *ad-hoc* reach conclusions and make inferences about the future trajectory of the system with high probability. The importance of such metrics cannot be stressed enough since their use by governments and regulatory agencies have the power to save countless jobs and to alleviate both recession and depression periods or even avoid them completely.

The goals and subject matter of the seven parts are sufficiently different to merit individual introductions and summaries, which now follow without further ado.

- **Chapter 2: Birth or Burst of Financial Bubbles: which one is easier to di-**

agnose? Within a Rational Expectations framework of [18], there is a view proposed by [1] that financial bubbles persist due the lack of knowledge of agents about their beginning rather than burst date. Rather than the beginning of bubbles, Chapter 1 argues that it is their ends that are devilishly difficult to forecast with any useful accuracy. The contribution of this Chapter is to propose a methodology to quantify these intuitions that is based on a series of works from Professor James Sethna and his group [74]. We make use of the LPPLS model in order to diagnose bubble regimes and endogenize the bubble starting date (i.e. t_1) thus allowing us to directly compare the level of indeterminacy of t_1 and the end of bubbles (i.e. t_c). Results on both synthetic and 8 real world time-series provide evidence that is the lack of certainty about the crash date rather than the beginning of bubbles, that allows bubbles to persist even when expectations are rational.

- **Chapter 3: Lagrange regularisation approach to compare nested data sets and determine objectively financial bubbles' inceptions.** Inspired by the question of identifying the start time τ of financial bubbles, we address the calibration of time series in which the inception of the latest regime of interest is unknown. By taking into account the tendency of a given model to overfit data, we introduce the Lagrange regularisation of the normalised sum of the squared residuals, $\chi^2_{np}(\mathbf{\Phi})$, to endogenously detect the optimal fitting window size $:= w* \in [\tau : \bar{t}_2]$ that should be used for calibration purposes for a fixed pseudo present time $\bar{t}_2$. The performance of the Lagrange regularisation of $\chi^2_{np}(\mathbf{\Phi})$ defined as $\chi^2_{\lambda}(\mathbf{\Phi})$ is exemplified on a simple Linear Regression problem with a change point and compared against the Residual Sum of Squares (RSS) $:= \chi^2(\mathbf{\Phi})$ and RSS/(N-p)$:= \chi^2_{np}(\mathbf{\Phi})$, where N is the sample size and p is the number of degrees of freedom. Applied to synthetic models of financial bubbles with a well-defined transition regime and to a number of financial time series

(US S&P500, Brazil IBovespa and China SSEC Indices), the Lagrange regularisation of $\chi^2_\lambda(\mathbf{\Phi})$ is found to provide well-defined reasonable determinations of the starting times for major bubbles such as the bubbles ending with the 1987 Black-Monday, the 2008 Sub-prime crisis and minor speculative bubbles on other Indexes, without any further exogenous information. It thus allows one to endogenise the determination of the beginning time of bubbles, a problem that had not received previously a systematic objective solution.

- **Chapter 4: On the Predictability of Stock Market Bubbles: Evidence from LPPLS Confidence$^{\mathbf{TM}}$ Multi-scale Indicators.** We examine the predictive ability of market-based indicators over the positive and negative bubbles in the S&P 500 index using the LPPLS Confidence^{TM} Multi-scale Indicators to the S&P-500 index. We find that the LPPLS framework is able to successfully capture, ex-ante, some of the prominent bubbles across different time scales, such as the Black Monday, Dot-com, and Subprime Crisis periods. We then show that measures of short selling activity have robust predictive power over negative bubbles, in line with the previous studies that short sellers have predictive ability over stock price crash risks. Interestingly, however, the liquidity premium embedded in stock prices is found to have robust predictive power over both the negative and positive bubbles, suggesting that the risk premium associated with exposure to liquidity shocks can be used to predict the occurrence of booms and crashes. The evidence overall points to the predictability of stock market bubbles, while positive and negative bubbles exhibit significantly different predictability patterns.

- **Chapter 5: Modified Profile Likelihood and Interval Forecast of the End of Financial Bubbles.** We present a detailed methodological study of the application of the modified profile likelihood method for the calibration of nonlinear financial models characterised by a large number of parameters. We apply the general approach to the Log-Periodic Power Law Singularity (LPPLS) model of financial bubbles. This model is particularly relevant because one of its parameters, the critical time t_c signalling the burst of the bubble, is arguably the target of choice for dynamical risk management. However, previous calibrations of the LPPLS model have shown that the estimation of t_c is in general quite unstable. Here, we provide a rigorous likelihood inference approach to determine t_c, which takes into account the impact of the other nonlinear (so-called "nuisance") parameters for the correct adjustment of the uncertainty on t_c. This provides a rigorous interval estimation for the critical time, rather than a point estimation in previous approaches. As a bonus, the interval estimations can also be obtained for the nuisance parameters (m, ω, damping), which can be used to improve filtering of the calibration results. We show that the use of the modified profile likelihood method dramatically reduces the number of local extrema by constructing much simpler smoother log-likelihood landscapes. The remaining distinct solutions can be interpreted as genuine scenarios that unfold as the time of the analysis flows, which can be compared directly via their likelihood ratio. Finally, we develop a multi-scale profile likelihood analysis to visualize the structure of the financial data at different scales (typically from 100 to 750 days). We test the methodology successfully on synthetic price time series and on three well-known historical financial bubbles.

- **Chapter 6: Hierarchical calibration of sloppy statistical models based on information geometry.** Model, data and optimisation approach should be considered as a "new complex system" in need of a general theory. We propose a powerful step in this direction by constructing a natural parametric hierarchy that is ubiquitous for multi-parameter models, and which provides a roadmap on how models should be calibrated to experimental data in order to greatly reduce estimation uncertainty. By employing sophisticated parameter estimation techniques with information from eigenvalues and eigenvectors of the Fisher Information Matrix, the proposed calibration scheme is capable of turning non-identifiable parameters into identifiable ones. Using a GARCH(1,1) model, our Monte Carlo simulation results show a factor of 8 reduction on estimates uncertainty relative to the standard Quasi Maximum Likelihood approach. The performance is notoriously good when using small sample (i.e. N=100, ..., 200).

- **Chapter 7: Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash.** We assess the performance of the real-time diagnostic, openly presented to the public on the website of the Financial Crisis Observatory (FCO) at ETH Zurich, of the bubble regime that developed in Chinese stock markets since mid-2014 and that started to burst in June 2015. The analysis is based on (i) the economic theory of rational expectation bubbles, (ii) behavioural mechanisms of imitation and herding of investors and traders and (iii) the mathematical formulation of the Log-Periodic Power Law Singularity (LPPLS) that describes the critical approach towards a tipping point in complex systems. The authors document how the real-time predictions were presented in the automated analysis of the FCO, as well as in our monthly FCO Cockpit report of June 2015. A complementary post-mortem analysis on the nature and value of the LPPLS methodology to diagnose the SSEC bubble and its termination is also given.

- **Chapter 8: Conclusion.**

1.2 The Log-Periodic Power-Law Singularity Model

The working-horse model of this thesis for detecting bubbles and predicting it's future path is the LPPLS model [56]. In this section, both theory and the mathematical formulation of the model are presented in detail. A large portion of this Chapter is also devoted to studying the process of calibrating the LPPLS model and stressing the challenges surrounding parameter estimates and model fitting. It is here that two major concepts (rigidity and sloppiness) are presented to the reader. These concepts are central building blocks of the present thesis.

The LPPLS model we use was proposed by [56] and is usually referred to in the literature of financial bubbles as the JLS model. It starts from the rational expectation settings of [18], where the observed price P of an asset can be written as

$$P = P_{fundamental} + P_{bubble} \tag{1.1}$$

where $P_{fundamental}$ and P_{bubble} represent respectively the fundamental value and the bubble component. The equations shows that the price is a linear combination of the fundamental value and the bubble component. The JLS Log-Periodic Power-Law (LPPLS) model specifies the dynamics of the bubble component independently of the dynamics of the fundamental price. The later can be specified according to standard valuation models, for instance leading to the usual geometrical random walk benchmark [8]. The model adds to this featureless fundamental price the so-called log-periodic power law structure, which is used to diagnose the presence of bubbles.

The LPPLS model is based on the standard jump-diffusion model, where the logarithm of the asset price $p(t)$ follows a random walk with a varying drift $\mu(t)$ in the presence of discrete discontinuous jumps:

$$\frac{dp}{p} = \mu(t)dt + \sigma(t)dW - \kappa dj. \tag{1.2}$$

Here, $\sigma(t)$ denotes the volatility, dW is the infinitesimal increment of a standard Wiener process and dj represents a discontinuous jump such as $j = \chi(t-t_c)$, where $\chi(\cdot)$ is a Heaviside function and t_c denotes the time of the jump. Within the "bubble-crash" framework, t_c defines the "critical time", which is defined within the rational expectations framework as the most probable time for the crash or change of regime to occur. The parameter κ then quantifies the amplitude of the crash when it occurs. The expected value of dj defines the crash hazard rate $h(t)$: $\mathrm{E}[dj] = h(t)dt$.

According to the Johansen-Ledoit-Sornette (JLS) model [58, 56, 57], the complex actions of noise traders can be aggregated into the following dynamics of the hazard rate:

$$h(t) = \alpha(t_c - t)^{m-1}\big(1 + \beta\cos(\omega\ln(t_c - t) - \phi')\big), \tag{1.3}$$

where α, β, ω and ϕ' are parameters. The core of the model is the singular power law behavior $(t_c - t)^{m-1}$ that embodies the mechanism of the positive feedback at the origin of the formation of bubble leading to a super-exponential price growth. The oscillatory dressing $1+\beta\cos(\omega\ln(t_c-t)-\phi')$ takes into account the existence of a possible hierarchical cascade of panic acceleration punctuating the course of the bubble. The particular form of the log-periodic function $\cos(\omega\ln(t_c-t)-\phi')$ in (1.3) is a first-order expansion of the general class of Weierstrass-type functions [46, 128] that describes the discrete-scale invariance around tipping points in complex natural and socio-economic systems [108, 109].

Under the no-arbitrage condition ($E[dp] = 0$), the excess return $\mu(t)$ is proportional to the crash hazard rate $h(t)$: $\mu(t) = \kappa h(t)$. Then direct solution of the equation (1.2) with the given dynamics of the hazard rate (1.3) under the condition that no crash has yet occurred ($dj = 0$) leads to the following Log-Periodic Power Law Singularity (LPPLS) equation for the expected value of a log-price:

$$\text{LPPLS}(t) \equiv \mathrm{E}[\ln p(t)] = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi), \tag{1.4}$$

where $B = -\kappa\alpha/m$ and $C = -\kappa\alpha\beta/\sqrt{m^2 + \omega^2}$. It is important to stress that the exact solution (1.4) describes the dynamics of the average log-price only up to critical time t_c and cannot be used beyond it. This critical time t_c corresponds to the termination of the bubble and indicates the change to another regime, which could be a large crash or a change of the average growth rate.

The LPPLS model in its original form (1.4) is described by three linear parameters (A, B, C) and four nonlinear parameters (m, ω, t_c, ϕ). As discussed further in Section 1.2.2, the calibration of the model can be performed using a two-stage procedure. First, the linear parameters A, B, C for fixed values of m, ω, t_c, ϕ can be obtained directly via the solution of a matrix equation. Second, the non-linear parameters m, ω, t_c, ϕ can be found using a nonlinear optimization method. Notwithstanding the reduction from 7 to 4 of the number of parameters to determine, the numerical optimization is not straightforward, as the cost function possesses a quasi-periodic structure with many local minima. Any local optimization algorithm fails here and an extra layer involving so-called metaheuristic algorithms [119] is needed in order to find the global optimum. Thus, the *taboo search* [30] has often been used to perform this metaheuristic determination of the 4 nonlinear parameters of the LPPLS function (1.4).

A better approach has been suggested by [38], which consists in reformulating the model (1.4) in a way that significantly simplifies the calibration procedure. The reformulation is based on the variable change

$$C_1 = C \cos \phi, \quad C_2 = C \sin \phi, \tag{1.5}$$

so that equation (1.4) becomes

$$\text{LPPLS}(t) := E[\ln P(t)] = A + B(f) + C_1(g) + C_2(h), \tag{1.6}$$

where, $ln[P(t)]$ is the price logarithm vector of size N and

$$f \equiv (tc - t)^m, \tag{1.7}$$

$$g \equiv (t_c - t)^m cos(\omega \; ln(t_c - t)), \tag{1.8}$$

$$h \equiv (t_c - t)^m sin(\omega \; ln(tc - t)). \tag{1.9}$$

In this form, the LPPLS function has only 3 nonlinear (t_c, ω, m) and 4 linear (A, B, C_1, C_2) parameters. As shown in [38], this transformation significantly decreases the complexity of the fitting procedure and improves its stability tremendously. This is because the modified cost function for (1.6) is now free from quasi-periodicity and enjoys good smooth properties with one or a few local minima in the case where the model is appropriate to the empirical data.

1.2.1 LPPLS Filtering Conditions

Additional constraints have been proposed, based on compilations of extensive analyses of historical bubbles [114, 60, 71]. Previous calibrations of the JLS model have further shown the value of additional constraints imposed on the nonlinear parameters in order to remove spurious calibrations (false positive identification of bubbles) [106, 56, 55, 44].

1. **exponent** $0.1 < m < 0.9$: This more stringent constraint than the one ($0 < m < 1$) discussed above improves the power of discriminating bubbles by removing calibrations that select parameters too close from the bounds.

2. **angular log-periodic frequency** $6 < \omega < 13$: This condition ensures that the preferred scaling ratio $:= e^{\frac{2\pi}{\omega}}$ of the log-periodicity [102] is of the order of 2, as suggested by general theoretical arguments [86].

3. **Search interval for t_c given by** $t_2 + 1 < t_c < t_2 + \eta||t_2 - t_1||$, where time is in units of days. $[t_1, t_2]$ is the time window in which the calibration is performed. One can think of t_2 as the "present" time, in the sense that it is the latest time at which information is used to calibrate the model. The beginning t_1 of the fitting interval sets the time scale $t_2 - t_1$ of the analysis. The factor η is of the order of 1, often chosen about 1/3 or 1/2 to ensure that the remaining bubble lifetime is within an horizon of predictability estimated to scale proportionally to the width of the analysing window.

4. **"Damping" condition:** Within the RE bubble framework used by the JLS model, the crash hazard rate is found to be proportional to the expected return during the bubble, conditional on the fact that the bubble has not yet burst. Since the crash hazard rate is by definition a non-negative quantity, this imposes that the expected return during the bubble should be non-negative. This can be shown to impose that condition $D \geq 1$ where $D := \frac{|B|m}{\omega\sqrt{C_1^2+C_1^2}}$ [47]. Intuitively, the amplitude of the log-periodic oscillations should be not too large compared with the acceleration of the price so that their combination ensures the non-negativity of the expected return at all times during the bubble. Of course, the realized stochastic return can be negative. It is only its expectation that should be non-negative.

5. **Number of oscillations:** The number of oscillations (half-periods) of the log-periodic component is given by $O := \frac{\omega}{\pi} \ln\left(\frac{t_c - t_1}{t_c - t_2}\right)$. [52] showed that a minimum number of oscillations is needed to distinguish a genuine log-periodic signal from one that could be generated by noise. In the present implementation, we qualify a fit only if $O \geq 2.5$.

6. **residuals** $r \sim$ **AR(1):** [70] have emphasised the need for the residuals of the LPPLS calibration to be mean-reverting in order for the calibration of the log-price by the LPPLS model to be consistent. Hence, we test for the mean-reverting (O-U) property of the residuals of the fit, using both the standard unit-root Augmented Dickey-Fuller and Phillips-Perron tests.

1.2.2 Estimating the LPPLS model

Fitting (1.4) envolves minimising the cost function,

$$F(t_c, m, \omega, A, B, C_1, C_2) = \frac{1}{2} \sum_{i=1}^{N} \Big[ln[P(t_i)] - A - B(f_i) - C_1(g_i) - C_2(h_i) \Big]^2. \quad (1.10)$$

Slaving the intrinsic linear (IL) parameters $\{A, B, C_1, C_2\}$ to the remaining nonlinears (NL) $\{t_c, m, \omega\}$, yields

$$\{\widehat{t_c}, \widehat{m}, \widehat{\omega}\} = arg \min_{\{t_c, m, \omega\}} F_1(t_c, m, \omega), \quad (1.11)$$

where $(\widehat{\bullet})$ denote estimated parameters. The cost function $F_1(t_c, m, \omega)$ is thus given by

$$\chi^2 := F_1(t_c, m, \omega) = \min_{\{A, B, C_1, C_2\}} F(t_c, m, \omega, A, B, C_1, C_2). \quad (1.12)$$

Using only 3 NL parameters, the optimization problem

$$\{\widehat{A}, \widehat{B}, \widehat{C}_1, \widehat{C}_2\} = arg \min_{\{A,B,C_1,C_2\}} F(t_c, m, \omega, A, B, C_1, C_2), \tag{1.13}$$

can be obtained analytically using ordinary least squares:

$$\begin{bmatrix} N & \sum f_i & \sum g_i & \sum h_i \\ \sum f_i & \sum f_i^2 & \sum f_i g_i & \sum f_i h_i \\ \sum g_i & \sum f_i g_i & \sum g_i^2 & \sum g_i h_i \\ \sum h_i & \sum f_i h_i & \sum g_i h_i & \sum h_i^2 \end{bmatrix} \begin{bmatrix} \widehat{A} \\ \widehat{B} \\ \widehat{C}_1 \\ \widehat{C}_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i f_i \\ \sum y_i g_i \\ \sum y_i h_i \end{bmatrix}.$$

1.2.3 Reaching the best-fit

Let us denote $\boldsymbol{\Phi} = [m, \omega, t_c, A, B, C_1, C_2]$ as the full parameter set and $\boldsymbol{\theta} = [m, \omega, t_c]$ as the NL parameter set of length m. Also, allow every residual $r_i := \sum_{i=1}^{N} \frac{1}{2}(y(t)_i - LPPLS(\theta; t)_i)$, of Eq. (1.11) be collected into a vector $\boldsymbol{r} : \Re^m \to \Re^N$ which reads $\boldsymbol{r} = (r_1(\theta), r_2(\theta), ..., r_N(\theta))^\intercal$. From now on, bold letters denote either matrices or vectors.

Embedded within a *m-dimensional* space, the slaving procedure displayed by Eq. (1.12) allows one to search for the smallest distance $\frac{1}{2}||\mathbf{r}(\boldsymbol{\theta})||_\mathbf{2}^\mathbf{2}$,where $|| \bullet ||$ denotes the Euclidean norm. *Taylor expansion* is a wide-used mechanism by optimization algorithms for approximating the value of a function at a given point. Assuming our cost function (χ^2), completely known at $\theta_k = \theta^*$, and $m \in \Re^{++}$, for the case of $m = \boldsymbol{\theta}$

variables, the Taylor expansion reads

$$\chi^2(\theta_i^* + \Delta\theta_i, \theta_j^* + \Delta\theta_j) = \chi^2(\theta_i^*, \theta_j^*) + \left[\frac{\partial \chi^2}{\partial \theta_i}\bigg|_{(\theta_i^*, \theta_j^*)} \Delta\theta_i + \frac{\partial \chi^2}{\partial \theta_j}\bigg|_{(\theta_i^*, \theta_j^*)} \Delta\theta_j \right] + \frac{1}{2}\left[\frac{\partial^2 \chi^2}{\partial^2 \theta_i}\bigg|_{(\theta_i^*, \theta_j^*)} \ldots \right. \tag{1.14}$$

$$\left. (\Delta\theta_i)^2 + 2\frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j}\bigg|_{(\theta_i^*, \theta_j^*)} \Delta\theta_i, \Delta\theta_j + \frac{\partial^2 \chi^2}{\partial_j^2}\bigg|_{(\theta_i^*, \theta_j^*)} (\Delta\theta_j)^2 \right]. \tag{1.15}$$

Organizing the displacements into column vectors $[\Delta\theta_i \, \Delta\theta_j]^\intercal$, Expression (1.15) can be written matricially in terms of fitting residuals as

$$\chi^2(\boldsymbol{\theta}^* + \boldsymbol{\Delta\theta}^*) = r(\boldsymbol{\theta}^*) + \vec{\bigtriangledown} r(\boldsymbol{\theta}^*)^\intercal \boldsymbol{\Delta\theta}^* + \frac{1}{2}\Delta(\boldsymbol{\theta}^*)^\intercal \mathbf{H}(\boldsymbol{\theta}^*)\boldsymbol{\Delta\theta}^*. \tag{1.16}$$

Expression (1.16), consists on the summation of a first order/linear variation and a second order/quadratic variation given in terms of the *Gradient* ($\vec{\bigtriangledown} r$), and *Hessian* ($\boldsymbol{H}$), respectively. The Gradient reads,

$$\bigtriangledown\chi^2(\boldsymbol{\theta}) := d\chi^2 = \frac{\partial r_i}{\partial \theta_i} d(\theta_i) \, + \, \frac{\partial r_i}{\partial \theta_j} d(\theta_j), \tag{1.17}$$

and is collected by a $(1{\times}m)$ matrix known as the Jacobian,

$$J(\boldsymbol{\theta}) \;\; = \begin{bmatrix} \frac{\partial r_i}{\partial \theta_i} \\ \\ \vdots \\ \\ \frac{\partial r_i}{\partial \theta_j} \end{bmatrix} \equiv \begin{bmatrix} \frac{\partial r_i}{\partial m} & \frac{\partial r_i}{\partial \omega} & \frac{\partial r_i}{\partial t_c} \end{bmatrix}^\intercal_{i,j=1,2,\ldots,M;}.$$

The slope of $\chi^2|_{\theta_k=\theta^*}$ at the point $\theta_k = \theta_*$, is a column vector of first-order partial derivatives (Eq. 1.17), representing the direction in which the the *cost* will increase more rapidly. This direction guides algorithms for correctly finding the solution (1.11) and uncertainty on estimated parameter may arise when this landscape is degenerate [74].

The LM algorithm, commonly used for solving nonlinear optimisation problems, is a Newton-type algorithm. When fitting residuals are small about the best fit it approximates the objective function $\chi^2\big|_*$ by expanding the quadratic term in Eq. (1.16) and thus infinitesimaly updating parameters $\delta(\boldsymbol{\theta})$, according to

$$\delta(\boldsymbol{\theta}) \;=\; -\,(J^T J + \lambda D^T D)^{-1} \bigtriangledown \chi^2(\boldsymbol{\theta}). \tag{1.18}$$

where D and λ are respectively, a matrix containing diagonal elements of the second-order derivatives matrix and a damping parameter. Moreover, λ rotates the search direction of the algorithm from the Gauss-Newton to the Gradient direction, thus controlling the step size used in the Taylor expansion for finding the roots of the objective function.

The Jacobian can be calculated using a semi-analytical method like the *centered finite-difference scheme*, which reads

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}, \tag{1.19}$$

where h is the time step. The Hessian matrix is then approximated in terms of the residuals as follows

$$\frac{\partial \chi^2(\boldsymbol{\theta})}{\partial \theta_i \theta_j} \;=\; \sum_{i=1}^{m} \left(r_i \frac{\partial^2 r_i(\theta)}{\partial \theta_i \partial \theta_j} + \frac{\partial r_i(\theta)}{\partial \theta_i} \frac{\partial r_i(\theta)}{\partial \theta_j} \right), \tag{1.20}$$

which is equivalent to

$$\boldsymbol{J}^\intercal \boldsymbol{J} \ := \ \boldsymbol{H} \ = \ \sum_{i=1}^{N} (\bigtriangledown r(\boldsymbol{\theta})_i)(\bigtriangledown r(\boldsymbol{\theta})_i)^\intercal. \tag{1.21}$$

It is known from PCA *(Principal Component Analysis)* that the output sensitivity of a model with respect to input parameters can be measured by eingenvalues λ and eigenvectors $\vec{\lambda}$ of the Hessian matrix $\boldsymbol{H}$ (Eq. 1.21), which embody the varying important (stiff) and unimportant (sloppy) combination of parameters.

1.2.4 Rigid vs. sloppy directions in parameter space

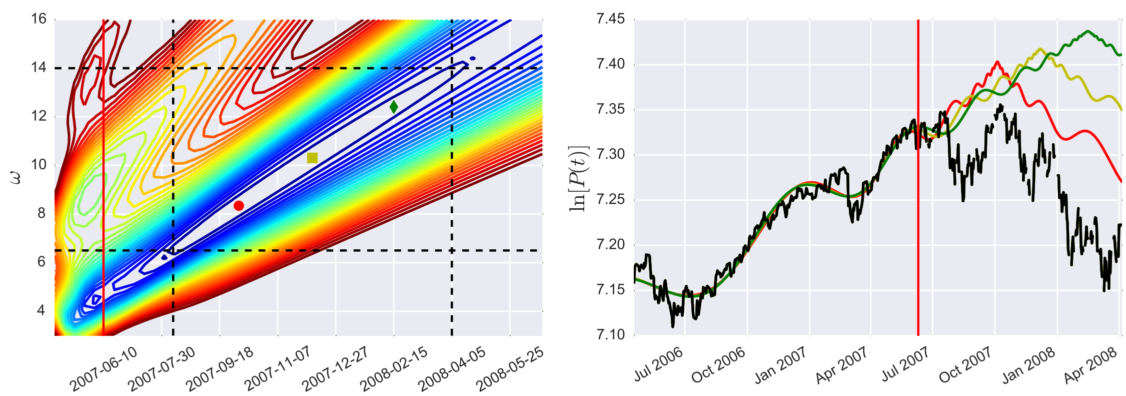


Figure 1.1: Uncertainty surrounding $\hat{t}_c$ and $\widehat{\omega}$: The cost function χ^2 of t_c and ω (x and y-axis of the left hand-side figure) form ellipses about the best-fit region. Notice that varying t_c together with ω has little influence on square residuals. Looking at the right hand-side figure, each coloured line corresponds to the extrapolation of our model using values of t_c denoted by dot, square and diamond markers.

Notwithstanding a number of improvements concerning the calibration of the LPPLS model, including meta-search heuristics [117] and reformulation of the equations to reduce the number of nonlinear parameters [38], the calibration of the LPPLS model remains a bottleneck towards achieving robust forecasts and a matter of contention [20, 116]. The main reason for this is the parameteric geometry which can be undestood via the Hessian matrix approximated at the best fit parameters and the Jacobian term (Eq. 1.21 and Eq. 1.18 respectively). These metrics allows one to understand the uncertainty permeating the LPPLS model and it's parametric sensitivity when fitted to data. How this uncertainty affects prediction can be better visualised through Figure 7.6. The cost function χ^2 of t_c and ω (x and y-axis of the left hand-side figure) displays ellipses about the best-fit region. This feature leads to uncertainty on parameter values since each parameter can adapt to changes of the other while keeping the cost χ^2 stable. We illustrate this feature using the log-price time series of the $S\&P$-500 Index on Fig. 1.1. Looking at the right hand-side figure, each coloured line corresponds to the extrapolation of our model using values of t_c denoted by dot, square and diamond markers (corresponding to a calendar date of $\approx$ Oct. 07, Dec. 07 and Feb. 08, respectively). All fits mimic the time-series in the same manner because of the trade-off between ω and t_c while the forecasted change of regime date t_c differs substantially.

This behaviour is referred to from now on as sloppiness, in accordance to a number of studies from [74]. Overcoming sloppiness and therefore obtaining significant estimates of the LPPLS parameters (specially t_c), constitute the core of this thesis. As we shall see, the exploration of the geometry of non-linear squares fits leads to a number of practical insights for the proper calibration of sloppy statistical models, including the LPPLS. Besides understanding and quantifying what are the important and unimportant parametric directions of a model, Information Geometry allows us to proper detect the beginning of financial bubbles and to turn practical non-identifiable parameters (i.e. optimisation problems such as the one depicted by Fig. 1.1) into practical identifiable ones (i.e. a convex-like cost function as observed in Oridinary Least Squares optimisation problems) [118].

Chapter 2

Birth or Burst of Financial Bubbles: Which one is easier to diagnose?

2.1 Introduction

Financial bubbles and their subsequent crashes provide arguably the most visible departures from well-functional efficient markets. There is an extensive literature (see e.g. the reviews [64, 25]) on the causes of bubbles as well as the reasons for bubbles to be sustained over surprising long period of times. One prominent question is: why are not arbitragers quelling bubbles in the same way that investment opportunities based on market imperfections tend to be removed by arbitrage? Among many arguments advanced in the literature, we here focus on the one suggested by [1]. In the authors' set-up, market participants become aware about the mispricing developing during the bubble in a disorganised fashion, one after another. Thus, opinions differ about the date when the bubble really started, preventing a coordinated arbitrage that would have put an end to the burgeoning bubble.

We question this view and mechanism. Indeed, from informal private exchanges with a large number of practitioners during massive bubbles (such as the dot-com bubble that terminated in 2000), it appears that the existence of a bubble permeates rather early in the consciousness of professionals. For instance, hedge-funds tend to correctly diagnosed bubbles in their early stages [49]. Think for instance of the famous "irrational exuberance" quip of Federal Reserve chairman, Alan Greenspan, issued on Dec. 5, 1996, more than three years before the bubble finally burst. One could have thought that such a strong statement from such a respected authority (at the time) would have synchronised the attention of arbitragers, leading to the demise of this bubble. In fact, after a minor volatility spike lasting no more than a few days following Greenspan's statement, the markets roared unabated to new highs, with the years of 1998 and 1999 enjoying a sky-rocketing cumulative return of more than 1100% for the technology sector. This suggests that bubbles do not persist and flourish because of a lack of general perception of their presence at early times. Rather than shorting the bubble, it is in the interest of investors to "surf" it as long as it lasts [24], given that the gains are so extraordinary that they typically attract less and less sophisticated investors as well as generally less well-informed foreigners, further fuelling the ascent of the bubble [97, 110]. Moreover, it often happens that value investors opportunistically transform into momentum followers during bubble regimes.

Rather than the beginning of bubbles, we argue that it is their ends that are devilishly difficult to forecast with any useful accuracy. As Keynes famously commented: "The market can stay irrational longer than you can stay solvent." And Greenspan's statement mentioned above also illustrates the discrepancy between an early diagnostic and the much delayed conclusion of the bubble. The contribution of the present article is to propose a methodology to quantify these intuitions. We put numbers on how precise can be the early diagnostic of a nucleating bubble and the estimation of its lifetime. For this, we use the log-periodic power law singularity (LPPLS) model [56, 110] to represent bubble regimes. We endogenize t_1 as a parameter to calibrate and estimate the LPPLS model on synthetic time series generated by the LPPLS model with noise and on eight historical bubble cases. In order to determine the uncertainties of the calibrated parameters we calculate the eigenvalues and eigenvectors of the Fisher Information Matrix defined from the cost function of the calibration exercise, following [19], [123] and [74]. Comparing the eigenvalues and the parameters that contribute to the corresponding eigenvalues, we can estimate the "rigid" parameters versus the "sloppy" ones, the later being characterised by very small eigenvalues so that the cost function is essentially insensitive to their specific values, making them impossible to determine precisely. We find that the eigenvalues of the Hessian matrix approximated at the best-fit parameters whose eigenvectors are dominated by t_1 tend to be three orders of magnitude larger than those controlled by t_c. As a rough estimate, this implies that the errors on t_c are about 30 times larger that on t_1. The determination of the end of the bubble is thus much more difficult that its beginning. This suggests that the lack of knowledge regarding the end of the bubble t_c should play a more important role on asset price inflation than the lack of agents consensus about t_1.

The article is structured as follows. Section 2.1.4 presents an initial result of the sloppy-rigid analysis performed on synthetic data. We then extend the application of the method to real world data in Sec. 2.2 and conclude in Section 2.3.

2.1.1 Expanded parameter space to endogenize t_1

Because the goal of the present Chapter is to determine which of the beginning or the end of a bubble is best estimated, we propose to endogenize the search for the beginning of the bubble in the calibration of the parameters. For this, we define the beginning of the bubble as the 'best' beginning time t_1 of the interval $[t_1, t_2]$ in which the calibration is performed, 'best' in the sense of minimising an appropriate scaled goodness of fit in combination with filtering conditions 1-6 presented on Section 1.2.1. As we shall see, it is not possible to rely solely on the scaled goodness of fit measure because this metric is essentially degenerate as the number degrees of freedom is reduced. More specifically, as $t_1 \to t_2$ the metric tends to yield smaller and smaller values even when scaled due the reduced number of data points used for fitting the model. However, since plateaus are observed in certain values of t_1 our results can be trustworthy when analysing the cost function through a local perspective rather than a global one. This leads to an expanded nonlinear parameter set $\mathbf{\Phi} = \{m, \omega, t_c, t_1\}$. The L^2 cost function to minimise generalises expression (1.10) in Appendix A to make explicit the beginning time t_1 of the time interval in the optimisation problem. In order to make different windows $[t_1, t_2]$ comparable, we normalise the sum of squares of the residuals by the number $t_2 - t_1$ of points in the sum [124]

$$\chi^2(\mathbf{\Phi}) := F_2(m, \omega, t_c, t_1) = \frac{1}{t_2 - t_1} \sum_{i=t_1}^{t_2} r_i(\mathbf{\Phi})^2 \ , \quad \text{with } r_i(\mathbf{\Phi}) = y(t)_i - flppls(\mathbf{\Phi}, t)_i \ , \tag{2.1}$$

where $flppls(\mathbf{\Phi}, t)_i \equiv LPPLS(t)$ is defined by expression (1.6) in Chapter 1. The calibration of the parameters $\mathbf{\Phi}$ is performed by following the same procedure as explained in Subsection (1.2.3) of Chapter (1).

2.1.2 Hessian matrix, its eigenvalues and eigenvectors quantifying rigidity vs. sloppiness

Since we are interested in the *relative parameter variations* and their corresponding impact upon χ^2, each parameter $[m, \omega, t_c, t_1] \in \mathbf{\Phi}$ was standardized using

$$\tilde{\Phi}_i = \frac{\Phi_i - \langle \Phi_i \rangle}{\sigma(\Phi_i)}, \tag{2.2}$$

where $\langle \Phi_i \rangle$ and $\sigma(\Phi_i)$ denotes respectively the median and the standard deviation of parameter i within its respective theoretical bound (see conditions 1-3 in Sec. 1.2.1). This transformation ensures that the mean of $\tilde{\Phi}_i$ is 0 and its standard deviation is equal to 1 for all 4 nonlinear parameters. This allows us to compare their sloppiness [23] based on a study of the Hessian matrix given by Eq. (2.13) in Appendix A.

Once the optimal parameter set $\mathbf{\Phi^*}$ is obtained and after standardisation, we computed the Jacobian $J(\mathbf{\Phi^*})$ defined by (2.9) using Eq. (2.14) in Appendix A. For small residuals $\boldsymbol{r}$, the Hessian matrix about the best-fit is then approximated by Eq. (2.12) using $\boldsymbol{J}^\intercal \boldsymbol{J}$. Recall that the Hessian matrix provides a quantification of the shape of the cost function in parameter space. The spectrum of eigenvalues $\{\lambda_i\}$ and their associated eigenvectors $\{\vec{v}_i\}$ of the Hessian matrix embody the information on the directions in parameter space that are best determined. Large eigenvalues correspond to rigid parameters (or combination of parameters), i.e., for which a small variation entails a relatively large change of the cost function. In contrast, small eigenvalues identify the "sloppy" parameters (or combination of parameters) that are poorly constrained, so that a rather large variation of their value does not change much the cost function. Geometrically, one can picture the cost function as made of wide valleys with elongated flat profiles along the eigenvectors with small eigenvalues, and of sharp narrow valleys along the eigenvectors with large eigenvalues. If, as we find below, a given eigenvector is dominated by one of the parameters, its corresponding eigenvalue determines the rigidity vs. sloppiness of that parameter (rigid if the eigenvalue is relatively large and sloppy if small).

In practice, fits whose spread between the largest and the smallest eigenvalue surpasses roughly three orders of magnitude are considered *sloppy* [74]. Generally, this results from a near degenerate Hessian matrix (see Eq. (2.10)), where changes in a given parameter do not lead to an increase of the cost function since this change can be compensated by that of another parameter along the sloppy direction. It is important to keep in mind, however, that even though large parameter uncertainty does in general exist, the macroscopic behavior of the system is not necessarily compromised [9, 122], given that predictions rely on rigid directions of the model.

2.1.3 Construction of synthetic LPPLS bubbles

To gain insight about the parameter structure of the extended LPPLS model and thus establish a solid background to our empirical analysis, we generate synthetic price time series. The synthetic price time series (a realisation is depicted in figure 2.1(a)) are obtained by using formula (1.10) with parameters given by the best LPPLS fit within the window $w \in [t_1 = Jan.\ 1981\text{: } t_2 = Aug.\ 1987]$ of the bubble that ended with the Black Monday 19 Oct. 1987 crash. These parameters are ($m = 0.44$, $\omega = 6.5$, $C_1 = -0.0001$, $C_2 = 0.0005$, $A = 1.8259$, $B = -0.0094, t_c = 1194$). To the deterministic component describing the expected log-price given by expression (1.10) and denoted by $flpplss(\boldsymbol{\phi}, t)$, we add a stochastic element according to

$$\ln[P(t)] = flpplss(\phi, t)\left(\frac{1 + \varepsilon(t)\ \sigma}{max(lppls(t))}\right), \tag{2.3}$$

where $t = [1, \dots, N = 1100]$. This corresponds to multiplicative (or proportional) noise term with $\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$ and $\sigma = 0.1$, $\mu = 0$. The black stochastic line in figure 2.1(a) represent $\ln[P(t)]$ given by (2.3).

For each synthetic bubble price time series, we calibrated it with Eq. (1.10) by minimizing expression (2.1) in windows $[t_1, t_2]$, scanning t_2 from 1981/01/01 to $t_2 = 1987/08/12$, with t_1 varying from $t_1 = Jan.\ 1981$ up to 60 business days before t_2, i.e. up to $t_{1,max} = t_2 - 60$ for each fixed t_2. Then, the Hessian matrix was approximated about the best-fit parameters for each t_1 and its corresponding eigenvalues were calculated.

2.1.4 Sloppiness and Rigidity of t_c vs. t_1 using synthetic price time series

We now extend the initial sloppy-rigid analysis performed in the LPPLS framework by [19], in order to test which one of the two parameters, beginning t_1 or the forecasted end $\widehat{t_c}$ of a bubble, is the most sloppy, i.e., has the largest uncertainty.

For illustration of the typical situation found in these synthetic tests (to be extended below), we perform the calibration of synthetic noisy price time series in the full window shown in figure 2.1 corresponding to $[t_1 = Jan.\ 1981$ to $t_2 = Aug.\ 1987]$ (represented as $[Date = 1 : Date = 1100]$). For a given realisation, we calculate the Hessian matrix and the corresponding eigenvalues and eigenvectors. We consider both the case of a fixed t_1 with cost function (1.10) (bottom rows of table 2.1) and of the extended cost function (2.1) endogenizing t_1 as a parameter to be determined (top rows of table 2.1). The entries of the two Hessian matrices are given in the columns on the left of the table and their eigenvalues in the central column λ.

Note first that the smallest eigenvalue is of the order 10^{-8} times the largest eigenvalue, exemplifying the sloppy nature of the calibration. In the right columns of the table, the components of the corresponding eigenvectors show that the largest eigenvalue is mainly determined by parameter m, the second largest one is mainly controlled by parameter ω and the smallest eigenvalue is always mainly associated with parameter t_c, confirming that the end of the bubble is very difficult to determine as the cost function is essentially degenerate in the direction of t_c in parameter space.

Interestingly, for the extended cost function, one can observe that the eigenvalue dominated by parameter t_1, while being smaller than the two others associated with m and ω, is approximately 4000 times larger that the eigenvalue describing the sloppiness of t_c. This is a first illustration that t_1 is much more rigidly determined (or depending on taste much less sloppy) than t_c. Indeed, using a geometrical intuitive interpretation, contour lines of the cost function form approximate ellipses whose axis lengths are inversely proportional to the square root of their corresponding eigenvalue λ (axis length $\simeq 1/\lambda^{\frac{1}{2}}$). Thus a factor 4000 translates into a size of the axis along t_c about 60 times larger than the axis along t_1. In other words, in this illustration, the uncertainty on t_1 is about 60 times smaller than on t_c.

Does this result hold for other time windows and in particular far from the end of the bubble? To investigate this question, we perform the same exercise of calibrating the LPPLS formula in windows $[t_1, t_2]$, varying t_1 from $Date = 1$ (about four years before the crash) to $Date = 1040$ days (close to the crash). The blue circles and red squares in Fig (2.1) display the obtained normalised eigenvalues λ_3 and λ_4 in logarithm scale, associated respectively to t_1 and t_c. By 'normalised', we mean that each eigenvalue obtained for a given calibration is divided by the largest one (i.e. that associated mainly with m). By repeating this process 100 times, confidence bounds for the eigenvalues for each t_1 can be determined and are depicted by the error bars. Figure 2.1 confirms that t_1 is always more rigid that t_c by far at all times t_2. It is particularly noteworthy that the situation for t_c does not improve in absolute terms or relative to t_1 even when getting closer and closer to the true end of the bubble: while the determination of the beginning t_1 of the bubble can be reasonably estimated, that of the end t_c remains much more elusive. One can also observe a rather stable behavior of these two eigenvalues when t_1 spans 1 to 600 days. Interestingly, the vast majority of fits performed during this time window correctly qualify the underlying time series as being in a bubble regime.

For Date $= t_1 > 600$, one can observe a fast drop of the eigenvalues associated with t_1 and t_c, which can be explained by the average negative curvature of the log-price associated with the first large log-periodic oscillation. This negative curvature confounds the information on the existence of the supposed super-exponential bubble and thus on the determination of both its beginning (t_1) and its end (t_c). When the average curvature of the log-price becomes positive again, one can observe a jump of the two eigenvalues upward, back to almost the same level as before the first jump down. This is followed by a further decrease of the eigenvalues as t_1 approaches too close to the true t_c.

The lower part of panel (a) of figure 2.1 checks that the two smallest eigenvalues λ_3 and λ_4 are indeed mostly representing respectively the directions along t_1 and t_c of the cost function. To quantify that this is the case, let us denote $\vec{v}\ (t_{1,\lambda_3})$ the component on t_1 of the third vector associated with the third largest eigenvalue λ_3. Similarly, let $\vec{v}\ (t_{c,\lambda_4})$ be the component on t_c of the fourth vector associated with the smallest eigenvalue λ_4. Analogously, $\vec{v}(\Phi_j, \lambda_i)$ is the component on parameter Φ_j of the eigenvector associated with the eigenvalue λ_i. Let us introduce the weights

$$C_{t_1,\lambda_3} = \frac{|\vec{v}\ (t_{1,\lambda_3})|}{\sum_{j=1}^{4} \sqrt{\vec{v}\ (\Phi_j, \lambda_3)^2}}, \qquad C_{t_c,\lambda_4} = \frac{|\vec{v}\ (t_{c,\lambda_4})|}{\sum_{j=1}^{4} \sqrt{\vec{v}\ (\Phi_j, \lambda_4)^2}}\ . \tag{2.4}$$

These weights C_{t_1,λ_3} and C_{t_c,λ_4} are shown in the lower part of panel (a) of figure 2.1 and confirm that the relative contributions of parameter t_1 (resp. t_c) in the eigenvector along

λ_3 (resp. λ_4) is never smaller than $\approx 90\%$ (respec. 99%).

The distributions of the normalized eigenvalues λ_2/λ_1, λ_3/λ_1 and λ_4/λ_1 over the ensemble of t_1 values for a fixed t_2 and a single log-price realisation are depicted in figure 2.1(b). The dashed lines show the mean values of the three distributions. This confirms that t_1 is much more rigid than t_c across all windows and that the hierarchy from rigid to sloppy is from m, ω, t_1 to t_c.

For the fixed Date $t_2 = 150$, figures 2.1(c)-2.1(h) show cross-sections of the cost function for parameters m, ω, t_1 and t_c. The shape of the cost function exhibits the valley patterns whose relative extensions in different directions are quantified by the eigenvalues, as discussed above. Figs. 2.1(c) and 2.1(d) show that, for parameters m *vs.* ω and m *vs.* t_c, the elliptic contour lines close to the cost function minimum are aligned approximately along the parameter axes. In contrast, for the parameter space of t_c *vs.* ω, the largest eigen-direction is along the diagonal direction. This feature expresses the fact that parameters ω and t_c are strongly correlated in the calibration process. In practice, this implies the existence of several values of t_c that are consistent with a low cost value, given that parameter ω (see Fig 2.1(e)) can be tuned to take this variation into account. Thus, this interdependency should be considered properly when constructing confidence intervals for ω and t_c, as correctly pointed out by [19].

For robustness, a sensitive analysis was performed around the best solutions of $\widehat{t}_1^*$ and $\widehat{t}_c^*$. By taking into account the sample size, the box-plot shown in figure 2.1(a) gives the corresponding χ^2 variation when each $t_1 \in \widehat{t}_1^*$ and $t_c \in \widehat{t}_c^*$ are used as input parameters in Eq. 2.1. Results confirm that changes in t_1 leads to cost variations ranging from -4% to 4% while changes in $\widehat{t}_c^*$ yields a negligible change. This is due to the compensation provided by the correlation between ω and the critical time parameter (see figure 2.1(e)) previously mentioned.

Finally, the coloured $pdf's$ in Fig 2.1(a) were constructed over the qualified fits (according to the filtering criteria of Sec 1.2.1) of the ensemble of noise realisations for a fixed $\bar{t}_2$. The fact that the $pdf(\widehat{t}_1^*)$ is wider than $pdf(\widehat{t}_c^*)$ is not contradicting our key result that t_1 is much better estimated than t_c *for a specific realisation*. In a real life situation, one can only analyse one single realisation given by history. In contrast, the coloured $pdf's$ in Fig 2.1(a) provide artifactual information, i.e. on an ensemble of statistically equivalent price trajectories differentiated only by the different realisations of the noise process. The fact that the $pdf(\widehat{t}_1^*)$ is wider than $pdf(\widehat{t}_c^*)$ thus inform us that there is more variability of t_1 than t_c from one realisation to another one. This results from the structure of log-periodicity, with slow oscillations at early times and fast oscillations close to t_c.

2.2 Empirical tests

2.2.1 Data: eight historical bubbles

In this section, we perform the same procedure as described in the previous section on real bubble events. Scanning a time interval extending from several years before until the burst of each bubble, we first determine the time $\bar{t}_2$ at which the price reached its maximum before the crash starts to develop. We fix this time as the end of our time window of analysis. Obviously, this procedure is not correct for forecasts as it uses future information (the fact that $\bar{t}_2$ was the maximum before the crash) but is useful to reduce the number of degrees of freedom for our purposes. Further down, we drop this assumption and incorporate random $t_2's$ into the analysis (see 2.2.3). For this fixed $\bar{t}_2$, our fitting windows scan t_1 using daily observations ranging from some $t_{1,i=1}$ until $t_{1,\bar{t}_2-60}$, where the unit of time is one day, which are used as input for the t_1 parameter during the calibration process (2.1).

We have chosen a representative set of eight well-known bubbles, augmented by the synthetic bubbles given by (2.3) for comparison.

- **Shanghai Stock Exchange Composite Index bubble ending in October 2007 (SSEC):**

  This bubble has been documented and studied by [55]. For $\bar{t}_2 = 10\ Oct$ 2007, our fitting windows scan t_1 using daily observations ranging from $t_{1,i=1} = 2005/01/01$ until $t_{1,\bar{t}_2-60} = 2007/08/10$. Converting from calendar time to time counted in unit of days, we have $t_{1,i=1} = 1$ and $\bar{t}_2 = 676$. Thus, there are 616 different possible windows for the fixed end time $\bar{t}_2 = 676$, the different windows corresponding to the different $t_{1,1} = 1,\ t_{1,2} = 2,\ \ldots,\ t_{1,\bar{t}_2-60} = 616$, which are used as input for the t_1 parameter during the calibration process (2.1). Note that the smallest (resp. largest) window has a duration of 60 (resp. 676) days.

- **S&P500 Index bubble ending in October 2007 (SP):**

  This bubble has been documented and studied in [115, 104, 105]. The same procedure as described for the previous bubble was applied to the $S\&P500$. With the choice $\bar{t}_2 = Jul.15$ 2007, our fitting windows scan t_1 using daily observations ranging from $t_{1,i=1} = 2002/01/01$ to $t_{1,\bar{t}_2-60} = 2007/05/01$. This yields a total of $N = 1332$ different possible values of t_1 over which to perform the calibration process (2.1).

- **S&P500 Index bubble ending on Black Monday, October 19, 1987 (BM):**

  This bubble has been documented and studied by [107]. Choosing $\bar{t}_2 = Aug.15$ 1987, our fitting windows scan t_1 using daily observations ranging from $t_{1,i=1} = 1981/01/01$ until $t_{1,\bar{t}_2-60} = 1987/06/12$. This yields a total of $N = 1674$ different possible values of t_1 over which to perform the calibration process (2.1).

- **Bovespa Index bubble ending in December 2003:**

  Choosing $\bar{t}_2 = Dec.15$ 2003, our fitting windows scan t_1 using daily observations ranging from $t_{1,i=1} = 1998/06/22$ until $t_{1,\bar{t}_2-60} = 2003/10/14$. This yields a total of $N = 1940$ different possible values of t_1 over which to perform the calibration process (2.1).

- **S&P500 Index short-term bubble ending in February 1997:**

  With the choice $\bar{t}_2 = Feb.13$ 1997, our fitting windows scan t_1 using daily observations ranging from $t_{1,i=1} = 1991/08/24$ to $t_{1,\bar{t}_2-60} = 1997/05/01$. This yields a total of $N = 1940$ different possible values of t_1 over which to perform the calibration process (2.1).

- **Shanghai Stock Exchange Composite Index bubble ending in June 2015:**

  A detailed real-time diagnosis and post-mortem analysis of the Shanghai 2015 bubble and subsequent crash can be found in [112]. With $\bar{t}_2 = Jun.10$ 2015, our fitting window scan t_1 using daily observations ranging from $t_{1,i=1} = 2009/12/18$ until $t_{1,\bar{t}_2-60} = 2015/04/11$ thus yielding a total of $N = 1940$ different possible values of t_1 over which to perform the calibration process (2.1).

- **NASDAQ (dot-com bubble) ending in February 2000**

  Searching for the beginning of this bubble at $\bar{t}_2 = Feb.20$ 2000, we scan t_1 using daily observations ranging from $t_{1,i=1} = 1994/09/01$ until $t_{1,\bar{t}_2-60} = 1999/12/24$ thus yielding a total of $N = 1940$ different possible values over which to perform the calibration process (2.1).

- **NIKKEI Index bubble ending in January 1999:**

  With the choice $\bar{t}_2 = Dec.10$ 1989, our fitting windows scan t_1 using daily observations ranging from $t_{1,i=1} = 1984/06/19$ to $t_{1,\bar{t}_2-60} = 1989/10/11$. This yields a total of $N = 1940$ different possible values of t_1 over which to perform the calibration process (2.1).

For each of these eight empirical time series, the corresponding residuals $r(t)$'s defined in (2.1) are obtained for each window $[t_1, \bar{t}_2]$. Except from moments where fits are not qualified, the residuals are well-behaved and are approximately normally distributed, as one can see via the normality testes given in Table (2.5).

2.2.2 Analysis of the cost function $\chi^2(\Phi)$ and its Hessian

For the eight empirical bubbles, figure 2.2 presents the pdf's of the normalized eigenvalues $\lambda_2/\lambda_1, \lambda_3/\lambda_1$ and λ_4/λ_1 of the Hessian matrix $H|^*_{(\mathbf{\Phi})}$, over the ensemble of t_1 values scanned in the analysis. As in table 2.1 obtained for a synthetic bubble, we find almost systematically that the largest eigenvalue is dominated by parameter m. The colors of the pdf's encode the parameter that dominates its corresponding eigenvalue: green for m, grey for ω, blue for t_1 and red for t_c. One can observe that the normalized λ_3 (blue pdf) associated predominantly with t_1 are systematically much larger than the values of the normalized λ_4 (red pdf) associated predominantly with t_c. This confirms for these eight empirical bubbles the conclusion obtained in the synthetic tests. Moreover, we find that more than 80% of the windows give a normalized $\lambda_2 > 10^{-3}$, allowing us to conclude that ω (and a fortiori m) is a relatively rigid parameter.

Overall, the eigenvalues $\lambda_i^E \equiv \lambda_i$, for $i = [m, \omega, t_1, t_c]$ exhibit a mixture of jumps intercalated with intervals of stable downward trends as the window length shrinks and t_1 approaches t_2. When using the original cost function (1.10), the same pattern for the contribution of the parameters to the eigenvectors is observed as well as the same behaviour of the eigenvalues as a function of t_1. The overall sloppiness does not change substantially when using the extended cost function F_2 (expression (2.1)) instead of F_1 (see definition (1.12)).

Figure 2.4 displays three two-dimensional cross-sections of the cost function $\chi^2(\mathbf{\Phi}^*)$

for the eight empirical bubbles along the three different pairs of structural parameters (m, ω), (t_c, m) and (t_c, ω). The corresponding values of the Hessian matrix, eigenvectors and eigenvalues are summarised in Table 2.3. Under the same conditions, Table 2.4 gives the values of the Hessian matrix, of the eigenvalues λ and associated eigenvectors $\vec{v}$ using the original cost function (1.12) without t_1. Applying the standard unit-root Augmented Dickey-Fuller and Phillips-Perron tests to the residuals of the best fits confirms that they are mean-reversing, in agreement with the condition proposed by [70] (see Table 2.5). Figure 2.4 shows the same ellipse-shaped structure as obtained previously in our synthetic tests described in section 2.1.3 and the same order of the importance of the parameters m, ω, t_1 and t_c associated respectively from the largest eigenvalue to the smallest one. This structure is also robust when using the LPPLS cost function without t_1 as a parameter to calibrate (see Table 2.4).

The dependence of the cost function $\chi^2(\Phi)$ for the same eight studied empirical bubbles along the three different pairs of parameters (t_1, m), (t_1, ω) and (t_1, t_c) is shown in figure 2.5. One can observe a much more complicated landscape than in figure 2.4 with multiple local minima associated with the introduction of the parameter t_1. Figure 2.5 demonstrates that the normalised cost function exhibits an overall decrease as t_1 increases, because the calibration has a smaller number of degrees of freedom to explain for smaller $t_2 - t_1$. This tends to bias the estimation of t_1 upward, i.e., to underestimate the true duration of the bubble. Figures 2.5 and 2.6 exemplify that t_c has very little impact on the value of $\chi^2(\Phi)$ once t_1 is given. In other words, one can really visualise here the sloppiness of t_c compared with the relatively much larger rigidity of t_1. This feature is absent for the pairs (t_1, m) and (t_1, ω), as small changes in m and ω often lead to significant variations of the cost function, in agreement with the information contained in the corresponding eigenvalues. One can also observe that, for certain values of t_1, the cost function exhibits clear minima as a function of t_c, supporting previous claims that the end of financial bubbles may be predictable only at certain time intervals, i.e., in "pockets of predictability" [105].

2.2.3 Visualisation of the relative rigidity of t_1 vs. the slopiness of t_c using their pdf's

The stability of t_1 relative to t_c for the the eight empirical bubbles described in section 2.2.1 is visualised in figure 2.7. For a given $\bar{t}_2$ (black vertical dashed line in each panel), the blue filled function represents the *pdf* of the qualified calibrated $\widehat{t}_1^*$'s in a search over the grey interval up to $\bar{t}_2 \; - \; 30$ days. The red filled function represents the corresponding pdf of qualified calibrated $\widehat{t}_c^*$'s. It is pleasant to observe that the pdf of $\widehat{t}_1^*$ is where one would visually locate a priori the start of the bubble, as this is the time when the price starts to present evidence of a faster-than-exponential growth. The pdf of $\widehat{t}_c^*$ is not too far from the true time of the change of regime but is often too late, except for the bubble on the Ibovespa index.

The sensitivity analysis performed around the best solution of $\widehat{t}_1^*$ and $\widehat{t}_c^*$ demonstrates that t_1 is even more important than t_c for the calibration, relative to the study performed on synthetic data. Specifically, changes in t_1 lead to fluctuations of the normalised sum of squared residuals spanning the interval (-2.0% to 2.0%), (0.0% to 1.0%), (-0.5% to 1.0%), (-2.0% to 2.0%), (0.0% to 1.75%), (0.0% to 2.5%), (0.0% to 2.0%), (0.0% to 3.0%), respectively for the SSEC, SP, BM and IBovespa, SP-1997, SSEC-2015, NASDAQ and NIKKEI bubbles. In contrast, changes in t_c have negligible impact on the cost function.

Even accounting for their errors bars, one can observe that λ_3 is at least two orders of magnitude larger that λ_4 for all t_1's, corroborating the results of Sec. 2.1.4. The parameters t_1 and t_c contribute respectively to λ_3 and λ_4 more than 35% throughout the analysed periods. Taken together, these features strengthen the evidence of the greater sloppiness of t_c compared to t_1, which makes the determination of the end of bubbles much more difficult that their beginning.

Finally, it is important to check how the above results generalise for different "present" times t_2, mimicking a real time situation of a developing bubble. As seen in figure 2.8, we consider four different values of t_2 shown by the vertical dashed black lines in each panel. Note that these four choices cover most of the duration of the bubbles. For each t_2, we search the optimal t_1 that minimises the cost function up to 600 days prior t_2. This intervals in which t_1 is scanned are represented by the different shades of grey, one per value of t_2.

One can observe that the pdf's of $\widehat{t_c^*}$ are quite wide and with several modes, except for the bubble on the S&P500 that burst in 2007. These modes are localised at times when the markets corrected, in addition to finding a neighbourhood of the true ends of the bubbles. In contrast, the pdf's of $\widehat{t_1^*}$ are monomodal, very narrow and pinpoint a time when the markets start their significant ascent. For instance, for the $S\&P500$ Index during the 1980's, our analysis identifies two possible modes for the end of the bubble, one occurring in the last quarter of 1985 associated with a significant drawdown and the other being close to the crash on Oct. 19, 1987. Conversely, the narrow pdf of $\widehat{t_1^*}$ identifies the beginning of the bubble around mid-1984. A similar situation is observed for the SSEC Index as well as for the synthetic case.

The existence of several modes of the pdf's of $\widehat{t_c^*}$ can be traced to the discrete scale invariance of the log-periodic oscillations [102], associated with the the occurrence of corrections or plateaux decorating the super-exponential growth, each of them being interpreted as a possible candidate for the end of the bubble by the calibration procedure. When used to diagnose bubbles, this confirms that the LPPLS model can determine t_1 significantly more accurately than t_c.

2.3 Conclusion

We have presented systematic tests of the precision and reliability with which the beginning and end of a bubble can be determined. This has required using a specific bubble model, the log-periodic power law singularity (LPPLS) model, which represents a bubble as a transient noisy super-exponential price trajectory decorated by accelerated volatility oscillations. One of the quality of the LPPLS model is to contain the end time t_c of the bubble as a defining parameter, which can thus be estimated over various time windows. In order to estimate the beginning of a bubble, we have proposed to endogenise in the cost function the parameter t_1 defining the beginning of the time window that provides the best goodness of fit and satisfy the LPPLS model conditions. The cost function quantifying the quality of fit of the LPPLS model has been extended to include t_1. Using the Fisher Information matrix, we have quantified the parameter uncertainty associated with the determination of both t_1 and t_c. Using both synthetic data and eight historical bubble cases, we find overwhelming evidence that the beginning of bubbles is much better constrained that their end. This is quantified by calculating the eigenvalues of the Hessian matrix, which characterise the shape of the cost function in the different directions in parameter space. Parameters associated with large eigenvalues are "rigid", i.e. they tend to be well estimated. In contrast, parameters associated with small eigenvalues are "sloppy", as large changes of their values do not impact the cost function, which is degenerate along their direction in parameter space. We find that the eigenvalues for which t_1 contributes most are several order of magnitudes larger than the eigenvalues for which t_c contributes most. Practically, this implies that the beginning of financial bubbles are comparatively much easier to determine using LPPLS than their ending time t_c. Our results are robust over all eight empirical bubbles and many synthetic tests, as well as when changing the time t_2 (the "present") of analysis.

Building on the initial sloppy-rigid analysis of the LPPLS model performed by [19], our results corroborate and extend those obtained by [70] when studying the beginning of Black-Monday bubble of 1987. Our analysis gives support to the empirical evidence among practitioners (in particular hedge-funds [49]), who tend to correctly diagnose on-going bubbles but in general fail to time their end.

As a side result, we have found that two structural parameters of the LPPLS model, the exponent m controlling the super-exponential growth of price and the angular log-periodic frequency ω describing the log-periodic acceleration of volatility, are very rigid. This suggests that the LPPLS model is a reasonable candidate for describing the generating process of prices during bubbles [110].

Finally, contrary to the claim in the literature that bubbles are not suppressed by arbitrageurs because they fail to agree on the beginning of the bubble [1], our findings suggests that bubbles persist due to the difficulty of synchronizing on the end of bubbles.

Acknowledgments

This work was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). The authors would like to acknowledge V. Filimonov. All calculations and figures were elaborated using open source software IPython [83]. The authors would like to thank the editor and two anonymous referees.

Appendix A: calibration of the extended set of 4 nonlinear parameters $\mathbf{\Phi} = \{m, \omega, t_c, t_1\}$

We compute the first-order partial derivatives of Eq. (2.1) - i.e. the Gradient vector $\vec{\bigtriangledown}\chi^2(\mathbf{\Phi})$ - with respect to each parameter as follows:

$$\begin{aligned}\frac{\partial\chi^2(\Phi)}{\partial m} =& \frac{2\ln(t_c-t)(t_c-t)^m\,[\,\sin\left(\ln(t_c-t)\,\omega\right)+\cos(\ln(t_c-t)\omega)-B\,]}{t_2-t_1}\\ &\times\frac{[\,(t_c-t)^m\ \sin\left(\ln(t_c-t)\,\omega+\cos(\ln(t_c-t)\omega)-B\right)}{t_2-t_1}\\ &+\frac{C_1+C_2+A\,]}{t_2-t_1}\end{aligned} \tag{2.5}$$

$$\begin{aligned}\frac{\partial\chi^2(\Phi)}{\partial\omega} =& \frac{-2\ln(t_c-t)(t_c-t)^m\,[\,\sin\left(\ln(t_c-t)\,\omega-\cos(\ln(t_c-t)\omega)\,]\right)}{t_2-t_1}\\ &\times\frac{[\,(t_c-t)^m\ \sin\left(\ln(t_c-t)\,\omega+\cos(\ln(t_c-t)\omega)-B\right)}{t_2-t_1}\\ &+\frac{C_1+C_2+A\,]}{t_2-t_1}\end{aligned} \tag{2.6}$$

$$\begin{aligned}\frac{\partial\chi^2(\Phi)}{\partial t_c} =& \frac{-2(t_c-t)^{m-1}\,[\,(C_1\omega-C_2m)\ \cos\left(\omega\ln(t_c-t)\right)+Bm)(t_c-t)^m}{t_2-t_1}\\ &\times\frac{(C_2\sin(\omega\ln(t_c-t))+C_1\cos(\omega\ln(t_c-t))}{t_2-t_1}\\ &+\frac{-B)+A]}{t_2-t_1}\end{aligned} \tag{2.7}$$

$$\begin{aligned}\frac{\partial\chi^2(\Phi)}{\partial t_1} =& \frac{(-y+C_2(t_c-t)^m\ \sin(\ln(t_c-t)\omega)\ +\ C_1(t_c-t)^m\ \cos\left(\ln(t_c-t)\omega\right)}{(t_2-t_1)^2}\\ &+\frac{B(t_c-t)^m+A)^2}{(t_2-t_1)^2}.\end{aligned} \tag{2.8}$$

These first-order partial derivatives are collected by the Jacobian matrix $J(\mathbf{\Phi})_{(N\times m)}$,

$$J(\mathbf{\Phi}) = \begin{bmatrix} \frac{\partial r_i}{\partial \Phi_i} \\ \vdots \\ \frac{\partial r_i}{\partial \Phi_j} \end{bmatrix} \equiv \begin{bmatrix} \frac{\partial r_i}{\partial m} \; \frac{\partial r_i}{\partial \omega} \; \frac{\partial r_i}{\partial t_c} \; \frac{\partial r_i}{\partial t_1} \end{bmatrix}^{\intercal}_{\substack{j=1,2,\ldots,m; \\ i=1,2,\ldots,N;}}, \tag{2.9}$$

where $r_i(\mathbf{\Phi}) = y(t)_i - flppls(\mathbf{\Phi}, t)_i$ are defined in Eq. (2.1).

Embedded within a m-dimensional Euclidean space, the Gradient vector points towards directions in which the cost increases more rapidly. At a given local minimum, the minimisation problem should not only display small $\vec{\bigtriangledown}\chi^2(\mathbf{\Phi})$ values but also, the cost curvature should be convex and the residuals approximately zero. Formally, this idea is expressed as

$$\Delta\chi^2(\mathbf{\Phi}) = \frac{1}{2}\vec{\bigtriangledown}(\mathbf{\Phi})^{\intercal}\boldsymbol{H}(\mathbf{\Phi}^*)\vec{\bigtriangledown}(\mathbf{\Phi}) \geq 0, \tag{2.10}$$

where $\boldsymbol{H}(\mathbf{\Phi}^*)$ denotes the Hessian matrix expressed at the best-fit parameters. Bold letters denote either matrices or vectors. This special structure of the sum of squares of residuals allows one to approximate $\boldsymbol{H}(\mathbf{\Phi}^*)$ in terms of the fitting residuals using,

$$\frac{\partial\chi^2(\mathbf{\Phi})}{\partial\Phi_i\Phi_j} = \sum_{i=1}^{m}\left(r_i\frac{\partial^2 r_i(\Phi)}{\partial\Phi_i\partial\Phi_j} + \frac{\partial r_i(\Phi)}{\partial\Phi_i}\frac{\partial r_i(\Phi)}{\partial\Phi_j}\right). \tag{2.11}$$

Since Eq. (2.10) holds, we therefore skip the calculation of second-order derivatives and compute $\boldsymbol{H}(\mathbf{\Phi}^*)$ solely using the Jacobian [74, 124],

$$\boldsymbol{H}|_* \equiv \boldsymbol{J}^{\intercal}\boldsymbol{J} = \sum_{i=1}^{N}(\vec{\bigtriangledown}r_i(\mathbf{\Phi}^*))(\vec{\bigtriangledown}r_j(\mathbf{\Phi}^*))^{\intercal}. \tag{2.12}$$

At the best-fit, the Hessian is always *symmetric*, positive-definite and has entries according to

$$H_{i,j}^{\chi^2(\Phi^*)} = \begin{bmatrix} \frac{\partial^2 r_i(\Phi)}{\partial m, \partial m} & \frac{\partial^2 r_i(\Phi)}{\partial m, \partial \omega} & \frac{\partial^2 r_i(\Phi)}{\partial m, \partial t_c} & \frac{\partial^2 r_i(\Phi)}{\partial m, \partial t_1} \\ \frac{\partial^2 r_i(\Phi)}{\partial \omega, \partial m} & \frac{\partial^2 r_i(\Phi)}{\partial \omega, \partial \omega} & \frac{\partial^2 r_i(\Phi)}{\partial \omega, \partial t_c} & \frac{\partial^2 r_i(\Phi)}{\partial \omega, \partial t_1} \\ \frac{\partial^2 r_i(\Phi)}{\partial t_c, \partial m} & \frac{\partial^2 r_i(\Phi)}{\partial t_c, \partial \omega} & \frac{\partial^2 r_i(\Phi)}{\partial t_c, \partial t_c} & \frac{\partial^2 r_i(\Phi)}{\partial t_c, \partial t_1} \\ \frac{\partial^2 r_i(\Phi)}{\partial t_1, \partial m} & \frac{\partial^2 r_i(\Phi)}{\partial t_1, \partial \omega} & \frac{\partial^2 r_i(\Phi)}{\partial t_1, \partial t_c} & \frac{\partial^2 r_i(\Phi)}{\partial t_1, \partial t_1} \end{bmatrix}. \tag{2.13}$$

Derivatives are calculated using a *centered finite-difference scheme*,

$$\chi^2(\mathbf{\Phi}^*)' \approx \frac{\chi^2(\mathbf{\Phi}^* + h) - \chi^2(\mathbf{\Phi}^* - h)}{2h}, \tag{2.14}$$

with step-size h varying according to information provided by the Gradient. The choice of h should neither be too small or too large.

lppls synthetic

	Hessian, $\mathbf{H}(\Phi^*)$				λ/λ_{max}	Eigenvectors, $\vec{v}$			
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	9.20e-1		8		1.00e0	**-9.99e-1**	-4.40e-2	-9.73e-5	-4.37e-4
w	4.05e-2	3.70e-03			2.07e-3	-4.40e-02	**9.98e-1**	-2.05e-3	-9.22e-3
tc	8.98e-5	2.02e-08	5.12e-07		1.06e-4	8.43e-4	-9.19e-3	5.27e-4	**-9.99e-1**
t1	4.04e-4	1.07e-06	2.40e-08	9.87e-05	5.37e-7	1.88e-4	-2.05e-3	**-9.99e-1**	-5.08e-4

	Hessian, $\mathbf{H}(\phi^*)$			λ	Eigenvectors, $\vec{v}$		
	m	ω	t_c		m	ω	t_c
m	5.91e-2			1.00e0	**9.98e-1**	3.59e-2	1.69e-3
ω	-1.96e-3	4.71e-3		7.84e-2	3.60e-2	**9.99e-1**	1.54e-2
t_c	9.78e-5	-7.53e-5	1.41e-6	2.35e-6	1.13e-3	1.55e-2	**9.99e-1**

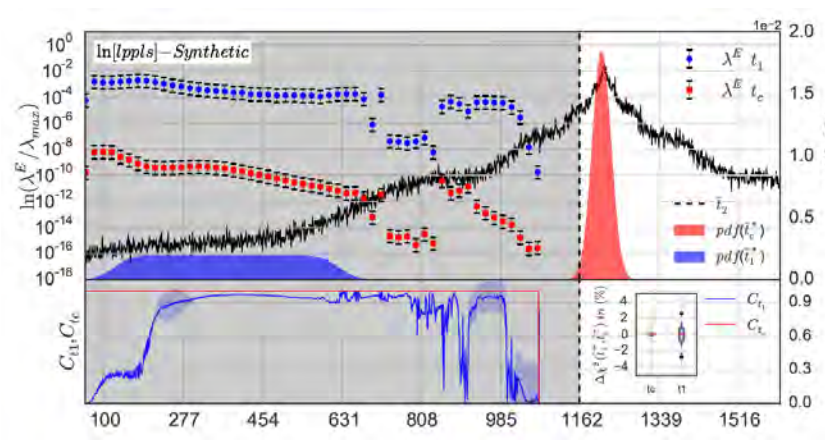
Table 2.1: Hessian matrix (left columns), eigenvalues λ (middle column) and corresponding eigenvectors $\vec{v}$ (right columns), for the best LPPLS fits of the synthetic price time series generated as explained in section 2.1.4 using $\chi^2(\phi)$ and $\chi^2(\Phi)$ in the full window shown in Fig. 2.1(a) corresponding to $[t_1 = Jan.\ 1981$ to $t_2 = Aug.\ 1987]$ (represented as $[Date = 1 : Date = 1100]$). The top rows correspond to the extended cost function (2.1) endogenizing t_1 as a parameter to be determined. The bottom rows correspond to the case of a fixed t_1 with cost function (1.10).

	Hessian, $\mathbf{H}(\Phi^*)$				λ/λ_{max}	Eigenvectors, $\vec{v}$			
					S&P 500(BM)				
					$t_1^* \approx Mar.\,1984$, $t^2 \approx Aug.\,1987$				
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	9.20e-1				1.00e0	**-9.99e-1**	-1.90e-2	-3.81e-4	-9.81e-3
ω	1.68e-2	3.70e-2			4.05e-2	-2.02e-2	**9.90e-1**	2.07e-3	1.36e-1
t_c	3.45e-4	8.20e-7	9.87e-4		4.90e-3	7.07e-3	1.35e-1	-1.64e-1	**-9.77e-1**
t_1	8.90e-3	4.63e-3	5.96e-4	5.12e-3	9.65e-4	-8.34e-4	-2.05e-2	**-9.86e-1**	1.63e-1
					S&P 500(SP)				
					$t_1^* \approx Dec.\,2003$, $t^2 \approx Jul.\,2007$				
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	9.20e-1				1.00e0	**-9.99e-1**	-2.12e-2	-6.61e-4	-1.21e-3
ω	1.320e-2	3.00e-1			3.26e-1	-2.13e-2	**9.99e-1**	-8.57e-6	3.18e-2
t_c	6.09e-4	7.85e-6	1.87e-5		2.31e-3	-5.59e-4	-3.18e-2	3.78e-2	**9.98e-1**
t_1	9.12e-4	9.50e-3	8.07e-5	2.43e-3	1.65e-5	-6.41e-4	1.20e-3	**9.99e-1**	-3.78e-2
					SSEC				
					$t_1^* \approx Feb.\,2005$, $t^2 \approx Oct.\,2007$				
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	5.22e-1				1.00e0	**-9.99e-1**	-1.26e-2	-3.75e-5	-1.74e-3
ω	6.32e-3	2.28e-2			4.35e-2	-1.26e-2	**9.99e-1**	3.01e-4	2.33e-2
t_c	1.95e-5	7.03e-6	9.81e-5		3.67e-3	1.44e-3	2.34e-2	-9.82e-4	**-9.99e-1**
t_1	9.02e-4	4.98e-4	1.98e-6	1.93e-3	1.87e-4	3.22e-5	2.78e-04	**-9.99e-1**	9.89e-4
					IBovespa				
					$t_1^* \approx Jun.\,2003$, $t^2 \approx Jan.\,2004$				
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	2.60e-1				1.00e0	**-9.99e-1**	-4.00e-3	-3.84e-6	-3.86e-5
ω	1.01e-3	2.03e-02			3.95e-2	-4.00e-3	**9.99e-1**	2.30e-7	2.12e-4
t_c	2.09e-6	6.16e-9	9.82e-6		3.77e-3	3.77e-5	2.12e-4	-1.02e-3	**-9.99e-1**
t_1	1.12e-05	2.02e-6	1.10e-6	9.82e-4	3.77e-5	3.80e-6	2.67e-8	**-9.99e-1**	1.02e-3

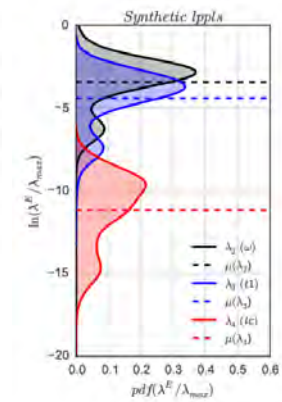
Table 2.2: Hessian matrix, eigenvalues and corresponding eigenvectors for the best LPPLS fits using Eq (2.1) of the financial bubbles described in section 2.2.1 in the time windows $[t_1^* : t_2]$ given in the table for each bubble. The numbers in boldface indicate the parameters that contribute the most to their eigenvectors.

	Hessian, $\mathbf{H}(\Phi^*)$				λ/λ_{max}	**Eigenvectors, $\vec{v}$**			
$S\&P\,500(1997)$, $t_1^* \approx Jul.\,1996$, $t^2 \approx Feb.\,1997$									
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	1.98e-1				1.00e0	**-9.66e-1**	-2.58e-1	-6.15e-5	-7.89e-3
ω	3.85e-2	6.40e-2			2.60e-1	-2.57e-1	**9.61e-1**	4.34e-4	1.00e-1
t_c	9.31e-6	9.77e-6	1.03e-4		2.05e-2	1.82e-2	-9.88e-2	3.91e-2	**9.94e-1**
t_1	2.65e-4	5.23e-3	1.65e-4	4.78e-3	4.63e-4	-6.63e-4	3.43e-3	**9.99e-1**	-3.90e-2
$SSEC\,(2015)$, $t_1^* \approx Mar.\,2014$, $t^2 \approx Jun.\,2015$									
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	6.55e-1				1.00e0	**-9.80e-1**	-1.98e-1	-3.94e-3	-7.08e-4
ω	1.25e-1	6.40e-2			5.68e-2	-1.98e-1	**9.79e-1**	-1.38e-2	3.59e-2
t_c	2.74e-3	8.25e-6	7.53e-5		5.73e-3	6.39e-3	-3.5e-2	7.76e-3	**9.99e-1**
t_1	2.21e-4	1.32e-3	1.25e-5	3.95e-3	8.38e-5	6.67e-3	-1.30e-2	**-9.99e-1**	7.26e-3
$NASDAQ$, $t_1^* \approx Jan.\,1996$, $t^2 \approx Feb.\,2000$									
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	9.07e-1				1.00e0	**-9.93e-1**	-1.12e-1	-8.27e-4	-1.01e-2
ω	1.01e-1	2.55e-2			1.53e-2	-1.12e-1	**9.92e-01**	-6.00e-03	5.01e-2
t_c	7.65e-4	2.05e-7	4.01e-6		7.56e-4	4.52e-3	5.07e-2	-3.45e-2	**-9.98e-1**
t_1	9.22e-3	1.72e-3	2.74e-5	8.23e-4	2.22e-6	1.34e-3	-7.62e-3	**-9.99e-1**	3.42e-02
$NIKKEI$, $t_1^* \approx Jan.\,1988$, $t^2 \approx Dec.\,1989$									
	m	ω	t_c	t_1		m	ω	t_c	t_1
m	9.20e-1				1.00e0	**-9.99e-1**	-4.40e-2	-9.73e-5	-4.38e-4
ω	4.05e-2	3.70e-3			2.07e-3	-4.40e-2	**9.98e-1**	-2.05e-3	-1.80e-2
t_c	8.98e-5	2.02e-8	5.12e-7		1.07e-3	1.23e-3	-1.80e-2	8.76e-5	**-9.99e-1**
t_1	4.04e-4	1.07e-6	2.40e-8	9.87e-4	5.37e-7	1.87e-4	-2.05e-3	**-9.99e-1**	-5.04e-5

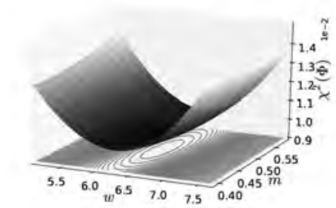
Table 2.3: Hessian matrix, eigenvalues and corresponding eigenvectors for the best LPPLS fits using Eq (2.1) of the financial bubbles described in section 2.2.1 in the time windows $[t_1^* : t_2]$ given in the table for each bubble. The numbers in boldface indicate the parameters that contribute the most to their eigenvectors.



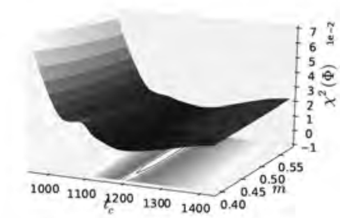
(a) Synthetic LPPLS: λ_{t1} vs. λ_{tc}



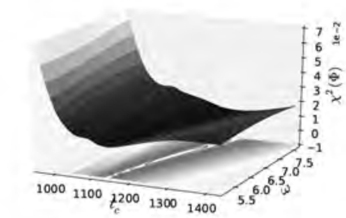
(b) $pdf(\lambda_2, \lambda_3, \lambda_4)$



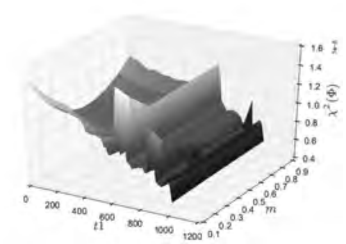
(c) $\chi^2(t_c = \text{'}1194\text{'}, m, \omega, t1 = 150)$



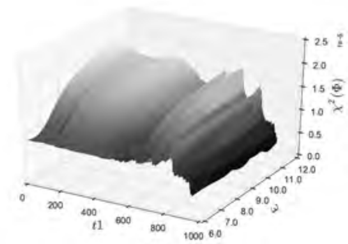
(d) $\chi^2(t_c, m, \omega \approx 6.3, t1 = 150)$



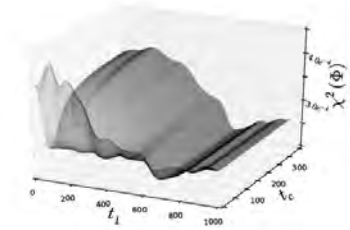
(e) $\chi^2(t_c, m \approx 0.44, \omega, t1 = 150)$



(f) $\chi^2(t_c = \text{'}1194\text{'}, m, \omega \approx 6.3, t_1)$



(g) $\chi^2(t_c = \text{'}1194\text{'}, m \approx 0.44, \omega, t_1)$



(h) $\chi^2(t_c, m \approx 0.44, \omega \approx 6.3, t_1)$

Figure 2.1: Quantification of the relative sloppiness of t_c and t_1 in synthetic data: (panel 2.1(a)): a synthetic noisy LPPLS price time series generated as described in the main text is shown with the black continuous line. Here, the *x-axis* denote values of the beginning t_1 of the window in which the calibration is performed at a fixed $\bar{t}_2$. The blue circles and red squares show the normalised eigenvalues (λ) of the Hessian matrix $H(\Phi^*)$ estimated at the best-fit, which correspond mainly to the directions t_1 and t_c in parameter space, (See Fig (2.7) for precise quantification). Errors bars for each parameter represent $\lambda_i^E \pm \sigma(\lambda_i^E)$ over 100 realisations (corresponding to 100 generations of the noise). The relative contribution (see Eq. 2.4) of t_1 and t_c at eigenvalues λ_3 and λ_4 is depicted using coloured lines within the grey-shaded area. Coloured *pdf*'s denote qualified values of t_1 and t_c at $\bar{t}_2$. The inset box-plot shows the normalised cost change in % terms when t_1 (respectively t_c) are perturbed over the ensemble of $\widehat{t}_1^*$ (respectively $\widehat{t}_c^*$). Panel 2.1(b) shows the distributions over the ensemble of t_1 values of the Hessian eigenvalues for the aggregated LPPLS fits performed in the synthetic data. From top to bottom, the *y-axis* display the hierarchy (rigid to sloppy) of parameters that govern the model output. Panels 2.1(c)-2.1(h): cross-sections of the cost function $\chi^2(\Phi)$ at the best-fit window ($w \equiv ||t1^* : t2^*||$, i.e. for the window starting at the fitted t_1 for the fixed $t_2 = 150$) in the planes (m, ω), (t_c, ω), (t_c, m), (t_1, m), (t_c, ω) and (t_1, t_c) respectively.

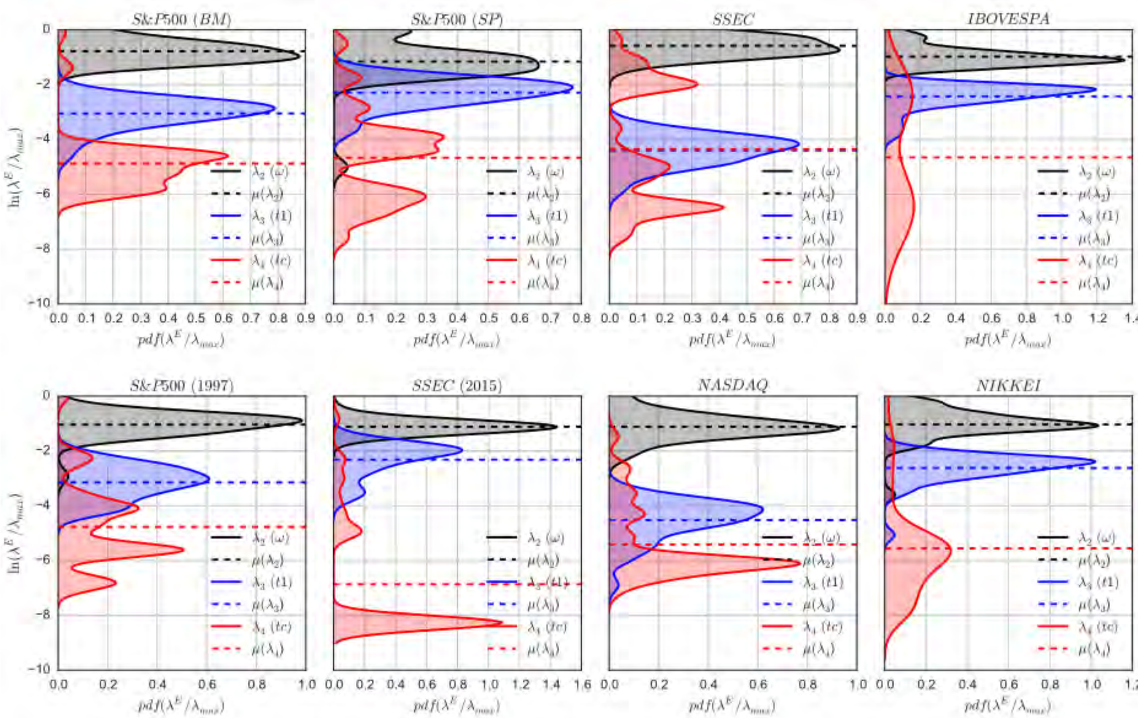


Figure 2.2: Probability density functions (pdf) of the normalized eigenvalues $\lambda_2/\lambda_1, \lambda_3/\lambda_1$ and λ_4/λ_1 over the ensemble of windows obtained by scanning t_1 for the eight empirical bubbles described in section 2.2.1. The horizontal dashed lines indicate the mean values of the distributions. The colors of the *pdf*'s encode the parameter that dominates its corresponding eigenvalue: green for m ($\lambda/\lambda_{max} = 0 = 1.00e0$), grey for ω, blue for t_1 and red for t_c.

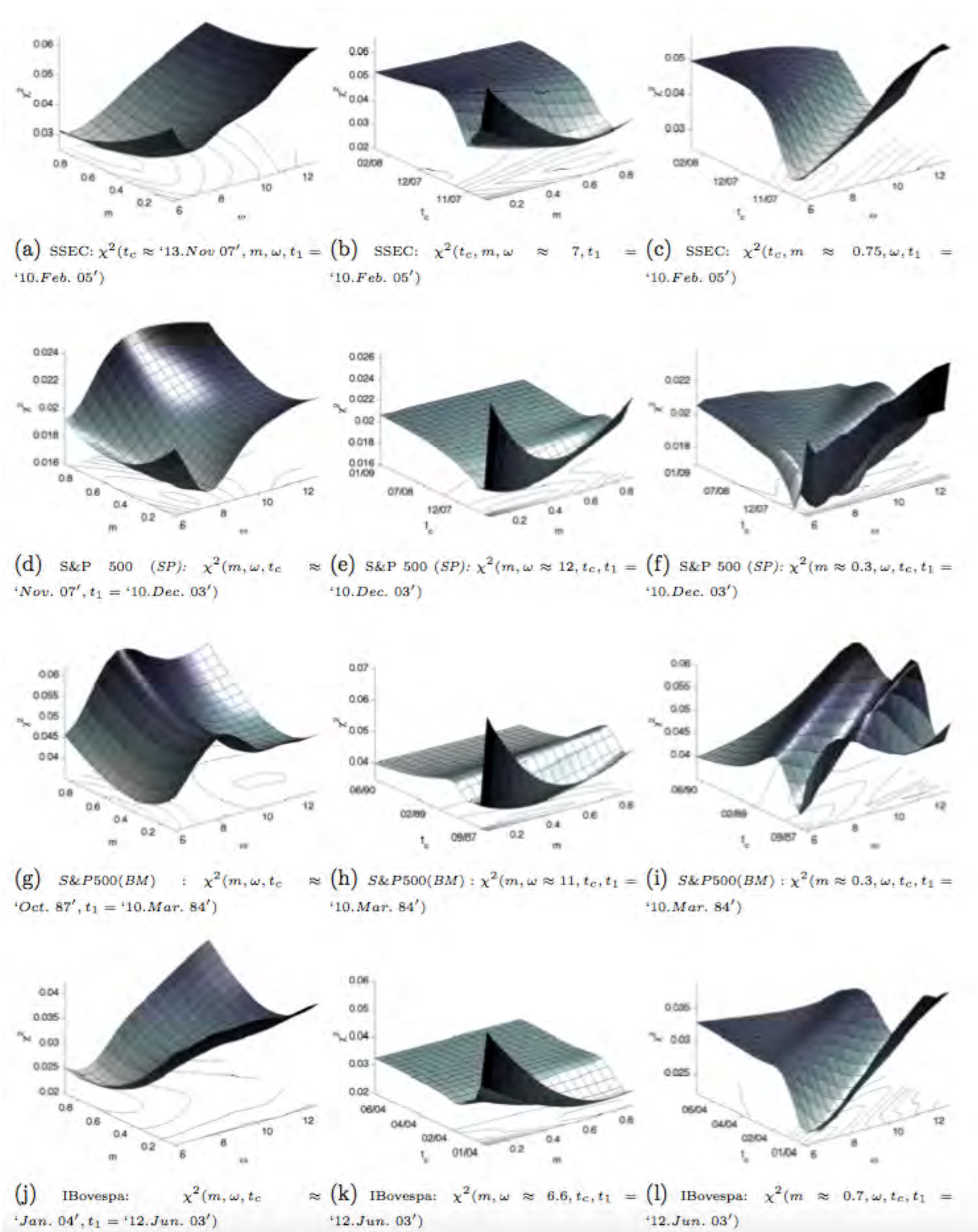
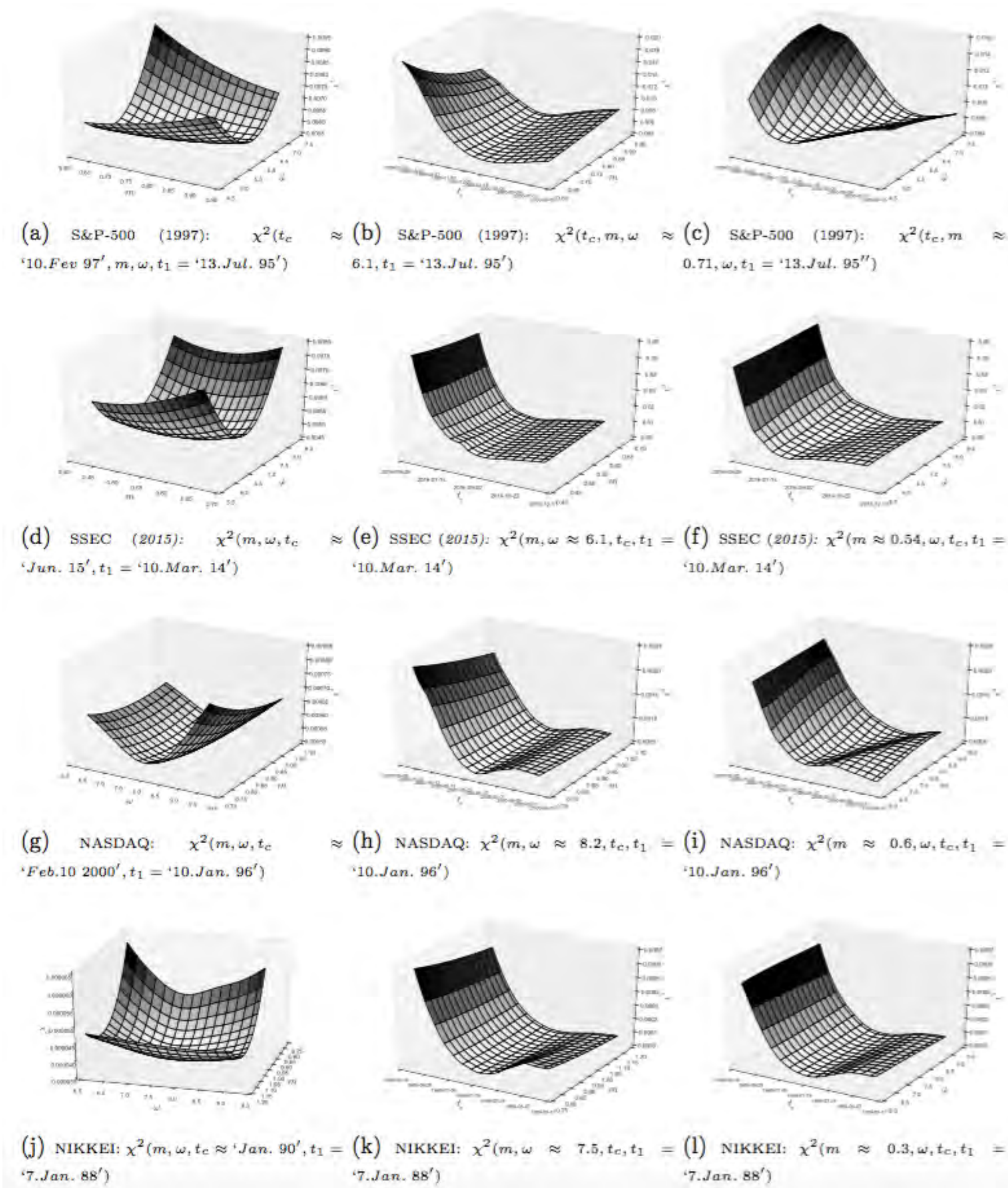


Figure 2.3: Cross-sections of the cost function $\chi^2(\Phi)$ for the eight studied empirical bubbles described in section 2.2.1 along the three different pairs of structural parameters (m, ω), (t_c, m) and (t_c, ω). Small (resp. large) values for the cost function correspond to lighter (resp. darker) colours and depict parameter domains giving better fits.



(a) S&P-500 (1997): $\chi^2(t_c \approx$ '$10.Fev\ 97'$, $m, \omega, t_1 =$ '$13.Jul.\ 95'$)

(b) S&P-500 (1997): $\chi^2(t_c, m, \omega \approx 6.1, t_1 =$ '$13.Jul.\ 95'$)

(c) S&P-500 (1997): $\chi^2(t_c, m \approx 0.71, \omega, t_1 =$ '$13.Jul.\ 95''$)

(d) SSEC (*2015*): $\chi^2(m, \omega, t_c \approx$ '$Jun.\ 15'$, $t_1 =$ '$10.Mar.\ 14'$)

(e) SSEC (*2015*): $\chi^2(m, \omega \approx 6.1, t_c, t_1 =$ '$10.Mar.\ 14'$)

(f) SSEC (*2015*): $\chi^2(m \approx 0.54, \omega, t_c, t_1 =$ '$10.Mar.\ 14'$)

(g) NASDAQ: $\chi^2(m, \omega, t_c \approx$ '$Feb.10\ 2000'$, $t_1 =$ '$10.Jan.\ 96'$)

(h) NASDAQ: $\chi^2(m, \omega \approx 8.2, t_c, t_1 =$ '$10.Jan.\ 96'$)

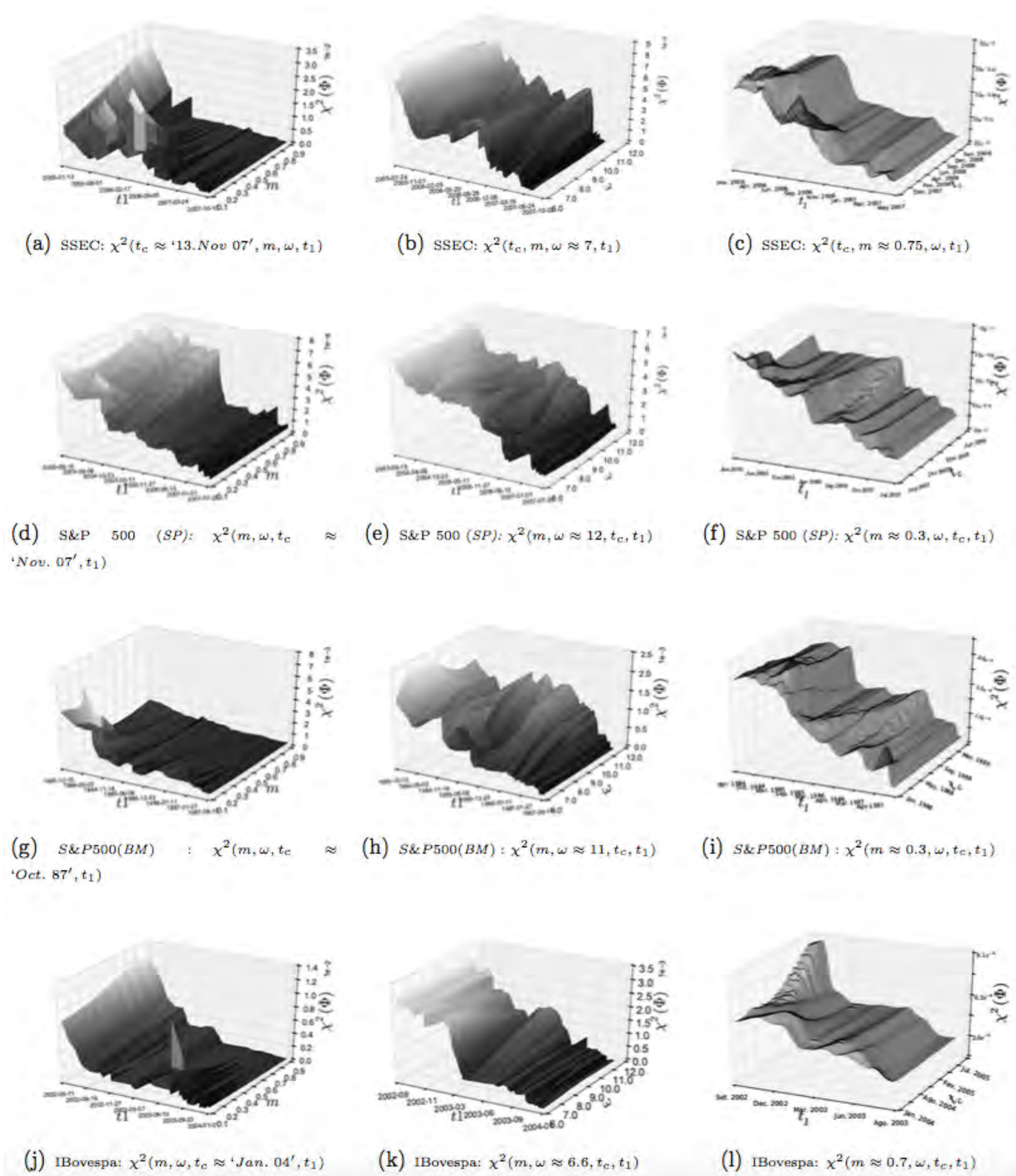
(i) NASDAQ: $\chi^2(m \approx 0.6, \omega, t_c, t_1 =$ '$10.Jan.\ 96'$)

(j) NIKKEI: $\chi^2(m, \omega, t_c \approx$ '$Jan.\ 90'$, $t_1 =$ '$7.Jan.\ 88'$)

(k) NIKKEI: $\chi^2(m, \omega \approx 7.5, t_c, t_1 =$ '$7.Jan.\ 88'$)

(l) NIKKEI: $\chi^2(m \approx 0.3, \omega, t_c, t_1 =$ '$7.Jan.\ 88'$)

Figure 2.4: Cross-sections of the cost function $\chi^2(\Phi)$ for the $S\&P500$, SSEC, NASDAQ and NIKKEI Indexes empirical bubbles described in section 2.2.1 along the three different pairs of structural parameters (m, ω), (t_c, m) and (t_c, ω). Small (resp. large) values for the cost function correspond to lighter (resp. darker) colours and depict parameter domains giving better fits.



(a) SSEC: $\chi^2(t_c \approx \text{'13.Nov 07'}, m, \omega, t_1)$

(b) SSEC: $\chi^2(t_c, m, \omega \approx 7, t_1)$

(c) SSEC: $\chi^2(t_c, m \approx 0.75, \omega, t_1)$

(d) S&P 500 *(SP):* $\chi^2(m, \omega, t_c \approx \text{'Nov. 07'}, t_1)$

(e) S&P 500 *(SP):* $\chi^2(m, \omega \approx 12, t_c, t_1)$

(f) S&P 500 *(SP):* $\chi^2(m \approx 0.3, \omega, t_c, t_1)$

(g) *S&P500(BM)* : $\chi^2(m, \omega, t_c \approx \text{'Oct. 87'}, t_1)$

(h) *S&P500(BM)* : $\chi^2(m, \omega \approx 11, t_c, t_1)$

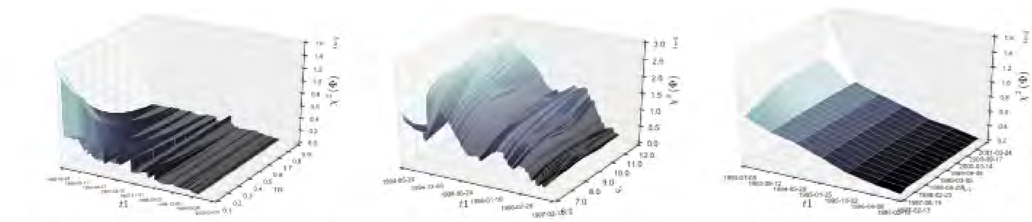
(i) *S&P500(BM)* : $\chi^2(m \approx 0.3, \omega, t_c, t_1)$

(j) IBovespa: $\chi^2(m, \omega, t_c \approx \text{'Jan. 04'}, t_1)$

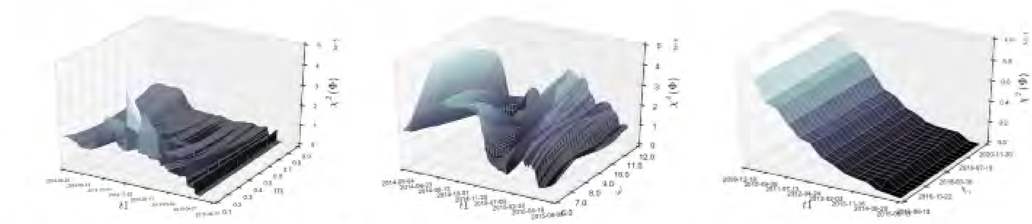
(k) IBovespa: $\chi^2(m, \omega \approx 6.6, t_c, t_1)$

(l) IBovespa: $\chi^2(m \approx 0.7, \omega, t_c, t_1)$

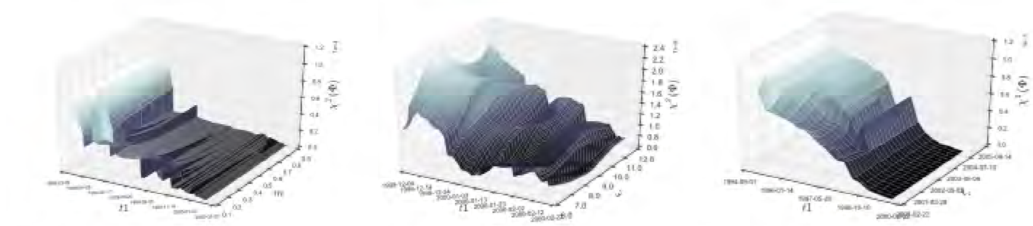
Figure 2.5: Cross-sections of the cost function $\chi^2(\Phi)$ for the eight studied empirical bubbles described in section 2.2.1 along the three different pairs of parameters (t_1, m), (t_1, ω) and (t_1, t_c), outlining the dependence on the parameter t_1 representing the beginning of the bubble. Small (resp. large) values for the cost function correspond to lighter (resp. darker) colours and depict parameter domains giving better fits.



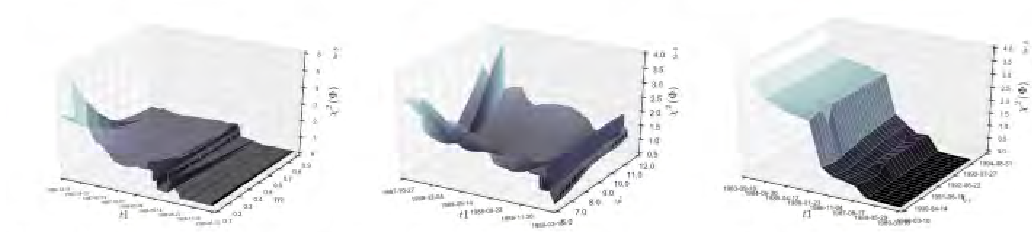
(a) S&P-500 (1997):$\chi^2(t_c \approx$ '$10.Fev\ 97'$, $m, \omega, t_1 =$ '$13.Jul.\ 95'$) (b) S&P-500 (1997): $\chi^2(t_c, m, \omega \approx 6.1, t_1 =$ '$13.Jul.\ 95'$) (c) S&P-500 (1997): $\chi^2(t_c, m \approx 0.71, \omega, t_1 =$ '$13.Jul.\ 95''$)



(d) SSEC *(2015)*: $\chi^2(m, \omega, t_c \approx$ '$Jun.\ 15'$, $t_1 =$ '$10.Mar.\ 14'$) (e) SSEC *(2015)*: $\chi^2(m, \omega \approx 6.1, t_c, t_1 =$ '$10.Mar.\ 14'$) (f) SSEC *(2015)*: $\chi^2(m \approx 0.54, \omega, t_c, t_1 =$ '$10.Mar.\ 14'$)



(g) NASDAQ: $\chi^2(m, \omega, t_c \approx$ '$Feb.10\ 2000'$, $t_1 =$ '$10.Jan.\ 96'$) (h) NASDAQ: $\chi^2(m, \omega \approx 8.2, t_c, t_1 =$ '$10.Jan.\ 96'$) (i) NASDAQ: $\chi^2(m \approx 0.6, \omega, t_c, t_1 =$ '$10.Jan.\ 96'$)



(j) NIKKEI: $\chi^2(m, \omega, t_c \approx$ '$Jan.\ 90'$, $t_1 =$ '$7.Jan.\ 88'$) (k) NIKKEI: $\chi^2(m \approx 0.3, \omega, t_c, t_1 =$ '$7.Jan.\ 88'$) (l) NIKKEI: $\chi^2(m \approx 0.7, \omega, t_c, t_1)$

Figure 2.6: Cross-sections of the cost function $\chi^2(\Phi)$ for the eight studied empirical bubbles described in section 2.2.1 along the three different pairs of parameters (t_1, m), (t_1, ω) and (t_1, t_c), outlining the dependence on the parameter t_1 representing the beginning of the bubble. Small (resp. large) values for the cost function correspond to lighter (resp. darker) colours and depict parameter domains giving better fits.

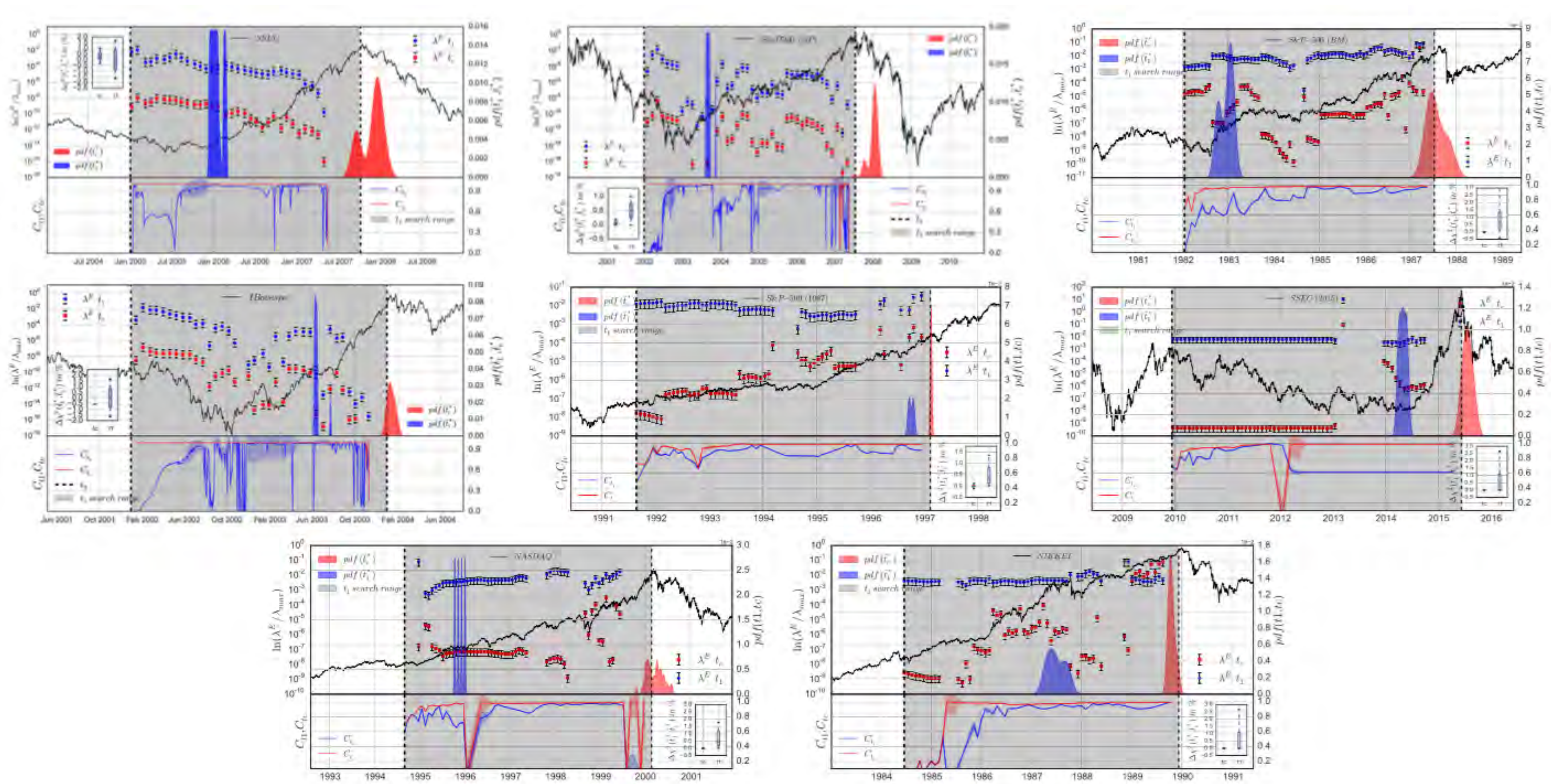


Figure 2.7: : Quantification of the relative sloppiness of t_c and t_1 for the eight empirical bubbles described in section 2.2.1. The logarithm of their price as a function of time is shown by the black continuous line in each panel. The blue circles and red squares show the normalised eigenvalues λ_3^E and λ_4^E of the Hessian matrix $H(\Phi^*)$ estimated for the best-fit, respectively mainly associated with t_1 and t_c, as a function of t_1 for a fixed $\bar{t}_2$ indicated by the black vertical dashed line. The errors bars represent $\pm$ one-sigma interval around the mean values, obtained over 100 estimations. The relative contribution defined by expression (2.4) of t_1 to the eigenvalue λ_3 and of t_c to λ_4 are shown using vertical coloured lines inside the grey-shaded area. The box-plots in the inset show the normalised cost change in percent when t_1 (respectively t_c) are sampled over the ensemble of calibrated $\widehat{t_1^*}$ (respectively $\widehat{t_c^*}$).

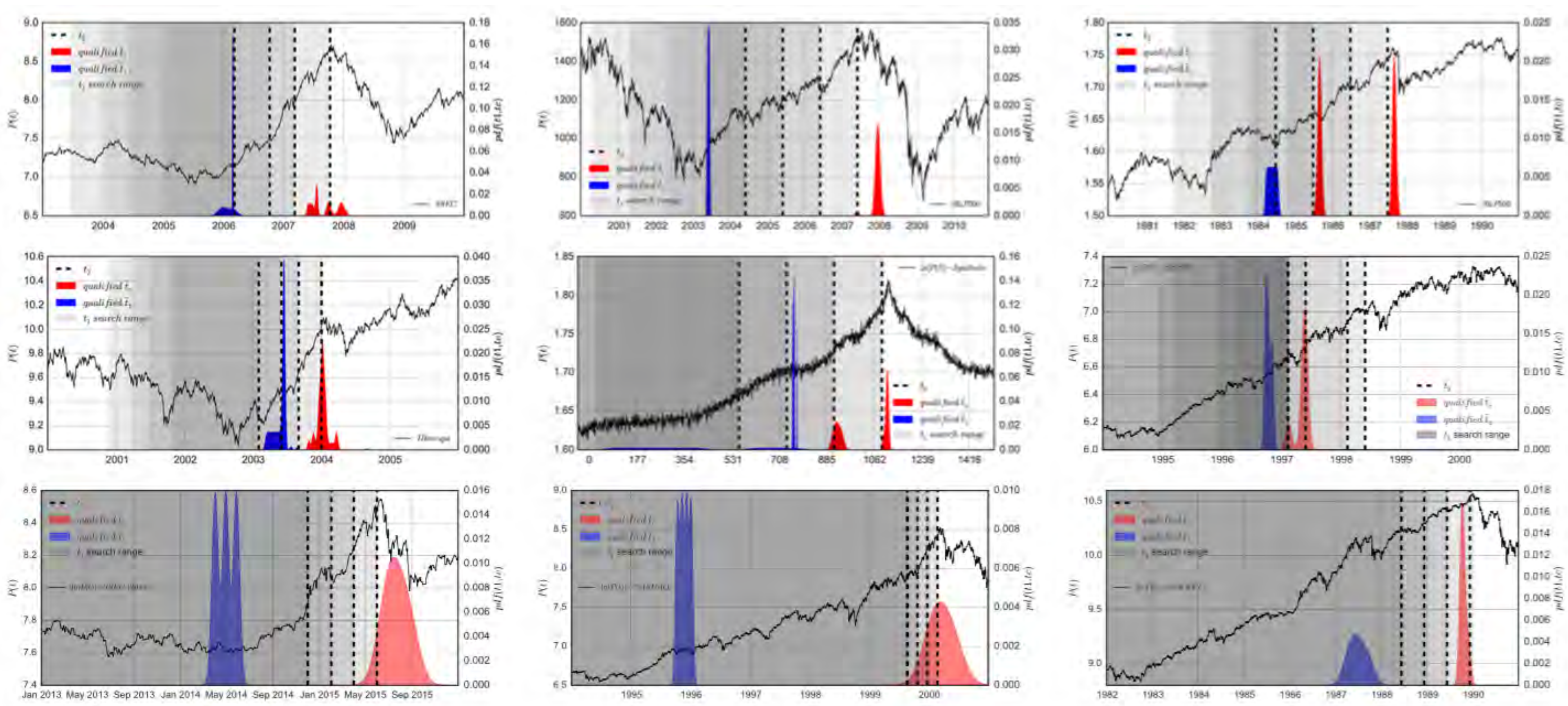


Figure 2.8: Analysis of the eight bubbles described in section 2.2.1 as well as for a synthetic bubble, for four different values of t_2 shown by the vertical dashed black lines in each panel. The red (resp. blue) filled curve represents the pdf of $\widehat{t_c^*}$ (resp. $\widehat{t_1^*}$) over the set of time windows obtained from the four t_2's by scanning t_1 over their corresponding grey shaded area, each of the same width of 600 days. The pdf's are represented using a kernel method with bandwidth ≈ 0.1 year.

	Hessian $\mathbf{H}(\phi^*)$				Eigenvectors $\vec{v}$		
				SSEC $t_1^* \approx Feb.\,2005$, $t^2 \approx Oct.\,2007$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	1.46e-1			1.00e0	**9.99e-1**	2.32e-2	1.24e-4
ω	3.32e-3	3.41e-3		2.28e-2	2.31e-2	9.98e-1	4.62e-2
t_c	1.46e-5	1.55e-4	1.00e-5	1.93e-5	9.49e-4	4.62e-2	9.98e-1
				S&P500 (SP) $t_1^* \approx Dec.\,2003$, $t^2 \approx Jul.\,2007$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	1.21e-2			1.00e0	9.99e-1	3.15e-2	1.80e-3
ω	3.46e-4	1.22e-3		9.95e-2	3.15e-2	9.99e-1	2.54e-2
t_c	2.10e-5	3.15e-5	8.98e-7	5.81e-6	9.98e-4	2.55e-2	9.99e-1
				S&P500 (BM) $t_1^* \approx Mar.\,1984$, $t^2 \approx Aug.\,1987$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	3.24e-1			1.00e0	9.99e-1	4.07e-2	3.02e-3
ω	1.22e-2	2.54e-2		7.70e-2	4.08e-2	9.98e-1	4.56e-2
t_c	9.33e-4	1.10e-3	1.71e-3	5.12e-3	1.16e-3	4.57e-2	9.98e-1
				IBovespa $t_1^* \approx Jun.\,2003$, $t^2 \approx Jan.\,2004$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	1.21e-2			1.00e0	9.83e-1	1.81e-1	-8.095009e-03
ω	1.80e-3	2.67e-3		1.88e-1	1.80e-1	9.82e-1	4.578182e-02
t_c	1.17e-4	8.36e-5	9.23e-5	6.96e-3	1.62e-2	4.35e-2	9.989187e-01
				S&P500 (1997) $t_1^* \approx Jul.\,1996$, $t^2 \approx Feb.\,1997$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	9.16e-4			1.00e0	**9.99e-1**	3.12e-2	1.33e-2
ω	2.62e-5	6.83e-5		7.86e-2	3.35e-2	**9.98e-1**	2.55e-1
t_c	1.16e-5	-1.77e-5	6.72e-6	2.15e-3	2.15e-3	2.56e-1	**9.66e-1**
				SSEC (2015) $t_1^* \approx Mar.\,2014$, $t^2 \approx Jun.\,2015$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	1.57e-1			1.00e0	9.99e-1	3.67e-2	4.09e-3
ω	5.55e-3	6.71e-3		4.13e-2	3.69e-2	9.98e-1	4.58e-2
t_c	6.34e-4	3.17e-4	1.25e-4	6.95e-4	2.39e-3	4.60e-2	9.98e-1
				NASDAQ $t_1^* \approx Jan.\,1996$, $t^2 \approx Feb.\,2000$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	4.11e-1			1.00e0	9.99e-1	1.65e-2	1.27e-3
ω	6.46e-3	2.13e-2		5.14e-2	1.65e-2	9.99e-1	1.43e-2
t_c	5.19e-4	3.13e-4	5.44e-6	9.39e-7	1.03e-3	1.44e-2	9.99e-1
				NIKKEI $t_1^* \approx Jan.\,1988$, $t^2 \approx Dec.\,1989$ λ/λ_{max}			
	m	ω	t_c		m	ω	t_c
m	8.72e-4			1.00e0	9.85e-1	1.68e-1	4.14e-3
ω	1.30e-4	1.32e-4		1.23e-1	1.69e-1	9.85e-1	-1.87e-2
t_c	-3.30e-6	-2.65e-6	5.44e-8	4.48e-7	9.21e-4	1.91e-2	9.99e-1

Table 2.4: Hessian matrix, eigenvalues and corresponding eigenvectors for the best LPPLS fits for the four studied empirical bubbles described in section 2.2.1 using Eq (1.12). The numbers represented in boldface identify the parameters that contribute most to their corresponding eigenvalue.

Index	Damping	Number of oscillations	LPPLS conditions	AR(1) residuals ($\vec{r}$) test	
				Dickey-Fuller	Phillips-Perron
Synthetic LPPLS	3.36	5.27	Satisfied	-3.420 (10^{-3})	-3.610 (10^{-3})
SSEC	1.04	3.91	Satisfied	-4.653 (10^{-3})	-4.642 (10^{-3})
$S\&P$ 500(SP)	1.15	5.13	Satisfied	-4.710 (10^{-3})	-4.702 (10^{-3})
$S\&P$ 500(BM)	1.41	4.02	Satisfied	-3.641 (10^{-3})	-3.641 (10^{-3})
$IBovespa$	1.85	3.47	Satisfied	-4.620 (10^{-3})	-4.610 (110^{-3})
$S\&P$ 500(1997)	1.15	2.32	Satisfied	-4.710 (10^{-3})	-4.702 (10^{-3})
SSEC (2015)	1.39	2.86	Satisfied	-4.710 (10^{-3})	-4.702 (10^{-3})
NASDAQ	1.32	2.63	Satisfied	-4.710 (10^{-3})	-4.702 (10^{-3})
NIKKEI	1.03	2.51	Satisfied	-4.710 (10^{-3})	-4.702 (10^{-3})

Table 2.5: Diagnosis of the LPPLS fits with calibrated parameters Φ^* for the four studied empirical bubbles described in section 2.2.1 and one synthetic data. We easily reject the hypothesis that the residuals are unit-root at the 99% confidence level. Values in brackets give the *p-values* of the test statistic.

Chapter 3

Lagrange regularisation approach to compare nested data sets and determine objectively financial bubbles' inceptions

3.1 Introduction

There is an inverse relationship between the tendency of a model to overfit data and the sample size used. In other words, the smaller the data sample size, the larger the number of degrees of freedom, the larger is the possibility of overfit [73]. Due this characteristic feature, one cannot compare directly goodness-of-fit metrics, such as the Residual Sum of Squares (RSS) := $\chi^2(\mathbf{\Phi})$ or its normalized version RSS/(N-p) := $\chi^2_{np}(\mathbf{\Phi})$, of statistical models over unequal sized samples for a given parametrisation Φ. Here, N denotes the sample size while p is the number of degrees of freedom of a model. This is particularly problematic when one is specifically interested in selecting the optimal sub-sample of a dataset to calibrate a model. This is a common problem when calibrating time series, when the model is only valid in a specific time window, which is unknown a priori. Our motivation stems from the question of determining the beginning of a financial bubble, but this question is more generally applicable to time series exhibiting regime shifts that one is interested in localising precisely.

In the literature, there are solutions for proper model selection such as the Lasso [121] and Ridge regressions [78], where the cost function contains an additional penalisation for large values of the estimated parameters. Well-known metrics such as the AIC and BIC are also standard tools for quantifying goodness-of-fit of different models [2] and for selecting the one with the best compromise between goodness-of-fit and complexity. However, results stemming from these methodologies are only comparable within the same data set.

There seems to be a gap in the literature about the proper procedure one should follow when comparing goodness-of-fit metrics of a model calibrated to different batches of a given data set. In order to fill this gap, we propose a novel metric for calibrating endogenised end points and compare nested data sets. The method empirically computes the tendency of a model to overfit a data set via what we term the "*Lagrange regulariser term*" λ. Once λ has been estimated empirically, the cost function can be corrected accordingly as a function of sample size, giving the Lagrange regularisation of $\chi^2_{np}(\mathbf{\Phi})$. As the number of data points or the window beginning- or end-point is now endogeneised, the optimal sample length can then be determined. We empirically test the performance of the Lagrange regularisation of $\chi^2_{np}(\mathbf{\Phi})$, which defined $\chi^2_{\lambda}(\mathbf{\Phi})$ as the regularised Residual Sum of Squares, in comparison with the naive $\chi^2(\mathbf{\Phi})$ and $\chi^2_{np}(\mathbf{\Phi})$ itself using both linear and non-linear models as well as synthetic and real-world time-series.

This paper is structured as follows. Section (3.2) explains the motivation behind the proposed Lagrange regularising term. Moreover, we provide details of the derivation of λ as well as the analytical expression for computing the tendency of a model to overfit data. In Section (3.3), we make use of a simple OLS regression to test the empirical performance of the Lagrange regularisation of $\chi^2_{np}(\mathbf{\Phi})$ on the problem of optimal sub-sample selection. Section (3.4) shows how the regulariser can be used alongside with the LPPLS model of financial bubbles in order to diagnose the beginning of financial bubbles. Empirical findings are given in Sec. (3.4.1) and Section (3.5) concludes.

3.2 Formulation of calibration with varying window sizes: How to endogenize t_1 and make different window sizes comparable

Let us consider the normalised mean-squared residuals, defined as the sum of squares of the residuals divided by the number $t_2 - t_1$ of points in the sum corrected by the number of degrees of freedom p of the model,

$$\chi^2_{np}(\mathbf{\Phi}) := \frac{1}{(t_2 - t_1) - p} \sum_{i=t_1}^{t_2} r_i(\mathbf{\Phi})^2 \ , \tag{3.1}$$

with

$$r_i(\mathbf{\Phi}) = y_i^{data} - y_i^{model}(\mathbf{\Phi}) \ , \tag{3.2}$$

where $\mathbf{\Phi}$ denotes the set of model parameters to fit including a priori the left end point t_1 of the calibration window. The term $y_i^{model}(\mathbf{\Phi})$ corresponds to the theoretical model and y_i^{data} is the empirical value of the time-series at time i.

For a fixed right end point t_2 of the calibration window, we are interested in comparing the results of the fit of the model to the empirical data for various left end points t_1 of the calibration window. The standard approach assumes a fixed calibration window $[t_1, t_2]$ with $N = t_2 - t_1 + 1$ data points. In order to relate the two problems, we consider the minimisation of $\chi^2_{np}(\mathbf{\Phi})$ at fixed $t_2 - t_1$ (for a fixed t_2) as minimising a general problem involving t_1 as a fit parameter augmented by the condition that $t_2 - t_1 + 1 = N$ is fixed. This reads

$$\mathrm{Min} \quad \chi^2_\lambda(\mathbf{\Phi}) \ , \tag{3.3}$$

with

$$\chi^2_\lambda(\mathbf{\Phi}) := \frac{1}{(t_2 - t_1) - p} \sum_{i=t_1}^{t_2} r_i(\mathbf{\Phi})^2 + \lambda(t_2 - t_1) \ , \tag{3.4}$$

where we have introduced the Lagrange parameter λ, which is conjugate to the constraint $t_2 - t_1 + 1 = N$. Once the parameters $\mathbf{\Phi}$ are determined, λ is obtained by the condition that the constraint $t_2 - t_1 + 1 = N$ is verified.

Since data points are discrete, the minimisation of (3.4) with respect to t_1 reads

$$\begin{aligned} 0 = \quad & \chi^2_\lambda(\mathbf{\Phi})(t_1+1) - \chi^2_\lambda(\mathbf{\Phi})(t_1) = \frac{1}{(t_2 - t_1 - p - 1)} \sum_{i=t_1+1}^{t_2} r_i(\mathbf{\Phi})^2 - \frac{1}{t_2 - t_1 - p} \sum_{i=t_1}^{t_2} r_i(\mathbf{\Phi})^2 - \lambda \\ = \quad & \frac{1}{t_2 - t_1 - p}\left(1 + \frac{1}{t_2 - t_1 - p} + \mathcal{O}\left(\frac{1}{(t_2 - t_1 - p)^2}\right)\right) \sum_{i=t_1+1}^{t_2} r_i(\mathbf{\Phi})^2 - \frac{1}{t_2 - t_1 - p} \sum_{i=t_1}^{t_2} r_i(\mathbf{\Phi})^2 - \lambda \ , \\ = \quad & -\frac{1}{t_2 - t_1 - p} r_{t_1}(\mathbf{\Phi})^2 \left(1 + \mathcal{O}\left(\frac{1}{t_2 - t_1 - p}\right)\right) + \frac{1}{t_2 - t_1 - p} \chi^2(\mathbf{\Phi}) \left(1 + \mathcal{O}\left(\frac{1}{t_2 - t_1 - p}\right)\right) - \lambda \ . \end{aligned} \tag{3.5}$$

Neglecting the small terms $\mathcal{O}\left(\frac{1}{t_2 - t_1 - p}\right)$ leads to

$$\chi^2_\lambda(\mathbf{\Phi}) = r_{t_1}(\mathbf{\Phi})^2 + \lambda(t_2 - t_1 - p) \ . \tag{3.6}$$

Expression (3.6) has the following implications. Consider the case where all squared terms $r_i(\mathbf{\Phi})^2$ in the sum (3.1) defining $\chi^2_\lambda(\mathbf{\Phi})$ are approximately the same and independent of t_1, which occurs when the residuals are thin-tailed distributed and the model is well specified. Then, we have

$$r_i(\mathbf{\Phi})^2 \approx r^2 \ , \ \ \forall i \ , \ \text{including } r_{t_1}(\mathbf{\Phi})^2 = r^2 \ , \tag{3.7}$$

and thus

$$\chi^2_{np}(\mathbf{\Phi}) \approx r^2 \ . \tag{3.8}$$

Expressing (3.6) with the estimation (3.7) yields

$$\chi^2_\lambda(\mathbf{\Phi}) \approx r^2 + \lambda(t_2 - t_1 - p) \ . \tag{3.9}$$

Comparing with (3.8), this suggests that varying t_1 is expected in general to introduce a linear bias of the normalised sum $\chi^2_{np}(\mathbf{\Phi})$ of squares of the residuals, which is proportional to the size of the calibration window (up to the small correction by the number p of degrees of freedom of the model). If we want to compare the calibrations over different window sizes, we need to correct for this bias.

More specifically, rather than fixing the window size $t_2 - t_1 + 1 = N$, we want to determine the 'best' t_1, thus comparing calibrations for varying window sizes, for a fixed right end point t_2. As a consequence, the Lagrange multiplier λ is no more fixed to ensure that the constraint $t_2 - t_1 + 1 = N$ holds, but now quantifies the average bias or "cost" associated with changing the window sizes. This bias is appreciable for small data sample sizes. It vanishes asymptotically as $N \to \infty$, i.e. $\lim_{N\to\infty}\lambda = 0$.

In statistical physics, this is analogous to the change from the canonical to the grand canonical ensemble, where the condition of a fixed number of particles (fixed number of points in a fixed window size) is relaxed to a varying number of particles with an energy cost per particle determined by the chemical potential (the Lagrange parameter λ) [45]. It is well-known that the canonical ensemble is recovered from the grand canonical ensemble by fixing the chemical potential (Lagrange multiplier) so that the number of particles is equal to the imposed constraints. Idem here.

How to determine the crucial Lagrange parameter λ? We propose an empirical approach. When plotting $\chi^2_{np}(\mathbf{\Phi})$ as a function of t_1 for various instances, we observe that a linearly decreasing function of t_1 provides a good approximation of it, as predicted by (3.6) (for $\lambda > 0$). The slope can then be interpreted as quantifying the average bias of the scaled goodness-of-fit $\chi^2_{np}(\mathbf{\Phi})$ due to the reduced number of data points as t_1 is increased. This average bias is clearly dependent on the data and of the model used to calibrate it. We can thus interpret the average linear trend observed empirically as determining the effective Lagrange regulariser term λ that quantifies the impact on the goodness-of-fit resulting from the addition of data points in the calibration, given the specific realisation of the data and the model to calibrate. Thus, to make all the calibrations performed for different t_1 comparable for the determination of the optimal window size, we propose to correct expression (3.1) by subtracting the term $\lambda(t_2 - t_1)$ from the normalised sum of squared residuals $\chi^2_{np}(\mathbf{\Phi})$ given by Eq. (3.1), where λ is estimated empirically as the large scale linear trend. Here, we omit the p correction since it leads to a constant translation for a given model with given number of degrees of freedom. Such a large scale linear trend of $\chi^2_{np}(\mathbf{\Phi})$ as a function of t_1 has been reported for a number of financial bubble calibrations in [31]. Our proposed procedure thus amounts simply to detrend $\chi^2_{np}(\mathbf{\Phi})$, which has the effect of making more pronounced the minima of $\chi^2_{np}(\mathbf{\Phi})$, as we shall see below for different models.

To summarise, endogenising t_1 in the set of parameters to calibrate requires to minimize

$$\chi^2_\lambda(\boldsymbol{\Phi}) = \chi^2_{np}(\boldsymbol{\Phi}) - \lambda(t_2 - t_1) \tag{3.10}$$

$$= \frac{1}{(t_2 - t_1) - p} \sum_{i=t_1}^{t_2} r_i(\boldsymbol{\Phi})^2 - \lambda(t_2 - t_1) \ , \tag{3.11}$$

with,

$$r_i(\boldsymbol{\Phi}) = y_i^{data} - y_i^{model}(\boldsymbol{\Phi}) \ , \tag{3.12}$$

where λ is determined empirically so that $\chi^2_{np}(\boldsymbol{\Phi}) - \lambda(t_2 - t_1)$ has zero drift as a function of t_1 over the set of scanned values. The obtained empirical value of λ can be used as a diagnostic parameter quantifying the tendency of the model to over-fit the data. We can thus also refer to λ as the "overfit measure". When it is large, the goodness-of-fit $\chi^2(\boldsymbol{\Phi})$ changes a lot with the number of data points, indicating a poor overall ability of the model to account for the data. [31] observed other cases where $\chi^2(\boldsymbol{\Phi})$ is constant as a function of t_1 (corresponding to a vanishing λ), which can be interpreted in a regime where the model fits robustly the data, "synchronizing" on its characteristic features in a way mostly independent of the number of data points.

3.3 Application of the Lagrange regularisation method to a simple linear-regression problem

Consider the following linear model:

$$\boldsymbol{Y} = \beta \boldsymbol{X} + \boldsymbol{\varepsilon}, \tag{3.13}$$

with explanatory variable of length $(N \times 1)$ denoted by $X = \{x_1, x_2, \ldots, x_N\}$, regressand $Y = \{y_1, y_2, \ldots, y_N\}$ and error vector $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$. Bold variables denote either matrices or vectors. Fitting Eq. (3.13) to a given data set Y^{data} consists on solving the quadratic minimisation problem

$$\hat{\beta} = \underset{\beta}{arg\,min}\ \chi^2(\boldsymbol{\Phi}), \tag{3.14}$$

where $\boldsymbol{\Phi}$ are parameters to be estimated and the objective function $\chi^2(\boldsymbol{\Phi})$ is given by

$$\chi^2(\boldsymbol{\Phi}) = \sum_{i=1}^{N} |Y_i^{data} - (Y_i^{model} - \beta X_i)|^2 \tag{3.15}$$

$$= ||\boldsymbol{Y}^{data} - (\boldsymbol{Y}^{model} - \boldsymbol{\beta X})||. \tag{3.16}$$

The solution of Eq. (3.14) with (3.16) for a given data set of length N reads

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{3.17}$$

Let $w^* \subseteq Y^{data}$ and have length $\leq$ N. $w^* \in [\tau : \bar{t}_2]$ thus denotes the optimal window size one should use for fitting a model into a data set of length N for a fixed end point $:= t_2$ and an optimal starting point $:= \tau$.

In order to show how the goodness-of-fit metric $\chi^2(\boldsymbol{\Phi})$ fails to flag the optimal τ-portion of the data set where the regime of interest exists and how delicate is $\chi^2_{np}(\boldsymbol{\Phi})$ for diagnosing the true value of the transition time τ, 20000 synthetic realisations of the process (3.13) were generated, with $X := t \in [-200, +1]$, in such a way that Y^{data} displays a sudden change of regime at $\tau = -100$. In the first half of the dataset $[-200, -100]$, the data points are generated with $\beta = 0.3$. In the second half of the dataset $[-101, 0]$, the data points are generated with $\beta = 0.6$. After the addition of random noise $\epsilon \sim \mathcal{N}(0, 1)$, each single resulting time-series was fitted for a fixed end time $\bar{t}_2 = 1$ while shrinking the left-most portion of the data (t_1) towards t_2, starting at $t_1 = -200, -199, \ldots, t_2 - 3$. For the largest window with $t_1 = -200$, there are $t_2 - t_1 + 1 = 1 - (-200) + 1 = 202$ data points to fit. For the smallest window with $t_1 = t_2 - 3$, there are $t_2 - t_1 + 1 = 4$ data points to fit. For each window size w, the process of generating synthetic data and fitting the model was repeated 20000 times, allowing us to obtain confidence intervals.

As depicted by Fig. (3.1), the proposed methodology is able to correctly diagnose the optimal starting point $:= \tau$ associated with the change of slope. While the $\chi^2(\Phi)$ metric monotonously decreases and the $\chi^2_{np}(\Phi)$ metric plateaus from $t = -100$ onwards, $\chi^2_{np} - \lambda(t_2 - t_1)$ monotonously increases over the same interval, thus marking a clear minimum. The variance of the metric $\chi^2_{\lambda}(\mathbf{\Phi})$ also increases over this interval. Specifically, the metric $\chi^2(\Phi)$ tends to favor the smallest windows and therefore overfitting is prone to develop and remain undetected. The metric $\chi^2_{np}(\mathbf{\Phi})$ suggests $\tau \approx -90$ after 20000 simulations, which is 10% away from the true value $\tau = 100$. Moreover, the dependence of $\chi^2_{np}(\mathbf{\Phi})$ as a function of t_1 is so flat for $t_1 \in [-100 : -40]$ that any given value of τ within this period is statistically significant. For this simulation study, $\chi^2_{np}(\mathbf{\Phi})$ ranges for 0.134 to 0.135 for $t_1 \in [-100 : -60]$, so as to be almost undistinguishable over this interval of possible τ values. As we shall see later on, the performance of $\chi^2_{np}(\mathbf{\Phi})$ degrades further to resemble that of the $\chi^2(\mathbf{\Phi})$ metric when dealing with more complex nonlinear models such as the LPPLS model. On the other hand, our proposed correction via the Lagrange regulariser λ provides a simple and effective method to identify the change of regime and the largest window size compatible with the second regime. The minimum is very pronounced and clear, which is not the case for $\chi^2_{np}(\mathbf{\Phi})$.

3.4 Using the Lagrange regularisation method for Detecting the Beginning of Financial Bubbles

In the previous Section, we have proposed a novel goodness-of-fit metric for inferring the optimal beginning point or change point τ (for a fixed end point $\bar{t}_2$) in the calibration of a simple linear model. The application of the Lagrange regulariser λ allowed us to find the optimal window length $w^* = [\tau : t_2]$ for fitting the model by enabling the comparison of the goodness-of-fits across different w values. We now extend the application of the methodology to a more complex non-linear model, which requires one to compare fits across different window sizes in order to diagnose bubble periods on financial instruments such as equity prices and price indexes.

3.4.1 Empirical analysis

We apply our novel goodness-of-fit metric to the problem of finding the beginning times of financial bubbles, defined as the optimal starting time t_1 obtained by endogenising t_1 and calibrating it. We first illustrate and test the method on synthetic time series and then apply it to real-world financial bubbles. A Python implementation of the algorithm is provided in the appendix.

Construction of synthetic LPPLS bubbles

To gain insight about the application of our proposed calibration methodology on a controlled framework and thus establish a solid background to our empirical analysis, we generate synthetic price time series that mimic the salient properties of financial bubbles, namely, a power law-like acceleration decorated by oscillations. The synthetic price time series are obtained by using formula (1.6) with parameters given by the best LPPLS fit within the window $w \in [t_1 = 1\ Jan.\ 1981\text{: } t_2 = 30\ Aug.\ 1987]$ of the bubble that ended with the Black Monday 19 Oct. 1987 crash. These parameters are m = 0.44, ω=6.5, C_1 = -0.0001, C_2=0.0005, A=1.8259, B= -0.0094, $t_c = 1194$ (corresponding to 1987/11/14), where days are counted since an origin put at $t_1 = Jan.\ 1981$. To the deterministic component describing the expected log-price given by expression (1.6) and denoted by $fLPPLS(\boldsymbol{\phi}, t)$, we add a stochastic element to obtain the synthetic price time series

$$\ln[P(t)] = fLPPLS(\boldsymbol{\phi}, t) + \sigma\epsilon(t), \tag{3.18}$$

where $\epsilon(t) \sim \mathcal{N}(0, \sigma^0)$ noise, $\sigma^0 = 0.03$ and $t = [1, \ldots, N = 1100]$.

To create a price time series with a well-defined transition point corresponding to the beginning of a bubble, we take the first 500 points generated with expression (3.18) and mirror them via a $t \to t_1 - t$ reflection across the time $t_1 = 1\ Jan.\ 1981$. We concatenate this reflected sequence of 500 prices to the 1100 prices obtained with (3.18) for $t \geq t_1$, so that the true transition point corresponding to the start of the bubble described by the LPPLS pattern is $t_1 = 1\ Jan.\ 1981$. The black stochastic line on the top of figure (3.2) represent this union of the two time-series. This union constitutes the whole synthetic time series on which we are going to apply our Lagrange regularisation of $\chi^2_\lambda(\mathbf{\Phi})$ in order to attempt recovering the true start time, denoted by the hypothetical time $t_1 = 1\ Jul.\ 1911$.

For each synthetic bubble price time series, we thus calibrated it with Eq. (1.6) by minimizing expression (3.1) in windows $w = [t_1, t_2]$, varying t_2 from 1912/07/01 to $t_2 = 1913/01/01$, with t_1 scanned from $t_1 = Jan.\ 1910$ up to 30 business days before t_2, i.e. up to $t_{1,max} = t_2 - 30$ for each fixed t_2. The goal is to determine whether the transition point τ we determine is close (or even equal to) the true hypothetical value $t_1 = 01\ Jul.\ 1911$ for different maturation times t_2 of the bubble. The number of degrees of freedom used for this exercise as well as for the real-world time series is $p = 8$, which includes the 7 parameters of the LPPLS model augmented by the extra parameter t_1.

Real-world data: analysing bubble periods of different financial Indices

The real-world data sets used consists on bubble periods that have occurred on the following major Indexes: $S\&P$-500[1], IBovespa[2] and $SSEC$[3]. For each data set and for each fixed pseudo present time t_2 depicted by red vertical dashed lines on Fig. (3.2), our search for the bubble beginning time τ consists in fitting the LPPLS model using a shrinking estimation window w with $t_1 = [t_2 - 30 : t_2 - 1600]$ with incremental step-size of 3 business days. This yields a total of 514 fits per t_2.

Analysis

Let us start with the analysis of the synthetic time-series[4] depicted in Fig. (3.3). For the earliest $t_2 = 1912/07/01$, our proposed goodness-of-fit scheme is already capable of roughly diagnosing correctly the bubble beginning time, finding the optimal τ to be $\approx May$ 1911. In contrast, the competing metric ($\chi^2_{np}(\mathbf{\Phi})$) is degenerate as $t_1 \rightarrow t_2$ and is thus blind to the beginning of the bubble. For t_2 closer to the end of the bubble, $\chi^2_{np}(\mathbf{\Phi})$ continues to deliver very small optimal windows, proposing the incorrect conclusion that the bubble has started very recently (i..e close to the pseudo present time t_2). This is a signature of strong overfitting, which is quantified via λ and depicted in the title of the figure alongside with the bubble beginning time and t_2. The Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi})$ locks into the true value of $\tau \approx Jul.1911$ as $t_2 \rightarrow t_c$, i.e., as t_2 moves closer and

[1] t_2's = {1987.07.15; 1997.06.01; 2000.01.01; 2007.06.01}

[2] t_2's = {2000.01.01; 2004.01.01; 2006.01.01; 2007.12.01}

[3] t_2's = {2000.08.01; 2007.05.01; 2009.07.01; 2015.05.01}

[4] t_2's = {1912.07.01; 1912.10.01; 1912.11.15; 1913.01.01}

closer to January 1913 and the LPPLS signal becomes stronger.

We now switch to the real-world time-series. For the $S\&P$-500 Index, see Fig. (3.4), the results obtained are even more pronounced. While again $\chi^2_{np}(\mathbf{\Phi})$ is unable to diagnose the optimal starting date of a faster than exponential log-price growth $\tau \equiv t_1$, the Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi})$ depicted by blank triangles in the lower box of the figure is capable of overcoming the tendency of the model to overfit data as $t_1 \to t_2$. Specifically, the method diagnoses the start of the Black-Monday bubble at $t_1 \approx March$ 1984 and the beginning of the Sub-Prime bubble at $\approx$ Aug. 2003 in accordance with [130].

We also picked two pseudo present times $t_2's$ at random in order to check how consistent are the results. To our delight, the method is found capable of capturing the different time-scales present of bubble formation in an endogenous manner. For $t_2 = 1997.06.01$, the method suggests the presence of a bubble that nucleated more than five years earlier. This recovers the bubble and change of regime in September 1992, documented in Chapter 9 of [110] as a "false alarm" in terms of being followed by a crash. Nevertheless, it was a genuine change of regime as the market stopped its ascent and plateaued for the three following months. For $t_2 = 2000.01.01$, $\chi^2_\lambda(\mathbf{\Phi})$ diagnoses a bubble with a shorter duration, which started in November 1998. The starting time is coherent with the recovery after the so-called Russian crisis of August-September 1998 when the US stock markets dropped by about 20%. And this bubble is nothing but the echo in the S&P500 of the huge dotcom bubble that crashed in March-April 2000. More generally, scanning t_2 and different intervals for t_1, the Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi})$ can endogenously identify a hierarchy of bubbles of different time-scales, reflecting their multi-scale structure [110, 36].

For the IBovespa and the SSEC Index (Figures (3.5) and (3.6) respectively), the huge superiority of the Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi})$ vs. the $\chi^2_{np}(\mathbf{\Phi})$ metric is again obvious. For each of the four chosen t_2's in each figure, $\chi^2_\lambda(\mathbf{\Phi})$ exhibits a well-marked minimum corresponding to a well-defined starting time for the corresponding bubble. These objectively identified t_1 correspond pleasantly to what the eye would have chosen. They pass the "smell test" [101]. In contrast, the $\chi^2_{np}(\mathbf{\Phi})$ metric provides essentially no guidance on the determination of t_1.

3.5 Conclusion

We have presented a novel goodness-of-fit metric, aimed at comparing goodnesses-of-fit across a nested hierarchy of data sets of shrinking sizes. This is motivated by the question of identifying the start time of financial bubbles, but applies more generally to any calibration of time series in which the start time of the latest regime of interest is unknown. We have introduced a simple and physically motivated way to correct for the overfitting bias associated with shrinking data sets, which we refer to at the Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi}) := \frac{1}{N-p} SSR$. We have suggested that the bias can be captured by a Lagrange regularisation parameter λ. In addition to helping remove or alleviate the bias, this parameter can be used as a diagnostic parameter, or "overfit measure", quantifying the tendency of the model to overfit the data. It is a function of both the specific realisation of the data and of how the model matches the generating process of the data.

Applying the Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi})$ to simple linear regressions with a change point, synthetic models of financial bubbles with a well-defined transition regime and to a number of financial time series (US S&P500, Brazil IBovespa and China SSEC Indices), we document its impressive superiority compared with the $\chi^2_{np}(\mathbf{\Phi})$ metric. In absolute sense, the Lagrange regularisation of the $\chi^2_{np}(\mathbf{\Phi})$ is found to provide very reasonable and well-defined determinations of the starting times for major bubbles such as the bubbles ending with the 1987 Black-Monday, the 2008 Sub-prime crisis and minor speculative bubbles on other Indexes, without any further exogenous information.

Appendix

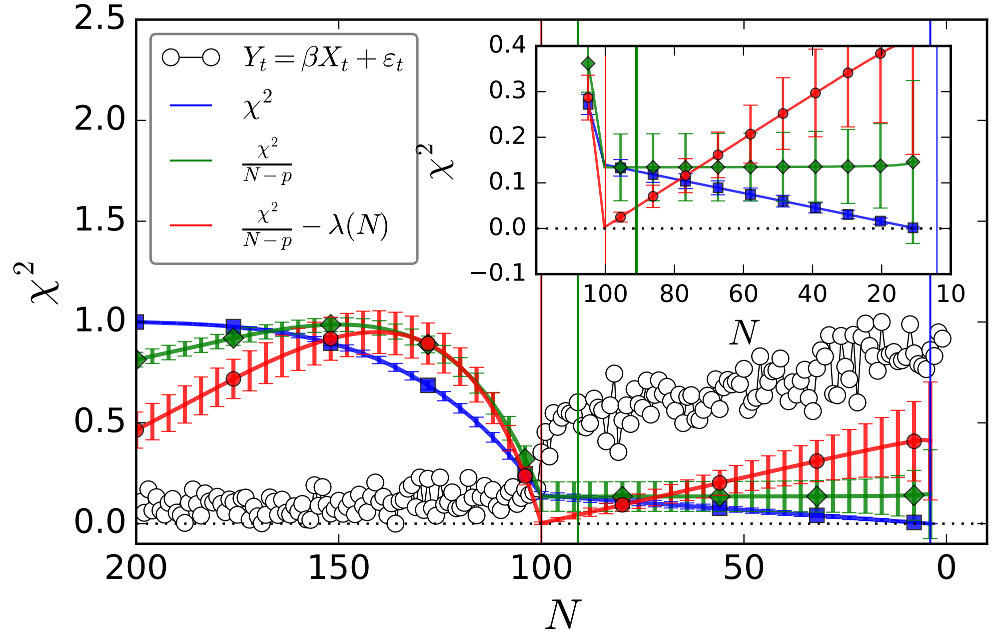


Figure 3.1: **Different goodness-of-fit measures applied to a shrinking-window linear regression problem (Eq. 3.13) in order to diagnose the optimal calibration window length:** We simulated synthetic time-series with length N=200 (white circles) using expression (3.13) with a sudden change of regime at $t = -100$. We then fitted the same model (3.13) within shrinking windows (from left to right), i.e. for a fixed $t_2 = 1$, we shrink t_1 from t_1=-200 to t_1=-3 and show the values of $\chi^2(\mathbf{\Phi})$ (blue), $\chi^2_{np}(\mathbf{\Phi})$ (green) and $\chi^2_{\lambda}(\mathbf{\Phi})$ (red) metrics as a function of this shrinking estimation window. For each pair $[t_2 : t_1]$ (i.e. for each N), the process of generating synthetic data and fitting the model was repeated 20000 times (resulting on confidence bounds for each metric). For t=[-200:-100], Y_t was simulated with $\beta = 0.3$ while from t = [-100:1], $\beta = 0.6$ was used. Without loss of generality, both the data and the cost functions had their values divided by their respectively maximum value in order to be bounded within the interval $[0, 1]$. A Python script for generating the figure and performing all calculations can be found on Appendix.

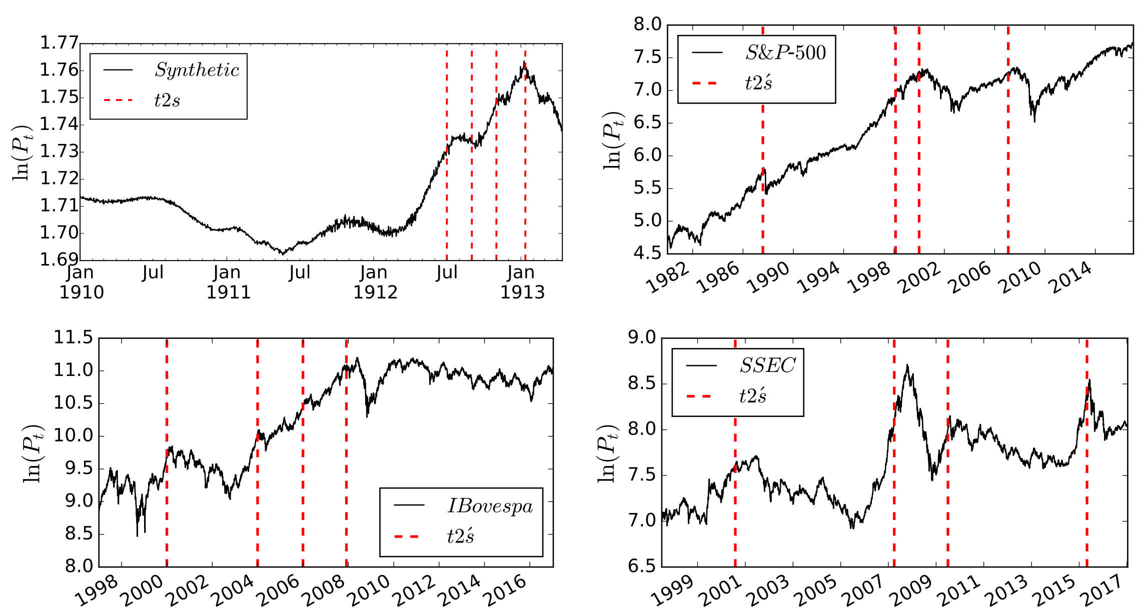


Figure 3.2: **Synthetic and real-world Time-series used in this study for measuring the performance of different goodness-of-fit metrics at different t_2's (red lines):** Synthetic time-series and Indexes $S\&P$-500, $IBovespa$ and $SSEC$ with $t_2's = \{1912.07.01; 1912.10.01; 1912.11.15; 1913.01.01\}$, $t_2's = \{1987.07.15;\ 1997.06.01;\ 2000.01.01;\ 2007.06.01\}$, $t_2's = \{2000.01.01; 2004.01.01; 2006.01.01; 2007.12.01\}$ and $t_2's = \{2000.08.01; 2007.05.01; 2009.07.01; 2015.05.01\}$ respectively (red dashed vertical lines).

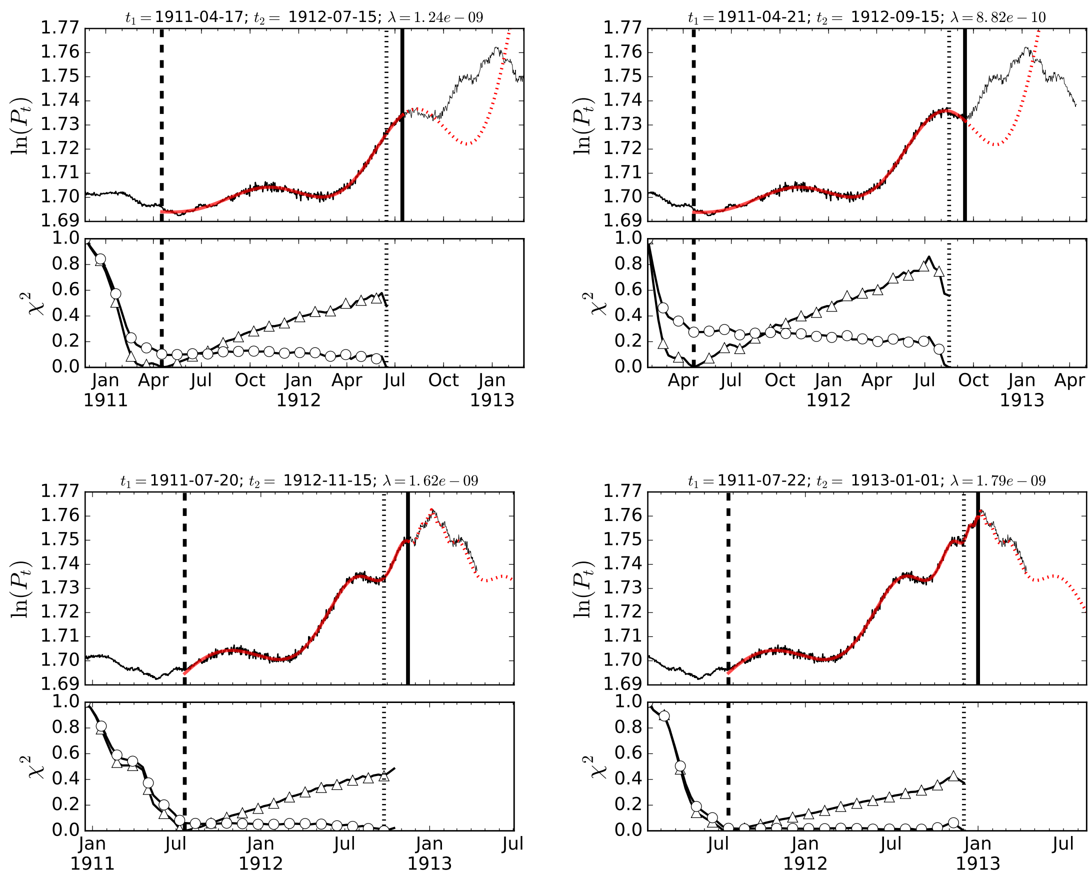


Figure 3.3: **Diagnosing the beginning of financial bubbles by comparing two goodness-of-fit metrics $\chi^2_{np}(\mathbf{\Phi})$ vs. $\chi^2_{\lambda}(\mathbf{\Phi})$ using the LPPLS model on Synthetic Time-Series:** $\chi^2_{np}(\mathbf{\Phi})$ is depicted by blank circles in the lower plot while our proposed metric is depicted by blank triangles. The dashed black vertical lines denotes the minimum of each goodness of fit metric and therefore represents the optimal $\tau \equiv t_1$ for $\chi^2_{np}(\mathbf{\Phi})$ and $\chi^2_{\lambda}(\mathbf{\Phi})$. For a fixed t_2, the log-price time-series of the Index was fitted using a shrinking window from $t_1 = [t_2 - 30 : t_2 - 1600]$ sampled every 3 days. For a fixed t_2 and t_1, we display the resulting fit of the LPPLS model (red line) obtained with the parameters solving Eq. (1.11).

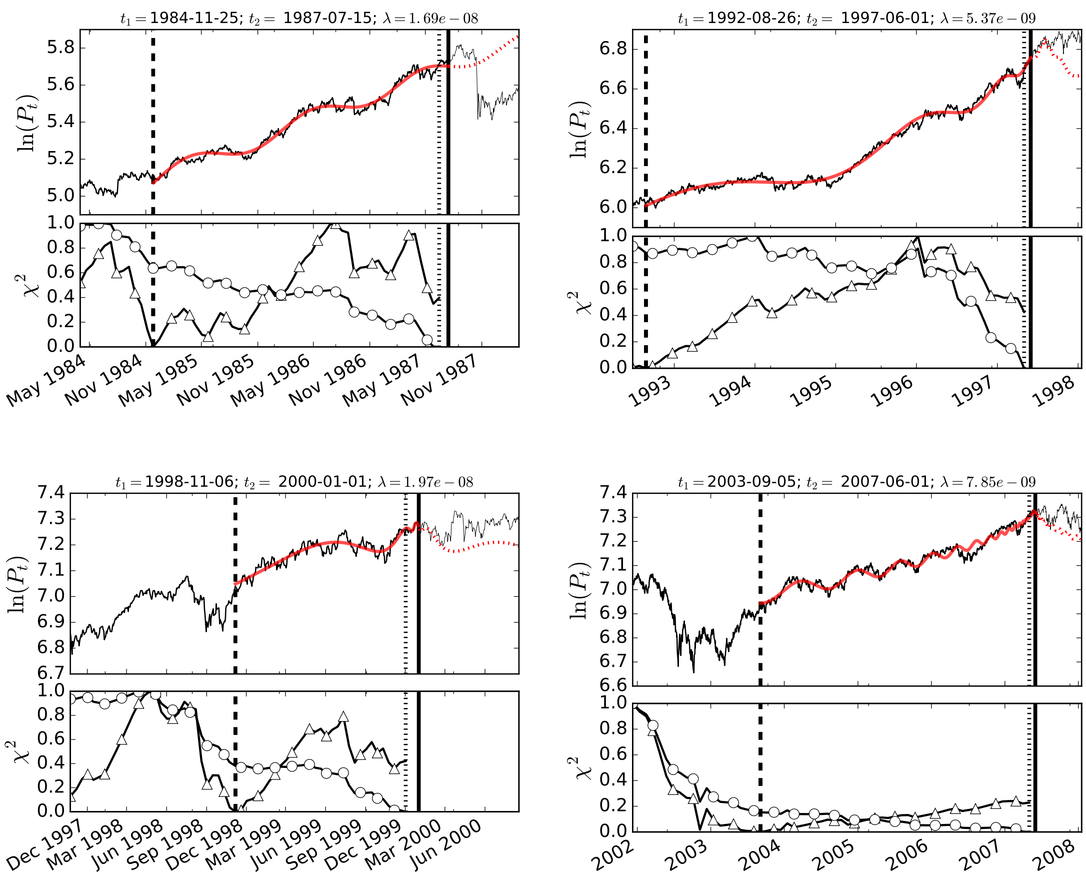


Figure 3.4: **Same as figure 3.3 for the US** $S\&P$**-500 Index.**

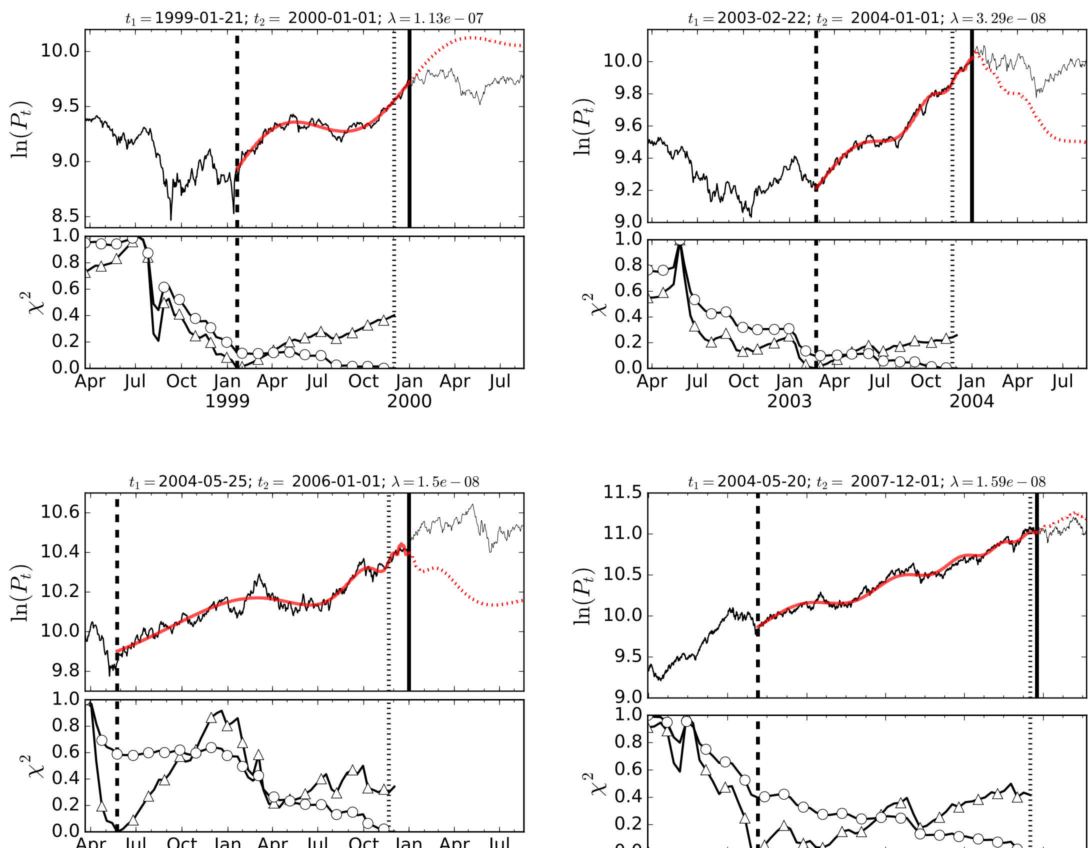


Figure 3.5: **Same as figure 3.3 for the Brazilian** ***IBovespa*** **index.**

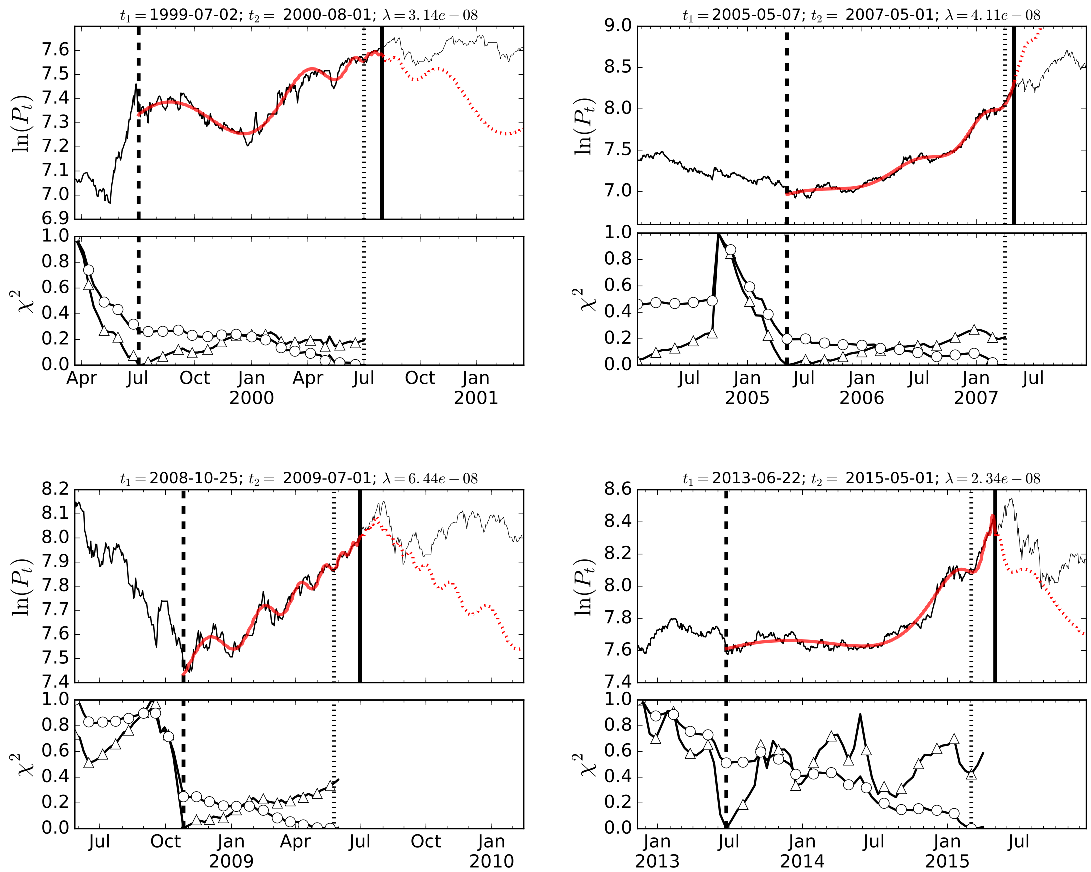


Figure 3.6: **Same as figure 3.3 for the Chinese *SSEC* index.**

```
// Python script for computing the Lambda regulariser metric - OLS case.

// Copyright: G.Demos @ ETH-Zurich - 2017.


########################

def simulateOLS():

    """ Generate synthetic OLS as presented in the paper """

    nobs = 200

    X    = np.arange(0,nobs,1)

    e    = np.random.normal(0, 10, nobs)

    beta = 0.5

    Y    = [beta*X[i] + e[i] for i in range(len(X))]

    Y = np.array(Y)

    X = np.array(X)

    Y[:100] = Y[:100] + 4*e[:100]

    Y[100:200] = Y[100:200]*8


    return X, Y



########################

def fitDataViaOlsGetBetaAndLine(X,Y):

    """ Fit synthetic OLS """

    beta_hat = np.dot(X.T,X)**-1. * np.dot(X.T,Y) # get beta

    Y = [beta_hat*X[i] for i in range(len(X))] # generate fit


    return Y
```

```
########################
def getSSE(Y, Yhat, p=1, normed=False):
    """
    Obtain SSE (chi^2)
    p -> No. of parameters
    Y -> Data
    Yhat -> Model
    """
    error = (Y-Yhat)**2.
    obj = np.sum(error)
    if normed == False:
        obj = np.sum(error)
    else:
        obj = 1/np.float(len(Y) - p) * np.sum(error)

    return obj



########################
def getSSE_and_SSEN_as_a_func_of_dt(normed=False, plot=False):
    """ Obtain SSE and SSE/N for a given shrinking fitting window w """
    # Simulate Initial Data
    X, Y = simulateOLS()

    # Get a piece of it: Shrinking Window
    _sse, _ssen = [], []
```

```
    for i in range(len(X)-10): # loop t1 until: t1 = (t2 - 10):
        xBatch = X[i:-1]
        yBatch = Y[i:-1]
        YhatBatch = fitDataViaOlsGetBetaAndLine(xBatch, yBatch)
        sse = getSSE(yBatch, YhatBatch, normed=False)
        sseN = getSSE(yBatch, YhatBatch, normed=True)
        _sse.append(sse); _ssen.append(sseN)

    if plot == False:
        pass
    else:
        f, ax = plt.subplots(1,1,figsize=(6,3))
        ax.plot(_sse, color='k')
        a = ax.twinx()
        a.plot(_ssen, color='b')
        plt.tight_layout()
    if normed==False:
        return _sse, _ssen, X, Y # returns results + data
    else:
        return _sse/max(_sse), _ssen/max(_ssen), X, Y # returns results + data


########################
def LagrangeMethod(sse):
    """ Obtain the Lagrange regulariser for a given SSE/N"""
    # Fit the decreasing trend of the cost function
    slope = calculate_slope_of_normed_cost(sse)
```

```
    return slope[0]



########################

def calculate_slope_of_normed_cost(sse):

    #Create linear regression object using statsmodels package

    regr = linear_model.LinearRegression(fit_intercept=False)


    # create x range for the sse_ds

    x_sse = np.arange(len(sse))

    x_sse = x_sse.reshape(len(sse),1)


    # Train the model using the training sets

    res = regr.fit(x_sse, sse)


    return res.coef_



########################

def obtainLagrangeRegularizedNormedCost(X, Y, slope):

     """ Obtain the Lagrange regulariser for a given SSE/N Pt. III"""

    Yhat = fitDataViaOlsGetBetaAndLine(X,Y) # Get Model fit

    ssrn_reg = getSSE(Y, Yhat, normed=True) # Classical SSE

    ssrn_lgrn = ssrn_reg - slope*len(Y) # SSE lagrange


    return ssrn_lgrn
```

```
########################

def GetSSEREGvectorForLagrangeMethod(X, Y, slope):


    """

    X and Y used for calculating the original SSEN

    slope is the beta of fitting OLS to the SSEN

    """


    # Estimate the cost function pondered by lambda using a Shrinking Window.

    _ssenReg = []

    for i in range(len(X)-10):

        xBatch = X[i:-1]

        yBatch = Y[i:-1]

        regLag = obtainLagrangeRegularizedNormedCost(xBatch,

              yBatch,

              slope)

        _ssenReg.append(regLag)


    return _ssenReg
```

Chapter 4

On the Predictability of Stock Market Bubbles: Evidence from LPPLS Confidence$^{\mathrm{TM}}$ Multi-scale Indicators

Bubble formation in financial markets has always been a topic of great interest, not only from an academic perspective regarding the informational efficiency of markets, but also for practitioners and policy makers who try to mitigate the negative effects of wild price fluctuations and subsequent crashes. Clearly, large sums are at stake when one deals with stock market bubbles. The U.S. market capitalization of firms, as of June 2017, is nearly $25.3 trillion USD (Bloomberg) or 133% of GDP, while the S&P 500 Total Market Capitalization was $21.3 trillion USD on March 31, 2017. Correspondingly, at the end of 2016, $15.9 trillion in corporate equities were held by households and non-profit organizations in the U.S. (Balance Sheet of Households and Nonprofit Organizations

(B.101) in Financial Accounts of the United States.). Given the level of penetration of the stock market investment in the U.S., a future crash in the stock market is likely to have widespread negative effects on the U.S. economy ([35, 77]), as it did in the aftermath of the dotcom bubble in 2000 or of the financial crisis of 2008. Therefore, it is not surprising that a large literature has been devoted to detecting bubbles in the U.S. stock market, especially since the 2008 financial crisis (see for example [21, 126, 25, 125] for detailed reviews in this regard). While significant effort has been spent on explaining how and why bubbles emerge and sustain over long periods[63], a large number of studies have instead focused on developing models to reliably detect bubbles. Consequently, the literature provides various methodologies to detect bubbles that aim to build on the drawback of the existing (earlier) ones (see for example, [22, 4, 84, 93, 5] for detailed discussions in this regard).

Against this backdrop and given that the $S\&P500$ is one of the most frequently used stock market indexes to gauge the U.S. stock market performance ([34]), representing more than 80% of available market capitalization, the objective of this paper is twofold. First, we present a methodology to detect both positive and negative bubbles [1] for the $S\&P500$ index using the Log-Periodic Power Law Singularity (LPPLS) model ([56, 110]), not otherwise possible based on the various bubble detection models cited earlier. We then introduce the Multi-scale LPPLS confidence$^{\text{TM}}$ indicators to characterise bubbles at different time scales. Second, having constructed indicators that describe positive and negative bubbles, we examine the predictive power of short selling activity and market liquidity over the bubble indicators and provide insight to the predictability of market booms and crashes using market-based indicators. In our predictive tests, we specifically focus on measures of short selling activity and market liquidity as recent studies suggest that

[1] A positive (resp. negative) bubble is defined as an upward (resp. downward) accelerating price followed by a crash (resp. rally).

short sellers are informed traders who are able to anticipate future aggregate cash flows and that short interest is positively related to stock price crash risk (e.g. [26] and [85]). We are also interested in the predictive ability of market liquidity over boom and crash indicators as the literature establishes a link between liquidity spikes and market downturns (e.g. [28, 81]). To the best of our knowledge, this is the first attempt in examining the predictive ability of market-based indicators over the positive and negative bubbles in the $S\&P500$ index using the LPPLS model. An important finding of this paper is thus to identify some of the factors contributing to the LPPLS-based bubble indicators, namely short selling and liquidity. The LPPLS model, which makes it possible to identify positive and negative bubbles, presents a valuable opening, allowing us to examine the predictability patterns of market booms and crashes separately.

Our findings show that the LPPLS framework, in the implementation presented here, is able to successfully capture, ex-ante, some of the prominent bubbles across different time scales. We observe that some of the great bubbles and subsequent crashes experienced during the Black Monday, Dot-com, and Subprime Crisis periods are successfully captured by the bubble indicators, while the long-term negative bubble indicator diagnoses correctly the transition from a sluggish market to a fast accelerating positive bubble during the mid-90s when the demand for the “New Economy” stocks developed in full force. One can also observe that the medium-term negative bubble indicator also shows a strong signal that the 2008 crisis was ending, providing a precursor to the strong rebound that started in March 2009. Examining the predictability of the negative and positive bubble indicators, our predictive tests reveal several interesting observations. First, we observe that the predictability patterns differ significantly for booms and busts represented by the positive and negative bubble indicators, respectively. We find that measures of short selling activity have robust predictive power over negative bubbles, in line with the view that short sellers are able to detect bad news hoarding by managers. The predictive power of short selling activity is robust to alternative measures of short selling employed in our predictive tests and is positively related to the negative bubble indicator, predicting the occurrence of negative bubbles one-month ahead. The predictive power of short selling proxies holds for both the short- and the long-term horizons which is consistent with the recent finding by [**?**] that short interest is positively related to one-year ahead stock price crash risk. On the other hand, our tests show that market liquidity also has robust predictive power over both the negative and positive bubbles in the short-term, suggesting that market liquidity measures can be used to predict the occurrence of both booms and crashes for short horizons. In short, the evidence points to the predictability of both positive and negative stock market bubbles via market-based proxies of trading activity.

The remainder of the paper is organized as follows. Section 4.1 explains the methodology to construct the LPPLS Confidence$^{\text{TM}}$ indicator, its application to the $S\&P500$ Index and the Markov Switching (MS) model employed in our predictive tests. Section 4.2 presents the empirical findings and Section 4.3 concludes the paper.

4.1 Positive and Negative Bubbles and LPPLS Confidence$^{\text{TM}}$ indicators

4.1.1 Capturing Positive and Negative Bubbles

As mentioned earlier, the methodology presented in this paper not only permits one to decouple the analysis of bubbles into different time-scales, but also allows one to focus on positive or negative bubbles separately. In the case of positive bubbles, the asset price grows super-exponentially towards t_c and ends with a change of regime (in general a crash), whereas negative bubbles are the exact $y \to -y$ mirror of positive bubbles with respect to the horizontal axis and exhibit an accelerating price drop ending with a change of regime, in general a potential "negative" crash, i.e. a substantial price appreciation (i.e. price rebound). This feature is captured by the LPPLS model through parameter B with the estimated parameter $\hat{B} < 0$ indicating a positive bubble and $\hat{B} > 0$ indicating a negative bubble. In both positive and negative bubbles, the critical time t_c denotes the time at which the bubble ends.

4.1.2 Definition of LPPLS Confidence$^{\text{TM}}$ indicators

The LPPLS Confidence$^{\text{TM}}$ indicator was introduced by [112] and It is also one of the key indicators powering the Financial Crisis Observatory[2] at ETH Zurich. It is defined as the fraction of fitting windows whose calibrations meet the filtering condition depicted in Table (7.1). It thus measures the sensitivity of the observed bubble pattern to the 142 time windows of duration from 30 to 750 trading days. A large value indicates that the LPPLS pattern is found at most scales and is thus more reliable. If the value is close to one, the pattern is practically insensitive to the choice of the window size $dt := t_2 - t_1$. A small value of the indicator signals a possible fragility since it is presented in a few fitting windows.

4.1.3 Multi-scale Indicators

In order to incorporate bubbles of different scales into the analysis, we introduce the Multi-scale LPPLS Confidence$^{\text{TM}}$ Indicator which is constructed as follows:

- <u>Short-term bubble</u>: The short-term bubble indicator at time t_2 is a number $\in [0, 1]$ which denotes the fraction of qualified fits for estimation windows of length $dt := t_2 - t_1 \in [30 : 90]$ business days for this t_2. As an example, if a fit is qualified

[2] http://tasmania.ethz.ch/pubfco/fco.html

at a given window i (i.e. the filtering conditions are met) then we set its index to $Q_i = 1$. If that is not the case, $Q_i = 0$. For a total of 13 fits $((90-30)/5+1)$, the short-term indicator is simply the average over these 13 windows of their index: $Short_{ind} = \frac{1}{13}\sum_{i=1}^{13} Q_i$.

- Medium-term bubble: The medium-term bubble indicator at time t_2 is a number $\in [0,1]$ which denotes the fraction of qualified fits for estimation windows of length $dt := t_2 - t_1 \in [90:300]$ business days for this t_2. For a total of 43 fits $((300-90)/5+1)$, using the same definition of the index Q_i for each of these 43 time windows, the medium-term bubble indicator is simply $Medium_{ind} = \frac{1}{43}\sum_{i=1}^{43} Q_i$.

- Long-term bubble: The long-term bubble indicator at time t_2 is a number $\in [0,1]$ which denotes the fraction of qualified fits for estimation windows of length $\in [300:745]$ business days for this t_2. For a total of 90 fits $((745-300)/5+1)$, using the same definition of the index Q_i for each of these 90 time windows, the long-term bubble indicator is simply $Long_{ind} = \frac{1}{90}\sum_{i=1}^{90} Q_i$.

4.1.4 Smoothed LPPLS Confidence$^{\mathbf{TM}}$ Multi-scale Indicators

The above defined short-term / medium term/ long-term bubble indicators exhibit significant statistical fluctuations. For the purpose of facilitating the visual interpretation of these indicators, we perform an exponential smoothing of these LPPLS confidence indicators via $AR(1)$ moving averages as follows

$$CLPPLS_{Short}(t) = \alpha_{short}\, CLPPLS_{short}(t-1) + (1-\alpha_{short}) Short_{ind}(t) \ , \tag{4.1}$$

$$CLPPLS_{Medium}(t) = \alpha_{medium}\, CLPPLS_{medium}(t-1) + (1-\alpha_{medium}) Medium_{ind}(t) \ , \tag{4.2}$$

$$CLPPLS_{Long}(t) = \alpha_{long}\, CLPPLS_{long}(t-1) + (1-\alpha_{long}) Long_{ind}(t) \ . \tag{4.3}$$

where $\alpha_{short} = 0.980, \alpha_{medium} = 0.995$ and $\alpha_{long} = 0.998$ corresponding respectively to time scales of 50, 200 and 500 days that are in synchrony with the respective time scales of the short-term / medium term/ long-term bubble indicators. In other words, given the fact that the short-term bubble indicator is constructed by using time windows of size $\in [30 : 90]$ business days, we perform a smoothing exponential averaging over the last 50 days for each t_2. Similarly, given the fact that the medium-term bubble indicator is constructed by using time windows of size $\in [90 : 300]$ business days, we perform a smoothing exponential averaging over the last 200 days for each t_2. Lastly, given the fact that the long-term bubble indicator is constructed by using time windows of size $\in [300 : 745]$ business days, we perform a smoothing exponential averaging over the last 500 days for each t_2.

The time series of these three smoothed bubble indicators, both for positive and negative bubbles, are shown in Fig. (4.1) for the period $t \in [Jan.1973 : Dec.2014]$ and for the financial time series obtained by taking the ratio of the $S\&P500$ Index divided by the capital weighted dividends of the constituting firms.

4.1.5 Predictive Tests

As mentioned earlier, a large value for a confidence indicator suggests that the LPPLS pattern is found over several time windows and is thus more reliable whereas a small value for the indicator signals a possible fragility since it is present in only a few fitting windows. Taking into account the specification of the bubble indicator in which greater values indicate the presence of a bubble, we utilize a regime switching model that incorporates market states representing bubble and non-bubble regimes. Therefore, having computed the positive and negative bubble indicators for short and long time horizons, we examine the predictive ability of short selling and liquidity-based indicators by estimating a Markov Switching predictive model specified as

$$Ind(t) = \gamma_{0,S_t} + \gamma_{1,S_t} X_{t-1} + \epsilon_t, \tag{4.4}$$

where $Ind(t)$ is either $Short_{ind}(t)$, $Medium_{ind}(t)$, or $Long_{ind}(t)$, S_t is a discrete regime variable taking values in $(0, 1)$, following a two-state Markov process and ϵ_t is the error term. X_{t-1} is a vector of the predictors measured at the end of month $t-1$. As explained in the next section, the predictive model is applied to alternative proxies for short selling activity in order to check the robustness of the findings.

Here, we stress that we use the non-smoothed indicators $Short_{ind}(t)$, $Medium_{ind}(t)$ and $Long_{ind}(t)$ as the dependent variables in (4.4), and not the smoothed ones $CLPPLS(t)$. Using the later would lead to spurious regressions and inaccurate p-values due to their build-in correlation structure.

4.2 Data and Empirical Findings

4.2.1 Data

The dataset used to construct the LPPLS Confidence$^{\text{TM}}$ Indicators includes monthly price-to-dividend (P/D) ratios for the $S\&P500$ Index over the period January 1973 through December 2014. As mentioned earlier, we focus on the predictive power of short selling and market liquidity measures over the confidence indicators representing positive and negative bubbles in the index. For this purpose, we examine various alternative proxies for each market-based predictor. Short selling activity is measured via two proxies. The first is the short interest index (SII) of [85] as an aggregate measure of short interest, constructed using firm-level short interest data. The authors argue that short sellers are informed traders and show that short interest is arguably the strongest predictor of aggregate stock returns, both in- and out-of-sample. The data is available on David Rapach's website. The second proxy for short selling activity is the short interest ratio (SIR) by [26], defined as the total number of shares sold short divided by total shares outstanding from the last month of the fiscal year. The authors show that this ratio is positively related to one-year ahead stock price crash risk. Following [26], we calculate this ratio using short interest data from Compustat.

Evidence that associates high stock market volume with periods of high market volatility has already been well-established in the literature (e.g. [66, 42, 62]). Therefore, we use liquidity as a control variable in our predictive tests in order to check the robustness of the predictive ability of short selling measures. Following a number of studies including [3] and [7], we use the stock market turnover (TURNO) as a proxy for market liquidity. We compute monthly turnover values as the number of shares traded divided by shares outstanding for all NYSE and AMEX firms from the CRSP files. Following [27], we detrend the monthly log turnover series by subtracting a one-year backward moving average of log turnover, yielding a triangular moving average of turnover growth rates.

4.2.2 Empirical Findings

Figure (4.1) presents the estimated positive and negative multi-scale LPPLS confidence^{TM} bubble indicators for the $S\&P500$ index divided by dividends. The short, medium and long-term bubble indicators are depicted in different colors and the log price-to-dividend ratio for the $S\&P500$ index is represented as the black solid line. Note that a large value for the indicator indicates that the LPPLS pattern is found for many windows in the corresponding scale range (of short-term, medium-term and long-term) and is thus more reliable. Looking at Figure (4.1), we observe remarkable "*spikes*" in the smoothed indicators at the eve of regime changes. For example, the long-term indicator successfully captures, ex-ante, all the great bubbles and subsequent crashes suffered by the $S\&P500$ index (Black-Monday - 1987, Dot-com - 2000 and Subprime - 2008) when using a threshold $\geq 50\%$. Similarly, the negative long-term indicator remarkably shows the start of a positive bubble at the beginning of 1995 where its value reaches ≈ 1. The exponential damping structure after each peak is due to the AR(1) smoothing explained in section 4.1.4.

It is also interesting to notice the number of small bubbles (green shaded region on the upper panel) permeating the bubbly period that stretches from 1994 to the burst of the dot-com bubble in 2000. Note also that throughout this period, the positive long-term indicator is ever increasing as well as the medium-term indicator, thus suggesting the maturation of the bubble towards instability across several distinct time-scales. Overall, these results support our claim that the LPPLS framework is a flexible tool for detecting bubbles across different time-scales.

Having constructed the series of positive and negative bubble indicators, we next examine their predictability using the regime-switching specification in Eq.(4.4). Table 4.2 reports the estimates for the Markov Switching model for the short-term bubble indicator, $Short_{ind}$.[3] Panels A and B report the findings when short selling activity is measured by the short interest index of [85] and the short interest ratio of [26], respectively. The two-state specification identifies two distinct regimes corresponding to bubble and non-bubble market states for each indicator series. Examining the findings for the negative bubble indicators, we see that short sellers indeed have significant predictive power over market crashes, consistent across both measures of short selling activity in Panels A and B. The model yields positive and highly significant estimates for both short selling measures, suggesting that higher level of short selling activity predicts the occurrence of negative

[3]The model is estimated using the non-smoothed bubble indicators explained in section 4.1.3 as the smoothed indicators may lead to spurious regressions and inaccurate p-values, as already mentioned.

bubbles in the short term. As expected, none of the short selling proxies have predictive significance in the case of the non-bubble regimes. Consistent with the positive coefficients observed for the negative bubble indicators, we see that the short interest ratio, defined as the total number of shares sold short divided by total shares outstanding, has predictive power over the positive bubble indicator with a negative coefficient, suggesting that higher short selling activity predicts lower occurrence for a positive bubble.

Similarly, examining the estimated coefficients for turnover, we see that market liquidity also commands significant predictive ability over both the negative and positive bubble indicators. The significant predictive power observed for turnover is consistent with the finding by [79] that market liquidity has a prevalent effect on stock bubbles and that liquidity shocks provide warning signals of impending bubble collapses. Interestingly however, the highly significant and positive estimates observed for turnover indicate that high market turnover can serve as a predictor of bubble occurrence in either direction, i.e. a booming or collapsing market condition. [96] also highlights the connection between available liquidity and rising asset prices. His argument stresses that strong balance sheets induce banks to increase their lending which, In turn, raises asset prices, leading to stronger balance sheets and so forth.

The findings for the medium- and long-term bubble indicators reported in Tables 4.3 and 4.4 further confirm the predictive power of short selling proxies over dropping markets (i.e. negative bubbles) across both the short and long horizons. We observe highly significant and positive coefficient estimates for both short selling proxies in the models for the negative bubble indicator, suggesting that short selling activity predicts greater occurrences of negatively trending markets over both short and long time scales. This finding is not inconsistent with [26] who document that short interest is positively related to one-year ahead stock price crash risk. To that end, our results confirm short sellers' predictive ability over developing market loss risks which are successfully captured by our implementation of the LPPLS framework presented in this study.

Interestingly, however, while market turnover retains its predictive ability in the medium term, we observe that the sign of the estimated coefficients for turnover in Table 4.3 flips to negative, suggesting that high turnover predicts lower occurrence of bubbles (in either direction) in the medium term. Similarly, in the case of the long term bubble indicator reported in Table 4.4, turnover loses its significance for the negative bubble indicator. These observations suggest that market liquidity has only a transient and relatively short-term impact on prices, which is detected for the short-term indicator, but not for the longer time scales of the bubble indicators, which are themselves more robust to detect the overall bubble sentiment at long-time scales.

Overall, our findings suggest that market-based indicators can indeed be utilized to predict the occurrence of market booms and collapsing market regimes, implied by the significant predictive ability observed for short selling proxies for negative bubble regimes across both the short and long horizons. On the other hand, market liquidity is found to predict the occurrence of both decreasing and booming market conditions while its predictive power is limited to shorter time horizons. These findings are encouraging news for market regulators as the results show that short selling proxies can be used to model and monitor negative bubble market conditions, while market liquidity can be used to supplement forecasting models for both boom and bust market conditions.

4.3 Conclusion

This paper has examined the predictability of stock market booms and crashes via an application of the LPPLS Confidence$^{\text{TM}}$ Multi-scale Indicators to the $S\&P500$ index.

First, we presented a methodology to detect positive and negative bubbles for the $S\&P500$ index using the Log-Periodic Power Law Singularity (LPPLS) model ([56, 110]), something not possible by other bubble detection models. Next, we provided insight to the predictability of market booms and crashes using market-based indicators by examining the predictive power of short selling activity and market liquidity over the constructed bubble indicators. To the best of our knowledge, this is the first attempt in examining the predictive ability of market-based indicators over the positive and negative bubbles in the $S\&P500$ index using the LPPLS model.

Our findings suggest that the LPPLS framework is able to successfully capture, ex-ante, some of the prominent bubbles across different time scales. We show that some of the great bubbles and subsequent crashes experienced during the Black Monday, Dot-com, and Subprime Crisis periods are successfully captured by the constructed bubble indicators. Our predictive tests indicate that measures of short selling activity have robust predictive power over negative bubbles, in line with the previous studies that short sellers have predictive ability over stock price crash risks. The predictive ability of short selling activity is robust to alternative measures of short selling as well as to short and long time horizons, consistent with the recent finding by [26] that short interest is positively related to one-year ahead stock price crash risk. On the other hand, our tests show that market liquidity has robust predictive power over both the negative and positive bubbles, however in the short-term, suggesting that market liquidity measures can be used to predict the occurrence of both booms and collapses for short horizons. We have thus identified short selling and liquidity as two important factors contributing to the LPPLS-based bubble indicators. The evidence overall points to the predictability of both positive and negative stock market bubbles via market-based proxies of trading activity and can be used as a guideline to model and monitor bubble conditions in stock markets.

Figure 4.1: Positive (upper panel) and negative (lower panel) multi-scale LPPLS Confidence bubble indicator. The black continuous line denotes the logarithm of the monthly Price over Dividend (P/D) time-series for the $S\&P500$ Index from January 1973 to December 2014. The short, medium and long-term bubble indicators are depicted in green, magenta and red respectively. We refer to Sec. (4.1) for the construction of the indexes.

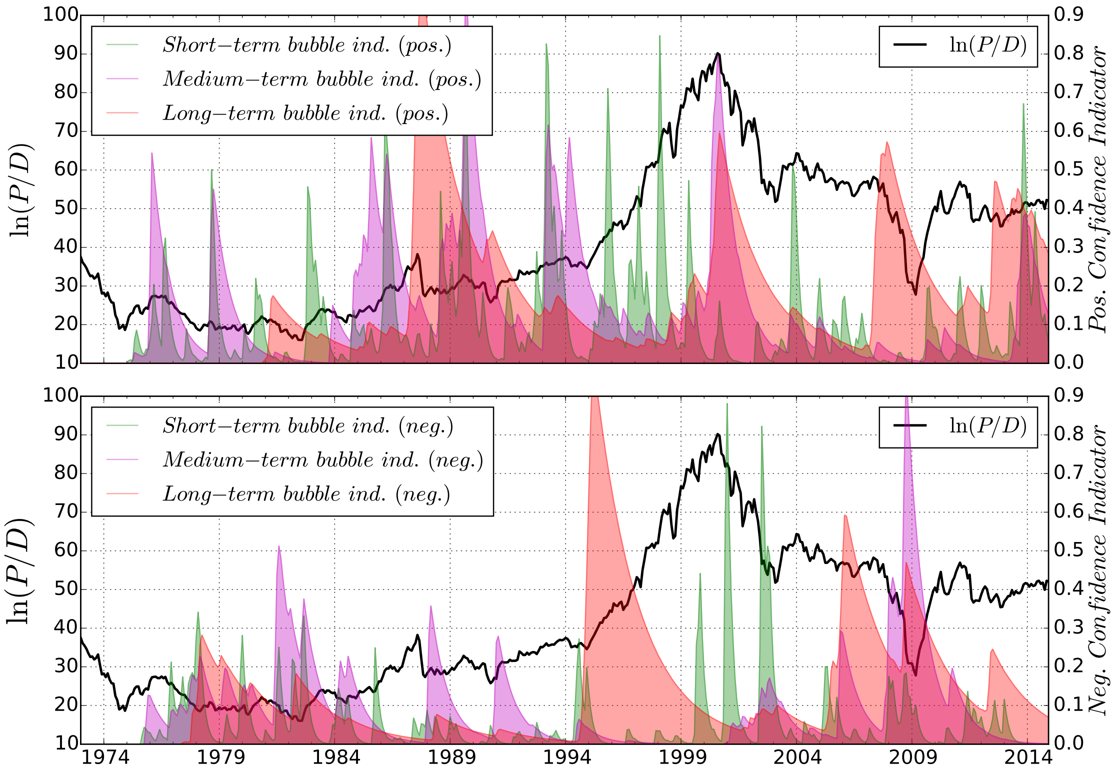


Table 4.1: Search space and filter conditions for the qualification of valid LPPLS fits. Within the JLS framework, the condition that the crash hazard rate $h(t)$ is non-negative by definition translates into a value of the Damping parameter $\frac{m|B|}{\omega|C|}$ larger than or equal to 1.

Item	Notation	Search space	Filtering condition 1	Filtering condition 2
3 nonlinear parameters	m	[0, 2]	[0.01, 1.2]	[0.01, 0.99]
	ω	[1, 50]	[6, 13]	[6, 13]
	t_c	$[t_2 - 0.2dt,$ $t_2 + 0.2dt]$	$[t_2 - 0.05dt,$ $t_2 + 0.1dt]$	$[t_2 - 0.05dt,$ $t_2 + 0.1dt]$
Number of oscillations	$\frac{\omega}{2} \ln \lvert \frac{t_c - t_1}{t_2 - t_1} \rvert$	—	$[2.5, +\infty)$	$[2.5, +\infty)$
Damping	$\frac{m\lvert B\rvert}{\omega\lvert C\rvert}$	—	$[0.8, +\infty)$	$[1, +\infty)$
Relative error	$\frac{p_t - \widehat{p}_t}{\widehat{p}_t}$	—	[0, 0.05]	[0, 0.2]

Table 4.2: The predictive ability of short interest and turnover on the *short–term* bubble indicator, $Short_{ind}(t)$.

Panel A: Short selling activity measured by the short interest index of Rapach et al. (2016)				
	Negative Bubble		**Positive Bubble**	
	Regime 1 (bubble regime)			
Constant	0.2988***	(0.0122)	0.4502***	(0.0131)
Short interest index, SII	0.0870***	(0.0122)	0.0167	(0.0127)
Turnover	0.5402*	(0.2317)	2.1897***	(0.2292)
	Regime 2 (non-bubble regime)			
Constant	0.0082***	(0.0012)	0.0099***	(0.0026)
Short interest index, SII	0.0007	(0.0012)	-0.001	(0.0029)
Turnover	-0.0167	(0.0167)	0.0838**	(0.0365)
AIC	-4.211		-2.593	
log L	1067.960		661.050	
Panel B: Short selling activity measured by the short interest ratio of Callen and Fang (2015)				
	Negative Bubble		**Positive Bubble**	
	Regime 1 (bubble regime)			
Constant	0.1235***	(0.0241)	0.5046***	(0.0199)
Short interest ratio, SIR	9.9473***	(1.7216)	-2.0706***	(0.5316)
Turnover	0.5620***	(0.2010)	2.0469***	(0.2273)
	Regime 2 (non-bubble regime)			
Constant	0.00650***	(0.0018)	0.0137***	(0.0037)
Short interest ratio, SIR	0.071153	(0.0478)	-0.129462	(0.0966)
Turnover	-0.016927	(0.0180)	0.0815**	(0.0356)
AIC	-4.136		-2.616	
log L	1042.885		662.960	

Note: This table reports the estimates for the Markov Switching model specified in Eq.(4.4). Market liquidity is measured by stock market turnover, computed as the number of shares traded divided by shares outstanding for all NYSE and AMEX
firms from the CRSP files. Following Campbell et al. (1993), we detrend the log turnover series by subtracting a one-year backward moving average of log turnover. Panels A and B report the findings for when short selling activity is measured by the short interest index of Rapach et al. (2016) and the short interest ratio of Callen and Fang (2015), respectively. The numbers in parentheses are the standard errors. ***, **, and * represent significance at 1, 5, and 10 percent, respectively.

Table 4.3: The predictive ability of short interest and turnover on the *medium–term* bubble indicator, $Medium_{ind}(t)$.

Panel A: Short selling activity measured by the short interest index of Rapach et al. (2016)				
	Negative Bubble		**Positive Bubble**	
	Regime 1 (bubble regime)			
Constant	0.1613***	(0.0100)	0.571***	(0.0164)
Short interest index, SII	0.0421***	(0.0065)	-0.0303*	(0.0171)
Turnover	-0.5756***	(0.0806)	-2.6188***	(0.2745)
	Regime 2 (non-bubble regime)			
Constant	0.0046***	(0.0009)	0.0101***	(0.0018)
Short interest index, SII	0.0008	(0.0009)	-0.0021	(0.0018)
Turnover	-0.0239*	(0.0126)	-0.0123	(0.0291)
AIC	-4.840		-3.389	
log L	1226.355		861.355	
Panel B: Short selling activity measured by the short interest ratio of Callen and Fang (2015)				
	Negative Bubble		**Positive Bubble**	
	Regime 1 (bubble regime)			
Constant	0.107***	(0.0099)	0.6521***	(0.0184)
Short interest ratio, SIR	1.9061***	(0.1914)	-5.8897***	(0.4582)
Turnover	-0.7668***	(0.1086)	-2.3249***	(0.1473)
	Regime 2 (non-bubble regime)			
Constant	0.0036***	(0.0012)	0.0108***	(0.0026)
Short interest ratio, SIR	0.0242	(0.0304)	-0.0576	(0.0687)
Turnover	-0.0202*	(0.0116)	-0.0194	(0.0255)
AIC	-4.958		-3.473	
log L	1248.589		877.131	

Note: This table reports the estimates for the Markov Switching model specified in Eq.(4.4). Market liquidity is measured by stock market turnover, computed as the number of shares traded divided by shares outstanding for all NYSE and AMEX
firms from the CRSP files. Following Campbell et al. (1993), we detrend the log turnover series by subtracting a one-year backward moving average of log turnover. Panels A and B report the findings for when short selling activity is measured by the short interest index of Rapach et al. (2016) and the short interest ratio of Callen and Fang (2015), respectively. The numbers in parentheses are the standard errors. ***, **, and * represent significance at 1, 5, and 10 percent, respectively.

Table 4.4: The predictive ability of short interest and turnover on the *long–term* bubble indicator, $Long_{ind}(t)$.

Panel A: Short selling activity measured by the short interest index of Rapach et al. (2016)				
	Negative Bubble		**Positive Bubble**	
	Regime 1 (bubble regime)			
Constant	0.5143***	(0.0128)	0.5297***	(0.0111)
Short interest index, SII	0.0379***	(0.0128)	0.0152	(0.0085)
Turnover	-0.3156	(0.2553)	-1.4654***	(0.1753)
	Regime 2 (non-bubble regime)			
Constant	0.0069***	(0.0020)	0.0092	(0.0020)
Short interest index, SII	0.0014	(0.0019)	-0.002	(0.0019)
Turnover	-0.0217	(0.0275)	0.0074	(0.0228)
AIC	-3.186		-3.217	
log L	810.157		818.126	
Panel B: Short selling activity measured by the short interest ratio of Callen and Fang (2015)				
	Negative Bubble		**Positive Bubble**	
	Regime 1 (bubble regime)			
Constant	0.4347***	(0.0203)	0.5447***	(0.0158)
Short interest ratio, SIR	2.516***	(0.4817)	-0.3782	(0.3294)
Turnover	0.2515	(0.2661)	-1.3063***	(0.1675)
	Regime 2 (non-bubble regime)			
Constant	0.0039	(0.0030)	0.0062**	(0.0029)
Short interest ratio, SIR	0.0908	(0.0768)	0.1091	(0.0770)
Turnover	-0.0139	(0.0269)	0.0063	(0.0282)
AIC	-3.202		-3.208	
log L	809.589		811.064	

Note: This table reports the estimates for the Markov Switching model specified in Eq.(4.4). Market liquidity is measured by stock market turnover, computed as the number of shares traded divided by shares outstanding for all NYSE and AMEX
firms from the CRSP files. Following Campbell et al. (1993), we detrend the log turnover series by subtracting a one-year backward moving average of log turnover. Panels A and B report the findings for when short selling activity is measured by the short interest index of Rapach et al. (2016) and the short interest ratio of Callen and Fang (2015), respectively. The numbers in parentheses are the standard errors. ***, **, and * represent significance at 1, 5, and 10 percent, respectively.

Chapter 5

Modified Profile Likelihood and Interval Forecast of the End of Financial Bubbles

5.1 Introduction

Financial bubbles and their subsequent crashes provide arguably the most visible departures from well-functional efficient markets. There is an extensive literature (see e.g. the reviews of [64][1],[55, 25, 125]) on the causes of bubbles as well as the reasons for bubbles to be sustained over surprising long period of times. One of these views emphasises the role of herding behaviour on bubble inflation [58]. When imitation is sufficiently strong, a high demand for the asset pushes the price upwards, which itself, and somewhat paradoxically, increases the demand, propelling further the price upward, and so on, in self-fulfilling positive feedback loops. In such regimes, the market is mainly driven by sentiment and becomes

[1]Long version at `http://arXiv.org/abs/0812.2449`

detached from any underlying economic value. This process is intrinsically unsustainable and the mispricing ends at a critical time, either smoothly (with a correction phase) or abruptly (via a crash). The formulation of this hypothesis of collective herding behavior within rational expectations theory resulted in the so-called Log-Periodic Power-Law Singularity (LPPLS) model, which has been used for many successful ex-post and ex-ante predictions of bubble bursts (see e.g. a partial list in [116] and a recent implementation for the Chinese bubble and its burst in 2015 [112]).

Notwithstanding a number of improvements concerning the calibration of the LPPLS model, including meta-search heuristics [117] and reformulation of the equations to reduce the number of nonlinear parameters [38], the calibration of the LPPLS model remains a bottleneck towards achieving robust forecasts and a matter of contention [20, 116]. In this context, the aim of the present paper is to present a fundamental revision of the calibration procedure of the LPPLS model. Specifically, we deviate from the traditional ordinary least squares (OLS) calibration that provides *point estimates* of parameters, which has been used since the introduction of the model in 1999 [56, 57]. Instead, we employ a rigorous likelihood approach and, for the first time to the best of our knowledge, we provide *interval estimates* of the parameters, including the most important critical times of market regime changes.

We deliberately avoid dwelling on the derivation of the model and its foundations, and take it as given. We do not discuss supporting evidence and critiques of the model, nor address how to apply the LPPLS model to construct robust signals for extensive backtests or real-time ex-ante predictions. These questions require extensive analyses and are beyond the scope of the present manuscript. See [60, 55, 116, 112, 127] for investigations in these directions.

The purpose of the present chapter is methodological, and the main focus is on the statistical aspects of the theory and the corresponding mathematical derivations. One of the major advances of this paper is to formulate the calibration procedure so that the critical time t_c is the major parameter of interest in the likelihood inference, while other model parameters are treated as so-called *nuisance parameters*. Of course, these other parameters are also intrinsic to the model but their existence contributes to the variance of the parameter of key interest. Such reformulation of the calibration procedure has its roots in an original idea proposed by [38], which was however developed in a crude way and without the proper statistical methodology.

The problem of dealing with nuisance parameters and of quantifying their impact on the uncertainty of the parameter of interest is not new in Statistics. However, to our knowledge, it has not been elaborated before in quantitative finance. Frequentist and Bayesian statistical schools have different views on this problem. The main debate between the supporters of likelihood-based versus Bayesian approaches is whether one should maximize over nuisance parameters (such as in simple profile likelihood) or integrate them out. Both approaches have their pros and cons. In general, the method of profile likelihood is known to often provide biased estimations. However, use of the Bayesian (or integration) approach requires specification of the prior distribution of the parameters, which leads to an extra uncertainty in inference. Under certain conditions, when the full likelihood function has a complex structure, the two methods can lead to dramatically different estimations [100, 16].

We will base our approach on the so-called modified profile likelihood proposed by [11] as a higher-order approximation to either a marginal or conditional likelihood function. Being unable to calculate the modified profile likelihood exactly due to strong model nonlinearity, we will employ the approximation suggested by [88], which is equivalent to the exact form up to errors of order $\mathcal{O}(n^{-1})$ for moderate deviations and of order $\mathcal{O}(n^{-1/2})$

in the large deviation sense, where n is the number of data points. The advantage of this method is that it takes the middle ground in the maximization-vs-integration debates: as shown by [92], the modified profile likelihood arises naturally from a non-Bayesian inference with an integrated likelihood and could even be considered as an approximation to a certain class of integrated likelihood functions. At the same time, it does not require specifying a prior density of the nuisance parameters, which makes it perfectly suitable in our case.

In the following, we will guide the reader from the well known OLS calibration procedure and its formulation as a likelihood problem, to the lesser known "profile likelihood" and then to the "modified profile likelihood", which has been essentially ignored in the applied literature. The modified profile likelihood allows one to improve the likelihood inference by accounting for the uncertainty of the nuisance parameters. Having a strong methodological emphasis, we will discuss all concepts and, more important, their assumptions and limitations in all necessary details. While this paper focuses on the LPPLS model, our general presentation and its specific implementation on the LPPLS model makes it useful as a general guide for likelihood inference in many other models of quantitative finance.

The paper is organized as follows. Section 5.2 presents the Ordinary Least Squares (OLS) method that has been used until now as the standard calibration tool of the LPPLS model, in particular for the estimation of the critical time t_c of the end of the bubble. Section 5.3 introduces the Likelihood and Profile Likelihood approaches. Section 5.4 presents the general concept of the modified profile likelihood and provides a very useful approximated expression for it. Parameter estimation uncertainties and the corresponding likelihood intervals are then derived. Section 5.5 applies the modified likelihood profile to estimate confidence intervals of the nuisance parameters m and ω as well as the damping variable. Section 5.6 presents the method of aggregation of the calibrations from different scales and illustrates the whole methodology on synthetic price time series. This section ends with the application of the method on three well-known historical financial bubbles. Section 5.7 concludes.

5.2 Nonlinear regression and Ordinary Least Squares fitting

5.2.1 Estimation of the critical time

In practical applications, the calibration of the LPPLS model often aims at forecasting the critical time t_c, because it is, by construction of the LPPLS model, the end of the bubble regime. This suggests to develop a special treatment for t_c. In this spirit, [38] suggested to reformulate the optimization problem (1.13) by subordinating the logperiodic angular frequency ω and power law exponent m to t_c:

$$\hat{t}_c = \arg\min_{t_c} F_2(t_c), \tag{5.1}$$

where

$$F_2(t_c) = \min_{\omega,m} F_1(t_c, m, \omega), \quad \{\hat{m}(t_c), \hat{\omega}(t_c)\} = \arg\min_{m,\omega} F_1(t_c, m, \omega) \tag{5.2}$$

and $F_1(t_c, m, \omega)$ is given by (1.12).

In general, such extra subordination dramatically reduces the number of local extrema of the cost-function. As we will see later from Figure 5.4, when the price trajectory displays a pronounced increase, the function $F_1(t_c, m, \omega)$ almost always presents just one minimum along the m direction and 3-4 local minima along the ω direction in the range $2 < \omega < 20$ (which may actually be relevant to capture higher harmonics of the logperiodicity structure [129, 128]). Further, this method allows one to avoid sloppy directions in the

(t_c, ω) plane, where the cost-function has a very long valley along the diagonal $t_c \sim \omega$, as illustrated in Figure 3b of [38] and in Figure 1.1.

At the expense of a small increase of computational complexity, beyond its simplification, the cost-function given by equation (5.2) provides a substantial improvement in inference from the model. Namely, in addition to the point estimate (5.1), expression (5.2) allows one to analyze the whole profile of the cost function $F_2(t_c)$ and the dependence of the estimates $\hat{m}$ and $\hat{\omega}$ as a function of the critical time t_c. In particular, one can identify all the extrema of $F_2(t_c)$ and their corresponding $m(t_c)$ and $\omega(t_c)$, from which expert judgment of the plausible scenarios can follow.

As an example, we consider the recent bubble and following collapse of the Chinese market, when the Shanghai Composite Index (SSE Composite) appreciated by approximately 150% between mid-2014 and mid-2015, peaked on June 12, 2015 and then lost 32% to its first well-defined bottom reached on July 8, 2015. This bubble was detected by the Financial Crisis Observatory (FCO) at ETH Zürich and further documented and dissected in [112]. We use data provided by Thomson Reuters Dataworks Enterprise (see Section 5.6.3 for discussions). Figure 5.1 presents the dynamics of the SSE Composite index together with the best LPPLS fit according to the OLS regression within the time window of $t_2 - t_1 = 180$ calendar days ending at the date of $t_2 =$June 12, 2015 when the market peaked.

In order to understand the "microstructure" of LPPLS fits, we employ the three-step subordination procedure (1.10)–(1.12), (5.1)–(5.2) and study the dependence of the cost-function $F_2(t_c)$ as well as $\hat{m}(t_c)$, $\hat{\omega}(t_c)$ and damping $\hat{D}(t_c)$ (see Fig. 5.2). One can see that the global (best) solution with estimated critical time $\hat{t}_c$ of July 7, 2015 (with $F_2 = 0.0597$) is not the only minimum, and a second local minimum is found at $t_c =$ June 18, 2015 (with $F_2 = 0.0604$), which suggests a second plausible scenario. Despite almost

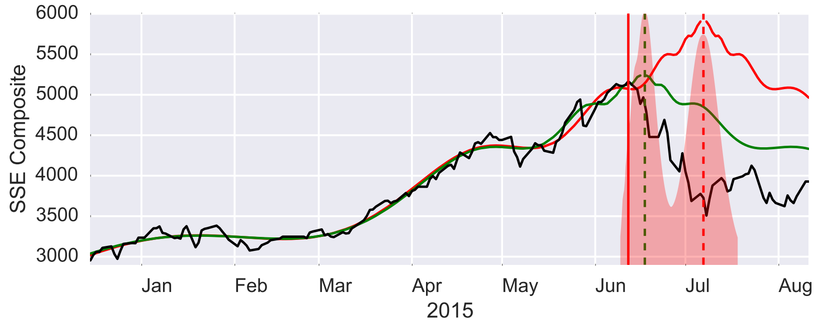


Figure 5.1: Price trajectory of the SSE Composite Index during the bubble of 2014–2015. The red vertical line denotes the date of the analysis (t_2 = 2015-06-12). Red and green solid lines correspond to the best and second best (see Figure 5.2) LPPLS fit in the window [2014-12-15, 2015-06-12] and their extrapolations to $t > t_2$. The vertical red and green dashed lines indicate the position of the critical times t_c for these two fits: 2015-07-08 and 2015-06-18 respectively. The shaded red areas delineate the likelihood interval of t_c at a 5% cutoff together with the shape of the modified profile likelihood (see Sections 5.4–5.4.4 and Figure 5.3).

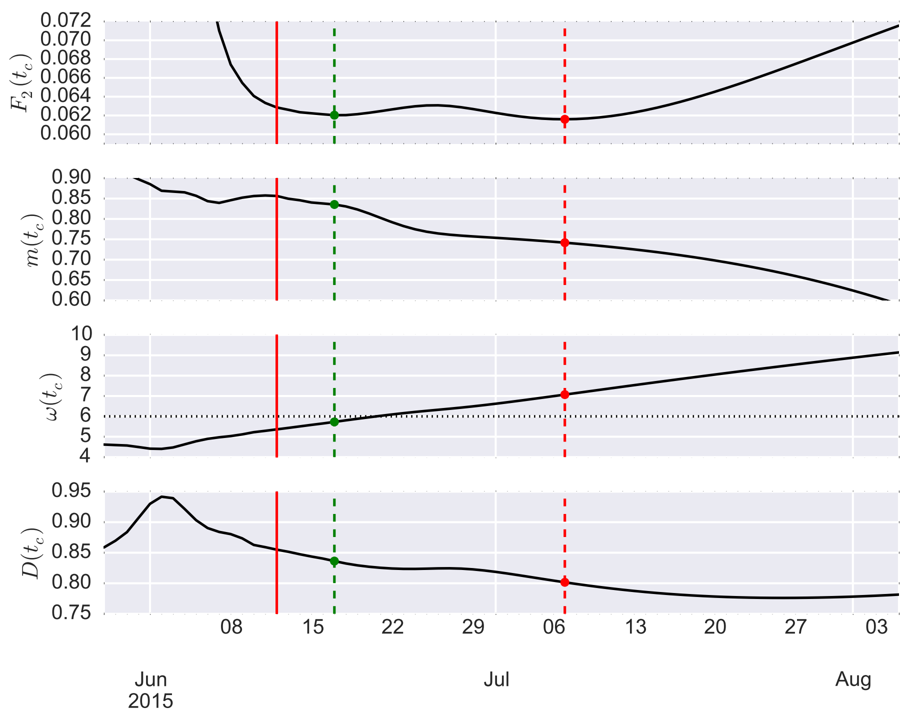


Figure 5.2: Profiles of the cost function $F_2(t_c)$ and parameters $\hat{m}$, $\hat{\omega}$ and damping $\hat{D} = m|B|/\omega|C|$ as a function of t_c. The solid red vertical line indicates the date of analysis (t_2 = 2015-06-12), the red and green dotted vertical lines correspond to the dates of the best and alternative solutions (2015-07-08 and 2015-06-18 respectively). The horizontal dotted line gives the level of the threshold (1.2.1) for the logperiodic angular frequency parameter ($\omega > 6$) that separates “qualified” fits from “not qualified” ones. The filled red and green circles show the point estimates of the model parameters — for the best and alternative fits, respectively.

identical values of the cost functions (sum of squared errors of residuals), we might reject the suboptimal solution on the basis of the fact that its logperiodic angular frequency falls outside of the empirical constraint (1.2.1) ($\hat{\omega} = 7.18$ for the optimal solution and $\hat{\omega} = 5.85$ for the suboptimal). Both solutions are associated with damping parameters that are below the constraint $D \geq 1$ ($\hat{D} = 0.8$ for the optimal and $\hat{D} = 0.83$ for the suboptimal solution), but they are both compatible with the relaxed constraint (1.2.1) (note that the value for the optimal solution is very close to the boundary of this constraint).

This case study exemplifies the essence of the problem of dealing with multiple and almost equivalent optimal solutions that point to quite different future scenarios. Above we have invoked previous experience [60] to reject the second scenario. However, this is not fully satisfactory from a theoretical view point. Moreover, past experience can be tainted by the use of the sub-optimal calibration procedure based on the original formulation of the model (1.4). To boot, past experience may not contain all possible situations, and surprises that are superficially of the "unknown unknown" type [69, 120] from the point of view of past experience might actually be understandable and knowable with the appropriate conceptual and theoretical framework [111].

The question we further investigate below is: How can we resolve between these two scenarios if we do not have (or do not want or trust to use) any prior information on what are plausible parameter values? In other words, how can we provide a quantitative estimation of how much one scenario is less likely than another?

5.3 Likelihood and Profile Likelihood

The OLS regression (1.10) represents the so-called *normal estimation* of the model parameters, i.e. provides *Maximum Likelihood Estimates (MLE)* under the assumption that the error term $\varepsilon(\tau_i; t_c, \psi)$ is normally distributed. The likelihood has then a well-known form:

$$L(t_c, \psi, s) = (2\pi s)^{-n/2} \exp\left(-\frac{\mathrm{SSE}(t_c, \psi)}{2s}\right) \to \max_{t_c, \psi, s}, \tag{5.3}$$

where $s = \sigma^2$ is a variance of the residuals $\varepsilon(\tau_i; t_c, \psi)$ and n is the number of data points. By definition the likelihood is meaningful only up to an arbitrary positive constant, thus below we will omit such constant pre-factors. The MLE of the parameters $\{\hat{t}_c, \hat{\psi}\}$, where $\psi = \{m, \omega, A, B, C_1, C_2\}$, is obtained straight from (5.3): considering the logarithm of the likelihood ($\ln L(t_c, \psi, s)$), one immediately arrives at (1.10) and an estimate for σ^2 is

$$\hat{\sigma}^2 \equiv \hat{s} = \frac{1}{n}\mathrm{SSE}(\hat{t}_c, \hat{\psi}). \tag{5.4}$$

Despite the equivalence of the MLE and OLS approaches in terms of computations, the MLE requires an explicit distributional assumption for the error term ε. This implies that the inference of ψ is implicit with the likelihood approach, while further work with some sampling method is needed in the least squares approach.

As discussed above, we are mostly interested in the inference of the critical time t_c while the other parameters $\eta = \{\psi, s\} \equiv \{m, \omega, A, B, C_1, C_2, s\}$ can be considered as *nuisanse* parameters that are useful insofar that they allow to adapt the model to the variability of the data. The elimination of nuisance parameters is a well-known statistical problem, which amounts to concentrating the likelihood around a single parameter of interest while accounting for the extra uncertainty resulting from the estimation of the nuisance parameters. Unfortunately, there is no technique that is efficient for all situations [15], in particular because it is not always meaningful to discuss the uncertainty in one parameter independently from that of all others.

As already mentioned, in the Bayesian approach, the elimination of the nuisance parameters corresponds to integrating them out. Its proper implementation requires specifying the prior distribution of all parameters $\{t_c, \eta\}$, calculating the posterior and then integrating out the nuisance parameters η from the posterior to derive the posterior marginal distribution of t_c. The major limitation of the Bayesian approach is indeed a specification of the prior. We will not pursue this way directly. However, as shown in Section 5.4, we will be able to capture the idea of integration over the nuisance parameters within a non-Bayesian framework.

One commonly used practice of elimination of nuisance parameters is based on a factorization of the complete likelihood into a product of the so-called *marginal* and *conditional likelihood functions* [65]. When available, this approach results in a genuine likelihood, i.e. the genuine probability of the observed data conditional on the parameter of interest (t_c). However, this approach requires transforming the sufficient statistics into a minimal sufficient statistics that has to be factored into two terms T and A. One of these terms, either the marginal distribution of T or the conditional distribution of T conditioned on A (which is then called *ancillary* for t_c), depends only on t_c, but not on η (see discussions in [82] and for example [14, 89]). Given that the LPPLS model (1.6) is highly nonlinear, it is not possible to find such factorization.

A simpler method is to construct the so-called *profile likelihood* (also known as a *concentrated likelihood* in the econometric applications). It consists in replacing the nuisance parameters by their MLE at each fixed value of the parameter of interest. Given the joint likelihood $L(t_c, \eta)$, the profile likelihood $L_p(t_c)$ is defined as

$$L_p(t_c) = \max_{\eta} L(t_c, \eta) \equiv L(t_c, \hat{\eta}_{t_c}), \tag{5.5}$$

where $\hat{\eta}_{t_c} = \arg\max_{\eta} L(t_c, \eta)$ is a MLE for η for a fixed value of t_c. The profile likelihood is often treated as a regular likelihood for further inference of t_c, i.e. one can normalize it, compute likelihood intervals or compare likelihood ratios.

The profile likelihood approach is technically identical to the analysis of the profile cost function $F_2(t_c)$ discussed in Section 5.2. Indeed, the MLE of $\hat{\psi}_{t_c}$ is given by the solution of the OLS (1.10): the estimates of $\hat{m}_{t_c}$ and $\hat{\omega}_{t_c}$ are derived from (5.2) where $\hat{A}_{t_c}, \hat{B}_{t_c}, \hat{C}_{1,t_c}, \hat{C}_{2,t_c}$ are given by (1.13). Finally, the form of $\hat{s}_{t_c}$ is similar to (5.4), where $\hat{t}_c$ is replaced by t_c. Moreover, the value of $L_p(t_c)$ can be directly derived from $F_2(t_c)$. Indeed, the estimation of $\hat{s}_{t_c}$ defined in (5.4) can be represented as

$$\hat{s}_{t_c} = \frac{1}{n}\mathrm{SSE}(t_c, \hat{\psi}) \equiv \frac{1}{n}F_2(t_c), , \tag{5.6}$$

and, after plugging (5.6) to (5.3) according to (6.6) we obtain:

$$L_p(t_c) \propto \left(\hat{s}_{t_c}\right)^{-n/2} \propto \Big(F_2(t_c)\Big)^{-n/2}, \tag{5.7}$$

where we have omitted all constant terms. Since the likelihood (5.3) is meaningful only up to a constant, one usually considers the *relative likelihood* (respectively, *relative profile likelihood* or *relative modified profile likelihood* that will be defined later), which is normalized to 1 by its maximum and thus takes value in $[0, 1]$:

$$R(t_c) = \frac{L(t_c)}{\max_{t_c} L(t_c)}. \tag{5.8}$$

Figure 5.3 (blue curves) presents an example of the relative profile likelihood $R(t_c)$ calculated for the case discussed in Section 5.2 and presented in Figure 5.2. One can observe the same two extrema found with function $F_2(t_c)$, which correspond to very close values of the likelihood ($R_p(t_c^{(1)}) = 1$ for the best solution $t_c^{(1)} = $ 2015-07-08 and $R_p(t_c^{(2)}) = 0.64$ for

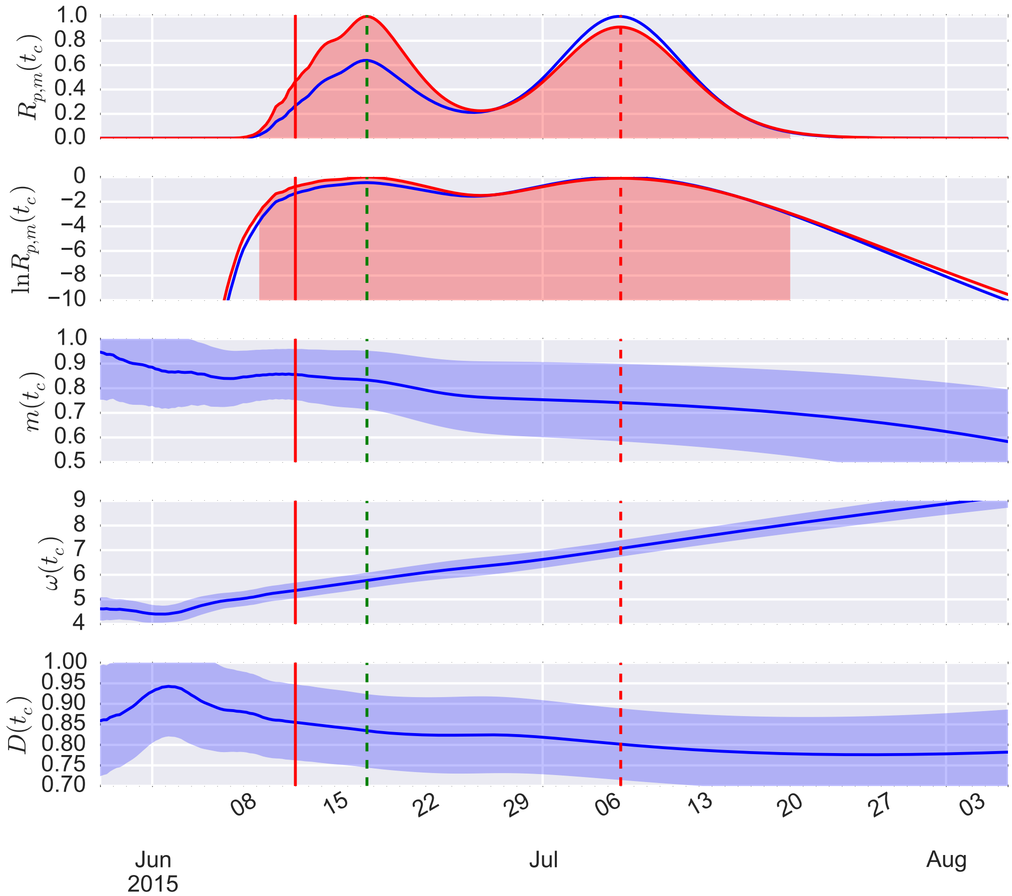


Figure 5.3: The top panel shows the relative profile likelihood (blue) and modified profile likelihood (red) as a function of the critical time t_c. The second panel from the top shows the corresponding log-likelihoods. The red shaded area corresponds to the likelihood interval of t_c at the 5% probability level (see Section 5.4.4). The three bottom panels give the point MLE parameter estimates of the model, m, ω and D, as a function of the critical time t_c (same as in Figure 5.2). Blue lines present point MLE estimates of the parameters and blue shaded areas correspond to their approximated likelihood intervals at the 5% probability level (see Section 5.5). The red continuous vertical line indicates the date of analysis ($t_2 =$ 2015-06-12). The dashed red and green vertical lines correspond to the dates of the best and alternative solutions (2015-07-08 and 2015-06-18 respectively).

the alternative solution $t_c^{(2)} =$ 2015-06-18). The likelihood ratio $R_p(t_c^{(1)})/R_p(t_c^{(2)}) = 1.56$ is not large enough to warrant preferring one maximum over the other. The inference based on a point OLS (or MLE) estimate can thus be quite misleading. In fact, the interval of "acceptable" values for t_c (the likelihood interval to be discussed in Section 5.4.4) is very broad, which confirms that a point estimation is far from reflecting the full picture.

5.4 Modified Profile Likelihood

5.4.1 General form of the modified profile likelihood

As discussed above, the profile likelihood is often treated as a regular likelihood but in fact it is not a genuine likelihood function. Specifically, it treats the nuisance parameters at a fixed value $\hat{\eta}_{t_c}$ as if they were known. It may thus overstate the amount of information about t_c and the inference on t_c based on $L_p(t_c)$ may be grossly misleading if the data contain insufficient information about η (in particular when η is high-dimensional as in our case, which can lead to an *overprecise* profile likelihood). Moreover, under certain conditions, the profile likelihood can provide unstable estimates with respect to small changes in the observed data. At the same time, more robust marginal and conditional likelihoods are not available in cases like ours.

In order to overcome this fundamental limitation of the profile likelihood, a series of adjusted versions have been proposed (see for instance, [29, 40, 13, 33] for general discussions). Most of them require orthogonality between the parameter of interest (t_c) and nuisance parameter (η). In our case, orthogonality does not hold and, in order to come up with a parametrization $\tilde{\eta}$ that would be orthogonal to t_c, one needs to solve a system of differential equations [29], which is nearly impossible to do analytically in our multi-dimensional non-linear case. Then, the most flexible approach is arguably the one proposed by [11], who introduced the so-called *modified profile likelihood* as a higher-order approximation to either a marginal or a conditional likelihood function (both derivations are possible).

The modified profile likelihood amounts to introducing an extra modulating factor

$M(t_c)$ to the profile likelihood:

$$L_m(t_c) = M(t_c)L_p(t_c) = |I(\hat{\eta}_{t_c})|^{-1/2} \left| \frac{\partial \hat{\eta}}{\partial \hat{\eta}_{t_c}} \right| L_p(t_c), \tag{5.9}$$

where $\hat{\eta}_{t_c}$ is a MLE of the nuisance parameters η at a fixed value of t_c; $I(\hat{\eta}_{t_c})$ is the corresponding *observed Fisher information* matrix on η assuming that t_c is known:

$$I(\hat{\eta}_{t_c}) = - \left. \frac{\partial^2 \ln L(t_c, \eta)}{\partial \eta \partial \eta^T} \right|_{\eta = \hat{\eta}_{t_c}}, \tag{5.10}$$

where η^T stands for the transpose of η; $\partial \hat{\eta} / \partial \hat{\eta}_{t_c}$ denotes a matrix of the first partial derivatives of the full MLE of the nuisance parameters η with respect to the MLE calculated at a fixed value of t_c; finally, $| \cdot |$ denotes the absolute value of a matrix determinant. Here and in the following, we assume that the parameters form a column vector, thus second order derivatives of the form (5.10) define a matrix.

The term $|I(\hat{\eta}_{t_c})|^{-1/2}$, which describes the curvature of the likelihood, can be considered as a penalty that subtracts from the profile log-likelihood "undeserved" information due to the estimation of the nuisance parameter η. And the Jacobian term $J(t_c) = |\partial \hat{\eta} / \partial \hat{\eta}_{t_c}|$ is needed to make the modified profile likelihood invariant with respect to the transformations of the nuisance parameters [82]. Practically, this term is extremely difficult to evaluate as its calculation requires writing down the MLE of the nuisance parameter $\hat{\eta}$ as a function of $\hat{\eta}_{t_c}$ and some ancillary statistic. In the general case, finding a suitable ancillary statistic is highly non-trivial, which dramatically hinders the analytical derivation of the correction term and thus limits the direct application of (5.9) .

Though the modified profile likelihood is not a genuine likelihood function, it is a rather good approximation to it, and it possesses a number of important properties that distinguishes it from the profile likelihood. First, its score function is asymptotically unbiased, having bias and information bias of order of $\mathcal{O}(n^{-1})$, i.e. the first and second Bartlett identities are satisfied up to the first order [33]. In contrast, for the regular profile likelihood, this is not true: these biases do not vanish and are typically of the order of $\mathcal{O}(1)$. In particular, the failure of the second Bartlett identity for the profile likelihood means that the log-likelihood ratio statistic based on the profile likelihood will not be asymptotically chi-square distributed as the standard log-likelihood ratio [67]. Second, as just mentioned, due to the Jacobian term, $L_m(t_c)$ is invariant with respect to a reparametrisation of the model (such as the variable change (1.5)). Further, the modified profile likelihood does not require orthogonality of t_c and η and, as we see further, its approximation can be calculated without specification of the ancillary statistic. Finally, [92] has shown that the modified profile likelihood can be considered as an approximation to a class of integrated likelihood functions and very naturally arises from a non-Bayesian inference with integrated likelihood. But, in contrast to the Bayesian approach or integrated likelihood functions, the modified profile likelihood does not require specification of a prior density for the nuisance parameters — the main limitation that hampered us from pursuing this direction.

5.4.2 Inference on the errors variance

The importance of the modified profile likelihood cannot be overstated, given that it is considered one of the breakthroughs in modern parametric inference [32]. Perhaps the best illustration of the power of this method relates to the estimation of the variance $s = \sigma^2$ in nonlinear regressions such as (5.3). It is well known that the standard estimation (5.4) or (5.6) is biased and it should be corrected to account for the number of degrees of freedom, i.e. the number of free parameters to estimate. The modified profile likelihood provides this correction as follows.

For the time being, let us consider s as a parameter of interest and all the other parameters $\lambda = \{t_c, m, \omega, A, B, C_1, C_2\}$ as nuisance parameters. Parameters s and λ are not only informationally orthogonal, but the estimation of λ does not depend on s at all (since λ is given straightforwardly from the OLS method). Thus, $\hat{\lambda}_s \equiv \hat{\lambda}$ and $|\partial\hat{\lambda}/\partial\hat{\lambda}_s| \equiv 1$. Having taken care of the Jacobian, we only need to calculate the observed Fisher information $I(\hat{s}_\lambda)$.

Straight from (5.3), we can derive the vector of first derivatives of the log-likelihood — the so-called *score function* $S(\lambda)$:

$$S(\lambda) = \frac{\partial \ln L(s,\lambda)}{\partial\lambda} = -\frac{1}{2s}\frac{\partial \mathrm{SSE}(\lambda)}{\partial\lambda}. \tag{5.11}$$

The negative second derivative gives us the observed Fisher information matrix whose determinant reads

$$|I(\lambda)| = \left|-\frac{\partial^2 \ln L(s,\lambda)}{\partial\lambda\partial\lambda^T}\right| = \left(\frac{1}{2s}\right)^{p_\lambda}\left|\frac{\partial^2 \mathrm{SSE}(\lambda)}{\partial\lambda\partial\lambda^T}\right|, \tag{5.12}$$

where $p_\lambda = \dim\lambda = 7$ is the dimension of the nuisance parameter space. Before plugging the expression (5.12) into (5.9) in order to obtain the modified profile likelihood of s, notice that (i) the matrix of second-order derivatives $\partial^2\mathrm{SSE}(\lambda)/\partial\lambda\partial\lambda^T$ in (5.12) does not depend on the parameter of interest s explicitly and (ii) the OLS estimation $\hat{\lambda}_s \equiv \hat{\lambda} = \{\hat{t}_c, \hat{m}, \hat{\omega}, \hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\}$ also does not depend on s. Thus, the determinant of $\partial^2\mathrm{SSE}(\lambda)/\partial\lambda\partial\lambda^T$ is a constant with respect to the variable s and therefore can be omitted. Then, the modified profile likelihood of s can be expressed in the following form:

$$L_m(s) \propto s^{(n-p_\lambda)/2}\exp\left(-\frac{\mathrm{SSE}(\hat{\lambda})}{2s}\right), \tag{5.13}$$

which leads to the following MLE for s:

$$\hat{s} = \frac{1}{n-p_\lambda}\mathrm{SSE}(\hat{\lambda}). \tag{5.14}$$

The denominator $n - p_\lambda$, which is different from n in (5.4), not only removes the bias of the estimator, but also results in a better likelihood-based inference of s when it is needed.

5.4.3 Approximation of the modified profile likelihood

Given all the remarkable properties of the modified profile likelihood, it has one very serious limitation, briefly mentioned above. Namely, for many realistic models, it is extremely difficult to calculate the Jacobian in (5.9). In order to get an intuition about the nature of the difficulty, it is useful to express it in the following form (see e.g. [82]):

$$J(t_c) \equiv \left| \frac{\partial \hat{\eta}}{\partial \hat{\eta}_{t_c}} \right| = \frac{|I(\hat{\eta}_{t_c})|}{|C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})|}, \tag{5.15}$$

where the matrix $C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})$ is given by the second-order derivatives of a log-likelihood $L(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}, a)$ that includes a new parameter a that is ancillary for $\{\hat{t}_c, \hat{\eta}\}$, i.e. $\{\hat{t}_c, \hat{\eta}, a\}$ is a sufficient statistic of the model:

$$C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}) = \frac{\partial^2 \ln L(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}, a)}{\partial \hat{\eta}_{t_c} \partial \hat{\eta}^T}. \tag{5.16}$$

In contrast to the observed Fisher information, which is also defined as a second-order derivative (5.10) calculated at a specific MLE $\hat{\eta}_{t_c}$, the calculation of C (5.16) is much more complicated because, in the general case, it requires a reformulation of the log-likelihood in order to introduce an explicit dependence on the MLEs $\hat{\eta}_{t_c}$ and $\hat{\eta}$. In the case of inference of the variance s presented in Section 5.4.2, we used the orthogonality of s and λ, which resulted in $\hat{\lambda}_s \equiv \hat{\lambda}$. In contrast, for the inference on t_c, there is no closed form expression for $J(t_c)$. And, as discussed above, we cannot use the *adjusted profile likelihood* [29] because orthogonalization of the nuisance parameters with respect to t_c is not feasible either.

In order to calculate expression (5.9), several approximation of L_m were proposed (see e.g. [12, 98, 88, 41, 99] and [91, 80] for reviews). We will use the approximation to the modified profile likelihood proposed by [88]. This approximation requires only the covariance of score functions of the nuisance parameters and is thus fairly easy to compute. As shown in [88], this approximation is invariant under the reparametrization of the model, is stable in the sense of conditional inference and agrees with the exact $J(t_c)$ (5.15) to order $\mathcal{O}(n^{-1})$ in the moderate deviation sense and to order $\mathcal{O}(n^{-1/2})$ in the large deviation sense, where n is the number of data points. Another famous approximation by [12] agrees with the exact form of L_m only to $\mathcal{O}(1)$ in the large deviation sense and thus is not asymptotically better than the simple profile likelihood L_p.

[88] suggested to approximate the matrix (5.16) with the covariance matrix of score functions of the following form:

$$C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}) \approx \Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}) \tag{5.17}$$

where

$$\Sigma\left(t_{c;1}, \eta_1; t_{c;2}, \eta_2\right) = \mathrm{E}_{(2)}\left[\left.\frac{\partial \ln L(t_c, \eta)}{\partial \eta}\right|_{\substack{t_c = t_{c;1} \\ \eta = \eta_1}} \left.\frac{\partial \ln L(t_c, \eta)}{\partial \eta^T}\right|_{\substack{t_c = t_{c;2} \\ \eta = \eta_2}}\right]. \tag{5.18}$$

Here the expectation $\mathrm{E}_{(2)}[\cdot]$ is taken with respect to the probability distribution of error term $\varepsilon(\tau; t_{c;2}, \eta_2)$ that corresponds to the parameters $\{t_{c;2}, \eta_2\}$. In contrast to the exact form (5.16), here we need only the score functions, which have expressions similar to (5.11). When calculation of (6.23) is too complicated, one can exploit the independence of observations $\varepsilon(\tau_i; t_c, \eta)$ and replace the covariance matrix (6.23) by its asymptotically equivalent sample estimation [90]:

$$\widehat{\Sigma}\left(t_{c;1}, \eta_1; t_{c;2}, \eta_2\right) = \sum_{i=1}^{n} \left.\frac{\partial f(\tau_i; t_c, \eta)}{\partial \eta}\right|_{\substack{t_c = t_{c;1} \\ \eta = \eta_1}} \left.\frac{\partial f(\tau_i; t_c, \eta)}{\partial \eta^T}\right|_{\substack{t_c = t_{c;2} \\ \eta = \eta_2}}, \tag{5.19}$$

where

$$f(\tau;t_c,\eta) = -\frac{1}{2}\ln(2\pi s) - \frac{1}{2s}\big(\ln p(\tau) - \mathrm{LPPLS}(\tau;t_c,\psi)\big)^2 \tag{5.20}$$

is a contribution from an individual observation to the log-likelihood. Of course, the adjustment (6.23) based on the theoretical covariance is superior to the sample-based estimation (5.19), in particular in cases of small sample size [90, 17]. For our purposes, we will use the exact form (6.23), which can be calculated in closed form. Finally, plugging (6.23) into (5.15) and (5.9), we obtain the desired approximated expression for $L_m(t_c)$:

$$L_m(t_c) \approx \frac{|I(\hat{\eta}_{t_c})|^{1/2}}{\left|\Sigma(t_c,\hat{\eta}_{t_c};\hat{t}_c,\hat{\eta})\right|} L_p(t_c)\ . \tag{5.21}$$

In this expression (6.24), the profile likelihood $L_p(t_c)$ is given by the previously calculated expression (5.7). The observed Fisher information $I(\hat{\eta}_{t_c})$ and the covariance matrix $\Sigma(t_c,\hat{\eta}_{t_c};\hat{t}_c,\hat{\eta})$ are given in Appendix 5.7. Omitting terms that do not depend on t_c, the final expression for $L_m(t_c)$ is given by:

$$L_m(t_c) \propto \frac{\left(\hat{s}_{t_c}\right)^{-(n-p-2)/2}\left|\sum_{i=1}^{n}\frac{\partial^2\mathrm{LPPLS}(\tau_i;t_c,\psi)}{\partial\psi\partial\psi^T}\right|_{\psi=\hat{\psi}_{t_c}}^{1/2}}{\left|\sum_{i=1}^{n}\frac{\partial\mathrm{LPPLS}(\tau_i;t_c,\psi)}{\partial\psi}\bigg|_{\substack{t_c=t_c\\ \psi=\hat{\psi}_{t_c}}}\frac{\partial\mathrm{LPPLS}(\tau_i;t_c,\psi)}{\partial\psi^T}\bigg|_{\substack{t_c=\hat{t}_c\\ \psi=\hat{\psi}}}\right|}, \tag{5.22}$$

where $p = \dim\psi = 6$. Following [90], let us introduce the rectangular $n\times p$ matrix

$$X_{ij}(t_c,\psi) = \frac{\partial\mathrm{LPPLS}(\tau_i;t_c,\psi)}{\partial\psi_j} \tag{5.23}$$

and the square $p \times p$ matrix

$$H_{ij}(t_c, \psi) = \sum_{k=1}^{n} \big(\ln p(\tau_k) - \text{LPPLS}(\tau_k; t_c, \psi) \big) \frac{\partial^2 \text{LPPLS}(\tau_k; t_c, \psi)}{\partial \psi_i \partial \psi_j}, \tag{5.24}$$

where ψ_j denotes the j-th element of the nuisance parameter vector $\{m, \omega, A, B, C_1, C_2\}$. Then, expression (5.22) simplifies into

$$L_m(t_c) \propto \frac{\big| X^T(t_c, \hat{\psi}_{t_c}) X(t_c, \hat{\psi}_{t_c}) - H(t_c, \hat{\psi}_{t_c}) \big|^{1/2}}{\big| X^T(\hat{t}_c, \hat{\psi}) X(t_c, \hat{\psi}_{t_c}) \big|} \big(\hat{s}_{t_c} \big)^{-(n-p-2)/2}, \tag{5.25}$$

where $\hat{s}_{t_c}$ is the MLE estimate of the variance (5.6) (it is not the adjusted estimate (5.14)), $\hat{\psi}_{t_c}$ is a vector of MLE estimates for the LPPLS parameters at a fixed value of t_c and $\{\hat{t}_c, \hat{\psi}\}$ are full MLE estimates of the parameters. The expressions of the first-order and second-order partial derivatives that are needed for (5.23) and (5.24) are given by (5.48) and (5.49) in Appendix B.

As a concrete illustration, we consider the 2015 bubble in Chinese markets already discussed in Sections 5.2–5.3. The red curves in the two top panels of figure 5.3 show the modified profile likelihood obtained from expression (5.25). It is particularly interesting that the adjustments to $L_m(t_c)$ have significantly changed the picture, since the "alternative" extremum has now a higher likelihood than the best OLS solution ($R_m(t_c^{(2)}) = 1.$ versus $R_m(t_c^{(2)}) = 0.91$). Thus, accepting the OLS point estimate would bias $\hat{t}_c$ by 19 days. The likelihood ratio is now even smaller than for the simple profile likelihood, $R_m(t_c^{(2)})/R_m(t_c^{(1)}) = 1.096$, and both extrema are almost equally likely.

5.4.4 Likelihood Intervals and Confidence Intervals

The major improvement of the standard MLE interpretation (5.3) over the OLS (1.10) is the fact that MLE provides a direct estimation of the uncertainty in estimated parameters. In other words, MLE can provide not just the *point* estimate of θ but a *range estimate* of values that are possible given the observed data. Such inference is based on the *likelihood ratio* $R(\theta)$, introduced earlier in the form of the relative likelihood (5.8) defined as the ratio of the likelihood normalized by its maximum value. When $R(\theta_0)$ is sufficiently small, the hypothesis that the parameter could have a value $\theta = \theta_0$ can be rejected as "unsupported by the data".

However, the question of "how small is sufficiently small?" often does not have a rigorous solution and strongly depends on the problem. Many authors suggest to choose some rather arbitrary *cutoff* and consider values of likelihood ratio above this cutoff to define a so-called *likelihood-based confidence interval* or *likelihood interval (LI)*. For example, many authors including [39]suggested that parameter values θ for which $L(\hat{\theta})/L(\theta) = 1/R(\theta) > 15$ should be declared "implausible", where $\hat{\theta}$ is the standard MLE.

In regular one-parameter models, one can create a frequentist confidence interval, based on a probability-based calibration. For example, the log-likelihood ratio test statistic $-2\ln R(\theta)$ can be then approximated using Wilk's theorem, and an approximate p-value is given by the χ^2-distribution with one degree of freedom. Further, for regular likelihood functions, i.e. those that are well-approximated by a quadratic function, one can define a *confidence interval (CI)* around MLE $\hat{\theta}$ solely based on the observed Fisher information. For example, a standard error would have the form $I^{-1/2}(\hat{\theta})$ and 95% CI would be given by $\hat{\theta} \pm 1.96 I^{-1/2}(\hat{\theta})$ (*Wald confidence interval*).

For our applications, these approaches are not perfectly suited. First, as can be seen in Figure 5.3, the profile and modified profile likelihoods are not regular: they are asymmetric and can be even multi-modal, so that the Wald CI does not provide a meaningful representation of parameter uncertainty. For the same reason, the calibration of the distribution of the test statistic under the null hypothesis is not straightforward and would be computationally very difficult given the dimensionality of the parameter space and the complexity of the LPPLS model (1.6). Finally, within our domain of application, an interpretation of the frequentist probability-based confidence intervals is not very intuitive. Indeed, giving the idiosyncratic nature of a bubble, in order to make sense out of the probabilistic intervals, one needs to involve a many-worlds interpretation, where price trajectory is shared among multiple universes.

For all the reasons mentioned above, we choose to operate with likelihood intervals that are more intuitive in our context and are not subjected to the assumptions of regularity. Following Fisher's suggestion, we define the likelihood interval at the 5% cutoff:

$$\mathrm{LI}(t_c) = \left\{ t_c : R_m(t_c) = \frac{L_m(t_c)}{L_m(\hat{t}_c)} > 0.05 \right\}. \tag{5.26}$$

The two top panels of Figure 5.3 show such 5% modified profile likelihood intervals for the case of 2015 Chinese bubble.

5.5 Filtering and likelihood intervals for nuisance parameters

Similarly to the inference on the critical time t_c, let us apply the modified profile likelihood approach to estimate the likelihood intervals (LIs) of parameters m and ω. This is of interest in particular because m and ω are used in the filtering conditions (1.2.1).

Three different ways of inference on m and ω exist. First, we could consider m (respectively, ω) as the sole parameter of interest and $\eta_m = \{t_c, \omega, A, B, C_1, C_2, s\}$ (respectively, $\eta_\omega = \{t_c, m, A, B, C_1, C_2, s\}$) as the vector of nuisance parameters. Then, an analysis similar to that developed in Sections 5.4–5.4.4 would provide the corresponding modified profile likelihood and LIs for these two parameters. However, keeping in mind that the parameter of main interest is the critical time t_c, we would need to somehow associate the inferred LIs for m (ω) with the corresponding values of t_c.

A second approach consists in targeting the vector $\theta = \{t_c, m, \omega\}$, while $\eta_\theta = \{A, B, C_1, C_2, s\}$ becomes the vector of nuisance parameters. The general framework remains the same as before. However, the computational complexity increases substantially, since the modified profile likelihood $L_m(\theta)$ is a 3-dimensional function. And the analysis of such function is not straightforward, with many 2D-cross-sections needed to obtain a suitable understanding of the topology in four dimensional space. Or we would need another layer of profile or modified profile likelihood to be calculated.

Here, we employ a third approach. For any fixed value of t_c, we consider a reduced LPPLS formula that is parameterized solely with the vector $\{m, \omega, A, B, C_1, C_2\}$. We then calculate a modified profile likelihood $L_m(m; t_c)$ (respectively $L_\omega(\omega; t_c)$) with $\eta_m = \{\omega, A, B, C_1, C_2, s\}$ (respectively, $\eta_\omega = \{m, A, B, C_1, C_2, s\}$) as the vector of nuisance parameters. The expression for $L_m(\cdot)$ is then similar to (5.25). For example, $L_m(m; t_c)$ has the form

$$L_m(m; t_c) \propto \frac{\left|X_m^T(t_c, m, \hat{\phi}_{t_c,m}) X_m(t_c, m, \hat{\phi}_{t_c,m}) - H_m(t_c, m, \hat{\phi}_{t_c,m})\right|^{1/2}}{\left|X_m^T(\hat{t}_c, \hat{m}, \hat{\phi}) X_m(t_c, m, \hat{\phi}_{t_c,m})\right|} \left(\hat{s}_{t_c,m}\right)^{-(n-p-2)/2}, \tag{5.27}$$

where $\phi = \{\omega, A, B, C_1, C_2\}$, $p = \dim \phi = 5$, $\{\hat{t}_c, \hat{m}, \hat{\phi}\}$ is the full MLE of all parameters of the LPPLS model and $\hat{\phi}_{t_c,m}$ is the MLE of ϕ at fixed values of $\{t_c, m\}$. Finally $\hat{s}_{t_c,m} = \mathrm{SSE}(t_c, m, \hat{\phi}_{t_c,m})/n$ and the matrice X_m is obtained from X (5.23) by removing the first column and the matrix H_m is the principal submatrix of H (5.24), obtained by removing its first row and first column. Targeting ω, the expression for $L_m(\omega; t_c)$ is also given by(5.27) up to a replacement of m by ω, where X_ω is obtained from X by removing the second column, and H_ω is obtained from H by removing the second column and second row.

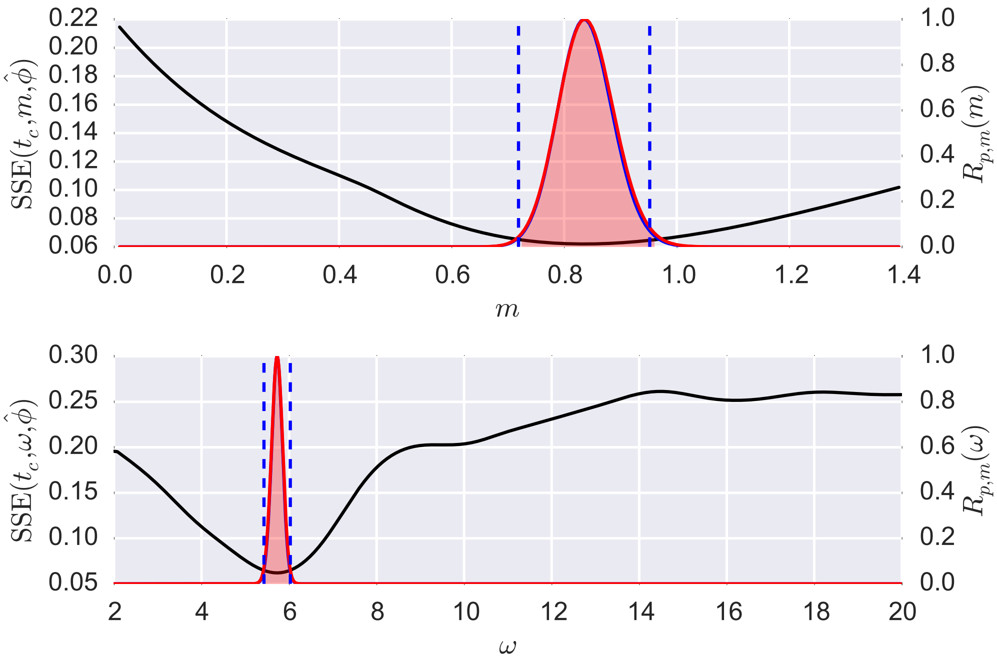


Figure 5.4: Profile of the cost function $F(\cdot)$ (black line, left scale), profile likelihood $L_p(\cdot)$ (blue line, right scale) and modified profile likelihood (red line, right scale) for the power law exponent m (top panel) and logperiodic angular frequency ω (bottom panel) for t_c =2015-06-17. Note that the profile and modified profile likelihood almost coincide. The red shaded intervals represent the likelihood intervals LI$(\cdot)$ at the 5% cutoff. The vertical blue dashed lines delineate the approximated likelihood intervals (5.29) at the 5% cutoff.

Figure (5.4) presents the profile and modified profile likelihoods for the parameters m and ω in the case considered before (Figures 5.1–5.3) for the fixed value of t_c =2015-06-17. It is interesting to note that the SSE profile of parameter m at a fixed t_c is unimodal in the range of interest. Moreover, our tests show that this is typically the case for a broad range of values $0 < m \lesssim 3$. The SSE profile for ω is multimodal, but when the price trajectory exhibits a clear upward trend with a substantial price appreciation over the window of calibration $[t_1, t_2]$ (e.g. when the price increase is substantially larger than the volatility), then the best solution $\hat{\omega}$ is often clearly delineated and the likelihood profile is essentially unimodal, i.e. the alternative solutions are implausible (as in Figure 5.4).

In Figure (5.4), it is almost impossible to distinguish the profile likelihood from the modified profile likelihood in the visible range of values. The values at which the log-likelihoods start to disagree, i.e. for $R_m(\cdot;t_c) = \ln L_m(\cdot;t_c)/\ln L_m(\hat{\cdot};t_c) \lesssim -15$, cannot be been seen in this linear scale representation. We have found that this situation is typical for many other cases. This very close agreement means that the profile likelihood $L_p(\cdot;t_c)$ is already a good approximation to either the marginal or the conditional likelihood so that we could use it directly for the inference of likelihood intervals. Moreover, the peak of the profile likelihood can often be well approximated by a quadratic function, allowing use to use this approximation for an analytical evaluation of LI[2].

In contrast to the estimated likelihood, the negative curvature of the profile likelihood function of a parameter η_i is not equal to $[I(\hat{\eta}_{t_c})]_{i,i}$, where I is the observed Fisher information matrix (5.38), but to $([I^{-1}(\hat{\eta}_{t_c})]_{i,i})^{-1}$ (see e.g. derivations in [51]). One can prove that $[I(\hat{\eta}_{t_c})]_{i,i} \geq ([I^{-1}(\hat{\eta}_{t_c})]_{i,i})^{-1}$, which means that the observed Fisher information of the profile likelihood is smaller than or equal to the observed Fisher information on the estimated likelihood. This illustrates the fact that the nuisance parameter η has to be estimated and thus adds to the uncertainty of the parameter of interest. Taking this approximation of the curvature into account, we can write the following Taylor expansion for the profile likelihood of m and ω at a fixed t_c:

$$\begin{aligned} \ln L_p(m;t_c) &\approx \ln L(t_c,\hat{\eta}_{t_c}) & -\tfrac{1}{2}([I^{-1}(\hat{\eta}_{t_c})]_{1,1})^{-1}(m-\hat{m}_{t_c})^2, \\ \ln L_p(\omega;t_c) &\approx \ln L(t_c,\hat{\eta}_{t_c}) & -\tfrac{1}{2}([I^{-1}(\hat{\eta}_{t_c})]_{2,2})^{-1}(\omega-\hat{\omega}_{t_c})^2, \end{aligned} \tag{5.28}$$

[2]We need to mention that this is not always the case, and a bi-modal structure of both profiles on m and ω is also possible, though rare. Moreover, in some cases, the second-order approximation of the modified profile likelihood might completely change the estimation of these parameters (see Appendix 5.7).

and thus the likelihood intervals at a cutoff of level c are given by

$$\begin{aligned}
\mathrm{LI}(m;t_c) &= \left\{m : \tfrac{L_p(m;t_c)}{L_p(\hat{m};t_c)} > c\right\} &= \left\{m : |m - \hat{m}_{t_c}| < \Delta_{m;t_c}\right\}, \quad \Delta_{m;t_c} &= \sqrt{-2\ln c\,[I^{-1}(\hat{\eta}_{t_c})]_{1,1}},\\
\mathrm{LI}(\omega;t_c) &= \left\{\omega : \tfrac{L_p(\omega;t_c)}{L(\hat{\omega};t_c)} > c\right\} &= \left\{\omega : |\omega - \hat{\omega}_{t_c}| < \Delta_{\omega;t_c}\right\}, \quad \Delta_{\omega;t_c} &= \sqrt{-2\ln c\,[I^{-1}(\hat{\eta}_{t_c})]_{2,2}}.
\end{aligned} \tag{5.29}$$

Here, $I(\hat{\eta}_{t_c})$ has the form (5.38) (Appendix A), and its submatrix of partial derivatives can be written in a matrix form similar to the numerator in (5.25). These likelihood intervals for $c = 0.05$ are indicated with dashed vertical lines in Figure 5.4, and one can see that they provide a very good approximation for the true LIs based on the modified profile likelihood for m and ω at a fixed t_c (red shaded areas).

The likelihood interval for the damping parameter $D = m|B|/\omega|C|$ is slightly more difficult to calculate. Because D does not enter LPPLS expression (1.6) directly, we first need to perform a variable change, e.g. by replacing the vector $\eta = \{m, \omega, A, B, C_1, C_2, s\}$ with $\zeta = \{D, \omega, A, B, C_1, C_2, s\}$. Under such reparametrization, the observed Fisher information matrix (5.38) is transformed into

$$I(\zeta) = J_D^T I_\eta(\eta(\zeta)) J_D, \tag{5.30}$$

where $J_D = \partial\eta/\partial\zeta$ is the Jacobian matrix of the transform from η to ζ, whose its full expression is given by (5.50) in Appendix C. Finally, the likelihood interval for the damping parameter is

$$\mathrm{LI}(D;t_c) = \left\{D : |D - \hat{D}_{t_c}| < \Delta_{D;t_c}\right\},\ \Delta_{D;t_c} = \sqrt{-2\ln c\left[I^{-1}(\hat{\zeta}_{t_c})\right]_{1,1}}. \tag{5.31}$$

As discussed above, the modified profile likelihood (5.9) for the main parameter t_c of interest is invariant with respect to such transformations of the nuisance parameter vector η.

We are now in position to discuss the overall results presented in Figure 5.3. The first important observation is that, in view of the determined likelihood intervals, the rejection of the "suboptimal" solution is no more warranted (given that $\mathrm{LI}(\omega) = \{5.47 < \omega < 6.07\}$). Observe that the optimal solution now easily fits in the extended interval of the damping parameter constraint (1.2.1). Second, it is interesting to compare the interval widths (2Δ) representing the uncertainty of the different parameters. In the particular example presented in Figure 5.3, the power law exponent m is the most uncertain parameter with $2\Delta_{m;t_c} \approx 0.3$, which is about 40% of the estimated value $\hat{m}_{t_c}$. The damping parameter D, which is proportional to m, also has a fairly broad likelihood interval with width $2\Delta_{D;t_c} \approx 0.17$, which is about 20% of the estimated value $\hat{D}_{t_c}$. Finally, the uncertainty of the logperiodic angular frequency is $2\Delta_{\omega;t_c} \approx 0.63$, which is about 7% of $\hat{\omega}_{t_c}$. In general, the widths 2Δ of the likelihood intervals strictly depend on the specific realisation of the data, but our extensive tests have shown that the above observations typically hold. Finally, it is interesting to document that such rather large uncertainty in the nuisance parameters does not result in a dramatic change of the likelihood intervals for the parameter of interest t_c. And while the modified profile likelihood corrects the shape of the distribution, the intervals (5.26) for the profile and modified profile likelihoods at a 5% cutoff agree rather well in this and many another cases.

5.6 Application of the methodology

In the previous Sections 5.3–5.5, we have developed a framework to infer the critical time t_c from the LPPLS model, which includes parameter estimation together with its confidence interval, as well as the confidence intervals of the relevant nuisance parameters *within a fixed calibration window* $[t_1, t_2]$. However, for real life applications, one cannot limit oneself to the analysis of a single time-scale, because financial time-series result from complex generating processes, from volatility clustering of the simplest form to multifractal models, subjected to regime-shifts leading to non-trivial scaling structures. In order to understand the complexity of such phenomena through the prism of some model like LPPLS, one needs to apply this model at different scales simultaneously, and also consider the evolution of the model parameters in time.

In this section, we extend the analysis of the LPPLS model to the scale-domain $t_2 - t_1$ and provide illustrations of the application of the methodology both to synthetic case and real price series.

5.6.1 Aggregation of time-scales

By time scale, we mean the width $\Delta_t \equiv t_2 - t_2$ of the time window in which the analysis is performed. The aggregation of analyses performed at different time-scales is not a trivial problem, whose difficulty starts with the mere computational complexity of non-linear models. Usually, the application of the model at several time scales proportionally increases the computational time, and the output data that needs to be analyzed also increases manifold. Further, in order to make the analysis operational, one needs a method for aggregating the massive amount of parameter information for the construction of the predictive features or signals. Then, the next step is to perform a full-scale back-testing of the constructed signals for understanding their predictive power. These challenging operational steps go beyond the scope of the present methodological paper, and will be reported elsewhere. Some practical aspects are already discussed in [112], where multi-scale signals were used for ex-ante forecasting the crash in Chinese markets in June 2015. [117] also presented a multi-scale analysis with LPPLS, in which the different scales were combined via a pattern recognition algorithm. Here, we will focus on the descriptive analysis and visualization aspects.

The analysis of the modified profile likelihood in Sections 5.4–5.5 was aimed at estimating the likelihood intervals (LI) of the critical time t_c as well as of the logperiodic angular frequency ω, power law exponent m and damping D, which contribute to the filtering criteria (1.2.1). A multi-scale approach would require analysis of these outputs for different values of window sizes $\Delta_t = t_2 - t_1$. One of the most natural ways is to construct the modified profile likelihood $L_m(t_c; \Delta_t)$ independently for different window sizes Δ_t. Because the absolute value of the likelihood depends on the amount of data, it does not make sense to compare value of L_m for different Δ_t directly. For comparison, we will use the normalization as in (5.8), and will apply it for each window size Δ_t independently, constructing the *relative multi-scale modified profile likelihood* $R(t_c, \Delta_t) = L_m(t_c; \Delta_t)/\max_{t_c} L_m(t_c; \Delta_t)$. The structure of $R(t_c, \Delta_t)$ then directly provides with scale-dependent likelihood intervals $\mathrm{LI}(t_c; \Delta_t)$ for the critical time t_c.

In order to illustrate this approach, we construct modified profile likelihoods for Δ_t varying from 60 to 700 days and for t_c varying from $t_2 - 50$ to $t_2 + 150$ days. The scale-dependent likelihood intervals are presented in red color in Figure 5.5 (red profiles are identical in both panels) for the same t_2=2015-06-12 that was used for illustration earlier in the paper. One can see the same bi-modal structure for $\Delta_t = 180$ as reported earlier, which suggests two possible scenarios for the end of the bubble: $t_c = t_2 + 5$ and $t_c = t_2 + 25$ (days). Figure 5.5 gives an illustrative overview of the structure of the inferred critical time t_c broken down in three time scales: (i) short time scales ($\Delta_t \leq 160$) suggest that the price trajectory is at its peak already and the critical time is close to the date of analysis $t_c \approx t_2$ (with the MLE of t_c being a few days before t_2); (ii) intermediate time scales ($180 \leq \Delta_t \leq 350$) suggest two main scenarios in which the critical time is clustered around 20-30 or 60-90 days in the future; and (iii) large scales $\Delta_t > 350$ do not give stable clusters.

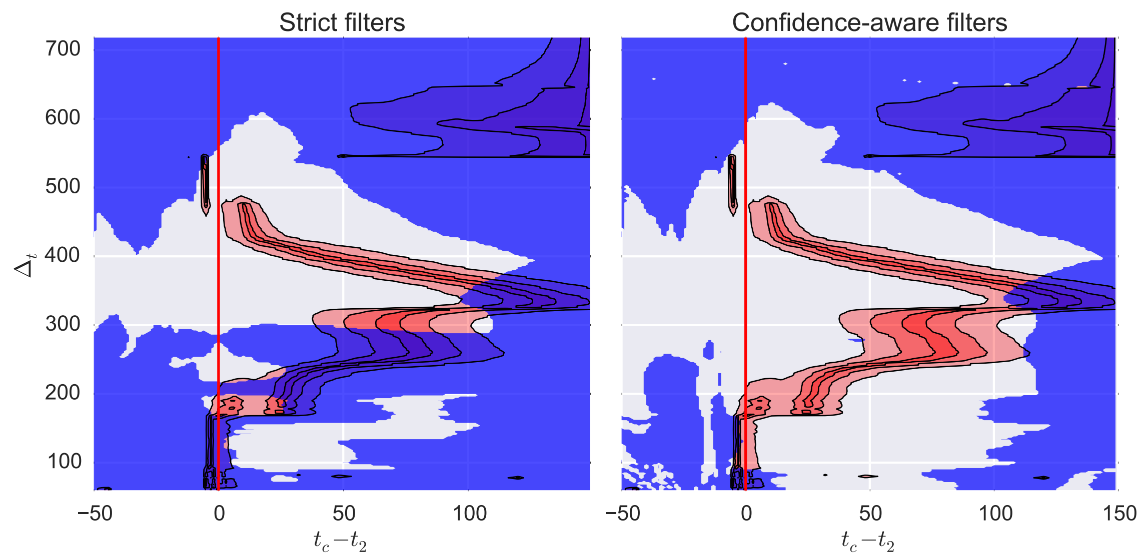


Figure 5.5: Two-dimensional structure of the relative multi-scale modified profile likelihood $R(t_c, \Delta_t)$ at the date t_2=2015-06-12, corresponding to multiple calibrations of the model with different window sizes Δ_t. The horizontal axis gives the value $t_c - t_2$ with the solid red vertical line indicating the case where t_c is coincident with the date of analysis. Each horizontal slice of the plot gives in color code the dependence of the individual modified profile likelihoods of the model (as in Figure 5.4) calculated for a given window size $\Delta_t = t_2 - t_1$ (vertical axis) as a function of $t_c - t_2$. The shaded red area corresponds to likelihood intervals $\mathrm{LI}(t_c)$ of the critical time at the 5%, 50% and 95% cutoff (from lightest to darkest colors). The shaded blue area denotes values of (t_c, Δ_t), where the constraints on the nuisance parameters (1.2.1) are not met. The left panel corresponds to the case when only MLE parameters are considered for filtering, the right panel corresponds to the case when likelihood intervals (5.29) and (5.31) are taken into account.

Figure 5.5 also provides important insights on the range of values $(t_c; \Delta_t)$ for which the parameters obey the theoretical constrains (1.2.1) — below we will refer to them as *qualified fits*. In the left panel, the blue shaded area indicates when LPPLS fits can be rejected based solely on the MLE values of m, ω and D ("strict filtering"). In the right panel, we take additionally into account the likelihood intervals of these parameters (see Section 5.5) and show only the region where these intervals have no overlap with the constraints (1.2.1) ("confidence-aware filtering"). These cases differ quite dramatically, in the sense that strict filtering falsely rejects a substantial number of fits that correspond to credible alternative scenarios. As discussed in [112], choosing the proper filters is one of the key ingredients for constructing successful signals. Being a very broad subject, constructing and testing useful filters goes beyond the scope of present paper. For the time being, we stress how crucial it is to take into account data-induced uncertainty when constructing signal filters.

Let us now describe some of the potential numerical issues that often arise in such complicated optimization problems. First, because the search for t_c is constrained in a pre-defined bounded interval, the real $\max_{t_c} L_m(t_c; \Delta_t)$ might lie outside of it, so that the numerical procedure might pick up a value at the boundary of the search space on t_c. Normalising $L_m(t_c)$ to 1 at this boundary point, this may result in having a wide range of high values of $R(t_c, \Delta_t)$ close to this boundary, leading to a spurious likelihood interval $\mathrm{LI}(t_c)$. In the example above, this is exactly what happens for $\Delta_t > 550$ (red profiles at the top-right of Figure 5.5), where the maximum of the modified profile likelihood is beyond the search range ($t_c > t_2 + 150$) and the inference on the likelihood intervals is completely misleading.

Another problem is the potential bad convergence of the optimization of the nuisance parameters in (6.6), which dramatically affects the value of $L_p(t_c)$ and thus $L_m(t_c)$. Usually, this situation occurs for large values of $t_c - t_2$, especially when $t_c - t_2$ is not small compared with the window size Δ_t. However, it highly depend on the structure of the residuals and the numerical method might not converge even for moderate values of $t_c - t_2$. What makes this issue complicated is that there is no simple way of detecting bad convergence, neither algorithmically nor even visually in plots like Figure 5.5. It often results in some kind of discontinuities in the plot, but not always. Take for instance the case $\Delta_t = 470 - 480$, where an apparent discontinuity of likelihood intervals is in fact the consequence of a continuous transition of one maximum of the likelihood to another when increasing Δ_t.

As with all non-linear optimization problems, there is no "silver bullet" to address such numerical issues. Measures such as increasing the region of search or the precision of numerical methods do not always help. Especially when one performs fully automated analyses, it is highly recommended to carefully validate each step of the procedure and take outputs with a grain of salt, not hesitating to "triple-check" any suspicious results.

5.6.2 Synthetic tests

In order to gain insight about the likelihood inference of the critical time (t_c) during a growing bubble and establish a solid background for our empirical analysis, we first test our methodology on synthetic time series, where the underlying process follows the LPPLS structure. Specifically, we generate the log-price as

$$\ln[P(t)] = LPPLS(t) + \sigma\epsilon(t), \tag{5.32}$$

where $LPPLS(t)$ is given by (1.6) with t_c^0=1975-02-09, $m^0 = 0.8$, $\omega^0 = 9$, $\phi^0 = 0$ and $A^0 = 8$, $B^0 = -0.015$, $C^0 = 0.0015$ (i.e. with low damping $D^0 = 0.88$), $\epsilon(t)$ is an iid N(0,1) noise and $\sigma^0 = 0.03$. The resulting price trajectory is illustrated in Figure 5.6.

With the goal of understandings the evolution of the parameters as a function of the "present" time t_2 (the time of analysis) for such a synthetic bubble, we apply our methodology to construct a multi-scale modified profile likelihood (see Section 5.6.1 and Figure 5.5) at different dates t_2 increasing towards the end of the bubble at t_c^0. The resulting multi-scale profiles are shown in insets of Figure 5.6. Far from the critical time (insets 1 and 2: $t_c^0 - t_2 = 192$ and 161 days respectively), the critical time cannot be identified even in such a clean synthetic case with weak noise (partially because we are limiting the search space to

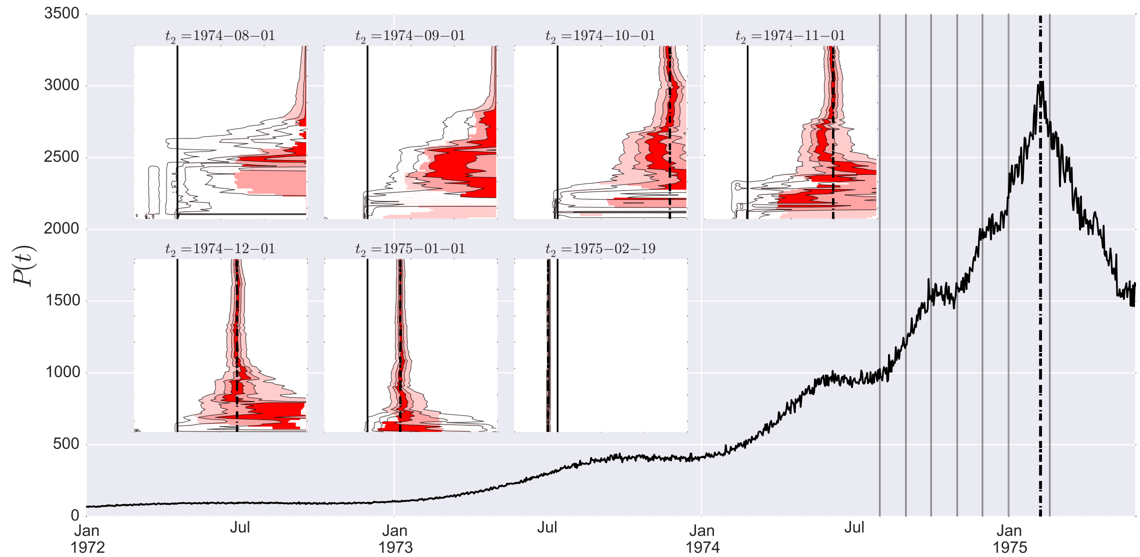


Figure 5.6: Synthetic price time series (5.32) together with multi-scale modified profile likelihoods calculated at various dates t_2. Each inset shows contour plots of the likelihood intervals at 5%, 50% and 95% cutoff levels (as in Figure 5.5). The red shaded area denotes values of (t_c, Δ_t) where the constrains on the nuisance parameters (1.2.1) are met when likelihood intervals (5.29) and (5.31) are taken into account (i.e. red color denote these parts of the contour plots that are not covered by blue area in Figure 5.5). The solid vertical line corresponds to $t_c = t_2$, and the dashed vertical line shows the true critical time $t_c = t_c^0$. Values of t_2 used for the analyses are indicated in the inset titles. They are also shown with vertical gray lines in the plot with the price trajectory.

$t_c < t_2 + 150$ days). Even when $t_c^0 - t_2$ enters the range < 150, the parameters continue to exhibit a large uncertainty. One can observe that fits for different scales progressively build a consensus, as the likelihood peaks aggregate around the true critical time with a narrow likelihood interval around it. This is first observed for the large scales $\Delta_t > 300 - 400$ (insets 3 and 4 for $t_c^0 - t_2 = 131$ and 100 days respectively), and this consensus spreads to smaller scales of $\Delta_t \sim 100 - 300$ (insets 5 and 6 for $t_c^0 - t_2 = 70$ and 39 days respectively).

Another remarkable fact is that, once the critical time is passed and the price trajectory switched to a crashing regime ($t_2 > t_c^0$, inset 7), all scales confirm this occurrence by fixing the MLE $\hat{t}_c \approx t_c^0$ with an extremely narrow likelihood interval, and this anchoring holds for a large time interval. The same effect is observed in the analysis of real data

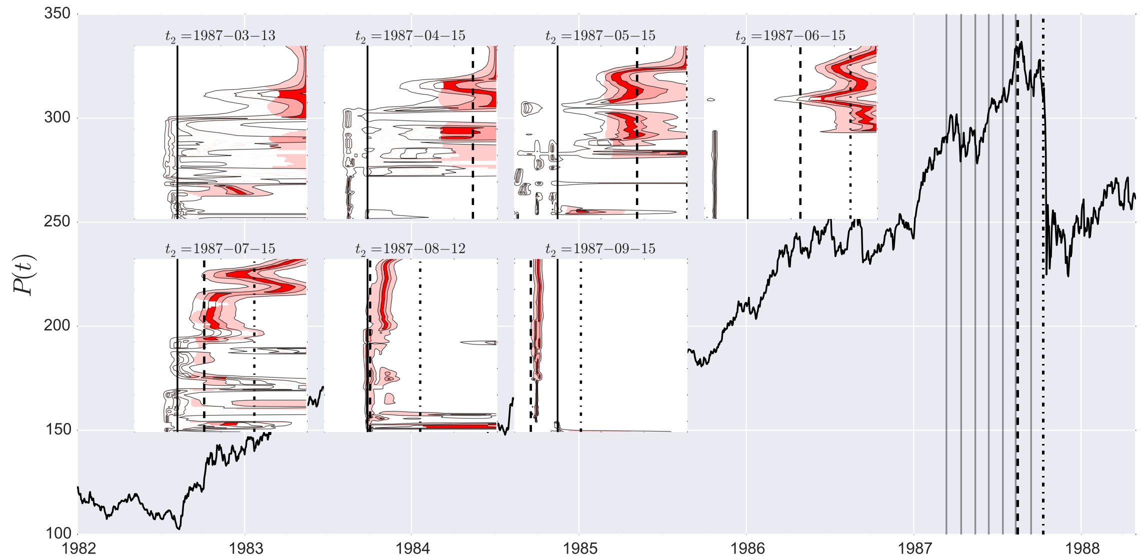


Figure 5.7: Same as Figure 5.6 for the $S\&P500$ index that shows a strong bubble of US markets developing in the second half of the 1980's, which culminated with the Black Monday crash of Oct. 19. 1987.

presented in Section 5.6.3 — even when the ex-ante forecast of the end of the bubble might be difficult or inconclusive, the change of the price direction can be identified quite reliably within a few days of the switching point.

5.6.3 Case-studies

We now provide examples of the application of the procedure described in previous sections to several well-known historical bubbles: (i) the rally in the US markets in the second half of the 1980's culminating with the Black Monday crash of Oct. 19. 1987, (ii) the dot-com bubble in the IT sector in the US culminating with a crash in April 2000, (iii) the Chinese bubble of 2014-2015 that peaked in June 2015.

We use the daily closing prices of the S&P 500, NASDAQ and SSEC indices provided by Thomson Reuters Dataworks Enterprise (DWE). We only consider business days, ignoring weekends and one-day holidays. However, for extended holidays (such as the

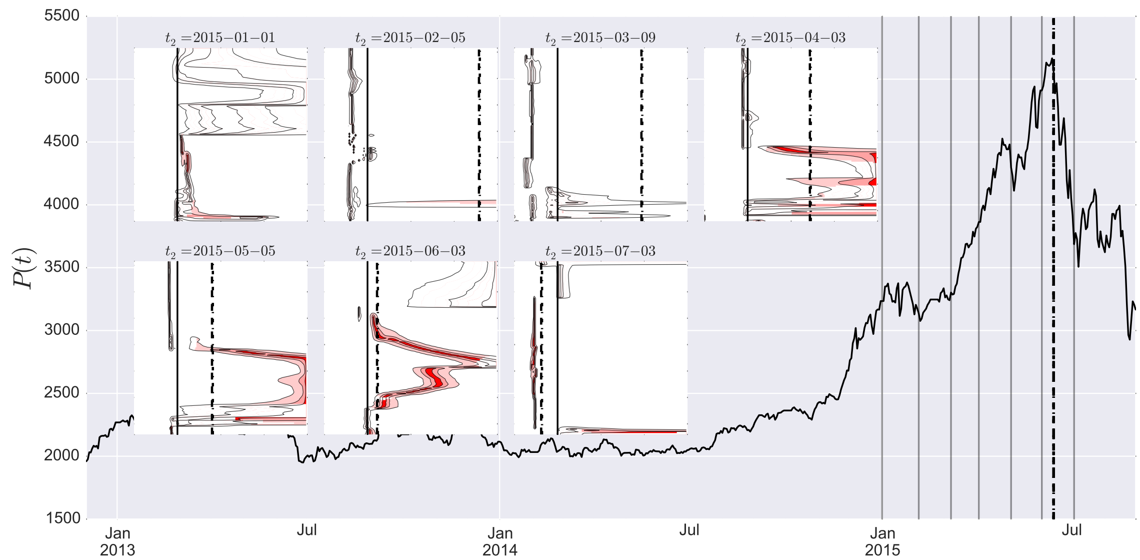


Figure 5.8: Same as Figure 5.6 for the Chinese bubble of 2014-2015 that peaked in June 2015.

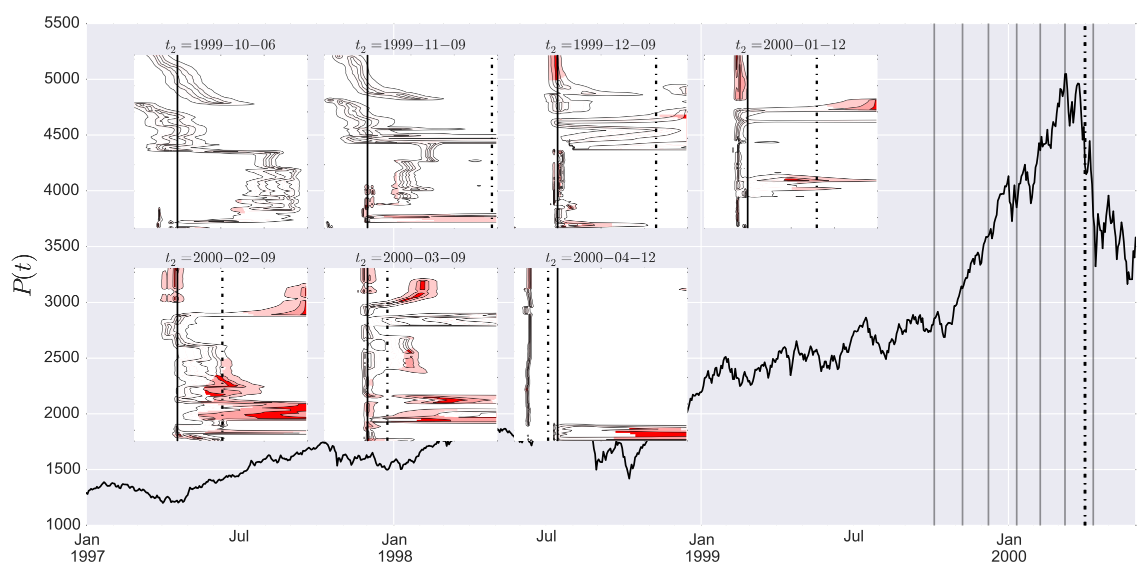


Figure 5.9: Same as Figure 5.6 for the dot-com bubble in the IT sector in the US culminating with a crash in April 2000

Chinese New Year in 2015, when exchanges were closed over February 7-13), we fill the gaps with the closing price of the previous day. For calibrations using business time such data preprocessing would not be necessary.

Employing the procedure explained in Sec. 5.6.2, we obtain Figures 5.7-5.9. In each of these three figures, the main graph shows the price time series $P(t)$ together with vertical dashes lines that identify remarkable turning points of the price dynamics. In the case of the S&P 500, we show two different vertical dashed lines associated with the two peaks of the index preceding the crash. The seven thin vertical lines indicate the position of the seven t_2 values chosen for the construction of the Likelihood intervals of t_c. The seven insets show contour plots of the likelihood intervals at 5%, 50% and 95% cutoff levels (as in Figure 5.6).

Figures 5.7 for the S&P 500 shows that the Profile Likelihhod of t_c as a function of time scale $\Delta_t \equiv t_2 - t_1$ and "present time" t_2 is very similar to those obtained in synthetic tests. As early as $t_2 =$ 1987-04-15, one can visualize the high Likelihood of $\widehat{t_c} \approx Oct.$ 1987 over almost all time scales. Interestingly, the Likelihood interval narrows down as t_2 approaches the end of the bubble. Moreover, there is an increase of the number of qualified fits (those where constrains on the model parameters (1.2.1) are met when likelihood intervals (5.29) and (5.31) are taken into account — shown as the red-shaded region) at t_2 increases. These two results can be rationalized by the fact that more information relevant to the identification of the bubble become available as more data are used.

As shown in figure 5.8, similar observations carry over to the SSEC bubble ending in June 2015, albeit with a smaller number of qualified fits. One can observe that the analyses performed for the time scales $\Delta_t \in [400, 500]$ and $\Delta_t \in [100, 200]$ provide a correct diagnostic of the end of the bubble with a narrow confidence interval. The time scale $\Delta_t \in [400, 500]$ correctly locks in on the true peak as early as April 2015.

Figure 5.9 shows the same analysis for the dotcom bubble that developed in the NASDAQ Index. At $t_2 = 2000\text{-}02\text{-}09$, the time scales $\Delta_t \in [100, 350]$ correctly lock in on the true peak $\approx$ April 2000. The other intermediate time scales give an estimation $\widehat{t}_c$ of the end of the bubble that agrees with the empirical value within the 95% confidence of the likelihood intervals.. All estimates on different t_2's appear to be either unqualified or signalling a different value for the change of regime to occur.

Overall, these empirical results exhibit the following behaviors: (i) for t_2 far from t_c, there are fewer qualified fits and different scales tend to provide distinct estimates $\widehat{t}_c$; (ii) when approaching the true t_c, the Likelihood intervals for $\widehat{t}_c$ start to align, with the formation of clusters associated with different possible scenarios; (iii) rather close to the true t_c, one can often observe a strong cluster around $\widehat{t}_c \approx t_2$ and a narrow likelihood interval.

On the other hand, the fact that different time scales used for fitting the LPPLS model tend to suggest different values of $\hat{t}_c$ is important to keep in mind, as this observation is in contrast with the behavior obtained for synthetic time-series. This is likely due to "model error", i.e., the simple LPPLS model (1.4) is only an approximation of the unknown true generating process of the price dynamics. For instance, earlier works [113, 59, 46, 128] have pointed out the important of including higher harmonics and more complex forms generalising this simple first-order LPPLS formula (1.4).

We thus stress the importance of employing filtering criteria to decrease the probability of the occurrence of errors of type I ("false positives"). The Likelihood Method has been shown to provide more reliable interval estimates for the critical time than simple OLS point estimates, in particular as t_2 approaches t_c.

5.7 Concluding remarks

We have presented a detailed methodological study of the application of the modified profile likelihood method proposed by [11], with the goal of tackling the instabilities and uncertainties occurring in the calibration of nonlinear financial models characterised by a large number of parameters. We have taken the Log-Periodic Power Law Singularity (LPPLS) model as an example for the application of the methodology. This is motivated by the claims of the LPPLS model to provide useful estimations of the end of bubbles and their crashes, which can be interpreted as critical times t_c. One of our major advances has been to formulate the calibration procedure in a way such that the critical time t_c of a given bubble becomes the major parameter of interest in the likelihood inference. In contrast, the other model parameters are treated as nuisance parameters. While the problem of dealing with nuisance parameters is not new in Statistics, the present article is, to our knowledge, the first one in quantitative finance that elaborate in details how to deal with them to obtain better inference on the target parameter (here t_c). We have shown that it is possible to bypass the strong nonlinearity of the model by using a very precise approximation for the modified profile likelihood. This has allowed us to provide a systematic construction of the parameter estimation uncertainties and of the corresponding likelihood intervals, both for the target parameter t_c and for the other so-called nuisance parameters. We have also introduced the importance of performing the calibrations at multiple time scales, i.e., in time windows of many different sizes typically from 100 to 750 days. This has led us to provide representations to aggregate the results obtained from the calibrations at different time scales, thus obtaining a multi-scale picture of the possible scenarios for the development of on-going bubbles. We have tested the methodology on synthetic price time series and on three well-known historical financial bubbles.

Acknowledgments

We are grateful to Professor Thomas A. Severini for helpful discussions about the approximation of the modified profile likelihood function. We also thank Diego Ardila Alvarez for many fruitful discussions while preparing the manuscript. Finally, we are grateful to the anonymous reviewer for constructive and useful feedbacks on the draft of the paper.

The analysis in the paper was performed using open source software: Python 2.7 (http://www.python.org) and libraries: Pandas [75], NumPy (http://www.numpy.org/), SciPy (http://www.scipy.org/), IPython [83], Matplotlib [53] and Seaborn (https://www.stanford.edu/~mwaskom/software/seaborn).

Derivation of the approximated modified profile likelihood for the LPPLS model

We derive the approximated expression for the modified profile likelihood (6.24) of the LPPLS model (1.6). The parameter of interest is the critical time t_c and nuisance parameters $\eta = \{\psi, s\}$ include both the vector $\psi = \{m, \omega, A, B, C_1, C_2\}$ of other LPPLS parameters and the variance s of the error term.

Observed Fisher information

The calculation of the observed Fisher information matrix $I(\hat{\eta}_{t_c})$ is straightforward. According to (5.10), it can be written in the form of a block matrix:

$$I(\hat{\eta}_{t_c}) = -\begin{pmatrix} \ell_{\psi,\psi}(\hat{\eta}_{t_c}) & \ell_{\psi,s}(\hat{\eta}_{t_c}) \\ \ell^T_{s,\psi}(\hat{\eta}_{t_c}) & \ell_{s,s}(\hat{\eta}_{t_c}) \end{pmatrix}, \tag{5.33}$$

where $\ell_{\cdot,\cdot}(\hat{\eta}_{t_c})$ denotes the respective second partial derivatives of the log-likelihood $\ln L(t_c, \psi, s)$ evaluated at the point $\hat{\eta}_{t_c} = \{\hat{\psi}_{t_c}, \hat{s}_{t_c}\}$. For the likelihood (5.3), the first partial derivatives are given by:

$$\begin{aligned} \frac{\partial \ell}{\partial \psi} &= -\frac{1}{2s}\frac{\partial \mathrm{SSE}(t_c,\psi)}{\partial \psi}; \\ \frac{\partial \ell}{\partial s} &= -\frac{n}{2s} + \frac{\mathrm{SSE}(t_c,\psi)}{2s^2}\ . \end{aligned} \tag{5.34}$$

In turn, the second partial derivatives read:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \psi \partial \psi^T} &= -\frac{1}{2s}\frac{\partial^2 \mathrm{SSE}(t_c,\psi)}{\partial \psi \partial \psi^T}; \\ \frac{\partial^2 \ell}{\partial \psi \partial s} &= \frac{1}{2s^2}\frac{\partial \mathrm{SSE}(t_c,\psi)}{\partial \psi}; \\ \frac{\partial^2 \ell}{\partial s^2} &= \frac{n}{2s^2} - \frac{\mathrm{SSE}(t_c,\psi)}{s^3}. \end{aligned} \tag{5.35}$$

The MLE $\hat{\eta}_{t_c}$ is given by the global maximum of $\ln L(t_c, \eta)$ for fixed t_c, so $\hat{\psi}_{t_c}$ is given by a global minimum of $\mathrm{SSE}(t_c, \psi)$, thus:

$$\ell_{\psi,s}(\hat{\eta}_{t_c}) = \left.\frac{\partial^2 \ell}{\partial \psi \partial s}\right|_{\eta=\hat{\eta}_{t_c}} = \frac{1}{2\hat{s}^2_{t_c}}\left.\frac{\partial \mathrm{SSE}(t_c,\psi)}{\partial \psi}\right|_{\psi=\hat{\psi}_{t_c}} = \Theta, \tag{5.36}$$

where $\Theta = \{0, 0, 0, 0, 0, 0\}^T$ is the vector of zeros. Taking into account (5.6), we can write for the third term in (5.35):

$$\ell_{s,s}(\hat{\eta}_{t_c}) = \left.\frac{\partial^2 \ell}{\partial s^2}\right|_{\eta=\hat{\eta}_{t_c}} = \frac{n}{2\hat{s}^2_{t_c}} - \frac{\mathrm{SSE}(t_c,\hat{\psi}_{t_c})}{\hat{s}^3_{t_c}} = -\frac{n}{2\hat{s}^2_{t_c}}. \tag{5.37}$$

Finally, plugging (5.36) and (5.37) into (5.33), we obtain the following form for the observed Fisher information:

$$I(\hat{\eta}_{t_c}) = \begin{pmatrix} \frac{1}{2\hat{s}_{t_c}} \left.\frac{\partial^2 \mathrm{SSE}(t_c, \psi)}{\partial\psi\partial\psi^T}\right|_{\psi=\hat{\psi}_{t_c}} & \Theta \\ \Theta^T & \frac{n}{2\hat{s}_{t_c}^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\hat{s}_{t_c}} \sum_{i=1}^{n} \left.\frac{\partial^2 \mathrm{LPPLS}(\tau_i; t_c, \psi)}{\partial\psi\partial\psi^T}\right|_{\psi=\hat{\psi}_{t_c}} & \Theta \\ \Theta^T & \frac{n}{2\hat{s}_{t_c}^2} \end{pmatrix}, \tag{5.38}$$

and its determinant

$$|I(\hat{\eta}_{t_c})| = \frac{n}{2\hat{s}_{t_c}^{p+2}} \left| \sum_{i=1}^{n} \frac{\partial^2 \mathrm{LPPLS}(\tau_i; t_c, \psi)}{\partial\psi\partial\psi^T} \right|_{\psi=\hat{\psi}_{t_c}}, \tag{5.39}$$

where $p = \dim\psi = 6$.

Covariance matrix

Here, we calculate the covariance matrix $\Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})$ (6.23). For this, we will first evaluate the general form of the matrix (6.23) and then substitute $(t_c, \hat{\eta}_{t_c})$ and $(\hat{t}_c, \hat{\eta})$. Similarly to the Fisher information, the matrix Σ (6.23) can be written in a block form:

$$\Sigma\left(t_{c;1}, \eta_1; t_{c;2}, \eta_2\right) = \mathrm{E}_{(2)} \left[\begin{pmatrix} \ell_\psi(1)\ell_\psi^T(2) & \ell_\psi(1)\ell_s(2) \\ \ell_s(1)\ell_\psi^T(2) & \ell_s(1)\ell_s(2) \end{pmatrix} \right], \tag{5.40}$$

where $\ell.(1)$ symbolically denotes the first partial derivative (5.34) of the log-likelihood evaluated at $(t_{c;1}, \eta_1) = (t_{c;1}, \psi_1, s_1)$; similarly, $\ell.(2)$ is evaluated at $(t_{c;2}, \eta_2) = (t_{c;2}, \psi_2, s_2)$. Given (1.10), the partial derivative of the SSE that enters (5.34) has the form:

$$\frac{\partial \mathrm{SSE}(t_c, \psi)}{\partial\psi} = -2\sum_{i=1}^{n} \varepsilon_i \frac{\partial \mathrm{LPPLS}_i}{\partial\psi}, \tag{5.41}$$

where we have denoted $\varepsilon_i = \varepsilon(\tau_i; t_c, \psi)$ and $\mathrm{LPPLS}_i = \mathrm{LPPLS}(\tau_i; t_c, \psi)$.

Let us first consider the cross-terms in (5.40). We substitute (5.41) into (5.34) and then into (5.40). Then, after replacing the product of sums with the double sum and using the linearity of the expectation operation, we have:

$$\mathrm{E}_{(2)}[\ell_\psi(1)\ell_s(2)] = -\frac{n}{2s_1s_2}\sum_{i=1}^{n}\mathrm{E}_{(2)}[\varepsilon_i]\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi}\right|_{\substack{t_c=t_{c;1}\\ \psi=\psi_1}} + \frac{1}{2s_1s_2^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}_{(2)}[\varepsilon_i\varepsilon_j^2]\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi}\right|_{\substack{t_c=t_{c;1}\\ \psi=\psi_1}}. \tag{5.42}$$

As discussed in Section 5.4.3, the expectations in (5.42) are taken with respect to the probability distribution that corresponds to the parameters $\{t_{c;2},\eta_2\}$, in other words under the assumption that $\varepsilon\sim N(0,s_2)$. Thus $\mathrm{E}_{(2)}[\varepsilon_i]=\mathrm{E}_{(2)}[\varepsilon_i\varepsilon_j^2]=0$, and the cross-term is equal to the zero-vector: $\mathrm{E}_{(2)}[\ell_\psi(1)\ell_s(2)]=\Theta$. Similarly for the second cross-term of (5.40):

$$\mathrm{E}_{(2)}[\ell_s(1)\ell_\psi^T(2)] = \left(\mathrm{E}_{(2)}[\ell_\psi(1)\ell_s(2)]\right)^T = \Theta^T. \tag{5.43}$$

Let us now consider the second derivatives with respect to the variance parameter s. Proceeding in the same way as above, we obtain:

$$\mathrm{E}_{(2)}[\ell_s(1)\ell_s(2)] = \frac{n^2}{4s_1s_2} - \frac{n}{4s_1^2s_2}\sum_{i=1}^{n}\mathrm{E}_{(2)}[\varepsilon_i^2] - \frac{n}{4s_1s_2^2}\sum_{j=1}^{n}\mathrm{E}_{(2)}[\varepsilon_j^2] + \frac{1}{4s_1^2s_2^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}_{(2)}[\varepsilon_i^2\varepsilon_j^2]. \tag{5.44}$$

Taking into account that $\mathrm{E}_{(2)}[\varepsilon_i^2]=s_2$, $\mathrm{E}_{(2)}[\varepsilon_i^4]=3s_2^2$ and $\mathrm{E}_{(2)}[\varepsilon_i^2\varepsilon_j^2]=s_2^2$ (when $i\neq j$), we obtain:

$$\mathrm{E}_{(2)}[\ell_s(1)\ell_s(2)] = \frac{n}{2s_1^2}. \tag{5.45}$$

Finally, the submatrix term reads:

$$\begin{aligned}
\mathrm{E}_{(2)}[\ell_\psi(1)\ell_\psi^T(2)] &= \frac{1}{s_1 s_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}_{(2)}[\varepsilon_i\varepsilon_j]\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi}\right|_{\substack{t_c=t_{c;1}\\ \psi=\psi_1}}\left.\frac{\partial \mathrm{LPPLS}_j}{\partial\psi^T}\right|_{\substack{t_c=t_{c;2}\\ \psi=\psi_2}} \\
&= \frac{1}{s_1}\sum_{i=1}^{n}\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi}\right|_{\substack{t_c=t_{c;1}\\ \psi=\psi_1}}\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi^T}\right|_{\substack{t_c=t_{c;2}\\ \psi=\psi_2}},
\end{aligned} \tag{5.46}$$

where we have accounted for the fact that $\mathrm{E}_{(2)}[\varepsilon_i\varepsilon_j] = 0$ when $i \neq j$.

The final expression is obtained by combining (5.43),(5.45) and (5.46) to (5.40) and evaluating the result at $(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})$:

$$\begin{aligned}
|\Sigma(t_c,\hat{\eta}_{t_c};\hat{t}_c,\hat{\eta})| &= \begin{vmatrix} \dfrac{1}{\hat{s}_{t_c}}\displaystyle\sum_{i=1}^{n}\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi}\right|_{\substack{t_c=t_c\\ \psi=\hat{\psi}_{t_c}}}\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi^T}\right|_{\substack{t_c=\hat{t}_c\\ \psi=\hat{\psi}}} & \Theta \\ \Theta^T & \dfrac{n}{2\hat{s}_{t_c}^2} \end{vmatrix} \\
&= \frac{n}{2\hat{s}_{t_c}^{2+p}}\left|\sum_{i=1}^{n}\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi}\right|_{\substack{t_c=t_c\\ \psi=\hat{\psi}_{t_c}}}\left.\frac{\partial \mathrm{LPPLS}_i}{\partial\psi^T}\right|_{\substack{t_c=\hat{t}_c\\ \psi=\hat{\psi}}}\right|,
\end{aligned} \tag{5.47}$$

where $p = \dim\psi = 6$. Note that a similar expression presented in [90] contains a typographical error in the power of the variance term.

Partial derivatives of the LPPLS function

We present here the analytical expressions of the first and second partial derivatives of the LPPLS function (1.6), which are necessary for the calculation of the modified profile likelihood (5.22)–(5.24).

The first-order derivatives have the following forms:

$$
\begin{aligned}
\partial \mathrm{LPPLS}/\partial m &= |t_c - t|^m \ln|t_c - t| \Big[B + C_1 \cos\big(\omega \ln|t_c - t|\big) + C_2 \sin\big(\omega \ln|t_c - t|\big) \Big]; \\
\partial \mathrm{LPPLS}/\partial \omega &= |t_c - t|^m \ln|t_c - t| \Big[-C_1 \sin\big(\omega \ln|t_c - t|\big) + C_2 \cos\big(\omega \ln|t_c - t|\big) \Big]; \\
\partial \mathrm{LPPLS}/\partial A &= 1; \\
\partial \mathrm{LPPLS}/\partial B &= |t_c - t|^m; \\
\partial \mathrm{LPPLS}/\partial C_1 &= |t_c - t|^m \cos\big(\omega \ln|t_c - t|\big); \\
\partial \mathrm{LPPLS}/\partial C_2 &= |t_c - t|^m \sin\big(\omega \ln|t_c - t|\big).
\end{aligned}
\tag{5.48}
$$

The second-order derivatives $\partial^2 \mathrm{LPPLS}/\partial\psi_i \partial\psi_j$, which are needed for the calculation of the matrix H (5.24), have the following form (omitting equivalent symmetrical entries, i.e.: $\partial^2 \mathrm{LPPLS}/\partial m \partial\omega \equiv \partial^2 \mathrm{LPPLS}/\partial\omega\partial m$):

$$
\begin{aligned}
\partial^2 \mathrm{LPPLS} \;/\; \partial m^2 &= |t_c - t|^m \left(\ln|t_c - t|\right)^2 \Big[B + C_1 \cos\big(\omega \ln|t_c - t|\big) + C_2 \sin\big(\omega \ln|t_c - t|\big) \Big]; \\
\partial^2 \mathrm{LPPLS} \;/\; \partial m \partial \omega &= |t_c - t|^m \left(\ln|t_c - t|\right)^2 \Big[-C_1 \sin\big(\omega \ln|t_c - t|\big) + C_2 \cos\big(\omega \ln|t_c - t|\big) \Big]; \\
\partial^2 \mathrm{LPPLS} \;/\; \partial m \partial B &= |t_c - t|^m \ln|t_c - t|; \\
\partial^2 \mathrm{LPPLS} \;/\; \partial m \partial C_1 &= |t_c - t|^m \ln|t_c - t| \cos\big(\omega \ln|t_c - t|\big); \\
\partial^2 \mathrm{LPPLS} \;/\; \partial m \partial C_2 &= |t_c - t|^m \ln|t_c - t| \sin\big(\omega \ln|t_c - t|\big); \\
\partial^2 \mathrm{LPPLS} \;/\; \partial \omega^2 &= -|t_c - t|^m \left(\ln|t_c - t|\right)^2 \Big[C_1 \cos\big(\omega \ln|t_c - t|\big) + C_2 \sin\big(\omega \ln|t_c - t|\big) \Big]; \\
\partial^2 \mathrm{LPPLS} \;/\; \partial \omega \partial C_1 &= -|t_c - t|^m \ln|t_c - t| \sin\big(\omega \ln|t_c - t|\big); \\
\partial^2 \mathrm{LPPLS} \;/\; \partial \omega \partial C_2 &= |t_c - t|^m \ln|t_c - t| \cos\big(\omega \ln|t_c - t|\big).
\end{aligned}
\tag{5.49}
$$

All other second-order partial derivatives are equal to zero.

Jacobian matrix for the damping parameter

The Jacobian matrix for the parameter transformation from $\eta = \{m, \omega, A, B, C_1, C_2, s\}$ to $\zeta = \{D, \omega, A, B, C_1, C_2, s\}$, where $D = m|B|/\omega|C|$ has the following form

$$J_D = \frac{\partial \eta}{\partial \zeta} = \begin{pmatrix} \frac{\omega|C|}{|B|} & \frac{D|C|}{|B|} & 0 & -\frac{D\omega|C|}{B|B|} & \frac{D\omega C_1}{|B||C|} & \frac{D\omega C_2}{|B||C|} & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.50}$$

where $|C| = \sqrt{C_1^2 + C_2^2}$.

Illustration of the differences between profile and modified profile likelihood intervals of nuisance parameters

Figure (5.10) illustrates a situation in which the approximate likelihood intervals (5.29) are misleading. It presents the profile and modified profile likelihoods for the nuisance parameters m and ω obtained by calibrating the LPPLS model to the Chinese SSEC Index in the time window from t_1=2006-05-04, t_2=2007-10-31 and at the fixed t_c= 2007-11-20. In contrast to the typical situation shown in Figure 5.4, one can clearly observe a bi-modal structure of the profile likelihoods of the nuisance parameters. Such bi-modal structure cannot be well described by intervals derived from a Fisher information-based likelihood. Moreover, this figure illustrates a case when the second-order modified profile

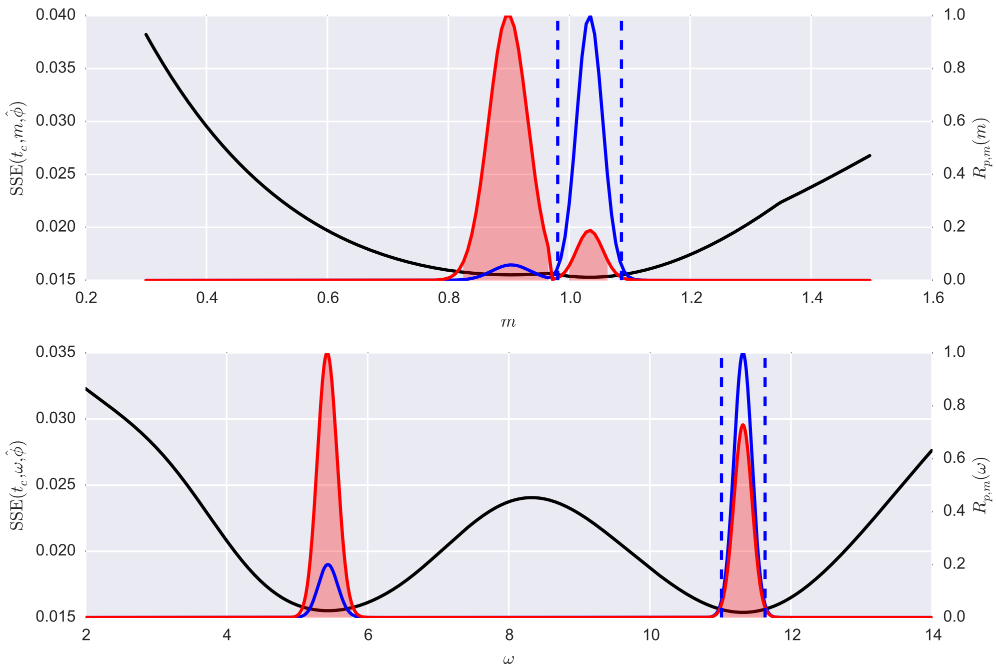


Figure 5.10: Profile of the cost function $F(\cdot)$ (black line, left scale), profile likelihood $L_p(\cdot)$ (blue line, right scale) and modified profile likelihood (red line, right scale) for the power law exponent m (top subplot) and the logperiodic angular frequency ω (bottom subplot) for t_c=2007-11-20. The red shaded intervals show the likelihood intervals LI$(\cdot)$ at the 5% cutoff level. Vertical blue dashed lines denote approximated likelihood intervals (5.29) at the 5% cutoff level. The calibration of the LPPLS model is performed on the Chinese SSEC Index for the bubble that bursts in June 2015.

likelihood suggests different estimated value of m and ω compared with the standard MLE: profile and modified profile likelihood have maxima at different points (similarly to the situation of the critical time in Figure 5.3).

While these situations are rather rare according to our experience, one needs to be aware that the approximate relations (5.29) might not reflect the full complexity of the structure of residuals.

Chapter 6

Hierarchical calibration of sloppy statistical models based on information geometry

6.1 Introduction

Parameter estimation in nonlinear dynamic models remains a very challenging problem due to its nonconvexity and ill-conditioning. In this paper, we stress that model, data and optimisation approach form a new complex system and point to the need of a theory that addresses this problem more generally. We show how information geometry-based estimations alongside with Profile Likelihood inference ensure the optimal trade-off between bias and variance allowing the incorporation of prior knowledge in a systematic way.

Properly calibrating multi-parameter models cannot be considered a trivial task. A successful calibration is not only data dependent but it is also a function of the proper choice the the minimization algorithm (derivative-based algorithms vs. simplex methods), the imposition (or not) of boundaries on parameter values, proper starting values and of-course on the structure of the model itself. Given the non-linear interaction among its different parts, model calibration can be viewed as a complex system and thus urges for a theory that addresses this problem more generally [118].

The more prone it is to a phenomena denoted sloppiness [74], i.e. parameter non-identifiability. A model it is said to be sloppy whenever it's output sensitivity to certain parameter variations is ~ 0 such as statistical inferences based on estimates is prone to significant uncertainty. This is the case whenever parameters have compensatory effects (i.e. collinearity) and thus cannot be individually identified. As we shall see, the sensitive (rigid) and sloppy (or poorly determined) parameters of a model can be diagnosed through the FIM eigenvalues.

According to [123], sloppiness is intrinsically a parametric issue rather than a model one. While the statistical uncertainty in each individual parameter might be infinite, the data places constraints on combinations of the parameters making the model still very useful. More precisely, on a macroscopic perspective a model it is still able to display low estimation uncertainty even with large individual parameter uncertainty if the directions of predictions on parameter space are constrained by data, which is often the case [74].

Instead of forgeting about precise estimates, In this chapter we argue that these parameter estimates should not be neglected. We stress that individual parameter uncertainty can be drastically reduced and practical non-identifiable parameters can be made identifiable via the employment of the proper parameter estimation technique. Our reasoning presented here is very much important for anyone who is not only interested on fitting models to data but rather obtaining precise parameter estimates.

The literature on how to enhance parametric estimation precision is vast [11, 12, 32]. Techniques such as profiling the likelihood function and concentrating the likelihood of a given parameter θ are examples where the nuisance parameters (i.e. unimportant) optimal values are constrained on fixed values of the parameter of interest.

By aligning information from Eigenvalues and Eigenvector of the FIM with a powerful estimation technique we show that calibrating models in a hierarchical fashion yield superior parameter estimates. The methodology is tested on a simple GARCH(p,q) model with distinct estimation techniques: Quasi-Likelihood (QML), Profile-Likelihood (LP) and Modified Profile-Likelihood (MPL). Results from a Monte Carlo simulation study suggests that I) profiling sloppy parameters with the MPL estimator yields superior estimates but when one uses MPL for each parameter of a practical non-identifiable model and then combine results of all estimates yields the supreme result.

The paper is structured as follows. Using a GARCH(1,1) model, Section (6.2) presents the reasoning behind the eigenvalues and eigenvectors of the Fisher Information Matrix and how these metrics can be used as a tool for picking the right focal parameter to "*rigidify*" a practical non-identifiable estimation problem. The results of our Monte-Carlo simulation studies are depicted in Sec. (6.2.3) and we conclude in Sec. (6.3).

6.2 The Hierarchical Calibration Technique: Models and Estimators

In this section we define the models, likelihood estimators and the Monte Carlo horse-race study employed in this paper in order to show that the way one proceeds to calibrate a model matters if he or she expects to obtain proper results. We explore the calibration of a GARCH(p,q) model using following estimators: Quasi-Maximum Likelihood, Profile Likelihood and Modified Profile Likelihood using different focal parameters.

6.2.1 The GARCH(p,q) model

The benchmark model in the literature for modelling conditional volatility processes observed on excess returns $:= r_t$ of financial time-series is the GARCH(p,q) model. Assuming a zero conditional mean for r_t, the model takes the following form

$$r_t = \sqrt{\sigma_t^2}\varepsilon_t, \quad \varepsilon_t \sim iid(0,1), \tag{6.1}$$

$$\sigma_t^2 = \omega + \gamma r_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{6.2}$$

for $t = [1, 2, \ldots, T]$ and p, q = 1. While parameter q denotes the order of the lagged (transformed) conditional variance, p refers to the symmetric innovation. The constrains $\omega > 0$, $\gamma \geq 0$, $\beta \geq 0$ and $(\gamma + \beta) < 1$ are imposed to guarantee that the unconditional variance of r_t remains finite, whereas its conditional variance σ_t^2 evolves over time [124].

The normal likelihood for the GARCH(1,1) model reads

$$f(\boldsymbol{r}|\boldsymbol{\theta}) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left(-\frac{r_t^2}{2\sigma_t^2}\right), \tag{6.3}$$

where $\boldsymbol{\theta} = [\omega, \gamma, \beta]$. We refer to estimates obtained by maximising (6.3) as the conditional maximum-likelihood estimates (MLEs) under normality assumption of innovations ε_t. In our implementation, we maximise the logarithm of (6.3) which is easier to handle. This yields the conditional log-likelihood function,

$$\mathcal{L}(\boldsymbol{r_t}|\boldsymbol{\theta}) = \sum_{t=p+1}^{T} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{r_t^2}{\sigma_t^2}\right], \tag{6.4}$$

$$\propto -\sum_{t=p+1}^{T} \left[\frac{1}{2}\ln(\sigma_t^2) + \frac{1}{2}\frac{r_t^2}{\sigma_t^2}\right], \tag{6.5}$$

where the term $\ln(2\pi)$ is omitted because it does not involve any parameter.

There is no closed-form solution for the MLE estimates so a common approach consists in the usage of numerical methods for solving the following optimisation problem:

$$\mathcal{L}(\hat{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{r_t}|\boldsymbol{\theta}), \tag{6.6}$$

Local search algorithms rely on information from the first and second order partial derivatives of the likelihood function 6.5 in order to update $\hat{\boldsymbol{\theta}}$. At the best-fit values of $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^*$, the first-order partial derivatives of Eq. 6.5 reads

$$\frac{\partial \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \omega} = \sum_{t=p+1}^{T} \left(\frac{r_t^2}{\sigma_t^2} - 1 \right) \frac{\partial \sigma_t^2}{\partial \omega} \frac{1}{\sigma_t^2}, \tag{6.7}$$

$$\frac{\partial \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \gamma} = \sum_{t=p+1}^{T} \left(\frac{r_t^2}{\sigma_t^2} - 1 \right) \frac{\partial \sigma_t^2}{\partial \gamma} \frac{1}{\sigma_t^2}, \tag{6.8}$$

$$\frac{\partial \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \beta} = \sum_{t=p+1}^{T} \left(\frac{r_t^2}{\sigma_t^2} - 1 \right) \frac{\partial \sigma_t^2}{\partial \beta} \frac{1}{\sigma_t^2}, \tag{6.9}$$

where,

$$\frac{\partial \sigma_t^2}{\partial \omega} = 1 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \omega}, \tag{6.10}$$

$$\frac{\partial \sigma_t^2}{\partial \gamma} = r_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \gamma}, \tag{6.11}$$

$$\frac{\partial \sigma_t^2}{\partial \beta} = \sigma_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \beta}. \tag{6.12}$$

Meanwhile, second-order derivates evaluated at the optimal parameter values $\boldsymbol{\theta^*}$,

$$\left. \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta\prime}} \right|_{\theta=\theta^*} = \sum_{t=p+1}^{T} \left[\frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta\prime}} \right], \tag{6.13}$$

are given by

$$\frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}\prime} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial^2 \omega} & \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \gamma \partial \omega} & \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \beta \partial \omega} \\ \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \gamma \partial \omega} & \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial^2 \gamma} & \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \beta \partial \gamma} \\ \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \beta \partial \omega} & \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \beta \partial \gamma} & \frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial^2 \beta} \end{bmatrix}$$

where each entry reads,

$$\frac{\partial^2 \mathcal{L}(r_t|\boldsymbol{\theta})}{\partial \theta \partial \theta\prime} = \left(\frac{r_t^2}{2\sigma^2} - \frac{1}{2\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \theta \partial \theta\prime} + \left(\frac{1}{2\sigma_t^2} \frac{r_t^2}{\sigma_t^3} \right) \frac{\partial \sigma_t}{\partial \theta} \frac{\partial \sigma_t}{\partial \theta\prime}. \tag{6.14}$$

At the best-fit values Eq. (6.13) is negative definite. Through the eigenvalues (λ) and eigenvectors of the Fisher Information Matrix, a natural parameter ordination in terms of importance appears and can be captured by the *Sloppy degree* $:= \mathcal{S}_d$ metric, defined as follows:

$$|\mathcal{S}_d| := \log\left(\frac{\lambda_i}{\lambda_{max}} \right) \forall\, i = \{1, \ldots, dim(\boldsymbol{\theta})\}, \tag{6.15}$$

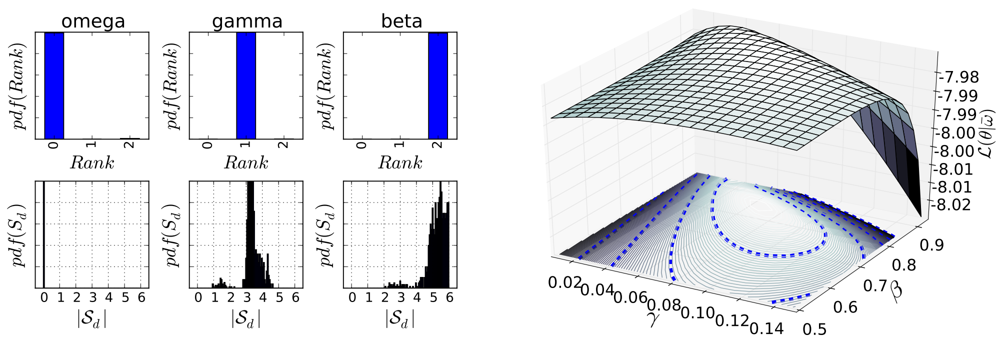
where λ is the eigenvalue of the *i-th* parameter $\in \boldsymbol{\theta}$. The metric is rich in information about the landscape of the likelihood function at $\hat{\boldsymbol{\theta}}$ and such information is the backbone of dimension-reduction techniques such as PCA [61]. $|\mathcal{S}_d|$ can be interpreted as follows: Assume parameter $\beta \in \boldsymbol{\theta}$ displays $|\mathcal{S}_d| = 4$. Embedded on the Euclidian Space at $\Re^{++}$, the metric stresses that the parametric region of a given model about the best-fit value is $1/\sqrt{10^{-\mathcal{S}_d}} \equiv 1/\sqrt{10^{-4}}$ longer than wider suggesting therefore a significant uncertainty on parameter inference due to multi-collinearity issues (parameter β and one or more predictor variables in a multiple regression model are highly correlated). Such statement holds true for any value of $|\mathcal{S}_d| \leq 3$ [74] but it reality the threshold value varies depending on the model. In order to visualise this feature, we show the contours of constant cost of the

Likelihood function (6.5) for different combinations of GARCH parameters γ and β while optimising $\hat{\omega}$ (Fig. 6.1(b)). Around the best-fit region (depicted by the blue dashed line closer to the centre of the maxima) we can observe that values of γ and β can be slightly exchanged without interfering significantly on the behaviour of the Log-Likelihood.

By simulating 10000 times the conditional volatility process given by Eq. (6.2) with distinct initial parametrisation $\boldsymbol{\theta} = \{\vec{\omega}, \vec{\gamma}, \vec{\beta}\}$[1] and different sample sizes $N \in [100{:}1000]$ we report $\mathit{pdf}(\mathcal{S}_d)$ and the parameter ranking of the model on the left-hand side of Fig. (6.1). A clearly hierarchy arises as a result: ω has Rank $= 0$ and therefore stands as the most rigid parameter, followed by γ and β. Such ordination suggests, as mentioned previously, that disturbances on rigid parameter ω yields a more severe impact upon the likelihood function than the remaining parameters hence parameter $\hat{\omega}$ can be estimated from data with very low uncertainty while $\hat{\beta}$ cannot.

To understand why this happens, recall that the GARCH process describes the conditional covariance at time t as a function of lagged cross-products of errors and conditional expectations of these cross-products. Obviously both "explanatory variables" share much common information and in the neighborhood of the optimum, where conditional expectations achieve the closest fit to their arguments, are deemed to be very similar. As a consequence, point-estimates of the parameters will often be highly correlated and imprecise, as suggested by the eigenvalues of the FIM (Fig. 6.1(a)) and the quasi-flat landscape of the log-likelihood plotted at the optimal value of ω and different combinations of γ and β (Fig. 6.1(b)).

[1] The parametrisation used was the following: $\vec{\omega} = \{0.03, 0.06, 0.09\}$, $\vec{\gamma} = \{0.06, 0.03, 0.01\}$ and $\vec{\beta} = \{0.9, 0.85, 0.86\}$.



(a) (b)

Figure 6.1: **GARCH(1,1) parameters (θ) ranked according to $|S_d|$ (Fig. A) and Contours of Constant Cost (Fig. B):** GARCH(1,1) parameters were ranked according to their sloppy degree (see Eq. 6.15) (left hand-side upper panel) from more important (Rank = 0) to less important (Rank = 2) based on the ensemble of $10e^3$ Monte Carlo simulations of a GARCH(1,1) process with different sample sizes T $\in$ [100:1000] and different parametrization for $\boldsymbol{\theta}$. The density of the sloppiness degree of each parameter θ calculated for each simulation of length T through the Hessian matrix ($|S_d|$) is shown at the lower layer of Fig. 6.1(a). Larger values of $|S_d|$ imply a larger parameter uncertainty and vice-versa.

6.2.2 The estimators of the hierarchical methodology

Information Geometry obtained via Eigenvalues and Eigenvectors of the FIM are indeed a very informative for understanding the parametric structure of a model and how it adapts to empirical data. Besides combining this geometrical information with Likelihood estimators, this section provides an in-depth explanation behind the Quasi-Likelihood, Profile Likelihood and Modified Profile Likelihood estimation methodologies. It will become clear for the reader why our proposed calibration technique is the way to go when one is interested on parametric inferences.

As mentioned in the previous Section, there are moments when one is interested on obtaining more robust and precise estimates of the parameter of interest := $\hat{\phi}$. In such scenarios the Profile-Likelihood estimator can be used. The method replaces the nuisance parameters (i.e. unimportant) by their MLE at each fixed value of the parameter of interest. Assume for the moment that $\phi = \gamma$, then nuisance parameters would be $\eta = [\omega, \beta]$ while for $\phi = \omega$, then $\eta = [\gamma, \beta]$.

Given the joint Likelihood $\mathcal{L}(\phi, \eta)$, the Profile Likelihood $\mathcal{L}_p(\phi)$ is formally defined as

$$\mathcal{L}_p(\phi|\eta) = \arg\max_{\eta} \ \mathcal{L}(\phi, \eta) \equiv \mathcal{L}(\phi, \widehat{\eta}_\phi), \tag{6.16}$$

where $\widehat{\eta}_\phi$ is a MLE for nuisance for a fixed value of ϕ. The profile likelihood of ϕ (also known as concentrated likelihood) thus reads

$$\mathcal{L}_p \propto (\hat{s}_\phi)^{-\frac{T}{2}} \propto \Big(\mathcal{L}_p(\phi|\eta)\Big)^{-\frac{T}{2}}, \tag{6.17}$$

with $s = \frac{1}{T}\mathcal{L}(\phi)$. We normalise the Log-Likelihood to 1 by its maximum so that $\mathcal{L}(\phi|\hat{\eta}) \in [0, 1]$,

$$\bar{\mathcal{L}_p}(\phi|\hat{\eta}) = \frac{\mathcal{L}(\phi|\hat{\eta})}{max_\phi \mathcal{L}(\phi|\hat{\eta})}. \tag{6.18}$$

Profiling the Likelihood function is prone to estimation bias, specially when the sample does not provide enough information about η [90]. A solution to overcome this issue consists on using the Modified Profile Likelihood [13]. Recently, this methodology was successfully applied for the problem of quantifying the uncertainty permeating changes of regime of financial time-series [36].

By adding a modulating factor $:= \mathcal{M}(\phi)$ to $\mathcal{L}_p(\phi|\hat{\eta})$, the Modified Profile Likelihood estimator reads

$$\mathcal{L}_{mp}(\phi) = \mathcal{M}(\phi)\mathcal{L}_p \tag{6.19}$$

$$= |\mathcal{I}(\hat{\eta}_\phi)|^{-\frac{1}{2}} \Big|\frac{\partial \hat{\eta}}{\partial \hat{\eta}_\phi}\Big| \mathcal{L}_p(\phi), \tag{6.20}$$

where $\mathcal{I}(\hat{\eta}_\phi)$ is the *Fisher Information matrix* (FIM) approximated at the best-fit parameters η assuming ϕ known. The FIM reads

$$\mathcal{I}(\hat{\eta}_\phi) = -\frac{\partial^2 \ln \mathcal{L}(\phi, \eta)}{\partial \eta \, \partial \eta'}\bigg|_{\eta=\hat{\eta}} . \tag{6.21}$$

Meanwhile, the Jacobian term $:= \left|\partial\hat{\eta}/\partial\hat{\eta}_\phi\right|$ works as an "*Invariance preserving*" quantity which keeps the modified profile likelihood invariant with respect to transformations of the nuisance parameters [80]. Precisely, this term denotes a matrix of the first partial derivatives of the full MLE of the η with respect to the MLE calculated at a fixed value of $\bar{\phi}$.

Calculating the Jacobian term is not an easy task. We follow [88] who suggested approximating the matrix $\left|\partial\hat{\eta}/\partial\hat{\eta}_\phi\right|$ via the covariance matrix of the score function (i.e. the first order derivatives of the likelihood function with respect to parameters, see Eq. 6.7), that is

$$C(\phi, \hat{\eta}_\phi; \hat{\phi}, \hat{\eta}) \approx \Sigma(\phi, \hat{\eta}_{t_c}; \hat{\phi}, \hat{\eta}) \tag{6.22}$$

where

$$\Sigma\left(\phi_1, \eta_1; \phi_2, \eta_2\right) = \mathrm{E}_{(2)}\left[\frac{\partial \ln L(\phi,\eta)}{\partial \eta}\bigg|_{\substack{\phi=\phi_1\\ \eta=\eta_1}} \frac{\partial \ln L(\phi,\eta)}{\partial \eta^T}\bigg|_{\substack{\phi=\phi_2\\ \eta=\eta_2}}\right] . \tag{6.23}$$

Given the non-linear nature of the GARCH process, the term $\mathcal{L}_m(\phi)$ thus has the following form,

$$L_m(\phi) \approx \frac{|I(\hat{\eta}_\phi)|^{-\frac{1}{2}}}{\left|\Sigma(\phi, \hat{\eta}_\phi; \hat{\phi}, \hat{\eta})\right|} L_p(\phi|\hat{\eta}) \ . \tag{6.24}$$

6.2.3 Comparsion of esimators using the GARCH(1,1) model

In order to show just how important the parameter ordination is when calibrating models, we proposed a simulation study where different calibration methods and Likelihood estimators are employed to the problem of fitting synthetic time-series. We generated synthetic time-series of a GARCH(1,1) model $:= \vec{Y}$ plus random noise $\varepsilon \sim \mathcal{N}(0,1)$ using random parametrization $\boldsymbol{\theta}$ satisfying restrictions enumerated on Sec. (6.2.1). The resulting time series $\vec{Y_g}$ was in turn fitted using the same model with different estimators / estimation methodologies.

Let us now define a family of distinct estimators $\boldsymbol{\Phi}$ that will be used for calibrating the GARCH(1,1) model unto $\vec{Y_g}$ and whose performance will be evaluated via the metric χ^2, defined below. Each estimator $e \in \boldsymbol{\Phi}$ reads:

- $\boldsymbol{e_1} := \mathcal{L}_{qml}(\theta)$: The simple Quasi-Maximum Likelihood estimator.
- $\boldsymbol{e_2} := \mathcal{L}_{mp}(\omega|\hat{\eta})$: The Modified Profile Likelihood estimator with $\phi = \omega$. Here the rigid parameter is the first one to be calibrated while $\eta = \{\gamma, \beta\}$ are in turn estimated jointly and conditional on a fixed value of $\bar{\omega}$.
- $\boldsymbol{e_3} := \mathcal{L}_{mp}(\beta|\hat{\eta})$: The Modified Profile Likelihood estimator with $\phi = \beta$. We now choose the sloppiest parameter of the GARCH(1,1) model namely β to be ϕ now. After the estimate of $\hat{\beta}$ is obtained via the Modified Profile Likelihood estimator, nuisance parameters are estimated jointly via Quasi-Maximum Likelihood and conditional on $\hat{\beta}$.
- $\boldsymbol{e_4} := \mathcal{L}_{mp}(\omega|\hat{\eta}; \gamma|\hat{\eta}; \beta|\hat{\eta})$: Here we estimate all and each parameter $\boldsymbol{\theta}$ via $\mathcal{L}_{mp}$ and regroup them in the end of the exercise in order to yield the set $\hat{\boldsymbol{\theta}}$. More specifically, each parameter estimation is performed independently of the others for the same synthetic realization and grouped at the end yielding $\hat{\boldsymbol{\theta}}$.
- $\boldsymbol{e_5} := \mathcal{L}_p(\omega|\hat{\eta})$: The Profile Likelihood estimator with $\phi = \omega$. Same as e_2 but parameter estimation is performed using the Profile Likelihood estimator instead of the Modified version.

- $\boldsymbol{e_6} := \mathcal{L}_p(\beta|\hat{\eta})$: The Profile Likelihood estimator with $\phi = \beta$. Same as e_3 but parameter estimation is performed using the Profile Likelihood estimator instead of the Modified version.

- $\boldsymbol{e_7} := \mathcal{L}_p(\omega|\hat{\eta};\gamma|\hat{\eta};\beta|\hat{\eta})$: A concurrent version of method e_4 but now the estimator $\mathcal{L}_p$ is used instead of $\mathcal{L}_{mp}$. Parameters $\in \boldsymbol{\theta}$ are estimated one at a time and then compiled at the end of the exercise by the set $\hat{\boldsymbol{\theta}}$.

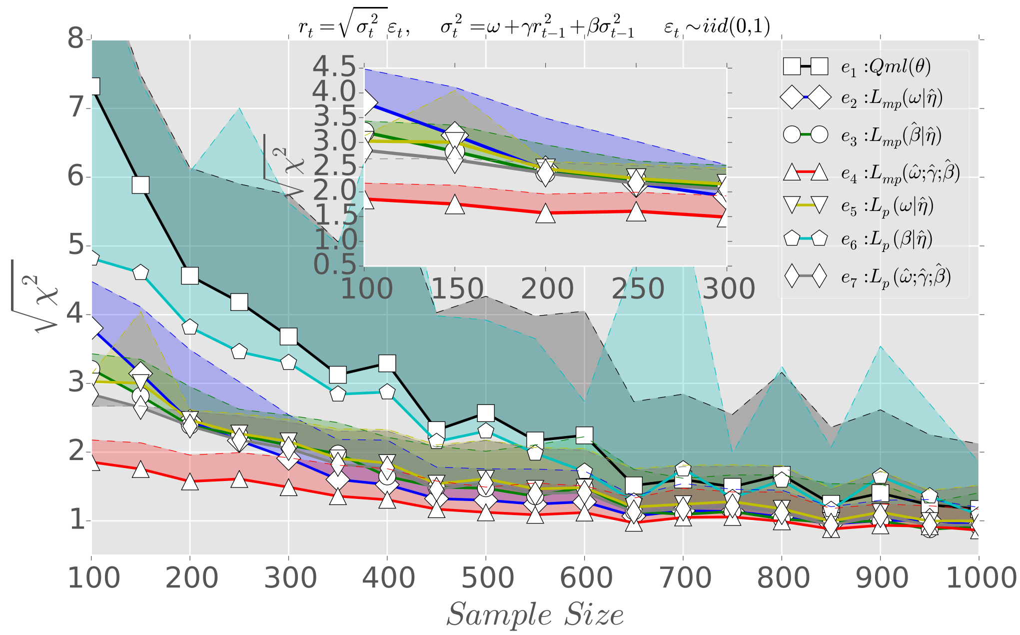


Figure 6.2: **Monte-Carlo simulation experiment for measuring the performance of estimators $e \in \Phi$ via metric 6.25:** For each sample size N $\in$ [100:1000] we simulate synthetic time-series of a GARCH(1,1) process from the DGP (6.2) plus random noise $\varepsilon \sim \mathcal{N}(0,1)$. The resulting synthetic data is in turn fitted using each estimator $e \in \phi$. For each N we repeat the previous mentioned process of generating and fitting synthetic data 400 times for increments on N of 50 units; N=$\{100, 150, 200, \ldots, 1000\}$. We depict the evolution of the first and second moments of the metric $\sqrt{\chi^2}$ for each value of N using continuous and dashed lines respectively. For clarity, the inset plot restrict the interval of N to $\{100 : 300\}$ and $\sqrt{\chi^2}$ to $\{0.5 : 4.5\}$.

Recall that our main interest is to validate some hypothesis: I) can the usage of information geometry yield superior parameter estimates than simply estimating all parameters of a model jointly? and II) can we enhance the performance of the Modified Profile Likelihood estimator depending on the choice of ϕ? In order to test these hypothesis we will make use of a performance metric for computing just how biased estimator $e \in \boldsymbol{\Phi}$ is for a given realisation of $\vec{Y_g}$. This can be achieved for each estimator e via

$$\chi^2[\mathcal{L}_e(\boldsymbol{\theta})] = \sum_{\theta_i}^{dim(\theta)} \left[\left(\frac{\mathcal{L}_e(\hat{\theta}_i) - \theta_i^*}{\theta_i^*}, \vec{Y_g} \right)^2 \right], \tag{6.25}$$

with $\boldsymbol{\theta}^* = \{\omega = 0.03; \gamma = 0.06; \beta = 0.9\}$. In order to obtain reliable statistics we undertook the following Monte-Carlo procedure: for each synthetic generated time-series $\vec{Y_g}$ of size N we fit the GARCH(1,1) model using each estimator $\boldsymbol{e} \in \boldsymbol{\Phi}$. The process is repeated 400 times for each combination of N and $\vec{Y_g}$ using $N = \{100, 150, 200, \ldots, 1000\}$. Results are depicted on Fig. (6.2) where the mean value of $\sqrt{\chi^2}$ for each estimator $\boldsymbol{e}$ is shown via continuous line and its corresponding standard deviation via dashed lines.

The superior performance clearly steams from estimator e_4. Specially for small samples, the estimator is able to retrieve true parameters from synthetic data roughly two times more precisely than the second best choice, e_8. Relative to the classical Quasi Maximum Likelihood estimator e_1, the best estimation technique yields a 6 order of magnitude precision gain, without taking the variance into account.

It is interesting to notice just how the choice of ϕ and the estimator technique impact the metric in terms of the first two moments (mean and variance), specially for small samples. Compare for example estimators 2 and 3; if one chooses to profile the sloppiest parameter (β) through the Modified Profile Likelihood estimator instead of the most rigid one (ω) the estimator variance turns out to be smaller than the mean of $\sqrt{\chi^2}$. This exercise clearly shows that prior information matters and that $L_{mp}(\beta|\hat{\eta})$ is as good as $L_p(\hat{\omega}, \hat{\gamma}, \hat{\beta})$ simply by taking geometrical information into consideration. In other words, the quality of results depend directly on the structure of the cost function of the model under consideration.

6.3 Conclusion

Our results suggest that profiling parameters always yield superior estimates in terms of precision and lower variance. However, depending on the nature of the problem one is facing these results can vary. More precisely, while the usage of information geometry for calibrating parameters of a practical non-identifiable model leads to superior results, this is not true when one is facing either identifiable problems (e.g. OLS or ARMA-like models) or a structural non-identifiable one (Sum of Decaying Exponentials), given their intrinsic parametric geometry.

We conclude stressing the existence of a hierarchy of parameter importance on statistical models that can be explored in order to improve estimates. Depending on the nature of the problem, information geometry can provide the necessary information for retrieving parameters from practical non-identifiable multi-parameter models (e.g. GARCH model and the LPPLS model).

It is important to stress that new models and more simulations studies with different parametrisation are required in order to attest the universality of our claim.

Chapter 7

Ongoing Diagnosis and Post-Mortem Analysis of the SSEC 2015 Bubble

7.1 Introduction

Mainly as a result of massive investments in real-estate and infrastructures over the past three decades, China's economy has moved from being largely closed to becoming a major global player. China's average annual GDP growth was 13% between 2000 and 2008, but has been slowing down to 7.8% in 2009 and to the estimate of 7% in 2015.

Accompanying this stellar growth, Chinese stock markets have experienced a roller coaster dynamics, with two large bubbles bursting respectively from May 2005 to Oct. 2007 and from Nov. 2008 to Aug. 2009 [55]. The latest bubble started around mid-2014 (see the analysis below for a qualification) and has recently started crashing in mid-June 2015. This last bubble corresponded to an approximate 150% growth in just one year. This growth of

Chinese stocks was all the more remarkable as it occurred at the time when the Chinese real estate market along with the overall economy was cooling significantly. This bubble can be seen as a result of a strong leverage that is disconnected from the realities of economic activity and corporate earnings. About 7% of China's population has been active in this stock market frenzy, profiting from the easy access to credit to invest in the stock markets. An interesting specificity of Chinese stock markets is that insurance firms and pension funds that are traditionally stabilising investors by their buy-and-hold strategies are essentially absent in the Chinese investing universe. As a consequence, about 90 millions of small and medium size Chinese investors constitute the main drivers of the stock markets, making it much more susceptible to rumours, imitation, speculations and crowd effects. In fact, there are many indications that the Chinese government has encouraged small retail investors to join in investing in the stock market, driving it up for a while but also catalysing its fragility. One should however temper this negative view by noting that Chinese households have most of their wealth in real-estate so that a crash cannot have the impact it would have in Western markets. Moreover, the spillover effect on Western markets should be minimal as most foreign investors cannot participate, removing their well-known unstable strategies of investing late and withdrawing at the time the crash to enhance the severity of corrections.

Since its peak on June 12, 2015, the SSEC (Shanghai composite index) has lost 32% to its bottom reached on July 8, 2015 and has since moved sideway with a large volatility. The smaller Shenzhen stock market has lost 41% over the same period. The Chinese government has taken unprecedented measures aimed at stopping the descent. Particularly, the benchmark lending and deposit rates have been cut several times by the People's Bank of China (PBoC) and are now at record lows, margin lending rules have been relaxed, a de-facto suspension of new IPOs has been declared, and "malicious short-selling" has been announced to be probed, and so on. Wild swings have continued to develop. On Monday July 27, 2015, Shanghai stocks lost 8.5%, thus suffering its worst one-day loss since 2007.

Fig. 7.1 compares visually the bubble that ended in 2007 to the present 2015 bubble, both on the Shanghai stock market (top panel). One can observe very similar price trajectories. In contrast, a similar comparison for the Hong Kong market (bottom panel) shows that the market development in the last year has not followed the same type of bubble behaviour as occurred before and until 2007. This is notwithstanding the fact that, on Nov. 10, 2014, the China Securities Regulatory Commission jointly with the Hong Kong Securities Regulatory Commission announced that a connection between the two stock markets would officially start on Nov.17 2014, which grants foreign investors unprecedented access to China's tightly controlled capital market. This is likely due to the fact that, in fact, mainland Chinese investors were the main traders in the Shanghai stock market while the Hong Kong market is much more open to foreign investors. The difference can thus be traced to the divergence on the views of China's prospects between domestic and foreign investors as well as the hope for quick gains fuelling speculation among retail Chinese investors. The decline in July 2015 suffered by the Hong Kong stock market demonstrates the spillover effect from the Shanghai market to the Hong Kong market, mainly via the so-called "Shanghai-Hong Kong stock connect" programme launched on Nov.17 2014.

In this chapter, we describe the remarkable success of the real-time diagnostic of the bubble regime that was obtained in our Financial Crisis Observatory (FCO: available at: `tasmania.ethz.ch/pubfco/fco.html`) that we are currently running at ETH Zürich. The analysis is based on the Johansen-Ledoit-Sornette (JLS) [56] model, which is built on (i) the economic theory of rational expectation bubbles, (ii) behavioral mechanisms of imitation and herding of investors and traders and (iii) the mathematical formulation

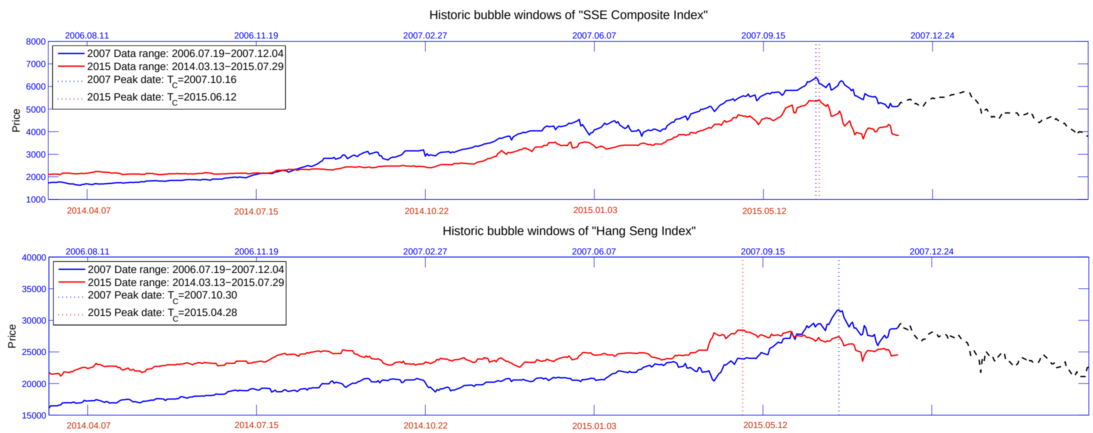


Figure 7.1: **top panel**: For the Shanghai stock market, comparison between the bubble that ended in October 2007 and the present one ending in June 2015. **bottom panel**: Same as top panel for the Hong Kong stock market. The blue line extended by dotted black line is the price time series within the window of [2006.07.19, 2008.03.25]. The red line corresponds to the window of [2014.03.13, 2015.07.29]. The blue dotted vertical lines $T_c = 2007.10.16$ (for Shanghai market) and $T_c = 2007.10.30$ (for Hong Kong market) represent the peak dates in 2007. The red dotted vertical lines $T_c = 2015.06.12$ (for Shanghai market) and $T_c = 2015.04.28$ (for Hong Kong market) represent the peak dates in 2015.

of the Log-Periodic Power Law Singularity (LPPLS) that describes the critical approach towards a tipping point in complex systems. The early warning signals that were reported in real time ex-ante provide strong additional supportive evidence for the relevance of the LPPLS-based methodology in the diagnostic of bubbles. We here document how the real-time predictions were presented in the automated analysis of the FCO, as well as in our monthly FCO Cockpit report of June 2015, where we have warned that Chinese equities were in bubble territory with the highest possible readings, that the price path followed was not sustainable and a correction was due. These results add to previous accounts of similar ex-ante advanced warning and predictions performed for the two previous financial bubbles in two most important Chinese stock indexes, Shanghai (US ticker symbol SSEC) and Shenzhen (SZSC), (i) from mid-2005, bursting in October 2007 and (ii) from November 2008, bursting in the beginning of August 2009 [55].

The article is organised as follows. Section 7.2.1 introduces the DS LPPLS Confidence and DS LPPLS Trust indicators, as well as the functioning of the FCO. Section 7.3 describes how the real-time diagnostic of the SSEC bubble was performed and communicated to the public and offers a summary of the evidence. Section 7.4 gives a complementary post-mortem analysis on the nature and value of the LPPLS methodology to diagnose the SSEC bubble and its termination. Section 7.5 concludes.

7.2 Methodology

7.2.1 LPPLS calibration and indicators

The model is calibrated on the data using the Ordinary Least Squares method, providing estimations of all parameters t_c, ω, m, A, B, C_1, C_2 in a given time window of analysis. We use the robust procedure proposed by [38], which reduces the estimation to just three nonlinear parameters m, ω and t_c. For each fixed data point t_2 (corresponding to a fictitious "present" up to which the data is recorded), we fit the price time series in shrinking windows (t_1, t_2) of length $dt := t_2 - t_1$ decreasing from 750 trading days to 125 trading days. We shift the start date t_1 in steps of 5 trading days, thus giving us 126 windows to analyse for each t_2. In order to minimise calibration problems and address the sloppiness of the model (1.10) with respect to some of its parameters (and in particular t_c), we use a number of filters to condition the solutions, which are summarised in Table 7.1. These filters derive from the empirical evidence gathered in investigations of previous bubbles [55]. Only those calibrations that meet the conditions given in Table 7.1 are considered valid and the others are discarded.

From the ensemble of qualified fits, we construct two DS bubble indicators as follows.

- **DS LPPLS Confidence:** It is the fraction of fitting windows for which the LPPLS

Item	Notation	Search space	Filtering condition 1	Filtering condition 2
3 nonlinear parameters	m	[0, 2]	[0.01, 1.2]	[0.01, 0.99]
	ω	[1, 50]	[2, 25]	[2, 25]
	t_c	$[t_2 - 0.2dt,$ $t_2 + 0.2dt]$	$[t_2 - 0.05dt,$ $t_2 + 0.1dt]$	$[t_2 - 0.05dt,$ $t_2 + 0.1dt]$
Number of oscillations	$\frac{\omega}{2} \ln \lvert \frac{t_c - t_1}{t_2 - t_1} \rvert$	—	$[2.5, +\infty)$	$[2.5, +\infty)$
Damping	$\frac{m\lvert B\rvert}{\omega\lvert C\rvert}$	—	$[0.8, +\infty)$	$[1, +\infty)$
Relative error	$\frac{p_t - \widehat{p}_t}{\widehat{p}_t}$	—	[0, 0.05]	[0, 0.2]

Table 7.1: Search space and filter conditions for the qualification of valid LPPLS fits. Within the JLS framework, the condition that the crash hazard rate $h(t)$ is non-negative by definition translates into a value of the Damping parameter $\frac{m|B|}{\omega|C|}$ larger than or equal to 1.

calibrations satisfy the filtering condition 1 in Table 7.1. It measures the sensitivity of the observed bubble pattern to the time scale dt. A large value indicates that the LPPLS pattern is present at most scales and is thus more reliable. A small value signals a possible fragility of the signal since it is present only in a few time windows.

- **DS LPPLS Trust:** Because the calibration is an attempt to disentangle the LPPLS signal from an unknown realisation of the residuals, it is important to assess the sensitivity of the results to different instances of these residuals. We thus resample 100 times the residuals and add them to the calibrated LPPLS structure to generate 100 synthetic price time series that proxy for 100 supposed independent realisations of equivalent price patterns. The DS LPPLS Trust indicator is the median level over the 126 time windows of the fraction among the 100 synthetic time series that satisfy the filtering condition 2 in Table 7.1. It measures how closely the theoretical LPPLS model matches the empirical price time series, 0 being a bad and 1 being a perfect match. As a rule of thumb, a value of DS LPPLS Trust larger than 5% indicates that the price process is not sustainable and there is a substantial risk for a critical transition to occur.

7.2.2 Financial Crisis Observatory

These two indicators 'DS LPPLS Confidence' and 'DS LPPLS Trust' constitute the core measures provided in the analyses presented daily on the website of the Financial Crisis Observatory (FCO) at ETH Zurich (`tasmania.ethz.ch/pubfco/fco.html`). Started in August 2008 in reaction to the on-going financial crisis, the FCO has the ambition to test and quantify rigorously, in a systematic way and on a large scale, the hypothesis that bubbles can be diagnosed before they burst and their end can be predicted probabilistically ex-ante. Before the launch of the daily watch of bubble indicators, a number of experiments have been performed, such as the financial bubble experiments.

The FCO provides the real-time values of the 'DS LPPLS Confidence' and 'DS LPPLS Trust' indicators, which are updated daily on 21 world stock markets, commodities, US sectors and US firms (our intra-group version developed for research purposes monitors a much larger number, approximately 25'000 assets). Every morning, the FCO system automatically acquires asset prices of the previous day and generates the DS LPPLS Confidence and Trust indicators for the previous day as described above. These daily updates have been freely available to anyone since 2012 and enjoy a suite of followers among private and professional investors. The philosophy driving the FCO is that only advanced forecasts can be free of data-snooping and of other statistical biases of standard ex-post tests. And it is up to everyone to assess the quality and usefulness of the indicators.

7.3 Real-time diagnostic of the 2015 SSEC bubble

The 2015 Shanghai bubble has been tracked and diagnosed in real-time with two different versions of our methodology that we now summarise.

7.3.1 The real-time daily FCO 'DS LPPLS Confidence' and 'DS LPPLS Trust' indicators

Fig. 7.2 shows a screenshot of the FCO website interface taken on Aug. 4, 2015, showing all monitored assets available to the public since 2012. Whenever our method detects a bubbly signature in the underlying price time series, a red bar is displayed. Each day, the screen is shifted by one day to include the new information, in a real causal time setting. For the 2015 SSEC bubble, one can observe that red bars form two clusters. The first one is from Dec. 1, 2014 to Jan. 1, 2015 and announced correctly the correction that occurred in January 2015. The second cluster starts on Apr. 4, 2015 and culminates on Jun. 17, 2015 with the start of the crash. This second cluster is actually made of two sub-structures, as shown in Fig.7.3.

Fig. 7.3 shows the value of the 'DS LPPLS Confidence' indicator in red together with the SSEC index in blue from July 2013 to July 2015. This graph could be retrieved each day up to the time of the observation, providing a real-time diagnostic of the on-going bubble. The website allows one to travel back in time to visualise the signal as was available at any time in the past. Note that the three peaks of the indicator anticipate shortly each the following most significant corrections exhibited by the SSEC index.

7.3.2 FCO cockpit reports

From February 2014 (with a gap for May to September 2014), and with an uninterrupted monthly periodicity since October 2014, we release so-called 'FCO cockpit reports' that provide a global status of the major bubbles developing in the world in all asset classes. The methodology is similar to that developed for the 'DS LPPLS Confidence' and 'DS LPPLS Trust' indicators presented above but is more coarse-grained to shape a general view of the developing bubble risks.

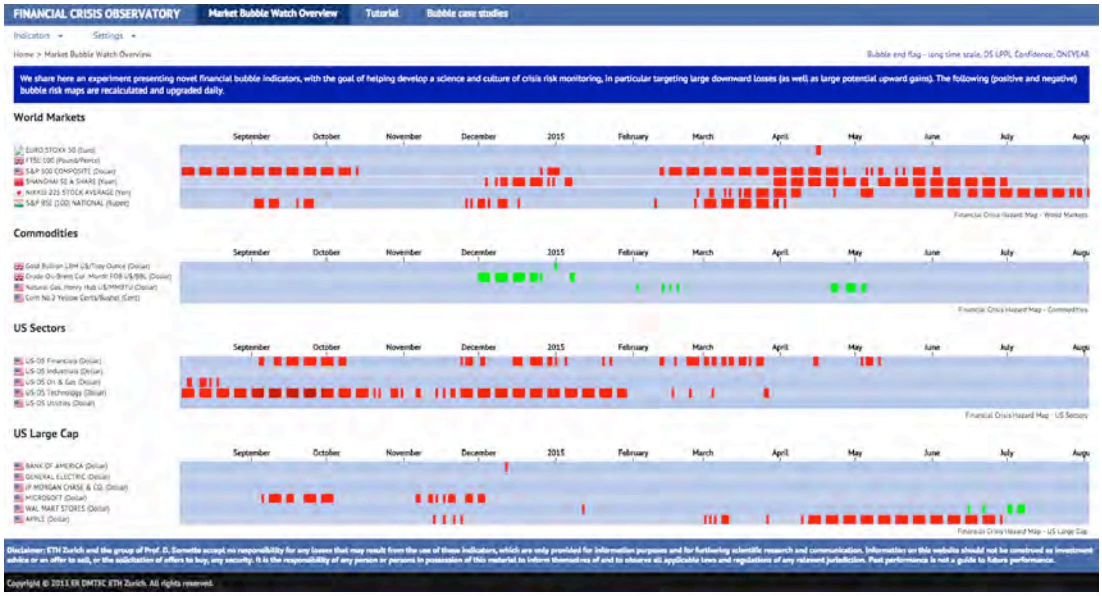


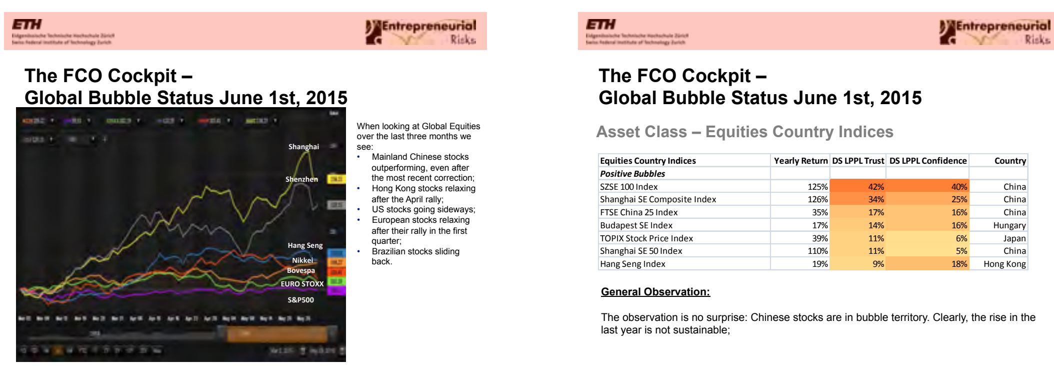
Figure 7.2: Snapshot of the FCO website (tasmania.ethz.ch/pubfco/fco.html) on Aug. 4, 2015. The SSEC index corresponds to the row emphasised by the rectangle with the white frame (fourth row in the 'World Markets' section), which covers the time interval from August 2014 to Aug. 4, 2015. Our first diagnostic of a "bubbly" SSEC Index occurred Dec. 2014 and persisted until Jan. 2015, when a change of regime indeed occurred (see Fig. 7.5). Afterwards, the signal re-appears even stronger on early April 2015 and persisted until the eventual burst of the bubble on June 16, 2015. Note the consistency of the bubble signal during this later period.

The FCO Report of June 1^{st}, 2015 warned about the overheated Chinese market (Fig. 7.4): "*... The observation is no surprise: Chinese stocks are in bubble territory. Clearly, the rise in the last year is not sustainable*". The table on the right side of Fig. 7.4 shows that the 'DS LPPLS Confidence' and 'DS LPPLS Trust' indicators were the highest among all other signals for world equity indices.

7.3.3 Summary of the FCO early warnings

Fig. 7.5 shows the SSEC index from May 2014 to July 2015 and illustrates how a LPPLS analysis works. The coloured curves covering the rainbow-coloured spectrum represent the LPPLS fits of the SSEC index from Dec. 2014 to May 25, 2015 (indicated by

Figure 7.3: 'DS LPPLS Confidence' indicator in red (right scale) together with the SSEC index in blue (left scale) from July 2013 to July 2015, as could be retrieved each day up to the time of the observation, providing a real-time diagnostic of the on-going bubble. After correctly diagnosing a change of regime to occur on early 2015, the DS-Confidence Indicator signalled an even larger upcoming correction by late April and early June 2015. Retrieved from the FCO website on June 16, 2015.



Equities Country Indices	Yearly Return	DS LPPL Trust	DS LPPL Confidence	Country
Positive Bubbles				
SZSE 100 Index	125%	42%	40%	China
Shanghai SE Composite Index	126%	34%	25%	China
FTSE China 25 Index	35%	17%	16%	China
Budapest SE Index	17%	14%	16%	Hungary
TOPIX Stock Price Index	39%	11%	6%	Japan
Shanghai SE 50 Index	110%	11%	5%	China
Hang Seng Index	19%	9%	18%	Hong Kong

Figure 7.4: Reproduction of the diagnostic of the SSEC bubble in the FCO Cockpit report of June 1, 2015.

the boundary with the grey domain), which illustrates how the LPPLS model could extract useful information on May 25, 2015 on the possible subsequent developments. Other earlier and later dates between May 2015 and June 2015 give similar results. The different coloured curves correspond to different weights given to the deviations from the pure LPPLS model, allowing us to assess the robustness of the calibration of the LPPLS model. They exemplify a quite remarkable match with the expected LPPLS structure. The smoothed double-humped curve filled with red represents the distribution of critical times t_c at which the bubble was anticipated on May 25, 2015 to burst. It turned out that the first peak of the estimated probability for the critical time coincides with the time when the SSEC peaks and then started to drop. The second high peak coincided with the time of fastest drop of the index. The bottom panel shows the returns of the SSEC index together with a robust estimation of the crash hazard rate (the probability that the crash will occur per unit time, conditional on not yet having occurred). The colors from green to red encode the strength of the crash hazard rate, and indicate an increasing danger for a crash to occur as the bubble was developing. The inset in the top panel shows the return time series as embodied by the deterministic component of the LPPLS fits and can be compared with the bottom panel.

7.4 Complementary post-mortem analysis

This section presents additional results that provide complementary insights on the nature and value of the LPPLS methodology to diagnose the SSEC bubble and its termination.

Fig. 7.6 summarises in its two panels the two ways of diagnosing bubbles using the LPPLS model. In the top panel, we show the distribution of predicted t_c's as well as estimated beginning of the bubble. This is obtained by scanning over several t_1's (grey shaded region) for different t_2's (black dashed vertical line). This allows us to determine that the SSEC bubble began as early as April 2014 (blue probability density function (*pdf*) in Fig 7.6) and that this unstable trajectory would be unlikely to persist past June 2015 (red *pdf* in Fig. 7.6).

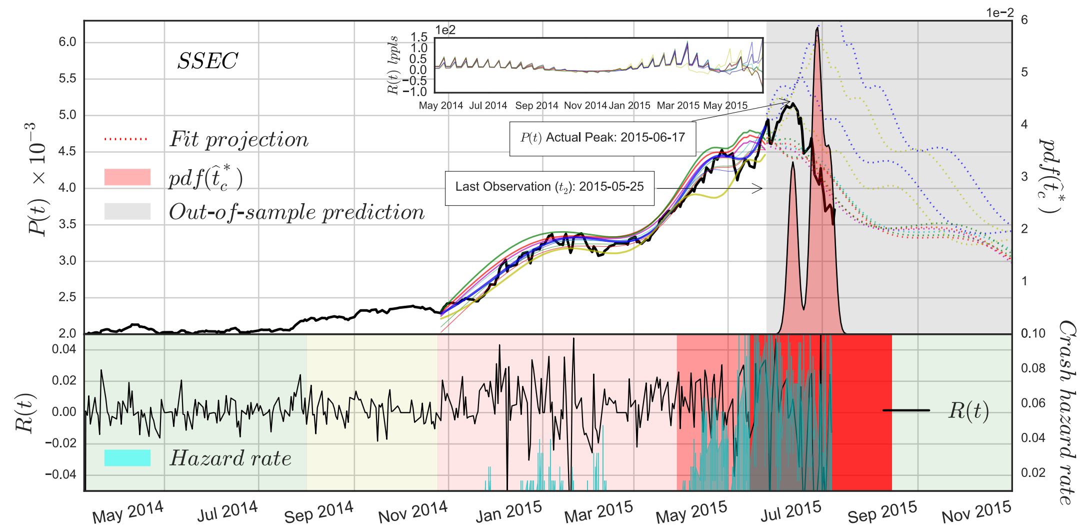


Figure 7.5: Summary of various indicators obtained by the LPPLS analysis performed at time $t_2 =$ May 25, 2015 on the SSEC index that crashed in June 2015. See text for details. Note that the distribution of t_c's does not match exactly that shown in figure 7.6, which is obtained by averaging over a number of different t_2, thus showing some dependence of the results on the "present" time used for the analysis. This should be expected since different times are associated with different information. Even if details are different, the most striking point is however the overall robustness of the diagnostic and the quite remarkable convergence to the correct period for the predicted crash.

In the lower panel of Fig. 7.6, the times at which the SSEC index is deemed to develop dangerously in a bubble are indicated by the DS LPPLS Confidence indicator shown as the black histogram together with the projections (coloured doted lines) of the LPPLS fits for different t_1's and for the t_2 indicated by the vertical dashed line. The "hazard zone" is a function of the DS LPPLS Confidence indicator: the darker the red, the higher is the probability for a significant change of regime to occur.

Fig. 7.7 provides additional information on the nature of the optimisation process via the two-dimensional cross-sections of its cost function in the space of the three nonlinear parameters m, ω and t_c of the LPPLS formula (1.10). This analysis allows us to provide an estimation of the typical confidence interval for t_c.

Fig. 7.8 suggests first a way to combine the 'DS LPPLS Confidence' and 'DS

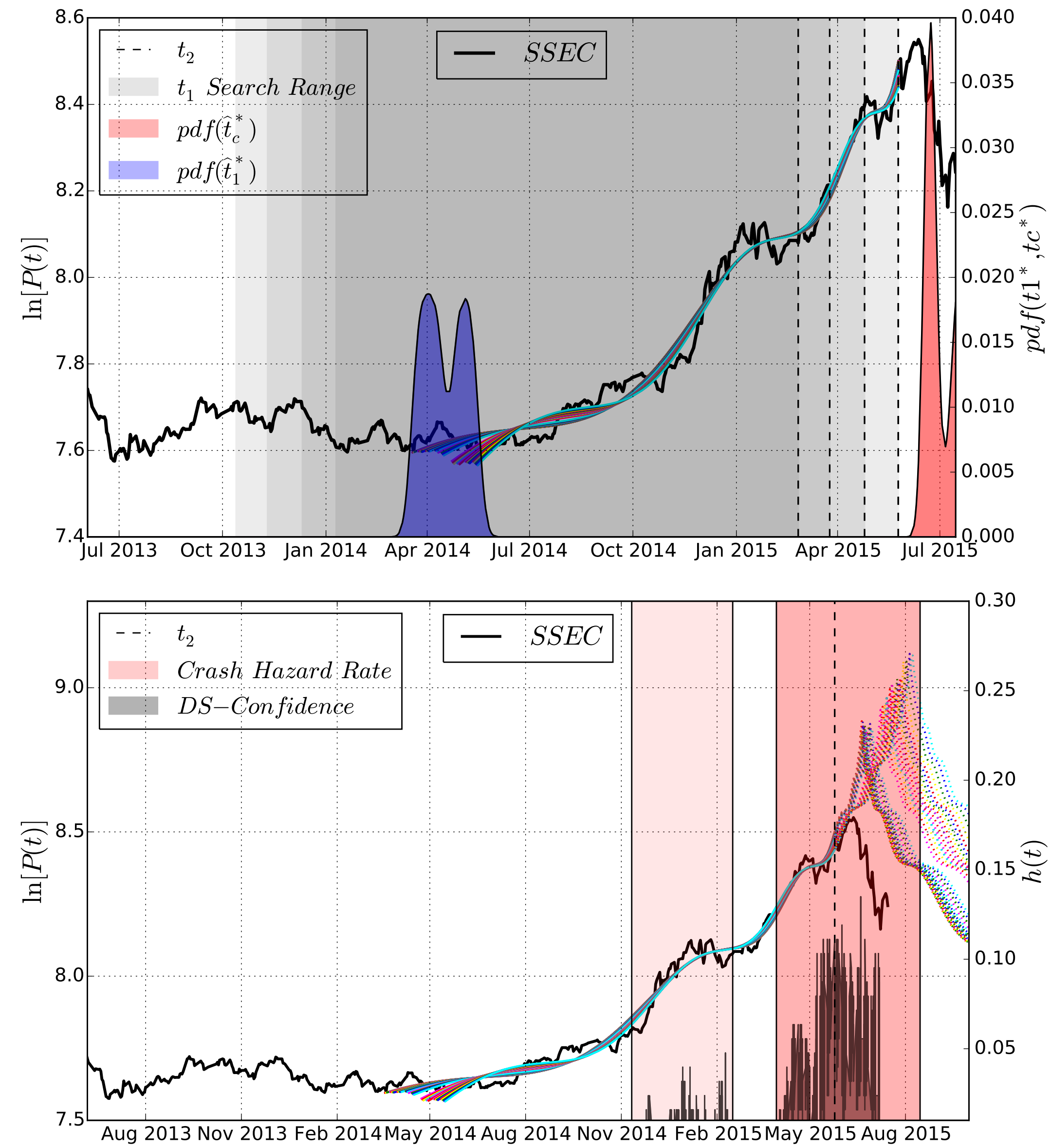


Figure 7.6: **top panel**: For each vertical dashed line t_2, we fit the LPPLS model for all windows of duration from 400 days to a minimum of 125 trading days prior to t_2, scanned daily. This search range for the beginning of the bubble t_1 is shown as the grey shaded regions. The optimal values for the bubble starting date t_1 are represented by the blue probability density distribution $pdf(\widehat{t}_1^*)$. It is pleasant to observe that this pdf is concentrated in a time interval where, visually, the acceleration of the SSEC index is starting. The forecasted critical times depicted by the red $pdf(\widehat{t}_c^*)$ present a strong probability measure at the time of the crash. The coloured lines illustrates the remarkable LPPLS characteristics of the SSEC index. **bottom panel**: The DS LPPLS Confidence (black *pdf*) successfully captures moments when the Index was bubbly. Values of the crash hazard rate $h(t)$ inform us about both the confidence and consistency of the critical time parameter t_c estimated over several windows (depicted here by coloured dashed lines). The darker the red shaded region, the higher is the probability of a change of regime to occur.

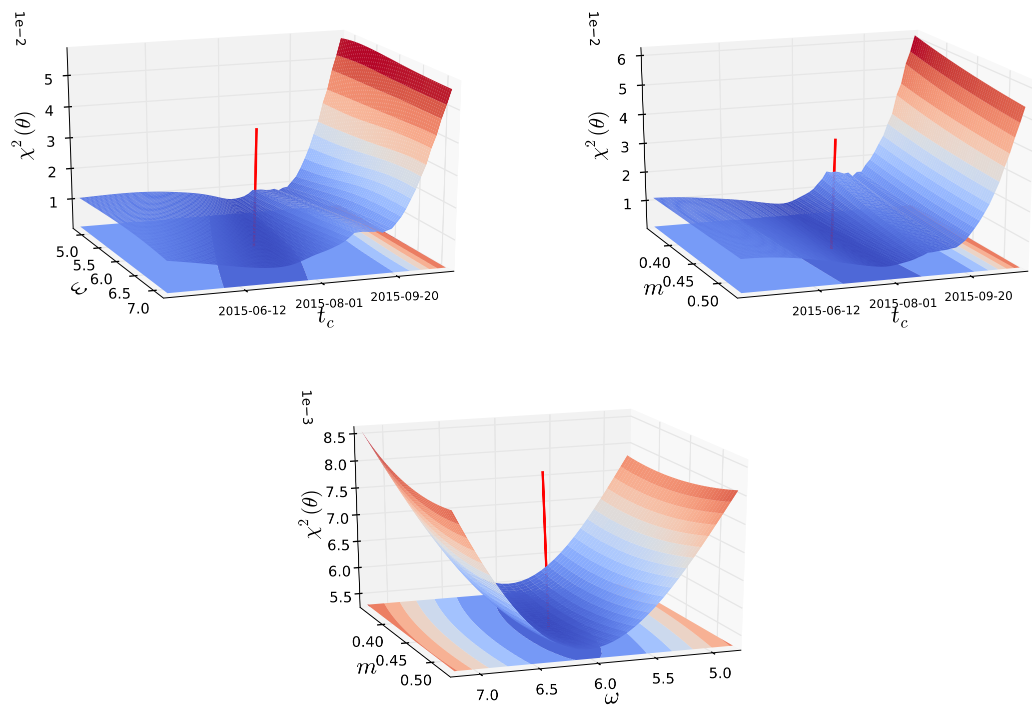


Figure 7.7: Three cross-sections of the cost function landscape as a function of pairs formed from the three nonlinear parameters m, ω and t_c of the LPPLS formula (1.6) obtained in the time window $w = [2014.06.20, 2015.05.12]$. The best-fit parameters are depicted by the red vertical lines. The cost function is convex in the space m, ω (bottom panel), showing a rather precise determination of the log-periodic angular frequency ω associated with the clear characteristic spells of log-price accelerations and corrections before the crash. In the two top panels, one can observe the existence of a secondary minimum for t_c, associated with the possibility of a scenario in which the crash of June-July 2015 would have been less severe and only an episode before a strong rebound followed by a latter even more severe crash. The shape of the cost function allows us to determine the following 90% confidence interval for the critical time t_c, namely [2015.06.01, 2015.08.01], as could be determined on 2015.05.12.

LPPLS Trust' indicators defined in section 7.2.1 to obtain perhaps a better diagnostic of the bubble development. This is performed by taking the product of DS LPPLS Confidence' indicator shown in the top panel and of the 'DS LPPLS Trust' indicator shown in the middle panel to form the confidence-trust product indicator shown in the bottom panel.

The bottom panel of Fig. 7.8 suggests a provoking interpretation. Here, the times when the People's Bank of China (PBoC) cut its benchmark lending rates are shown with the blue and green arrows. Notice how they coincide systematically with the end of plateaus or of the corrections that have punctuated the development of the bubble. Just looking at this data, one cannot escape the impression that the bubble may have been engineered, or at least catalysed, by the PBoC and its monetary policy. Given that the upbeat growth of the Chinese economy in the previous decades based on investment in real-estate and infrastructure has significantly abated, and given that the Chinese government is interested in diversifying the source of funding of both public, semi-public and private organisations, it is natural that the stock market would become both the witness as well as the engine of a transition towards a more capitalistic economy. Here, there is an analogy with the Western stock markets, which since 2008 have been supported and pushed up by the easy money policy of central banks. The reasoning is that a booming stock market is good for investors, for firms and for consumers, via the wealth spillover and the confidence effect. Pursuing this line of reasoning would however seem to be in contradiction with the theory underlying our claim in this article that the bubble was diagnosable and was actually diagnosed in advance. It would seem to tell that the bubble was due to the exogenous influence of the PBoC and not to the endogenous self-organisation of the markets resulting from positive feedbacks between herding investors. Indeed, this exogenous story would be valid if we would consider central banks as impervious to the travails of mere mortals agitating themselves on the stock markets. In fact, as argued already with the Russian crisis and crash in August 1998 [**?**], central banks tend to be actually 'slaved' to the stock markets [**?**, **?**], so that their actions have to be considered as an endogenous component of the whole.

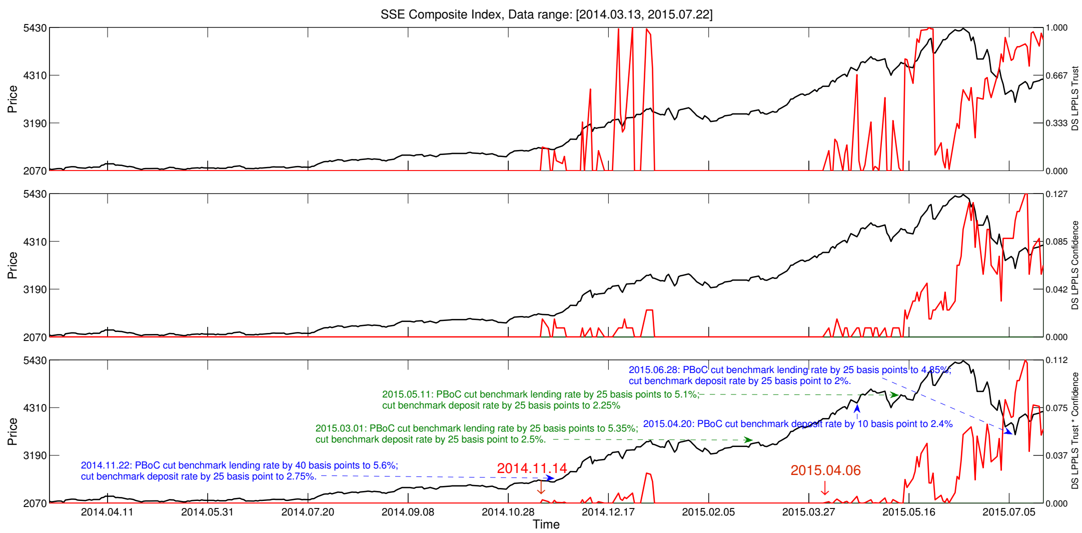


Figure 7.8: **top panel**: 'DS LPPLS Confidence' indicator (red) together with the SSEC index. **middle panel**: 'DS LPPLS Trust' indicator (red) together with the SSEC index. **bottom panel**: Product of the two above indicators and times when the People's Bank of China (PBoC) cut its benchmark lending rates (blue and green arrows).

7.5 Conclusion

Since the crash that started on June 17, 2015, the Chinese government has taken many unusual steps to stop it. For instance, on July 8, 2015, the central bank pledged to help maintain market stability, and Chinese stock regulators banned company officers and major shareholders from selling shares in listed companies. The following reasons can be invoked to understand the growing importance of stock markets for China, its economy, for the investors and people and for its government. First, the performance of the stock market will become increasingly correlated with economic growth, via the wealth effect mechanism and the use of stock markets by firms to fund their growth. The progressive trend towards privatisation of state-owned enterprises, the reduction of the reliance on bank lending, and the improvement of the pricing of capital relies on well-functioning and attractive stock markets. Moreover, in the medium term goal of the internationalisation of the renminbi, there is a need for the stock markets to rise in order to attract foreign investors. There are more political reasons also fostering the needs for a rising stock market, namely to reward China's elites for their support to the government's agenda by replacing previous sources of revenues associated with well-connected insider profits and booming property markets. The stock markets embody the "China dream" of successful entrepreneurship and the Chinese way towards capitalism, which can only succeed in the eyes of the Chinese public if the stock markets continue to grow in importance and in valuation.

For all these reasons, it is likely that the three large bubbles and crashes that have punctuated the ascent of the Chinese stock market in the last decade are just the vanguards of many more and larger bubbles to come. Hence, the FCO with its vision to provide advanced warnings is likely to become even more relevant and important in the future. We look forward to continuing improving the methodology and reporting our progress on the publicly accessible FCO website at ETH Zurich, as well as in specific publications.

Chapter 8

Conclusion

In this thesis, different technologies for diagnosing and forecasting speculative financial bubbles were pursued. Among many models proposed by the literature, we focused our efforts on refining the JLS-LPPLS model [56], which stands out from its peers due to its intrinsic parameter t_c, which represents in calendar days the most probable date of a future change of regime occurring on a financial instrument. Moreover, the model is rooted in an elegant complex-system framework in which financial markets are viewed as an out-of-equilibrium dynamical system and subject to both herding and positive feedback; major drivers of speculative bubbles.

Based on this framework, we were able to provide insightful contributions for the financial bubbles and crashes literature which are summarised below. Chapter 2 proposes a novel methodology for quantifying the uncertainty permeating the birth and burst of speculative bubbles (t_1 := birth of the bubble and t_c := burst). Our results show that speculative manias persist even in the presence of fully Rational Expectation agents (see [18]) due to the lack of consensus regarding t_c rather than t_1, as argued by [1]. Given different expectations with regards to t_c, a synchronisation among agents for selling the bubbly financial instrument is not observed and thus bubbles are allowed to grow. As a bonus, we find that the structural parameter of the LPPLS model, the exponent m controlling the super-exponential growth of price, are very "rigid" according the Hessian matrix analysis, which supports the LPPLS model as a reasonable candidate for describing the generating process of prices during bubbles.

Chapter 3 provides a novel technique which allows one to compare goodness of fit of a model using different sample sizes. By adding a penalty term to the cost function which is proportional to the overfit tendency of a model, one can select the optimal sub-sample in which models should be fitted. When applied to the LPPLS model, the method allows one to automatically make statistically significant inferences regarding the starting date of speculative bubbles.

On Chapter 4, a methodology for diagnosing in real-time the multi-scale aspect of price exuberance during speculative bubbles is proposed. A core result of this section consisted in showing that the LPPLS framework is able to successfully capture, ex-ante, some of the prominent bubbles across different time scales and that market-based proxies of trading activity can be used as a guideline to model and monitor bubble conditions in stock market.

While Chapters 2-4 main focus lays on diagnosing financial bubbles inceptions', Chapters 5-6 make use of Information Geometry in order to enhance the estimation accuracy of the crash date, t_c and to show how eigenvalues and eigenvectors of the Fisher Information Matrix can be explored in order to overcome the problem of parametric non-identifiability. Specifically, Chapter 5 proposes a three-step calibration approach for retrieving $\hat{t}_c$ where all remaining LPPLS' parameters are treated as nuisance. We then employ a sophisticated Likelihood estimator, namely, the Modified Profile Likelihood, which adjusts the original Profile Likelihood to take into account the sloppy nature of t_c. Results are excellent in terms of estimations accuracy, specially for small samples ($N \leq 200$). Another core results of this Section is that the methodology allows one to construct robust Likelihood Confidence Intervals around parameter estimates.

We then found that this hierarchical structure of how models should be calibrated goes beyond the LPPLS model. Moreover, in Chapter 6 we make use of a GARCH(p,q) model to show that the combination of a sophisticated Profile Likelihood estimator with the optimal focal parameter (chosen via the Eigenvalue - Eigenvector hierarchy of the Fisher Information Matrix) is extremely important to obtain precise parameter estimates, specially for small samples ($N \leq 200$). We have also found that practical non-identifiable optimisation problems can be made practical identifiable ones via this methodology. Testing the robustness of the methodology to different model classes is still an open question which is worthwhile to be pursued.

Finally, Chapter 7 provides supportive evidence for the relevance of the LPPLS-based methodology in the diagnostic of bubbles. The chapter documents how the real-time predictions of the Shanghai Composite Index (SSEC) December 2015 bubble were presented in the automated analysis of the FCO, as well as in their FCO Cockpit report of June 2015. A complementary post-mortem analysis of the nature and value of the LPPLS methodology in diagnosing the SSEC bubble and its termination is also given.

It is important to stress that our work was mainly data driven and contrary to the standard paradigm of efficient markets, which explains price movements by the arrival of new information. Instead, we have worked with the notion of speculative bubbles where the price of an asset systematically detaches from the fundamental value over an extended period of time and is severely subject to both positive feedback and herding mechanisms. Moreover, we focused solely on endogenous bubbles occurring on financials instruments and completely neglected any exogenous triggering factors even thought we are aware of the reflexivity nature of financial markets [37].

In order to enhance our current capacity to diagnose bubbles and forecast crashes, future research could make use of techniques such as Deep-Learning and Neural Networks on top of the LPPLS methodology in order to learn the intricate non-linear relationships between fundamental valuation metrics (i.e., earnings, debt to equity ratio, future earnings ...etc) and price fluctuations of a security. Finally, the results obtained in this thesis are of great practical importance to Portfolio Managers, Macroeconomists and Central Bankers alike given the unique capacity of the framework to dynamically detect bubbles and forecast major changes of regime.

Bibliography

[1] D. Abreu and K. Brunnermeier. Bubbles and crashes. *Econometrica*, 1:173–204, 2003.

[2] H Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19(6):716–723, 1974.

[3] Y. Amihud. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5:31–56, 2002.

[4] K. Anderson, C. Brooks, and A. Katsaris. Speculative bubbles in the s&p 500: Was the tech bubble confined to the tech sector? *Journal of Empirical Finance*, 17(3):345–361, 2010.

[5] V. Arora and S-P. Shi. Nonlinearities and tests of asset price bubbles. *Empirical Economics*, 50(4):1421–1433, 2016.

[6] S. Athey, I. Parashkevov, V. Sarukkai, and J. Xia. Bitcoin pricing, adoption, and usage: Theory and evidence. *SSRN Working Paper No. 2826674*, 2016.

[7] D. Avramov, T. Chordia, and A. Goyal. Liquidity and autocorrelations in individual stock returns. *Journal of Finance*, 61:2365–2394, 2006.

[8] L. Bachelier. Theorie de la speculation. *Annales Scientifiques de Ecole Normale Superieure*, 17(21), 1900.

[9] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality. *Phys. Rev. A*, 3:364–374, Jul 1988.

[10] Nicholas Barberis, Andrei Shleifer, and Robert Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49:307–343, 1998.

[11] Ole E Barndorff-Nielsen. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70(2):343–365, 1983.

[12] Ole E Barndorff-Nielsen. Adjusted Versions of Profile Likelihood and Directed Likelihood, and Extended Likelihood. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1):125–140, 1994.

[13] Ole E Barndorff-Nielsen and D R Cox. *Inference and Asymptotics* . Inference and Asymptotics, London, 1994.

[14] Debabrata Basu. On the Elimination of Nuisance Parameters. *Journal of the American Statistical Association*, 72(358):355–366, June 1977.

[15] M J Bayarri and Morris H DeGroot. Difficulties and ambiguities in the definition of a likelihood function. *Journal of the Italian Statistical Society*, 1(1):1–15, February 1992.

[16] James O Berger, Brunero Liseo, and Robert L Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical science*, 14(1):1–28, February 1999.

[17] C Alan Bester and Christian Hansen. A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects. *Journal of Business and Economic Statistics*, 27(2):131–148, April 2009.

[18] Olivier J. Blanchard and Mark W. Watson. Bubbles, rational expectations and financial markets. Working Paper 945, National Bureau of Economic Research, June 1982.

[19] D. Bree, D. Challet, and P. Peirano. Prediction accuracy and sloppiness of log-periodic functions. *Quantitative Finance*, 3:275–280, 2013.

[20] David S Brée, Damien Challet, and Pier Paolo Peirano. Prediction accuracy and sloppiness of log-periodic functions. *Quantitative Finance*, 13(2):275–280, 2013.

[21] J. Breitung and R Kruse. When bubbles burst: econometric tests based on structural breaks. *Statistical Papers*, 54:911–930, 2013.

[22] C. Brooks and A. Katsaris. A three-regime model of speculative behavior: Modelling the evolution of the s&p 500 composite index. *Economic Journal*, 115(505):767–797, 2005.

[23] Kevin S. Brown and James P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904, August 12 2003.

[24] Markus K. Brunnermeier and Stefan Nagel. Hedge funds and the technology bubble. *Journal of Finance*, 59:2013–2040, 2004.

[25] Markus K Brunnermeier and Martin Oehmke. Bubbles, Financial Crises, and Systemic Risk. In George M Constantinides, Milton Harris, and Rene M Stulz, editors, *Handbook of the Economics of Finance*, pages 1221–1288. Elsevier, 2012.

[26] J. F. Callen and X. Fang. Short interest and stock price crash risk. *Journal of Banking and Finance*, 60:181–194, 2015.

[27] J. Y. Campbell, S. J. Grossman, and J. Wang. Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, 108:905–939, 1993.

[28] Subrahmanyam A. Chordia, T. and Anshuman. Trading activity and expected stock returns. *Journal of Financial Economics*, 59:3–32, 2001.

[29] D R Cox and Nancy Reid. Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1):1–39, 1987.

[30] D Cvijovicacute and J Klinowski. Taboo Search: An Approach to the Multiple Minima Problem. *Science*, 267(5198):664–666, February 1995.

[31] G. Demos and D. Sornette. Birth or burst of financial bubbles: which one is easier to diagnose? *Quantitative Finance*, 5:657–675, 2017.

[32] Thomas J DiCiccio. Introduction to Barndorff-Nielsen (1983) On a Formula for the Distribution of the Maximum Likelihood Estimator. In Samuel Kotz and Norman L Johnson, editors, *Breakthroughs in Statistics*, pages 395–431. Springer New York, New York, NY, 1997.

[33] Thomas J DiCiccio, Michael A Martin, Steven E Stern, and G Alastair Young. Information Bias and Adjusted Profile Likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):189–203, 1996.

[34] Eugene Fama, Thomas Shiller, and Lars Peter Hansen. Understanding asset prices. *The Royal Swedish Academy of Sciences*, 2013.

[35] J.D. Farmer. The stock market crash really did cause the great recession. *Oxford Bulletin of Economics and Statistics*, 77(5):617–633, 2015.

[36] Vladimir Filimonov, Guilherme Demos, and Didier Sornette. Modified profile likelihood inference and interval forecast of the burst of financial bubbles. *Quantitative Finance*, 7(8):1167–1186, 2017.

[37] Vladimir Filimonov and Didier Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Phys. Rev. E*, 85(5):32–47, 2012.

[38] Vladimir Filimonov and Didier Sornette. A Stable and Robust Calibration Scheme of the Log-Periodic Power Law Model. *Physica A: Statistical Mechanics and its Applications*, 392(17):3698–3707, 2013.

[39] Ronald A Fisher. *Statistical Methods and Scientific Inference*. Oliver & Boyd, 1956.

[40] D A S Fraser and Nancy Reid. Adjustments to profile likelihood. *Biometrika*, 76(3):477–488, 1989.

[41] D A S Fraser, Nancy Reid, and J Wu. A Simple General Formula for Tail Probabilities for Frequentist and Bayesian Inference. *Biometrika*, 86(2):249–264, 1999.

[42] A. R. Gallant, P. E. Rossi, and G. Tauchen. Stock prices and volume. *Review of Financial Studies*, V:199–242, 1992.

[43] David Garcia and Frank Schweitzer. Social signals and algorithmic trading of bitcoin. *Royal Society Open Science*, 2(9):150288, 2015.

[44] Petr Geraskin and Dean Fantazzini. Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask. *The European Journal of Finance*, 19(5):366–391, 2011.

[45] Josiah Willard Gibbs. *Elementary Principles in Statistical Mechanics*. Dover Books on Physics. Dover Publications, 1902.

[46] S Gluzman and Didier Sornette. Log-periodic route to fractal functions. *Physical Review E*, 65(3):036142, March 2002.

[47] H.-C. Graf v. Bothmer and C. Meister. Predicting critical crashes? a new restriction for the free variables. *Physica A*, 320:539–547, 2003.

[48] Benjamin Graham and David Dodd. *Security Analysis*. McGraw-Hill, sixth-edition edition, 1988.

[49] R. Gurkaynak. Econometric tests of asset price bubbles: taking stock. *Journal of Economic Surveys*, 22:166–186, 2008.

[50] Michael Harrison and D. M. Kreps. Speculative investor behavior in a stock-market with heterogeneous expectations. *Quarterly Journal of Economics*, 92:323–336, 1978.

[51] Leonhard Held and Daniel Sabanés Bové. *Applied Statistical Inference: Likelihood and Bayes*. Springer, Berlin, 2013.

[52] Y. Huang, A. Johansen, M. W. Lee, H. Saleur, and D. Sornette. Artifactual log-periodicity in finite-size data: Relevance for earthquake aftershocks. *J. Geophys. Res.*, 105:25451–25471, 2000.

[53] John D Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9(3):90–95, 2007.

[54] A. Hüsler, D. Sornette, and C. H. Hommes. Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price. *Journal Economic Behavior & Organization*, 92:304–316, 2012.

[55] Zhi-Qiang Jiang, Wei-Xing Zhou, Didier Sornette, Ryan Woodard, Ken Bastiaensen, and Peter Cauwels. Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 Chinese stock market bubbles. *Journal of Economic Behavior & Organization*, 74(3):149–162, June 2010.

[56] A. Johansen, D. Sornette, and O. Ledoit. Predicting financial crashes using discrete scale invariance. *Journal of Risk*, 1(4):5–32, 1999.

[57] Anders Johansen, Olivier Ledoit, and Didier Sornette. Crashes as Critical Points. *International Journal of Theoretical and Applied Finance*, 3(2):219–255, 2000.

[58] Anders Johansen and Didier Sornette. Critical Crashes. *Risk*, 12(1):91–94, 1999.

[59] Anders Johansen and Didier Sornette. Financial "Anti-Bubbles": Log-Periodicity in Gold and Nikkei collapses. *International Journal of Modern Physics C*, 10(04):563–575, June 1999.

[60] Anders Johansen and Didier Sornette. Shocks, Crashes and Bubbles in Financial Markets. *Brussels Economic Review*, 53(2):201–253, 2010.

[61] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.

[62] C. M. Jones, G. Kaul, and M. L. Lipson. Transactions, volume, and volatility. *Review of Financial Studies*, 7(4):631–651, 1994.

[63] T. Kaizoji and D. Sornette. Market bubbles and crashes. *Encyclopedia of Quantitative Finance*, 3:144–156, 2010.

[64] Taisei Kaizoji and Didier Sornette. Market Bubbles and Crashes. In *Encyclopedia of Quantitative Finance*. Wiley, 2010.

[65] John D Kalbfleisch and D A Sprott. Application of Likelihood Methods to Models Involving Large Numbers of Parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(2):175–208, 1970.

[66] J. M. Karpoff. The relation between price changes and trading volume: a survey. *Journal of Financial and Quantitative Analysis*, XXII:109–126, 1987.

[67] John T. Kent. Robust properties of likelihood ratio test. *Biometrika*, pages 19–27, 1982.

[68] C. P. Kindleberger. *Manias, Panics and Crashes: A History of Financial Crises*. Macmillan, London, 3 edition, 1978.

[69] Frank H. Knight. *Risk, Uncertainty, and Profit*. Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA, 1921.

[70] L. Lin, R. E. Ren, and D. Sornette. The volatility-confined lppl model: A consistent model of 'explosive' financial bubbles with mean-reverting residuals. *International Review of Financial Analysis*, 33:210–222, 2014.

[71] Li Lin, Ruoen Ren, and Didier Sornette. A Consistent Model of 'Explosive' Financial Bubbles with Mean-Reversing Residuals. *International Review of Financial Analysis*, 33:210–225, 2014.

[72] Andrew W. Lo. The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30.5:15–29, 2004.

[73] S. Loscalzo, L. Yu, and C. Ding. Consensus group stable feature selection. *In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.*, pages 567–576, 2009.

[74] B. Machta, R. Chachra, M. Transtrum, and J. Sethna. Parameter space compression underlies emergent theories and predictive models. *Science*, 342:604–607, 2013.

[75] Wes McKinney. *Python for Data Analysis*. O'Reilly Media, 2012.

[76] Minsky and Hyman. The modeling of financial instability: An introduction. *Proceedings of the fifth annual Pittsburgh conference*, 1974.

[77] P.K. Narayan and S.S. Sharma. Asset price bubbles and economic welfare. *International Review of Financial Analysis*, 44:139–148, 2016.

[78] Andrew Y. Ng. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *Proceedings of the Twenty-first International Conference on Machine Learning*, pages 78–. ACM, 2004.

[79] O. Nneji. Liquidity shocks and stock bubbles. *Journal of International Financial Markets, Institutions & Money*, 35:132–146, 2015.

[80] Luigi Pace and Alessandra Salvan. Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, 136(10):3554–3564, October 2006.

[81] L. Pastor and R. F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111:642–685, 2003.

[82] Yudi Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, 2001.

[83] Fernando Pérez and Brian E Granger. IPython: A System for Interactive Scientific Computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

[84] P.C.B. Phillips and J. Yu. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491, 2011.

[85] D. E. Rapach, M. C. Ringgenberg, and G. Zhou. Short interest and aggregate stock returns. *Journal of Financial Economics*, 121:46–65, 2016.

[86] H. Saleur and D. Sornette. Complex exponents and log-periodic corrections in frustrated systems. *J.Phys.I France*, 6(3):327–355, 1996.

[87] Paul Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial management review*, 6:41–49, 1965.

[88] Thomas A Severini. An Approximation to the Modified Profile Likelihood Function. *Biometrika*, 85(2):403–411, 1998.

[89] Thomas A Severini. Likelihood functions for inference in the presence of a nuisance parameter. *Biometrika*, 85(3):507–522, September 1998.

[90] Thomas A Severini. An empirical adjustment to the likelihood ratio statistic. *Biometrika*, 86(2):235–247, June 1999.

[91] Thomas A Severini. *Likelihood Methods in Statistics*. Oxford Statistical Science Series. Oxford University Press, 2001.

[92] Thomas A Severini. Integrated likelihood functions for non-Bayesian inference. *Biometrika*, 94(3):529–542, August 2007.

[93] S.P. Shi. Specification sensitivities in the markov-switching unit root test for bubbles. *Empirical Economics*, 45(2):697–713, 2013.

[94] R. Shiller. *Irrational exuberance*. Princeton University Press, Princeton, NJ, 2000.

[95] R. J. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71:421–436, 1981.

[96] H. S. Shin. Risk and liquidity in a system context. *Jounal of Financial Intermediation*, 17(3):315–329, 2016.

[97] Andrei Shleifer. *Inefficient markets, an introduction to behavioral finance*. Clarendon Lectures in Economics, Oxford University Press, 2000.

[98] Ib M Skovgaard. An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, 2(2):145–165, June 1996.

[99] Ib M Skovgaard. Likelihood Asymptotics. *Scandinavian Journal of Statistics*, 28(1):3–32, March 2001.

[100] Richard L Smith and J C Naylor. A Comparison of Maximum Likelihood and Bayesian Estimators for the Three- Parameter Weibull Distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):358, 1987.

[101] Robert Solow. Building a science of economics for the real world. *House Committee on Science and Technology; Subcommittee on Investigations and Oversight (July 20)*, 2010.

[102] D. Sornette. Discrete scale invariance and complex dimensions. *Physics Reports*, 297(5):239–270, 1998. (extended version at http://xxx.lanl.gov/abs/cond-mat/9707012).

[103] D. Sornette. Physics and financial economics (1776-2014): Puzzles, ising and agent-based models. *Report of Progress in Physics*, 77:062001, 2014.

[104] D. Sornette and P. Cauwels. 1980-2008: The illusion of the perpetual money machine and what it bodes for the future. *Risks*, 2:103–131, 2014.

[105] D. Sornette and P. Cauwels. Financial bubbles: mechanisms, diagnostics and state of the world. *Review of Behavorial Economics*, 2(3):279–305, 2015.

[106] D. Sornette and A. Johansen. Significance of log-periodic precursors to financial crashes. *Quantitative Finance*, 1(4):452–471, 2001.

[107] D. Sornette, A. Johansen, and J. P. Bouchaud. Stock market crashes, precursors and replicas. *Journal of Physics*, 6:167–175, France, 09 Aout 1995.

[108] Didier Sornette. Discrete-scale invariance and complex dimensions. *Physics Reports*, 297(5):239–270, 1998.

[109] Didier Sornette. Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes and human birth. *Proceedings of the National Academy of Sciences USA*, 99(Supp. 1):2522–2529, 2002.

[110] Didier Sornette. *Why stock markets crash: Critical events in complex financial systems*. Princeton University Press, New Jersey, 2003.

[111] Didier Sornette. Dragon-Kings, Black Swans and the Prediction of Crises. *International Journal of Terraspace Science and Engineering*, 2(1), December 2009.

[112] Didier Sornette, Guilherme Demos, Qun Zhang, Peter Cauwels, Vladimir Filimonov, and Qunzhi Zhang. Real-Time Prediction and Post-Mortem Analysis of the Shanghai 2015 Stock Market Bubble and Crash. *Journal of Investment Strategies*, 4(4):77–95, 2015.

[113] Didier Sornette and Anders Johansen. Large financial crashes. *Physica A: Statistical Mechanics and its Applications*, 245(3-4):411–422, 1997.

[114] Didier Sornette and Anders Johansen. Significance of log-periodic precursors to financial crashes. *Quantitative Finance*, 1(4):452–471, 2001.

[115] Didier Sornette and Ryan Woodard. Financial bubbles, real estate bubbles, derivative bubbles, and the financial and economic crisis (2009). *Proceedings of APFA7 (Applications of Physics in Financial Analysis), "Econophysics Approaches to Large-Scale Business Data and Financial Crisis"*, Misako Takayasu, Tsutomu Watanabe and Hideki Takayasu, eds., Springer:101–148, 2010.

[116] Didier Sornette, Ryan Woodard, Wanfeng Yan, and Wei-Xing Zhou. Clarifications to questions and criticisms on the Johansen–Ledoit–Sornette financial bubble model. *Physica A: Statistical Mechanics and its Applications*, 392(19):4417–4428, October 2013.

[117] Didier Sornette and Wei-Xing Zhou. Predictability of Large Future Changes in major financial indices. *International Journal of Forecasting*, 22:153–168, 2006.

[118] Cecilia Sua, Guillermo Cecchi, Guillermo Marshall, and Diego Ferna. When the optimal is not the best: Parameter estimation in complex biological models. *PLoS One*, 5(10):3–19, 2010.

[119] El-Ghazali Talbi. *Metaheuristics: from design to implementation*. Wiley, Hoboken, NJ, 2009.

[120] Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.

[121] R. Tibshirani. Regression shrinkage and selection via the lasso. *Royal Statistical Society*, 58(1):267–288, 1996.

[122] M. Transtrum and J. Sethna. Improvements to the levenberg-marquardt algorithm for nonlinear least-squares minimization. *arXiv preprint*, 2012. http://arxiv.org/abs/1201.5885.

[123] Mark K. Transtrum, Benjamin B. Machta, and James P. Sethna. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E*, 83:673–701, Mar 2011.

[124] R. Tsay. *Analysis of financial time-series*. Wiley series in probability and statistics. Wiley, 3 edition, 2010.

[125] W Xiong. Bubbles, Crisis and Heterogeneous Beliefs. In Jean-Pierre Fouque and Joseph A Langsam, editors, *Handbook on Systemic Risk*, pages 663–713. Cambridge University Press, Cambridge, 2013.

[126] W. Yan, R. Rebib, R. Woodard, and D. Sornette. Detection of crashes and rebounds in major equity markets. *International Journal Portfolio Analysis and Management*, 1:59–79, 2010.

[127] Qun Zhang, Qunzhi Zhang, and Didier Sornette. Early warning signals of financial crises with multi-scale quantile regressions of Log-Periodic Power Law Singularities. *http://ssrn.com/abstract=2674128*, 2015.

[128] Wei-Xing Zhou and Didier Sornette. of the 2000-2002 anti-bubble in the US S&P 500 index: Explanation of the hierarchy of 5 crashes and Prediction. *Physica A: Statistical Mechanics and its Applications*, 330:584–604, 2003.

[129] Wei-Xing Zhou and Didier Sornette. Non-Parametric Analyses of Log-Periodic Precursors to Financial Crashes. *Int. J. Mod. Phys. C*, 14(8):1107–1126, 2003.

[130] Wei-Xing Zhou and Didier Sornette. Testing the stability of the 2000-2003 us stock market "antibubble". *Physica A: Statistical Mechanics and its Applications*, 348:428–452, 2005.