Aalto University
School of Business

THE LOG-PERIODIC POWER LAW

Evidence from the Finnish Stock Market

Master's Thesis
Jouni Koistinen
Aalto University School of Business
Finance
Spring 2020

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Jouni Koistinen

Title of thesis The Log-Periodic Power Law - Evidence from the Finnish Stock Market

Degree Master of Science (Economics and Business Administration)

Degree programme Master's Programme in Finance

Thesis advisor(s) Professor Vesa Puttonen

Year of approval 2020 **Number of pages** 74 **Language** English

Abstract

The log-periodic power law (LPPL) is a relatively new tool to model stock market crashes from the area of econophysics. The first version of the model was introduced by physicists Didier Sornette, Anders Johansen and Jean-Philippe Bouchaud in the late '90s. The model is based on behavioural models of herding and imitation among traders, the theory of critical phenomena in complex systems from physics, the mean-field theory and the Ising model from statistical mechanics.

The LPPL model can be used as a tool to predict the time of the forthcoming price crashes on various traded assets and market indices. This is accomplished by calibrating the LPPL model to the price data. If the log-periodic price oscillations are detected as the model predicts, we get estimates of future critical time (crash) and the price of the asset at that time.

In the thesis, we fit the LPPL model to the five most significant stock market crashes of OMX Helsinki -index and the ten most significant stock's price drawdowns, to find out, if the log-periodic price oscillations are found before the market downturns. We also study, how accurate the model's estimates are for the critical time and the price of the asset at that time.

We found a weak signal on OMXH calibrations only in the year 2000 crash. Further analysis shows that the OMXH index log-periodic oscillations arise from the oscillations on the Nokia stock price. Our results with the individual stocks are more robust. We have a very good fit with five out of ten stocks, and two of them meets all the filtering rules according to the model.

We conclude that the LPPL oscillations can be found from the Helsinki Stock Exchange on five out of ten tested stock events. The model's estimates for the crash time are very accurate on average. The average difference is only 1.4 days, although the number of estimates is low. The price estimates vary greatly compared to the actual.

Keywords econophysics, LPPL, stock market crash, Helsinki Stock Exchange

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Jouni Koistinen

Työn nimi Log-Periodic Power Law -malli Helsingin arvopaperipörssissä

Tutkinto Kauppatieteen maisteri

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Professori Vesa Puttonen

Hyväksymisvuosi 2020 **Sivumäärä** 74 **Kieli** Englanti

Tiivistelmä

Tutkin maisterin opinnäytetyössäni suhteellisen uutta mallia ekonofysiikan alueelta pörssiromahdusten mallintamiseen. Malli on nimeltään 'log-periodic power law' (LPPL) ja sen ensimmäisen version esittelivät 90-luvun lopulla fyysikot Didier Sornette, Anders Johansen ja Jean-Philippe Bouchaud. Malli pohjautuu sijoittajien käyttäytymistä mallintaviin jäljittely- ja laumakäyttäytymismalleihin, monimutkaisten järjestelmien kriittisen ilmiön malleihin fysiikassa sekä Ising-malliin tilastollisesta mekaniikasta.

Mallia voidaan käyttää työkaluna tulevien kurssiromahdusten ennakointiin eri markkinoilla. LPPL-malli sovitetaan hintadatan päälle. Jos malli havaitsee logaritmi-jaksollisia värähtelyjä hinnassa, saamme mallinnuksen tuloksena arvion tulevan kurssiromahduksen ajankohdasta ja kohteen hinnasta kriittisellä hetkellä.

Opinnäytetyössä sovitan LPPL-mallin viiteen merkittävimpään pörssiromahdukseen OMX Helsinki -indeksissä ja kymmeneen suurimpaan yksittäisen osakkeen kurssiromahdukseen Helsingin pörssissä. Tutkin myös, kuinka tarkkoja ovat mallin antamat ennusteet kurssiromahduksen ajankohdasta ja hintatasosta.

Malli antoi vain heikon tuloksen OMXH-indeksin tapahtumille. Heikko signaali löytyi vuoden 2000 kurssiromahdukselle ja tarkempi tutkimus osoitti kurssiromahdusta ennakoivan logaritmi-jaksollisen värähtelyn syntyvän vastaavista värähtelyistä Nokian osakekurssissa. Tulokset yksittäisille osakkeille olivat vahvempia. LPPL-malli saatiin sovitettua hyvin kurssidataan viidellä kymmenestä osakkeesta ja kaksi osakkeista täytti kaikki suodatussäännöt.

Johtopäätöksenä, LPPL-mallin mukaisia logaritmi-jaksollisia värähtelyjä löytyy puolesta testatuista kymmenestä osakkeesta ennen kurssiromahdusta. Mallin antamat ennusteet romahduksen ajankohdalle ovat keskimäärin hyvin tarkkoja. Ero todelliseen oli keskimäärin vain 1,4 päivää, joskin havaintoja oli lukumääräisesti vähän. Hinta-arviot vaihtelivat rajusti.

Avainsanat ekonofysiikka, LPPL, pörssiromahdus, Helsingin pörssi

Table of Contents

List of figures

List of tables

1 Introduction

In the thesis, we are going to take a leap of faith from the classical theory of finance into the world of econophysics. Classical theories of finance, and economics in general, are based on elegant equilibrium models (like famous CAPM in finance). They are essential tools of thinking, but they are normative. They tell how things ought to be; they do not explain the way they are.

There is one standard tool penetrating the whole universe of classical financial models, and it is also the cause of many problems[1]: Gaussian reasoning (normality). Random walks are based on normality, linear least squares algebra of market coefficients is based on normality (some distributions do not work with the central limit theorem) and even the famous Black-Scholes model derives option prices based on normality.

This is nothing new. Skewed fat-tailed distributions, shortly autocorrelated returns and extreme market events are all known by the academics and the practitioners as well; and they are being included in new more complex models.

The worst time for the classical models is a stock market crash. Normality does not apply any more; it is extremistan[2]. Falling stock prices become highly correlated, and the benefits of portfolio diversification are largely lost. Stocks may also fall many consecutive days in a magnitude unheard in the Gaussian world[3].

In the classical finance theory, the stock market crash phenomenon is largely dismissed, but in the new emerging econophysics, it is one of the key areas of research. Econophysics is a new school of thought that utilises theories and models from physics to solve complex problems in economics. Actually, many of the founding fathers of current financial theories are physicists, for example, Fisher Black was physicist and mathematician, and Black-

[1] Excellent history of Gaussian reasoning in finance is presented in the book 'The Physics of Wall Street: a brief history of predicting the unpredictable' by James Owen Weatherall.

[2] Term coined by Nassim Taleb in his book The Black Swan.

[3] If we assume normality of daily returns, DJIA drop of -22.6% in October 19, 1987, would correspond to one event in 520 million years! Sornette (2004).

Scholes option model differential equation was finally solved after they noticed that the solution is known from the thermodynamics[4].

In the thesis, we are going to study a relatively new tool to model stock market crashes, namely the log-periodic power law (LPPL). The first version of the model was introduced by physicists Sornette[5], Johansen[6] and Bouchaud[7] (Sornette et al., 1996). The LPPL model is based on behavioural models of herding and imitation among traders, the theory of critical phenomena in complex systems from physics, the mean-field theory and the Ising model from statistical mechanics.

From the perspective of classical finance theory, the LPPL model and its projections seem unwarranted. However, there is a plausible theory behind the model, and the results from various studies are very promising. As one would expect, LPPL studies are mostly published in the papers of theoretical and applied physics.

[4] Black (1989).

[5] Didier Sornette is Professor on the Chair of Entrepreneurial Risks at Swiss Federal Institute of Technology Zurich (ETH Zurich). He is also a professor of the Swiss Finance Institute, a professor associated with both the department of Physics and the Department of Earth Sciences at ETH Zurich, an Adjunct Professor of Geophysics at IGPP and ESS at UCLA. He was previously jointly a Professor of Geophysics at UCLA, Los Angeles California and a Research Director on the theory and prediction of complex systems at the French National Centre for Scientific Research.

[6] Anders Johansen, The Niels Bohr Institute at the University of Copenhagen.

[7] Jean-Philippe Bouchaud is founder and Chairman of Capital Fund Management (CFM) and professor of physics at École Polytechnique.

1.1 Background of the study

In this chapter, we aim to give a non-mathematical introduction to the subject. This description deals with the original LPPL model. The basic mathematical review and references can be found from the chapter' literature review'.

Derivation of the original LPPL model (Johansen et al., 2000) can be presented in four main steps:

1. Fundamental assumption: Market crash may be caused by a local self-reinforcing imitation between traders. The local imitation builds up gradually to a global co-operative behaviour. This imitation drives asset prices away from their fundamental values up to a critical point, when many traders place the same order (sell) at the same time, causing a crash.

The critical point is not just a phrase, but mathematically defined event in nonlinear complex dynamic systems, defined as an explosion to infinity of system's quantity.

2. Modelling imitation among traders: In the real world traders and investors get their investment ideas from various sources: advertisements, analysis, friends, investment clubs, salespeople, news media, to name a few. To model this complex interaction between market participants, the LPPL model uses a type of network: hierarchical diamond lattice. The lattice has a fractal quality (self-similarity) and some nodes (i.e. traders) are more connected than the others.

3. The LPPL model: Hazard rate of critical phenomenon (nodes converging to a sell-side and causing a crash) on a hierarchical diamond lattice can be modelled using the *mean field theory*[8].

By making an assumption that traders in the system are risk-neutral, have rational expectations and asset prices follow a martingale process (no-arbitrage condition), Johansen et al. (2000) were able to derive a price formula which is basically *'a deterministic time-*

[8] "In physics and probability theory, mean field theory (MFT) approximates the behaviour of a large number of small individual components which interact with each other. The effect of all the other individuals on any given individual is approximated by a single averaged effect, thus reducing a many-body problem to a one-body problem." (Chen et al. 2018)

varying drift decorated by a non-stationary stochastic random walk component'[9] (Lin et al., 2009).

The price dynamics of the model is explained by the imitation process and by a rational expectation that traders must increase their future price projections to be compensated for increased crash or down-side risk.

The dynamics of the system leads to the LPPL model, where the stability of the system is corrected by log-periodic price oscillations. Oscillations increase their frequency toward the critical point, and this observation indicates an increased risk for the crash. At the critical point, the system 'explodes', reaches the finite-time singularity.

The very same models are successfully used to forecast other complex systems like the structural integrity of heterogeneous materials, avalanches and even earthquakes. Observed log-periodic oscillations of such (physical) systems close to a critical point seem to be a consequence of physical laws, not some artificial residuals of the mathematical model. One may argue that behaviour of people cannot be predicted with such mechanical model – that may very well be so, but on the other hand, actions of a large group of people with complex interactions and feedback process (price) from the market may become predictable at least some extent on particular situations.

4. Predicting market crash: The LPPL model can be fitted over a financial time series to examine, if the LPPL signature is present, and have an early warning about the imminent crash. The model itself is nonlinear and multi-parameter. Calibrating, finding parameters, for this type of model can be a difficult task. One can say that 'all the parts are moving'.

The model does not 'know' when the market is going to crash, but it can give a probabilistic crash estimate (both time window and the market level) with some probability density function. Actually, the market may not crash at all, but change 'market regime' and prices deflate over a long period of time.

[9] The random walk component is later replaced, see: A Consistent Model of 'Explosive' Financial Bubbles with Mean-Reversing Residuals., Lin et al. (2009).

Following reasons make the LPPL model more attractive than other technical models:

1 There is a plausible theory behind the model (growing imitation among traders).

2 Two different modelling strategies lead to the same final model specification[10].

3 Modelling of the interactions between traders is based on a well documented mathematical properties of complex dynamic systems.

4 Log-periodic oscillations are observed on other similar systems close to a critical point.

Most of the academic studies have found the LPPL model to produce reasonable estimates of market crashes (albeit critique by Chang and Feigenbaum, 2006 & Bree and Joseph, 2010, see also improvements by Lin et al., 2009).

Being able to predict a stock market crash opens the door for making big profits on the options market. Before releasing the model to the public, Sornette and his colleague Olivier Ledoit did precisely that. They made a 400% profit from the US stock market crash on October 27 1997 (DJIA lost 554 points or 7.18%), and have released trading documents from Merrill Lynch to prove it[11].

The well-known critique against 'the money machine' is, of course, valid in this case too: once the money machine method is known by the investment community and many traders seek to profit from it, their activity change the market price (and dynamics?) and previous excess-returns disappear. If the LPPL model turns out to be a valid method to predict market crashes, it will be interesting to see how the markets will be changed, when it becomes commonly known. Perhaps we don't see a sudden large price crashes any more or do we see them earlier than before?

[10] "Two different modelling strategies lead to the same final model specification: (i) a rational-expectation (RE) model of rational bubbles with combined Wiener and Ornstein-Uhlenbeck innovations describing the dynamics of rational traders coexisting with noise traders driving the crash hazard rate; or (ii) a behavioural specification of the dynamics of the stochastic discount factor describing the overall combined decisions of both rational and noise traders." Lin et al. (2009)

[11] The whole story about the model is published in the book 'The Physics of Wall Street: a brief history of predicting the unpredictable' by James Owen Weatherall.

A model that estimates stock market crash risk and even possible timeline seems to be in contradiction with the efficient markets hypothesis. Market crashes are, of course, relatively rare incidents and one could easily claim any predictive power of such model due to luck. Fortunately, there is a growing body of academic LPPL model crash studies from different countries and markets.

1.2 Objectives and contribution

We have three main objectives for the thesis: (i) To examine ex-post if the Helsinki stock market index OMX Helsinki displays the log-periodic power law (LPPL) signatures before the most significant stock market drawdowns. (ii) We also test, if there are LPPL signals present before the largest drawdowns of individual stocks on the Helsinki Stock Exchange. (iii) We examine how accurate are the predictions of the model relative to the actual crash date and value level.

Contribution to the current literature: To the best of our knowledge, there are no previous LPPL studies with the Helsinki stock market data. The most recent development of the LPPL model promises very high confidence levels and robust model specification; and founders ask for more tests (Lin et al., 2009): *"Further validation will come by testing further on other known bubble cases and in real-time. These studies are currently underway and will be reported elsewhere."*

1.3 Organization of the thesis

The rest of the thesis is organized following way: Chapter 2 gives literature review, chapter 3 introduces the research questions, chapter 4 describes the methodology, chapter 5 presents the findings, and the conclusions and further research are discussed in the final chapter 6.

2 Literature review

The LPPL modelling of the stock market crash is based on a study of the complex systems and the mean field theory from statistical mechanics. In the following, we present a basic mathematical derivation of the log-periodic power law.

2.1 The log-periodic power law

Following is a basic derivation of the LPPL model based on Johansen et al. (2000) and Geraskin and Fantazzini (2013) papers. Some of the mathematics is omitted, i.e. more detailed derivation of susceptibility functions and ideas of the mean field theory, which are more complex mathematical physics and can be found from the original sources and their references.

2.1.1 Modeling of the interactions between the traders

The original LPPL model assumes two types of agents in the markets: rational investors ('know' fundamental value of assets) and noise traders, who are subject to irrational herding behaviour (drive prices out of fundamental value). The rational investors continue to invest even with the inflated prices as they are expected to be compensated with higher bubble returns, given that the market does not crash. According to Johansen et al. (2000), a stock market crash happens when imitation among investors grows to a critical point, and all the (noise) traders converge to the sell-side.

The hazard of such incidence is modelled using the mean field theory from statistical mechanics, and a simple way of describing an imitative process is assuming that the hazard rate can be quantified with the equation: $dh/dt = Ch^{\delta}$, where $C > 0$ is a constant and δ is the average number of interactions among traders minus one. If we integrate the previous, we get (Johansen et al., 2000):

$$h(t) = \left(\frac{h_0}{t_c - t} \right)^{\alpha}, \alpha = \frac{1}{\delta - 1} \tag{1}$$

, where t_c is the critical time when the system is expected to change state (traders converge to sell-side) and the exponent α must be between the zero and one; otherwise the price would not be finite when the system reaches the critical time.

In the microscopic modeling of the market, Johansen et al. (2000) connect traders in a network. The idea of the model is known as the Ising model in the statistical-mechanics[12]. Every trader has only two possible states: buy ($s_i = 1$) or sell ($s_i = -1$). State of a trader is determined by neighbouring nodes (traders) in the network as well as the global influence (i.e. news) and idiosyncratic behaviour. State of a trader (s_i) is modelled using the Markov process (Johansen et al., 2000):

$$s_i = sign\left(K \sum_{k \in N(i)} s_k + \sigma\varepsilon_i + G \right) \tag{2}$$

, where sign(x) is a function equal to +1 if $x > 1$ and -1 if $x < 1$. K is a positive constant governing the imitation among traders, and the error term is an iid standard normal variable for idiosyncratic behaviour. The N(i) represent the agents that have direct connections to agent i. The G marks the global influence available to all traders.

If K increases, the order in the network increases. If σ increases, the order decreases. There is a critical point K_c: When $K < K_c$ imitation between traders is low, and the system has a

[12] *"The two-dimensional square-lattice Ising model is one of the simplest statistical-mechanics models to show a phase transition. The model consists of discrete variables that represent magnetic dipole moments of atomic spin, and they are in one of two stages (+1 or -1). The spins are arranged in a lattice, and each interacts with its neighbours and with an external magnetic field."* Kamimura and Ohira (2019), 28.

low sensitivity to a small change in global influence. When $K > K_c$, the nodes can convert to the same state (sell) and thus cause a crash.

The susceptibility of the system (χ) defines a system's sensitivity to a small change in global influence. In our case, a change in the average stage of the traders per small change in global influence. In the simplest form, the susceptibility is defined as (Johansen et al., 2000):

$$\chi = \frac{dE\left[M\right]}{dG}\Big|_{G=0} \tag{3}$$

, where

χ is the susceptibility of the system.

M is the market, defined as the average state of traders ($\bar{s}$).

G is the global influence.

2.1.2 Derivation of the log-periodic power law

The dynamics of the asset price p_t is captured as a stochastic differential equation (Johansen et al., 2000):

$$dp_t = \mu_t p_t dt - kp_t dj \tag{4}$$

, where the first part is a price process with a drift μ_t and the second part is a conditional price crash by the size of percentage $k \in (0, 1)$. The condition $j = 0$ before the crash and $j = 1$ when the crash happens.

If we assume no arbitrage conditions and the price follows martingale process ($E[dp_t] = 0$), we get (Johansen et al., 2000):

$$dp_t = \mu_t p_t dt - kp_t h_t dt = 0 \tag{5}$$

, where h denotes *the hazard rate, the probability per unit of time of crash taking place in the next instant, given it has not yet occurred* (Geraskin and Fantazzini, 2013).

From the previous equation 5 we get $\mu_t = kh_t$, which leads to price dynamics before the crash $d(\ln p_t) = kh_t dt$, whose solution is (Johansen et al., 2000):

$$\ln\left(\frac{p_t}{p_{t_0}}\right) = k\int_{t_0}^{t} h_s \, ds \tag{6}$$

By following a theory of rational expectations, Johansen et al. (2000) expect traders to incorporate the risk of the market crash to their price expectations and market price should increase to compensate for the increased risk. "*The idea is that the higher the probability of the crash is, the faster the price should grow to compensate investors for the increased risk of a crash in the market*" (Geraskin and Fantazzini, 2013).

Using previously described price dynamics, Johansen et al. (2000) show that a system variables close to a critical point can be described by the power law, and the susceptibility of the systems diverges to:

$$\chi \approx A\left(K_c - K\right)^{-\gamma} \tag{7}$$

, where A is a positive constant, K_c is a tendency toward imitation at the critical point and $\gamma > 0$ is called the critical exponent of the susceptibility.

Critical exponents describe the behaviour of physical quantities near-continuous phase transitions (critical point). It is believed, though not proven, that they are universal, i.e. they do not depend on the details of the physical system (Gunasekaran, 2012). Johansen et al. (2000) have shown that using a physical model of critical phenomena is relevant, because stock market returns display similar power law probability distributions than physical systems near the critical point (figure 1 and 2).

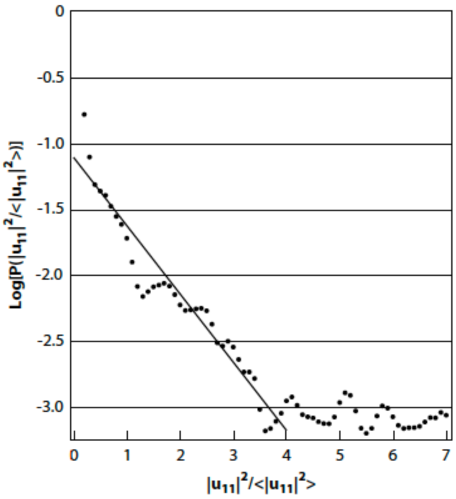

Figure 1: Probability distribution of the square of the fluid velocity Sornette (2004).

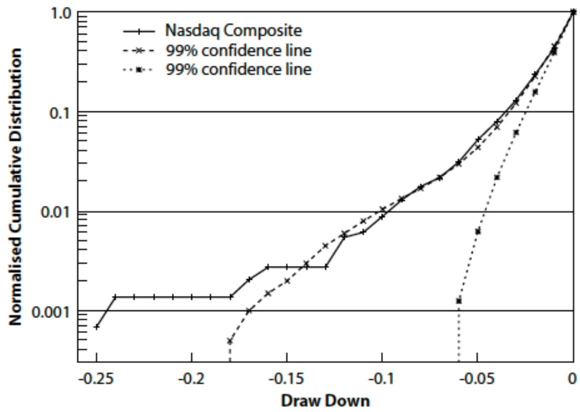

Figure 2: Normalized cumulative distribution of drawdowns in the Nasdaq composite since its establishment in 1971 until April 18, 2000. Sornette (2004)

Bi-dimensional Ising-model describes connections of the traders in a uniform way, while in the real markets, some agents are more connected than the others. To solve this, Johansen et al. (2000) introduce a hierarchical diamond lattice –model, where connections between agents vary in the lattice (figure 3).

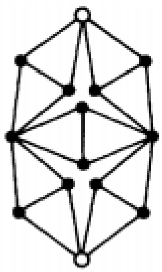

Figure 3: A small 2-dimensional hierarchical lattice, some of the nodes are more connected than the others.

Hierarchical diamond lattice has a fractal structure, and a version of this model was first solved by Derrida et al. (1983). The most significant difference is that the critical exponent (γ) can now be a complex number. The diamond lattice -model has a general solution of susceptibility as follows (Johansen et al., 2000).

$$\begin{aligned}\chi &\approx \Re\left[A_0\left(K_c - K\right)^{-\gamma} + A_1\left(K_c - K\right)^{-\gamma+i\omega} + \dots\right] \\ &\approx A_0'\left(K_c - K\right)^{-\gamma} + A_1'\left(K_c - K\right)^{-\gamma}\cos\left[\omega\ln\left(K_c - K\right) + \psi\right] + \dots\end{aligned} \tag{8}$$

, where A_0, A_1 are real numbers and $\Re[...]$ represent the real part of a complex number. If K evolves smoothly, we can have approximation near the critical point that $K_c - K$ varies as $t_c - t$, and we get the hazard rate of (Johansen et al., 2000):

$$h(t) \approx B_0\left(t_c - t\right)^{-\alpha} + B_1\left(t_c - t\right)^{-\alpha}\cos\left[\omega\ln\left(t_c - t\right) + \psi'\right] \tag{9}$$

The power law in this equation is corrected by log-periodic oscillations (Johansen et al., 2000). The origin of the oscillations is the hierarchical network of the diamond lattice. Traders imitate their closest neighbours on the lattice and state of individual traders oscillates through the network as some traders (nodes) are more connected than the others.

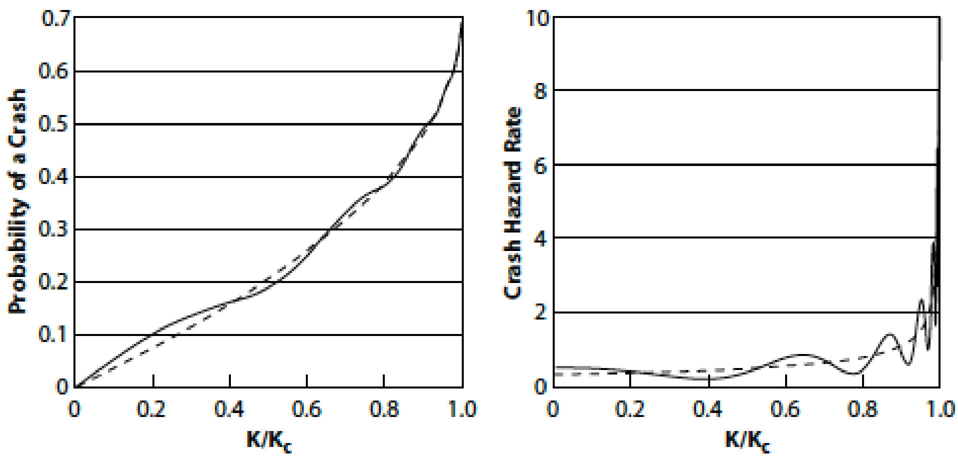

Figure 4: Log-periodic oscillations. The figure shows simulated log-periodic oscillations decorating non-linear trend. The frequency of the oscillations increases toward the critical point. Figure source: Sornette (2004).

The LPPL model connects imitation of traders to the positive feedback of the market price to a single model. The price dynamics is defined by the bullish or momentum traders (s' = 1) and bearish traders (s'' = -1) expecting the market crash.

By applying equation 9 to equation 6, we get the expected asset price dynamics before the crash (Johansen et al., 2000):

$$\ln\left[E(p_i)\right] \approx \ln(p_c) + \frac{k}{\beta}\left\{B_0\left(t_c - t_i\right)^\beta + B_1\left(t_c - t_i\right)^\beta \cos\left[\omega \ln\left(t_c - t_i\right) + \varphi\right]\right\} \tag{10}$$

, 'which can be rewritten in a more suitable form for fitting a financial time series' (Geraskin and Fantazzini, 2013):

$$\ln\left[E(p_i)\right] \approx A + B\left(t_c - t_i\right)^\beta \left\{1 + C\cos\left[\omega \ln\left(t_c - t_i\right) + \varphi\right]\right\} \tag{11}$$

The equation 11 is known as the log-periodic power law (LPPL), where A is the expected asset price (ln) in the critical moment, B is the increase of ln asset price over time before the critical moment t_c and β is an exponent quantifying the super-exponential growth path. The parameter C is a proportional magnitude of the oscillations around the growth path, while ω is the angular frequency of oscillations during the bubble, and φ is a phase parameter.

More discussion and the definition of the parameters can be found from the chapter ‘methodology’. The idea is that by fitting the model to price data, it reveals information about the phase of the underlying structure (mode of the traders) and how far the system is from the critical point. This information is revealed by the oscillations arising from mutual imitation of traders connected by the lattice.

3 Research questions

The thesis has three main research questions:

1. Do we find evidence that the log-periodic power law (LPPL) signatures are present before the most significant stock market drawdowns in the Helsinki Stock Exchange? We test this by fitting the LPPL model to the five most significant OMX Helsinki index drawdowns.

2. Do we find evidence that the log-periodic power law (LPPL) signatures are present before the most significant individual stock crashes in the Helsinki Stock Exchange? We test this by fitting the LPPL model to the ten most significant single stock drawdowns.

3. If we find the LPPL model signatures: How accurate are the modelled estimates for the critical time and peak values?

Data: OMX Helsinki index and the main list stock price data.

4 Methodology

Following is a presentation of the original methodology by Johansen et al. (2000) to fit the LPPL model to financial time series. We use the original methodology with the later mean-reverting residuals specification (Lin et al., 2009).

4.1 The LPPL model

The original LPPL model rewritten as (Geraskin and Fantazzini, 2013):

$$\ln\left[E(p_i)\right] = A + B\left(t_c - t_i\right)^{\beta} + C\left(t_c - t_i\right)^{\beta}\cos\left(\omega \ln\left(t_c - t_i\right) + \varphi\right) \tag{12}$$

, where

p_i is the asset price

A is the expected value (ln) of the asset at the critical time, A > 0, when $t = t_c$ (i.e. time = critical time) all the other terms vanish.

B is the correlation coefficient in term responsible of faster-than-exponential growth

C is the correlation coefficient in the term responsible of the oscillations

t_c is the expected time when the super-exponential price growth ends in crash or change of regime.

β is the exponent controlling faster than exponential price growth ($0.1 \leq \beta \leq 0.9$). Smaller values indicate stronger super-exponential price growth.

ω is the angular frequency of the log-periodic oscillations during the bubble.

φ is the phase parameter $0 < \varphi < 2\pi$

4.2 Estimating parameters of the LPPL model

The LPPL model can be written compactly into a form (Geraskin and Fantazzini, 2013):

$$y_i = A + Bf + Cg \tag{13}$$

, where

$$y_i = \ln\left[E(p_i)\right]$$
$$f = (t_c - t_i)^{\beta}$$
$$g = (t_c - t_i)^{\beta}\cos\left(\omega\ln(t_c - t_i) + \varphi\right)$$

Parameters of the linear subsystem (A, B and C) can be found by ordinary least squares (OLS) once the nonlinear system parameters (t_c, β, ω, φ) are fixed. Matrix notation of the OLS:

$$\begin{aligned} &X' y = (X' X) b \\ &s.t. \\ &\hat{b} = (X'X)^{-1} X' y \end{aligned} \tag{14}$$

, where

$$X = \begin{pmatrix} 1 & f_1 & g_1 \\ \vdots & \vdots & \vdots \\ 1 & f_n & g_n \end{pmatrix}, b = \begin{pmatrix} A \\ B \\ C \end{pmatrix} \tag{15}$$

Calibrating the LPPL model is a difficult and computationally expensive task: noisy data, relatively small number of data points (near the drawdown) and large number of parameters.

It is easy to see that the LPPL model fit is controlled by the four nonlinear parameters (t_c, β, ω, φ) and the linear subsystem (parameters A, B and C) has unique solution that can be found by the OLS once the four free parameters are fixed.

One cannot simply use any non-linear solver to find values for all the parameters at once. This will generate spurious fits. Calibration of the model is based on some combination of finding suggested solutions for parameters, freezing some of the parameters and using a nonlinear solver to find solutions for the free parameters. The fit may be improved working from non-linear to linear part and vice versa.

Many search algorithms, like Levenberg–Marquardt algorithm, are sensible to the selected initial values. They can find local minimas instead of global, when initial solutions are less optimal. Genetic algorithm used in this paper is less prone to this problem, but not totally immune.

4.1.1 Search space and filtering rules

The calibration algorithm selected on the thesis is a multi phase genetic algorithm fitting with following search space and filtering rules.

Table 1: The LPPL parameters search space and filtering rules used in the thesis.

The LPPL Parameters and statistics		
Item	**Search space**	**Filtering**
A	A > 0	A > 0
B	B < 0	B < 0
C	| C | < 1	| C | < 1
β	[0.1, 0.9]	[0.1, 0.9]
ω	[2, 25]	[2, 25]
φ	[0, 2π]	[0, 2π]
t_c	[t_2, t_2 + 1 year]	[t_2, t_2 + 0.25 year]
Number of oscillations	-	[2.5, +∞]
Damping factor	-	[1, +∞]
Relative error	-	[-0.2, 0.2]

The search space and filtering rules are based on the previous studies (Sornette et al., 2015) with slight modification to fit calibrations with fewer time windows.

The filtering restrictions for the parameters B, β, ω and C are 'stylized facts' of the LPPL model documented in numerous studies[13]. The limits on B and β ensures a faster than an exponential acceleration of the log-price at the critical time. The restriction of ω limits log-periodic oscillations not be neither too fast (fitting random noise) or too slow (contribute to the trend or momentum) (Lin et al., 2009). Previous studies have found the importance of fundamental log-periodic angular frequency of value $\omega \approx 6.4 \pm 1.5$ (Zhou and Sornette 2009,

13 Lin et al. (2009). For more suggestions for typical values see Sornette (2004), 335.

Johansen 2003, Johansen and Sornette 2002). The last restriction of C ensures that the hazard rate of the crash always remains positive (Bothmer and Meister, 2003).

We search the critical time within one year after the end of the calibration time window. In general, if the critical time is too far from the end of the calibration window, it is considered unreliable and it must be re-evaluated in future calibrations. Sornette et al. (2015) suggest in their recent work that the critical time should be within $t_2 + 0.1dt$ (dt = calibration time window's length).

Number of oscillations is a number of modelled oscillations calculated as $(\omega/2\pi)*\ln((t_c - t_1) / (t_c - t_2))$, where t_1 is the start of the calibration time window and t_2 is the end of it (Zhou and Sornette, 2009). Previous studies have show that about 2.5 modelled oscillations on the time window is sufficient to capture the real log-periodic oscillations (Sornette et al., 2015).

Limits on the damping factor = $(\beta|B|)/(\omega|C|)$ is based on the fact that the crash hazard rate h(t) is non-negative by definition (Bothmer and Meister, 2003). The relative error is $(p_t - p'_t)/ p'_t$ and limited to 20% on the single largest error (Sornette et al., 2015).

If the calibration matches with the filtering rules, model residuals are stationary and the critical time estimate is within reasonable time, the LPPL signature is present.

The LPPL model is susceptible to the value of free parameters. It can be modelled over any time series with some respect, so multiple tests are required to verify that we do not have only spurious regression. See parts Lomb-Scargle spectral analysis and residual testing.

4.1.2 Genetic algorithm

In the thesis, we found the LPPL parameters using the genetic algorithm as a search tool. Genetic algorithms (GA) were first suggested by Holland (1975).

The basic steps of the genetic algorithm (modified from Mitchell, 1999, 8-9):

1. Generate a random population of n solutions (chromosomes) or use initial values provided by the user.

2. Evaluate the fitness of each chromosome in the population.

3. Select chromosomes for preproduction based on fitness. Apply genetic operators and create a new off-springs.

4. Replace the current population with new off-springs.

5. Test, if the end conditions are satisfied: the maximum number of loops met, or fitness is not improving anymore according to the rules. If not, go back to step 2.

Genetic algorithms have their limitations. *Holland's schema theory states that the number of low-order, low defining-length schemata with above-average fitness increases exponentially between successive generations* (Banzhaf et al., 2018, 18). This means that the genetic algorithm is expected to find better solutions with consecutive generations. The schema theorem holds in un-finite populations (in theory), but may fail in practice. In general heuristic algorithms (like GA) produce good fits instead of the best fit.

At the first phase of the thesis, we 'beta tested' the LPPL calibrations with genetic algorithm tool NLP Solver by Sun Microsystems. The search space was heavily restricted using pre-found limits, and we managed to get consistent calibration results with this system. However, it was clear that it could not be used beyond testing. The tool version comes with a heavy front-end (OpenOffice), resulting in a poor calculation performance measured in computing time. A single approximation may take hours.

To improve computing time and have even more accurate results, we designed a multi-phase calibration program based on the R statistical program and 'genalg' genetic algorithm package for R by Egon Willighagen and Michel Ballings.

Multi-phase GA calibration program:

1. Following a suit by Liberatore (2010), we first fit a simple model by setting parameters C = 0 and β = 1. This is a computationally effective way to avoid the algorithm getting stuck on a local optimum later. The best fit solution for the function $\ln(p_t) = A + (t_c - t_i)$ is used as a seed on the second phase.

2. On the second phase, we loosen the previous limits and use the genetic algorithm to find candidates for the general solution within the search limits.

3. On the third phase, we freeze the other parameters and fine-tune the critical parameters in sequences to see, if the fit of the LPPL-model improves measured with the mean squared error (MSE). E.g. Parameters ω, φ and t_c have strong correlation near the optimal LPPL solution. If the parameter ω have been fixed, the LPPL fit would still improve by tweaking the values of φ and T_c (Liberatore, 2010).

4. Finally, the program reports the calibrated parameters, generates the plot files and test statistics to measure the fitness of the calibration.

The calibration program can be driven multiple times with different start up values (e.g. time windows) to find successful results.

In addition to test statistics, we get instant feedback of the fit with a visual examination of the LPPL-function over the price plot. This may feel unsound, but actually, visual pattern recognition is a human forte. Some researches have proposed using computer visual pattern recognition techniques to find better-suggested values for ω and φ (Liberatore, 2010), much of the same way human can gauge the fitness of the model. With some experience, it is straightforward to see if the LPPL-fit captures the price gyrations and if the fit, most likely, could be improved by running the GA with a different calibration time window. In the end, the MSE is, of course, unbiased guideline here.

Using the R based system, our calibration results were improved measured with the mean squared error and the computing time was enhanced. However, it still takes about 20 minutes to calculate a single calibration. The computing time can be improved substantially by using a newer calibration scheme of the LPPL model (Filimonov and Sornette, 2013) and more robust and faster search methods that it makes possible (e.g. Nelder–Mead or Levenberg–

Marquardt algorithm). A custom program can also utilise a computer's graphical processor unit (GPU) for parallel computing (see Using GPU to increase computing power).

4.1.3 Graphical analysis of the LPPL calibrations

We transfer the LPPL calibration data to Matlab software to analyse the object function non-linear search space and to have confirmation that the genetic algorithm (GA) system finds optimal or close to optimal parameter values.

Nokia data is used as an example on the following 3D-plots. While the LPPL calibration parameters are found using our genetic algorithm system on R statistical software, plots are generated with Matlab. Matlab programming is required due to different data dimensions. There are single LPPL parameter values, and for example, 750 pairs of date and price data for the calculus of every single data point on the 3D-plot.

The contour of the 3D-plots varies greatly depending on the underlying price data. More volatile single stock prices create more ridged and chaotic contours than less volatile stock indices. Thus single stock LPPL calibrations are prone to a more significant parameter estimation error and less accurate critical moment projections on average.

The LPPL parameters on x and y –axes are allowed to change while the others are frozen. The mean squared error (MSE) of the object function is calculated on every data point on the 3D-plot with actual price data within the calibration time window. The minima of the MSE is marked with the red dot. The Minima is compared with the calibrated values from the GA. Matlab displays the minima within plots drawing accuracy, and this is sufficient for the visualisation and testing the concept. Following calibration is used as an example.

Table 2: LPPL calibration to Nokia, local peak at 2000.333. 't.end' is the end of the calibration time window, trading days before the local peak [-130 = six months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error, and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-130	500	4.500	-1.677	2000.375	0.066	0.594	14.967	2.967	0.012	0.972

Following 3D-plots point out that our LPPL calibrations are optimal in pair by pair examination of non-linear parameters and the genetic algorithm system is working. However, this is not proof that the whole set of seven LPPL parameters is optimal. Due to the complexity of the calibration problem and heuristic search algorithms, the general optimal solution may be unattainable and unknown. In practice, good-enough-fit has to be accepted.

On the figure 5 omega (ω) and phi (φ) are let to change. There is a single global minima of the cost function MSE = 0.0123 on ω = 15.0 and φ = 3.0. This is in line with more accurate calibrated values of ω = 14.967 and φ = 2.967. There are also multiple local minimas on the search space, where the search algorithm could get trapped. One can also notice that ω and φ correlate generating valleys and hills. The LPPL fit can many times be slightly improved by freezing ω and let the search algorithm to find improved φ and T_c (Liberatore, 2010). This is implemented on the R code.

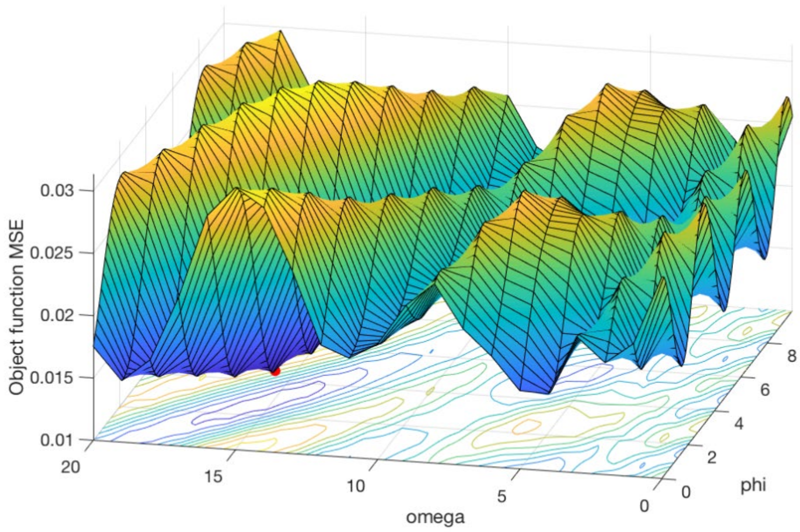

Figure 5: The cost function when omega (ω) and phi (φ) are varying.

On the figure 6 omega (ω) and critical time (t_c) are let to change. There is a single global minima of the cost function MSE = 0.0123 on ω = 15.0 and t_c = 2000.375. This is in line with more accurate calibrated values of ω = 14.967 and t_c = 2000.375.

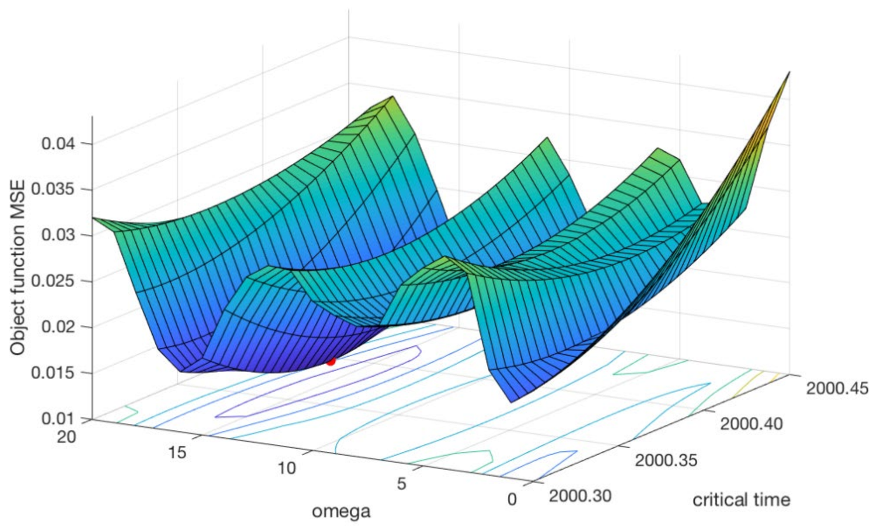

Figure 6: The cost function when omega (ω) and critical time (t_c) are varying.

On the figure 7 beta (β) and critical time (t_c) are let to change. There is a single global minima of the cost function MSE = 0.0123 on β = 0.6 and t_c = 2000.370. This is in line with more accurate calibrated values of β = 0.594 and t_c = 2000.375.

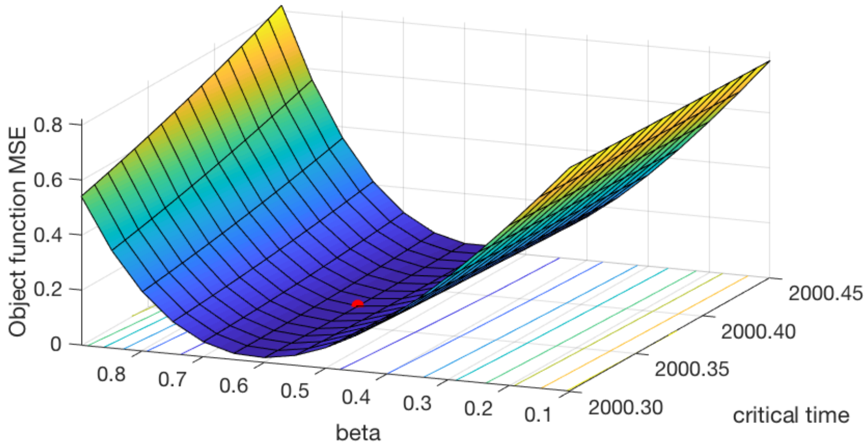

Figure 7: The cost function when beta (β) and critical time (t_c) are varying.

On the figure 8 beta (β) and omega (ω) are let to change There is a single global minima of the cost function MSE = 0.0123 on β = 0.6 and ω = 15.0. This is in line with more accurate calibrated values of β = 0.594 and ω = 14.967.

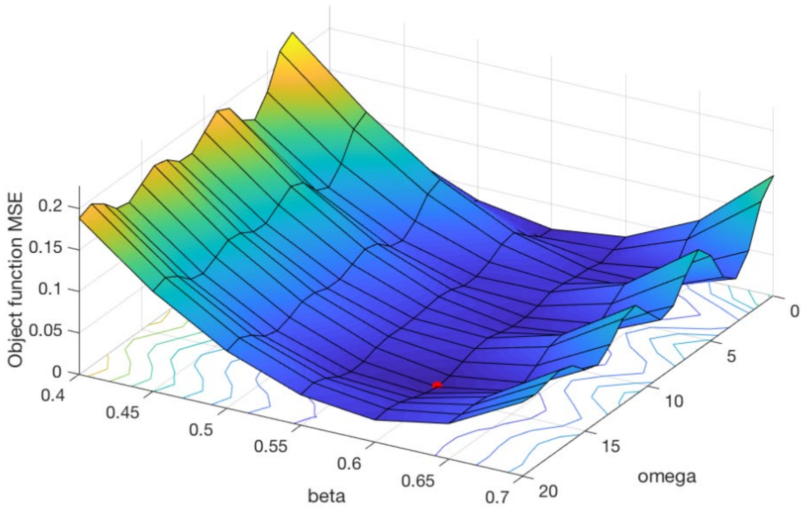

Figure 8: The cost function when beta (β) and omega (ω) are varying.

4.3 Calibration accuracy

The sloppiness of the LPPL model fits and difficulties with the search algorithms have made critics (Bree and Joseph, 2010) challenge the relevance of the LPPL fits. Time series of financial instruments' price data are noisy, and even if they carry LPPL signatures, the signatures are challenging to identify, according to the critics.

Sornette et al. (2013) admit the difficulties with the original LPPL model, but previous calibration algorithms are tested with (Sornette et al., 2013) synthetic data[14].

The LPPL signal was buried in fractional Brownian noise with various Hurst exponents from anti-persistent to persistent. Two types of noise were used: Gaussian noise and student t distribution noise with four degrees of freedom. Sornette et al. (2013) generated 200 synthetic time series for each type of noise. Each time series was calibrated with the LPPL model using the Levenberg-Marquart algorithm. Following table displays the results of estimates with ten best realisations of each time series.

Table 3: The parameter values used to generate the synthetic data are shown in the second column "Reference". The mean and standard deviation values of the parameters obtained by fitting the model to the synthetic LPPL time series with two types of noise discussed in the text are given in the last two columns. Reproduced from Sornette et al. (2013).

Finding the LPPL signal from a noisy data			
Parameter	**Reference**	**Mean (std) of Gaussian**	**Mean (std) of Student's t**
t_c	300	296.07 (22.44)	295.15 (20.81)
β	0.7	0.74 (0.15)	0.72 (0.18)
ω	10	9.75 (1.43)	9.71 (1.47)

The results show that the LPPL signal can be found from noisy time series reasonable well with the described search algorithm. The test shows a small bias on the values: critical time (t_c) and omega (ω) are underestimated on average while exponent beta (β) is overestimated.

[14] New improvements in calibration: Filimonov and Sornette (2013) suggest a new robust calibration procedure that can reduce nonlinear search space to tree dimensions. According to Filimonov and Sornette, the new procedure can find the best fit in most cases of actual bubble regime.

4.4 What kind of bubbles can be detected?

Explosive price bubbles and their critical times can be modelled with the LPPL model if the following criteria are met: (i) The bubble demonstrates super-exponential price growth near the critical point (the growth rate itself is affected, leading to the increasing instability of the system). (ii) The crash is preceded by a price bubble with fluctuations created by groupthink of investors (captured by the LPPL model). (iii) The crash must be endogenous – a result of investors' imitation and herding.

Explosive price bubbles in their extreme are relatively rare incidents in financial markets. Their impact, however, is massive to the long term investment returns and investors' financial well being. A model that can flag an early warning is thus well needed from the practical point of view.

4.5 Can LPPL signatures be created by pure chance?

Financial time series are typically noisy, and possible signals carried by time series are impaired by low signal-to-noise ratios. The efficient market theory suggests that markets follow a random walk and indications of predictive signals are only due to random fluctuations. The LPPL model has been tested against simulated stock market data to determine if the LPPL signatures are present in random walk time series and what is the rate of false positives (error of the first kind).

Lin et al. (2009) have tested the volatility-confined LPPL model on simulated data and compared the results with results on the real stock market data. They used the S&P 500 index from January 3. 1950 to November 21. 2008 as real data.

Simulated time series are generated based on the GARCH (1,1) model:

$$\begin{aligned} \ln I_t - \ln I_{t-1} &= \mu_0 + \sigma_t \zeta_t \\ \sigma_t^2 &= \sigma_0^2 + \alpha\left(\ln I_{t-1} - \ln I_{t-2} - \mu_0\right)^2 + \beta\sigma_{t-1}^2 \end{aligned} \tag{16}$$

The GARCH (1,1) -model parameters are estimated on the S&P 500 index from January 3. 1950 to November 21. 2008. Lin et al. (2009) generated two sets of 1000 synthetic GARCH time series: i) samples of random lengths from 750 days to 1500 days and ii) samples of the fixed length of 1500 days.

Only 0.2% per cent of simulated time series (i) match LPPL conditions (0.1% on fixed sample length ii). Tests on residuals show a remarkable low rate (0.2%) of false positives.

From S&P 500 data (the same time window as on parameters estimation) they generated two sets of time windows of 750 trading days. The first time window (respectively second) slides trough the S&P 500 time series in increments of 25 days (respectively 50 days). Results show that 2.49% of time windows I (25 days slide) and 2.84% of time windows II (50 days slide) obey the LPPL conditions indicating bubble regime. This is more than tenfold the LPPL signals received from the simulated time series. All the positive LPPL signals match with well-known bubble regimes, and the null hypothesis (not stationary) for residuals is rejected every time indicating a high level of model specification.

The test shows that the LPPL conditions are not easily met by chance. Lin et al. (2009) report a low rate (0.2%) of false positives when the LPPL is applied to the GARCH benchmark. Out of the sample case of seven known international bubble regime show very high confidence level (99.9%) of residual testing (all but one).

4.6 Calibrating prices or logarithmic prices

It is possible to calibrate the LPPL model to the price data instead of logarithmic prices. In general, logarithmic prices should be used instead of regular prices when calibrating the LPPL model (Sornette et al., 2013): If the expected crash is proportional to the price increase during the bubble, logarithmic prices should be used. If the data values change several orders of magnitude over time, the standard least square fit is not suitable anymore, and the normalized least-square minimization is recommended instead. In the thesis, we calibrate logarithmic prices only.

4.7 Calibrating multiple time windows

In the thesis, we use arbitrary calibration time windows ending 0, 22 (a month), 42 (2 months), 65 (3 months) and 130 (about six months) trading days before the known crash. The length of the calibration time window varies between 250, 500 and 750 trading days. If the previous time windows do not provide positive results, we seek to find individual time windows that may carry log-periodic oscillations. This process is naturally prone to data snooping, but we control the findings with filtering rules. More about data snooping can be found from the paragraph 'The problem of data snooping'. Successful or example calibrations are reported.

The selection of the time window parameters is based on similar values used on previous studies. The optimal time window's length seems to depend from the nature of the bubble regime in question (time span when the positive feedback mechanism and herding are manifested according to the LPPL model) and also from time series volatility and volume of the noise component.

The number of time windows in the thesis is sufficient to answer the research questions: do we find the LPPL signatures in the Helsinki Stock Exchange and how accurate they are in general. The accuracy of the LPPL critical time calibrations may be improved by introducing more time windows.

Professional applications use multiple shrinking time windows (for example +/- five trading days) to capture various estimates for the LPPL parameters. This was not feasible in the thesis due to slow computing performance of the estimation tools used.

In the case of professional use, more powerful methods and tools should be used. For example, *ETH Financial Crisis Observatory* uses the *Brutus* cluster supercomputer to fit the LPPL model in various financial time series.

4.8 Lomb-Scargle spectral analysis

We use Lomb-Scargle spectral analysis and residual analysis to examine the fitness of the calibration. This methodology is used in many previous studies (Jiang et al., 2010).

Lomb-Scargle spectral analysis is the de facto tool for analysing different time series, for example, in astrophysics. It can be used to find the dominant frequencies in unevenly sampled time series (for example, stock price data, no data points outside the trading days).

Angular frequency (ω) in physics is the rate of change of the phase of a sinusoidal waveform (in our case, a scalar measure of the expected LPPL oscillations in the price data, one complete wave cycle).

One 'revolution' of the LPPL cosine-part controlled waveform is equal to 2π radians [-π,+π],

$$\omega \equiv 2\pi / T = 2\pi f \tag{17}$$

, where

f is ordinary frequency (typically measured in hertz).

The angular frequency (ω) is the same ω-parameter as in the LPPL model.

In order to find out if the LPPL oscillations are present, we perform Lomb-Scargle spectral analysis to de-trended price residuals to find out dominant frequencies.

The normalized L-S periodogram at frequency ω is (Press and Rybicki, 1989):

$$P_n(\omega) = \frac{1}{2\sigma^2}\left\{ \frac{\left[\sum_j (h_j - \bar{h})\cos\omega\left(t_j - \tau\right)\right]^2}{\sum_j \cos^2\omega\left(t_j - \tau\right)} + \frac{\left[\sum_j (h_j - \bar{h})\sin\omega\left(t_j - \tau\right)\right]^2}{\sum_j \sin^2\omega\left(t_j - \tau\right)} \right\} \qquad (18)$$

, where

h_j = observation

t_j = observation time

σ^2 = variance of observations

We have to convert observation times as the LPPL model oscillations are log-periodic, thus $t_j' = \ln(t_c - t_j)$.

The frequency-dependent time offset τ is evaluated at each ω

$$\tan 2\omega\tau = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j} \qquad (19)$$

Lomb-Scargle spectral analysis is carried out to de-trended residuals data from the same time window as used for model calibration.

The series of de-trended 'lomb' residuals are calculated as (Jiang et al., 2010):

$$r(t) = x^{-\beta}\left(\ln\left[p(t)\right] - A - Bx^{\beta}\right) \tag{20}$$

, where $x \equiv t_c - t$ and A, B, t_c, β have been found via the LPPL model calibration. These residuals represent the log-periodic oscillations over the super-exponential price growth path.

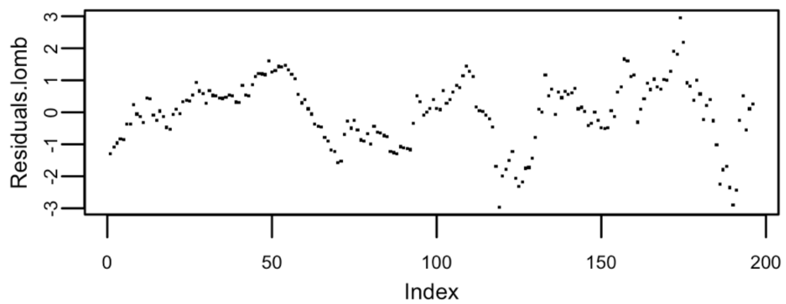

Figure 10: De-trended Valoe [0, 195] lomb residuals. The frequency of the oscillations increases toward the critical point as the LPPL model predicts. The calibration time window is in the square brackets, see individual stock calibrations.

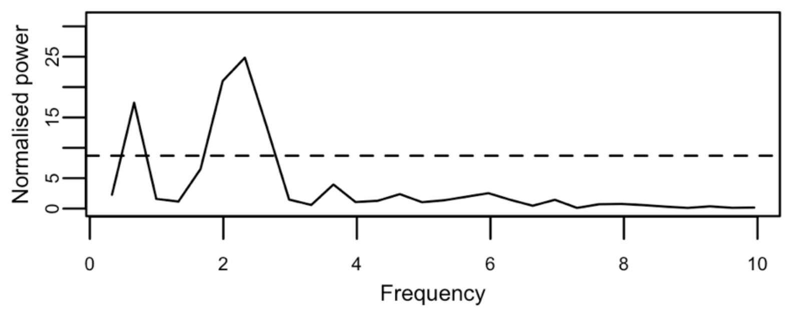

Figure 9: Lomb-Scargle periodogram on de-trended Valoe [0, 195] residuals. The dominant frequency in the Valoe example is 2.3218, which corresponds almost exactly to the calibrated LPPL ω-parameter value of 14.582 ($\omega_{lomb} \equiv 2\pi f \approx 2 \times 3.1416 \times 2.3218 \approx 14.588$). The horizontal line indicates the 99% confidence level. The calibration time window is in the square brackets, see individual stock calibrations.

The Lomb-Scargle dominant frequency should correspond to the calibrated angular frequency ($\omega \approx \omega_{lomb}$) to verify that the same frequency of oscillations is present in the price data (as in the figure 9). However, using a filter based solely on ω is challenging, particularly for an individual stock, due to noisy data.

Previous studies have found the importance of fundamental log-periodic angular frequency of value $\omega \approx 6.4 \pm 1.5$ (Johansen 2003, Johansen and Sornette 2002) and its first and second harmonics ($n\omega$). The importance of harmonics is a result of the concept of discrete scale invariance.

All the results of the Lomb-Scargle spectral analysis are reported in the 'summary of the results' -chapters.

4.9 Residual analysis

Analysing residuals (observed value – predicted value) is an important task to evaluate a model fit to data. Non-stationary residuals may indicate spurious regression, and it makes the statistical reasoning unsound.

Lin et al. (2009) suggest that the volatility-confined LPPL model residuals should follow a mean-reverting Ornstein-Uhlenbeck process (stationary mean-reverting process).

Introducing the Ornstein-Uhlenbeck process to the residuals Lin et al. (2009) make the price process (conditionally) stochastic, but the residuals have a tendency to 'mean revert' toward the LPPL trajectory.

> *"While keeping the structure of the model based on time-varying expectations of future returns, the daily logarithmic returns are no longer described by a deterministic drift decorated by a Gaussian-distributed white noise. Instead, specifying a mean-reversal noise component, the no-arbitrage condition predicts that the expected returns become stochastic, which represents the on-going reassessment by investors of the future returns."*
>
> Lin et al. (2009)

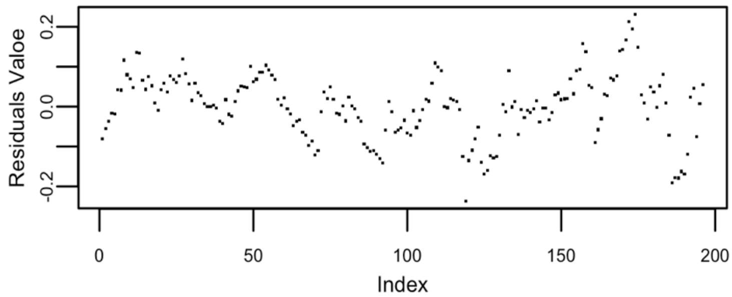

Figure 11: Residuals of the Valoe [0, 195] LPPL calibration. Residuals are stationary at 99% confidence level p(DF) = 0.01 and p(PP) = 0.01. The calibration time window is in the square brackets, see individual stock calibrations.

The figure 11 presents the Valoe residuals. Residuals display mean-reverting tendency. In the LPPL model framework, 'the deterministic LPPL price trajectory is the unbiased expectation of a representative rational agent in the market, while the stochastic (mean-reverting) component describes the estimation errors' (Lin et al., 2009).

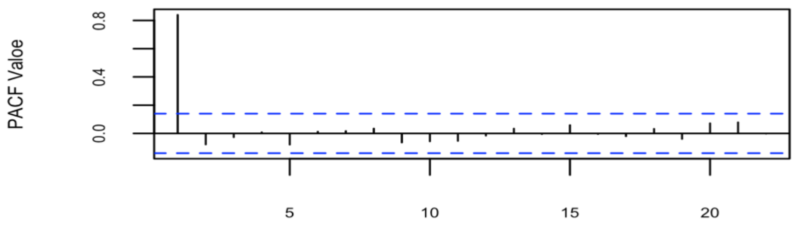

Figure 12: The partial autoregression correlation function (PACF) of the Valoe [0, 195] LPPL calibration. Horizontal lines indicate the 95% confidence level. All values with lags larger than one are non-significant, indicating the absence of linear dependence. The PACF proofs that the residuals are not white noise. The calibration time window is in the square brackets, see individual stock calibrations.

In the figure 12 we plot the partial autoregression correlation function (PACF) of Valoe residuals. We notice a strong correlation at the first lag and rest of the lags are non-significant. This indicates that the residuals are not white noise, but the result is in line with the Ornstein-Uhlenbeck process mean revert specification.

The stationarity of the residual series is tested using Dickey-Fuller (DF) and Phillips-Perron (PP) tests. Both tests have a null hypothesis 'not stationary', and small p-values indicate stationarity of the series. In the LPPL tradition, calibrations with the 99 % confidence level of stationarity of the residuals are considered successful (e.g. Jiang et al., 2010). This translates to p-values of 0.01 or less on the DF- and PP-tests.

The residual test combined results are reported in the 'summary of the results' -chapters.

4.10 The problem of data snooping

The first academic works in the area were single calibrations on price data, which can have multiple problems. Finding a single time window where the calibration filtering rules are met, can lead to data snooping and fragile signal findings that are not present on the other time windows and are not very trustworthy.

Data snooping, finding patterns from the data that seem meaningful but are spurious, is a severe problem in all financial data analysis:

> *" Data-snooping biases can never be completely eliminated; they are an unavoidable aspect of nonexperimental inference and are a particular problem in nonlinear models because of the large degrees of freedom involved in these models. Awareness of the influence of data snooping is the most important step in dealing with the problem. Solutions to the problem will almost never be statistical in nature; rather, analysts should look to some kind of framework to limit the number of possibilities in their search. The framework might come from economic theory, psychological theory, or the analyst's intuition, judgment, and experience."*
>
> Andrew W. Lo, Professor of Finance, MIT Sloan School of Management[15]

The data snooping problem is unavoidable with a single LPPL calibration. We are seeking a time window where the LPPL signal can be identified, and the process is very prone to data snooping. To exacerbate matters, the LPPL model is a nonlinear model with a total of seven parameters, and it can be fitted to almost any time series in some extend. The problem can be diminished with strict filtering, but not completely eliminated. The risk is naturally highest with short time windows and a limited number of modelled oscillations where the log-periodic oscillations may have been created by pure chance. In practice, a visual examination of the LPPL plot over the price data gives instant feedback on how well the LPPL captures the price gyrations and does the model describe price trajectory accurately enough. In the professional industry use, we, of course, require more automated and less subjective judgement.

[15] Quote retrieved from the internet: https://bit.ly/2W7Jdj4

Sornette et al. (2015) have addressed this issue by introducing *DS LPPLS Confidence* -indicator, which is the fraction of fitting windows satisfying the LPPL filtering conditions. Calibrating multiple shrinking time windows on the same data and getting the confidence indicator lessens the data snooping problem. A large confidence value indicates a more reliable LPPL signal that is present at most of the time windows. A small value indicates that the LPPL signal can be found from only a small number of time windows, and it is thus less trustworthy.

Calibrating shrinking time windows is not attainable in the theses due to extensive computing power required with the current heuristic approximation.

4.11 Using GPU to increase computing power

Calibrating the LPPL model is computationally expensive. Using heuristics algorithms to find the best-fit parameters is often limited by computing power, and we must weigh the estimation error against the computation speed.

On a personal computer, computational power may be significantly improved by using a graphical processor (GPU) to do computing instead of CPU (the main processor of a computer). Modern GPUs have large and fast working memory and a large number of computing cores. While a modern CPU typically has four to eight cores, have GPU hundreds of smaller cores. This opens a cost-effective way to utilize the GPU to have access to high-performance computing.

GPU accelerating has become an option on many high-level programming tools in recent years. Even the free R software environment used in the theses have now access to CUDA libraries[16] from graphics cards manufacturer Nvidia as well as some third-party and open-source GPU computing libraries. These libraries should provide easy programming access to GPU accelerated parallel computing.

Liberatore (2010) has studied the subject and introduces a new method to find initial solutions and use a parallel Levenberg-Marquardt algorithm (LMA) to fit the LPPL model.

[16] *CUDA is a parallel computing platform and programming model developed by Nvidia for general computing* (developer.nvidia.com).

He had speed increase by more than a factor of four on a general-purpose multi-core computer. This process can speed up substantially more on a GPU architecture. We leave this subject to further studies for now.

5 Analysis and results

In the following, we analyse the LPPL calibrations before the five most significant drawdowns in the OMX Helsinki index and ten most significant drawdowns on selected individual stocks. The data for the analysis is retrieved from the DataStream-database. The data is adjusted price time series to make it comparable trough time.

5.1 LPPL calibrations to the OMX Helsinki drawdowns

The five largest drawdowns in the OMX Helsinki index are fitted with the LPPL model.

Table 4: The five largest Drawdowns of the OMX Helsinki index (OMXH). The onset is a date after the peak value. The bottom is a date of the bottom value. The depth is the percentage change of the drawdown. Date format YYYY-MM-DD.

The largest drawdowns of the OMX Helsinki index[17]			
Incident	**Onset**	**Bottom**	**Depth**
1. After the Tech Bubble	2000-05-03	2003-03-10	-74.35 %
2. 90's Recession	1989-04-18	1992-09-07	-73.07 %
3. Financial Crisis	2007-11-08	2009-03-06	-67.52 %
4. Euro/ Debt Crisis	2011-02-10	2012-06-04	-39.31 %
5. Russian Crisis	1998-07-22	1998-10-08	-38.74 %

Following pages demonstrate the LPPL calibrations before the five largest OMX Helsinki index drawdowns. They are presented in the drawdown's size order [2000, 1989, 2007, 2011, 1998].

[17] Data: Adjusted time-series of the OMX Helsinki price -index values from the DataStream (1987-01-02 to 2019-12-30).

5.1.1 LPPL Calibration 2000 'Tech Bubble'

The index value trajectory of OMX Helsinki before the year 2000 is a prime example of super-exponential price growth. Our critical time estimates are very accurate, taking consideration that the index value growth path plateaued about two months or 42 trading days before the actual peak. Unfortunately, the residual stationarity tests fail in these calibrations. Only the oldest calibration, 130 trading days before the peak, is stationary (p = 0.05). There are probably multiple time windows where the residuals prove to be stationary, but we could not find them with our partly manual calibration system. We analyse the OMXH 2000 results more deeply on chapter 'Summary of the results on OMX Helsinki'.

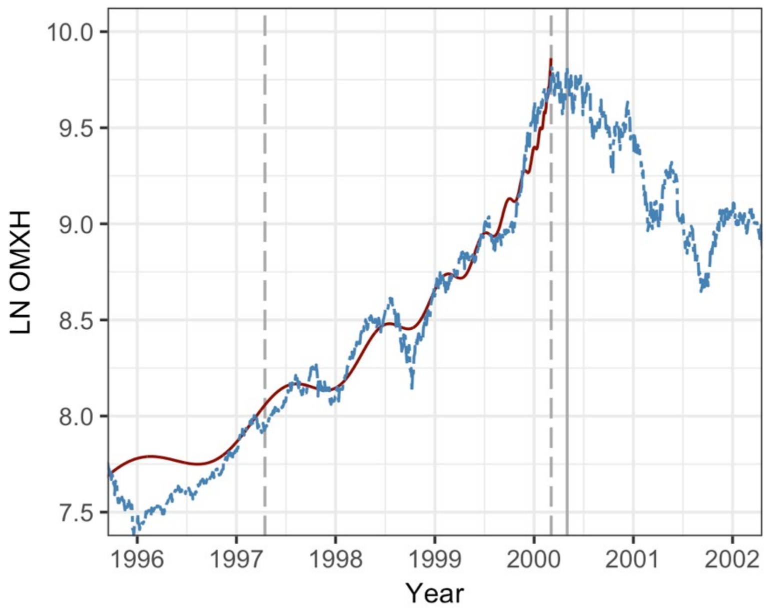

Figure 13: An example fit of the LPPL model to the OMX Helsinki 2000. Parameters: A = 9.983 (actual peak value 9.816), B = -1.270, C = 0.059, t.crit = 2000.173 (actual peak time 2000.333), β = 0.418, ω = 13.872 and φ = 5.102. Date format year decimal. Vertical dashed lines indicate the calibration time window [1997.29 - 2000.17]. The solid vertical line indicates the actual peak time.

Table 5: LPPL calibration to the OMX Helsinki 2000, local peak at 2000.333. 't.end' is the end of the calibration time window, trading days before the local peak [-42 = two months, -65 = three months, -130 = six months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-42	750	9.983	-1.270	2000.173	0.059	0.418	13.872	5.102	0.00728	0.964
-65	750	9.885	-1.233	2000.084	-0.057	0.403	15.766	1.581	0.00693	0.959
-130	750	9.623	-0.788	2000.437	-0.048	0.638	16.009	4.676	0.00491	0.959

5.1.2 LPPL Calibration 1989 '90's Recession'

We could not model the OMX Helsinki drawdown in 1989 with the LPPL model. The figure 14 shows the LPPL model calibration a half year before the peak. The calibration time window starts from the beginning of the OMXH time series available. At the example fit, the number of modelled oscillations is only 2.03 and not sufficient to confirm true log-periodic oscillations. The residuals are not stationary.

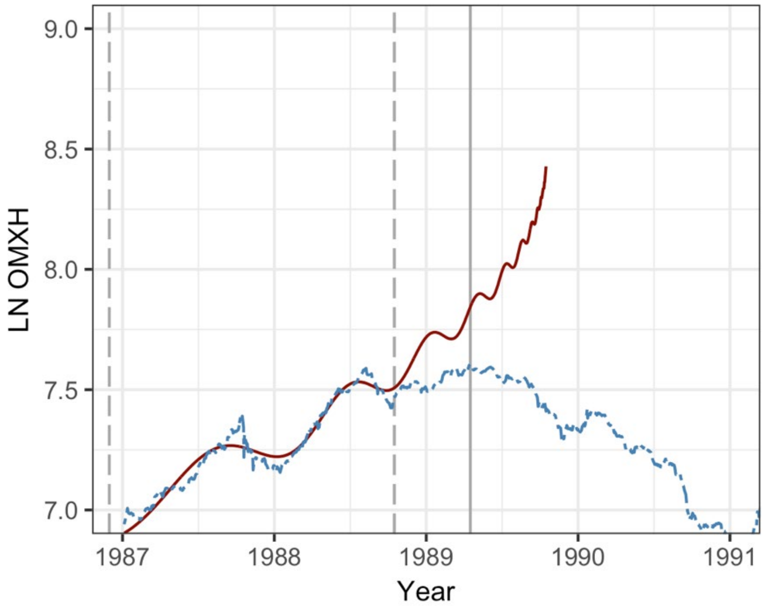

Figure 14: An example fit of the LPPL model to the OMX Helsinki 1989. Parameters: A = 8.464 (actual peak value 7.606), B = -0.894, C = 0.062, t.crit = 1989.79 (actual peak time 1989.293), β = 0.480, ω = 12.058 and φ = 3.166. Date format year decimal. Vertical dashed lines indicate the calibration time window [1986.91 - 1988.79]. The solid vertical line indicates the actual peak time.

Table 6: LPPL calibration to the OMX Helsinki 1989, local peak at 1989.293. 't.end' is the end of the calibration time window, trading days before the local peak [-130 = six months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-130	488	8.464	-0.894	1989.790	0.062	0.480	12.058	3.166	0.00138	0.955

5.1.3 LPPL Calibration 2007 'Financial Crisis'

We could not model the OMX Helsinki drawdown in 2007 with the LPPL model. At the example fit, the number of modelled oscillations is 2.52, but the residuals are not stationary, and the Lomb-Scargle spectral analysis is not matched. The calibrated critical time is by accident, precisely a year after the actual peak date.

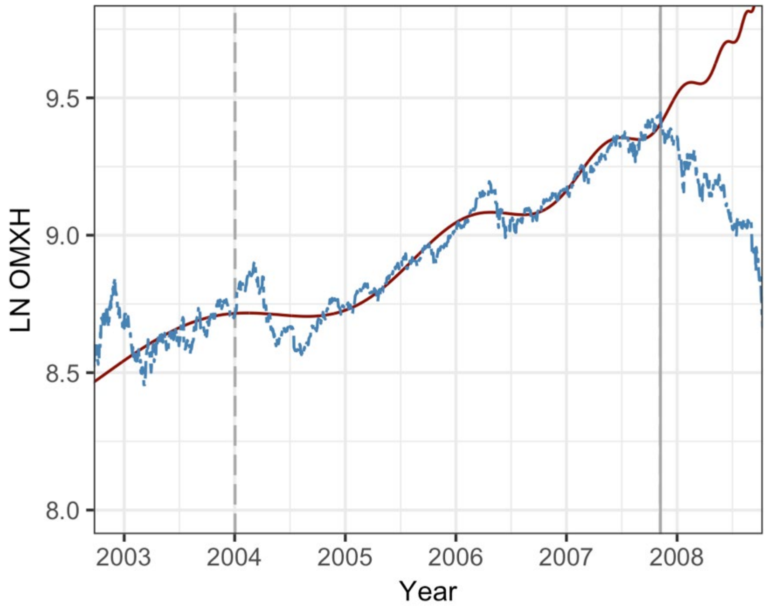

Figure 15: An example fit of the LPPL model to the OMX Helsinki 2007. Parameters: A = 10.131 (actual peak value 9.446), B = -0.697, C = 0.041, t.crit = 2008.849 (actual peak time 2007.849), β = 0.478, ω = 10.015 and φ = 2.357. Date format year decimal. Vertical dashed lines indicate the calibration time window [2004 - 2007.85]. The solid vertical line indicates the actual peak time.

Table 7: LPPL calibration to the OMX Helsinki 2007, local peak at 2007.849. 't.end' is the end of the calibration time window, trading days before the local peak. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	**t.window**	**A**	**B**	**t.crit**	**C**	**β**	**ω**	**φ**	**MSE**	**R^2**
0	1000	10.131	-0.697	2008.849	0.041	0.478	10.015	2.357	0.00272	0.949

5.1.4 LPPL Calibration 2011 'Euro/ Debt Crises'

We could not model the OMX Helsinki drawdown in 2011 with the LPPL model. At the example fit, the number of modelled oscillations is only 1.35, not sufficient to confirm true log-periodic oscillations. The residuals are stationary (p = 0.01), but the coefficient of determination is relatively low compared to more successful calibrations.

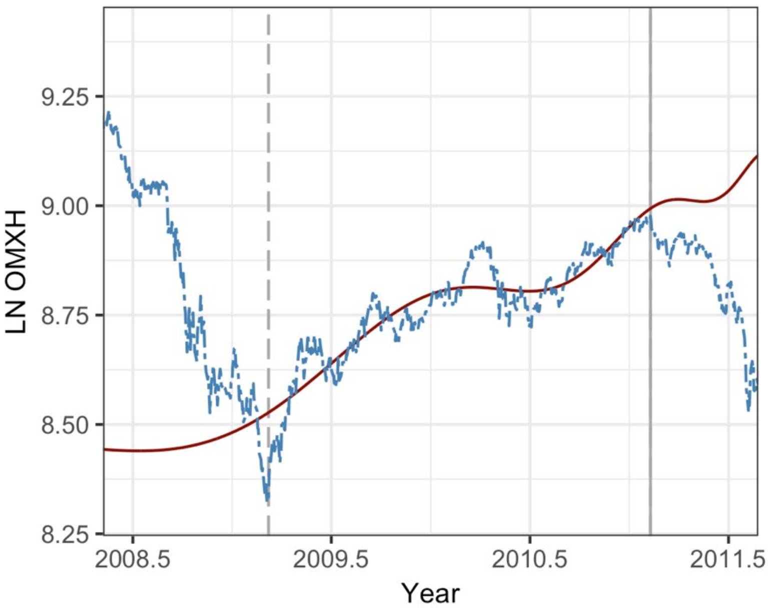

Figure 16: An example fit of the LPPL model to the OMX Helsinki 2011. Parameters: A = 9.271 (actual peak value 8.976), B = -0.314, C = 0.036, t.crit = 2012.093 (actual peak time 2011.107), β = 0.722, ω = 7.851 and φ = 0.477. Date format year decimal. Vertical dashed lines indicate the calibration time window [2009.18 - 2011.11]. The solid vertical line indicates the actual peak time.

Table 8: LPPL calibration to the OMX Helsinki 2011, local peak at 2011.107. 't.end' is the end of the calibration time window, trading days before the local peak. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
0	500	9.271	-0.314	2012.093	0.036	0.722	7.851	0.477	0.00200	0.862

5.1.5 LPPL Calibration 1998 'Russian Crises'

We could not model the OMX Helsinki drawdown in 1998 with the LPPL model. At the example fit, the number of modelled oscillations is only 1.75 and not sufficient to confirm true log-periodic oscillations. The residuals are stationary (p = 0.05). Further analysis shows that the price trajectory is part of the bubble regime ending at the 2000 price crash, as shown in chapter *LPPL Calibration 2000 'Tech Bubble'*.

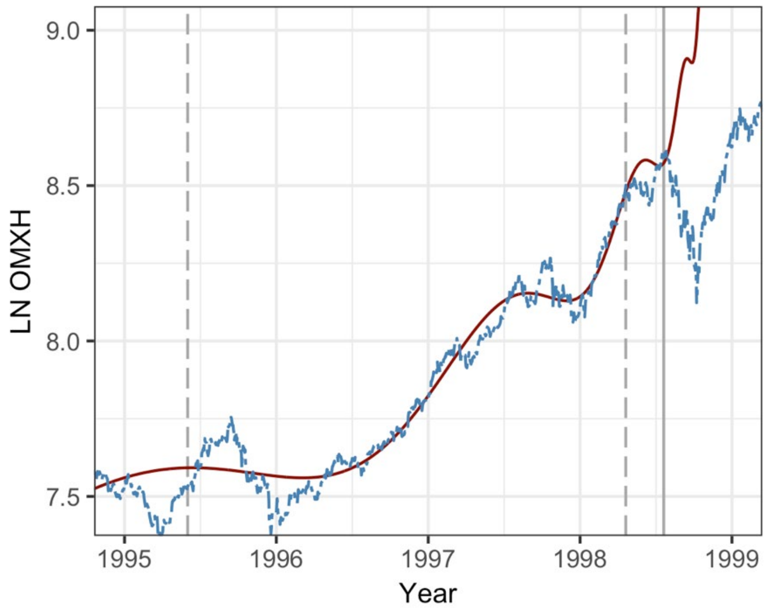

Figure 17: An example fit of the LPPL model to the OMX Helsinki 1998. Parameters: A = 9.958 (actual peak value 8.613), B = -1.794, C = -0.103, t.crit = 1998.852 (actual peak time 1998.551), β = 0.259, ω = 6.001 and φ = 1.267. Date format year decimal. Vertical dashed lines indicate the calibration time window [1995.42 - 1998.3]. The solid vertical line indicates the actual peak time.

Table 9: LPPL calibration to the OMX Helsinki 1998, local peak at 1998.551. 't.end' is the end of the calibration time window, trading days before the local peak [-65 = three months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-65	750	9.958	-1.794	1998.852	-0.103	0.259	6.001	1.267	0.00311	0.960

5.2 Summary of the results on the OMX Helsinki

We did not find solid results overall with the LPPL model on the five most significant drawdowns on OMX Helsinki index. The table 10 shows the fundamental values to analyse possible log-periodic signals. The best LPPL model results are found on the calibrations to the 2000 'Tech Bubble'. The calibrations on the other drawdowns are not successful: there are a too low number of modelled oscillations or Lomb-Scargle spectral analysis and residual analysis are poor. All the filtering rules are not met with any of the calibrations.

Table 10: The best LPPL OMXH calibrations and filtering values. Values β and ω are from the calibrated fit. R^2 is the coefficient of determination. Residuals indicate, if the they are stationary (p-levels: *** = 0.01, ** = 0.05, *=mixed results, - not stationary). N.osc is the number of modelled oscillations (>2.5 filtering). Damp. is the damping factor (>1 filtering). Error is the single largest relative error (< 0.2 filtering).

				Filtering rules				Lomb. analysis	
Calibration [timew]	**β**	**ω**	**R^2**	**Residuals**	**N.osc**	**Damp.**	**Error**	**ω.fit**	**ω.lomb**
OMXH 2000 [-42, 750]	0.418	13.872	0.964	*	17.44	0.65	0.041	13.87	14.02
OMXH 2000 [-65, 750]	0.403	15.766	0.959	-	20.12	0.55	0.040	15.77	13.23
OMXH 2000 [-130, 750]	0.638	16.009	0.959	**	4.47	0.65	0.042	16.01	14.35
OMXH 1989	0.480	12.058	0.955	-	2.03	0.57	0.011	12.06	12.37
OMXH 2007	0.478	10.015	0.949	-	2.52	0.82	0.017	10.02	19.92
OMXH 2011	0.722	7.851	0.862	***	1.35	0.80	0.017	7.85	5.81
OMXH 1998	0.259	6.001	0.960	**	1.75	0.75	0.026	6.00	6.88

Following is a discussion of the calibrations to the 2000 'Tech Bubble'.

The super-exponential growth of the index values is clearly evident before the bubble burst in early 2000. The low calibrated beta values capture this. The Lomb-Scargle spectral analysis shows with very high statistical significance that the dominant log-periodic frequency corresponds to the calibrated omega values (the first 2000 calibration). This can be seen as evidence that the index value data carries log-periodic oscillations.

Unfortunately, the residuals are stationary only on one calibration (p = 0.05). All three 2000 calibrations have a strong coefficient of determination values indicating that the model explains very well the variability of the OMXH index.

There might very well be multiple time windows before the 2000 price crash where all the filtering rules are met. Unfortunately, with the current calibration method, we are unable to test a large number of time windows due to the computing power required.

5.2.1 The accuracy of the OMXH critical time estimates

In the table 11 we compare estimated critical date to the actual peak date. Difference to actual peak date value is presented as trading days as well as the standard date formats. The number of trading days in a year used here is 250.

The closest estimate in the raw data is the oldest calibration, taken about a half year before the peak. The actual peak time, as the end of the super-exponential growth, is somewhat arbitrary. The index growth plateaued about two months or 42 trading days before the actual peak value at 2000.333. Taking the end of the super-exponential growth into consideration, the latest critical time estimate is very accurate.

Table 11: The accuracy of the critical time estimates. Differences of estimated critical time to the actual is presented in trading days (250 per year) and calendar days. Date formats are year decimal and YYYY-MM-DD.

Calibration [timew]	**Actual critical time**	**Est. t.crit**	**Diff. in trading days**	**Actual date**	**Est. Date**	**Diff in calendar days**
OMXH 2000 [-42, 750]	2000.333	2000.173	-40.0	2000-05-02	2000-04-03	-59
OMXH 2000 [-65, 750]	2000.333	2000.084	-62.3	2000-05-02	2000-01-31	-92
OMXH 2000 [-130, 750]	2000.333	2000.437	26.0	2000-05-02	2000-06-08	37
Average	2000.333	2000.231	-25.4			-38

5.2.2 The accuracy of the OMXH value estimates

How accurate are the index value estimates at a critical time? First, we have to convert logarithmic values back to the normal index values by taking *e* to the power of the LPPL A-parameter. OMX Helsinki index peak value was 9.816 (ln) or 18 331[18] at 2000-05-02.

Calibrated values for the A-parameter are 9.983, 9.885 and 9.623 from the latest to the oldest calibration and their regular rounded index value counterparts are 21 655, 19 634 and 15 108 respectively. The differences of the estimates to the actual are 18.13%, 7.11% and -17.58%.

The average of the A-parameter is 9.830, and the corresponding index value is 18 589, which differs only 1.41% from the actual. We cannot put too much emphasis on the statistical

[18] There are slight differences due to rounding. Actual values are from the database.

measures as the number of observations is too low. It is still reassuring to notice that the model and its calibrations create reasonable accurate estimates on average. We discuss more on use of the estimate means and confidence intervals under chapter 6 'Discussion of the findings'.

5.3 LPPL calibrations to individual stocks

We calibrate the LPPL-model to individual stocks to test if we find the LPPL signatures before the most significant stock crashes.

We select the ten most significant stock drawdowns from the Helsinki Stock Exchange main-list satisfying the following criteria: (i) A single price crash per stock. (ii) A stock has sufficient price history before the drawdown that can be used to calibrate the LPPL-model. There are a number of large stock drawdowns during the dot-com-bubble era, but they are relatively soon after IPO's and thus lack sufficient trading history. Stocks are sorted using a custom R-code and DataStream database.

Table 12: The largest individual stock drawdowns of the Helsinki Stock Exchange main-list meeting the criteria. The onset is a date after the peak value. The bottom is a date of the bottom value. The depth is the percentage change of the drawdown. Date format YYYY-MM-DD.

Finnish stocks with the largest drawdowns[19]			
Stock	**Onset**	**Bottom**	**Depth**
Valoe	2000-02-03	2017-07-28	-99,93 %
Bittium	2000-02-09	2008-12-19	-99,50 %
Componenta	2007-06-06	2016-12-05	-99,27 %
Uutechnic Group	1997-10-23	2015-01-28	-98,80 %
Talvivaara Mng.Co.	2011-02-10	2014-11-03	-98,47 %
Metsa Board B	2000-01-04	2009-03-09	-97,94 %
Nokia	2000-05-03	2012-07-18	-97,89 %
Outokumpu A	2008-05-20	2015-09-30	-97,35 %
Glaston	2000-03-24	2013-03-20	-97,12 %
Tulikivi A	2006-04-25	2015-04-30	-97,04 %

[19] Data: Adjusted price timeseries from the DataStream (start of the data to 2017-11-20).

5.3.1 Valoe

The LPPL model can be fitted to the Valoe price data before the peak at 2000.087. Visually inspecting the graph, the LPPL fit captures the fist five price oscillations nicely as an indication of a strong positive bubble regime forming. The residuals are stationary (DF and PP –tests p = 0.01) at the calibrations and the high coefficients of determinations indicate a good overall fit.

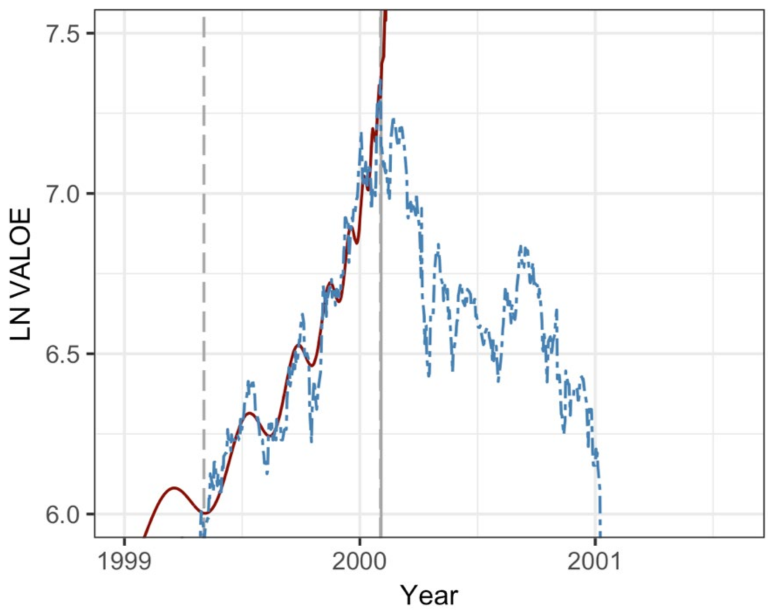

Figure 18: The best fit of the LPPL model to Valoe. Parameters: A = 8.716 (actual peak value 7.353), B = -2.774, C = 0.096, t.crit = 2000.126 (actual peak time 2000.087), β = 0.215, ω = 14.582 and φ = 0.882. Date format year decimal. Vertical dashed lines indicate the calibration time window [1999.34 - 2000.09]. The solid vertical line indicates the actual peak time.

Table 13: LPPL calibration to Valoe, local peak at 2000.087. 't.end' is the end of the calibration time window, trading days before the local peak [-42 = two months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-42	260	7.086	-1.316	2000.167	0.053	0.893	13.000	4.735	0.01050	0.916
-0	195	8.716	-2.774	2000.126	0.096	0.215	14.582	0.882	0.00633	0.943

5.3.2 Bittium

We could not find a statistically significant fit of the LPPL model to the Bittium price data before the peak at 2000.104. There are a limited number of modelled price oscillations (2.66), and the residual tests fail (DF-test p = 0.08 and PP-test p = 0.08)

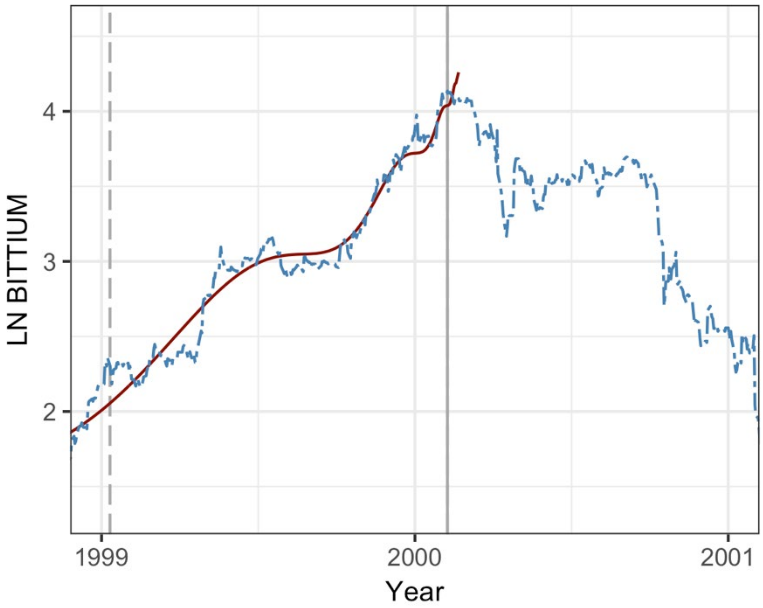

Figure 19: The best fit of the LPPL model to the Bittium. Parameters: A = 4.323 (actual peak value 4.135), B = -1.947, C = 0.220, t.crit = 2000.142 (actual peak time 2000.104), β = 0.591, ω = 4.948 and φ = 1.982. Date format year decimal. Vertical dashed lines indicate the calibration time window [1999.03 - 2000.1]. The solid vertical line indicates the actual peak time.

Table 14: LPPL calibration to Bittium, local peak at 2000.104. 't.end' is the end of the calibration time window, trading days before the local peak. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
0	280	4.323	-1.947	2000.142	0.220	0.591	4.948	1.982	0.01132	0.957

5.3.3 Componenta

We could not find statistically significant fit of the LPPL model to the Componenta price data before the peak at 2007.425. The number of modelled oscillations is too low and the residuals are not stationary.

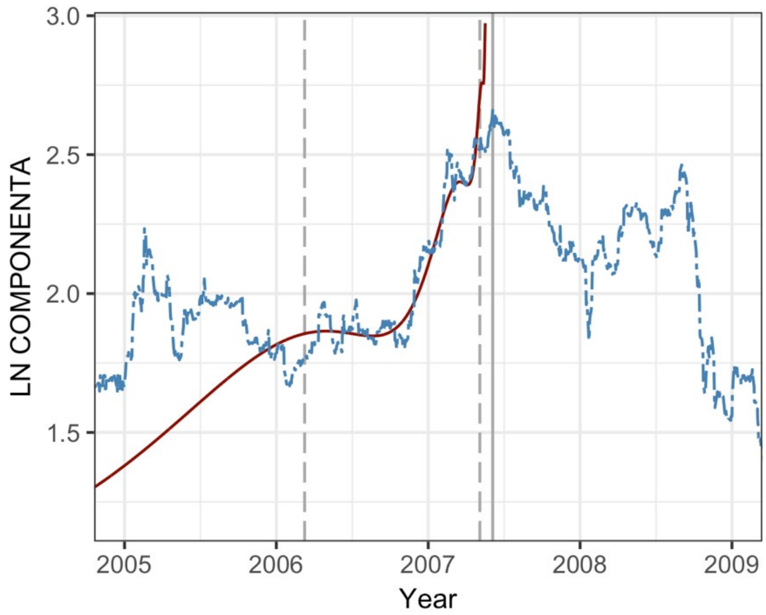

Figure 20: An example fit of the LPPL model to Componenta. Parameters: A = 3.458 (actual peak value 2.66), B = -1.656, C = -0.131, t.crit = 2007.382 (actual peak time 2007.425), β = 0.229, ω = 3.494 and φ = 2.055. Date format year decimal. Vertical dashed lines indicate the calibration time window [2006.19 - 2007.34]. The solid vertical line indicates the actual peak time.

Table 15: LPPL calibration to Componenta, local peak at 2007.425. 't.end' is the end of the calibration time window, trading days before the local peak [-22 = one month]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-22	300	3.458	-1.656	2007.382	-0.131	0.229	3.494	2.055	0.00355	0.942

5.3.4 Uutechnic Group

The LPPL model can be fitted to the Uutechnic Group price data before the peak at 1997.784. The number of modelled oscillations is 2.7, and the residuals are stationary (DF and PP – tests p = 0.01). High coefficient of determination indicates a very good overall fit of the model.

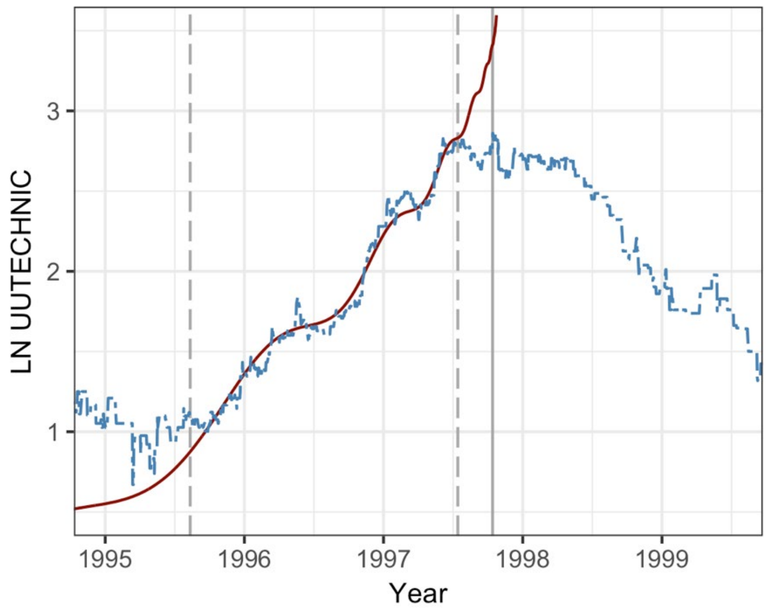

Figure 21: The best fit of the LPPL model to Uutechnic. Parameters: A = 3.609 (actual peak value 2.860) B = -1.638, C = -0.093, t.crit = 1997.811 (actual peak time 1997.784), β = 0.593, ω = 8.174 and φ = 5.484. Date format year decimal. Vertical dashed lines indicate the calibration time window [1995.61 - 1997.53]. The solid vertical line indicates the actual peak time.

Table 16: LPPL calibration to Uutechnic, local peak at 1997.784. 't.end' is the end of the calibration time window, trading days before the local peak [-22 = one month, -65 = three months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-65	500	3.609	-1.638	1997.811	-0.093	0.593	8.174	5.484	0.00500	0.982

5.3.5 Talvivaara Mng. Co.

The LPPL model can be fitted to the Talvivaara price data before the peak at 2011.107. Unfortunately, there is only limited number of modelled oscillations (1.96 and 1.97) on the calibrations meaning that the signal may have been created by pure chance. The residuals are stationary (DF and PP –tests p = 0.01).

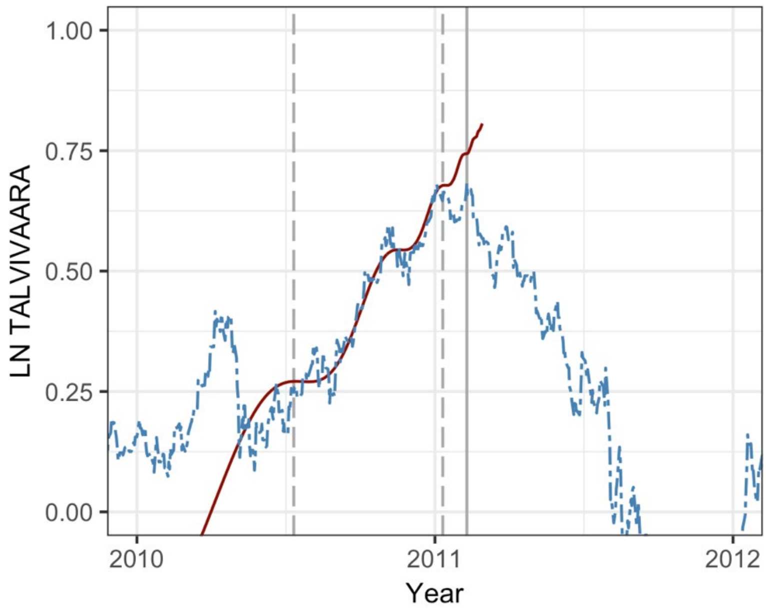

Figure 22: The best fit of the LPPL model to Talvivaara. Parameters: A = 0.807 (actual peak value 0.683), B = -0.855, C = 0.101, t.crit = 2011.159 (actual peak time 2011.107), β = 0.899, ω = 7.914 and φ = 2.536. Date format year decimal. Vertical dashed lines indicate the calibration time window [2010.53 - 2011.03]. The solid vertical line indicates the actual peak time.

Table 17: LPPL calibration to Talvivaara, local peak at 2011.107. 't.end' is the end of the calibration time window, trading days before the local peak [-22 = one month]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-22	151	0.745	-0.796	2011.112	0.079	0.893	6.001	1.356	0.00102	0.951
-22	130	0.807	-0.855	2011.159	0.101	0.899	7.914	2.536	0.00081	0.955

5.3.6 Metsa Board B

We could not fit the LPPL model to the Metsa Board B price data before the peak. The LPPL calibration fails on multiple time windows: the number of modelled oscillations is too low and the residuals are not stationary. At the example fit the residuals are not stationary, and the number of oscillations is only 1.31.

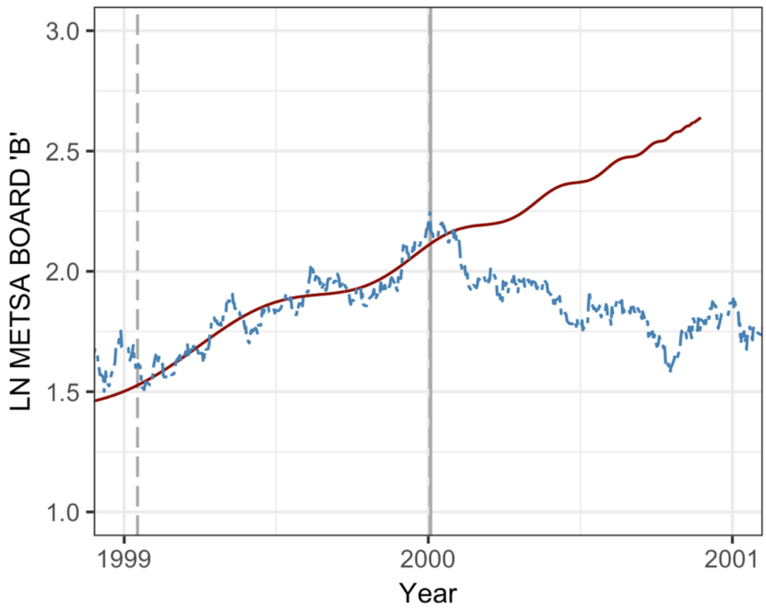

Figure 23: An example fit of Metsa Board B. Parameters: A = 2.641 (actual peak value 2.244), B = -0.606, C = -0.037, t.crit = 2000.898 (actual peak time 2000.005), β = 0.893, ω = 11.267 and φ = 5.393. Date format year decimal. Vertical dashed lines indicate the calibration time window [1999.04 - 2000.01]. The solid vertical line indicates the actual peak time.

Table 18: LPPL calibration to Metsa Board B, local peak at 2000.005. 't.end' is the end of the calibration time window, trading days before the local peak. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
0	250	2.641	-0.606	2000.898	-0.037	0.893	11.267	5.393	0.00255	0.897

5.3.7 Nokia

The LPPL model can be fitted to the Nokia price data before the peak at 2000.333. The LPPL calibration is successful on multiple time windows. The residuals are stationary (DF and PP –tests p = 0.01) on the first two calibrations and p = 0.05 level on the last. High coefficient of determination indicates a very good overall fit of the model.

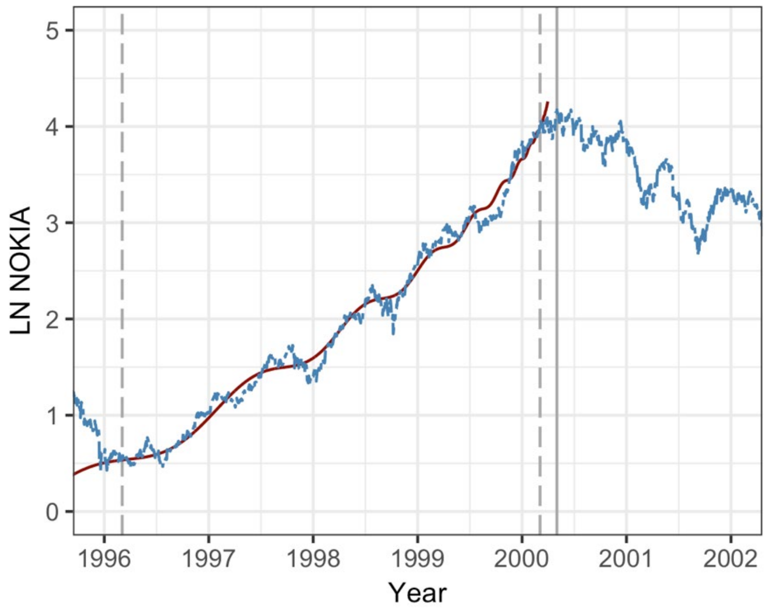

Figure 24: An example fit of the LPPL model to Nokia. Parameters: A = 4.315 (actual peak value 4.174), B = -1.576, C = 0.059, t.crit = 2000.251 (actual peak time 2000.333), β = 0.628, ω = 13.499 and φ = 4.813. Date format year decimal. Vertical dashed lines indicate the calibration time window [1996.17 - 2000.17]. The solid vertical line indicates the actual peak time.

Table 19: LPPL calibrations to Nokia, local peak at 2000.333. 't.end' is the end of the calibration time window, trading days before the local peak [-42 = two months]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-42	1040	4.315	-1.576	2000.251	0.059	0.628	13.499	4.813	0.01112	0.988
-42	750	3.998	-1.344	2000.172	0.086	0.703	14.371	5.121	0.01102	0.982
-42	500	4.648	-1.842	2000.309	-0.089	0.510	5.970	3.333	0.01005	0.970

5.3.8 Outokumpu A

We could not find a perfect LPPL fit on any of the time windows tested on the Outokumpu A price data before the price downturn. On the example fit the residual test fails (DF-test p = 0.07 and PP-test p = 0.01).

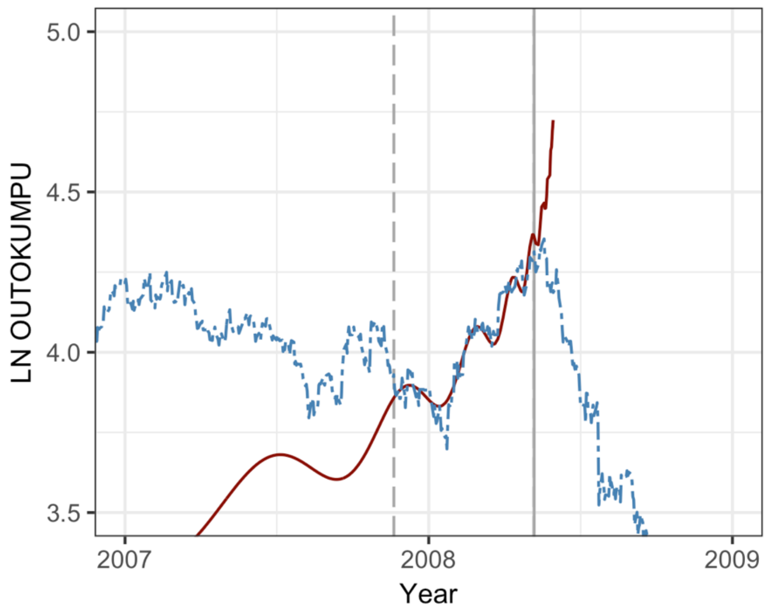

Figure 25: An example fit of the LPPL model to Outokumpu A. Parameters: A = 5.103 (actual peak value 4.313), B = -1.549, C = 0.096, t.crit = 2008.414 (actual peak time 2008.347), β = 0.259, ω = 9.877 and φ = 0.608. Date format year decimal. Vertical dashed lines indicate the calibration time window [2007.89 - 2008.35]. The solid vertical line indicates the actual peak time.

Table 20: LPPL calibration to Outokumpu A, local peak at 2008.38. 't.end' is the end of the calibration time window, trading days before the local peak. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
0	120	5.103	-1.549	2008.414	0.096	0.259	9.877	0.608	0.00207	0.921

5.3.9 Glaston

We could not find a significant fit of the LPPL model on the Glaston stock prior the price downturn at 2000.224. At the example fit the residuals are stationary (DF-test p = 0.01 and PP-test p = 0.03), but the number of calibrated oscillations is only 1.96.

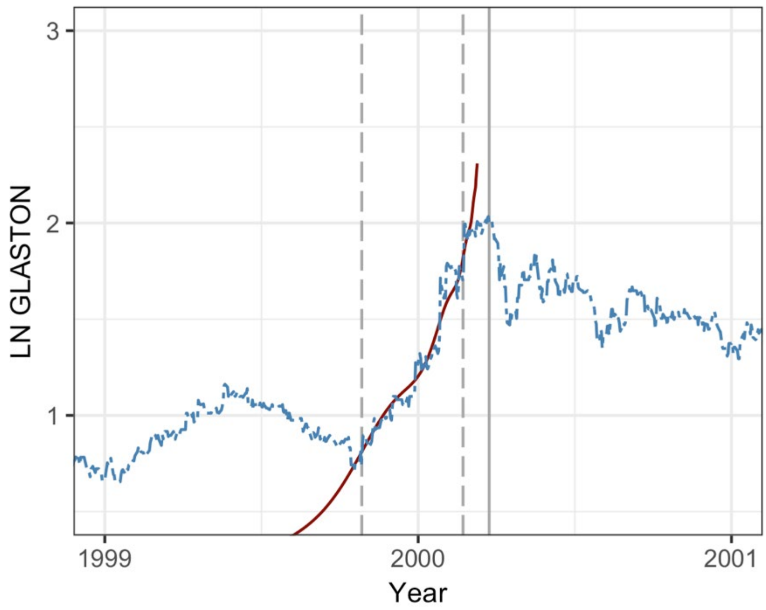

Figure 26: An example fit of the LPPL model to Glaston. Parameters: A = 2.499 (actual peak value 2.034), B = -2.646, C = -0.088, t.crit = 2000.191 (actual peak time 2000.224), β = 0.448, ω = 6.001 and φ = 4.237. Date format year decimal. Vertical dashed lines indicate the calibration time window [1999.82 - 2000.14]. The solid vertical line indicates the actual peak time.

Table 21: LPPL calibration to Glaston, local peak at 2007.425. 't.end' is the end of the calibration time window, trading days before the local peak [-22 = one month]. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
-22	84	2.499	-2.646	2000191	-0.088	0.448	6.001	4.237	0.00558	0.940

5.3.10 Tulikivi

The LPPL model can be fitted to the Tulikivi price data before the peak at 2006.31 on multiple time windows. The residuals are stationary ($p = 0.05$) at the first calibration (figure 27) and the second calibration ($p = 0.01$) with fewer oscillations (1.97). High coefficient of determination indicates a good overall fit of the model.

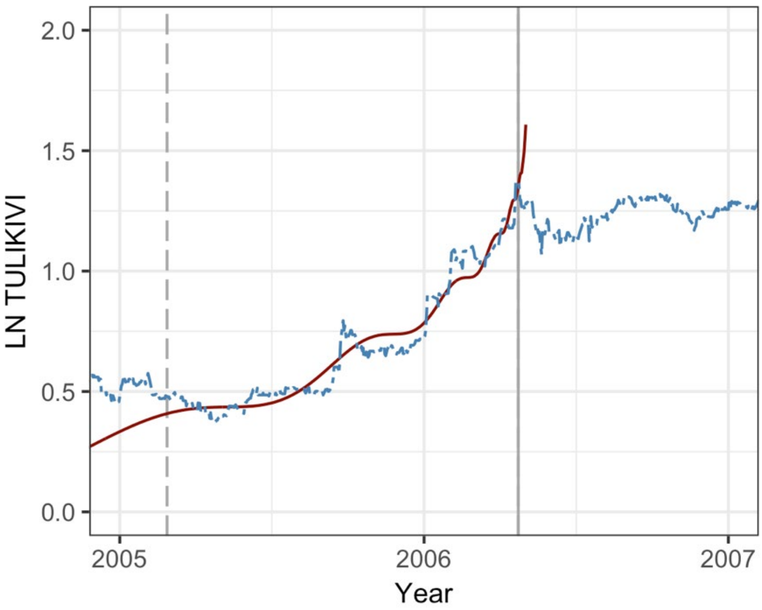

Figure 27: The best fit of the LPPL model to Tulikivi. Parameters: A = 1.786 (actual peak value 1.374), B = -1.361, C = -0.053, t.crit = 2006.336 (actual peak time 2006.31), β = 0.313, ω = 7.755 and φ = 1.766. Date format year decimal. Vertical dashed lines indicate the calibration time window [2005.16 - 2006.31]. The solid vertical line indicates the actual peak time.

Table 22: LPPL calibration to Tulikivi, local peak at 2006.31. 't.end' is the end of the calibration time window, trading days before the local peak. 't.window' is the length of the calibration time window in trading days. Calibrated LPPL parameters: [A, B, t.crit, C, β, ω, φ]. MSE is the mean squared error and R^2 is the coefficient of determination.

t.end	t.window	A	B	t.crit	C	β	ω	φ	MSE	R^2
0	300	1.786	-1.361	2006.336	-0.053	0.313	7.755	1.766	0.00321	0.949
0	126	1.445	-1.218	2006.394	-0.127	0.756	6.508	5.420	0.00131	0.968

5.4 Summary of the results on the selected stocks

The results on selected individual stocks are more robust than OMXH index results. Half of the stocks returned stationary residuals (p = 0.01) on the LPPL calibrations, and most of them (four) had a sufficient number of modelled oscillations on some of the calibrations.

The table 23 shows that it is not easy to meet all the filtering rules. Only Valoe and Uutechnic calibrations can be seen to meet all the requirements. It is possible that the other stocks could have met the criteria on some other time windows that have not been tested, but our current calibration system is not efficient enough to test all the time periods.

It can be seen from the data that successful calibrations are identified by stationary residuals companied with a high coefficient of determination and most likely with matching Lomb-Scargle spectral analysis. The Lomb-Scargle analysis can be distracted by multiple peak frequencies because of noisy data, and it is not simple to build exact filtering rule for that.

Table 23: The best LPPL stock calibrations and filtering values. Values beta and ω are from the calibrated fit. R^2 is the coefficient of determination. Residuals indicate, if the they are stationary (p-levels: *** = 0.01, ** = 0.05, *=mixed results, - not stationary). N.osc is the number of modelled oscillations (>2.5 filtering). Damp. is the damping factor (>1 filtering). Error is the single largest relative error (< 0.2 filtering).

				Filtering rules				Lomb. analysis	
Calibration	**β**	**ω**	**R^2**	**Residuals**	**N.osc**	**Damp.**	**Error**	**ω.fit**	**ω.lomb**
Valoe	0.215	14.582	0.943	***	7.03	0.43	0.038	14.58	14.58
Valoe	0.893	13.000	0.916	***	3.36	1.72	0.078	13.00	11.65
Bittium	0.591	4.948	0.957	-	2.66	1.06	0.139	4.95	5.59
Componenta	0.229	3.494	0.942	-	1.86	0.83	0.070	3.49	3.83
Uutechnic	0.593	8.174	0.982	***	2.69	1.28	0.128	8.17	9.15
Talvivaara	0.893	6.001	0.951	***	1.96	1.51	0.433	6.00	6.20
Talvivaara	0.899	7.914	0.955	***	1.97	0.97	0.270	7.91	8.13
Metsa Board B	0.893	11.267	0.897	-	1.31	1.29	0.073	11.23	25.83
Nokia	0.628	13.499	0.988	***	8.48	1.25	0.314	13.50	4.82
Nokia	0.703	14.371	0.982	***	30.66	0.76	0.192	14.37	4.48
Nokia	0.510	5.970	0.970	**	2.57	1.76	0.218	5.97	4.67
Outokumpu A	0.259	9.877	0.921	*	3.25	0.42	0.041	9.88	9.13
Glaston	0.448	6.001	0.940	**	1.96	2.23	0.099	6.00	6.22
Tulikivi	0.313	7.755	0.949	**	4.70	1.03	0.196	7.76	3.31
Tulikivi	0.756	6.508	0.968	***	1.97	0.82	0.073	6.51	6.62

The filtering rule for a single largest error may be a bit too much, at least for a limited number of calibrated time windows. It can filter out perfectly good fits based on a single price fluctuation on extreme market conditions. It may be valid for a multiple time windows calibrations where we seek to find out a number of positive signals over the total number of calibrations.

5.4.1 The accuracy of the stocks' critical time estimates

In the table 24 we compare estimated critical date of the individual stocks to the actual peak date. Stock calibrations with the stationary residuals (p = 0.01) are selected for the comparison. They all do not meet the filtering rules, but we examine the overall accuracy of the most successful calibrations. Difference to actual peak date value is presented as trading days as well as the standard date format. The number of trading days in a year used here is 250.

Table 24: The accuracy of the stocks' critical time estimates. Differences of estimated critical time to the actual is presented in trading days (250 per year) and calendar days. Date formats are year decimal and YYYY-MM-DD.

Calibration [timew]	Actual critical time	Est. t.crit	Diff. in trading days	Actual date	Est. Date	Diff. in calendar days
Valoe [-42, 260]	2000.087	2000.167	20.0	2000-02-02	2000-03-02	29
Valoe [0, 195]	2000.087	2000.126	9.7	2000-02-02	2000-02-16	14
Uutechnic [-65, 500]	1997.784	1997.811	6.7	1997-10-15	1997-10-24	9
Talvivaara [-22, 151]	2011.107	2011.112	1.3	2011-02-09	2011-02-10	1
Talvivaara [-22, 130]	2011.107	2011.159	13.0	2011-02-09	2011-02-28	19
Nokia [-42, 1040]	2000.333	2000.251	-20.5	2000-05-02	2000-04-01	-31
Nokia [-42, 750]	2000.333	2000.172	-40.3	2000-05-02	2000-03-03	-60
Tulikivi [0, 126]	2006.310	2006.394	21.0	2006-04-24	2006-05-24	30
Average			1.4			1.4

There are slight differences between a computed date and price values to actual values from the database due to rounding on year decimal and price logarithmic conversions. The actual calendar dates and prices on the tables are from the database.

We have to consider the super-exponential growth trajectory when we account the accuracy of the estimates. The LPPL model is expected to give an estimate, where the price growth singularity is achieved. In many times, the price process plateau to a horizontal price movement before the peak price is achieved. This is very well observed in the case of Nokia, where the super-exponential price growth ended about two months before the peak price.

The calibrations of Valoe and Uutechnic, which met all the filtering requirements, are among the most accurate with Talvivaara. The Talvivaara results were also strong, hindered only by the low number of modelled oscillations.

The averages of the differences are reported as well. They do not carry very much information as there is a low number of stocks, and some are presented by multiple calibrations. On our sample, differences happen to cancel each other out, and on average, the difference between the estimated critical date and actual is very low, only 1.4 calendar days.

5.4.2 The accuracy of the stocks' price estimates

From the practical point of view, we are more interested in the critical time estimate than the peak price estimate as the market price is available at the given moment. In the table 25 the LPPL model price estimates are compared to the actual peak values of the stocks. Stock calibrations with the stationary residuals (p = 0.01) are selected for the comparison.

Some of the estimates are extremely high, even when the critical time estimates are quite accurate as in cases of Valoe and Uutechnic. The LPPL model describes an explosive price process, and with some calibrations, unrealistic end prices are inevitable.

Table 25: The accuracy of the stocks' peak price estimates. Actual stock peak prices, estimated peak prices and percentage differences.

Calibration [timew]	**Actual peak price**	**LPPL estimate EXP(A)**	**Diff. %**
Valoe [-42, 260]	1560.68	1195.12	-23.42
Valoe [0, 195]	1560.68	6099.73	290.84
Uutechnic [-65, 500]	17.46	39.93	111.49
Talvivaara [-22, 151]	1.98	2.11	6.40
Talvivaara [-22, 130]	1.98	2.24	13.20
Nokia [-42, 1040]	64.95	74.81	15.19
Nokia [-42, 750]	64.95	54.49	-16.11
Tulikivi [0, 126]	3.95	4.24	7.36

6 Conclusions and further research

In the following chapters, we present a summary of the thesis, discuss the findings and limitations of the study. We also present topics for further research.

6.1 Research summary

The log-periodic power law (LPPL) is a relatively new tool to model stock market crashes. The first version of the model was introduced by Didier Sornette, Anders Johansen and Jean-Philippe Bouchaud in the late '90s (Sornette et al., 1996). The model is based on behavioural models of herding and imitation among traders, the theory of critical phenomena in complex systems from physics, the mean-field theory and the Ising model from statistical mechanics.

The LPPL-model can be used as a tool to predict the time of the forthcoming price crashes on various traded assets and market indices. This is accomplished by calibrating the LPPL model to the price data and use a set of filtering rules to examine if the LPPL model signature is present on the price data. If the log-periodic price oscillations are detected as the model predicts, we get an estimate of future critical time and price of the asset at that time. The critical time is the estimate of the moment when the price of an asset is expected to crash or change regime (i.e. end of the speculative bubble).

The founding fathers of the model, as well as other researchers, have used the model to successfully predict the price drawdowns on different markets, both ex-ante and ex-post. In the thesis, we seek to find out if the LPPL signatures can be found from the Finnish stock market before significant price drawdowns.

We formulate three research questions: (i) Do we find evidence that the LPPL signatures are present before the most significant stock market drawdowns in the Helsinki Stock Exchange? (ii) Do we find evidence that the LPPL signatures are present before the most significant individual stock crashes in the Helsinki Stock Exchange? (iii) If we find the LPPL model signatures: How accurate are the LPPL model estimates for the critical time and peak values?

We select the five most significant drawdowns of the OMX Helsinki stock index and ten most significant individual stock's drawdown for testing. These incidents are found from the DataStream-database using a custom R code.

Calibrating the LPPL model is a difficult task. The original model has a total of seven parameters, and four parameters specify the non-linear search space. Multiple computational methods have been suggested to the problem, and the area seems to be still evolving.

In the thesis, we design a custom program to calibrate the LPPL model. The program is a multi-phase genetic algorithm tool using the original LPPL methodology with later mean-reverting residuals specification (Lin et al., 2009). The R statistical computing environment is used as a programming tool. We also analyze the non-linear search space of the model graphically, with Matlab software, to illustrate the complexity of the minimization problem, and visually verify the results of the genetic algorithm tool.

The results of the calibrations are tested against the filtering rules, regular residual tests and the Lomb-Scargle spectral analysis to rule out spurious fits. Successful calibrations are analyzed for their estimate accuracy.

6.2 Discussion of the findings

In the following chapter, we discuss the findings on calibrations of the OMX Helsinki index and ten individual stocks from the Helsinki Stock Exchange.

6.2.1 OMX Helsinki calibrations

In order to answer our first research question, do we find the LPPL signatures before the most significant stock market drawdowns in the Helsinki Stock Exchange, we fit the LPPL model to the five biggest OMX Helsinki index drawdowns. They are in size order: 2000 'Tech Bubble', 90's 'Recession', 'Financial crisis' of 2007, 'Euro/ Debt crisis' of 2011 and 'Russian Crisis' of 1998.

Our success with the OMX Helsinki was only partial at best. We could not find a perfect fit with any of the drawdowns or time windows tested. The best LPPL model results are found on calibrations to the 2000 'Tech Bubble', however all the filtering rules are not met with any of the calibrations. The calibrations on the other drawdowns are not successful. They have too low number of modelled oscillations or residual analysis is failed.

The LPPL calibrations to the 2000 'Tech Bubble' gave satisfactory results in terms of accuracy. The estimates for the critical time varied from -40 days to +26 trading days compared to the actual. The latest calibration taken 42 days before the crash gave the -40 days result (i.e. imminent crash warning). The latest calibration is very accurate taking into consideration that the super exponential price growth died out about 42 trading days before the peak value. The estimates for the index value (parameter A) were also sufficient. Estimates varied from -17.58% to +18.13% compared to the actual, and the average estimate differs only 1.41% from the actual. The results of the calibrations were hindered only by poor results on the residual testing.

6.2.2 Use of statistical tools

The variance of the estimated parameters may seem substantial, but we have to consider the extreme market conditions that we are trying to predict. In general, we can use standard statistical tools to lower the estimation error by taking the average of multiple critical time estimates $(\overline{t_c})$ from different calibrations. We can also calculate the standard error of the mean ($\sigma_{\overline{t}_c}$), $\sigma_{\overline{t}_c} = \sigma/\sqrt{n}$, where σ is the standard deviation of the estimations and n is the number of estimations. Now we can calculate the mean estimate's confidence interval with a selected sigma (z) value: $\overline{t_c} \pm z\ \sigma_{\overline{t}_c}$.

To have theoretical bases for the previous mean and confidence interval, we must have confidence that we are estimating the same 'true' critical time over the different calibrations, i.e. calibration time windows cannot be too far away from each other (the end of the time window). Otherwise, the different calibrations may estimate different critical times, real or imaginary, and the mean of the estimates becomes meaningless.

In the thesis, we cannot put an emphasis on the statistical measures, because the number of estimates is so low due to computing power restrictions. However, it is still reassuring to notice that the model calibrations create reasonable accurate forecasts on average.

6.2.3 Results on OMXH are dominated by Nokia

One interesting observation comes from the fact that the OMX Helsinki index is a market capitalization-weighted index, and Nokia's weight on the index was about two-thirds in the tech bubble era. This leads to the assumption that Nokia and OMXH calibrations should have strong similarities during the tech bubble.

We have calibrated Nokia and OMXH index with the same time window ending -42 trading days before the drawdown and length of 750 trading days [1997.29 - 2000.17]. We found almost identical critical times 2000.172 for Nokia and 2000.173 for OMXH. Parameters quantifying the log-periodic oscillations are also very similar: the angular frequency ω for Nokia is 14.371 and 13.872 for OMXH, and the phase parameter φ is almost identical 5.121 (Nokia) and 5.102 (OMXH).

From the previous, we can conclude that the OMXH index log-periodic oscillations arise from the oscillations on the Nokia price data. The residuals of the Nokia calibration are stationary ($p = 0.01$), but the residuals of the OMXH calibration are not. The other stocks on the OMXH index interfere with the signal, and the perfect LPPL signature cannot be found in the index.

6.2.4 Calibrations on individual stocks

To answer our second research question, we seek the LPPL signatures before the crash of ten individual stocks. The stocks are selected based on the size of the crash; given that they have sufficient trading history before the crash to calibrate the model. The stocks are Valoe, Bittium, Componenta, Uutechic Group, Talvivaara, Metsa Board B, Nokia, Outokumpu A, Glaston and Tulikivi A.

Our results with individual stocks are stronger than with the OMXH. We had a very good fit and stationary residuals with five out of ten stocks [Valoe, Uutechnic, Talvivaara, Nokia and Tulikivi]. Two of the stocks, Valoe and Uutechnic, met all the filtering rules.

The estimates for critical time are quite sufficient, varying between -40.3 and +20 trading days in the whole data set. The average difference is only 1.4 trading days over all calibrations. The latest estimates of critical time for Valoe and Uutechic are quite accurate,

+9.7 days and +6.7 days. The estimates could be improved by taking multiple calibrations and calculating mean estimates as described in the previous chapter.

The estimates of the stock price at the critical time have substantial variation, from -23.42% to +290.84% compared to the actual. Valoe and Uutechnic, even though meeting all the filtering rules, are the worst. The LPPL model descries an explosive price process, and in some cases unrealistic end prices are inevitable.

6.3 Limitations of the study

The most significant limitation of the study is a limited number of time windows calibrated with the LPPL model. Calibrating the LPPL model is computationally expensive, and a large number of calibrations was not feasible with the current algorithm and programming tools used.

The results would have been more comprehensive if we could have calibrated a large number of varying time windows, and get statistical measures of the LPPL fits. These kinds of systems are used in professional settings with high-performance computing or super-computer architecture.

The computing time can be improved substantially by using a newer calibration scheme of the LPPL model (Filimonov and Sornette, 2013) and more robust and faster search methods that it makes possible (e.g. Nelder–Mead or Levenberg–Marquardt algorithm). In retrospect, we should have used the newer specification of the LPPL model to build calibration software. However, the programming implementation of the newer specification may be more demanding than the original model with the genetic algorithm tool.

6.4 Is the LPPL model valid?

The founding fathers of the model position the LPPL model as a universal tool to describe financial bubbles (Lin et al., 2009):

> *The above tests performed on these seven bubbles presented in Tables 6 and 7 suggest that our proposed volatility-confined LPPL model, first tested for the bubble*

and crash of October 1987, is not just fitting a single "story" but provide a consistent universal description of financial bubbles, namely a super-exponential acceleration of price decorated with log-periodic oscillations with mean-reverting residuals.

The possible predictive power of the LPPL model seems to be in contradiction with the efficient markets hypothesis, which is, in many ways, the cornerstone of modern finance thinking. We presume that there is a large number of researchers who do not see the LPPL as a valid model. However, some of the results, even in our small sample of stocks, are intriguing. For instance, Valoe stock displays evident log-periodic oscillations before the price crash of the year 2000. These oscillations can be visually seen from the plot as well as quantified perfectly with the Lomb-Scargle spectral analysis with very high statistical significance. We have the price oscillations, as the theory behind the model predicts, but can we trust them. Could these oscillations be created by the random walk? In the end, it may be a question of believing the model's postulates or not.

Stock market bubbles and crashes may not be as unpredictable as we used to think. A recent industry-level stock market bubble research, 'Bubbles for Fama' (Greenwood et al., 2019) studies stock performance after large price run-ups. In the paper, a bubble is defined as a 100% stock price increase of a US industry in terms of raw and net of market return over the past two years. They find 40 such cases in the US industry-level stock price data from 1926 to 2014. The research concentrates on 12 and 24 -months space after the price run-ups: What are average returns, what is the likelihood of a crash and are there some factors to help investors to forecast the crash beforehand. Greenwood et al. (2019) find at least some predictability on both crashes and returns after the price run-ups. Results are confirmed with the out-of-sample international data set as well.

Greenwood et al. (2019) conclude that *'the probability of a substantial crash after a 100% return is much higher than it is on average and, in fact, rises monotonically as past returns increase'*. This finding resonates with the LPPL model bubble definition. The LPPL model uses by definition a super-exponential price growth during a bubble regime, where the growth rate itself is affected. At the end of the bubble regime prices are expected to rise almost vertical and the system's instability grows to a finite-time singularity.

Greenwood et al. (2019) find that returns going forward after 100% price run-ups are hard to predict on average, as some of the industries crash and others keep performing at least for a time. This subject is not covered with the LPPL modelling. The LPPL model does not give any prediction of what happens after the critical time, other than the expected transition to a different market regime i.e. the end of the accelerated price process.

Greenwood et al. (2019) find that the industries that do crash, have some telling attributes during the bubble phase: '*They have higher volatility, stock issuance, especially rapid price increases, and disproportionate price rises among newer firms'.* Incorporating these telling parameters to a crash signalling system may increase the reliability of the crash alarms.

Based only on our small sample of stocks, we obviously cannot make a definitive judgement on the validity of the LPPL model. From the practical point of view, international studies indicate that the model could be a valuable signalling system for portfolio managers about the imminent crash, among other tools. However, for the active speculative trading, it may not offer accurate enough timing to be useful.

Given the statistical uncertainties of the LPPL model estimates, wrong signals are inevitable. The trustworthiness of the estimates could be improved by taking probabilistic view to the estimates. Improved accuracy and smaller confidence intervals of the estimates could be achieved by taking a large number of calibrations with different time windows. These systems require substantial computing power and are used in professional settings (like Financial Crisis Observatory, ETH Zürich). The accuracy of the system may also be increased combining the results with other common measures of increased market risk (see chapter ' Further research').

6.5 Short answers to the research questions

We present the summary of the results, in short, to answer the research questions:

(i) Do we find evidence that the LPPL signatures are present before the most significant stock market drawdowns in the Helsinki Stock Exchange? We found the LPPL signature only in one of the OMXH drawdowns out of five tested. The test results on the positive case are not perfect. Positive tests on the 2000 'Tech Bubble' seems to be a ghost of log-periodic signal found on the Nokia stock price.

(ii) Do we find evidence that the LPPL signatures are present before the most significant individual stock crashes in the Helsinki Stock Exchange? We have better success with the individual stocks than with the OMXH. We have a very good fit and stationary residuals with five out of ten stocks, and two of them meet all the filtering rules. Especially Valoe stock displays very prominent log-periodic oscillations, with very high statistical significance.

(iii) How accurate are the LPPL model estimates for the critical time and peak values? The LPPL model gives reasonably accurate estimates for the critical time on average. The estimates for the asset value on the critical time vary significantly in our small sample.

6.6 Further research

The LPPL model, in essence, quantifies the imitation and herding behaviour among traders[20]. The increase in the imitation tendency is measured with the increased frequency of modelled price oscillations towards the critical point. This insight gives an instant idea about further research: We have other measurements for the investors' groupthink and sentiment. Those could be tested in complementation with the LPPL model.

One intuitive indicator that can be tested simultaneously with the LPPL oscillations is the Google trends indicator. Previous studies have shown that the volume of certain Google search words may indicate a change in the market regime (Preis et al., 2013; Huang et al., 2019). The study that connects these two indicators of investors' groupthink and sentiment could carry out interesting results. The Google trends indicator has previously been used with the LPPL model as a tool to indicate traders' growing interest in gold and the LPPL calibration time window is selected based on the search volume index (see Geraskin and Fantazzini, 2013).

The other indicator that can also be tested for correlation with the LPPL results on the S&P 500 index is the VIX index. The VIX measures the implied 30-day volatility of the U.S. stock market, derived from the SPX call and put options (cboe.com). As the VIX is a direct

[20] Traders' imitation behaviour is quantified by the K constant in the Ising model (equation 2), carried through to the susceptibility function (equation 8) and transferred to the LPPL model by the approximation that K_c - K varies as $t_c - t$ near the critical point (equation 9 and 10).

measure of the market's expectation of the near future uncertainty, it could have interesting revelations combined with the LPPL model results.

It turns out that the VIX has been tested with the LPPL model as Zhou and Sornette (2006) have performed a factor analysis to examine the explanatory power of the different proposed fundamental factors compared to the LPPL model. These factors included interest rate, interest spread, historical volatility, implied volatility (VIX) and exchange rates. Further study could include the VIX from a different angle. There are also several other measures of market sentiment used by the practitioners that can be tested in complementation with the LPPL model.

The future research area, more of engineering calibre, can be building a computer rig designed to perform the LPPL calibrations efficiently. The parallel computing on the GPU or multi-core cluster computer architecture is perhaps the most efficient means for that. There are also new cloud computing options that can be used for advanced AI processing like Tensor processing unit (TPU) by Google introduced in 2016.

From the advanced programming point of view, we have an intuition that the LPPL model fitting is not very far away from visual recognition and neural-network computing may create surprising results with the model. There are new tools like Keras (keras.io) that make it easy to run the same code on normal CPU or the graphics processor (GPU), or even in Tensor processing unit on TensorFlow back-end. The R statistical software used in the thesis has an interface for Keras from 2017 (keras.rstudio.com). Christian Kindler has made preliminary tests for fitting the LPPL model with Keras[21].

Then there is, of course, the money: Is it possible to make systematically excess returns by trading based on LPPL signals? This is an under-examined subject. There are some studies with simulated data, but it would be an interesting subject for further study to examine if profitable trading strategies could be constructed using the real market data and the LPPL signals. Our intuition on the subject is that the estimation errors most likely hamper the LPPL traders' results and excess returns are difficult to realize. However, a more advanced option-based strategy designed to excel in special market conditions may be advantageous.

21 https://medium.com/@ch9.ki7/lppl-or-arbitrary-curve-fitting-with-tensorflow-and-keras-9a8123f504f2

References

Banzhaf, W., Olson, R. S., Tozier, W. A., Riolo, R. L. (2018). Genetic Programming Theory and Practice XV, [GPTP 2017, University of Michigan, Ann Arbor, USA, May 18-20, 2017]. *Genetic and Evolutionary Computation, Springer 2018*, ISBN 978-3-319-90511-2

Black, F. (1989). How we came up with the option formula. *Journal of Portfolio Management*, *15*(2), 4-8.

Bothmer, H. C. G. V., & Meister, C. (2003). Predicting critical crashes? A new restriction for the free variables. *Physica A: Statistical Mechanics and its Applications*, 320, 539-547.

Bree, D. S., & Joseph, N. L. (2010). Fitting the Log Periodic Power Law to financial crashes: a critical analysis. *arXiv preprint arXiv:1002.1010.*

Chang, G., & Feigenbaum, J. (2006). A Bayesian analysis of log-periodic precursors to financial crashes. *Quantitative Finance*, *6*(1), 15-36.

Chen, Y., Hashmi, M. U., Mathias, J., Bušić, A., & Meyn, S. (2018). Distributed control design for balancing the grid using flexible loads. In *Energy Markets and Responsive Grids* (pp. 383-411). Springer, New York, NY.

Derrida, B., De Seze, L., & Itzykson, C. (1983). Fractal structure of zeros in hierarchical models. *Journal of Statistical Physics*, *33*(3), 559-569.

Filimonov, V., & Sornette, D. (2013). A stable and robust calibration scheme of the log-periodic power law model. *Physica A: Statistical Mechanics and its Applications*, *392*(17), 3698-3707.

Geraskin, P., & Fantazzini, D. (2013). 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.

Greenwood, R., Shleifer, A., & You, Y. (2019). Bubbles for Fama. *Journal of Financial Economics*, 131(1), 20-43.

Gunasekaran, S., Kubizňák, D., & Mann, R. B. (2012). Extended phase space thermodynamics for charged and rotating black holes and Born-Infeld vacuum polarization. *Journal of High Energy Physics*, *2012*(11), 110.

Holland, J. H. (1975). An introductory analysis with applications to biology, control, and artificial intelligence. *Adaptation in Natural and Artificial Systems. First Edition, The University of Michigan, USA*.

Huang, M. Y., Rojas, R. R., & Convery, P. D. (2019). Forecasting stock market movements using Google Trend searches. *Empirical Economics*, 1-19.

Jiang, Z. Q., Zhou, W. X., Sornette, D., Woodard, R., Bastiaensen, K., & Cauwels, P. (2010). Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 Chinese stock market bubbles. *Journal of economic behavior & organization*, *74*(3), 149-162.

Johansen, A., Ledoit, O., & Sornette, D. (2000). Crashes as critical points. *International Journal of Theoretical and Applied Finance*, *3*(02), 219-255.

Johansen, A., & Sornette, D. (2002). Endogenous versus exogenous crashes in financial markets. *Available at SSRN 344980*.

Johansen, A. (2003). Characterization of large price variations in financial markets. *Physica A: Statistical Mechanics and its Applications*, *324*(1-2), 157-166.

Kamimura, A., & Ohira, T. (2019). *Group Chase and Escape: Fusion of Pursuits-Escapes and Collective Motions*. Springer Nature.

Liberatore, V. (2010). Computational LPPL fit to financial bubbles. *arXiv preprint arXiv:1003.2920*.

Lin, L., Ren, R. E., & Sornette, D. (2009). A Consistent Model of 'Explosive' Financial Bubbles with Mean-Reversing Residuals. SSRN Electronic Journal. 10.2139/ssrn.1596022.

Mitchell, M. (1999). *An introduction to genetic algorithms*. MIT press.

Press, W. H., & Rybicki, G. B. (1989). Fast algorithm for spectral analysis of unevenly sampled data. *The Astrophysical Journal*, *338*, 277-280.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, *3*, 1684.

Sornette, D., Johansen, A., & Bouchaud, J. P. (1996). Stock market crashes, precursors and replicas. *Journal de Physique I*, *6*(1), 167-175.

Sornette, D. (2004). *Why stock markets crash: critical events in complex financial systems*. Princeton University Press.

Sornette, D., Woodard, R., Yan, W., & Zhou, W. X. (2013). Clarifications to questions and criticisms on the Johansen–Ledoit–Sornette financial bubble model. *Physica A: Statistical Mechanics and its Applications*, *392*(19), 4417-4428.

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

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house.

Weatherall, J. O. (2013). *The physics of wall street: a brief history of predicting the unpredictable*. Houghton Mifflin Harcourt.

Zhou, W. X., & Sornette, D. (2006). Fundamental factors versus herding in the 2000–2005 US stock market and prediction. *Physica A: Statistical Mechanics and its Applications*, *360*(2), 459-482.

Zhou, W. X., & Sornette, D. (2009). A case study of speculative financial bubbles in the South African stock market 2003–2006. *Physica A: Statistical Mechanics and its Applications*, *388*(6), 869-880.