Modified Profile Likelihood Inference and Interval Forecast of the Burst of Financial Bubbles

Vladimir Filimonov[a,b], Guilherme Demos[a], Didier Sornette[a,c]

[a]*Dept. of Management, Technology and Economics, ETH Zürich, Zürich, Switzerland*
[b]*Department of Economics, Perm State University, Perm, Russia*
[c]*Swiss Finance Institute, c/o University of Geneva*

February 29, 2016

Abstract

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

Keywords: financial bubbles; crashes; inference; nuisance parameters; modified profile likelihood; nonlinear regression; JLS model; log-periodic power law; finite time singularity: nonlinear optimization.

Email addresses: `vfilimonov@ethz.ch` (Vladimir Filimonov), `gdemos@student.ethz.ch` (Guilherme Demos), `dsornette@ethz.ch` (Didier Sornette)

1. Introduction

Financial bubbles and their subsequent crashes provide arguably the most visible departures from well-functional efficient markets. There is an extensive literature (see e.g. the reviews of Kaizoji and Sornette (2010)[1], Jiang et al. (2010); Brunnermeier and Oehmke (2012); Xiong (2013)) on the causes of bubbles as well as the reasons for bubbles to be sustained over surprising long period of times. One of these views emphasises the role of herding behaviour on bubble inflation (Johansen and Sornette, 1999b). When imitation is sufficiently strong, a high demand for the asset pushes the price upwards, which itself, and somewhat paradoxically, increases the demand, propelling further the price upward, and so on, in self-fulfilling positive feedback loops. In such regimes, the market is mainly driven by sentiment and becomes detached from any underlying economic value. This process is intrinsically unsustainable and the mispricing ends at a critical time, either smoothly (with a correction phase) or abruptly (via a crash). The formulation of this hypothesis of collective herding behavior within rational expectations theory resulted in the so-called Log-Periodic Power-Law Singularity (LPPLS) model, which has been used for many successful ex-post and ex-ante predictions of bubble bursts (see e.g. a partial list in (Sornette et al., 2013) and a recent implementation for the Chinese bubble and its burst in 2015 (Sornette et al., 2015)).

Notwithstanding a number of improvements concerning the calibration of the LP-PLS model, including meta-search heuristics (Sornette and Zhou, 2006) and reformulation of the equations to reduce the number of nonlinear parameters (Filimonov and Sornette, 2013), the calibration of the LPPLS model remains a bottleneck towards achieving robust forecasts and a matter of contention (Brée et al., 2013; Sornette et al., 2013). In this context, the aim of the present paper is to present a fundamental revision of the calibration procedure of the LPPLS model. Specifically, we deviate from the traditional ordinary least squares (OLS) calibration that provides *point estimates* of parameters, which has been used since the introduction of the model in 1999 (Johansen et al., 1999, 2000). Instead, we employ a rigorous likelihood approach and, for the first time to the best of our knowledge, we provide *interval estimates* of the parameters, including the most important critical times of market regime changes.

We deliberately avoid dwelling on the derivation of the model and its foundations, and take it as given. We do not discuss supporting evidence and critiques of the model, nor address how to apply the LPPLS model to construct robust signals for extensive backtests or real-time ex-ante predictions. These questions require extensive analyses and are beyond the scope of the present manuscript. See (Johansen and Sornette, 2010; Jiang et al., 2010; Sornette et al., 2013, 2015; Zhang et al., 2015) for investigations in these directions.

The purpose of the present paper is methodological, and the main focus is on the statistical aspects of the theory and the corresponding mathematical derivations. One of the major advances of this paper is to formulate the calibration procedure so that the

---

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

critical time t_c is the major parameter of interest in the likelihood inference, while other model parameters are treated as so-called *nuisance parameters*. Of course, these other parameters are also intrinsic to the model but their existence contributes to the variance of the parameter of key interest. Such reformulation of the calibration procedure has its roots in an original idea proposed by Filimonov and Sornette (2013), which was however developed in a crude way and without the proper statistical methodology.

The problem of dealing with nuisance parameters and of quantifying their impact on the uncertainty of the parameter of interest is not new in Statistics. However, to our knowledge, it has not been elaborated before in quantitative finance. Frequentist and Bayesian statistical schools have different views on this problem. The main debate between the supporters of likelihood-based versus Bayesian approaches is whether one should maximize over nuisance parameters (such as in simple profile likelihood) or integrate them out. Both approaches have their pros and cons. In general, the method of profile likelihood is known to often provide biased estimations. However, use of the Bayesian (or integration) approach requires specification of the prior distribution of the parameters, which leads to an extra uncertainty in inference. Under certain conditions, when the full likelihood function has a complex structure, the two methods can lead to dramatically different estimations (Smith and Naylor, 1987; Berger et al., 1999).

We will base our approach on the so-called modified profile likelihood proposed by Barndorff-Nielsen (1983) as a higher-order approximation to either a marginal or conditional likelihood function. Being unable to calculate the modified profile likelihood exactly due to strong model nonlinearity, we will employ the approximation suggested by Severini (1998a), which is equivalent to the exact form up to errors of order $\mathcal{O}(n^{-1})$ for moderate deviations and of order $\mathcal{O}(n^{-1/2})$ in the large deviation sense, where n is the number of data points. The advantage of this method is that it takes the middle ground in the maximization-vs-integration debates: as shown by Severini (2007), the modified profile likelihood arises naturally from a non-Bayesian inference with an integrated likelihood and could even be considered as an approximation to a certain class of integrated likelihood functions. At the same time, it does not require specifying a prior density of the nuisance parameters, which makes it perfectly suitable in our case.

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

The paper is organized as follows. Section 2 presents the Log-Periodic Power Law Singularity model and discusses its structure and constraints. Section 3 presents the

Ordinary Least Squares (OLS) method that has been used until now as the standard calibration tool of the LPPLS model, in particular for the estimation of the critical time t_c of the end of the bubble. Section 4 introduces the Likelihood and Profile Likelihood approaches. Section 5 presents the general concept of the modified profile likelihood and provides a very useful approximated expression for it. Parameter estimation uncertainties and the corresponding likelihood intervals are then derived. Section 6 applies the modified likelihood profile to estimate confidence intervals of the nuisance parameters m and ω as well as the damping variable. Section 7 presents the method of aggregation of the calibrations from different scales and illustrates the whole methodology on synthetic price time series. This section ends with the application of the method on three well-known historical financial bubbles. Section 8 concludes.

2. Log-Periodic Power Law Singularity model

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

$$\frac{dp}{p} = \mu(t)dt + \sigma(t)dW - \kappa dj. \tag{1}$$

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

According to the Johansen-Ledoit-Sornette (JLS) model (Johansen and Sornette, 1999b; Johansen et al., 1999, 2000), the complex actions of noise traders can be aggregated into the following dynamics of the hazard rate:

$$h(t) = \alpha(t_c - t)^{m-1}\big(1 + \beta \cos(\omega \ln(t_c - t) - \phi')\big), \tag{2}$$

where α, β, ω and ϕ' are some parameters. The core of the model is the singular power law behavior $(t_c - t)^{m-1}$ that embodies the mechanism of the positive feedback at the origin of the formation of bubble leading to a super-exponential price growth. The oscillatory dressing $1 + \beta \cos(\omega \ln(t_c - t) - \phi')$ takes into account the existence of a possible hierarchical cascade of panic acceleration punctuating the course of the bubble. The particular form of the log-periodic function $\cos(\omega \ln(t_c - t) - \phi')$ in (2) is a first-order expansion of the general class of Weierstrass-type functions (Gluzman and Sornette, 2002; Zhou and Sornette, 2003b) that describes the discrete-scale invariance around tipping points in complex natural and socio-economic systems (Sornette, 1998, 2002).

4

Under the no-arbitrage condition ($\mathrm{E}[dp] = 0$), the excess return $\mu(t)$ is proportional to the crash hazard rate $h(t)$: $\mu(t) = \kappa h(t)$. Then direct solution of the equation (1) with the given dynamics of the hazard rate (2) under the condition that no crash has yet occurred ($dj = 0$) leads to the following Log-Periodic Power Law Singularity (LPPLS) equation for the expected value of a log-price:

$$\mathrm{LPPLS}(t) \equiv \mathrm{E}[\ln p(t)] = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi), \quad (3)$$

where $B = -\kappa\alpha/m$ and $C = -\kappa\alpha\beta/\sqrt{m^2 + \omega^2}$. It is important to stress that the exact solution (3) describes the dynamics of the average log-price only up to critical time t_c and cannot be used beyond it. This critical time t_c corresponds to the termination of the bubble and indicates the change to another regime, which could be a large crash or a change of the average growth rate. Nevertheless, in practical applications, one often heuristically extends (3) for $t > t_c$, assuming the validity of a time-inversion symmetry of the price trajectory around t_c (Johansen and Sornette, 1999a; Zhou and Sornette, 2003b):

$$\mathrm{LPPLS}(t) = A + B|t_c - t|^m + C|t_c - t|^m \cos(\omega \ln|t_c - t| - \phi). \quad (4)$$

The LPPLS model in its original form (3) or (4) is described by three linear parameters (A, B, C) and four nonlinear parameters (m, ω, t_c, ϕ). As discussed further in Section 3, the calibration of the model can be performed using a two-stage procedure. First, the linear parameters A, B, C for fixed values of m, ω, t_c, ϕ can be obtained directly via the solution of a matrix equation. Second, the non-linear parameters m, ω, t_c, ϕ can be found using a nonlinear optimization method. Notwithstanding the reduction from 7 to 4 of the number of parameters to determine, the numerical optimization is not straightforward, as the cost function possesses a quasi-periodic structure with many local minima. Any local optimization algorithm fails here and an extra layer involving so-called metaheuristic algorithms (Talbi, 2009) is needed in order to find the global optimum. Thus, the *taboo search* (Cvijovicacute and Klinowski, 1995) has often been used to perform this metaheuristic determination of the 4 nonlinear parameters of the LPPLS function (3).

A better approach has been suggested by Filimonov and Sornette (2013), which consists in reformulating the model (3) in a way that significantly simplifies the calibration procedure. The reformulation is based on the variable change

$$C_1 = C\cos\phi, \quad C_2 = C\sin\phi, \quad (5)$$

so that equation (3) becomes

$$\mathrm{LPPLS}(t) = A + B|t_c - t|^m + C_1|t_c - t|^m \cos(\omega \ln|t_c - t|) + C_2|t_c - t|^m \sin(\omega \ln|t_c - t|). \quad (6)$$

In this form, the LPPLS function has only 3 nonlinear (t_c, ω, m) and 4 linear A, B, C_1, C_2 parameters. As shown in (Filimonov and Sornette, 2013), this transformation significantly decreases the complexity of the fitting procedure and improves its stability

tremendously. This is because the modified cost function for (6) is now free from quasi-periodicity and enjoys good smooth properties with one or a few local minima in the case where the model is appropriate to the empirical data.

Let us complement this exposition by briefly discussing the constraints in parameter space. Since the integral $\int_{t_0}^{t_c} h(t)dt$ of the hazard rate (2) gives the probability of the occurrence of a crash, it should be bounded by 1, which yields the condition $m < 1$. At the same time, the log-price (3) should also remain finite for any $t \leq t_c$, which implies the condition $m > 0$. In addition, in order for the LPPLS formula to capture the super-exponential acceleration associated with a bubble, we need $B < 0$. Finally, the hazard rate $h(t)$ is non-negative by definition (van Bothmer and Meister, 2003), which translates into the constraint $D = m|B|/\omega|C| > 1$, where D is the so-called *damping* parameter.

Additional constraints have been proposed, based on compilations of extensive analyses of historical bubbles (Sornette and Johansen, 2001; Johansen and Sornette, 2010; Lin et al., 2014). Johansen and Sornette (2010) document approximate Gaussian distributions of ω and m with the corresponding mean and standard deviations: $\omega \approx 6.35 \pm 1.55$ and $m \approx 0.33 \pm 0.18$. In practical implementations, these constraints are slightly modified in order to minimize errors of type I (incorrect rejection of the LPPLS hypothesis). In particular, the constraints for ω are often pushed upward to avoid small angular log-frequencies that can spuriously appear as a result of improper fitting of trends. Finally, the strict theoretical constraint $D > 1$ on the damping parameter is derived under the assumption that the crash occurs in one immediate negative jump. As this is in general counterfactual (a crash has usually a duration of weeks to months, and is characterized by a large drawdown (Johansen and Sornette, 2001/02, 2010)), the constraint $D > 1$ can be relaxed (Sornette et al., 2015). To sum up, the following set of theoretical and empirical constraints on the parameters can be regarded as the stylized features of LPPLS:

$$0.1 \lesssim m \lesssim 0.9, \quad 6 \lesssim \omega \lesssim 13, \quad B < 0, \quad D = \frac{m|B|}{\omega\sqrt{C_2^2 + C_2^2}} \gtrsim 0.8. \tag{7}$$

The last important question we need to address is the "nature" of the proper time t to consider. The standard jump-diffusion equations (1) can be applied either to *calendar time*, where t continuously increases, or to *business time*, where weekends and non-trading days are omitted within a discrete version where t increases by 1 day from Friday to the next Monday. The standard discrete normal diffusion process (i.e. without the jump term) is invariant to this transformation of the calendar, assuming that the drift μ and volatility σ are rescaled properly. While the LPPLS model (3) is not invariant to such time change, both approaches are possible. In previous works, the calibration of the model has been performed using calendar time, assuming that the price variations over non-trading dates are non-observable but nevertheless embody an information flow that impacts the overall price dynamics. Here, we also use calendar time.

3. Nonlinear regression and Ordinary Least Squares fitting

3.1. Solution of the nonlinear regression problem

In the LPPLS framework, forecasting the termination of a bubble amounts to finding the best estimation of the critical time $\hat{t}_c$. This requires calibrating formula (6) on the observed price trajectory in order to determine t_c together with all the other parameters of the model, $\psi = \{m, \omega, A, B, C_1, C_2\}$. In previous works, this was done via a nonlinear regression of the vector of log-prices $Y = \{\ln p(\tau_i)\}$ on the vector of observation dates $X = \{\tau_i : \tau_i \in [t_1, t_2]\}$, where $[t_1, t_2]$ denotes the window of analysis. The *Ordinary Least Squares (OLS)* method amounts to minimizing the sum of squared residuals $(\varepsilon(\tau_i; t_c, \psi) = \ln p(\tau_i) - \text{LPPLS}(\tau_i; t_c, \psi))$ between Y and the LPPLS formula,

$$\{\hat{t}_c, \hat{\psi}\} = \arg \min_{t_c, \psi} \text{SSE}(t_c, \psi), \tag{8}$$

where the sum of squared errors (SSE) is given by

$$\text{SSE}(t_c, \psi) = \sum_{i=1}^{n} \left(\varepsilon(\tau_i; t_c, \psi) \right)^2 \equiv \sum_{i=1}^{n} \left(\ln p(\tau_i) - \text{LPPLS}(\tau_i; t_c, \psi) \right)^2. \tag{9}$$

Minimization of such nonlinear multivariate cost function is a highly non-trivial task due to presence of multiple local minima, where the local optimization algorithm can get trapped.

In the original formulation of the model (3), three linear parameters A, B, C can be slaved to the nonlinear parameters t_c, m, ω, ϕ (Johansen and Sornette, 1999b). This decreases significantly the complexity of the calibration problem, but does not remove the quasi-periodic structure of the cost-function with many minima. As mentioned above, this requires metaheuristic methods for the optimization. In the reformulated model (6), the complexity of the optimization problem is further decreased by transforming the non-linear phase into a linear parameter. Then, one can first slave the four linear parameters A, B, C_1, C_2 to the three remaining nonlinear parameters t_c, m, ω (Filimonov and Sornette, 2013). The minimization problem (8) is thus transformed into:

$$\{\hat{t}_c, \hat{m}, \hat{\omega}\} = \arg \min_{t_c, m, \omega} F_1(t_c, m, \omega), \tag{10}$$

where the cost function $F_1(t_c, m, \omega)$ is given by

$$F_1(t_c, m, \omega) = \min_{A, B, C_1, C_2} \text{SSE}(t_c, m, \omega, A, B, C_1, C_2). \tag{11}$$

The optimization problem in (11) has a unique solution obtained directly from the first-order condition:

$$\begin{pmatrix} n & \sum f_i & \sum g_i & \sum h_i \\ \sum f_i & \sum f_i^2 & \sum f_i g_i & \sum f_i h_i \\ \sum g_i & \sum f_i g_i & \sum g_i^2 & \sum g_i h_i \\ \sum h_i & \sum f_i h_i & \sum g_i h_i & \sum h_i^2 \end{pmatrix} \begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{C}_1 \\ \hat{C}_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i f_i \\ \sum y_i g_i \\ \sum y_i h_i \end{pmatrix} \tag{12}$$

where

$$y_i = \ln p(\tau_i),$$
$$f_i = |t_c - \tau_i|^m,$$
$$g_i = |t_c - \tau_i|^m \cos(\omega \ln |t_c - \tau_i|),$$
$$h_i = |t_c - \tau_i|^m \sin(\omega \ln |t_c - \tau_i|).$$

$$(13)$$

As discussed in Section 2, the reduction from 4 to 3 nonlinear parameters decreases dramatically the number of local extrema to only a few, so there is not much need for metaheuristic methods such as the Taboo search (Cvijovicacute and Klinowski, 1995), which was previously the main tool of the LPPLS calibration (Johansen et al., 2000). In most cases, a single "quasi-local" optimization algorithm such as the Nelder-Mead simplex method (Nelder and Mead, 1965) can reliably find the absolute minimum of $F_1(t_c, m, \omega)$ given by expression (11). In complicated cases the Nelder-Mead simplex method can be complemented by employing *repeated local searches*. This amounts to start the local search routine from multiple different initial points and then select the best solution.

3.2. Estimation of the critical time

In practical applications, the calibration of the LPPLS model often aims at forecasting the critical time t_c, because it is, by construction of the LPPLS model, the end of the bubble regime. This suggests to develop a special treatment for t_c. In this spirit, Filimonov and Sornette (2013) suggested to reformulate the optimization problem (10) by subordinating the logperiodic angular frequency ω and power law exponent m to t_c:

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

$$(14)$$

where

$$F_2(t_c) = \min_{\omega,m} F_1(t_c, m, \omega), \quad \{\hat{m}(t_c), \hat{\omega}(t_c)\} = \arg \min_{m,\omega} F_1(t_c, m, \omega) \qquad (15)$$

and $F_1(t_c, m, \omega)$ is given by (11).

In general, such extra subordination dramatically reduces the number of local extrema of the cost-function. As we will see later from Figure 4, when the price trajectory displays a pronounced increase, the function $F_1(t_c, m, \omega)$ almost always presents just one minimum along the m direction and 3-4 local minima along the ω direction in the range $2 < \omega < 20$ (which may actually be relevant to capture higher harmonics of the logperiodicity structure (Zhou and Sornette, 2003a,b)). Further, this method allows one to avoid sloppy directions in the (t_c, ω) plane, where the cost-function has a very long valley along the diagonal $t_c \sim \omega$, as illustrated in Figure 3b of (Filimonov and Sornette, 2013).

At the expense of a small increase of computational complexity, beyond its simplification, the cost-function given by equation (15) provides a substantial improvement in inference from the model. Namely, in addition to the point estimate (14), expression (15) allows one to analyze the whole profile of the cost function $F_2(t_c)$ and the

dependence of the estimates $\hat{m}$ and $\hat{\omega}$ as a function of the critical time t_c. In particular, one can identify all the extrema of $F_2(t_c)$ and their corresponding $m(t_c)$ and $\omega(t_c)$, from which expert judgment of the plausible scenarios can follow.

As an example, we consider the recent bubble and following collapse of the Chinese market, when the Shanghai Composite Index (SSE Composite) appreciated by approximately 150% between mid-2014 and mid-2015, peaked on June 12, 2015 and then lost 32% to its first well-defined bottom reached on July 8, 2015. This bubble was detected by the Financial Crisis Observatory (FCO) at ETH Zürich and further documented and dissected in (Sornette et al., 2015). We use data provided by Thomson Reuters Dataworks Enterprise (see Section 7.3 for discussions). Figure 1 presents the dynamics of the SSE Composite index together with the best LPPLS fit according to the OLS regression within the time window of $t_2 - t_1 = 180$ calendar days ending at the date of $t_2 =$June 12, 2015 when the market peaked.

In order to understand the "microstructure" of LPPLS fits, we employ the three-step subordination procedure (11)–(13), (14)–(15) and study the dependence of the cost-function $F_2(t_c)$ as well as $\hat{m}(t_c)$, $\hat{\omega}(t_c)$ and damping $\hat{D}(t_c)$ (see Fig. 2). One can see that the global (best) solution with estimated critical time $\hat{t}_c$ of July 7, 2015 (with $F_2 = 0.0597$) is not the only minimum, and a second local minimum is found at $t_c =$ June 18, 2015 (with $F_2 = 0.0604$), which suggests a second plausible scenario. Despite almost identical values of the cost functions (sum of squared errors of residuals), we might reject the suboptimal solution on the basis of the fact that its logperiodic angular frequency falls outside of the empirical constraint (7) ($\hat{\omega} = 7.18$ for the optimal solution and $\hat{\omega} = 5.85$ for the suboptimal). Both solutions are associated with damping parameters that are below the constraint $D \geq 1$ ($\hat{D} = 0.8$ for the optimal and $\hat{D} = 0.83$ for the suboptimal solution), but they are both compatible with the relaxed constraint (7) (note that the value for the optimal solution is very close to the boundary of this constraint).

This case study exemplifies the essence of the problem of dealing with multiple and almost equivalent optimal solutions that point to quite different future scenarios. Above we have invoked previous experience (Johansen and Sornette, 2010) to reject the second scenario. However, this is not fully satisfactory from a theoretical view point. Moreover, past experience can be tainted by the use of the sub-optimal calibration procedure based on the original formulation of the model (3). To boot, past experience may not contain all possible situations, and surprises that are superficially of the "unknown unknown" type (Knight, 1921; Taleb, 2007) from the point of view of past experience might actually be understandable and knowable with the appropriate conceptual and theoretical framework (Sornette, 2009).

The question we further investigate below is: How can we resolve between these two scenarios if we do not have (or do not want or trust to use) any prior information on what are plausible parameter values? In other words, how can we provide a quantitative estimation of how much one scenario is less likely than another?
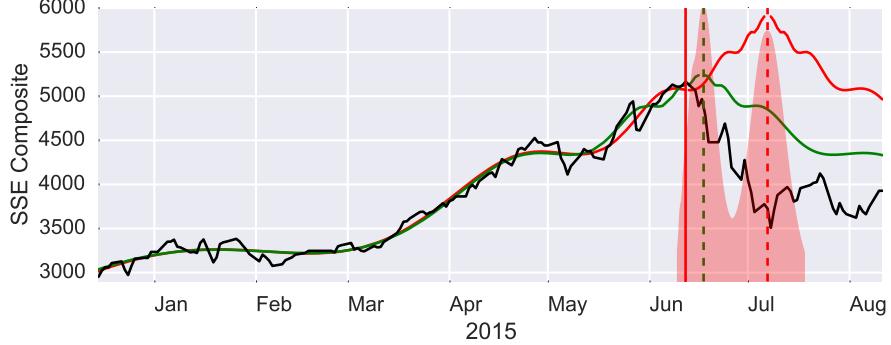
Figure 1: Price trajectory of the SSE Composite Index during the bubble of 2014–2015. The red vertical line denotes the date of the analysis (t_2 = 2015-06-12). Red and green solid lines correspond to the best and second best (see Figure 2) LPPLS fit in the window [2014-12-15, 2015-06-12] and their extrapolations to $t > t_2$. The vertical red and green dashed lines indicate the position of the critical times t_c for these two fits: 2015-07-08 and 2015-06-18 respectively. The shaded red areas delineate the likelihood interval of t_c at a 5% cutoff together with the shape of the modified profile likelihood (see Sections 5–5.4 and Figure 3).
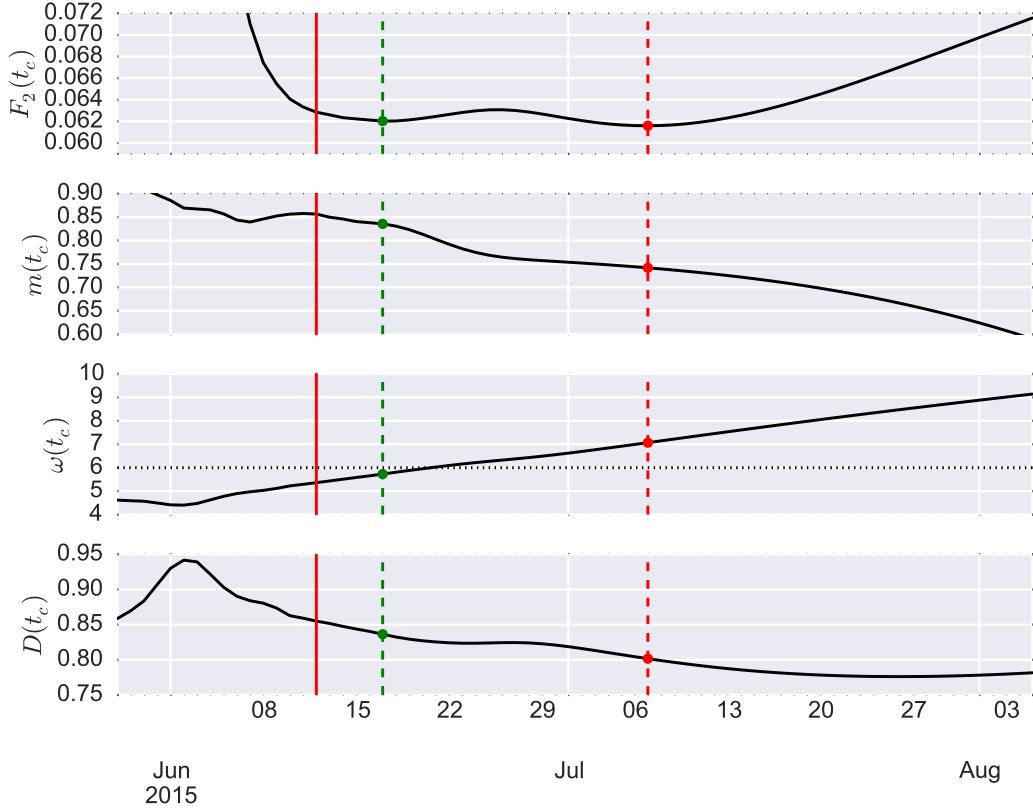

Figure 2: Profiles of the cost function $F_2(t_c)$ and parameters $\hat{m}$, $\hat{\omega}$ and damping $\hat{D} = m|B|/\omega|C|$ as a function of t_c. The solid red vertical line indicates the date of analysis (t_2 = 2015-06-12), the red and green dotted vertical lines correspond to the dates of the best and alternative solutions (2015-07-08 and 2015-06-18 respectively). The horizontal dotted line gives the level of the threshold (7) for the logperiodic angular frequency parameter ($\omega > 6$) that separates "qualified" fits from "not qualified" ones. The filled red and green circles show the point estimates of the model parameters — for the best and alternative fits, respectively.

10

4. Likelihood and Profile Likelihood

The OLS regression (8) represents the so-called *normal estimation* of the model parameters, i.e. provides *Maximum Likelihood Estimates (MLE)* under the assumption that the error term $\varepsilon(\tau_i; t_c, \psi)$ is normally distributed. The likelihood has then a well-known form:

$$L(t_c, \psi, s) = (2\pi s)^{-n/2} \exp\left(-\frac{\text{SSE}(t_c, \psi)}{2s}\right) \to \max_{t_c, \psi, s}, \qquad (16)$$

where $s = \sigma^2$ is a variance of the residuals $\varepsilon(\tau_i; t_c, \psi)$ and n is the number of data points. By definition the likelihood is meaningful only up to an arbitrary positive constant, thus below we will omit such constant pre-factors. The MLE of the parameters $\{\hat{t}_c, \hat{\psi}\}$ is obtained straight from (16): considering the logarithm of the likelihood ($\ln L(t_c, \psi, s)$), one immediately arrives at (8) and an estimate for σ^2 is

$$\hat{\sigma}^2 \equiv \hat{s} = \frac{1}{n}\text{SSE}(\hat{t}_c, \hat{\psi}). \qquad (17)$$

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

As discussed above, we are mostly interested in the inference of the critical time t_c while the other parameters $\eta = \{\psi, s\} \equiv \{m, \omega, A, B, C_1, C_2, s\}$ can be considered as *nuisanse* parameters that are useful insofar that they allow to adapt the model to the variability of the data. The elimination of nuisance parameters is a well-known statistical problem, which amounts to concentrating the likelihood around a single parameter of interest while accounting for the extra uncertainty resulting from the estimation of the nuisance parameters. Unfortunately, there is no technique that is efficient for all situations (Bayarri and DeGroot, 1992), in particular because it is not always meaningful to discuss the uncertainty in one parameter independently from that of all others.

As already mentioned, in the Bayesian approach, the elimination of the nuisance parameters corresponds to integrating them out. However, the likelihood is not a regular density function and does not obey probability laws. Therefore, the naive integration of the likelihood is not a meaningful operation. The proper implementation of the Bayesian approach requires specifying the prior distribution of all parameters $\{t_c, \eta\}$, calculating the posterior and then integrating out the nuisance parameters η from the posterior to derive the posterior marginal distribution of t_c. The major limitation of the Bayesian approach is indeed a specification of the prior. We will not pursue this way directly, however, as shown in Section 5, we will be able to capture the idea of integration over the nuisance parameters within a non-Bayesian framework.

One commonly used practice of elimination of nuisance parameters is based on a factorization of the complete likelihood into a product of the so-called *marginal* and *conditional likelihood functions* (Kalbfleisch and Sprott, 1970). When available, this

11

approach results in a genuine likelihood, i.e. the genuine probability of the observed data conditional on the parameter of interest (t_c). However, this approach requires transforming the sufficient statistics into a minimal sufficient statistics that has to be factored into two terms T and A. One of these terms, either the marginal distribution of T or the conditional distribution of T conditioned on A (which is then called *ancillary* for t_c), depends only on t_c, but not on η (see discussions in (Pawitan, 2001) and for example (Basu, 1977; Severini, 1998b; Qin, 2005)). Given that the LPPLS model (6) is highly nonlinear, it is not possible to find such factorization.

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

$$L_p(t_c) = \max_\eta L(t_c, \eta) \equiv L(t_c, \hat{\eta}_{t_c}), \tag{18}$$

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

The profile likelihood approach is technically identical to the analysis of the profile cost function $F_2(t_c)$ discussed in Section 3 Indeed, the MLE of $\hat{\psi}_{t_c}$ is given by the solution of the OLS (8): the estimates of $\hat{m}_{t_c}$ and $\hat{\omega}_{t_c}$ are derived from (15) where $\hat{A}_{t_c}, \hat{B}_{t_c}, \hat{C}_{1,t_c}, \hat{C}_{2,t_c}$ are given by (12). Finally, the form of $\hat{s}_{t_c}$ is similar to (17), where $\hat{t}_c$ is replaced by t_c. Moreover, the value of $L_p(t_c)$ can be directly derived from $F_2(t_c)$. Indeed, the estimation of $\hat{s}_{t_c}$ can be represented as

$$\hat{s}_{t_c} = \frac{1}{n}\text{SSE}(t_c, \hat{\psi}) \equiv \frac{1}{n}F_2(t_c), , \tag{19}$$

and, after plugging (19) to (16) according to (18) we obtain:

$$L_p(t_c) \propto \left(\hat{s}_{t_c}\right)^{-n/2} \propto \left(F_2(t_c)\right)^{-n/2}, \tag{20}$$

where we have omitted all constant terms.

Since the likelihood (16) is meaningful only up to a constant, one usually considers the *relative likelihood* (respectively, *relative profile likelihood* or *relative modified profile likelihood* that will be defined later), which is normalized to 1 by its maximum and thus takes value in $[0, 1]$:

$$R(t_c) = \frac{L(t_c)}{\max_{t_c} L(t_c)}. \tag{21}$$

Figure 3 (blue curves) presents an example of the relative profile likelihood $R(t_c)$ calculated for the case discussed in Section 3 and presented in Figure 2. One can observe the same two extrema found with function $F_2(t_c)$, which correspond to very close values of the likelihood ($R_p(t_c^{(1)}) = 1$ for the best solution $t_c^{(1)} = $ 2015-07-08 and $R_p(t_c^{(2)}) = 0.64$ for the alternative solution $t_c^{(2)} = $ 2015-06-18). The likelihood ratio
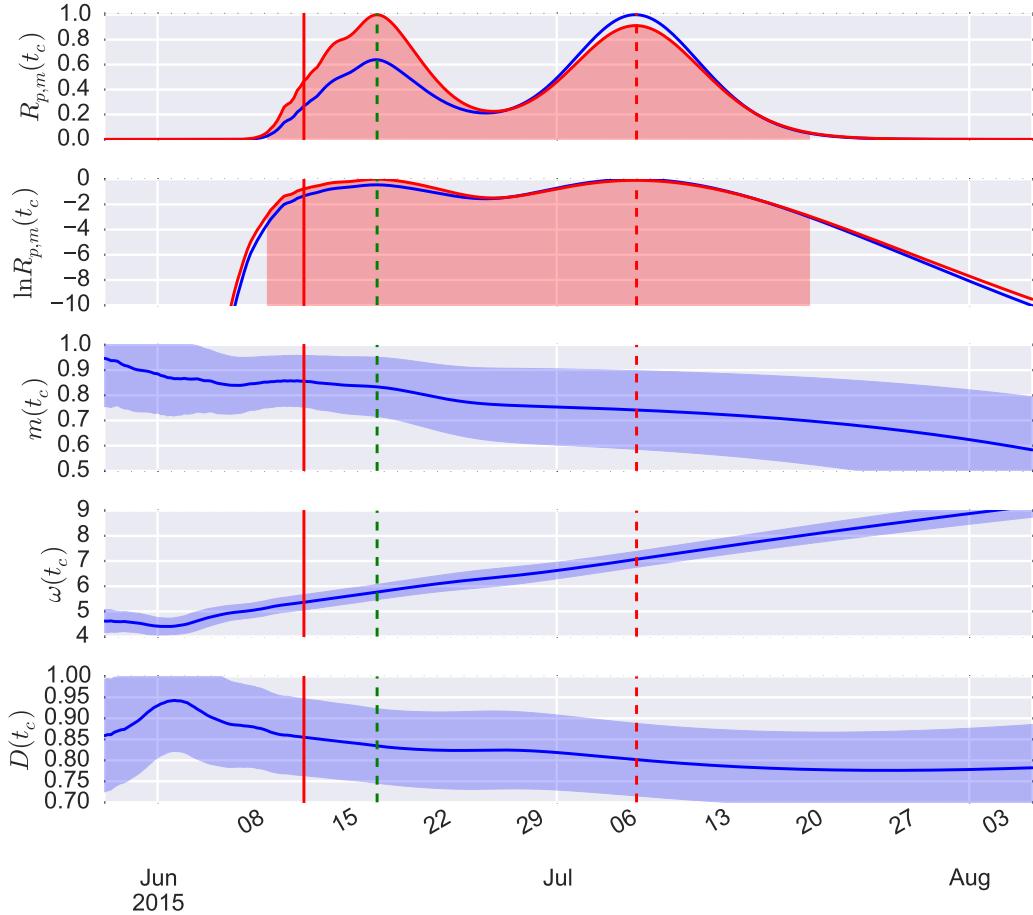
12

Figure 3: The top panel shows the relative profile likelihood (blue) and modified profile likelihood (red) as a function of the critical time t_c. The second panel from the top shows the corresponding log-likelihoods. The red shaded area corresponds to the likelihood interval of t_c at the 5% probability level (see Section 5.4). The three bottom panels give the point MLE parameter estimates of the model, m, ω and D, as a function of the critical time t_c (same as in Figure 2). Blue lines present point MLE estimates of the parameters and blue shaded areas correspond to their approximated likelihood intervals at the 5% probability level (see Section 6). The red continuous vertical line indicates the date of analysis ($t_2 = $ 2015-06-12). The dashed red and green vertical lines correspond to the dates of the best and alternative solutions (2015-07-08 and 2015-06-18 respectively).

$R_p(t_c^{(1)})/R_p(t_c^{(2)}) = 1.56$ is not large enough to warrant preferring one maximum over the other. The inference based on a point OLS (or MLE) estimate can thus be quite misleading. In fact, the interval of "acceptable" values for t_c (the likelihood interval to be discussed in Section 5.4) is very broad, which confirms that a point estimation is far from reflecting the full picture.

13

5. Modified Profile Likelihood

5.1. General form of the modified profile likelihood

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

In order to overcome this fundamental limitation of the profile likelihood, a series of adjusted versions have been proposed (see for instance, (Cox and Reid, 1987; Fraser and Reid, 1989; Barndorff-Nielsen and Cox, 1994; DiCiccio et al., 1996) for general discussions). Most of them require orthogonality between the parameter of interest (t_c) and nuisance parameter (η). In our case, orthogonality does not hold and, in order to come up with a parametrization $\tilde{\eta}$ that would be orthogonal to t_c, one needs to solve a system of differential equations (Cox and Reid, 1987), which is nearly impossible to do analytically in our multi-dimensional non-linear case. Then, the most flexible approach is arguably the one proposed by Barndorff-Nielsen (1983), who introduced the so-called *modified profile likelihood* as a higher-order approximation to either a marginal or a conditional likelihood function (both derivations are possible).

The modified profile likelihood amounts to introducing an extra modulating factor $M(t_c)$ to the profile likelihood:

$$L_m(t_c) = M(t_c)L_p(t_c) = |I(\hat{\eta}_{t_c})|^{-1/2} \left| \frac{\partial \hat{\eta}}{\partial \hat{\eta}_{t_c}} \right| L_p(t_c), \tag{22}$$

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

$$I(\hat{\eta}_{t_c}) = - \left. \frac{\partial^2 \ln L(t_c, \eta)}{\partial \eta \partial \eta^T} \right|_{\eta = \hat{\eta}_{t_c}}, \tag{23}$$

where η^T stands for the transpose of η; $\partial \hat{\eta}/\partial \hat{\eta}_{t_c}$ denotes a matrix of the first partial derivatives of the full MLE of the nuisance parameters η with respect to the MLE calculated at a fixed value of t_c; finally, $|\cdot|$ denotes the absolute value of a matrix determinant. Here and in the following, we assume that the parameters form a column vector, thus second order derivatives of the form (23) define a matrix.

The term $|I(\hat{\eta}_{t_c})|^{-1/2}$, which describes the curvature of the likelihood, can be considered as a penalty that subtracts from the profile log-likelihood "undeserved" information due to the estimation of the nuisance parameter η. And the Jacobian term

$J(t_c) = |\partial\hat{\eta}/\partial\hat{\eta}_{t_c}|$ is needed to make the modified profile likelihood invariant with respect to the transformations of the nuisance parameters (Pawitan, 2001). Practically, this term is extremely difficult to evaluate, which dramatically limits the application of (22).

Unlike profile likelihoods, the modified profile likelihood is a genuine likelihood function and has a number of important properties. First, as just mentioned, due to the Jacobian term, $L_m(t_c)$ is invariant with respect to a reparametrisation of the model (such as the variable change (5)). Second, it does not require orthogonality of t_c and η, neither does it require specification of an ancillary statistics. Finally, Severini (2007) has shown that the modified profile likelihood can be considered as an approximation to a class of integrated likelihood functions and very naturally arises from a non-Bayesian inference with integrated likelihood. But, in contrast to the Bayesian approach or integrated likelihood functions, the modified profile likelihood does not require specification of a prior density for the nuisance parameters — the main limitation that hampered us from pursuing this direction.

5.2. Inference on the errors variance

The importance of the modified profile likelihood cannot be overstated, given that it is considered one of the breakthroughs in modern parametric inference (DiCiccio, 1997). Perhaps the best illustration of the power of this method relates to the estimation of the variance $s = \sigma^2$ in nonlinear regressions such as (16). It is well known that the standard estimation (17) or (19) is biased and it should be corrected to account for the number of degrees of freedom, i.e. the number of free parameters to estimate. The modified profile likelihood provides this correction as follows.

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

Straight from (16), we can derive the vector of first derivatives of the log-likelihood — the so-called *score function $S(\lambda)$*:

$$S(\lambda) = \frac{\partial \ln L(s, \lambda)}{\partial \lambda} = -\frac{1}{2s}\frac{\partial \mathrm{SSE}(\lambda)}{\partial \lambda}. \tag{24}$$

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

$$|I(\lambda)| = \left|-\frac{\partial^2 \ln L(s, \lambda)}{\partial \lambda \partial \lambda^T}\right| = \left(\frac{1}{2s}\right)^{p_\lambda}\left|\frac{\partial^2 \mathrm{SSE}(\lambda)}{\partial \lambda \partial \lambda^T}\right|, \tag{25}$$

where $p_\lambda = \dim \lambda = 7$ is the dimension of the nuisance parameter space. Before plugging the expression (25) into (22) in order to obtain the modified profile likelihood of s, notice that (i) the matrix of second-order derivatives $\partial^2\mathrm{SSE}(\lambda)/\partial\lambda\partial\lambda^T$ in (25)

15

does not depend on the parameter of interest s explicitly and (ii) the OLS estimation $\hat{\lambda}_s \equiv \hat{\lambda} = \{\hat{t}_c, \hat{m}, \hat{\omega}, \hat{A}, \hat{B}, \hat{C}_1, \hat{C}_2\}$ also does not depend on s. Thus, the determinant of $\partial^2 \text{SSE}(\lambda)/\partial\lambda\partial\lambda^T$ is a constant with respect to the variable s and therefore can be omitted. Then, the modified profile likelihood of s can be expressed in the following form:

$$L_m(s) \propto s^{(n-p_\lambda)/2} \exp\left(-\frac{\text{SSE}(\hat{\lambda})}{2s}\right), \tag{26}$$

which leads to the following MLE for s:

$$\hat{s} = \frac{1}{n - p_\lambda}\text{SSE}(\hat{\lambda}). \tag{27}$$

The denominator $n - p_\lambda$, which is different from n in (17), not only removes the bias of the estimator, but also results in a better likelihood-based inference of s when it is needed.

5.3. Approximation of the modified profile likelihood

Given all the remarkable properties of the modified profile likelihood, it has one very serious limitation, briefly mentioned above. Namely, for many realistic models, it is extremely difficult to calculate the Jacobian in (22). In order to get an intuition about the nature of the difficulty, it is useful to express it in the following form (see e.g. (Pawitan, 2001)):

$$J(t_c) \equiv \left|\frac{\partial\hat{\eta}}{\partial\hat{\eta}_{t_c}}\right| = \frac{|I(\hat{\eta}_{t_c})|}{|C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})|}, \tag{28}$$

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

$$C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}) = \frac{\partial^2 \ln L(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}, a)}{\partial\hat{\eta}_{t_c}\partial\hat{\eta}^T}. \tag{29}$$

In contrast to the observed Fisher information, which is also defined as a second-order derivative (23) calculated at a specific MLE $\hat{\eta}_{t_c}$, the calculation of C (29) is much more complicated because, in the general case, it requires a reformulation of the log-likelihood in order to introduce an explicit dependence on the MLEs $\hat{\eta}_{t_c}$ and $\hat{\eta}$. In the case of inference of the variance s presented in Section 5.2, we used the orthogonality of s and λ, which resulted in $\hat{\lambda}_s \equiv \hat{\lambda}$. In contrast, for the inference on t_c, there is no closed form expression for $J(t_c)$. And, as discussed above, we cannot use the *adjusted profile likelihood* (Cox and Reid, 1987) because orthogonalization of the nuisance parameters with respect to t_c is not feasible either.

In order to calculate expression (22), several approximation of L_m were proposed (see e.g. (Barndorff-Nielsen, 1994; Skovgaard, 1996; Severini, 1998a; Fraser et al., 1999; Skovgaard, 2001) and (Severini, 2001; Pace and Salvan, 2006) for reviews). We will

16

use the approximation to the modified profile likelihood proposed by Severini (1998a). This approximation requires only the covariance of score functions of the nuisance parameters and is thus fairly easy to compute. As shown in (Severini, 1998a), this approximation is invariant under the reparametrization of the model, is stable in the sense of conditional inference and agrees with the exact $J(t_c)$ (28) to order $\mathcal{O}(n^{-1})$ in the moderate deviation sense and to order $\mathcal{O}(n^{-1/2})$ in the large deviation sense, where n is the number of data points. Another famous approximation by Barndorff-Nielsen (1994) agrees with the exact form of L_m only to $\mathcal{O}(1)$ in the large deviation sense and thus is not asymptotically better than the simple profile likelihood L_p.

Severini (1998a) suggested to approximate the matrix (29) with the covariance matrix of score functions of the following form:

$$C(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}) \approx \Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta}) \tag{30}$$

where

$$\Sigma(t_{c;1}, \eta_1; t_{c;2}, \eta_2) = \mathrm{E}_{(2)} \left[\left. \frac{\partial \ln L(t_c, \eta)}{\partial \eta} \right|_{\substack{t_c=t_{c;1} \\ \eta=\eta_1}} \left. \frac{\partial \ln L(t_c, \eta)}{\partial \eta^T} \right|_{\substack{t_c=t_{c;2} \\ \eta=\eta_2}} \right]. \tag{31}$$

Here the expectation $\mathrm{E}_{(2)}[\cdot]$ is taken with respect to the probability distribution of error term $\varepsilon(\tau; t_{c;2}, \eta_2)$ that corresponds to the parameters $\{t_{c;2}, \eta_2\}$. In contrast to the exact form (29), here we need only the score functions, which have expressions similar to (24). When calculation of (31) is too complicated, one can exploit the independence of observations $\varepsilon(\tau_i; t_c, \eta)$ and replace the covariance matrix (31) by its asymptotically equivalent sample estimation (Severini, 1999):

$$\widehat{\Sigma}(t_{c;1}, \eta_1; t_{c;2}, \eta_2) = \sum_{i=1}^{n} \left. \frac{\partial f(\tau_i; t_c, \eta)}{\partial \eta} \right|_{\substack{t_c=t_{c;1} \\ \eta=\eta_1}} \left. \frac{\partial f(\tau_i; t_c, \eta)}{\partial \eta^T} \right|_{\substack{t_c=t_{c;2} \\ \eta=\eta_2}}, \tag{32}$$

where

$$f(\tau; t_c, \eta) = -\frac{1}{2}\ln(2\pi s) - \frac{1}{2s}\left(\ln p(\tau) - \mathrm{LPPLS}(\tau; t_c, \psi) \right)^2 \tag{33}$$

is a contribution from an individual observation to the log-likelihood. Of course, the adjustment (31) based on the theoretical covariance is superior to the sample-based estimation (32), in particular in cases of small sample size (Severini, 1999; Bester and Hansen, 2009). For our purposes, we will use the exact form (31), which can be calculated in closed form. Finally, plugging (31) into (28) and (22), we obtain the desired approximated expression for $L_m(t_c)$:

$$L_m(t_c) \approx \frac{|I(\hat{\eta}_{t_c})|^{1/2}}{|\Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})|} L_p(t_c) . \tag{34}$$

In this expression (34), the profile likelihood $L_p(t_c)$ is given by the previously calculated expression (20). The observed Fisher information $I(\hat{\eta}_{t_c})$ and the covariance matrix $\Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})$ are given in Appendix Appendix A. Omitting terms that do not

depend on t_c, the final expression for $L_m(t_c)$ is given by:

$$L_m(t_c) \propto \frac{\left(\hat{s}_{t_c}\right)^{-(n-p-2)/2} \left| \sum_{i=1}^{n} \frac{\partial^2 \mathrm{LPPLS}(\tau_i; t_c, \psi)}{\partial \psi \partial \psi^T} \right|^{1/2}_{\psi=\hat{\psi}_{t_c}}}{\left| \sum_{i=1}^{n} \frac{\partial \mathrm{LPPLS}(\tau_i; t_c, \psi)}{\partial \psi} \right|_{\substack{t_c=t_c \\ \psi=\hat{\psi}_{t_c}}} \left. \frac{\partial \mathrm{LPPLS}(\tau_i; t_c, \psi)}{\partial \psi^T} \right|_{\substack{t_c=\hat{t}_c \\ \psi=\hat{\psi}}}}, \tag{35}$$

where $p = \dim \psi = 6$. Following (Severini, 1999), let us introduce the rectangular $n \times p$ matrix

$$X_{ij}(t_c, \psi) = \frac{\partial \mathrm{LPPLS}(\tau_i; t_c, \psi)}{\partial \psi_j} \tag{36}$$

and the square $p \times p$ matrix

$$H_{ij}(t_c, \psi) = \sum_{k=1}^{n} \left(\ln p(\tau_k) - \mathrm{LPPLS}(\tau_k; t_c, \psi) \right) \frac{\partial^2 \mathrm{LPPLS}(\tau_k; t_c, \psi)}{\partial \psi_i \partial \psi_j}, \tag{37}$$

where ψ_j denotes the j-th element of the nuisance parameter vector $\{m, \omega, A, B, C_1, C_2\}$. Then, expression (35) simplifies into

$$L_m(t_c) \propto \frac{\left| X^T(t_c, \hat{\psi}_{t_c}) X(t_c, \hat{\psi}_{t_c}) - H(t_c, \hat{\psi}_{t_c}) \right|^{1/2}}{\left| X^T(\hat{t}_c, \hat{\psi}) X(t_c, \hat{\psi}_{t_c}) \right|} \left(\hat{s}_{t_c} \right)^{-(n-p-2)/2}, \tag{38}$$

where $\hat{s}_{t_c}$ is the MLE estimate of the variance (19) (it is not the adjusted estimate (27)), $\hat{\psi}_{t_c}$ is a vector of MLE estimates for the LPPLS parameters at a fixed value of t_c and $\{\hat{t}_c, \hat{\psi}\}$ are full MLE estimates of the parameters. The expressions of the first-order and second-order partial derivatives that are needed for (36) and (37) are given by (B.1) and (B.2) in Appendix B.

As a concrete illustration, we consider the 2015 bubble in Chinese markets already discussed in Sections 3–4. The red curves in the two top panels of figure 3 show the modified profile likelihood obtained from expression (38). It is particularly interesting that the adjustments to $L_m(t_c)$ have significantly changed the picture, since the "alternative" extremum has now a higher likelihood than the best OLS solution ($R_m(t_c^{(2)}) = 1$. versus $R_m(t_c^{(2)}) = 0.91$). Thus, accepting the OLS point estimate would bias $\hat{t}_c$ by 19 days. The likelihood ratio is now even smaller than for the simple profile likelihood, $R_m(t_c^{(2)})/R_m(t_c^{(1)}) = 1.096$, and both extrema are almost equally likely.

5.4. Likelihood Intervals and Confidence Intervals

The major improvement of the standard MLE interpretation (16) over the OLS (8) is the fact that MLE provides a direct estimation of the uncertainty in estimated parameters. In other words, MLE can provide not just the *point* estimate of θ but a *range estimate* of values that are possible given the observed data. Such inference is based on

the *likelihood ratio* $R(\theta)$, introduced earlier in the form of the relative likelihood (21) defined as the ratio of the likelihood normalized by its maximum value. When $R(\theta_0)$ is sufficiently small, the hypothesis that the parameter could have a value $\theta = \theta_0$ can be rejected as "unsupported by the data".

However, the question of "how small is sufficiently small?" often does not have a rigorous solution and strongly depends on the problem. Many authors suggest to choose some rather arbitrary *cutoff* and consider values of likelihood ratio above this cutoff to define a so-called *likelihood-based confidence interval* or *likelihood interval (LI)*. For example, many authors including Fisher (1956)suggested that parameter values θ for which $L(\hat{\theta})/L(\theta) = 1/R(\theta) > 15$ should be declared "implausible", where $\hat{\theta}$ is the standard MLE.

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

For our applications, these approaches are not perfectly suited. First, as can be seen in Figure 3, the profile and modified profile likelihoods are not regular: they are asymmetric and can be even multi-modal, so that the Wald CI does not provide a meaningful representation of parameter uncertainty. For the same reason, the calibration of the distribution of the test statistic under the null hypothesis is not straightforward and would be computationally very difficult given the dimensionality of the parameter space and the complexity of the LPPLS model (6). Finally, within our domain of application, an interpretation of the frequentist probability-based confidence intervals is not very intuitive. Indeed, giving the idiosyncratic nature of a bubble, in order to make sense out of the probabilistic intervals, one needs to involve a many-worlds interpretation, where price trajectory is shared among multiple universes.

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

$$\text{LI}(t_c) = \left\{ t_c : R_m(t_c) = \frac{L_m(t_c)}{L_m(\hat{t}_c)} > 0.05 \right\}. \tag{39}$$

The two top panels of Figure 3 show such 5% modified profile likelihood intervals for the case of 2015 Chinese bubble.

6. Filtering and likelihood intervals for nuisance parameters

Similarly to the inference on the critical time t_c, let us apply the modified profile likelihood approach to estimate the likelihood intervals (LIs) of parameters m and ω.

This is of interest in particular because m and ω are used in the filtering conditions (7).

Three different ways of inference on m and ω exist. First, we could consider m (respectively, ω) as the sole parameter of interest and $\eta_m = \{t_c, \omega, A, B, C_1, C_2, s\}$ (respectively, $\eta_\omega = \{t_c, m, A, B, C_1, C_2, s\}$) as the vector of nuisance parameters. Then, an analysis similar to that developed in Sections 5–5.4 would provide the corresponding modified profile likelihood and LIs for these two parameters. However, keeping in mind that the parameter of main interest is the critical time t_c, we would need to somehow associate the inferred LIs for m (ω) with the corresponding values of t_c.

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

Here, we employ a third approach. For any fixed value of t_c, we consider a reduced LPPLS formula that is parameterized solely with the vector $\{m, \omega, A, B, C_1, C_2\}$. We then calculate a modified profile likelihood $L_m(m; t_c)$ (respectively $L_\omega(\omega; t_c)$) with $\eta_m = \{\omega, A, B, C_1, C_2, s\}$ (respectively, $\eta_\omega = \{m, A, B, C_1, C_2, s\}$) as the vector of nuisance parameters. The expression for $L_m(\cdot)$ is then similar to (38). For example, $L_m(m; t_c)$ has the form

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

(40)

where $\phi = \{\omega, A, B, C_1, C_2\}$, $p = \dim \phi = 5$, $\{\hat{t}_c, \hat{m}, \hat{\phi}\}$ is the full MLE of all parameters of the LPPLS model and $\hat{\phi}_{t_c,m}$ is the MLE of ϕ at fixed values of $\{t_c, m\}$. Finally $\hat{s}_{t_c,m} = \text{SSE}(t_c, m, \hat{\phi}_{t_c,m})/n$ and the matrice X_m is obtained from X (36) by removing the first column and the matrix H_m is the principal submatrix of H (37), obtained by removing its first row and first column. Targeting ω, the expression for $L_m(\omega; t_c)$ is also given by (40) up to a replacement of m by ω, where X_ω is obtained from X by removing the second column, and H_ω is obtained from H by removing the second column and second row.

Figure (4) presents the profile and modified profile likelihoods for the parameters m and ω in the case considered before (Figures 1–3) for the fixed value of $t_c =$2015-06-17. It is interesting to note that the SSE profile of parameter m at a fixed t_c is unimodal in the range of interest. Moreover, our tests show that this is typically the case for a broad range of values $0 < m \lesssim 3$. The SSE profile for ω is multimodal, but when the price trajectory exhibits a clear upward trend with a substantial price appreciation over the window of calibration $[t_1, t_2]$ (e.g. when the price increase is substantially larger than the volatility), then the best solution $\hat{\omega}$ is often clearly delineated and the likelihood profile is essentially unimodal, i.e. the alternative solutions are implausible (as in Figure 4).
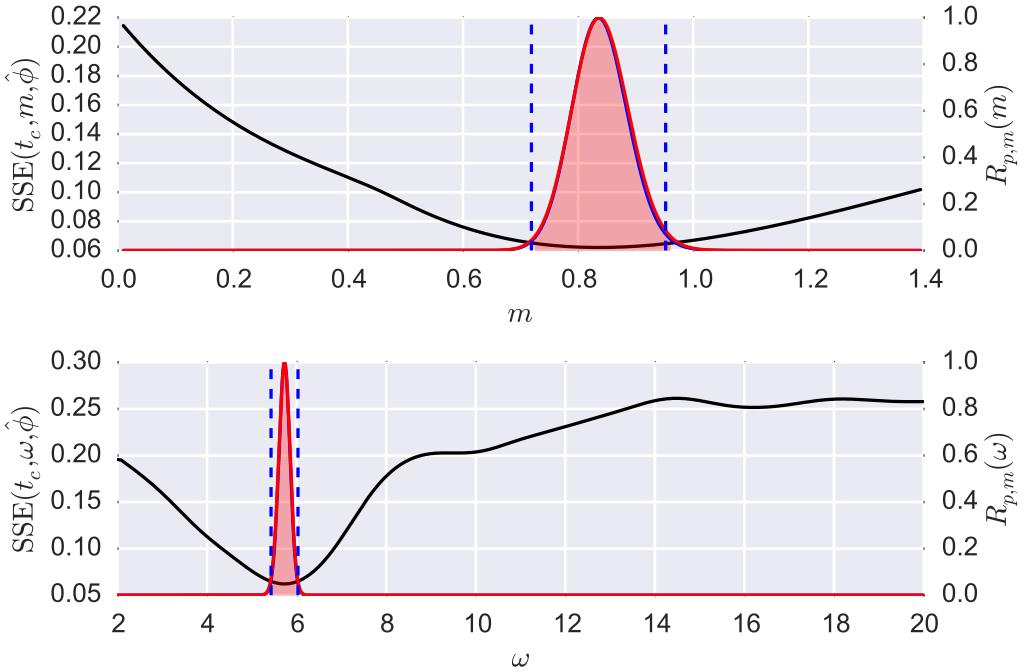
20

Figure 4: Profile of the cost function $F(\cdot)$ (black line, left scale), profile likelihood $L_p(\cdot)$ (blue line, right scale) and modified profile likelihood (red line, right scale) for the power law exponent m (top panel) and logperiodic angular frequency ω (bottom panel) for $t_c =$2015-06-17. Note that the profile and modified profile likelihood almost coincide. The red shaded intervals represent the likelihood intervals LI($\cdot$) at the 5% cutoff. The vertical blue dashed lines delineate the approximated likelihood intervals (42) at the 5% cutoff.

In Figure (4), it is almost impossible to distinguish the profile likelihood from the modified profile likelihood in the visible range of values. The values at which the log-likelihoods start to disagree, i.e. for $R_m(\cdot; t_c) = \ln L_m(\cdot; t_c)/ \ln L_m(\hat{\cdot}; t_c) \lesssim -15$, cannot be been seen in this linear scale representation. We have found that this situation is typical for many other cases. This very close agreement means that the profile likelihood $L_p(\cdot; t_c)$ is already a good approximation to either the marginal or the conditional likelihood so that we could use it directly for the inference of likelihood intervals. Moreover, the peak of the profile likelihood can often be well approximated by a quadratic function, allowing use to use this approximation for an analytical evaluation of LI[2].

In contrast to the estimated likelihood, the negative curvature of the profile likelihood function of a parameter η_i is not equal to $[I(\hat{\eta}_{t_c})]_{i,i}$, where I is the observed Fisher information matrix (A.6), but to $([I^{-1}(\hat{\eta}_{t_c})]_{i,i})^{-1}$ (see e.g. derivations in (Held and Bové, 2013)). One can prove that $[I(\hat{\eta}_{t_c})]_{i,i} \geq ([I^{-1}(\hat{\eta}_{t_c})]_{i,i})^{-1}$, which means that

---

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

the observed Fisher information of the profile likelihood is smaller than or equal to the observed Fisher information on the estimated likelihood. This illustrates the fact that the nuisance parameter η has to be estimated and thus adds to the uncertainty of the parameter of interest. Taking this approximation of the curvature into account, we can write the following Taylor expansion for the profile likelihood of m and ω at a fixed t_c:

$$
\begin{aligned}
\ln L_p(m; t_c) &\approx \ln L(t_c, \hat{\eta}_{t_c}) - \tfrac{1}{2}([I^{-1}(\hat{\eta}_{t_c})]_{1,1})^{-1}(m - \hat{m}_{t_c})^2, \\
\ln L_p(\omega; t_c) &\approx \ln L(t_c, \hat{\eta}_{t_c}) - \tfrac{1}{2}([I^{-1}(\hat{\eta}_{t_c})]_{2,2})^{-1}(\omega - \hat{\omega}_{t_c})^2,
\end{aligned}
\tag{41}
$$

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

$$
\begin{aligned}
\mathrm{LI}(m; t_c) &= \left\{ m : \tfrac{L_p(m; t_c)}{L_p(\hat{m}; t_c)} > c \right\} &= \{m : |m - \hat{m}_{t_c}| < \Delta_{m; t_c}\}, & \quad \Delta_{m; t_c} &= \sqrt{-2 \ln c \, [I^{-1}(\hat{\eta}_{t_c})]_{1,1}}, \\
\mathrm{LI}(\omega; t_c) &= \left\{ \omega : \tfrac{L_p(\omega; t_c)}{L(\hat{\omega}; t_c)} > c \right\} &= \{\omega : |\omega - \hat{\omega}_{t_c}| < \Delta_{\omega; t_c}\}, & \quad \Delta_{\omega; t_c} &= \sqrt{-2 \ln c \, [I^{-1}(\hat{\eta}_{t_c})]_{2,2}}.
\end{aligned}
\tag{42}
$$

Here, $I(\hat{\eta}_{t_c})$ has the form (A.6) (Appendix A), and its submatrix of partial derivatives can be written in a matrix form similar to the numerator in (38). These likelihood intervals for $c = 0.05$ are indicated with dashed vertical lines in Figure 4, and one can see that they provide a very good approximation for the true LIs based on the modified profile likelihood for m and ω at a fixed t_c (red shaded areas).

The likelihood interval for the damping parameter $D = m|B|/\omega|C|$ is slightly more difficult to calculate. Because D does not enter LPPLS expression (6) directly, we first need to perform a variable change, e.g. by replacing the vector $\eta = \{m, \omega, A, B, C_1, C_2, s\}$ with $\zeta = \{D, \omega, A, B, C_1, C_2, s\}$. Under such reparametrization, the observed Fisher information matrix (A.6) is transformed into

$$
I(\zeta) = J_D^T I_\eta(\eta(\zeta)) J_D,
\tag{43}
$$

where $J_D = \partial\eta/\partial\zeta$ is the Jacobian matrix of the transform from η to ζ, whose its full expression is given by (C.1) in Appendix C. Finally, the likelihood interval for the damping parameter is

$$
\mathrm{LI}(D; t_c) = \left\{ D : |D - \hat{D}_{t_c}| < \Delta_{D; t_c} \right\}, \quad \Delta_{D; t_c} = \sqrt{-2 \ln c \left[I^{-1}(\hat{\zeta}_{t_c}) \right]_{1,1}}.
\tag{44}
$$

As discussed above, the modified profile likelihood (22) for the main parameter t_c of interest is invariant with respect to such transformations of the nuisance parameter vector η.

We are now in position to discuss the overall results presented in Figure 3. The first important observation is that, in view of the determined likelihood intervals, the rejection of the "suboptimal" solution is no more warranted (given that $\mathrm{LI}(\omega) = \{5.47 < \omega < 6.07\}$). Observe that the optimal solution now easily fits in the extended interval of the damping parameter constraint (7). Second, it is interesting to compare the interval widths (2Δ) representing the uncertainty of the different parameters. In the particular example presented in Figure 3, the power law exponent m is the most uncertain parameter with $2\Delta_{m; t_c} \approx 0.3$, which is about 40% of the estimated value $\hat{m}_{t_c}$. The damping

22

parameter D, which is proportional to m, also has a fairly broad likelihood interval with width $2\Delta_{D;t_c} \approx 0.17$, which is about 20% of the estimated value $\hat{D}_{t_c}$. Finally, the uncertainty of the logperiodic angular frequency is $2\Delta_{\omega;t_c} \approx 0.63$, which is about 7% of $\hat{\omega}_{t_c}$. In general, the widths 2Δ of the likelihood intervals strictly depend on the specific realisation of the data, but our extensive tests have shown that the above observations typically hold. Finally, it is interesting to document that such rather large uncertainty in the nuisance parameters does not result in a dramatic change of the likelihood intervals for the parameter of interest t_c. And while the modified profile likelihood corrects the shape of the distribution, the intervals (39) for the profile and modified profile likelihoods at a 5% cutoff agree rather well in this and many another cases.

7. Application of the methodology

In the previous Sections 4–6, we have developed a framework to infer the critical time t_c from the LPPLS model, which includes parameter estimation together with its confidence interval, as well as the confidence intervals of the relevant nuisance parameters *within a fixed calibration window* $[t_1, t_2]$. However, for real life applications, one cannot limit oneself to the analysis of a single time-scale, because financial time-series result from complex generating processes, from volatility clustering of the simplest form to multifractal models, subjected to regime-shifts leading to non-trivial scaling structures. In order to understand the complexity of such phenomena through the prism of some model like LPPLS, one needs to apply this model at different scales simultaneously, and also consider the evolution of the model parameters in time.

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

7.1. Aggregation of time-scales

By time scale, we mean the width $\Delta_t \equiv t_2 - t_2$ of the time window in which the analysis is performed. The aggregation of analyses performed at different time-scales is not a trivial problem, whose difficulty starts with the mere computational complexity of non-linear models. Usually, the application of the model at several time scales proportionally increases the computational time, and the output data that needs to be analyzed also increases manifold. Further, in order to make the analysis operational, one needs a method for aggregating the massive amount of parameter information for the construction of the predictive features or signals. Then, the next step is to perform a full-scale back-testing of the constructed signals for understanding their predictive power. These challenging operational steps go beyond the scope of the present methodological paper, and will be reported elsewhere. Some practical aspects are already discussed in (Sornette et al., 2015), where multi-scale signals were used for ex-ante forecasting the crash in Chinese markets in June 2015. Sornette and Zhou (2006) also presented a multi-scale analysis with LPPLS, in which the different scales were

combined via a pattern recognition algorithm. Here, we will focus on the descriptive analysis and visualization aspects.

The analysis of the modified profile likelihood in Sections 5–6 was aimed at estimating the likelihood intervals (LI) of the critical time t_c as well as of the logperiodic angular frequency ω, power law exponent m and damping D, which contribute to the filtering criteria (7). A multi-scale approach would require analysis of these outputs for different values of window sizes $\Delta_t = t_2 - t_1$. One of the most natural ways is to construct the modified profile likelihood $L_m(t_c; \Delta_t)$ independently for different window sizes Δ_t. Because the absolute value of the likelihood depends on the amount of data, it does not make sense to compare value of L_m for different Δ_t directly. For comparison, we will use the normalization as in (21), and will apply it for each window size Δ_t independently, constructing the *relative multi-scale modified profile likelihood* $R(t_c, \Delta_t) = L_m(t_c; \Delta_t)/\max_{t_c} L_m(t_c; \Delta_t)$. The structure of $R(t_c, \Delta_t)$ then directly provides with scale-dependent likelihood intervals $\mathrm{LI}(t_c; \Delta_t)$ for the critical time t_c.
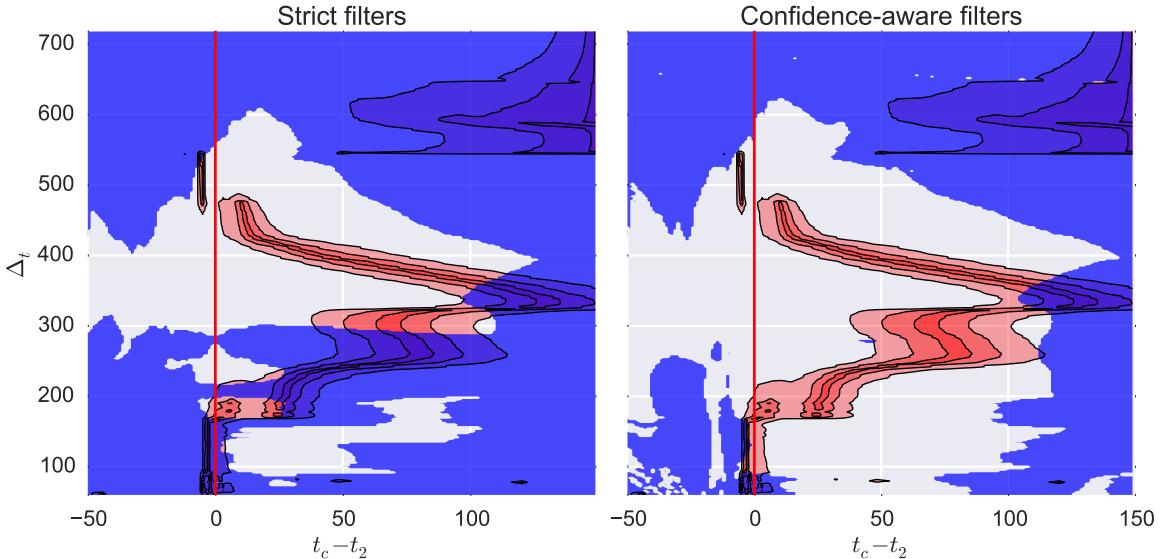

Figure 5: Two-dimensional structure of the relative multi-scale modified profile likelihood $R(t_c, \Delta_t)$ at the date t_2=2015-06-12, corresponding to multiple calibrations of the model with different window sizes Δ_t. The horizontal axis gives the value $t_c - t_2$ with the solid red vertical line indicating the case where t_c is coincident with the date of analysis. Each horizontal slice of the plot gives in color code the dependence of the individual modified profile likelihoods of the model (as in Figure 4) calculated for a given window size $\Delta_t = t_2 - t_1$ (vertical axis) as a function of $t_c - t_2$. The shaded red area corresponds to likelihood intervals $\mathrm{LI}(t_c)$ of the critical time at the 5%, 50% and 95% cutoff (from lightest to darkest colors). The shaded blue area denotes values of (t_c, Δ_t), where the constraints on the nuisance parameters (7) are not met. The left panel corresponds to the case when only MLE parameters are considered for filtering, the right panel corresponds to the case when likelihood intervals (42) and (44) are taken into account.

In order to illustrate this approach, we construct modified profile likelihoods for Δ_t varying from 60 to 700 days and for t_c varying from $t_2 - 50$ to $t_2 + 150$ days. The

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

Figure 5 also provides important insights on the range of values $(t_c; \Delta_t)$ for which the parameters obey the theoretical constraints (7) — below we will refer to them as *qualified fits*. In the left panel, the blue shaded area indicates when LPPLS fits can be rejected based solely on the MLE values of m, ω and D ("strict filtering"). In the right panel, we take additionally into account the likelihood intervals of these parameters (see Section 6) and show only the region where these intervals have no overlap with the constraints (7) ("confidence-aware filtering"). These cases differ quite dramatically, in the sense that strict filtering falsely rejects a substantial number of fits that correspond to credible alternative scenarios. As discussed in (Sornette et al., 2015), choosing the proper filters is one of the key ingredients for constructing successful signals. Being a very broad subject, constructing and testing useful filters goes beyond the scope of present paper. For the time being, we stress how crucial it is to take into account data-induced uncertainty when constructing signal filters.

Let us now describe some of the potential numerical issues that often arise in such complicated optimization problems. First, because the search for t_c is constrained in a pre-defined bounded interval, the real $\max_{t_c} L_m(t_c; \Delta_t)$ might lie outside of it, so that the numerical procedure might pick up a value at the boundary of the search space on t_c. Normalising $L_m(t_c)$ to 1 at this boundary point, this may result in having a wide range of high values of $R(t_c, \Delta_t)$ close to this boundary, leading to a spurious likelihood interval LI(t_c). In the example above, this is exactly what happens for $\Delta_t > 550$ (red profiles at the top-right of Figure 5), where the maximum of the modified profile likelihood is beyond the search range ($t_c > t_2 + 150$) and the inference on the likelihood intervals is completely misleading.

Another problem is the potential bad convergence of the optimization of the nuisance parameters in (18), which dramatically affects the value of $L_p(t_c)$ and thus $L_m(t_c)$. Usually, this situation occurs for large values of $t_c - t_2$, especially when $t_c - t_2$ is not small compared with the window size Δ_t. However, it highly depend on the structure of the residuals and the numerical method might not converge even for moderate values of $t_c - t_2$. What makes this issue complicated is that there is no simple way of detecting bad convergence, neither algorithmically nor even visually in plots like Figure 5. It often results in some kind of discontinuities in the plot, but not always. Take for instance the case $\Delta_t = 470 - 480$, where an apparent discontinuity of likelihood intervals is in

fact the consequence of a continuous transition of one maximum of the likelihood to another when increasing Δ_t.

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

7.2. Synthetic tests

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

$$\ln[P(t)] = LPPLS(t) + \sigma\epsilon(t), \tag{45}$$

where $LPPLS(t)$ is given by (4) with t_c^0=1975-02-09, $m^0 = 0.8$, $\omega^0 = 9$, $\phi^0 = 0$ and $A^0 = 8$, $B^0 = -0.015$, $C^0 = 0.0015$ (i.e. with low damping $D^0 = 0.88$), $\epsilon(t)$ is an iid N(0,1) noise and $\sigma^0 = 0.03$. The resulting price trajectory is illustrated in Figure 6.
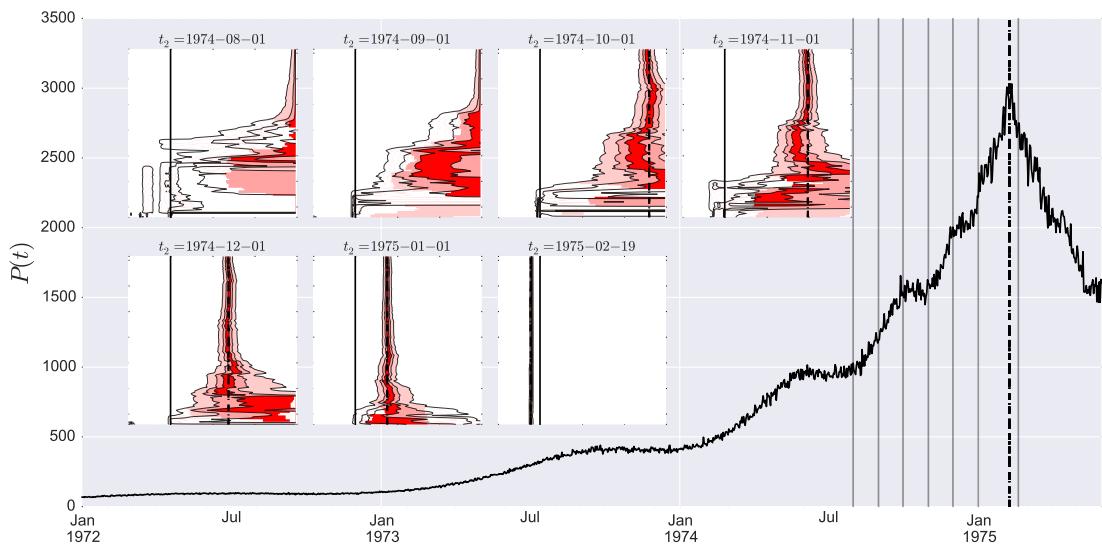

Figure 6: Synthetic price time series (45) together with multi-scale modified profile likelihoods calculated at various dates t_2. Each inset shows contour plots of the likelihood intervals at 5%, 50% and 95% cutoff levels (as in Figure 5). The red shaded area denotes values of (t_c, Δ_t) where the constraints on the nuisance parameters (7) are met when the likelihood intervals (42) and (44) are taken into account (i.e. red color denote these parts of the contour plots that are not covered by blue area in Figure 5). The solid vertical line corresponds to $t_c = t_2$, and the dashed vertical line shows the true critical time $t_c = t_c^0$. Values of t_2 used for the analyses are indicated in the inset titles. They are also shown with vertical gray lines in the plot with the price trajectory.

With the goal of understandings the evolution of the parameters as a function of the "present" time t_2 (the time of analysis) for such a synthetic bubble, we apply our

26

methodology to construct a multi-scale modified profile likelihood (see Section 7.1 and Figure 5) at different dates t_2 increasing towards the end of the bubble at t_c^0. The resulting multi-scale profiles are shown in insets of Figure 6. Far from the critical time (insets 1 and 2: $t_c^0 - t_2 = 192$ and 161 days respectively), the critical time cannot be identified even in such a clean synthetic case with weak noise (partially because we are limiting the search space to $t_c < t_2 + 150$ days). Even when $t_c^0 - t_2$ enters the range < 150, the parameters continue to exhibit a large uncertainty. One can observe that fits for different scales progressively build a consensus, as the likelihood peaks aggregate around the true critical time with a narrow likelihood interval around it. This is first observed for the large scales $\Delta_t > 300 - 400$ (insets 3 and 4 for $t_c^0 - t_2 = 131$ and 100 days respectively), and this consensus spreads to smaller scales of $\Delta_t \sim 100 - 300$ (insets 5 and 6 for $t_c^0 - t_2 = 70$ and 39 days respectively).

Another remarkable fact is that, once the critical time is passed and the price trajectory switched to a crashing regime ($t_2 > t_c^0$, inset 7), all scales confirm this occurrence by fixing the MLE $\hat{t}_c \approx t_c^0$ with an extremely narrow likelihood interval, and this anchoring holds for a large time interval. The same effect is observed in the analysis of real data presented in Section 7.3 — even when the ex-ante forecast of the end of the bubble might be difficult or inconclusive, the change of the price direction can be identified quite reliably within a few days of the switching point.

7.3. Case-studies

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

We use the daily closing prices of the S&P 500, NASDAQ and SSEC indices provided by Thomson Reuters Dataworks Enterprise (DWE). We only consider business days, ignoring weekends and one-day holidays. However, for extended holidays (such as the Chinese New Year in 2015, when exchanges were closed over February 7-13), we fill the gaps with the closing price of the previous day. For calibrations using business time (see Section 2), such data preprocessing would not be necessary.

Employing the procedure explained in Sec. 7.2, we obtain Figures 7-9. In each of these three figures, the main graph shows the price time series $P(t)$ together with vertical dashes lines that identify remarkable turning points of the price dynamics. In the case of the S&P 500, we show two different vertical dashed lines associated with the two peaks of the index preceding the crash. The seven thin vertical lines indicate the position of the seven t_2 values chosen for the construction of the Likelihood intervals of t_c. The seven insets show contour plots of the likelihood intervals at 5%, 50% and 95% cutoff levels (as in Figure 6).

Figures 7 for the S&P 500 shows that the Profile Likelihhod of t_c as a function of time scale $\Delta_t \equiv t_2 - t_1$ and "present time" t_2 is very similar to those obtained in synthetic tests. As early as $t_2 = $ 1987-04-15, one can visualize the high Likelihood

27

of $\widehat{t}_c \approx Oct.$ 1987 over almost all time scales. Interestingly, the Likelihood interval narrows down as t_2 approaches the end of the bubble. Moreover, there is an increase of the number of qualified fits (those where constraints on the model parameters (7) are met when the likelihood intervals (42) and (44) are taken into account — shown as the red-shaded region) at t_2 increases. These two results can be rationalized by the fact that more information relevant to the identification of the bubble become available as more data are used.

As shown in figure 8, similar observations carry over to the SSEC bubble ending in June 2015, albeit with a smaller number of qualified fits. One can observe that the analyses performed for the time scales $\Delta_t \in [400, 500]$ and $\Delta_t \in [100, 200]$ provide a correct diagnostic of the end of the bubble with a narrow confidence interval. The time scale $\Delta_t \in [400, 500]$ correctly locks in on the true peak as early as April 2015.

Figure 9 shows the same analysis for the dotcom bubble that developed in the NASDAQ Index. At $t_2 = 2000\text{-}02\text{-}09$, the time scales $\Delta_t \in [100, 350]$ correctly lock in on the true peak $\approx$ April 2000. The other intermediate time scales give an estimation $\widehat{t}_c$ of the end of the bubble that agrees with the empirical value within the 95% confidence of the likelihood intervals.. All estimates on different t_2's appear to be either unqualified or signalling a different value for the change of regime to occur.

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

On the other hand, the fact that different time scales used for fitting the LPPLS model tend to suggest different values of $\hat{t}_c$ is important to keep in mind, as this observation is in contrast with the behavior obtained for synthetic time-series. This is likely due to "model error", i.e., the simple LPPLS model (3) is only an approximation of the unknown true generating process of the price dynamics. For instance, earlier works (Sornette and Johansen, 1997; Johansen and Sornette, 1999a; Gluzman and Sornette, 2002; Zhou and Sornette, 2003b) have pointed out the important of including higher harmonics and more complex forms generalising this simple first-order LPPLS formula (3).

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

8. Concluding remarks

We have presented a detailed methodological study of the application of the modified profile likelihood method proposed by Barndorff-Nielsen (1983), with the goal of tackling the instabilities and uncertainties occurring in the calibration of nonlinear fi-
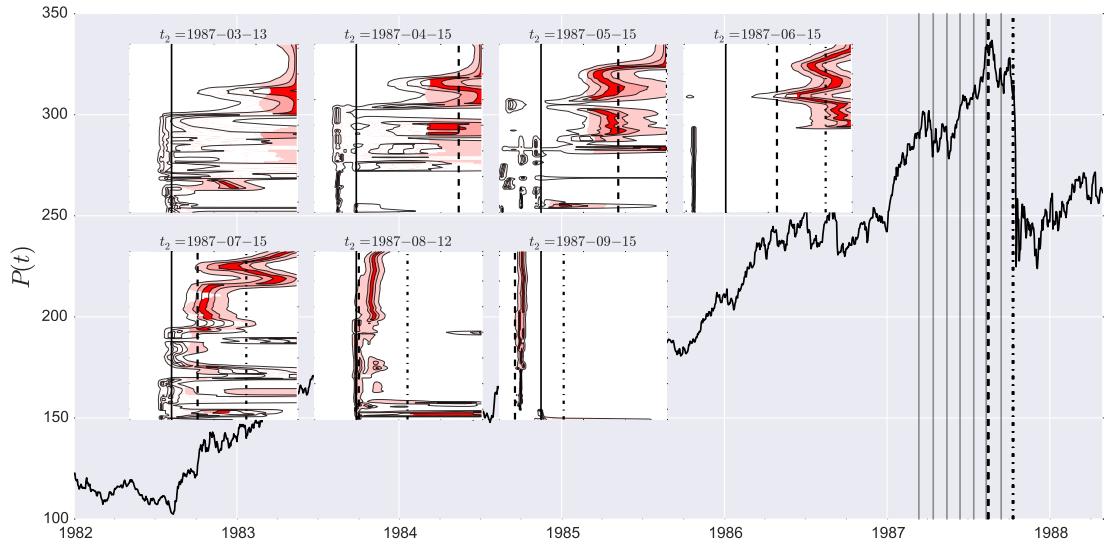
Figure 7: Same as Figure 6 for the $S\&P500$ index that shows a strong bubble of US markets developing in the second half of the 1980's, which culminated with the Black Monday crash of Oct. 19. 1987.
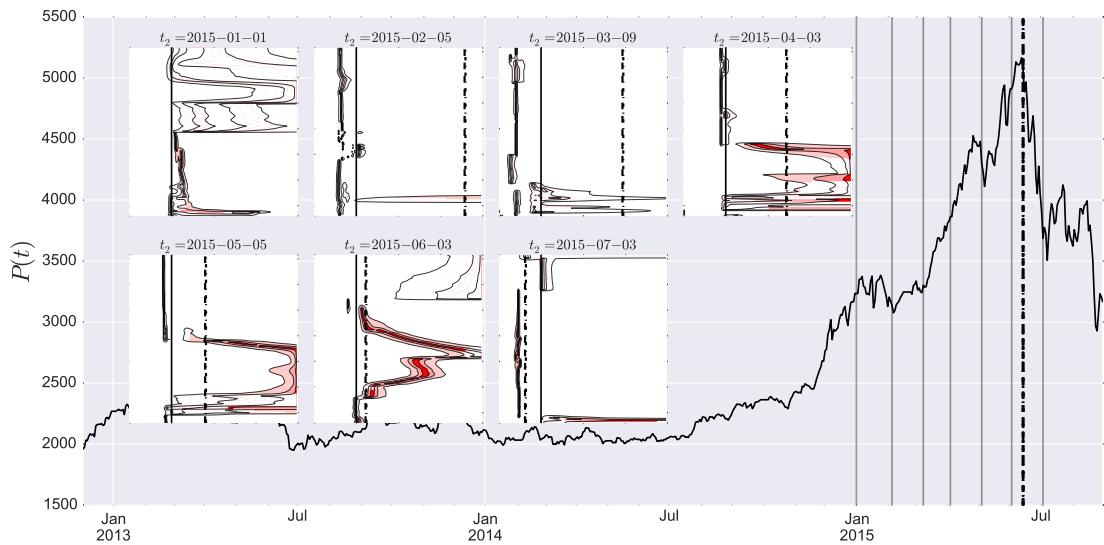

Figure 8: Same as Figure 6 for the Chinese bubble of 2014-2015 that peaked in June 2015.
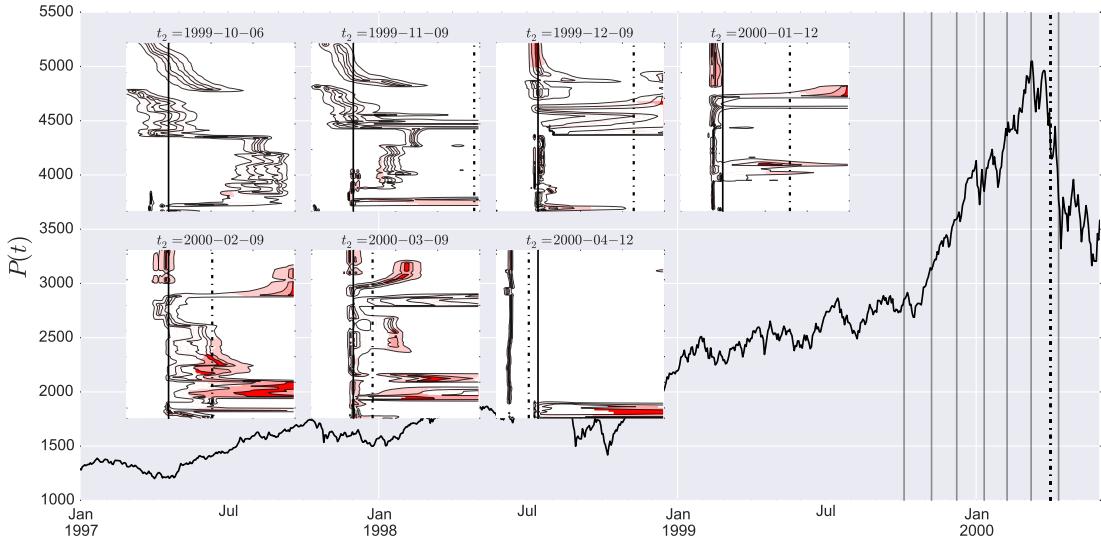
29

Figure 9: Same as Figure 6 for the dot-com bubble in the IT sector in the US culminating with a crash in April 2000

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

Acknowledgments

Alvarez for many fruitful discussions while preparing the manuscript.

Appendix A. Derivation of the approximated modified profile likelihood for the LPPLS model

We derive the approximated expression for the modified profile likelihood (34) of the LPPLS model (6). The parameter of interest is the critical time t_c and nuisance parameters $\eta = \{\psi, s\}$ include both the vector $\psi = \{m, \omega, A, B, C_1, C_2\}$ of other LPPLS parameters and the variance s of the error term.

Appendix A.1. Observed Fisher information

The calculation of the observed Fisher information matrix $I(\hat{\eta}_{t_c})$ is straightforward. According to (23), it can be written in the form of a block matrix:

$$I(\hat{\eta}_{t_c}) = - \begin{pmatrix} \ell_{\psi,\psi}(\hat{\eta}_{t_c}) & \ell_{\psi,s}(\hat{\eta}_{t_c}) \\ \ell_{s,\psi}^T(\hat{\eta}_{t_c}) & \ell_{s,s}(\hat{\eta}_{t_c}) \end{pmatrix}, \tag{A.1}$$

where $\ell_{\cdot,\cdot}(\hat{\eta}_{t_c})$ denotes the respective second partial derivatives of the log-likelihood $\ln L(t_c, \psi, s)$ evaluated at the point $\hat{\eta}_{t_c} = \{\hat{\psi}_{t_c}, \hat{s}_{t_c}\}$. For the likelihood (16), the first partial derivatives are given by:

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

In turn, the second partial derivatives read:

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

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

$$\ell_{\psi,s}(\hat{\eta}_{t_c}) = \left. \frac{\partial^2 \ell}{\partial \psi \partial s} \right|_{\eta = \hat{\eta}_{t_c}} = \frac{1}{2\hat{s}_{t_c}^2} \left. \frac{\partial \text{SSE}(t_c, \psi)}{\partial \psi} \right|_{\psi = \hat{\psi}_{t_c}} = \Theta, \tag{A.4}$$

where $\Theta = \{0, 0, 0, 0, 0, 0\}^T$ is the vector of zeros. Taking into account (19), we can write for the third term in (A.3):

$$\ell_{s,s}(\hat{\eta}_{t_c}) = \left. \frac{\partial^2 \ell}{\partial s^2} \right|_{\eta=\hat{\eta}_{t_c}} = \frac{n}{2\hat{s}_{t_c}^2} - \frac{\text{SSE}(t_c, \hat{\psi}_{t_c})}{\hat{s}_{t_c}^3} = -\frac{n}{2\hat{s}_{t_c}^2}. \tag{A.5}$$

Finally, plugging (A.4) and (A.5) into (A.1), we obtain the following form for the observed Fisher information:

$$I(\hat{\eta}_{t_c}) = \begin{pmatrix} \left. \frac{1}{2\hat{s}_{t_c}} \frac{\partial^2 \text{SSE}(t_c, \psi)}{\partial \psi \partial \psi^T} \right|_{\psi=\hat{\psi}_{t_c}} & \Theta \\ \Theta^T & \frac{n}{2\hat{s}_{t_c}^2} \end{pmatrix} = \begin{pmatrix} \left. \frac{1}{\hat{s}_{t_c}} \sum_{i=1}^{n} \frac{\partial^2 \text{LPPLS}(\tau_i; t_c, \psi)}{\partial \psi \partial \psi^T} \right|_{\psi=\hat{\psi}_{t_c}} & \Theta \\ \Theta^T & \frac{n}{2\hat{s}_{t_c}^2} \end{pmatrix}, \tag{A.6}$$

and its determinant

$$|I(\hat{\eta}_{t_c})| = \frac{n}{2\hat{s}_{t_c}^{p+2}} \left| \sum_{i=1}^{n} \frac{\partial^2 \text{LPPLS}(\tau_i; t_c, \psi)}{\partial \psi \partial \psi^T} \right|_{\psi=\hat{\psi}_{t_c}}, \tag{A.7}$$

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

Appendix A.2. Covariance matrix

Here, we calculate the covariance matrix $\Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})$ (31). For this, we will first evaluate the general form of the matrix (31) and then substitute $(t_c, \hat{\eta}_{t_c})$ and $(\hat{t}_c, \hat{\eta})$. Similarly to the Fisher information, the matrix Σ (31) can be written in a block form:

$$\Sigma(t_{c;1}, \eta_1; t_{c;2}, \eta_2) = \text{E}_{(2)} \left[\begin{pmatrix} \ell_\psi(1)\ell_\psi^T(2) & \ell_\psi(1)\ell_s(2) \\ \ell_s(1)\ell_\psi^T(2) & \ell_s(1)\ell_s(2) \end{pmatrix} \right], \tag{A.8}$$

where $\ell.(1)$ symbolically denotes the first partial derivative (A.2) of the log-likelihood evaluated at $(t_{c;1}, \eta_1) = (t_{c;1}, \psi_1, s_1)$; similarly, $\ell.(2)$ is evaluated at $(t_{c;2}, \eta_2) = (t_{c;2}, \psi_2, s_2)$. Given (9), the partial derivative of the SSE that enters (A.2) has the form:

$$\frac{\partial \text{SSE}(t_c, \psi)}{\partial \psi} = -2 \sum_{i=1}^{n} \varepsilon_i \frac{\partial \text{LPPLS}_i}{\partial \psi}, \tag{A.9}$$

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

Let us first consider the cross-terms in (A.8). We substitute (A.9) into (A.2) and then into (A.8). Then, after replacing the product of sums with the double sum and using the linearity of the expectation operation, we have:

$$\text{E}_{(2)}[\ell_\psi(1)\ell_s(2)] = -\frac{n}{2s_1 s_2} \sum_{i=1}^{n} \text{E}_{(2)}[\varepsilon_i] \left. \frac{\partial \text{LPPLS}_i}{\partial \psi} \right|_{\substack{t_c=t_{c;1} \\ \psi=\psi_1}} + \frac{1}{2s_1 s_2^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{E}_{(2)}[\varepsilon_i \varepsilon_j^2] \left. \frac{\partial \text{LPPLS}_i}{\partial \psi} \right|_{\substack{t_c=t_{c;1} \\ \psi=\psi_1}}. \tag{A.10}$$

As discussed in Section 5.3, the expectations in (A.10) are taken with respect to the probability distribution that corresponds to the parameters $\{t_{c;2}, \eta_2\}$, in other words under the assumption that $\varepsilon \sim N(0, s_2)$. Thus $\mathrm{E}_{(2)}[\varepsilon_i] = \mathrm{E}_{(2)}[\varepsilon_i \varepsilon_j^2] = 0$, and the cross-term is equal to the zero-vector: $\mathrm{E}_{(2)}[\ell_\psi(1)\ell_s(2)] = \Theta$. Similarly for the second cross-term of (A.8):

$$\mathrm{E}_{(2)}[\ell_s(1)\ell_\psi^T(2)] = \left(\mathrm{E}_{(2)}[\ell_\psi(1)\ell_s(2)]\right)^T = \Theta^T. \tag{A.11}$$

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

$$\mathrm{E}_{(2)}[\ell_s(1)\ell_s(2)] = \frac{n^2}{4s_1 s_2} - \frac{n}{4s_1^2 s_2} \sum_{i=1}^{n} \mathrm{E}_{(2)}[\varepsilon_i^2] - \frac{n}{4s_1 s_2^2} \sum_{j=1}^{n} \mathrm{E}_{(2)}[\varepsilon_j^2] + \frac{1}{4s_1^2 s_2^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}_{(2)}[\varepsilon_i^2 \varepsilon_j^2]. \tag{A.12}$$

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

$$\mathrm{E}_{(2)}[\ell_s(1)\ell_s(2)] = \frac{n}{2s_1^2}. \tag{A.13}$$

Finally, the submatrix term reads:

$$\begin{aligned}
\mathrm{E}_{(2)}[\ell_\psi(1)\ell_\psi^T(2)] &= \frac{1}{s_1 s_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}_{(2)}[\varepsilon_i \varepsilon_j] \left. \frac{\partial \mathrm{LPPLS}_i}{\partial \psi} \right|_{\substack{t_c = t_{c;1} \\ \psi = \psi_1}} \left. \frac{\partial \mathrm{LPPLS}_j}{\partial \psi^T} \right|_{\substack{t_c = t_{c;2} \\ \psi = \psi_2}} \\
&= \frac{1}{s_1} \sum_{i=1}^{n} \left. \frac{\partial \mathrm{LPPLS}_i}{\partial \psi} \right|_{\substack{t_c = t_{c;1} \\ \psi = \psi_1}} \left. \frac{\partial \mathrm{LPPLS}_i}{\partial \psi^T} \right|_{\substack{t_c = t_{c;2} \\ \psi = \psi_2}},
\end{aligned} \tag{A.14}$$

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

The final expression is obtained by combining (A.11),(A.13) and (A.14) to (A.8) and evaluating the result at $(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})$:

$$\begin{aligned}
|\Sigma(t_c, \hat{\eta}_{t_c}; \hat{t}_c, \hat{\eta})| &= \left| \begin{matrix} \dfrac{1}{\hat{s}_{t_c}} \sum\limits_{i=1}^{n} \left. \dfrac{\partial \mathrm{LPPLS}_i}{\partial \psi} \right|_{\substack{t_c = t_c \\ \psi = \hat{\psi}_{t_c}}} \left. \dfrac{\partial \mathrm{LPPLS}_i}{\partial \psi^T} \right|_{\substack{t_c = \hat{t}_c \\ \psi = \hat{\psi}}} & \Theta \\[2ex] \Theta^T & \dfrac{n}{2\hat{s}_{t_c}^2} \end{matrix} \right| \\[3ex]
&= \frac{n}{2\hat{s}_{t_c}^{2+p}} \left| \sum_{i=1}^{n} \left. \frac{\partial \mathrm{LPPLS}_i}{\partial \psi} \right|_{\substack{t_c = t_c \\ \psi = \hat{\psi}_{t_c}}} \left. \frac{\partial \mathrm{LPPLS}_i}{\partial \psi^T} \right|_{\substack{t_c = \hat{t}_c \\ \psi = \hat{\psi}}} \right|, \tag{A.15}
\end{aligned}$$

where $p = \dim \psi = 6$. Note that a similar expression presented in Severini (1999) contains a typographical error in the power of the variance term.

Appendix B. Partial derivatives of the LPPLS function

We present here the analytical expressions of the first and second partial derivatives of the LPPLS function (6), which are necessary for the calculation of the modified profile likelihood (35)–(37).

The first-order derivatives have the following forms:

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

(B.1)

The second-order derivatives $\partial^2 \text{LPPLS}/\partial \psi_i \partial \psi_j$, which are needed for the calculation of the matrix H (37), have the following form (omitting equivalent symmetrical entries, i.e.: $\partial^2 \text{LPPLS}/\partial m \partial \omega \equiv \partial^2 \text{LPPLS}/\partial \omega \partial m$):

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

(B.2)

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

Appendix C. Jacobian matrix for the damping parameter

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

$$
J_D = \frac{\partial \eta}{\partial \zeta} =
\begin{pmatrix}
\frac{\omega|C|}{|B|} & \frac{D|C|}{|B|} & 0 & -\frac{D\omega|C|}{B|B|} & \frac{D\omega C_1}{|B||C|} & \frac{D\omega C_2}{|B||C|} & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
\qquad \text{(C.1)}
$$

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

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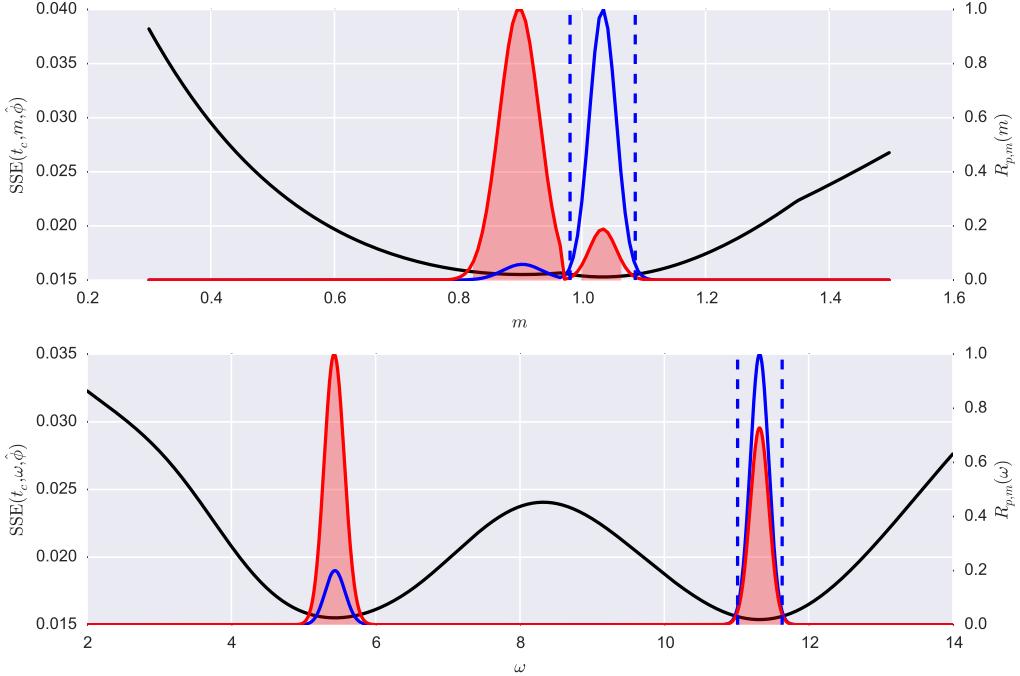

Figure D.10: Profile of the cost function $F(\cdot)$ (black line, left scale), profile likelihood $L_p(\cdot)$ (blue line, right scale) and modified profile likelihood (red line, right scale) for the power law exponent m (top subplot) and the logperiodic angular frequency ω (bottom subplot) for t_c=2007-11-20. The red shaded intervals show the likelihood intervals LI($\cdot$) at the 5% cutoff level. Vertical blue dashed lines denote approximated likelihood intervals (42) at the 5% cutoff level. The calibration of the LPPLS model is performed on the Chinese SSEC Index for the bubble that bursts in June 2015.

Figure (D.10) illustrates a situation in which the approximate likelihood intervals (42) are misleading. It presents the profile and modified profile likelihoods for the nuisance parameters m and ω obtained by calibrating the LPPLS model to the Chinese SSEC Index in the time window from t_1=2006-05-04, t_2=2007-10-31 and at the fixed t_c= 2007-11-20. In contrast to the typical situation shown in Figure 4, one can clearly observe a bi-modal structure of the profile likelihoods of the nuisance parameters. Such bi-modal structure cannot be well described by intervals derived from a Fisher information-based likelihood. Moreover, this figure illustrates a case when the second-order modified profile likelihood suggests different estimated value of m and ω compared with the standard MLE: profile and modified profile likelihood have maxima at different points (similarly to the situation of the critical time in Figure 3).

While these situations are rather rare according to our experience, one needs to be aware that the approximate relations (42) might not reflect the full complexity of the structure of residuals.

References

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Anders Johansen and Didier Sornette. Large Stock Market Price Drawdowns Are Outliers. *Journal of Risk*, 4(2):69–110, 2001/02.

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

Anders Johansen, Didier Sornette, and Olivier Ledoit. Predicting Financial Crashes Using Discrete Scale Invariance. *Journal of Risk*, 1(4):5–32, 1999.

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

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

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

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

Li Lin, Ruoen Ren, and Didier Sornette. A Consistent Model of 'Explosive' Financial Bubbles with Mean-Reversing Residuals. *International Review of Financial Analysis*, 33:210–225, 2014.

Wes McKinney. *Python for Data Analysis*. O'Reilly Media, 2012.

JA Nelder and R Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 1965.

Luigi Pace and Alessandra Salvan. Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, 136(10):3554–3564, October 2006.

Yudi Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, 2001.

Fernando Pérez and Brian E Granger. IPython: A System for Interactive Scientific Computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

J Qin. Marginal likelihood, conditional likelihood and empirical likelihood: Connections and applications. *Biometrika*, 92(2):251–270, June 2005.

Thomas A Severini. An Approximation to the Modified Profile Likelihood Function. *Biometrika*, 85(2):403–411, 1998a.

Thomas A Severini. Likelihood functions for inference in the presence of a nuisance parameter. *Biometrika*, 85(3):507–522, September 1998b.

Thomas A Severini. An empirical adjustment to the likelihood ratio statistic. *Biometrika*, 86(2):235–247, June 1999.

Thomas A Severini. *Likelihood Methods in Statistics*. Oxford Statistical Science Series. Oxford University Press, 2001.

Thomas A Severini. Integrated likelihood functions for non-Bayesian inference. *Biometrika*, 94(3):529–542, August 2007.

Ib M Skovgaard. An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, 2(2):145–165, June 1996.

Ib M Skovgaard. Likelihood Asymptotics. *Scandinavian Journal of Statistics*, 28(1): 3–32, March 2001.

Richard L Smith and J C Naylor. A Comparison of Maximum Likelihood and Bayesian Estimators for the Three- Parameter Weibull Distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):358, 1987.

Didier Sornette. Discrete-scale invariance and complex dimensions. *Physics Reports*, 297(5):239–270, 1998.

Didier Sornette. Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes and human birth. *Proceedings of the National Academy of Sciences USA*, 99(Supp. 1):2522–2529, 2002.

Didier Sornette. Dragon-Kings, Black Swans and the Prediction of Crises. *International Journal of Terraspace Science and Engineering*, 2(1), December 2009.

Didier Sornette and Anders Johansen. Large financial crashes. *Physica A: Statistical Mechanics and its Applications*, 245(3-4):411–422, 1997.

Didier Sornette and Anders Johansen. Significance of log-periodic precursors to financial crashes. *Quantitative Finance*, 1(4):452–471, 2001.

Didier Sornette and Wei-Xing Zhou. Predictability of Large Future Changes in major financial indices. *International Journal of Forecasting*, 22:153–168, 2006.

Didier Sornette, Ryan Woodard, Wanfeng Yan, and Wei-Xing Zhou. Clarifications to questions and criticisms on the Johansen–Ledoit–Sornette financial bubble model. *Physica A: Statistical Mechanics and its Applications*, 392(19):4417–4428, October 2013.

Didier Sornette, Guilherme Demos, Qun Zhang, Peter Cauwels, Vladimir Filimonov, and Qunzhi Zhang. Real-Time Prediction and Post-Mortem Analysis of the Shanghai 2015 Stock Market Bubble and Crash. *Journal of Investment Strategies*, 4(4):77–95, 2015.

El-Ghazali Talbi. *Metaheuristics: from design to implementation*. Wiley, Hoboken, NJ, 2009.

Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.

HCG van Bothmer and C Meister. Predicting critical crashes? A new restriction for the free variables. *Physica A: Statistical Mechanics and its Applications*, 320:539–547, 2003.

W Xiong. Bubbles, Crisis and Heterogeneous Beliefs. In Jean-Pierre Fouque and Joseph A Langsam, editors, *Handbook on Systemic Risk*, pages 663–713. Cambridge University Press, Cambridge, 2013.

Qun Zhang, Qunzhi Zhang, and Didier Sornette. Early warning signals of financial crises with multi-scale quantile regressions of Log-Periodic Power Law Singularities. *http://ssrn.com/abstract=2674128*, 2015.

Wei-Xing Zhou and Didier Sornette. Non-Parametric Analyses of Log-Periodic Precursors to Financial Crashes. *Int. J. Mod. Phys. C*, 14(8):1107–1126, 2003a.

Wei-Xing Zhou and Didier Sornette. of the 2000-2002 anti-bubble in the US S&P 500 index: Explanation of the hierarchy of 5 crashes and Prediction. *Physica A: Statistical Mechanics and its Applications*, 330:584–604, 2003b.