ETH zürich

Master's Thesis

Leave-out methods for selecting the optimal starting point of financial bubbles

A more flexible statistical model for the LPPLS model and leave-out methods for selecting the beginning of bubbles.

Author
S. H. Magnússon

Advisor
Dr. S. Wheatley

Supervisor
Prof. N. Meinshausen
Prof. D. Sornette

August 28, 2018

Abstract

Identifying financial bubbles and predicting the burst of them is of high theoretical but also practical interest. The Log-Periodic Power Law Singularity (LPPLS) model is an attempt to model unsustainable growth in financial markets, namely super-exponential growth, and predict the inevitable burst of such bubbles. This thesis aims to improve the statistical estimation of the LPPLS model by allowing the residuals of the parametric model to have an auto-regressive part and heteroskedasticity in the innovations. Further, new methods for selecting the optimal starting point of a bubble are investigated and compared to existing methods. Finally, the parameters in the LPPLS model need to satisfy specific constraints to describe a bubble. We extend these constraints partly to probabilistic boundaries.

The methods for selecting the optimal initial point of a bubble were tested and compared to results on synthetic and historical data. The proposed improvements on the residual structure are necessary for estimating the parameters and its confidence intervals. The suggested methods for selecting the optimal initial point of the bubble are conceptually more appealing than existing methods but require refinements. We suggest doing further experiments to compare and find the merits and drawbacks of each selection criteria.

Contents

1 Introduction

Understanding the dynamics of financial instruments has been the topic of research and debate since men started trading goods. With the introduction of modern currency, people have resolved to ever more sophisticated models to study and analyze these trends. The complex interaction between investors and exogenous shocks now coupled with increasingly many financial innovations makes the task of disentangling the signal from noise ever more challenging. The price dynamics of financial instruments can often be approximately described as exponential growth with added noise. The noise structure can be found to be quite complicated without the use of exogenous information. When this model holds true, no bubble is present, and we classify these times as regular times for that particular financial instrument. Occasionally though, we can see in financial markets that the price of an instrument grows super-exponentially as a function of time, which we classify as a bubble [7, 14, 19]. Super-exponential time dynamics occur during transient regimes when the price growth rate is growing. Such processes have been found in multiple historical bubbles [10, 11, 14, 21]. Under these circumstances, it is hardly plausible that the fundamental value is increasing at the same rate. Thus, these unsustainable super-exponential price dynamics usually lead to a crash or severe corrections, and the process swiftly changes regime. Ideally estimating the discrepancy between the price trajectory and the fundamental value would be an essential element in identifying bubbles. Formally the price of a financial instrument can be defined as

$$p_t = p_t^\star + p_t' \tag{1}$$

where $p_t^\star$ is the fundamental value at time t and p_t' is the additional bubble component at time t. However, the fundamental value is hard to estimate accurately [1, 15], and thus identifying bubbles based on deviations from fundamental value is tricky in general although perhaps achievable in some cases [23]. The Log-Periodic Power Law Singularity (LPPLS) model was proposed as a generic model for capturing bubbles [10, 11, 18] using only endogenous information, that is only prices of that particular instrument over time.

The LPPLS model is a parametric model which describes a positive feedback process which results in a finite-time singularity [8, 17, 20] with accelerating log-periodic oscillations [16]. It describes some financial bubbles in a simple but effective manner. The LPPLS model cannot explain crashes caused by exogenous shocks, for example, natural disasters.

1.1 Contributions

The contribution of the thesis starts with improving the statistical estimation of the parametric LPPLS model. The assumption of independent and identically distributed residuals is violated on real-world bubbles as shown for several distinct historical bubbles in the thesis. Thus, a more general model for the residuals is proposed, namely with auto-regressive residuals and heteroskedasticity. The proposed model yields more robust and efficient parameter estimates

and more accurate predictions, at least for synthetic data. The complex structure of the residuals needs to be accounted for to calculate confidence intervals for the parameter estimates.

Secondly, the thesis proposes partly probabilistic boundaries on the parameter space. In the parametric LPPLS model, we require the parameters to satisfy certain conditions. We derive probabilistic boundaries on the linear parameters of the model based on Bayesian regression methods. This model also describes approximate marginal and joint confidence interval for the full set of the linear parameters.

The final goal is to identify the beginning of a bubble. The initial time of bubble is a fuzzy concept and has to be dealt with in order to capture the bubble structure efficiently and robustly. The proposed model suffers from computational challenges as for each suggested initial point we need to fit the model. The likelihood of the model is non-linear and non-convex and thus requires a time-consuming search through the parameter space. Demos et al. proposed the Lagrange regularisation approach [4]. It assumes that the residuals have constant variance through time and no correlation which we show is violated in historical bubbles. It also depends on the window selected where the bubble is supposed to have started and all their corresponding fits. We suggest several improvements and compare their empirical performance. The proposed methods depend solely on prediction and out-of-sample generalization. First, we begin with simple mean-squared-error, which does not take into account the inherent heteroskedasticity of financial data. Then we try to correct for the heteroskedasticity with a generalized sum of squares and finally Kullback-Leibler divergence.

2 The LPPLS model

2.1 Theoretical derivation of the LPPLS model

When a bubble occurs, the price of a given asset decouples from its fundamental value[13, 18]. Under the Johansen-Ledoit-Sornette (JLS) model [10, 11], the asset price p_t follows a diffusive dynamics, with drift μ_t and discrete jumps described by the stochastic differential equation

$$\frac{dp_t}{p_t} = \mu_t dt + \sigma_t dB_t - \kappa dj_t \tag{2}$$

where κ is the loss amplitude associated with a crash and dB_t is the increments of a Brownian motion with zero mean and variance dt. The volatility σ_t describes the level of stochasticity at a certain time. Let t_c denote the time of the crash. Then, the discontinuous jumps are describes with $j_t = 0$ for $t < t_c$ and $j_t = 1$ for $t \geq t_c$. The dynamics of the jump process follows a hazard rate $h(t)$. Under this assumption, the conditional probability of a crash in the infinitesimal interval $[t, t + dt]$ is $h(t)dt$, given that it has not happened already. Thus we have that

by the low of total expectation

$$\mathbf{E}_t[dj_t] = \mathbf{E}_t[dj_t | t_c \in [t, t+dt]] + \mathbf{E}_t[dj_t | t_c \notin [t, t+dt]] = h(t)dt \tag{3}$$

where the expected value is with respect to the filtration set at time t. Due to the no-arbitrage condition the price process is a martingale meaning $\mathbf{E}_t[\frac{dp_t}{p_t}] = 0$ ignoring the risk free rate. Taking expected value with respect to time of the whole stochastic differential equation (2) and knowing that $\mathbf{E}_t[dB_t] = 0$ thus leads to $\mu_t = \kappa h(t)$.

The JLS model assumes that noise traders may destabilize the market due to mutual herding behaviors. According to the model, the cumulative effect of noise traders can be approximately described by the following dynamics of the crash hazard rate:

$$h(t) \approx B'|t_c - t|^{m-1} + C'|t_c - t|^{m-1} \cos(\omega \ln|t_c - t| + \phi'). \tag{4}$$

Here, $t < t_c$, t_c is the most probable time of the crash. Using the equation $\mu_t = \kappa h(t)$ and integrating from $t < t_c$ to t_c the expected value process yields the average log-price series, the LPPLS model,

$$\mathbf{E}[\ln p_t] = A + B|t_c - t|^m + C|t_c - t|^m \cos(\omega \ln|t_c - t| + \phi). \tag{5}$$

It is important to note that model is totally non-commital about what happens after t_c. Finally, Filiminov and Sornette [5] suggested a more stable representation of the model in equation (5), transforming the model to 4 linear parameters and 3 non-linear. Recall the trigonometry law:

$$\cos(u + v) = \cos(u)\cos(v) - \sin(u)\sin(v). \tag{6}$$

Using the rule, we get

$$C\cos(\omega \ln|t_c - t| + \phi) = C\cos(\omega \ln|t_c - t|)\cos(\phi) - C\sin(\omega \ln|t_c - t|)\sin(\phi). \tag{7}$$

Set $C_1 = C\cos\phi$ and $C_2 = -C\sin\phi$ yields

$$\begin{aligned}\mathbf{E}[\ln p_t] = A + B|t_c - t|^m + C_1|t_c - t|^m \cos(\omega \ln|t_c - t|) \\ + C_2|t_c - t|^m \sin(\omega \ln|t_c - t|).\end{aligned} \tag{8}$$

Here, $A = \mathbf{E}[\ln p_{t_c}]$ is the log-price at most probable time t_c of the burst of the bubble, $B = -\frac{B'\kappa}{m}$ quantifies the amplitude of the power law acceleration and $C = -\frac{C'\kappa}{\sqrt{m^2+\omega^2}}$ describes the amplitude of the log-periodic oscillations. In the final calibration, $C_1 = -\frac{C'\kappa\cos\phi}{\sqrt{m^2+\omega^2}}$ and $C_2 = \frac{C'\kappa\sin\phi}{\sqrt{m^2+\omega^2}}$. The crucial parameter m captures the rate of super-exponential growth and describes the acceleration of the process. ω is the log-periodic angular frequency and $\phi \in [0, 2\pi)$ is the phase correction describing the time scale of the oscillations. In what follows we set $\alpha = (m, \omega, t_c)^T$, the set of non-linear parameters in equation (8), $\beta = (A, B, C_1, C_2)^T$ is the set of linear parameters in the same representation and $\theta = (\alpha, \beta)$ is the full vector of parameters representing the expectation of the logarithm of prices process.

2.2 Natural constraints and search space

For a bubble to be present we require some theoretical constraints on our parameters [5, 9, 12, 22]. For the linear parameters in the model we require the constraints

$$A > 0 \tag{9}$$

$$B < 0 \tag{10}$$

$$\sqrt{C_1^2 + C_2^2} \leq |B|. \tag{11}$$

Furthermore, van Bothmer and Meister [2] derived a constraint on the variables of the JLS model from the statement that the crash rate should be non-negative:

$$-Bm - \sqrt{C_1^2 + C_2^2}\sqrt{m^2 + \omega^2} \geq 0. \tag{12}$$

The subspace $S \in \mathbb{R}^4$ is defined as the set satisfying equations (9)-(12).

Suppose we have a fitting window beginning at time τ_1 and ending at t_2 with $\Delta = t_2 - \tau_1$. For the non-linear parameters in the model, the constraints needed to be satisfied for a bubble are

$$m \in (0, 1) \tag{13}$$

$$\omega \in [\omega_1, \omega_2] \tag{14}$$

$$t_c \in [t_2 - \delta\Delta, t_2 + \xi\Delta]. \tag{15}$$

The constraint $m > 0$ is required due to a non-positive m corresponding to super-exponential time dynamics resulting in a non-finite price at the crash. Also, for $m \geq 1$, there is no acceleration and the growth is not super-exponential. The boundaries for ω are such that it is positive and the number of oscillations are within a certain range. The range sometimes depends on the size of the fitting window of the potential bubble. Sometimes a narrow filtering condition such as $\omega \in [6, 13]$ is used [9, 12, 5, 24]. The region for t_c depends upon application where $\delta, \xi \in [0, 1)$ are chosen appropriately. For example, $\xi = 0.2$ and $\delta = 0.2 \vee 0$. Call the space satisfying equations (14)-(16) R.

In the fitting procedure, we search over a space of the non-linear parameters. The search space $U \in \mathbb{R}^3$ must always contain R. In [24] U is defines as the space satisfying

$$m \in [0, 2] \tag{16}$$

$$\omega \in [1, 50] \tag{17}$$

$$t_c \in [t_2 - 0.2\Delta, t_2 + 0.2\Delta] \tag{18}$$

but the space is in general defined depending upon usage.

2.3 Statistical model

The LPPLS model assumes that the log-prices in a bubble phase have the form

$$\begin{aligned}\ln p_t = A + B|t_c - t|^m + C_1|t_c - t|^m \cos(\omega \ln|t_c - t|) \\ + C_2|t_c - t|^m \sin(\omega \ln|t_c - t|) + \epsilon_t\end{aligned} \tag{19}$$

where $t \in \{t_1, t_1 + 1, ..., t_c\}$, t_1 is the unknown true start of the bubble. At $t = t_c$ the limit of the right hand side is simply A. The joint distribution of the residuals ϵ_t is not known and needs to be estimated.

Under standard LPPLS assumptions, the error term is simply modeled as $\eta_t \sim \mathbf{N}(0, \sigma^2)$. We refer to this model as the ordinary least squares (OLS) LPPLS model.

The second model has the same parametric part but models the residuals with an auto-regressive part and a smoothly changing structure on the standard deviation in time, namely $\epsilon_t = \rho\epsilon_{t-1} + \eta_t$, $\eta_t \sim \mathbf{N}(0, \sigma_t^2)$. We refer to this model as the generalized least squares LPPLS model with auto-regressive part of order 1 and smooth heteroskedasticity (GLSH(1,0)).

The linear parameters β and the non-linear parameters α have to be estimated based on data along with the parameters describing the residuals.

2.4 Matrix formulation

Assume that the bubble has started at some point t_1 and that $t_1' \geq t_1$. Let $Y_{[t_1':t_2]} = (\ln p_{t_1'}, \ln p_{t_1'+1}, ..., \ln p_{t_2})^T$ for $t_2 \leq t_c$. Then we can write our model in vector notation as

$$Y_{[t_1':t_2]} = \mathbf{X}_{[t_1':t_2]}(\alpha)\beta + \epsilon_{[t_1':t_2]}. \tag{20}$$

The design matrix $\mathbf{X}_{[t_1':t_2]}(\alpha)$ is parameterized with parameters α and the subscript indicates the sequential rows selected from the matrix. The row vector for time t in the parameterized design matrix is

$$\begin{aligned}X_{[t:t]}(m, \omega, t_c) = \big(1, |t_c - t|^m, |t_c - t|^m \cos(\omega \ln|t_c - t|) \\ , |t_c - t|^m \sin(\omega \ln|t_c - t|)\big)\end{aligned} \tag{21}$$

and $\epsilon_{[t_1':t_2]} = (\epsilon_{t_1'}, \epsilon_{t_1'+1}, ..., \epsilon_{t_2})^T$ is the vector of residuals corresponding to each logarithm of price. When the time window of the data is known and also the non-linear parameters we simply write

$$Y = \mathbf{X}\beta + \epsilon \tag{22}$$

2.5 Generalized Least Squares (GLS)

Suppose we have the linear model

$$Y = \mathbf{X}\beta + \epsilon \tag{23}$$

where

$$\epsilon \sim N(0, \Sigma). \tag{24}$$

Using the Cholesky factorization, we have that $\Sigma = LL^T$, where L is a uniquely determined lower triangular. Covariance matrices are invertible by construction and we thus have $\Sigma^{-1} = (L^T)^{-1}L^{-1}$ and $(L^T)^{-1} = (L^{-1})^T$. Set

$$\tilde{Y} := L^{-1}Y \tag{25}$$

$$\tilde{X} := L^{-1}\mathbf{X}. \tag{26}$$

Then,

$$\tilde{Y} = \tilde{\mathbf{X}}\beta + \tilde{\epsilon} \tag{27}$$

where $\tilde{\epsilon} \sim N(0, I)$ where I is the identity matrix. Thus, knowing the covariance matrix of a GLS model we can transform it to the simple factorizable independent and identically distributed case. Under a fixed known design matrix $\mathbf{X}$ the OLS estimator minimizes the squared Euclidean distance

$$\hat{\beta} = \arg\min_{\beta} ||\tilde{Y} - \tilde{\mathbf{X}}\beta||^2. \tag{28}$$

Taking derivative of the objective function yields

$$-2\tilde{\mathbf{X}}^T(\tilde{Y} - \tilde{\mathbf{X}}\beta) = 0. \tag{29}$$

This is equivalent to the normal equations:

$$\tilde{\mathbf{X}}^T\tilde{Y} = \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\beta. \tag{30}$$

Assume full rank of $\tilde{\mathbf{X}}$. Then, the OLS estimator in the transformed space is the GLS estimator for β:

$$\hat{\beta} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{Y} = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}Y. \tag{31}$$

Under the OLS model, the covariance matrix is a diagonal matrix with fixed variance and the solution simplifies to

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TY. \tag{32}$$

2.6 Covariance construction

Suppose $t \geq t_1 + 1$ and the final time of the window is $t_2 \leq t_c$. Under the OLS model the covariance matrix is simply a diagonal matrix with fixed variance and the likelihood factorizes. Under the GLSH(1,0) model, $\epsilon_t = \rho\epsilon_{t-1} + \eta_t$, $\eta_t \sim \mathbf{N}(0, \sigma_t^2)$, thus

$$\text{Var}(\epsilon_t) = \text{Cov}(\rho\epsilon_{t-1} + \eta_t, \rho\epsilon_{t-1} + \eta_t) = \rho^2\text{Var}(\epsilon_{t-1}) + \sigma_t^2 \tag{33}$$

since η_t and η_{t-1}, the innovations, are independent by assumption. Also as the innovations are independent

$$\text{Cov}(\epsilon_t, \epsilon_{t+k}) = \rho^k\text{Var}(\epsilon_t) \tag{34}$$

for $t + k \leq t_2$. Equations 33 and 34 fully define our covariance C matrix for finite data:

$$C_{i,j} = \text{Cov}(\epsilon_i, \epsilon_j) \tag{35}$$

for $i, j \in \{t_1, ..., t_2\}$. Due to the simple parametric form the inverse exists given that $\sigma_t^2 > 0$ for all t. It is a tridiagonal matrix, where on the diagonal we have

$$C^{-1}_{t_2,t_2} = \frac{1}{\sigma^2_{t_2}} \tag{36}$$

and

$$C^{-1}_{i,i} = \frac{1}{\sigma_i^2} + \frac{\rho^2}{\sigma^2_{i+1}} \tag{37}$$

for $i \in \{t_1, ..., t_2 - 1\}$. For $i - j = 1$ we have that

$$C^{-1}_{i,j} = C^{-1}_{j,i} = -\frac{\rho}{\sigma_i^2}. \tag{38}$$

The rest of the inverse covariance matix equals zero.

Efficient estimator for ρ can be found by using the principles of maximum likelihood estimation. Fix $w_t = \frac{1}{\sigma_t^2}$. Minimizing the sum of squares of the weighted regression is the solution of the equation

$$\frac{d}{d\rho} \sum_{t=t_1+1}^{t_2} w_t(\epsilon_t - \rho\epsilon_{t-1})^2 = 0 \tag{39}$$

with the solution

$$\hat{\rho} = \sum_{t=t_1+1}^{t_2} \frac{w_t \epsilon_t \epsilon_{t-1}}{w_t \epsilon^2_{t-1}}. \tag{40}$$

Having estimates of σ_t^2 and using the auto-regressive corrected estimate for $\hat{\text{Var}}(\epsilon_{t_1}) = \frac{\hat{\sigma}^2_{t_1}}{1-\hat{\rho}^2}$ yields an iterative algorithm for estimating the residual covariance matrix Σ which is completely determined by ρ and σ_t^2.

2.7 Bayesian OLS model

Under the LPPLS OLS model, the data distribution is

$$p_{OLS}(y|\theta, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\big(\frac{1}{2\sigma^2}(y - \mathbf{X}\beta)^T(y - \mathbf{X}\beta)\big). \tag{41}$$

The joint posterior distribution of the parameters and the variance under the OLS assumption is

$$p(\theta, \sigma^2|y) \propto p_{OLS}(y|\theta, \sigma^2)p(\theta, \sigma^2). \tag{42}$$

Recall that the indicator function of a subset U on a set X is defines as

$$\mathbf{1}_U(x) = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \in X \setminus U. \end{cases} \tag{43}$$

Assume that the joint prior for our parameters is $p(\theta, \sigma^2) \propto \mathbf{1}_{U_{OLS}}(\alpha, \sigma^2)$ where U_{OLS} describes the search space for the non-linear parameters and the fact that $\sigma^2 > 0$.

2.8 Bayesian GLS model

Under the GLSH(1,0) LPPLS model, the data distribution has the density

$$p_{GLS}(y|\theta, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\big(\frac{1}{2}(y - \mathbf{X}\beta)^T\Sigma^{-1}(y - \mathbf{X}\beta)\big). \tag{44}$$

The joint posterior distribution of the parameters under the GLS assumption is

$$p(\theta, \Sigma|y) \propto p_{GLS}(y|\theta, \Sigma)p(\theta, \Sigma). \tag{45}$$

Assume that the prior for our parameters is $p(\theta, \Sigma) \propto \mathbf{1}_{U_{GLS}}(\alpha, \rho, \sigma_t)$ where the set U_{GLS} describes the search space for the non-linear parameters and $\rho \in (-1, 1)$ and each $\sigma_t^2 > 0$. The prior is a non-informative and improper prior. An improper prior density is such that the integral over the parameter space does not exist. Then, the maximum a posteriori estimation for both models is simply restricted maximization of the log-likelihood. The common search space U for the non-linear parameters is

$$m \in [0, 2] \tag{46}$$

$$\omega \in [1, 50] \tag{47}$$

and the space for t_c is wider than the assumed constraints in the LPPLS model.

Probabilistic description of β arises naturally from the Bayesian model. The conditional posterior distribution of the linear parameters β is

$$p(\beta|\Sigma, y, \alpha) \propto p_{GLS}(y|\theta, \Sigma)p(\beta|\Sigma, \alpha). \tag{48}$$

Assuming that $p(\beta|\Sigma, \alpha) \propto 1$ yields

$$p(\beta|\Sigma, y, \alpha) \propto p_{GLS}(y|\theta, \Sigma). \tag{49}$$

Filling the square shows that the conditional distribution of β is Gaussian. The quadratic term inside the exponential reveals that the covariance matrix of the conditional distribution for β is $\Sigma_\beta = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}$. From the linear term $\mu_\beta = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}Y$ which is the usual GLS estimator for β. The conditional distribution is thus

$$\beta|\Sigma, y, \alpha \sim N(\mu_b, \Sigma_\beta). \tag{50}$$

Using the point estimates for the non-linear parameters and the data yields an approximate distribution of the linear parameters. Integration over the constraints space S in order gives a measure of confidence for the bubble signal. This can be one of the tools to identify if a true bubble is present or not.

2.9 Posterior maximization

For the two models, parameter estimation is based on likelihood maximization. In the case of the OLS model, the likelihood factorizes. The parameters are selected as

$$(\hat{\theta}_{OLS}, \hat{\sigma}^2) := \arg\max_{\theta,\sigma^2} p_{OLS}(y|\theta,\sigma^2)\mathbf{1}_{U_{OLS}}(\alpha,\sigma^2). \tag{51}$$

The optimization requires some type of search over U_{OLS} and at each candidate triplet α we use the least squares estimate $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$ and estimate the residual variance. The selection is then based on which parameters jointly optimize the objective.

For the GLS, the likelihood does not factorize and the selection is found by

$$(\hat{\theta}_{GLS}, \hat{\Sigma}) := \arg\max_{\theta,\Sigma} p_{GLS}(y|\theta,\Sigma)\mathbf{1}_{U_{GLS}}(\alpha,\rho,\sigma_t^2). \tag{52}$$

The optimization requires some type of search over U_{GLS}. At candidate α the generalized least squares estimator is used $\hat{\beta} = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}Y$ and the structured covariance matrix are estimated in a combinatorical fashion to maximize the objective.

2.10 Algorithmic comparison

For a given t_1 and a non-linear triplet (m,ω,t_c) the linear model is fitted. Due

Algorithm 1 Fit the OLS LPPLS model for a given set of non-linear parameters

Require: t_1
Require: A triplet (m,ω,t_c)
1: Compute the design matrix $\mathbf{X}$ based on the given triplet (m,ω,t_c).
2: Calculate β from equation (32)

to the simple structure of the inverse covariance matrix, namely a tridiagonal matrix(see equations (36)-(38)), in the LPPLS GLSH(1,0) model the calculation of β simplifies greatly. The algorithms have the same computational complexity, given that the while loop has constant time (often 3-6 iterations in practice). The different LPPLS algorithms only differ by a scaling factor not depending on the length of the time series in computational time. Hence, the GLSH(1,0) is reasonable in practice but a initial search over the non-linear space with the simple OLS solution is recommended for speeding up the computation. An even cheaper version is to store the M best solutions according to the OLS fits and fit the GLSH(1,0) model based on these fits.

Algorithm 2 Fit the GLSH(1,0) LPPLS model for a given set of non-linear parameters

Require: t_1
Require: A triplet (m, ω, t_c)

1: Compute the design matrix $\mathbf{X}$ based on the given triplet (m, ω, t_c).
2: Calculate initial estimate of β from equation (32).
3: Get initial estimate of ρ, the auto-regressive parameter.
4: Fit the filtered residuals with a smooth curve for initial estimate of the variance structure.
5: Use the estimates of ρ and variance structure to calculate Σ and Σ^{-1}.
6: **while** not converged in Σ,β **do**
7: Estimate β with equation (31)
8: Estimate ρ using equation (40)
9: Estimate the variance structure with a smooth curve on the new filtered residuals.
10: Use the estimates of ρ and variance structure to update estimates of Σ and Σ^{-1}.
11: **end while**

3 Model selection based on empirical data

3.1 Goodness of fit measures

A widely used model selection criterion in the context of likelihood literature is the Akaike information criterion (AIC). Given a fixed set of data, the AIC statistic estimates the relative quality of statistical models. Under this fixed set of data, the AIC estimates the goodness of each model relative to each of the other models. It is derived under the principles of information theory and estimates the relative information lost by a given model and the selected model is the one that loses the minimum amount of information [3]. The first order estimate for the information loss for a given maximum value of the likelihood $\hat{L}$ and k degrees of freedom for the model is defined:

$$\mathrm{AIC} := 2k - 2\ln\hat{L}. \tag{53}$$

As we usually deal with reasonably large sample sizes the second order corrections are not needed, known as AIC with a correction for small sample sizes (AICc).

The relative comparison for the two models is

$$\begin{aligned} 2\big(- p_{GLS}(y|\hat{\theta}_{GLS}, \hat{\Sigma}) + p_{OLS}(y|\hat{\theta}_{OLS}, \hat{\sigma^2}) \\ + k_{GLS} - k_{OLS}\big) \end{aligned} \tag{54}$$

where k_{OLS}, k_{GLS} are the number of parameters in the corresponding model. The quantity

$$Q := \exp\Big(\frac{\mathrm{AIC}_{GLS} - \mathrm{AIC}_{OLS}}{2}\Big) \tag{55}$$

Table 1: For the six empirical bubbles considered we compare the AIC and the BIC for the LPPLS OLS and LPPLS GLSH(1,0) fit. The parameter n is the number of data points in each fit and the quantity Q from equation (55) is proportional to the probability that the OLS model minimizes the estimated information loss. The bubbles considered are the Argentinian MERVAL index, the Brazilian IBOVESPA index, the Philippines PSEi index, the Thailand Bangkok SET index, SSEC Shanghai index and U.S. Sugar prices.

	Argentina 1994	Brazil 1997	Philippines 1994	Thailand 1994	Shanghai 2007	Sugar 2005
n	187	328	250	145	449	168
AIC_{OLS}	-727.7	-1491.5	-1081.9	-648.2	-1404.6	-773.8
AIC_{GLS}	-1016.8	-1916.1	-1537.0	-865.4	-2465.7	-940.9
Q	$1.7*10^{-63}$	$6.4*10^{-94}$	$1.5*10^{-99}$	$6.7*10^{-48}$	$3.8*10^{-231}$	$5.4*10^{-37}$
BIC_{OLS}	-705.1	-1464.9	-1057.3	-627.3	-1375.9	-752.0
BIC_{GLS}	-980.9	-1865.0	-1494.2	-834.4	-2402.1	-907.1

can be interpreted as being proportional to the probability that the OLS model minimizes the estimated information loss.

A simple alternative to the AIC statistic is the Bayesian Information Criterion

$$\text{BIC} := \ln(N)k - 2\ln\hat{L} \tag{56}$$

where k is the degrees of freedom in the model and N is the sample size. From an information theory standpoint the BIC approximates the minimum description length criterion but with a negative sign. Further, it is independent of the prior being used. While AIC tries to select the model that most adequately describes an unknown, high dimensional reality the BIC tries to find the true model among the set of candidates by assumption.

In summary, when fitting models it is often possible to increase the likelihood by introducing new parameters. However, doing so may result in overfitting. Both BIC and AIC try to resolve this problem by introducing a penalty term for the number of parameters in the model. For sample size $N > 7$, the penalty term is larger in BIC than in AIC and thus in general BIC favors simpler models than AIC.

3.2 Empirical Comparison of the LPPLS OLS and LPPLS GLSH(1,0)

We consider several known empirical bubbles [10, 11, 14, 21]. The window selected to fit each bubble is such that it is clear that the bubble has started and the endpoint is the local maximum of the log-price series before a severe correction. A comparison of the two different approaches can be seen on figures 1-12 and figures 48-107 from the Appendix. As an example we take the Philippines

index in 1994, the corresponding diagnostics plots and fits for the other bubbles are shown in the Appendix. The residuals in the GLSH(1,0) model are fitted with a locally weighted smoothing with $\frac{\Delta-1}{60}$ degrees of freedom where Δ is the length of the fitting window. In general, the signal estimates from the LPPLS OLS model and the LPPLS GLSH(1,0) model are very similar.

The assumptions of the LPPLS OLS model are violated as the raw residuals found by subtracting the fitted LPPLS OLS model from the log prices, in general, have a strong auto-regressive part. This is evident from the autocorrelation(figures 2,49,61,73,85,97) and the partial-autocorrelation(figures 3,50,62,74 ,86,98) plots for each bubble. The partial-autocorrelation is found by controlling for the values of the time series at all shorter lags. The strong auto-regressive part is evident from the auto-correlation and partial auto-correlation plots while there is a weak or non-existing moving average part. The auto-correlation function seems to be declining exponentially fast in general although in some cases it is close to being linear, an indication of long memory in the residuals. After correcting for auto-regressive residuals, the innovations, in general, require correcting for the heteroskedasticity under the normal assumption, see figures 11,12,58,59,70,71,82,83 ,94,95,106,107. The improvement upon the standard ordinary least squares estimates for the empirical bubbles considered in AIC and BIC is astronomical as seen from Q in table 1 and the relative difference for both AIC and BIC for all bubbles.

In the time series literature, visual inspection of model assumptions is common practice. Several standard diagnostics plots are known. For our empirical bubbles, we consider the residuals as a function of the fitted signal, known as Tukey-Anscombe plot, which should have no structure, see figures 4,9,51,56,63,68,75,80 ,87,92,99,104. Further, we plot the standardized residuals as a function of time, which should look like random noise, see figures 5,10,52,57,64,69,76,81,88,93,100,105. Finally, the quantile-quantile plots are considered to check if the normal assumption is valid. Some of the quantile-quantile plots on the fits look reasonable before accounting for heteroskedasticity, but commonly there are outliers which we do not want to bias our parameter estimation. The fitted heteroskedasticity often improves the situation. The standard diagnostics plots look much more reasonable for the GLSH(1,0) model than the OLS model as the assumption of no correlation, constant variance, and normality of the residuals is violated in general for the empirical bubbles considered.

Figure 1: The fit of the full bubble in Philippines in 1994(PSEi index, closing prices) for the OLS LPPLS model in blue and the GLSH(1,0) LPPLS model in red. As usual, the GLSH(1,0) fit is a bit smoother, but the signals are otherwise similar. The fitted parameters for the LPPLS signals($t_2 = 1994/01/04$ corresponds to the end of the fitting window, the true peak of the bubble) are $\hat{A}_{OLS} = 8.6582, \hat{B}_{OLS} = -0.4265, \hat{C}_{1,OLS} = 0.0039, \hat{C}_{2,OLS} = 0.0120, \hat{m}_{OLS} = 0.2224, \hat{\omega}_{OLS} = 7.5459, \hat{t_c}_{,OLS} = t_2 + 2$ and $\hat{A}_{GLS} = 8.4524, \hat{B}_{GLS} = -0.2788, \hat{C}_{1,GLS} = 0.0021, \hat{C}_{2,GLS} = 0.0080, \hat{m}_{GLS} = 0.2734, \hat{\omega}_{GLS} = 7.5459, \hat{t_c}_{,GLS} = t_2 + 2, \hat{\rho} = 0.9303$.

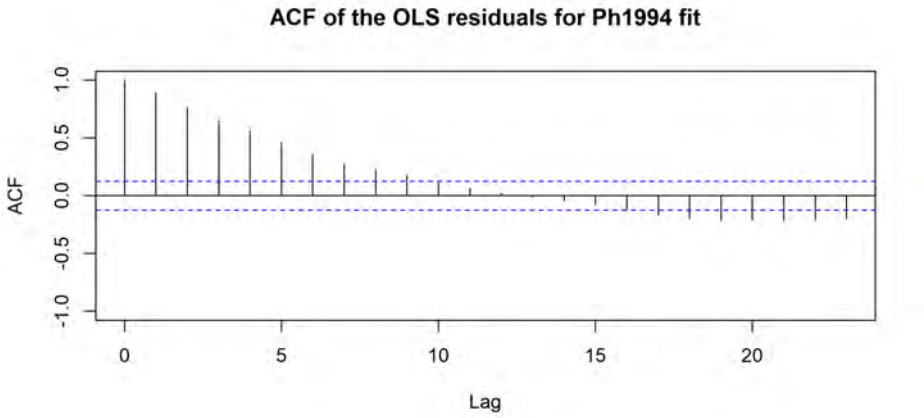

Figure 2: The autocorrelation function for the LPPLS OLS fit for the full bubble in Philippines(PSEi Index) in 1994. Strong auto-regressive part is evident. The exponential decay of the coefficients indicates short memory.

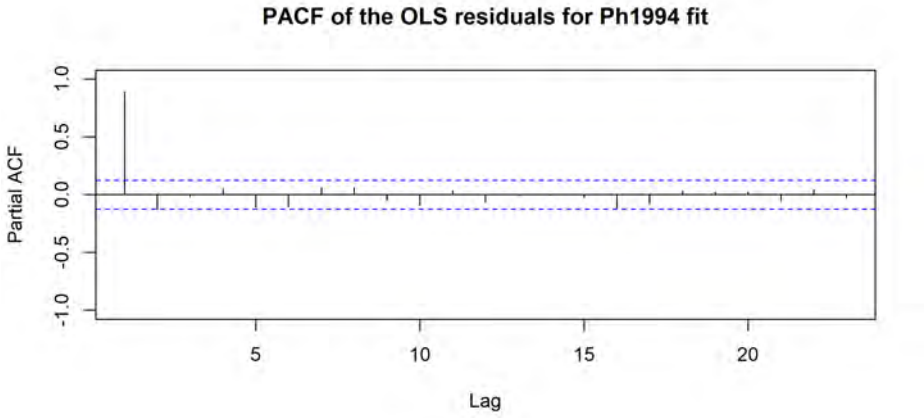

Figure 3: The partial autocorrelation function for the LPPLS OLS fit for the bubble in Philippines(PSEi Index) in 1994. Strong auto-regressive part is evident of order 1.

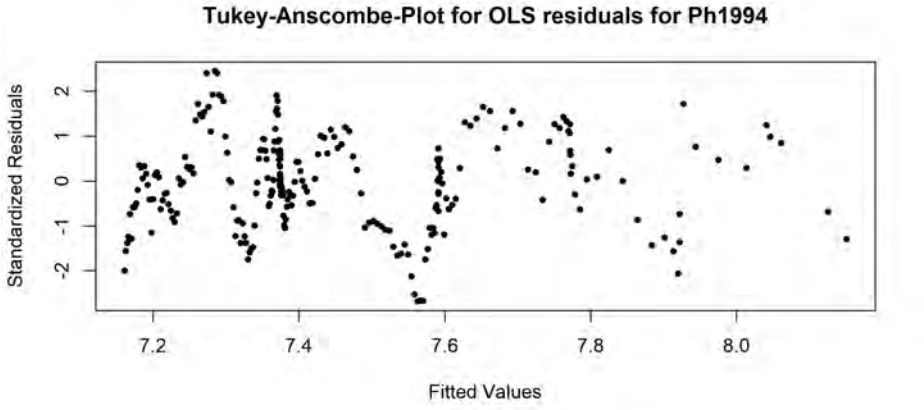

Figure 4: Tukey-Anscombe plot for the LPPLS OLS fit of the standardized residuals for the bubble in Philippines(PSEi Index) in 1994. A structure is evident, and thus the model assumptions violated.

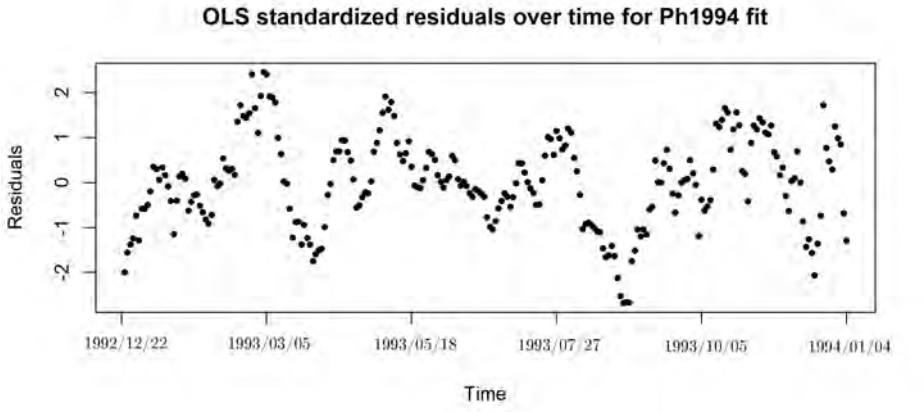

Figure 5: The plot shows standardized residuals over time for the LPPLS OLS fit for the bubble in Philippines(PSEi Index) in 1994. A structure is evident, and thus the model assumptions violated.

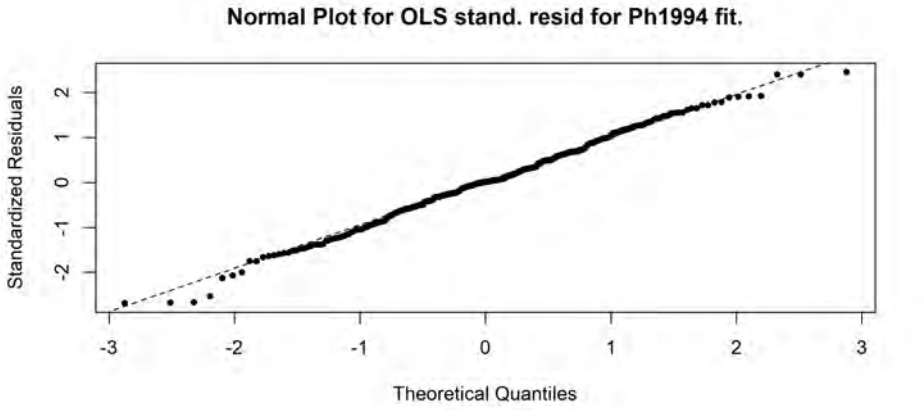

Figure 6: The quantile-quantile normal plot for the LPPLS OLS fit for the bubble in Philippines(PSEi Index) in 1994. The normality is not violated. However, the strong auto-regressive structure needs to be corrected for before considering the Q-Q plot.

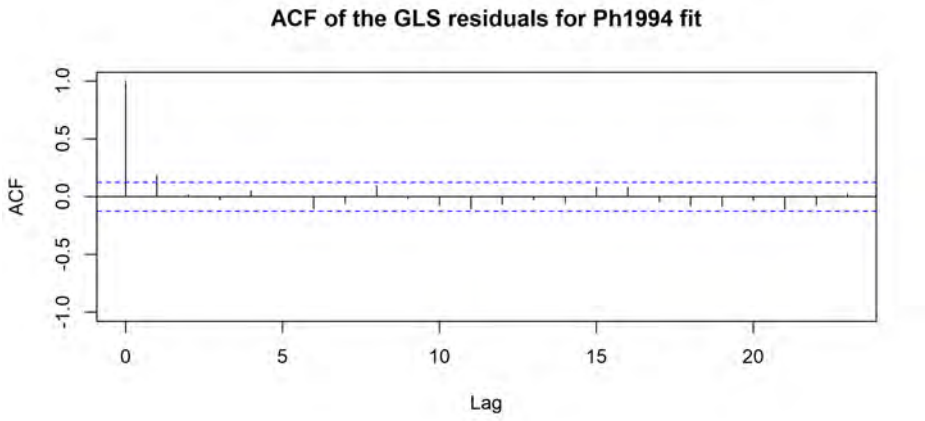

Figure 7: The autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Philippines(PSEi Index) in 1994. The transformation seems to have uncorrelated the residuals for the most part.

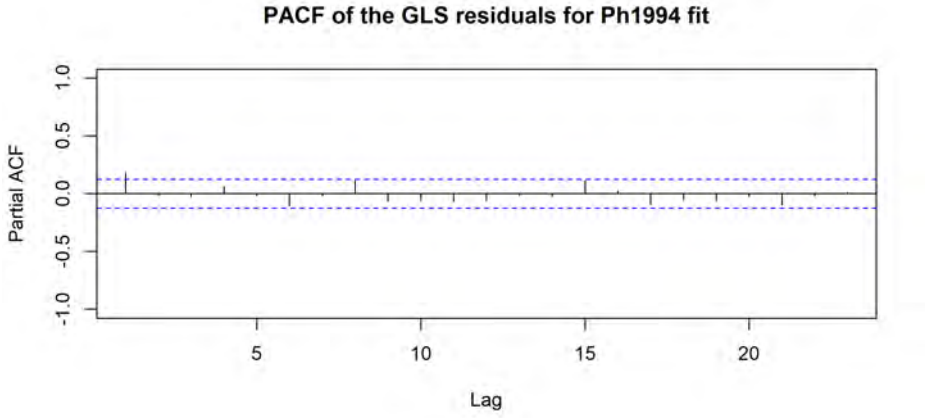

Figure 8: The partial autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Philippines(PSEi Index) in 1994. The transformation seems to have uncorrelated the residuals for the most part.

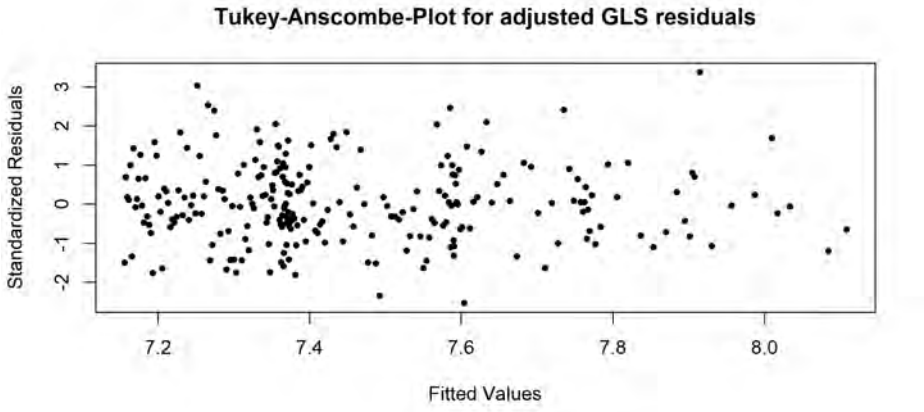

Figure 9: Tukey-Anscombe plot for the filtered residuals for the LPPLS GLSH(1,0) fit of the standardized residuals for the bubble in Philippines(PSEi Index) in 1994. There is hardly any structure in the residuals.

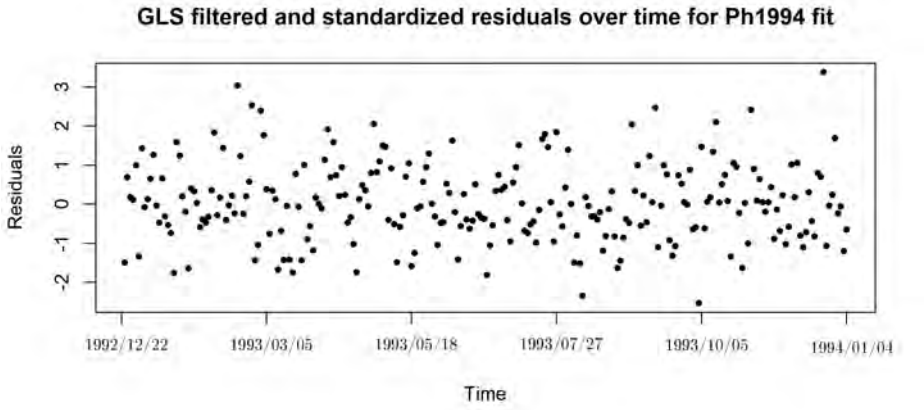

Figure 10: Standardized residuals over time plot for the LPPLS GLSH(1,0) fit for the bubble in Philippines(PSEi Index) in 1994. There is hardly any structure in the residuals.

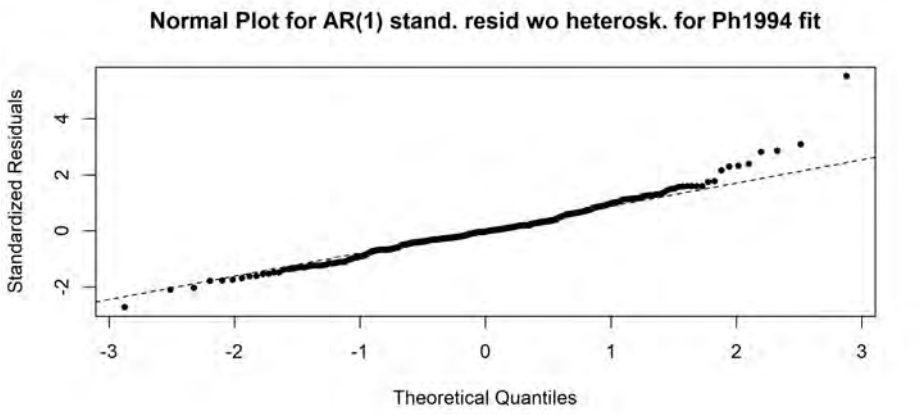

Figure 11: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit before accounting for heteroskedasticity for the bubble in Philippines(PSEi Index) in 1994. Heavy positive tails of the residuals indicate violation of the normality assumption.

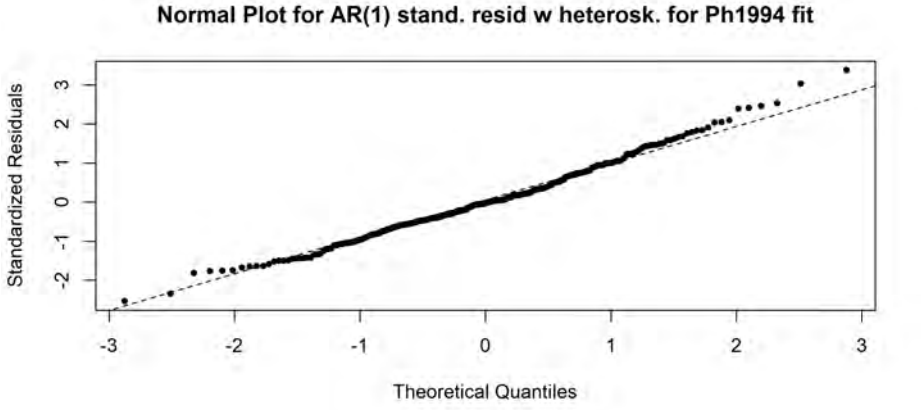

Figure 12: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit after accounting for heteroskedasticity for the bubble in Philippines(PSEi Index) in 1994. The fitted heteroskedasticity is estimated by using a loess smoother on the empirical residuals after filtering with $\Delta/60$ degrees of freedom. The normality assumption can hardly be rejected.

4 Selection of t_1 for the LPPLS model

In the previous chapters, the initial time of a bubble t_1 has been assumed to be known. In practice, however, selecting a reasonable t_1 can be quite challenging. Initial times of bubbles are a fuzzy concept and algorithms for selecting t_1 depend on the dataset, namely on the final observed time $t_2 \leq t_c$. The problem of selecting optimal sub-dataset from a nested sequence is not a standard problem. An attempt to solve the problem was proposed by Demos and Sornette[4] using regularisation on size of the fitting windows in order to correct for overfitting on smaller windows.

4.1 Lagrange regularisation

Set the possible beginning of fitting windows as $t_1' \in [\tau_1 : T_1]$. Set

$$\hat{Y}_{[t_1':t_2],OLS} = \mathbf{X}_{[t_1':t_2]}(\hat{\alpha}_{[t_1':t_2],OLS})\hat{\beta}_{[t_1':t_2],OLS} \tag{57}$$

$$\hat{Y}_{[t_1':t_2],GLS} = \mathbf{X}_{[t_1':t_2]}(\hat{\alpha}_{[t_1':t_2],GLS})\hat{\beta}_{[t_1':t_2],GLS} \tag{58}$$

where $\hat{\beta}_{[t_1':t_2],OLS}$ and $\hat{\alpha}_{[t_1':t_2],OLS}$ are the set of parameters maximizing the OLS log-likelihood and $\hat{\beta}_{[t_1':t_2],GLS}$ and $\hat{\alpha}_{[t_1':t_2],GLS}$ are the set of parameters maximizing the GLSH(1,0) log-likelihood. Then, with t_2 the soonest observation assumed to be $t_2 \leq t_c$ we have the cost function for both the models as

$$\chi^2_{\lambda,t_1'} = \frac{(Y_{[t_1':t_2]} - \hat{Y}_{[t_1':t_2]})^T(Y_{[t_1':t_2]} - \hat{Y}_{[t_1':t_2]})}{(t_2 - t_1') - p} + \lambda(t_2 - t_1') \tag{59}$$

where p is the number of parameters in the corresponding models. The parameter λ is defined as the slope of the line fitted with intercept through all $\chi^2_{\lambda,t_1'}$ which penalizes smaller windows since the slope is usually negative. The selection of λ is somewhat arbitrary, depends heavily on the whole set of fits at each point of the window $[\tau_1 : T_1]$. The problem with this approach is that the error metric sum of squares assumes identical variance over the sample which is clearly not the case in general for empirical bubbles. Thus, periods with lower volatility are more likely to be selected even though the signal might begin much earlier. Also, the selection is based on in sample measure and is thus prone to select t_1 based on overfitting instead of generalization performance. Hence, a natural improvement involves selecting the optimal t_1 based on predictive performance, which we propose.

4.2 Cross-validation for time dependent data

Using the idea of cross-validation for time dependent data the selection of the optimal t_1 is decided by best predictive performance in some metric. In the econometric literature such methods are known as backtesting.

Backtesting criteria for selecting t_1 aim to find the sub-dataset that provides the most accurate forecasts in some metric. The last part of the time

series $Y_{[t_2-n+1:t_2]}$ is known as the validation set. Typically, a cross-validation procedure for time dependent data given some summable error function E and assuming $t_c \geq t_2$ can be described by the following steps:

1. For each candidate $t_1' \in [\tau_1 : T_1]$ with t_2 fixed we select the same $n < t_2 - T_1$ and we fit the model on the data $Y_{[t_1':t_2-n]}$. The best linear unbiased estimator [6] to predict

$$y^{\star}_{t_2-n+1} = \mathbf{X}_{[t_2-n+1:t_2-n+1]}(\hat{\alpha}_{[t_1':t_2-n],GLS})\hat{\beta}_{[t_1':t_2-n],GLS} + \hat{\rho}_{[t_1':t_2-n]}\hat{\epsilon}_{t_2-n,[t_1':t_2-n]} \tag{60}$$

and calculate the error associated with that prediction

$$e_{t_2-n+1,t_1'} = E(y_{t_2-n+1}, y^{\star}_{t_2-n+1}). \tag{61}$$

2. For $t \in \{t_2 - n + 2, ..., t_2\}$)(correspondingly, $i \in 2, ..., n$) we fit the model on $Y_{[t_1':t-1]}$, use the best linear unbiased estimator [6] to predict

$$y^{\star}_{t_2-n+i} = \mathbf{X}_{[t:t]}(\hat{\alpha}_{[t_1':t-1],GLS})\hat{\beta}_{[t_1':t-1],GLS} + \hat{\rho}_{[t_1':t-1]}\hat{\epsilon}_{t-1,[t_1':t-1]} \tag{62}$$

and calculate the error associated with that prediction

$$e_{t,t_1'} = E(y_{t_2-n+i}, \hat{y}_{t_2-n+i}). \tag{63}$$

3. Select $t_1^{\star} = \arg\min_{t_1'} \sum_{t=t_2-n+1}^{t_2} e_{t,t_1'}$.

4. Fit the model on $Y_{[t_1^{\star}:t_2]}$ and use that model as our estimation of the process.

In practice, it is time consuming to fit for every t_1' a new model to all of $\{t_2 - n, t_2 - 1\}$. Then a computationally cheaper modification simply fits once on $y_{[t_1:t_2-n]}$ and then uses the best linear unbiased predictions [6]

$$y_t^{\star} = \mathbf{X}_{[t]}(\hat{\alpha}_{[t_1':t_2-n],GLS})\hat{\beta}_{[t_1':t_2-n],GLS} + \hat{\rho}^{t-t_2+n}_{[t_1':t_2-n]}\hat{\epsilon}_{t_2-n,[t_1':t_2-n]} \tag{64}$$

for $t \in \{t_2 - n + 1, t_2\}$) and calculates the sum of the error function as before.

The size of the validation set n for a given bubble needs to be selected. Robust estimate for the error function is called for with low variance but also accurate parameter estimates. This has to be estimated empirically and a single global answer is unlikely to exist, since it is unlikely that it is independent of all of the sample size in the true bubble, the signal to noise ratio and the strength of the bubble.

Selection of the error function is crucial. The mean square error on predicted data (MSPE) for n dimensional out of sample Z and prediction $Z^{\star}$ can be defined as

$$\text{MSPE}(Z, Z^{\star}) := \frac{(Z - Z^{\star})^T(Z - Z^{\star})}{n}. \tag{65}$$

MSPE and similar summable functions are a simple solution but as the model does contain heteroskedasticity perhaps a weighted version is more appropriate. However, that requires modeling the validation set.

To fit the LPPLS model on the validation set or similar non-parametric function n has to be large enough to fit the bubble on that as well. The standard way of correcting for the residual structure under the Gaussian assumption is to use generalized sum of squares (GSS) where we need to estimate the validation set covariance matrix, Σ_q. Formally the GSS is defined as

$$\text{GSS}(Z, \mu^\star) = \frac{(Z - \mu^\star)^T \Sigma_q^{-1} (Z - \mu^\star)}{n} \tag{66}$$

where Σ_q is the covariance matrix of the validation data Z and $\mu^\star$ is the unconditional prediction based on the training data. Another solution is to minimize the Kullback-Leibler divergence between the predictions and the fit on the validation data.

In that case we fit the model on the validation data. This requires n to be substantially large. Then the Kullback-Leibler divergence (KL) between the predicted distribution p under normal assumption and equal covariances and the Gaussian distribution of the validation data $q \sim N(\mu_q, \Sigma_q)$ is

$$KL(p||q) = \frac{1}{2}(\mu_q - \mu^\star)^T \Sigma_q^{-1} (\mu_q - \mu^\star) + K \tag{67}$$

where K is a fixed constant.

In the testing fit the model on $Z = Y_{[t_2-n+1:t_2]}$ and estimate

$$\hat{\mu}_q = \mathbf{X}_{[t_2-n+1:t_2]}(\hat{\alpha}_{[t_2-n+1:t_2],GLS})\hat{\beta}_{[t_2-n+1:t_2],GLS} \tag{68}$$

and construct the covariance matrix $\hat{\Sigma_q}$ based on the usual assumption and the fitted parameters $\hat{\theta_q}$, $\hat{\rho}$ and $\hat{\sigma_t}^2$. For the predicted means for each $t_1' \in \{\tau_1, \tau_1 + 1, ..., T_1\}$ predict $\mu_{t_1'}^\star = \mathbf{X}_{[t]}(\hat{\alpha}_{[t_1':t_2-n],GLS})\hat{\beta}_{[t_1':t_2-n],GLS}$. Then for each loss function E the optimal initial point is found as

$$t_1^\star = \arg \min_{t_1' \in \{\tau_1, ..., T_1\}} E(n, t_1'). \tag{69}$$

Non-parametric methods to estimate both μ_q and Σ_q are also possible, however they can not capture estimates of the critical t_c.

5 Comparison of the different decision criteria for selecting t_1

5.1 Simulation study

The synthetic dataset are 100 realizations generated from a time series model. The first 150 data points are generated from a long memory process with a first order auto-regressive part. At $t = 151$ there is a regime shift and the latter 250 data points are generated from a parametric LPPLS signal coupled with noise structure. The parameters for the simulation study are estimates

Table 2: Parameter values for the LPPLS signal and variance structure in the simulation model. There is a regime shift at $t_1 = 151$ and each parameter only corresponds to one of the two regimes.

Parameter	Value	Time interval
M	0.2735	$t \geq 151$
W	7.5459	$t \geq 151$
T_c	402	$t \geq 151$
a	8.4524	$t \geq 151$
b	-0.2788	$t \geq 151$
c_1	0.0021	$t \geq 151$
c_2	0.0080	$t \geq 151$
σ_t	variable	$t \geq 151$
s	$0.05 * \sigma_{151}$	$t \geq 151$
ρ	0.930	$t \geq 151$
d	0.4	$t < 151$
ϕ	0.875	$t < 151$
σ_v	σ_{151}	$t < 151$

for the Philippines 1994 bubble. The noise structure in the bubble phase has a strong auto-regressive part. In addition, there is random white noise on top of the signal with the auto-regressive signal in order to make it harder to detect. Formally, the process, which it is simulated from, is defined as

$$\begin{aligned} \ln p(t) = a + b|T_c - t|^M + c_1|T_c - t|^M \cos(W \ln|T_c - t|) \\ + c_2|T_c - t|^M \sin(W \ln|T_c - t|) + \epsilon_t + e_t \end{aligned} \tag{70}$$

for $t \geq 151$. Here, all the parameters are defined as before and the added white-noise is $e_t \overset{iid}{\sim} \mathbf{N}(0, s^2)$. The heteroskedasticity of the innovations is shown in figure 13. For the initial 150 points, the log-price series was simulated from an ARFIMA(1,d,0) model with no drift which was scaled such that a smooth fitting curve had a comparable residual variance for before and after t_1 and shifted such that the change of regime at $t = 151$ was reasonably smooth. Formally, a stochastic process Z_t is an ARFIMA(1,d,0) if it follows

$$\big(1 - \phi B\big)(1 - B)^d Z_t = v_t \tag{71}$$

where v_t is white-noise with variance σ_v^2 and B is the backshift operator. All the parameter values and which time interval they correspond to are shown in table 2. For visualization, four different realization of the process are shown in figure 14.

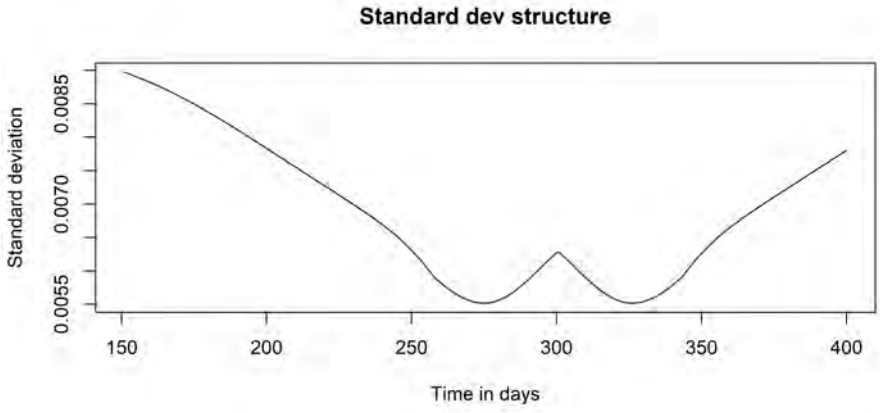

Figure 13: The structure of σ_t for the bubble phase in the simulation study, $t \in \{151, 152, ..., 400\}$.

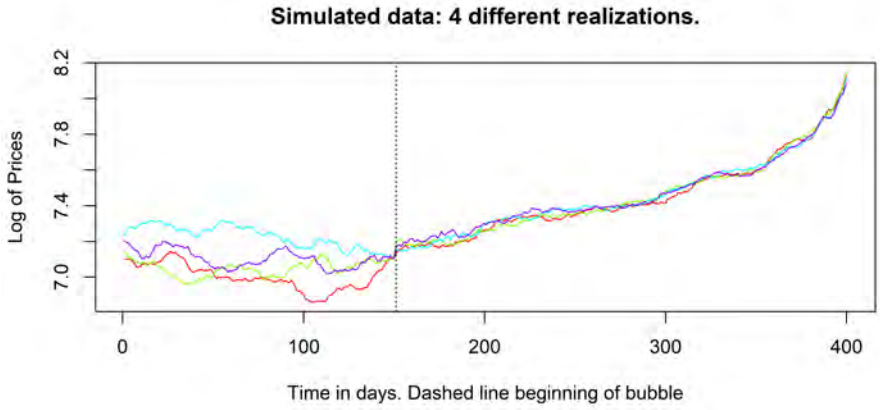

Figure 14: 4 different realizations of the simulated data. The black dashed line marks the beginning of the bubble phase at $t = 151$.

5.2 Proposed methods

In the simulation study the following methods are compared:

1. The Lagrange regularisation method as a benchmark.
2. MSPE out of sample minimization for both OLS and GLSH(1,0) for $n \in \{10, 20, 30, 40, 50\}$.
3. GSS out of sample minimization for GLSH(1,0) for $n \in \{30, 40, 50\}$.
4. KL out of sample minimization for GLSH(1,0) for $n \in \{30, 40, 50\}$.

In the study, we fit both the LPPLS OLS model and LPPLS GLSH(1,0) model for each $t_1' \in \{1, 6, 11, ..., 321\}$ where the true $t_1 = 151$ and leave out $n \in \{0, 10, 20, 30, 40, 50\}$. Then we minimize each of the error functions to select the optimal t_1. The search space in the study is $t_c \in [400, 450]$, $m \in [0, 2]$ and $\omega \in [4, 25]$.

Table 3: Estimated T_c for the fit at the optimally selected t_1 for each criteria for the dataset of 100 synthetic bubbles. Here, the GLS LPPLS type corresponds to the LPPLS GLSH(1,0) model. True $T_c = 402$. Empirical summary statistics are presented in the columns of the table. MSE($\hat{T}_c$) is the empirical mean squared error.

Selection criteria	LPPLS type	n	$E(\hat{T}_c)$	sd($\hat{T}_c$)	$\min \hat{T}_c$	$\max \hat{T}_c$	MSE($\hat{T}_c$)
Lagrange	OLS	0	402.29	2.02	400	409	4.13
MSPE	OLS	10	402.51	2.13	400	411	4.77
MSPE	GLS	10	402.21	1.04	401	411	1.11
MSPE	OLS	20	402.34	1.92	400	410	3.78
MSPE	GLS	20	402.15	1.05	401	410	1.11
MSPE	OLS	30	402.68	2.44	400	411	6.34
MSPE	GLS	30	402.37	1.84	401	414	3.49
MSPE	OLS	40	403.17	2.49	400	410	7.49
MSPE	GLS	40	402.53	1.70	401	411	3.15
MSPE	OLS	50	403.60	5.42	400	450	31.66
MSPE	GLS	50	402.65	1.75	401	410	3.45
GSS	GLS	30	402.49	2.02	401	414	4.27
GSS	GLS	40	402.78	2.00	401	411	4.56
GSS	GLS	50	402.80	2.03	401	411	4.72
KL	GLS	30	402.36	1.71	401	411	3.04
KL	GLS	40	402.78	1.99	401	411	4.54
KL	GLS	50	402.73	1.90	401	411	4.09

5.3 Results

Histograms of each selection criteria are shown in figures 15-31. Most methods for selecting t_1 seem to be relatively unbiased, but the variance seems to increase in general as n, the size of leave-out data, grows. The Lagrange regularisation method seems to be performing best from the histograms. However, that does not necessarily indicate that it is optimal for all metrics. Selecting t_1 too soon or too late might still yield a good fit.

In tables 3 and 4 we can see summary statistics for the simulation study. The crucial parameter T_c seems to be better captured by the GLSH(1,0) model compared to the OLS fits, evident from the summary statistics. Further n between 10 and 20 seems to be optimal for the leave-out methods. As the GSS and KL divergence can only be reliably used for larger n they are outperformed by MSPE for the GLSH(1,0) model. However, for a comparable n, the KL method seems to have a similar performance as the MSPE method. The process which the simulations are sampled from is a process similar to the LPPLS GLSH(1,0) model but with a non-smooth variance structure and added white-noise. Hence, it is encouraging to see that it outperforms the LPPLS OLS model, and its model selecting criteria, in predicting the crucial T_c.

Table 4: Empirical median of the parameters M, W,a, b, c_1, c_2 for the fit at the optimally selected t_1 for each criteria for the dataset of 100 synthetic bubbles. Here, the GLS LPPLS type corresponds to the LPPLS GLSH(1,0) model. True values for the parameters are $M = 0.2735$, $W = 7.5459$, $a = 8.4524$, $b = -0.2788$, $c_1 = 0.0021$, $c_2 = 0.0080$. GLS corresponds to the GLSH(1,0) model.

Selection criteria	LPPLS type	n	Med($\hat{M}$)	Med($\hat{W}$)	Med($\hat{a}$)	Med($\hat{b}$)	Med($\hat{c}_1$)	Med($\hat{c}_2$)
Lagrange	OLS	0	0.294	7.49	8.40	-0.240	0.0044	0.0032
MSPE	OLS	10	0.271	7.53	8.45	-0.278	0.0026	0.0017
MSPE	GLS	10	0.273	7.56	8.45	-0.278	0.0016	0.0068
MSPE	OLS	20	0.288	7.50	8.41	-0.252	0.0023	0.0006
MSPE	GLS	20	0.277	7.56	8.45	-0.273	0.0013	0.0053
MSPE	OLS	30	0.272	7.52	8.44	-0.265	0.0004	0.0022
MSPE	GLS	30	0.268	7.57	8.47	-0.295	0.0011	0.0065
MSPE	OLS	40	0.255	7.72	8.52	-0.328	0.0007	0.0024
MSPE	GLS	40	0.263	7.59	8.49	-0.299	0.0003	0.0067
MSPE	OLS	50	0.267	7.61	8.46	-0.288	0.0034	0.0003
MSPE	GLS	50	0.263	7.64	8.50	-0.307	0.0008	0.0062
GSS	GLS	30	0.263	7.58	8.48	-0.299	0.0010	0.0067
GSS	GLS	40	0.259	7.65	8.51	-0.322	0.0001	0.0067
GSS	GLS	50	0.255	7.66	8.50	-0.318	0.0007	0.0063
KL	GLS)	30	0.263	7.57	8.48	-0.297	0.0012	0.0066
KL	GLS	40	0.255	7.65	8.51	-0.328	0.0001	0.0064
KL	GLS	50	0.259	7.66	8.50	-0.311	0.0003	0.0062

For the additional parameters, the median of the parameter M seems to be estimated robustly for most methods and the same applies to W. Still, we note that the acceleration parameter m is best captured by out of sample methods with $n \leq 20$ and with the GLSH(1,0) fit. The other parameters are less critical, but both a and b are captured well in the median for our particular model structure.

Finally, the correlation plot indicates that the value of n, the size of the leave-out dataset, is crucial and matters more than the particular error function (see figure 32). The Lagrange regularisation method stands alone with low correlation with all other methods.

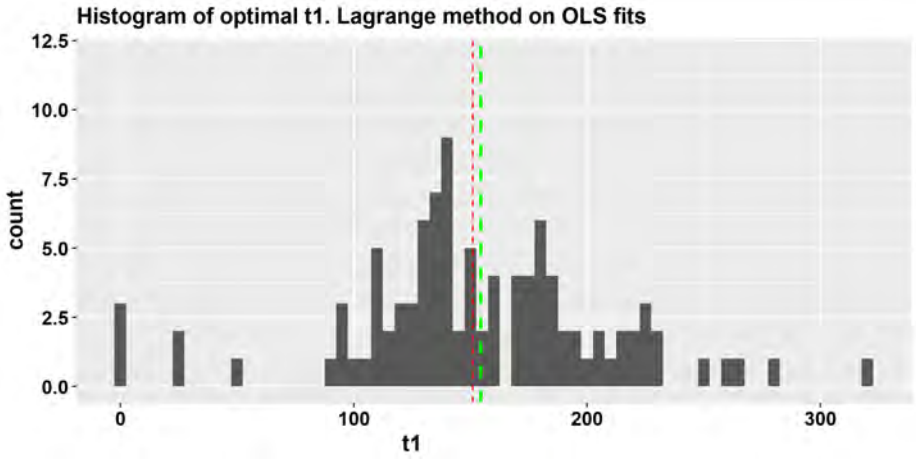

Figure 15: Histogram of the optimally selected t_1 for the Lagrange regularisation method based on the LPPLS OLS fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

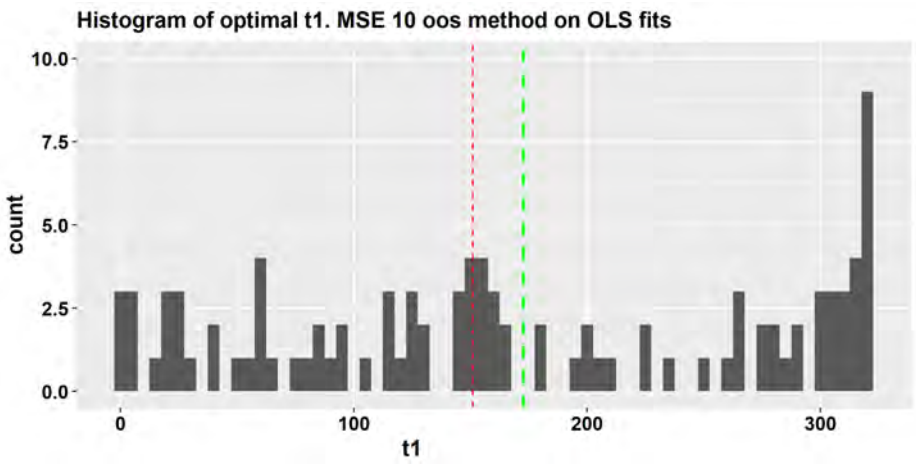

Figure 16: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 10$ leave-out method based on the LPPLS OLS fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

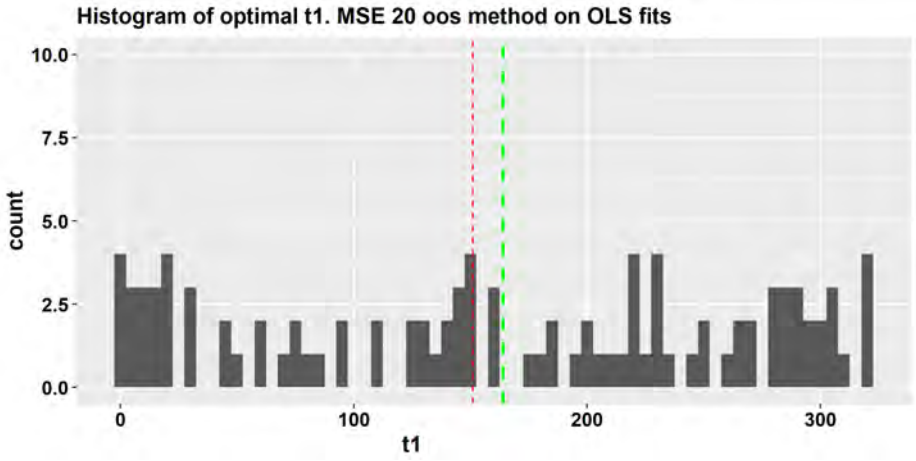

Figure 17: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 20$ leave-out method based on the LPPLS OLS fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

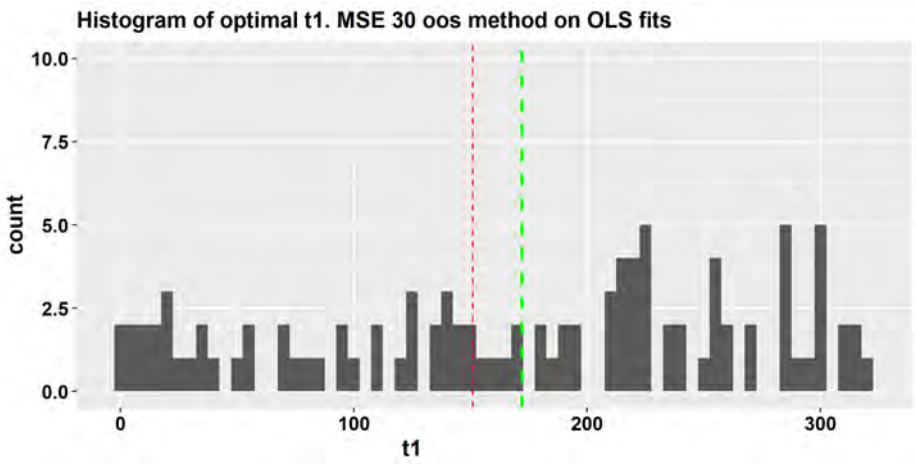

Figure 18: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 30$ leave-out method based on the LPPLS OLS fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

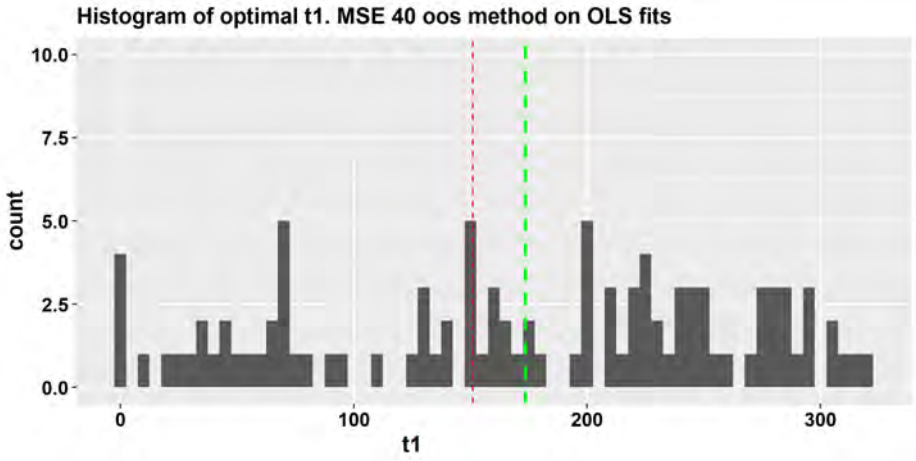

Figure 19: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 40$ leave-out method based on the LPPLS OLS fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

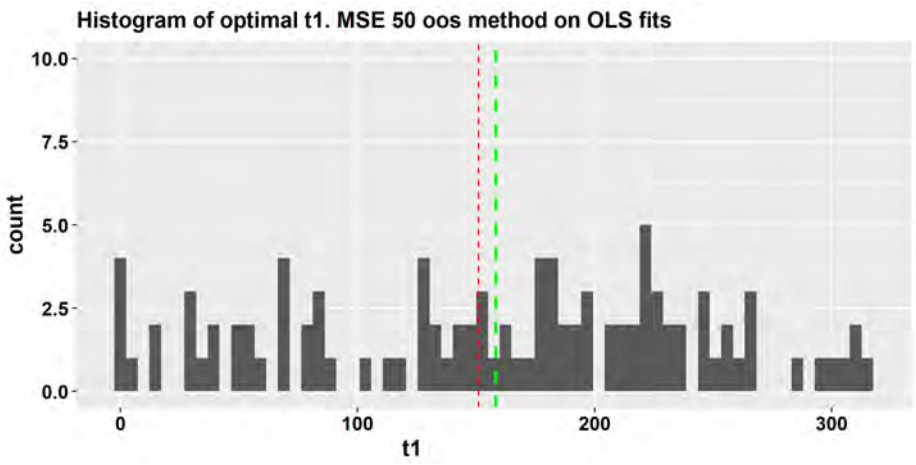

Figure 20: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 50$ leave-out method based on the LPPLS OLS fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

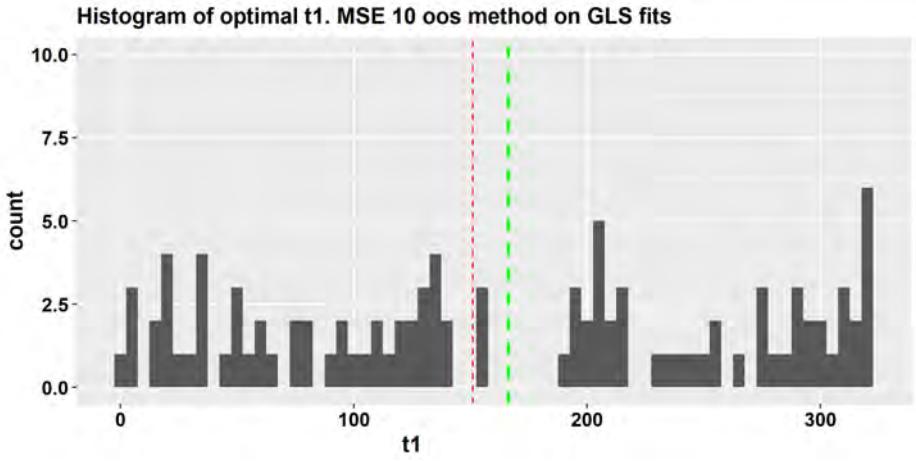

Figure 21: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 10$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

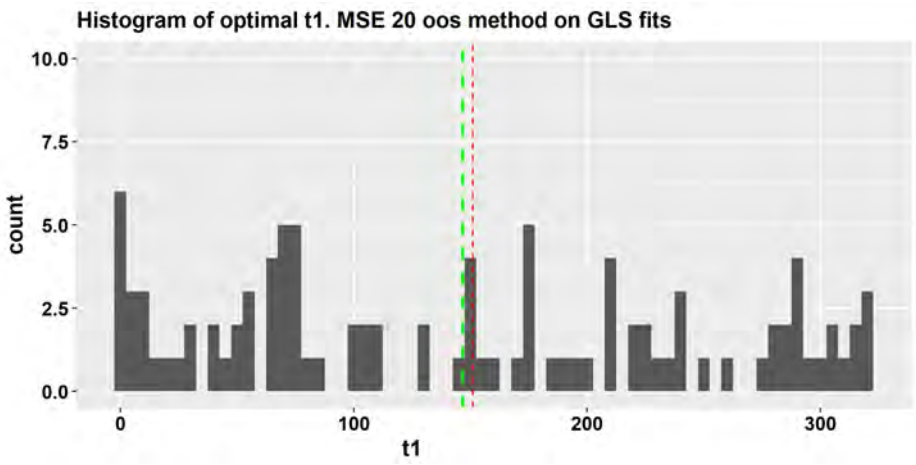

Figure 22: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 20$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

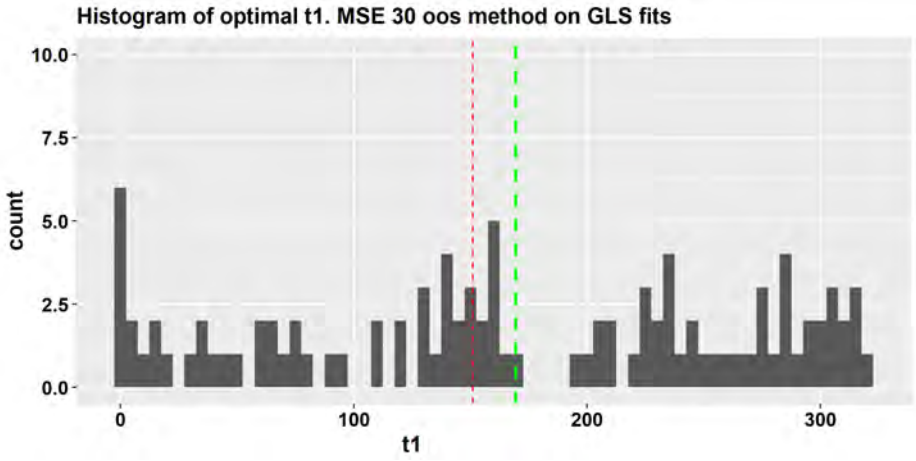

Figure 23: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 30$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

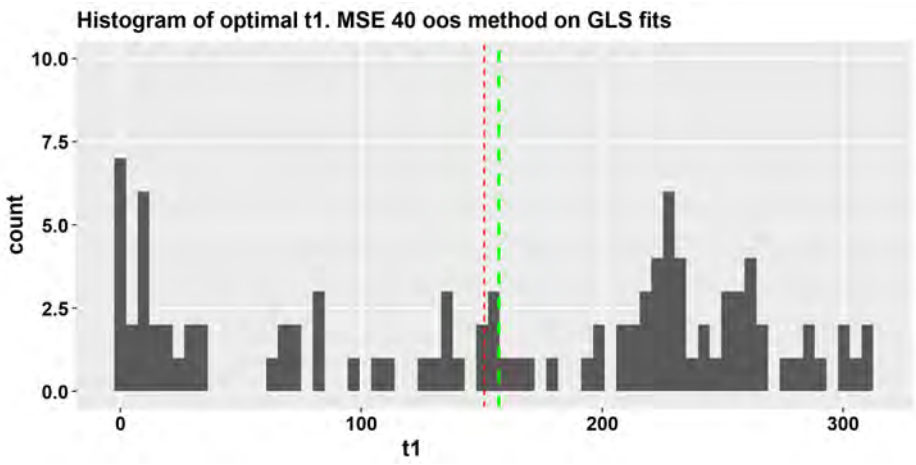

Figure 24: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 40$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

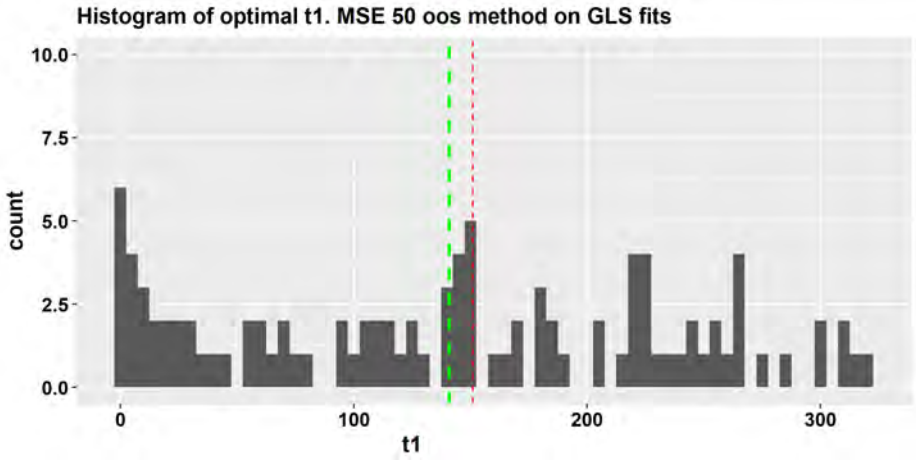

Figure 25: Histogram of the optimally selected t_1 for the MSPE prediction with $n = 50$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

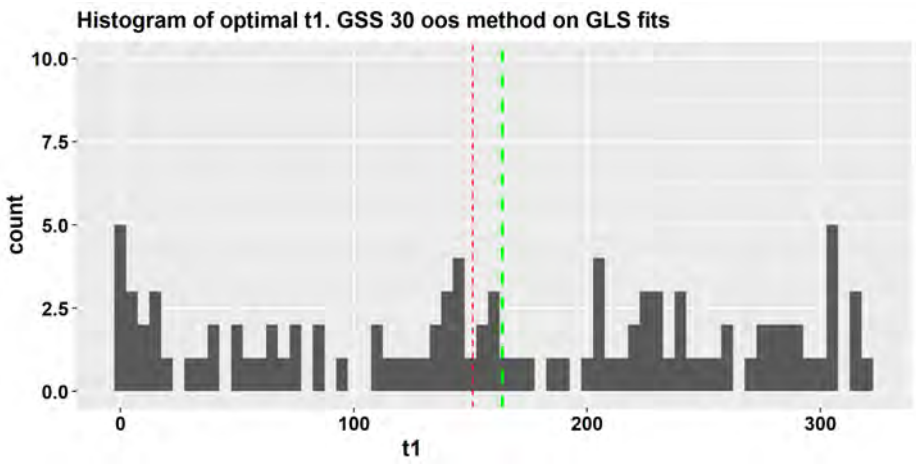

Figure 26: Histogram of the optimally selected t_1 for the GSS prediction with $n = 30$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

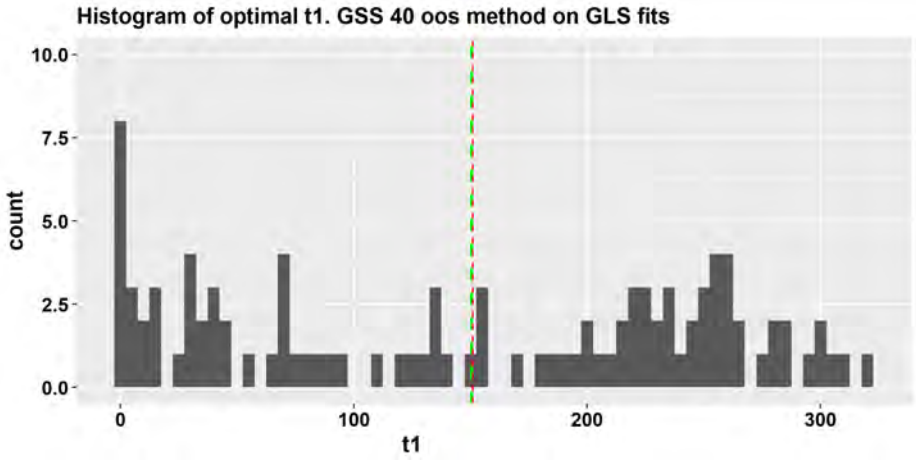

Figure 27: Histogram of the optimally selected t_1 for the GSS prediction with $n = 40$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

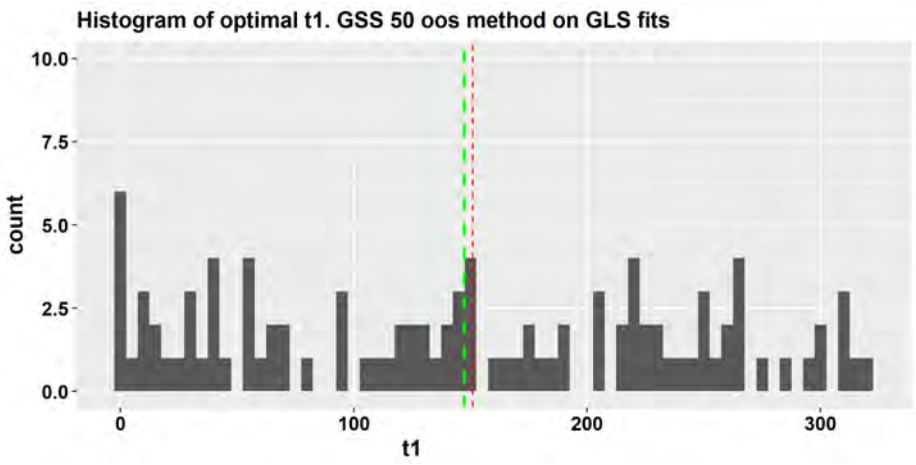

Figure 28: Histogram of the optimally selected t_1 for the GSS prediction with $n = 50$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

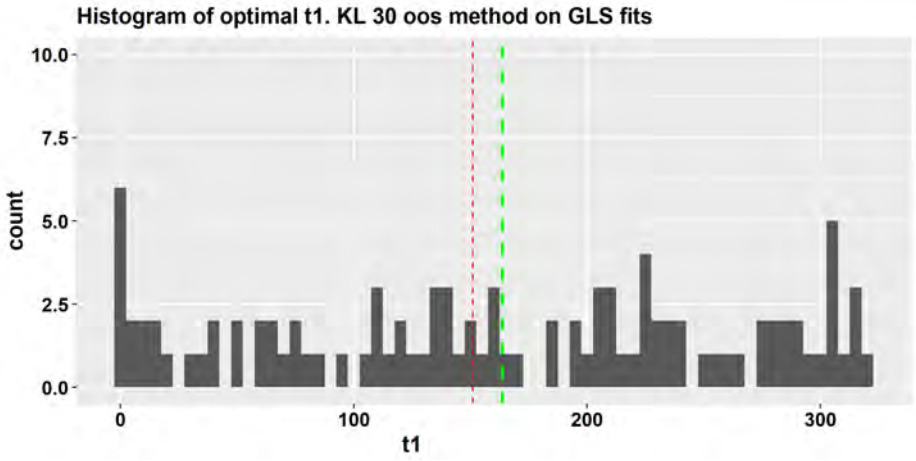

Figure 29: Histogram of the optimally selected t_1 for the KL prediction with $n = 30$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

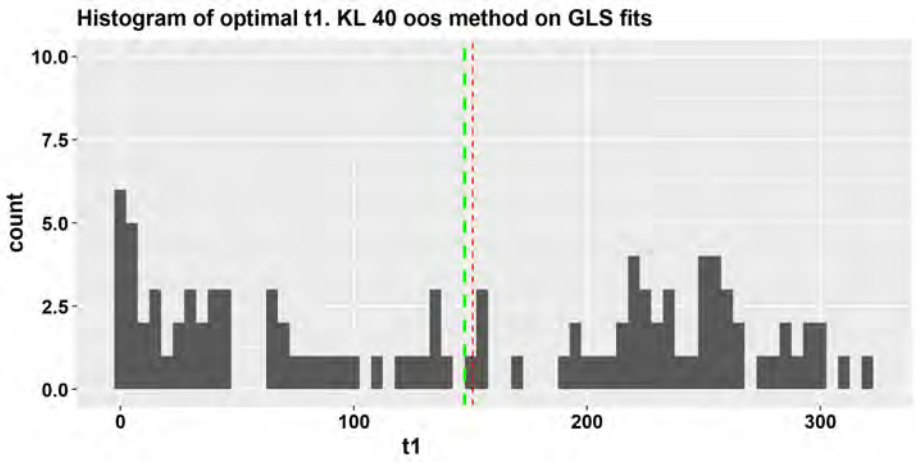

Figure 30: Histogram of the optimally selected t_1 for the KL prediction with $n = 40$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

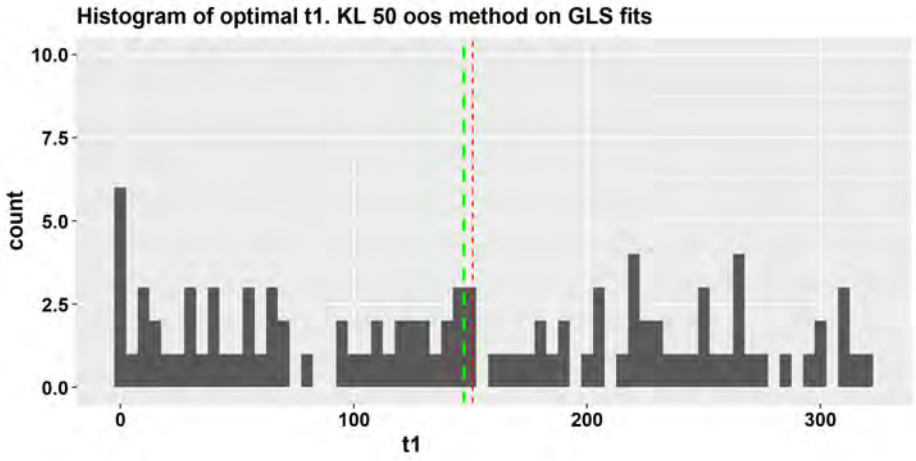

Figure 31: Histogram of the optimally selected t_1 for the KL prediction with $n = 50$ leave-out method based on the LPPLS GLSH(1,0) fits. The red line indicates the true $t_1 = 151$ and the green the empirical mean of the selected t_1.

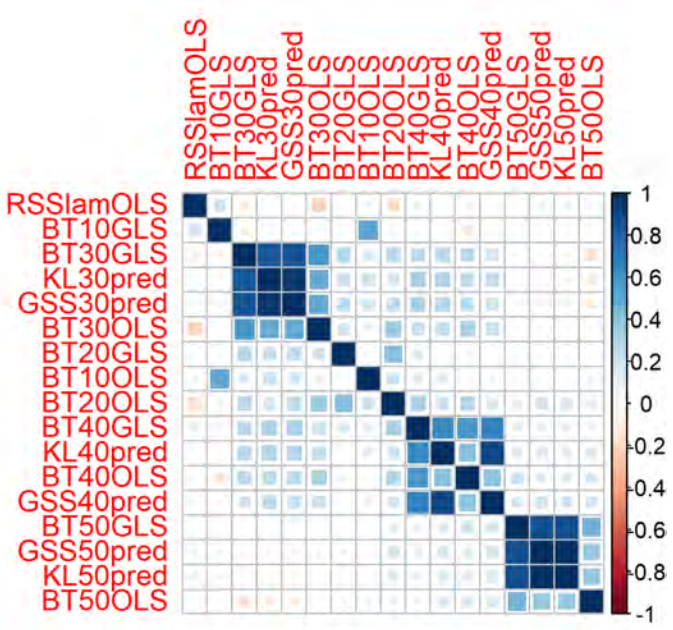

Figure 32: Correlation plot for the optimally selected t_1 for each method over the whole simulated dataset. The columns/rows of the matrix are reordered according to the angular order of the two most meaningful eigenvectors. RSSlamOLS stands for the Lagrange regularisation method, BT stands for back-testing with MSPE, GSS stands for the generalized sum of squares and KL for the Kullback-Leibler divergence error function. The size of leave-out data n is also included in the names. From the plot, we see that clustering is often based on identical n, the size of leave-out data, at least for $n \geq 30$ but not on the particular model selection criterion. The Lagrange regularisation method is isolated.

5.4 t_1 selection for empirical bubbles

We selected the optimal t_1 for each of the historical bubbles using our models, the LPPLS OLS and the LPPLS GLSH(1,0), and the different loss functions. The candidate $t_1' \in \{1, 2, ..., t_2 - 80\}$ were considered. The beginning was selected such that it was early enough to contain the full bubble signal. The parameter t_2, the final value of the data used for fitting, was selected 10 days before the peak of the corresponding bubble, that is two weeks in trading days. The window for candidate t_c was $t_c' \in [t_2, t_2 + 50]$ where the true peak is at $t_2 + 10$. For the other parameters, $m \in [0, 2]$ and $\omega \in [4, 25]$ is the space searched over.

The results of this study are inconclusive(see table 5). The methods select quite different initial points in general. However, as in the simulation study, selection criteria with identical n, the size of leave-out data, perform similarly. The selected window for each bubble and predicted t_c corresponding to the optimally selected t_1 are shown in table 5 and figures 33-47.

For the bubble in Argentina(MERVAL index) in 1994, most methods agree that the bubble is coming soon. For the other bubbles, predictions vary greatly, and no method is consistently more accurate than the others.

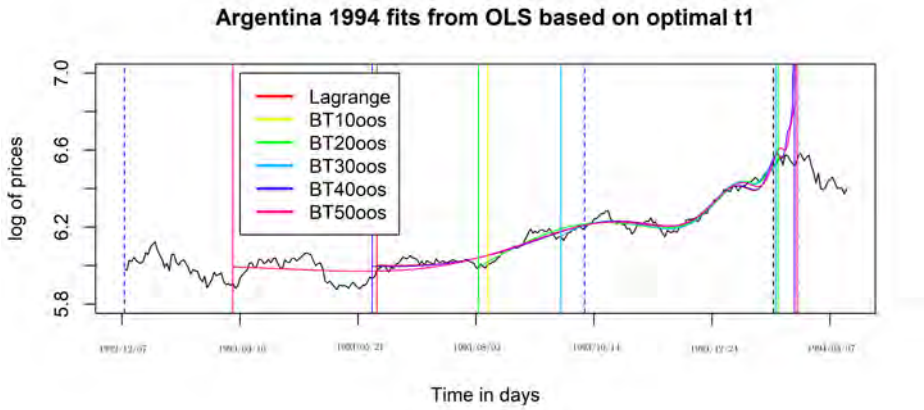

Figure 33: Optimal OLS LPPLS fits on the Argentina(MERVAL index) 1994 bubble for each of i) Lagrange regularisation method ii) MSPE with $n = 10$ iii) MSPE with $n = 20$ iv) MSPE with $n = 30$ v) MSPE with $n = 40$ vi) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 1992/12/07$, $t_2 = 1994/02/02$ and the true peak of the bubble at $t_c = 1994/02/16$.

Table 5: The table shows the optimally selected t_1 and the corresponding t_c from that particular fit for each of the empirical bubbles considered in the format $(\delta_{t_1^\star}, \delta_{\hat{t}_c})$ where $\delta_{t_1^\star} = t_1^\star - t_0$, t_0 being the beginning of the search window for t_1, and $\delta_{\hat{t}_c} = \hat{t}_c - t_c$ is the difference in trading days between the predicted peak and the actual peak of the corresponding bubble. Here, GLS corresponds to the GLSH(1,0) LPPLS model. Soonest search date for t_1 for the bubbles are at i) Argentina: $t_0 = 1992/12/07$ ii) Brazil: $t_0 = 1996/01/04$ iii) Philippines: $t_0 = 1992/07/23$ iv) Thailand: $t_0 = 1992/12/18$ v) Sugar: $t_0 = 2005/01/05$. The end point of the data used for fitting(t_2) are at i) Argentina: $t_2 = 1994/02/02$ ii) Brazil: $t_2 = 1997/06/24$ iii) Philippines: $t_2 = 1993/12/16$ iv) Thailand: $t_2 = 1993/12/16$ v) Sugar: $t_2 = 2006/01/20$. True peak of the bubbles are at i) Argentina: $t_c = 1994/02/16$ ii) Brazil: $t_c = 1997/07/08$ iii) Philippines: $t_c = 1994/01/04$ iv) Thailand: $t_c = 1993/12/30$ v) Sugar: $t_c = 2006/02/03$.

Method	n	Type	Argentina 1994	Brazil 1997	Philippines 1994	Thailand 1994	Sugar 2005
Lagrange	0	OLS	(107,-2)	(56,-10)	(93,40)	(93,-6)	(52,9)
MSPE	10	OLS	(154,-9)	(125,-10)	(92,40)	(124,32)	(108,40)
MSPE	20	OLS	(150,-9)	(105,-10)	(91,40)	(148,40)	(162,24)
MSPE	30	OLS	(185,-10)	(251,-7)	(198,7)	(153,40)	(29,7)
MSPE	40	OLS	(105,-2)	(62,-10)	(205,9)	(98,-5)	(174,20)
MSPE	50	OLS	(46,-1)	(61,-10)	(237,-9)	(74,-2)	(161,23)
MSPE	10	GLS	(154,-8)	(25,40)	(3,0)	(88,-7)	(124,40)
MSPE	20	GLS	(155,-8)	(110,-8)	(27,11)	(148,35)	(162,22)
MSPE	30	GLS	(185,-7)	(158,-4)	(200,8)	(153,40)	(26,8)
MSPE	40	GLS	(106,-4)	(259,40)	(206,8)	(64,5)	(40,10)
MSPE	50	GLS	(40,-5)	(68,-7)	(237,11)	(138,17)	(162,22)
GSS	30	GLS	(185,-7)	(158,-4)	(200,8)	(153,40)	(26,8)
GSS	40	GLS	(106,-4)	(259,40)	(206,8)	(64,5)	(42,10)
GSS	50	GLS	(39,-5)	(68,-7)	(237,11)	(150,40)	(134,34)
KL	30	GLS	(185,-7)	(158,-4)	(200,8)	(153,40)	(26,8)
KL	40	GLS	(106,-4)	(259,40)	(206,8)	(64,5)	(81,33)
KL	50	GLS	(39,-5)	(68,-7)	(237,11)	(150,40)	(162,22)

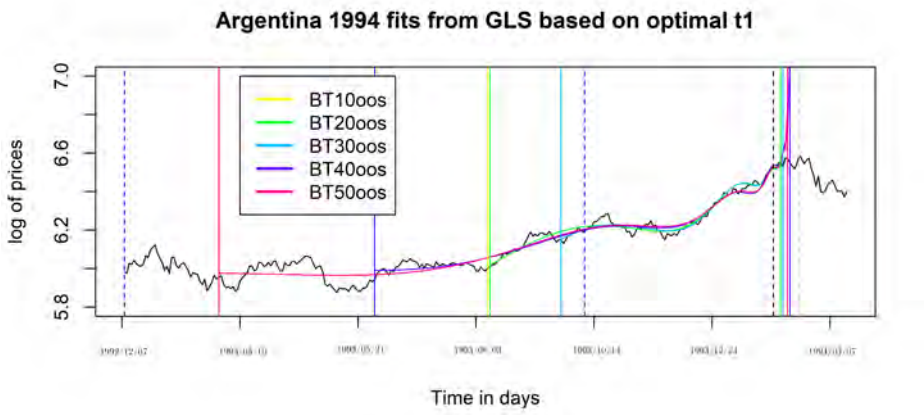

Figure 34: Optimal GLSH(1,0) LPPLS fits on the Argentina(MERVAL index) 1994 bubble for each of i) MSPE with $n = 10$ ii) MSPE with $n = 20$ iii) MSPE with $n = 30$ iv) MSPE with $n = 40$ v) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 1992/12/07$, $t_2 = 1994/02/02$ and the true peak of the bubble at $t_c = 1994/02/16$.

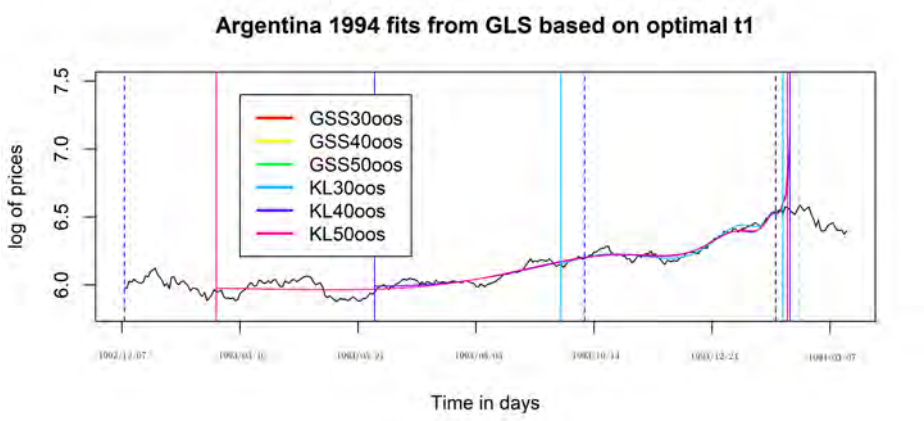

Figure 35: Optimal GLSH(1,0) LPPLS fits on the Argentina(MERVAL index) 1994 bubble for each of i) GSS with $n = 30$ ii) GSS with $n = 40$ iii) GSS with $n = 50$ iv) KL with $n = 30$ v) KL with $n = 40$ iv) KL with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Note here for each n the KL and GSS selected identical t_1. Here, the search space for t_1 starts at $t_0 = 1992/12/07$, $t_2 = 1994/02/02$ and the true peak of the bubble at $t_c = 1994/02/16$.

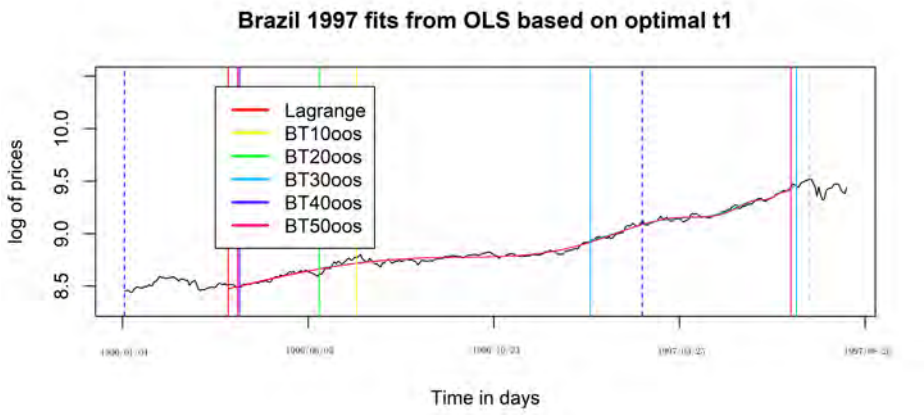

Figure 36: Optimal OLS LPPLS fits on the Brazil(IBOVESPA Index) 1997 bubble for each of i) Lagrange regularisation method ii) MSPE with $n = 10$ iii) MSPE with $n = 20$ iv) MSPE with $n = 30$ v) MSPE with $n = 40$ vi) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 1996/01/04$, $t_2 = 1997/06/24$ and the true peak of the bubble at $t_c = 1997/07/08$.

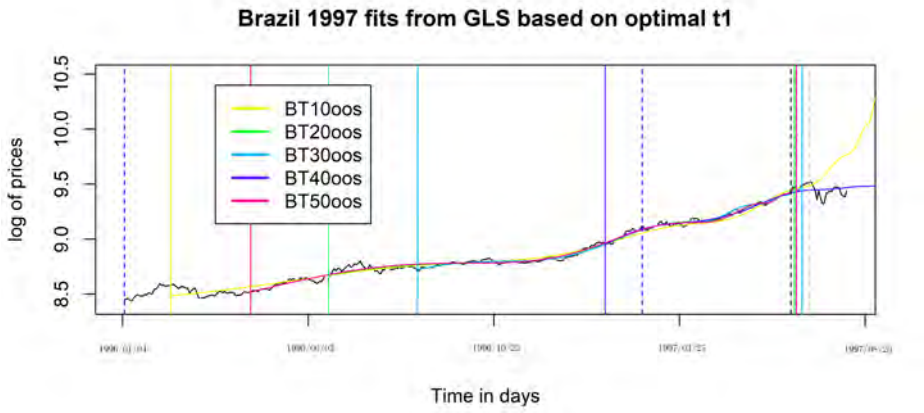

Figure 37: Optimal GLSH(1,0) LPPLS fits on the Brazil(IBOVESPA Index) 1997 bubble for each of i) MSPE with $n = 10$ ii) MSPE with $n = 20$ iii) MSPE with $n = 30$ iv) MSPE with $n = 40$ v) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 1996/01/04$, $t_2 = 1997/06/24$ and the true peak of the bubble at $t_c = 1997/07/08$.

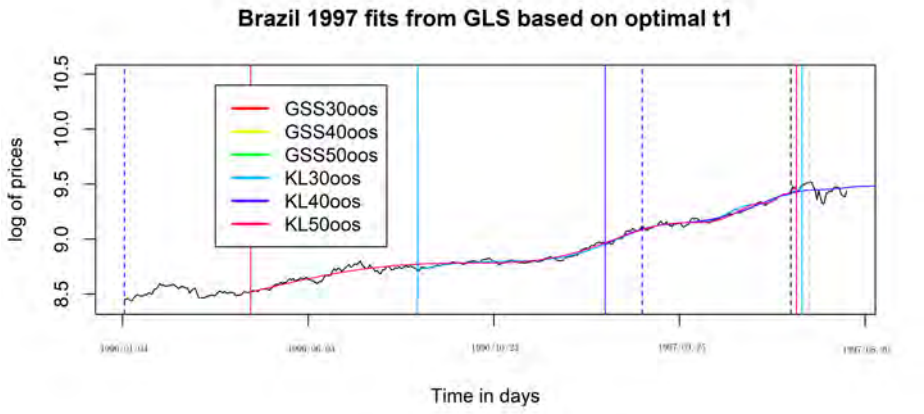

Figure 38: Optimal GLSH(1,0) LPPLS fits on the Brazil(IBOVESPA Index) 1997 bubble for each of i) GSS with $n = 30$ ii) GSS with $n = 40$ iii) GSS with $n = 50$ iv) KL with $n = 30$ v) KL with $n = 40$ iv) KL with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Note here for each n the KL and GSS selected identical t_1. Here, the search space for t_1 starts at $t_0 = 1996/01/04$, $t_2 = 1997/06/24$ and the true peak of the bubble at $t_c = 1997/07/08$.

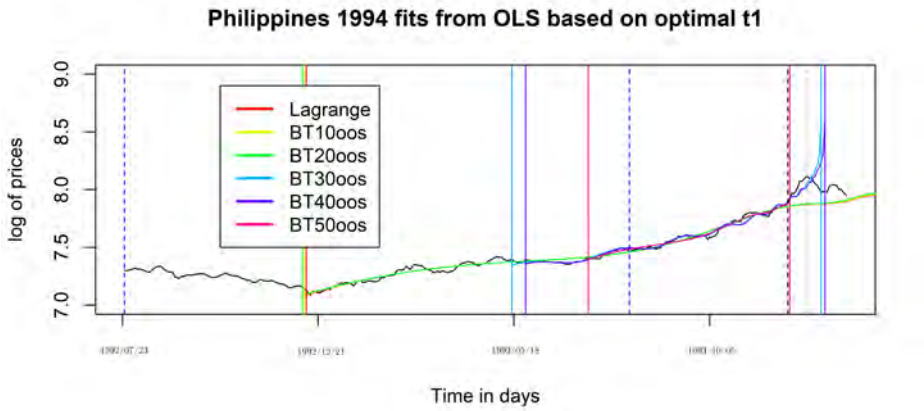

Figure 39: Optimal OLS LPPLS fits on the Philippines(PSEi Index) 1994 bubble for each of i) Lagrange regularisation method ii) MSPE with $n = 10$ iii) MSPE with $n = 20$ iv) MSPE with $n = 30$ v) MSPE with $n = 40$ vi) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 1992/07/23$, $t_2 = 1993/12/16$ and the true peak of the bubble at $t_c = 1994/01/04$.

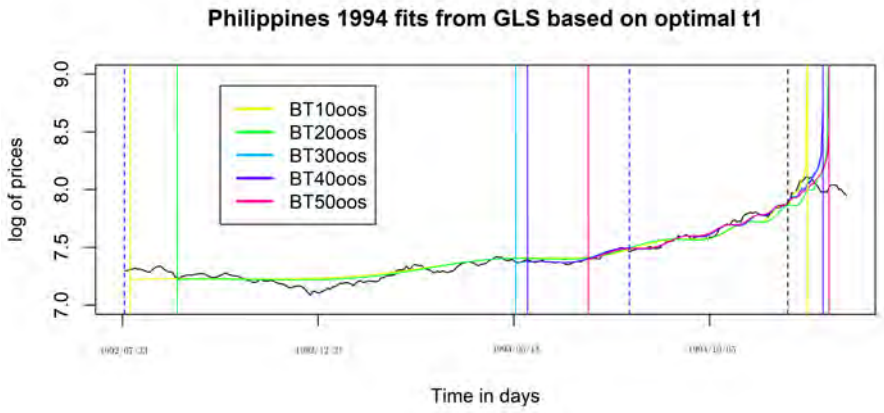

Figure 40: Optimal GLSH(1,0) LPPLS fits on the Philippines(PSEi Index) 1994 bubble for each of i) MSPE with $n = 10$ ii) MSPE with $n = 20$ iii) MSPE with $n = 30$ iv) MSPE with $n = 40$ v) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 1992/07/23$, $t_2 = 1993/12/16$ and the true peak of the bubble at $t_c = 1994/01/04$.

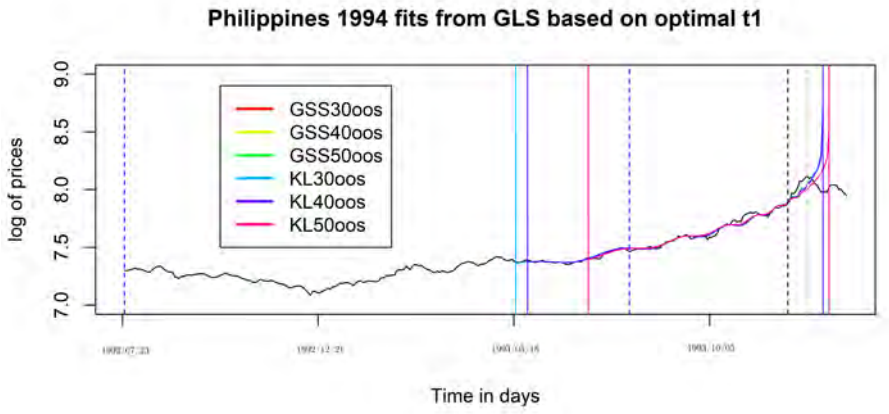

Figure 41: Optimal GLSH(1,0) LPPLS fits on the Philippines(PSEi Index) 1994 bubble for each of i) GSS with $n = 30$ ii) GSS with $n = 40$ iii) GSS with $n = 50$ iv) KL with $n = 30$ v) KL with $n = 40$ iv) KL with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Note here for each n the KL and GSS selected identical t_1. Here, the search space for t_1 starts at $t_0 = 1992/07/23$, $t_2 = 1993/12/16$ and the true peak of the bubble at $t_c = 1994/01/04$.

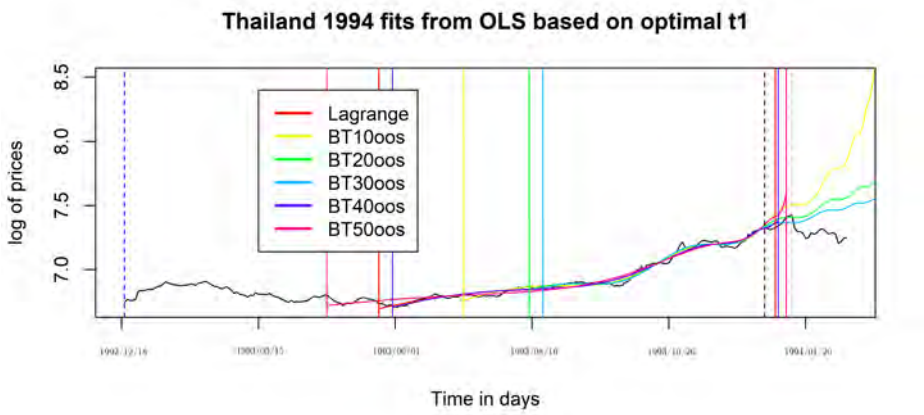

Figure 42: Optimal OLS LPPLS fits on the Thailand(Bangkok SET Index) 1994 bubble for each of i) Lagrange regularisation method ii) MSPE with $n = 10$ iii) MSPE with $n = 20$ iv) MSPE with $n = 30$ v) MSPE with $n = 40$ vi) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Also, the time series omits data from 1993/12/30 to 1994/01/14. Here, the search space for t_1 starts at $t_0 = 1992/12/18$, $t_2 = 1993/12/16$ and the true peak of the bubble at $t_c = 1993/12/30$.

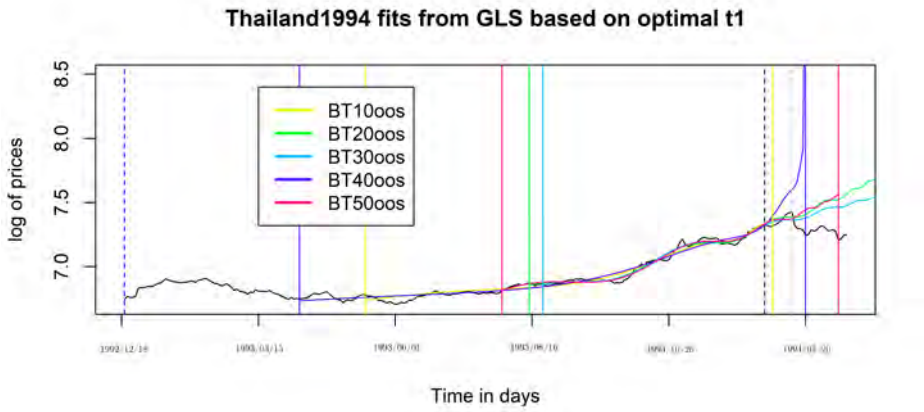

Figure 43: Optimal GLSH(1,0) LPPLS fits on the Thailand(Bangkok SET Index) 1994 bubble for each of i) MSPE with $n = 10$ ii) MSPE with $n = 20$ iii) MSPE with $n = 30$ iv) MSPE with $n = 40$ v) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Also, the time series omits data from 1993/12/30 to 1994/01/14. Here, the search space for t_1 starts at $t_0 = 1992/12/18$, $t_2 = 1993/12/16$ and the true peak of the bubble at $t_c = 1993/12/30$.

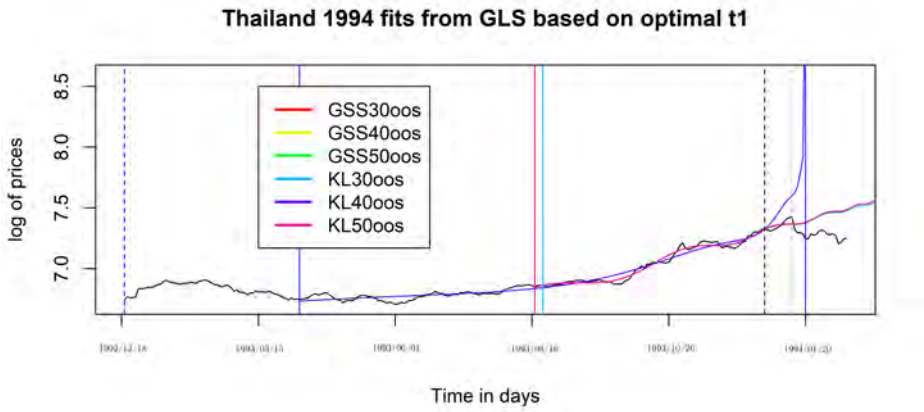

Figure 44: Optimal GLSH(1,0) LPPLS fits on the Thailand(Bangkok SET Index) 1994 bubble for each of i) GSS with $n = 30$ ii) GSS with $n = 40$ iii) GSS with $n = 50$ iv) KL with $n = 30$ v) KL with $n = 40$ iv) KL with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Note here for each n the KL and GSS selected identical t_1. Also, the time series omits data from 1993/12/30 to 1994/01/14. Here, the search space for t_1 starts at $t_0 = 1992/12/18$, $t_2 = 1993/12/16$ and the true peak of the bubble at $t_c = 1993/12/30$.

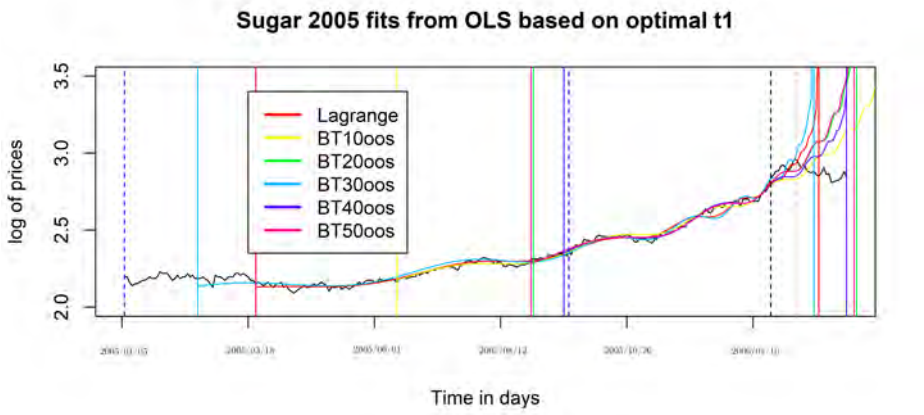

Figure 45: Optimal OLS LPPLS fits on the Sugar 2005 bubble for each of i) Lagrange regularisation method ii) MSPE with $n = 10$ iii) MSPE with $n = 20$ iv) MSPE with $n = 30$ v) MSPE with $n = 40$ vi) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 2005/01/05$, $t_2 = 2006/01/20$ and the true peak of the bubble at $t_c = 2006/02/03$.

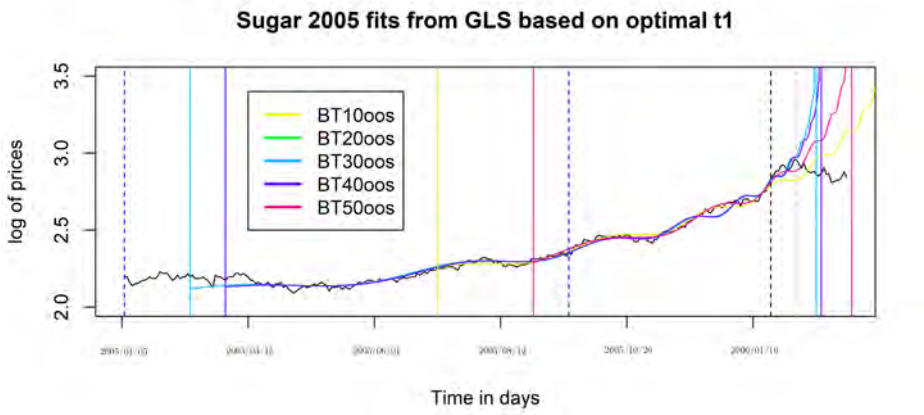

Figure 46: Optimal GLSH(1,0) LPPLS fits on the Sugar 2005 bubble for each of i) MSPE with $n = 10$ ii) MSPE with $n = 20$ iii) MSPE with $n = 30$ iv) MSPE with $n = 40$ v) MSPE with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 2005/01/05$, $t_2 = 2006/01/20$ and the true peak of the bubble at $t_c = 2006/02/03$.

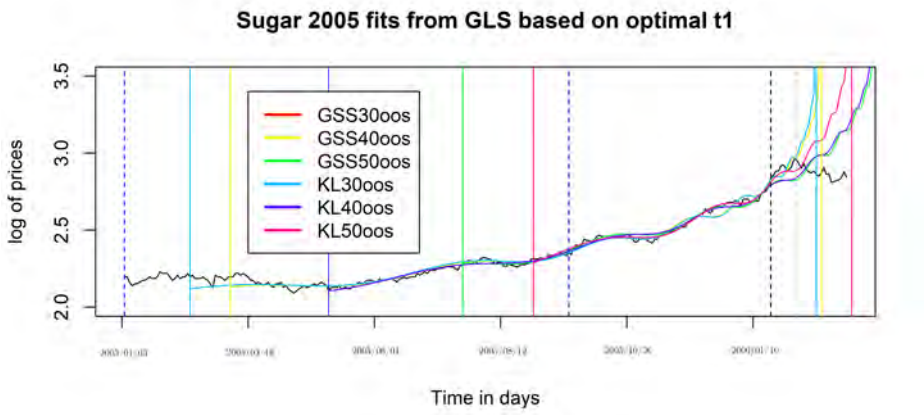

Figure 47: Optimal GLSH(1,0) LPPLS fits on the Sugar 2005 bubble for each of i) GSS with $n = 30$ ii) GSS with $n = 40$ iii) GSS with $n = 50$ iv) KL with $n = 30$ v) KL with $n = 40$ iv) KL with $n = 50$. Here the blue dashed lines indicate the window for each t_1 is searched on, the black dashed line is the t_2, end of the data used in the fitting. Ten days later is the grey dashed line indicating the real peak of the bubble. The colored lines indicate the selected t_1 and predicted t_c for each fit. Here, the search space for t_1 starts at $t_0 = 2005/01/05$, $t_2 = 2006/01/20$ and the true peak of the bubble at $t_c = 2006/02/03$.

6 Summary

The goal of this thesis was to compare existing estimation methods for the LPPLS method to new methods and improve the selection criteria for the beginning of a bubble. Further, the model was extended such that the linear parameters have a probabilistic description which can be used, among other methods, to access the strength of the bubble signal and which corrects appropriately for different sample sizes and variance structures.

The likelihood for both the LPPLS GLSH(1,0) model and the LPPLS OLS model were derived in a Bayesian setting and the two models were compared on several known empirical bubbles. From standard diagnostics plots and standard model selection criteria, the GLSH(1,0) was a significant improvement on the LPPLS OLS model although its model assumptions were sometimes violated. Then, we proceeded to propose selection criteria for the optimal starting points of bubbles. This was based on predictive performance instead of the existing methods which depend on in-sample performance. We compared the two competing models on synthetic data with long memory ARFIMA model coupled with a LPPLS signal with strong auto-regressive residuals and non-smooth variance structure of the residuals, with white noise added on top. For the simulation study, the LPPLS GLSH(1,0) was superior in selecting t_1 and predicting t_c, especially when leaving smaller validation set out to access predictive performance. Finally, the competing selection criteria were tested on some empirical bubbles, but as the sample size is small, the results do not give any firm conclusions.

In future work it would be interesting to compare the different t_1 selection methods in a much wider range of situations and see the merits and drawbacks of each of the methods. Further, the computational challenge for fitting the model needs to be accessed. As we need to fit our model for each proposed t_1' on a window of length Δ, our algorithm is linear in Δ. Perhaps we could reduce the search space, for example, reduce the computational complexity by using binary search, cutting the computational complexity of the fitting window to $\ln \Delta$. The LPPLS GLSH(1,0) model can easily be extended, for example, increase the number of auto-regressive terms as sometimes indicated by the auto-correlation plots or a different structure of the variance of the innovations. Finally, in the Bayesian setting, priors can easily be added and perhaps introduce the random nature of the design matrix $\mathbf{X}$ into our likelihood.

References

[1] Fischer Black. "Noise". In: *The journal of finance* 41.3 (1986), pp. 528–543.

[2] Hans-Christian Graf v Bothmer and Christian Meister. "Predicting critical crashes? A new restriction for the free variables". In: *Physica A: Statistical Mechanics and its Applications* 320 (2003), pp. 539–547.

[3] Kenneth P Burnham and David R Anderson. "Multimodel inference: understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2 (2004), pp. 261–304.

[4] Guilherme Demos and Didier Sornette. "Lagrange regularisation approach to compare nested data sets and determine objectively financial bubbles' inceptions". In: *arXiv preprint arXiv:1707.07162* (2017).

[5] Vladimir Filimonov and Didier Sornette. "A stable and robust calibration scheme of the log-periodic power law model". In: *Physica A: Statistical Mechanics and its Applications* 392.17 (2013), pp. 3698–3707.

[6] Arthur S Goldberger. "Best linear unbiased prediction in the generalized linear regression model". In: *Journal of the American Statistical Association* 57.298 (1962), pp. 369–375.

[7] Andreas Hüsler, Didier Sornette, and Cars H Hommes. "Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price". In: *Journal of Economic Behavior & Organization* 92 (2013), pp. 304–316.

[8] Kayo Ide and Didier Sornette. "Oscillatory finite-time singularities in finance, population and rupture". In: *Physica A: Statistical Mechanics and its Applications* 307.1-2 (2002), pp. 63–106.

[9] Zhi-Qiang Jiang et al. "Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 Chinese stock market bubbles". In: *Journal of economic behavior & organization* 74.3 (2010), pp. 149–162.

[10] Anders Johansen, Olivier Ledoit, and Didier Sornette. "Crashes as critical points". In: *International Journal of Theoretical and Applied Finance* 3.02 (2000), pp. 219–255.

[11] Anders Johansen, Didier Sornette, and Olivier Ledoit. "Predicting financial crashes using discrete scale invariance". In: *Journal of Risk 1 (4), 5-32* (1999).

[12] Anders Johansen, Didier Sornette, et al. "Shocks, crashes and bubbles in financial markets". In: *Brussels Economic Review* 53.2 (2010), pp. 201–253.

[13] Charles Poor Kindleberger and Charles P Kindleberber. *Economic response: comparative studies in trade, finance, and growth*. Harvard University Press, 1978.

[14] Matthias Leiss, Heinrich H Nax, and Didier Sornette. "Super-exponential growth expectations and the global financial crisis". In: *Journal of Economic Dynamics and Control* 55 (2015), pp. 1–13.

[15] Jeremy J Siegel. "What is an asset price bubble? An operational definition". In: *European financial management* 9.1 (2003), pp. 11–24.

[16] Didier Sornette. "Discrete-scale invariance and complex dimensions". In: *Physics reports* 297.5 (1998), pp. 239–270.

[17] Didier Sornette. "Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth". In: *Proceedings of the National Academy of Sciences* 99.suppl 1 (2002), pp. 2522–2529.

[18] Didier Sornette. *Why stock markets crash: critical events in complex financial systems*. Princeton University Press, 2017.

[19] Didier Sornette and Peter Cauwels. "Financial bubbles: Mechanism, diagnostic and state of the world." In: *Review of Behavioral Economics.* 2(3) (2015), pp. 279–3055.

[20] Didier Sornette and Peter Cauwels. "Managing risk in a creepy world". In: *Journal of Risk Management in Financial Institutions* 8.1 (2015), pp. 83–108.

[21] Didier Sornette, Anders Johansen, and Jean-Philippe Bouchaud. "Stock market crashes, precursors and replicas". In: *Journal de Physique I* 6.1 (1996), pp. 167–175.

[22] Didier Sornette et al. "Clarifications to questions and criticisms on the Johansen–Ledoit–Sornette financial bubble model". In: *Physica A: Statistical Mechanics and its Applications* 392.19 (2013), pp. 4417–4428.

[23] Spencer Wheatley et al. "Are Bitcoin Bubbles Predictable? Combining a Generalized Metcalfe's Law and the LPPLS Model". In: (2018).

[24] Qun Zhang, Qunzhi Zhang, and Didier Sornette. "Early warning signals of financial crises with multi-scale quantile regressions of Log-Periodic Power Law Singularities". In: *PloS one* 11.11 (2016), e0165819.

A Appendix

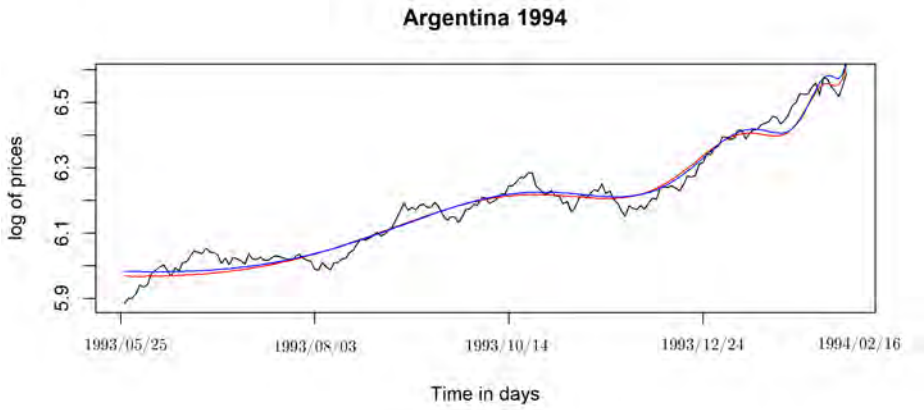

Figure 48: The fit of the full bubble in Argentina(MERVAL index) in 1994(closing prices) for the OLS LPPLS model in blue and the GLSH(1,0) LPPLS model in red. As usual, the GLSH(1,0) fit is a bit smoother, but the signals are otherwise similar. The fitted parameters for the LPPLS signals($t_2 = 1994/02/16$ corresponds to the end of the fitting window, the true peak of the bubble) are $\hat{A}_{OLS} = 7.5271, \hat{B}_{OLS} = -0.6718, \hat{C}_{1,OLS} = -0.0201, \hat{C}_{2,OLS} = 0.0137, \hat{m}_{OLS} = 0.1551, \hat{\omega}_{OLS} = 6.0663, \hat{t_c}_{,OLS} = t_2 + 6$ and $\hat{A}_{GLS} = 7.1689, \hat{B}_{GLS} = -0.3754, \hat{C}_{1,GLS} = -0.0177, \hat{C}_{2,GLS} = 0.0018, \hat{m}_{GLS} = 0.2163, \hat{\omega}_{GLS} = 6.1429, \hat{t_c}_{,GLS} = t_2 + 6, \hat{\rho} = 0.9019$.

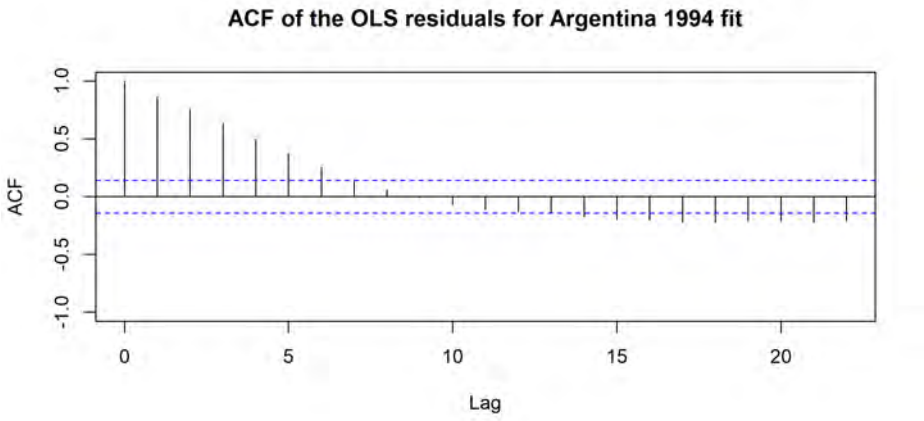

Figure 49: The autocorrelation function for the LPPLS OLS fit for the full bubble in Argentina in 1994. Strong auto-regressive part is evident. Here it is not clear if the decay is exponential or linear, at least the estimated auto-regressive parameter is close to 1.

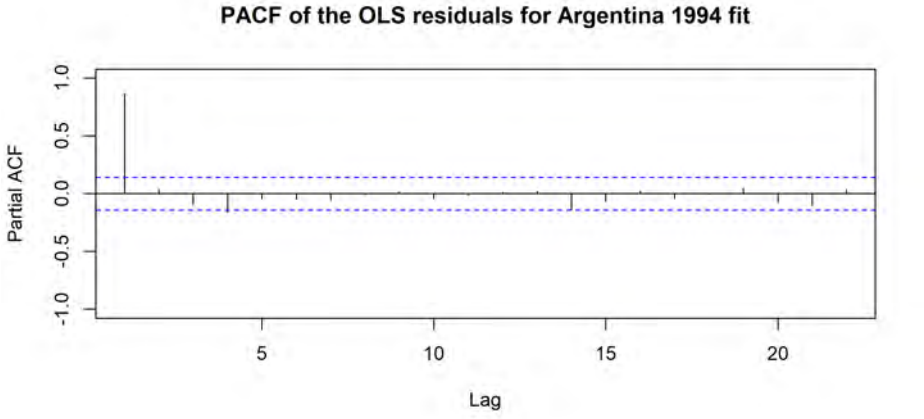

Figure 50: The partial autocorrelation function for the LPPLS OLS fit for the bubble in Argentina in 1994. Strong auto-regressive part is evident of order 1.

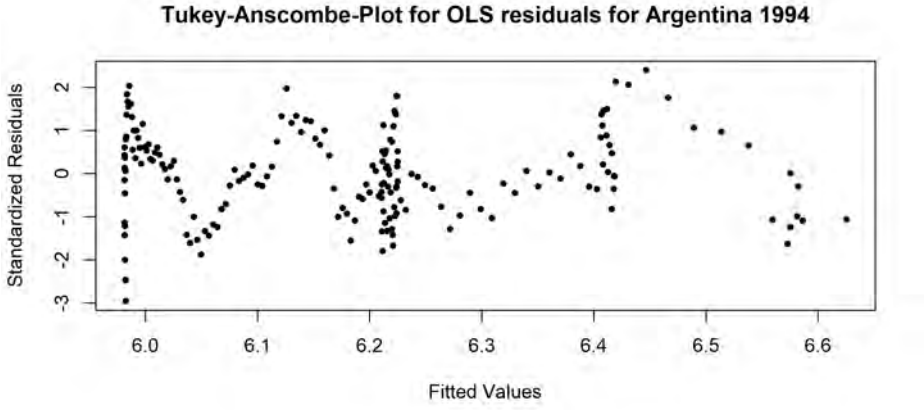

Figure 51: Tukey-Anscombe plot for the LPPLS OLS fit of the standardized residuals for the bubble in Argentina in 1994. A structure is evident, and thus the model assumptions violated.

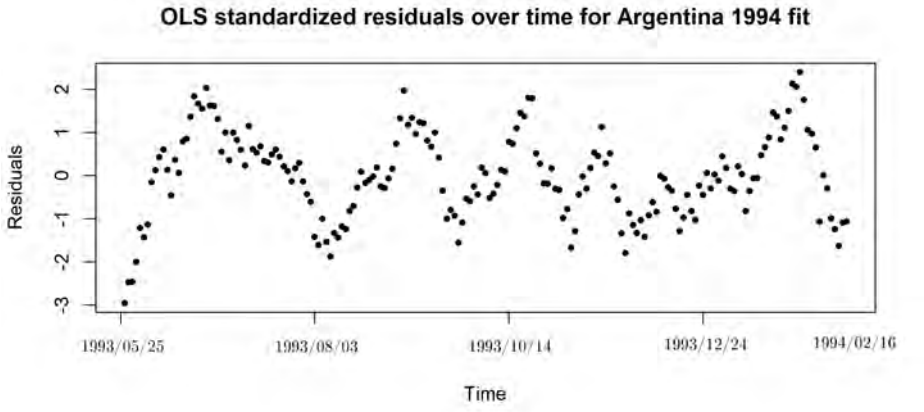

Figure 52: The plot shows standardized residuals over time for the LPPLS OLS fit for the bubble in Argentina in 1994. A structure is evident, and thus the model assumptions violated.

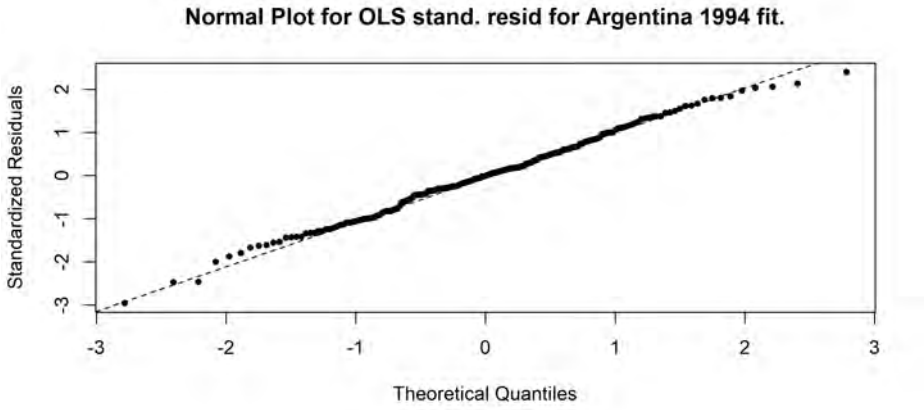

Figure 53: The quantile-quantile normal plot for the LPPLS OLS fit for the bubble in Argentina in 1994. The normality is not violated. However, the strong auto-regressive structure needs to be corrected for before considering the Q-Q plot.

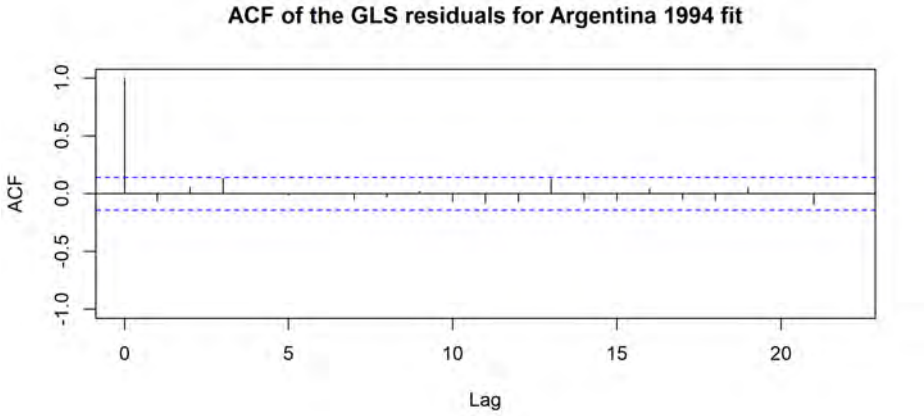

Figure 54: The autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Argentina in 1994. The transformation seems to have uncorrelated the residuals for the most part.

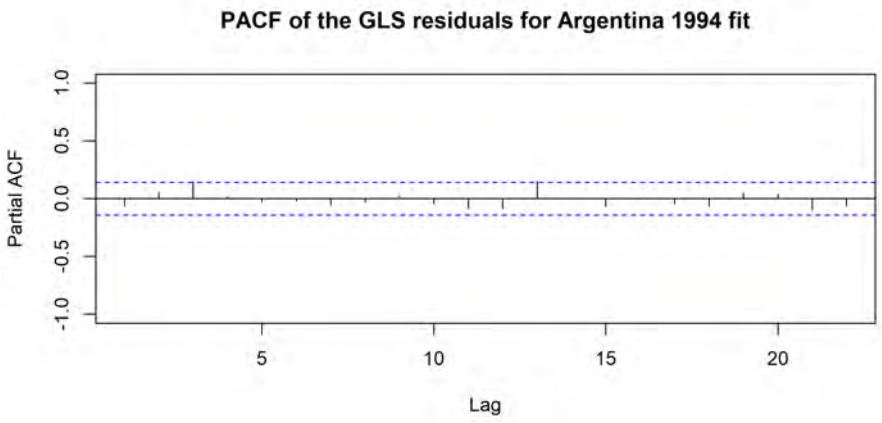

Figure 55: The partial autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Argentina in 1994. The transformation seems to have uncorrelated the residuals for the most part.

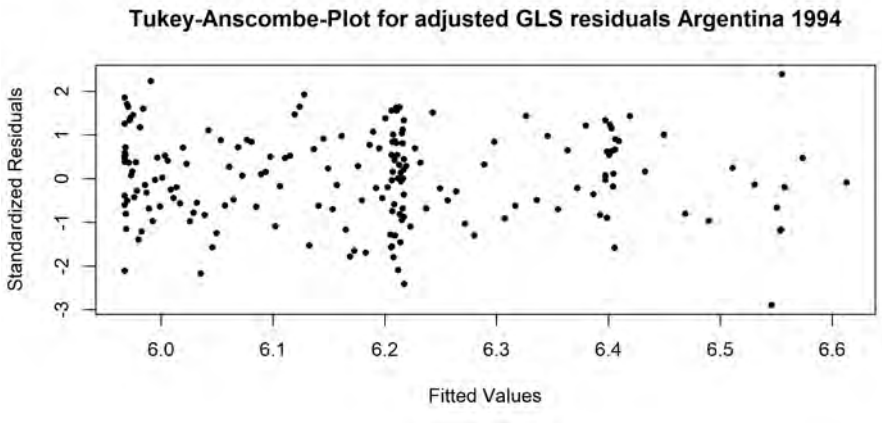

Figure 56: Tukey-Anscombe plot for the filtered residuals for the LPPLS GLSH(1,0) fit of the standardized residuals for the bubble in Argentina in 1994. There is hardly any structure in the residuals.

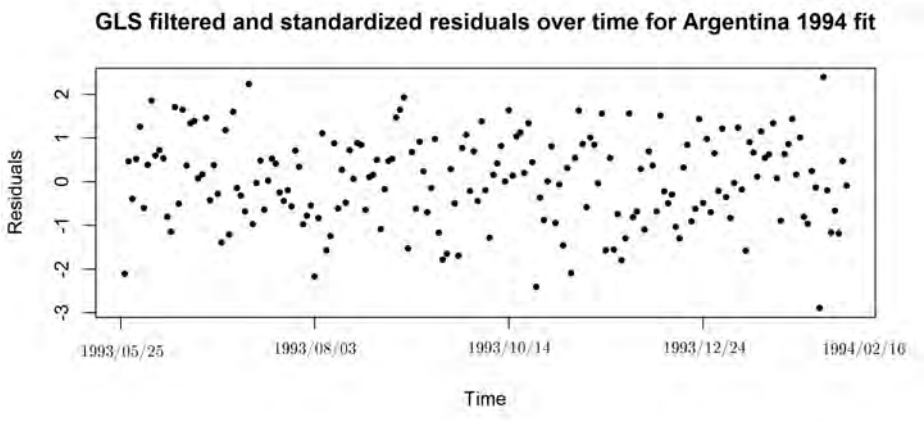

Figure 57: Standardized residuals over time plot for the LPPLS GLSH(1,0) fit for the bubble in Argentina in 1994. There is hardly any structure in the residuals.

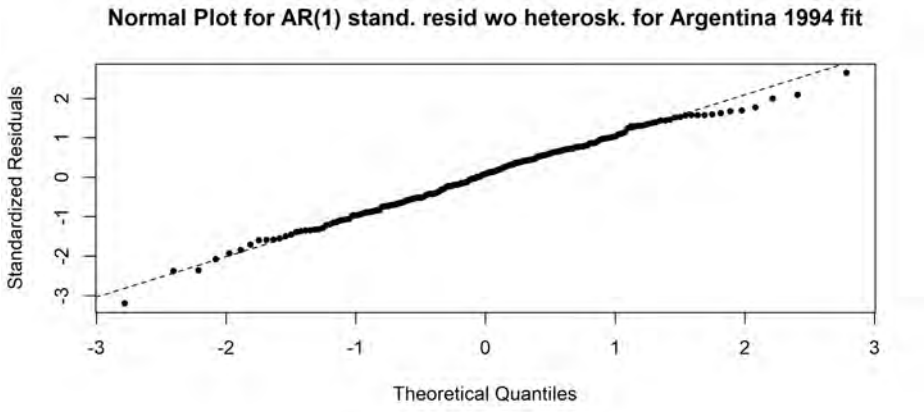

Figure 58: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit before accounting for heteroskedasticity for the bubble in Argentina in 1994. We cannot reject the assumption of normality.

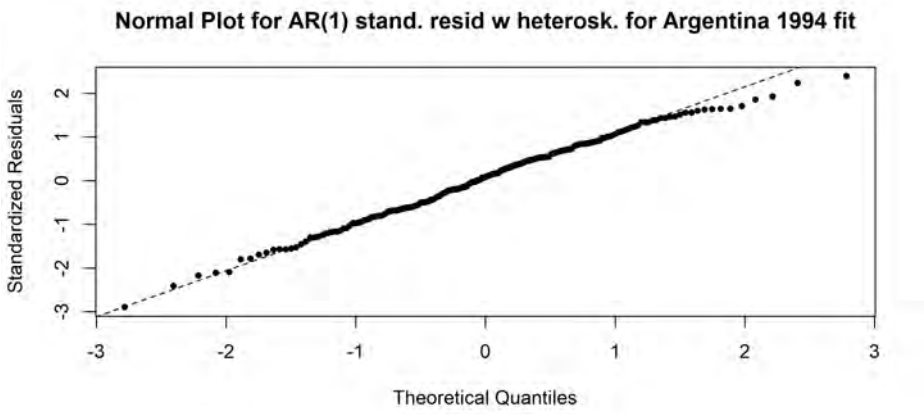

Figure 59: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit after accounting for heteroskedasticity for the bubble in Argentina in 1994. The fitted heteroskedasticity is estimated by using a loess smoother on the empirical residuals after filtering with $\Delta/60$ degrees of freedom, where Δ is the number of trading days in the fitting window. The normality assumption can hardly be rejected.

Figure 60: The fit of the full bubble in Brazil(IBOVESPA) in 1997 for the OLS LPPLS model in blue and the GLSH(1,0) LPPLS model in red. As usual, the GLSH(1,0) fit is a bit smoother, but the signals are otherwise similar. The fitted parameters for the LPPLS signals(t_2 = 1997/07/08 corresponds to the end of the fitting window, the true peak of the bubble) are $\hat{A}_{OLS} = 9.59104, \hat{B}_{OLS} = -0.0412, \hat{C}_{1,OLS} = 0.0035, \hat{C}_{2,OLS} = 0.0008, \hat{m}_{OLS} = 0.5714, \hat{\omega}_{OLS} = 5.7602, \hat{t_c}_{,OLS} = t_2 + 1$ and $\hat{A}_{GLS} = 9.6215, \hat{B}_{GLS} = -0.0531, \hat{C}_{1,GLS} = -0.0005, \hat{C}_{2,GLS} = 0.0039, \hat{m}_{GLS} = 0.5286, \hat{\omega}_{GLS} = 6.0357, \hat{t_c}_{,GLS} = t_2 + 2, \hat{\rho} = 0.8621$.

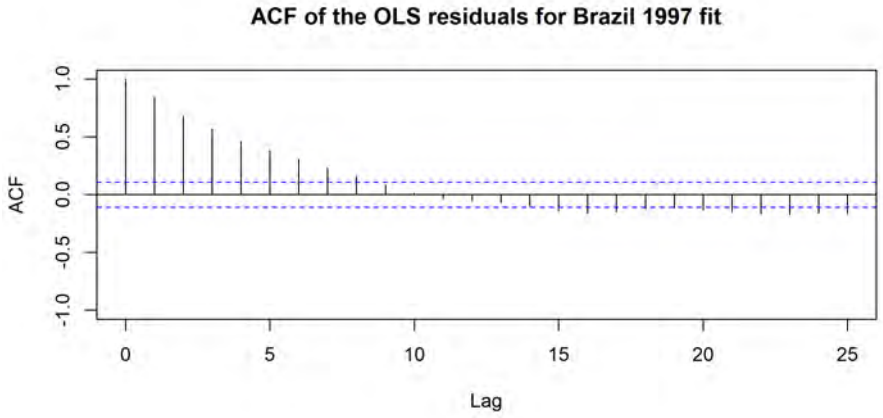

Figure 61: The autocorrelation function for the LPPLS OLS fit for the full bubble in Brazil(IBOVESPA) in 1997. Strong auto-regressive part is evident. The exponential decay indicates short memory residuals.

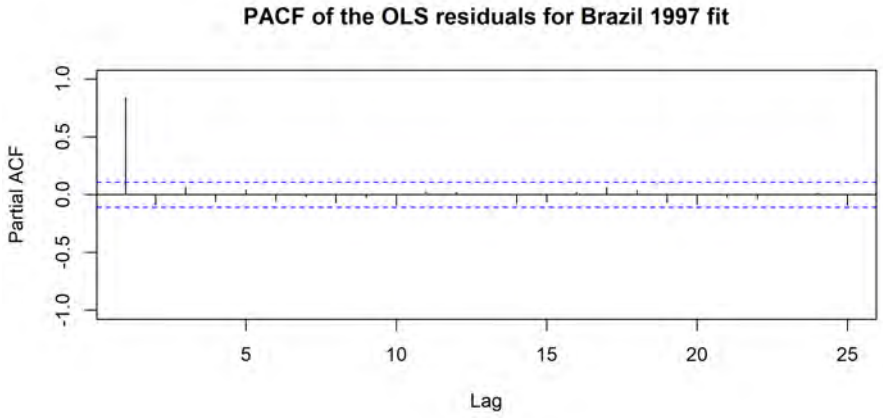

Figure 62: The partial autocorrelation function for the LPPLS OLS fit for the bubble in Brazil(IBOVESPA) in 1997. Strong auto-regressive part is evident of order 1.

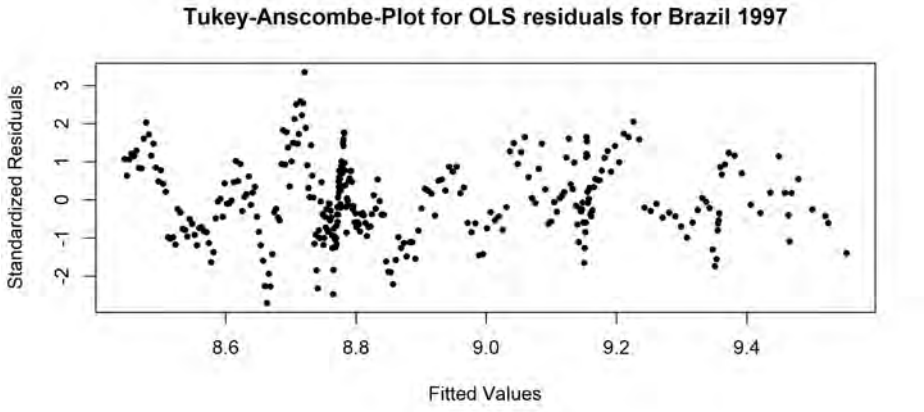

Figure 63: Tukey-Anscombe plot for the LPPLS OLS fit of the standardized residuals for the bubble in Brazil(IBOVESPA) in 1997. A structure is evident, and thus the model assumptions violated.

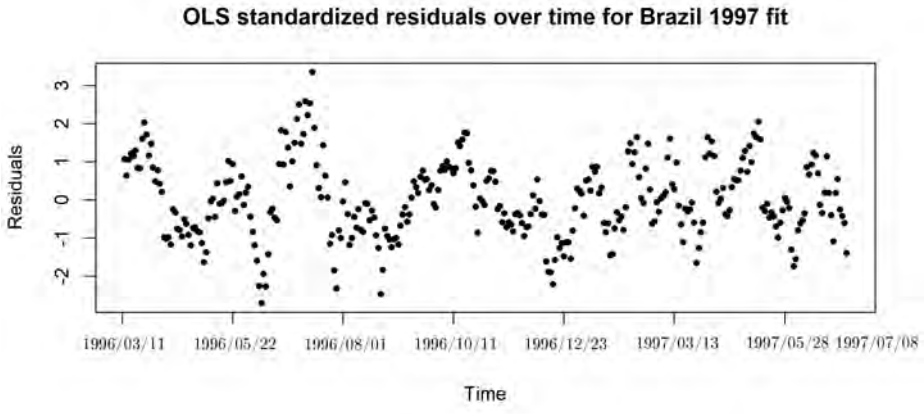

Figure 64: The plot shows standardized residuals over time for the LPPLS OLS fit for the bubble in Brazil(IBOVESPA) in 1997. A structure is evident, and thus the model assumptions violated.

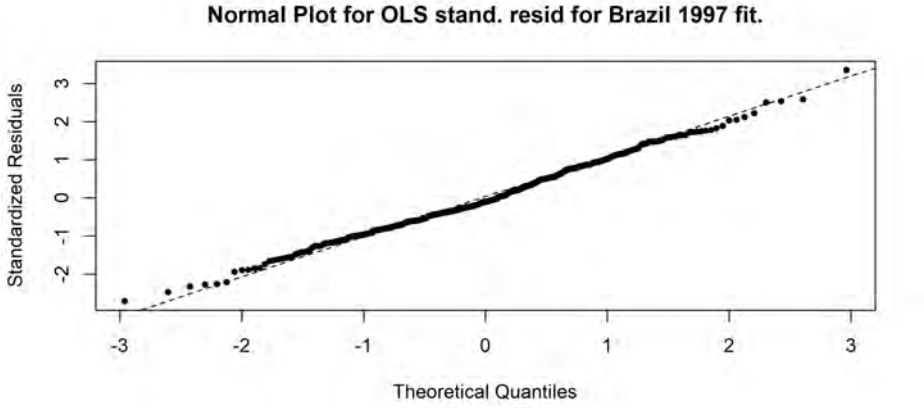

Figure 65: The quantile-quantile normal plot for the LPPLS OLS fit for the bubble in Brazil(IBOVESPA) in 1997. The normality is not violated. However, the strong auto-regressive structure needs to be corrected for before considering the Q-Q plot.

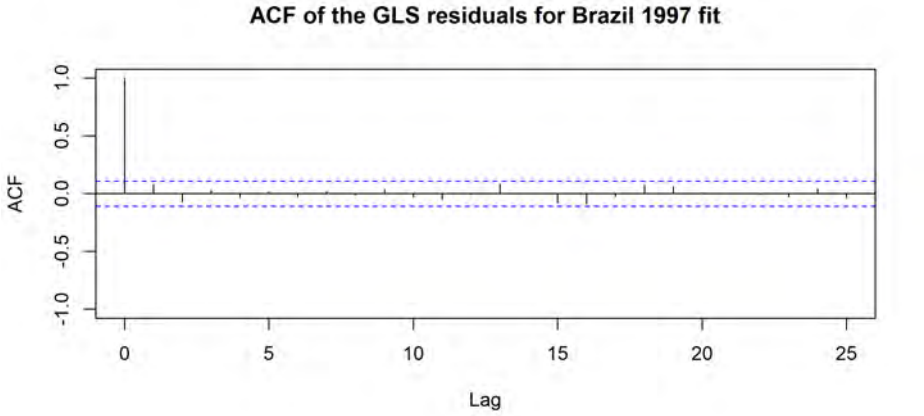

Figure 66: The autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Brazil(IBOVESPA) in 1997. The transformation seems to have uncorrelated the residuals for the most part.

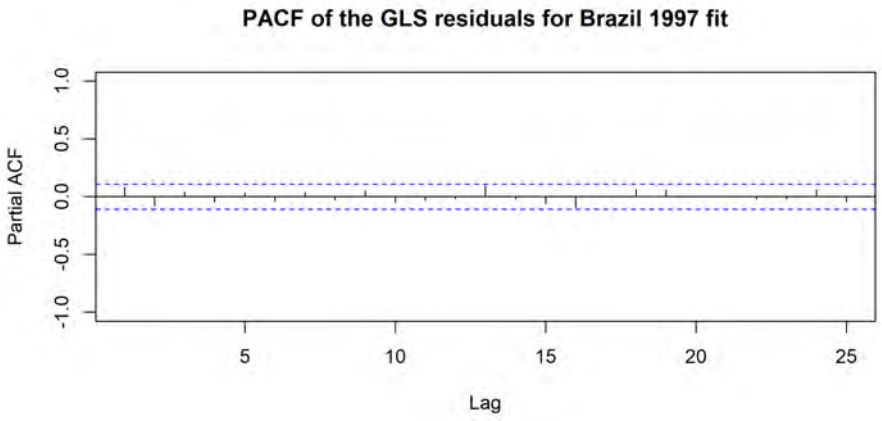

Figure 67: The partial autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Brazil(IBOVESPA) in 1997. The transformation seems to have uncorrelated the residuals for the most part.

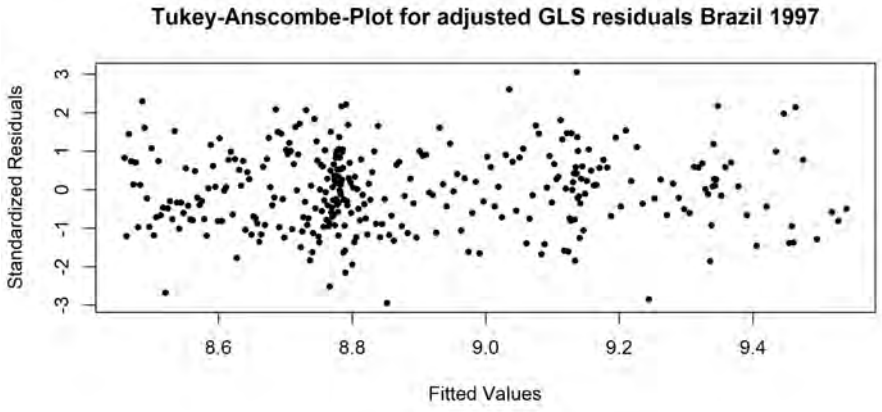

Figure 68: Tukey-Anscombe plot for the filtered residuals for the LPPLS GLSH(1,0) fit of the standardized residuals for the bubble in Brazil(IBOVESPA) in 1997. There is hardly any structure in the residuals.

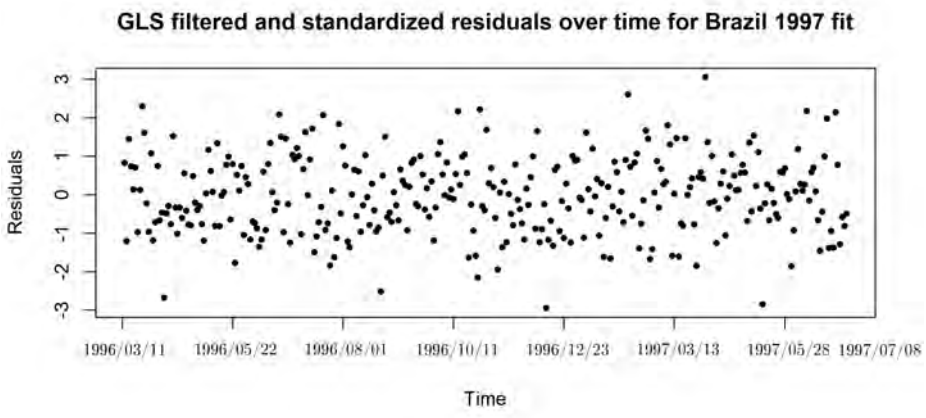

Figure 69: Standardized residuals over time plot for the LPPLS GLSH(1,0) fit for the bubble in Brazil in 1997. There is hardly any structure in the residuals.

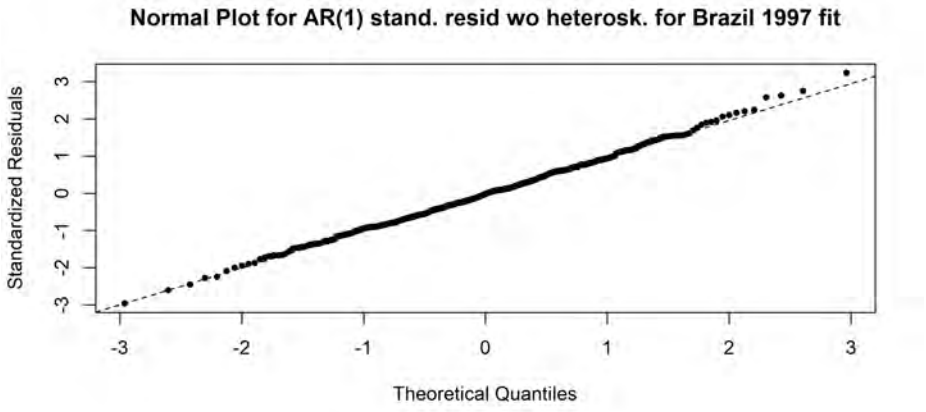

Figure 70: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit before accounting for heteroskedasticity for the bubble in Brazil in 1997. We cannot reject the assumption of normality.

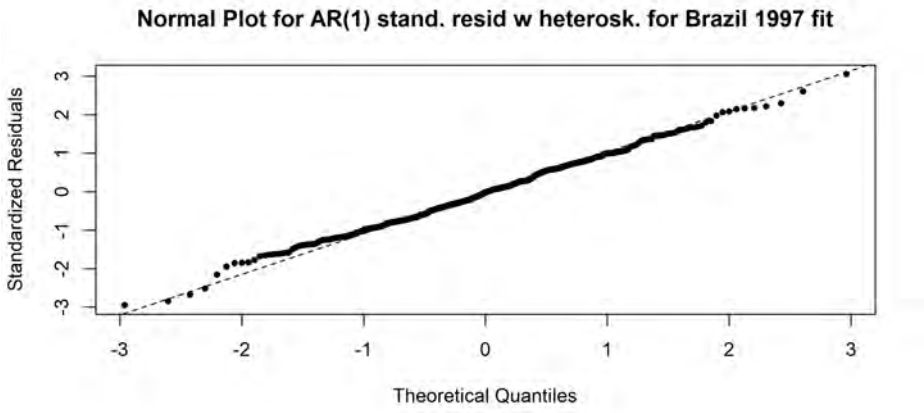

Figure 71: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit after accounting for heteroskedasticity for the bubble in Brazil in 1997. The fitted heteroskedasticity is estimated by using a loess smoother on the empirical residuals after filtering with $\Delta/60$ degrees of freedom, where Δ is the number of trading days in the fitting window. The normality assumption can hardly be rejected, but our fitted residual structure seems to make the situation a bit worse.

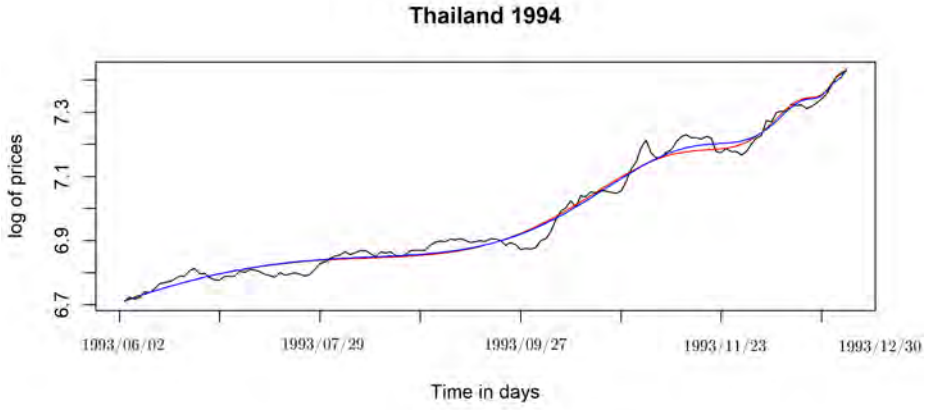

Figure 72: The fit of the full bubble in Thailand(Bangkok SET Index) in 1994 for the OLS LPPLS model in blue and the GLSH(1,0) LPPLS model in red. As usual, the GLSH(1,0) fit is a bit smoother, but the signals are otherwise similar. The fitted parameters for the LPPLS signals($t_2 = 1993/12/30$ corresponds to the end of the fitting window, the true peak of the bubble) are $\hat{A}_{OLS} = 7.4347, \hat{B}_{OLS} = -0.0240, \hat{C}_{1,OLS} = -0.0008, \hat{C}_{2,OLS} = -0.0029, \hat{m}_{OLS} = 0.7061, \hat{\omega}_{OLS} = 4.8112, \hat{t_c}_{,OLS} = t_2$ and $\hat{A}_{GLS} = 7.4975, \hat{B}_{GLS} = -0.0397, \hat{C}_{1,GLS} = 0.0033, \hat{C}_{2,GLS} = 0.0021, \hat{m}_{GLS} = 0.6143, \hat{\omega}_{GLS} = 5.3010, \hat{t_c}_{,GLS} = t_2 + 2, \hat{\rho} = 0.8921$.

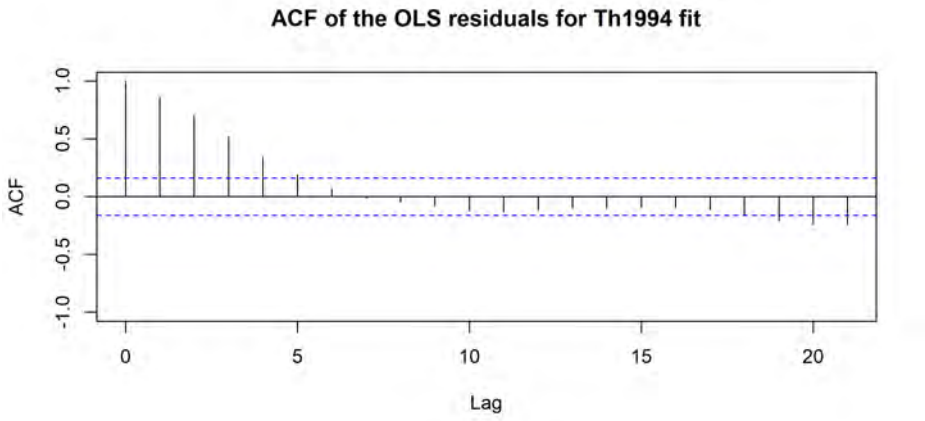

Figure 73: The autocorrelation function for the LPPLS OLS fit for the full bubble in Thailand(Bangkok SET Index) in 1994. Strong auto-regressive part is evident. Here it is not clear if the decay is exponential or linear, at least the estimated auto-regressive parameter is close to 1.

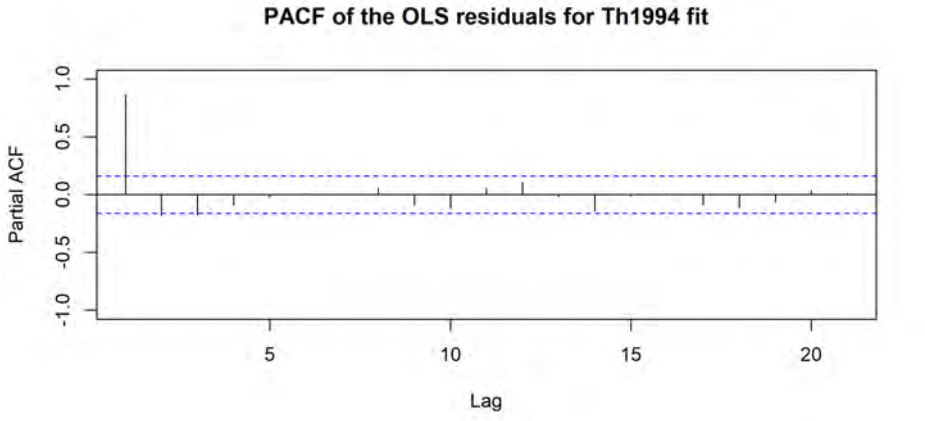

Figure 74: The partial autocorrelation function for the LPPLS OLS fit for the bubble in Thailand(Bangkok SET Index) in 1994. Strong auto-regressive part is evident of order 1.

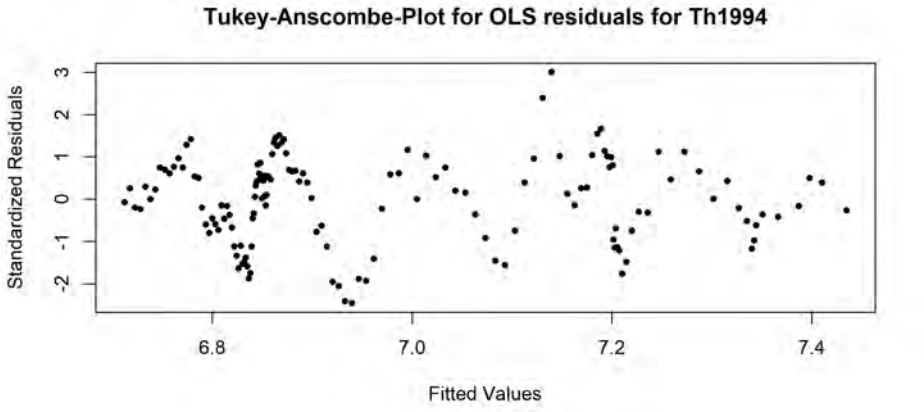

Figure 75: Tukey-Anscombe plot for the LPPLS OLS fit of the standardized residuals for the bubble in Thailand(Bangkok SET Index) in 1994. A structure is evident, and thus the model assumptions violated.

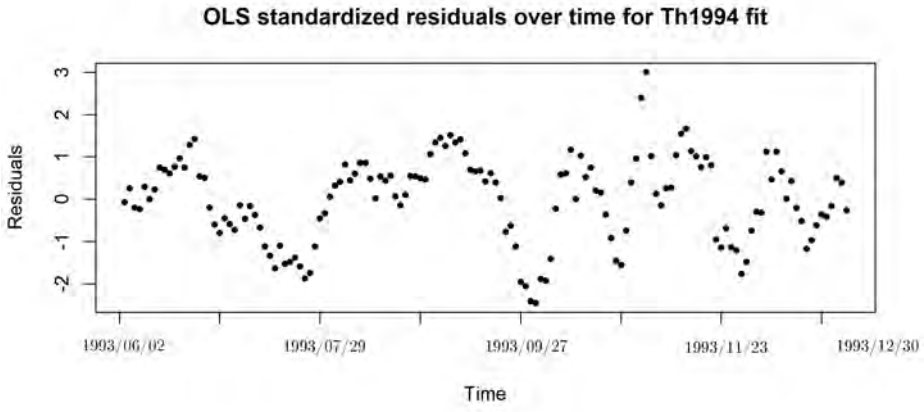

Figure 76: The plot shows standardized residuals over time for the LPPLS OLS fit for the bubble in Thailand(Bangkok SET Index) in 1994. A structure is evident, and thus the model assumptions violated.

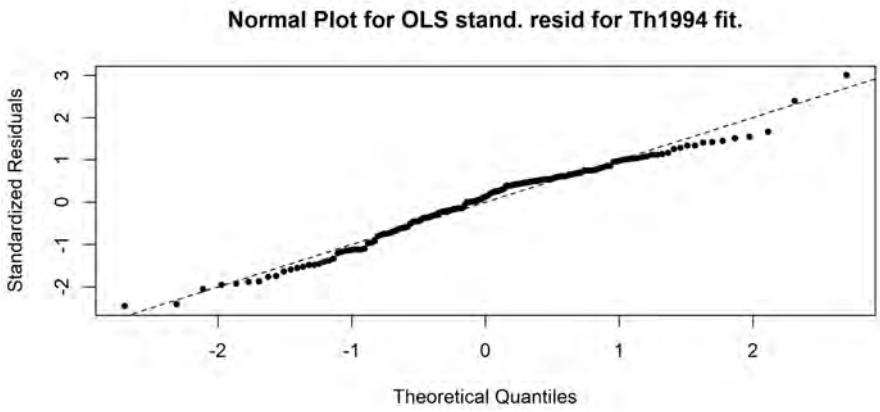

Figure 77: The quantile-quantile normal plot for the LPPLS OLS fit for the bubble in Thailand(Bangkok SET Index) in 1994. The normality is not violated. However, the strong auto-regressive structure needs to be corrected for before considering the Q-Q plot.

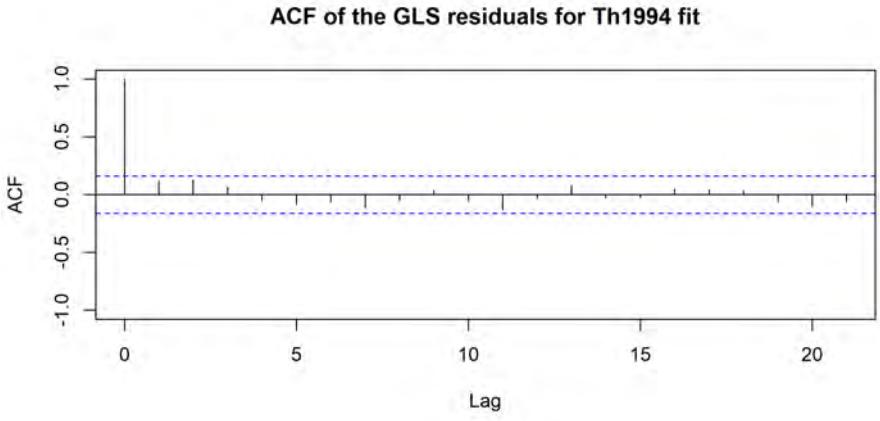

Figure 78: The autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Thailand(Bangkok SET Index) in 1994. The transformation seems to have uncorrelated the residuals for the most part.

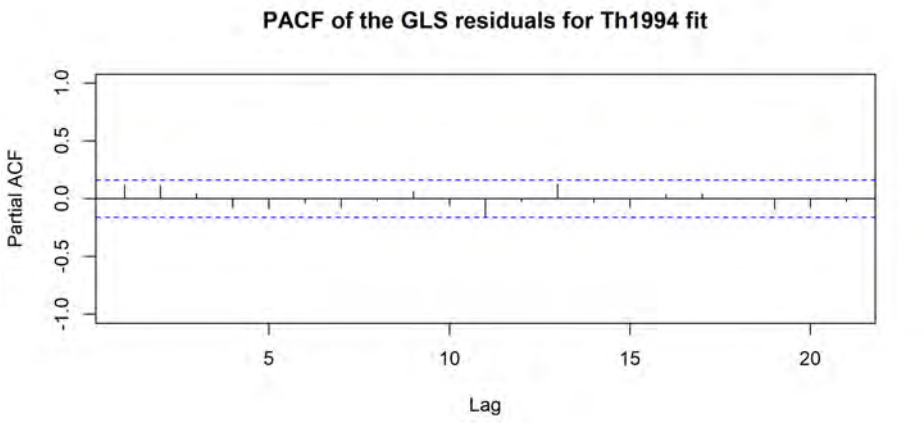

Figure 79: The partial autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Thailand(Bangkok SET Index) in 1994. The transformation seems to have uncorrelated the residuals.

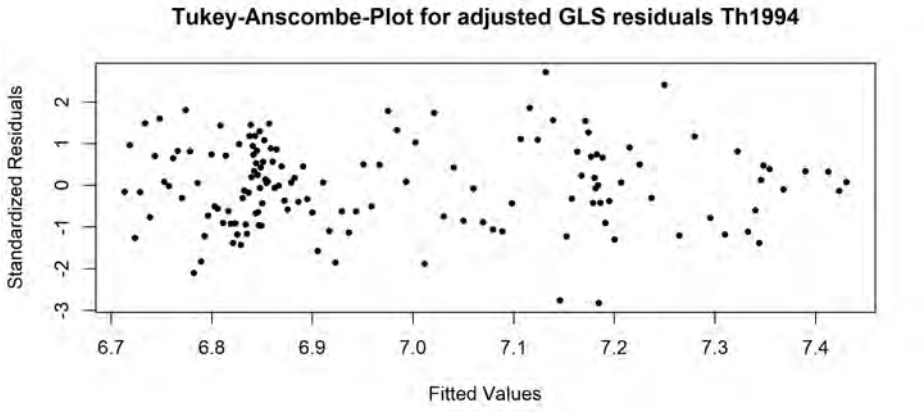

Figure 80: Tukey-Anscombe plot for the filtered residuals for the LPPLS GLSH(1,0) fit of the standardized residuals for the bubble in Thailand(Bangkok SET Index) in 1994. There is hardly any structure in the residuals.

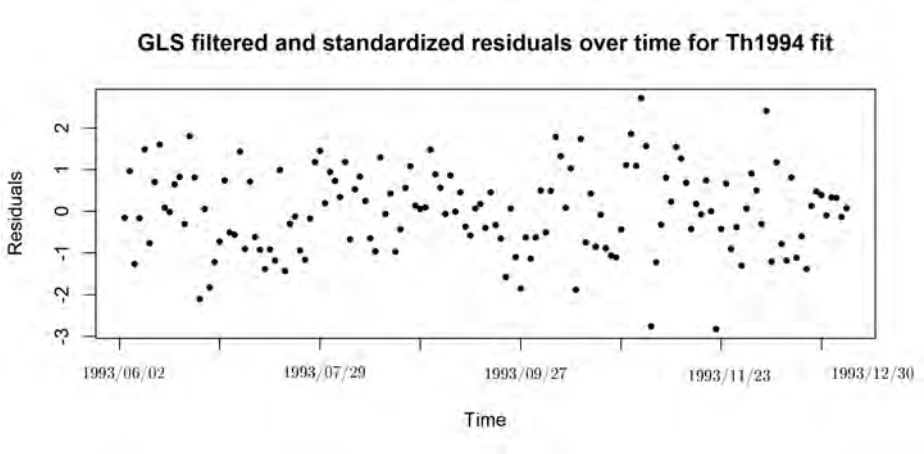

Figure 81: Standardized residuals over time plot for the LPPLS GLSH(1,0) fit for the bubble in Thailand(Bangkok SET Index) in 1994. There is hardly any structure in the residuals.

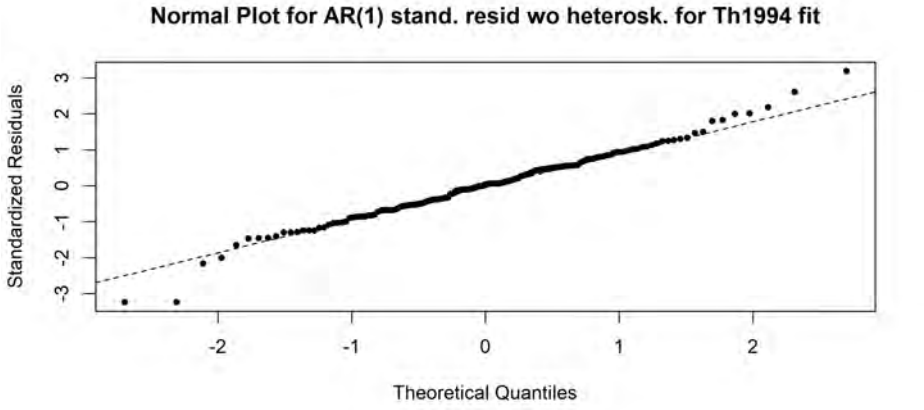

Figure 82: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit before accounting for heteroskedasticity for the bubble in Thailand(Bangkok SET Index) in 1994. The residuals seem slightly heavy tailed but the distribution is approximately symmetric.

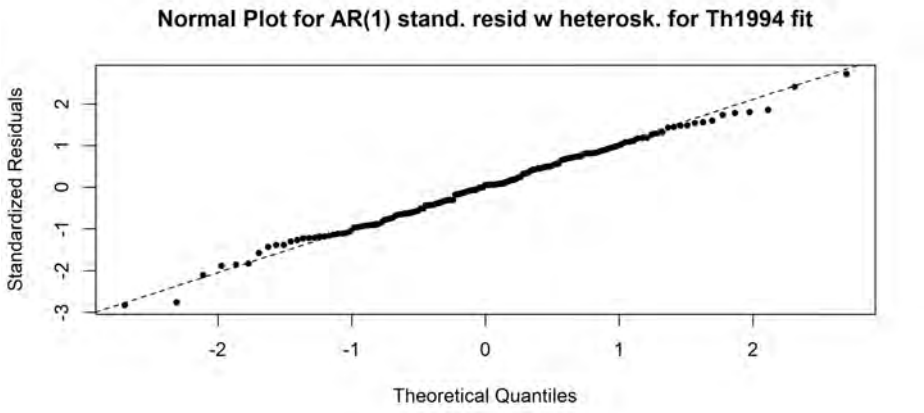

Figure 83: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit after accounting for heteroskedasticity for the bubble in Thailand(Bangkok SET Index) in 1994. The fitted heteroskedasticity is estimated by using a loess smoother on the empirical residuals after filtering with $\Delta/60$ degrees of freedom, where Δ is the number of trading days in the fitting window. The normality assumption can hardly be rejected and the heavy tails have been accounted for.

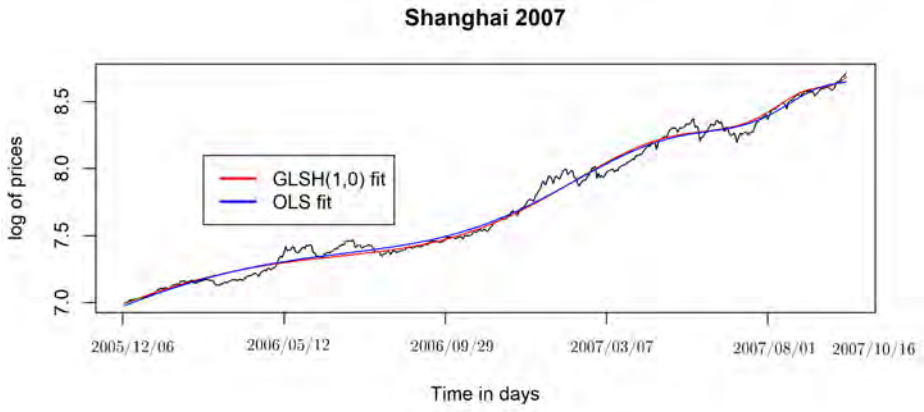

Figure 84: The fit of the full bubble in Shanghai(SSEC) in 2007 for the OLS LPPLS model in blue and the GLSH(1,0) LPPLS model in red. As usual, the GLSH(1,0) fit is a bit smoother, but the signals are otherwise similar. The fitted parameters for the LPPLS signals(t_2 = 2007/10/16 corresponds to the end of the fitting window, the true peak of the bubble) are $\hat{A}_{OLS} = 8.8750, \hat{B}_{OLS} = -0.0080, \hat{C}_{1,OLS} = 0.0005, \hat{C}_{2,OLS} = -0.0005, \hat{m}_{OLS} = 0.8959, \hat{\omega}_{OLS} = 6.0714, \hat{t_c}_{,OLS} = t_2 + 39$ and $\hat{A}_{GLS} = 8.7842, \hat{B}_{GLS} = -0.0072, \hat{C}_{1,GLS} = -0.0003, \hat{C}_{2,GLS} = 0.0007, \hat{m}_{GLS} = 0.9163, \hat{\omega}_{GLS} = 5.5000, \hat{t_c}_{,GLS} = t_2 + 18, \hat{\rho} = 0.9400$.

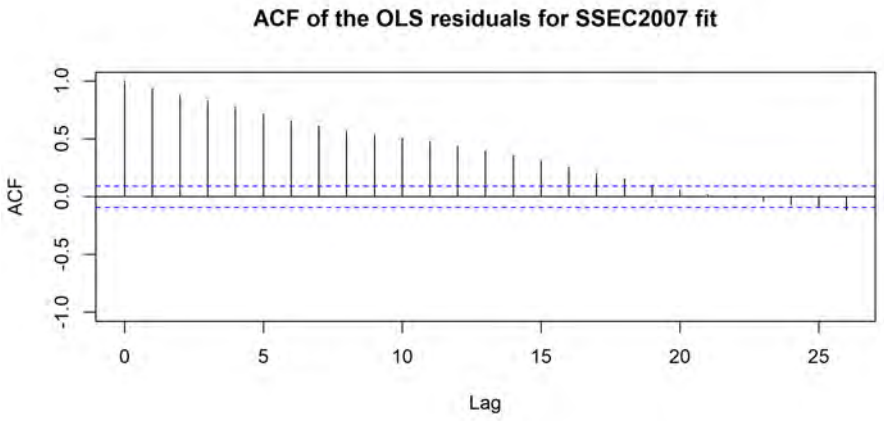

Figure 85: The autocorrelation function for the LPPLS OLS fit for the full bubble in Shanghai(SSEC) in 2007. Strong auto-regressive part is evident. Here it is not clear if the decay is exponential or linear, at least the estimated auto-regressive parameter is close to 1.

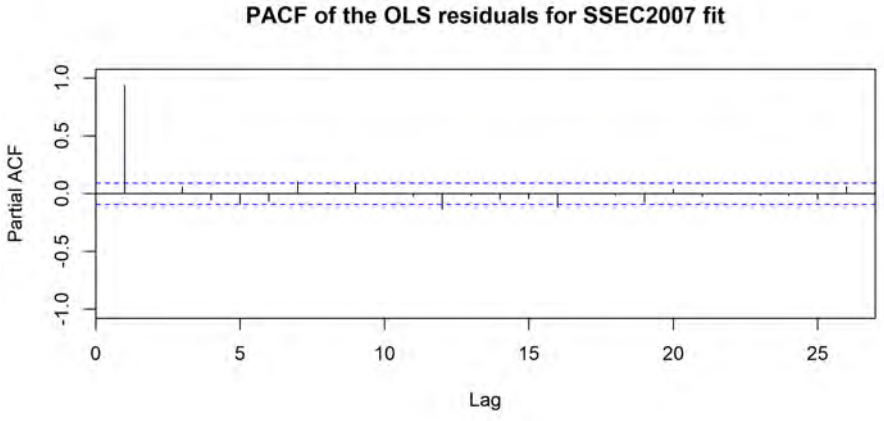

Figure 86: The partial autocorrelation function for the LPPLS OLS fit for the bubble in Shanghai(SSEC) in 2007. Strong auto-regressive part is evident of order 1.

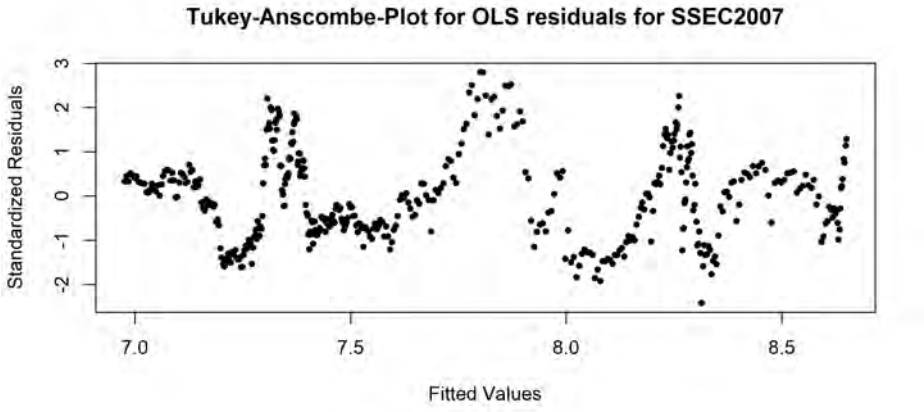

Figure 87: Tukey-Anscombe plot for the LPPLS OLS fit of the standardized residuals for the bubble in Shanghai(SSEC) in 2007. A structure is evident, and thus the model assumptions violated.

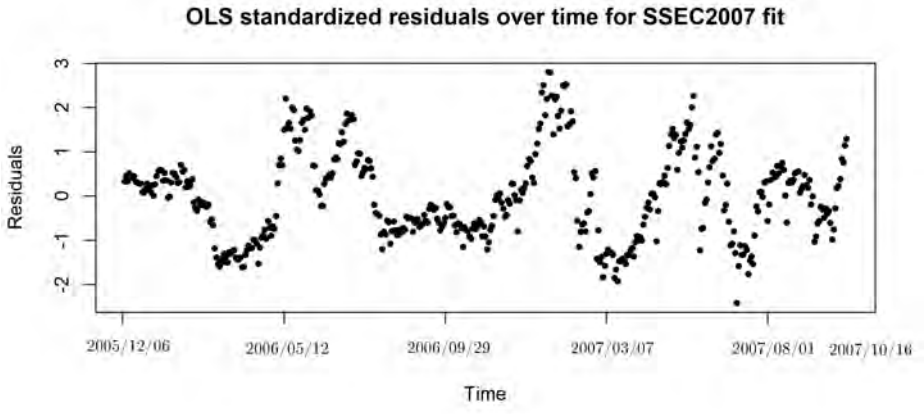

Figure 88: The plot shows standardized residuals over time for the LPPLS OLS fit for the bubble in Shanghai(SSEC) in 2007. A structure is evident, and thus the model assumptions violated.

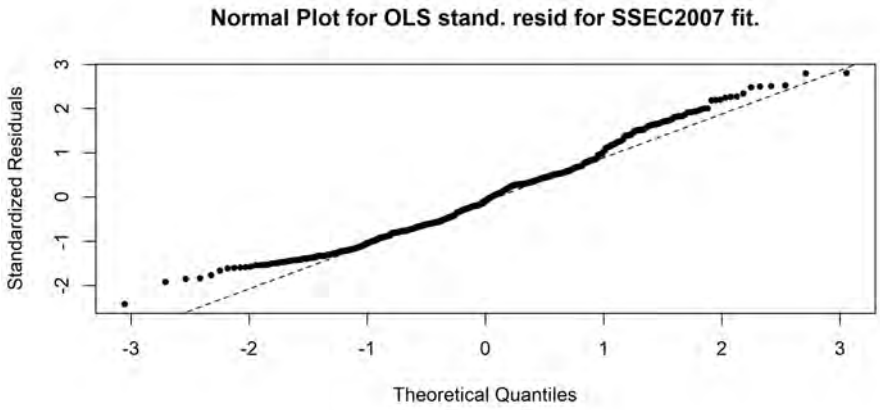

Figure 89: The quantile-quantile normal plot for the LPPLS OLS fit for the bubble in Shanghai(SSEC) in 2007. The normality is assumption is violated. However, the strong auto-regressive structure needs to be corrected for before considering the Q-Q plot.

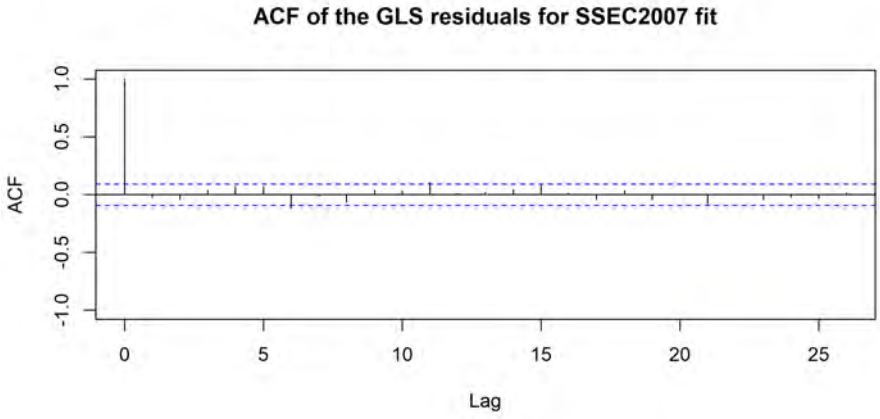

Figure 90: The autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Shanghai(SSEC) in 2007. The transformation seems to have uncorrelated the residuals for the most part.

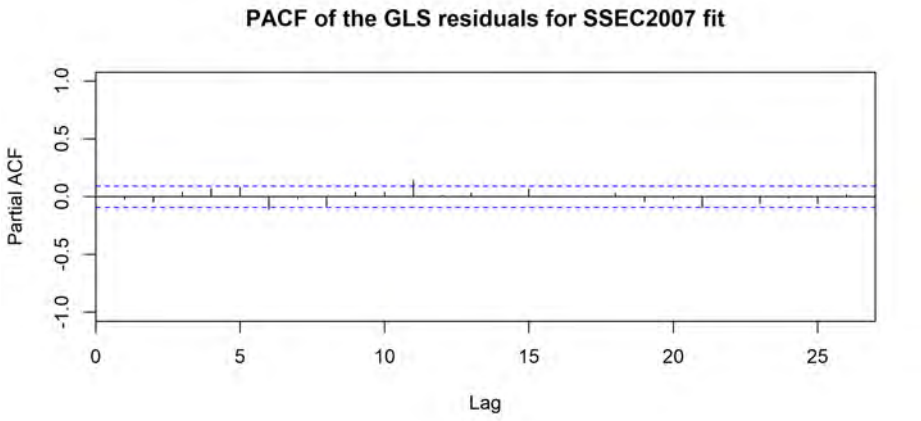

Figure 91: The partial autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Shanghai(SSEC) in 2007. The transformation seems to have uncorrelated the residuals for the most part.

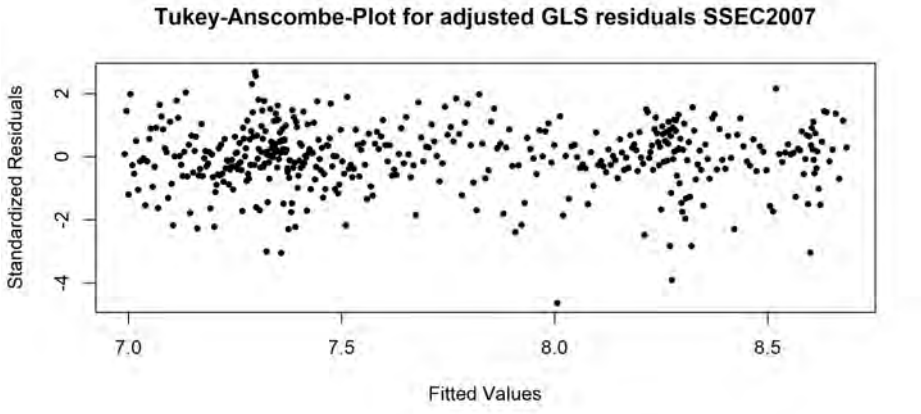

Figure 92: Tukey-Anscombe plot for the filtered residuals for the LPPLS GLSH(1,0) fit of the standardized residuals for the bubble in Shanghai(SSEC) in 2007. There is hardly any structure in the residuals.

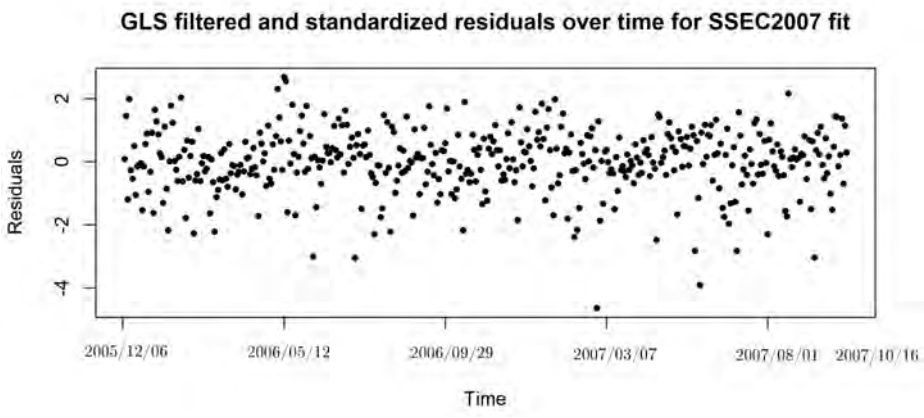

Figure 93: Standardized residuals over time plot for the LPPLS GLSH(1,0) fit for the bubble in Shanghai(SSEC) in 2007. There is hardly any structure in the residuals.

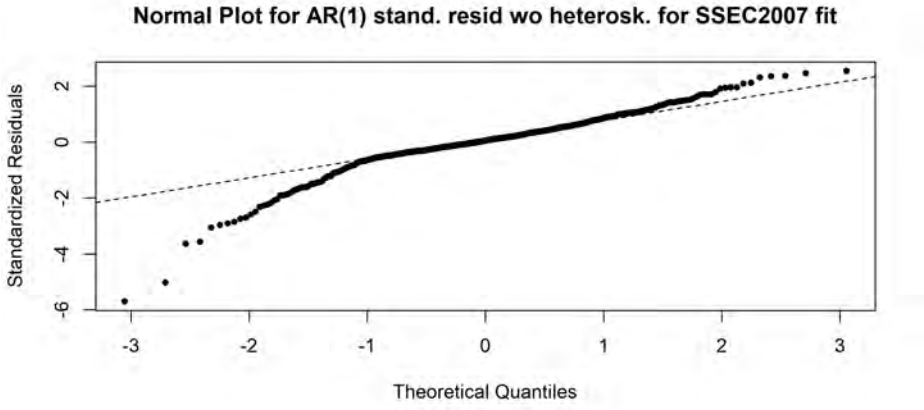

Figure 94: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit before accounting for heteroskedasticity for the bubble in Shanghai(SSEC) in 2007. The assumption of normality is violated as there are heavy-tailed negative residuals.

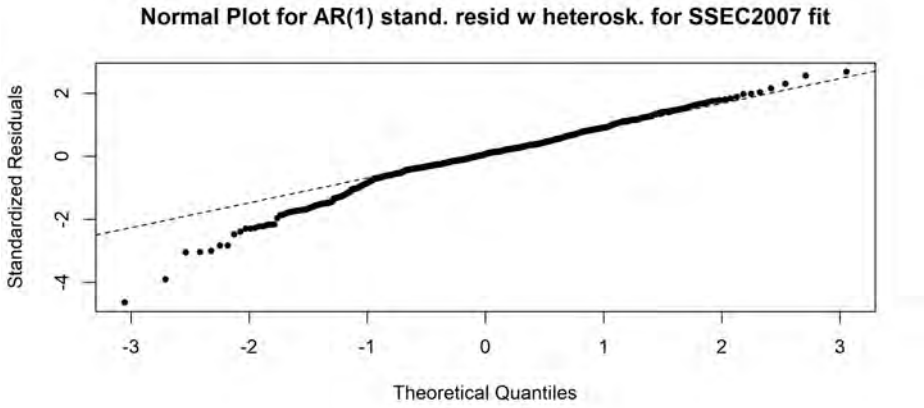

Figure 95: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit after accounting for heteroskedasticity for the bubble in Shanghai(SSEC) in 2007. The fitted heteroskedasticity is estimated by using a loess smoother on the empirical residuals after filtering with $\Delta/60$ degrees of freedom, where Δ is the number of trading days in the fitting window. The normality assumption is not satisfied but the fitted heteroskedasticity improves the situation.

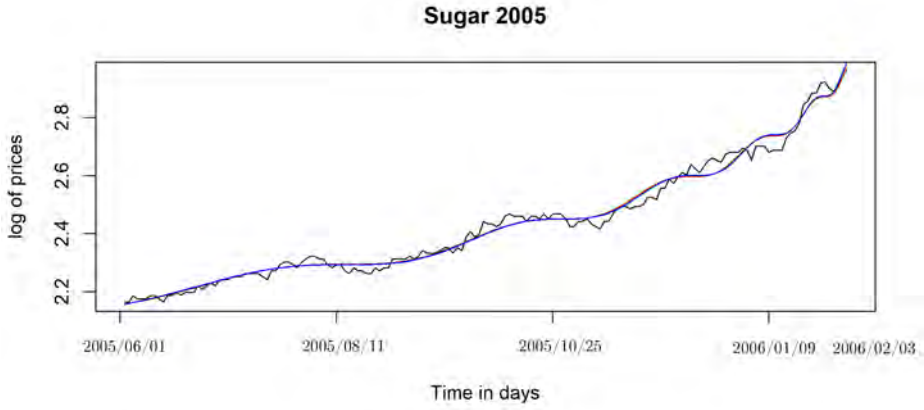

Figure 96: The fit of the full bubble in Sugar 2005 for the OLS LPPLS model in blue and the GLSH(1,0) LPPLS model in red. The fitted signals are very similar. The fitted parameters for the LPPLS signals($t_2 = 2006/02/03$ corresponds to the end of the fitting window, the true peak of the bubble) are $\hat{A}_{OLS} = 5.0917, \hat{B}_{OLS} = -1.5927, \hat{C}_{1,OLS} = 0.0135, \hat{C}_{2,OLS} = -0.0067, \hat{m}_{OLS} = 0.1163, \hat{\omega}_{OLS} = 12.5357, \hat{t_c}_{,OLS} = t_2 + 12$ and $\hat{A}_{GLS} = 5.6042, \hat{B}_{GLS} = -2.0351, \hat{C}_{1,GLS} = -0.0130, \hat{C}_{2,GLS} = 0.0085, \hat{m}_{GLS} = 0.1000, \hat{\omega}_{GLS} = 13.1225, \hat{t_c}_{,GLS} = t_2 + 14, \hat{\rho} = 0.7562$.

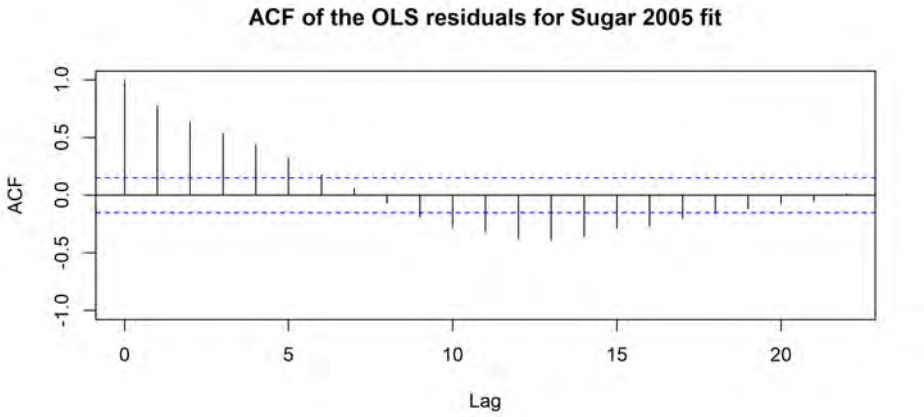

Figure 97: The autocorrelation function for the LPPLS OLS fit for the full bubble in Sugar 2005. Strong auto-regressive part is evident. Here it is not clear if the decay is exponential or linear, at least the estimated auto-regressive parameter is close to 1.

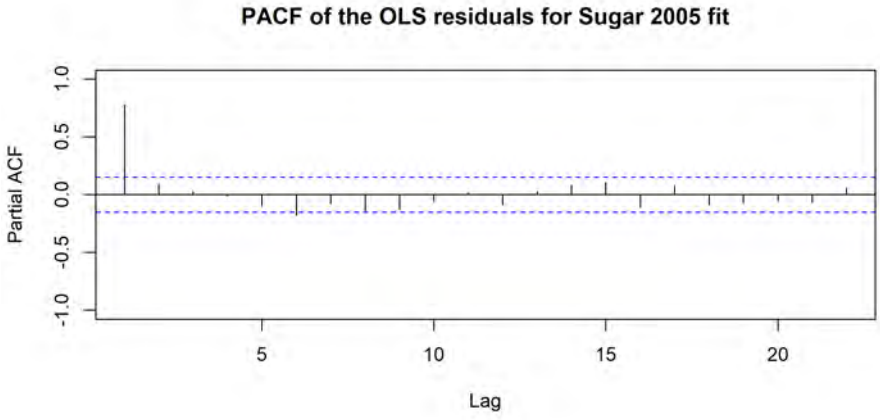

Figure 98: The partial autocorrelation function for the LPPLS OLS fit for the bubble in Sugar 2005. Strong auto-regressive part is evident of order 1.

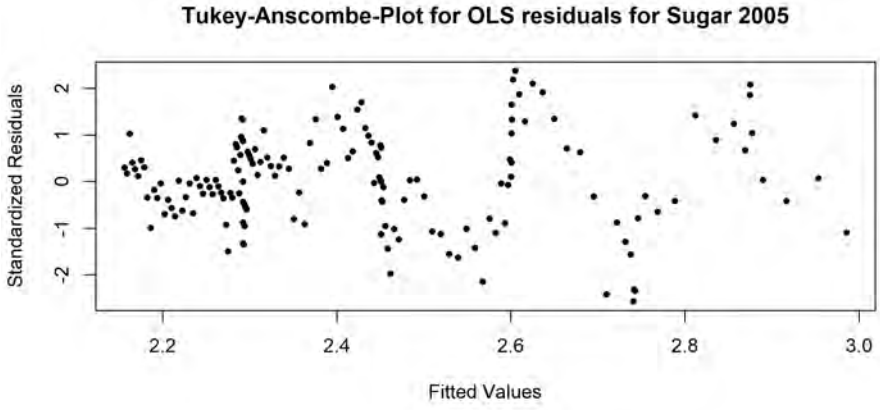

Figure 99: Tukey-Anscombe plot for the LPPLS OLS fit of the standardized residuals for the bubble in Sugar 2005. A structure is evident, and thus the model assumptions violated.

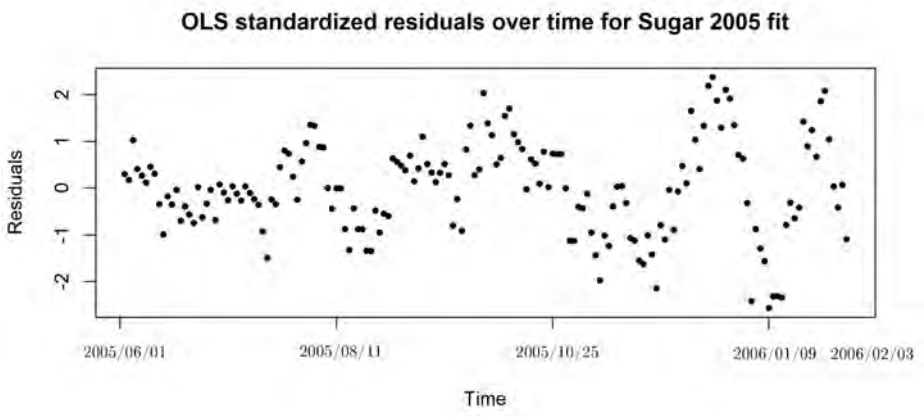

Figure 100: The plot shows standardized residuals over time for the LPPLS OLS fit for the bubble in Sugar 2005. A structure is evident, and thus the model assumptions violated.

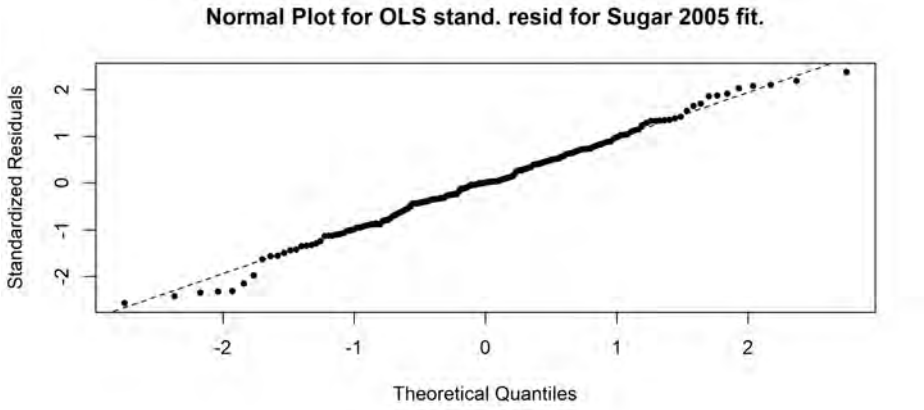

Figure 101: The quantile-quantile normal plot for the LPPLS OLS fit for the bubble in Sugar 2005. The normality is not clearly violated. However, the strong auto-regressive structure needs to be corrected for before considering the Q-Q plot.

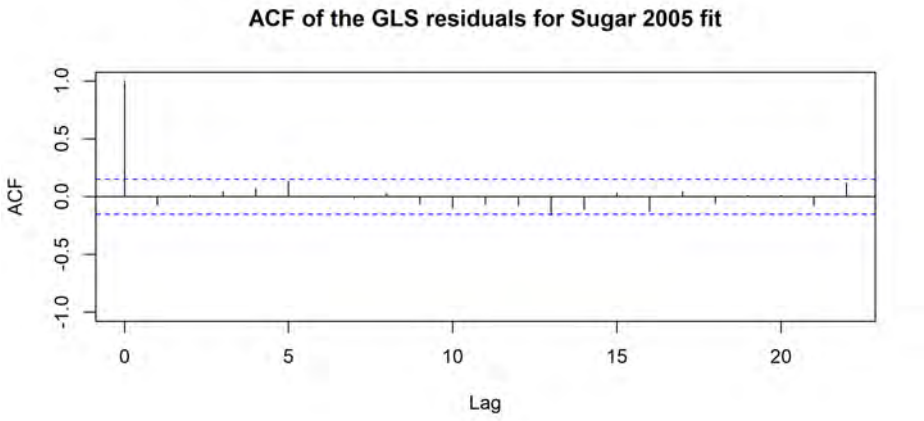

Figure 102: The autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Sugar 2005. The transformation seems to have uncorrelated the residuals.

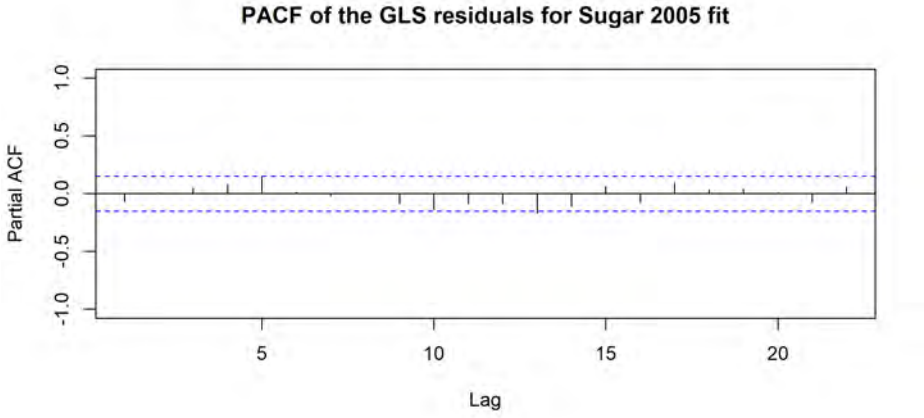

Figure 103: The partial autocorrelation function of the filtered residuals for the LPPLS GLSH(1,0) fit for the bubble in Sugar 2005. The transformation seems to have uncorrelated the residuals.

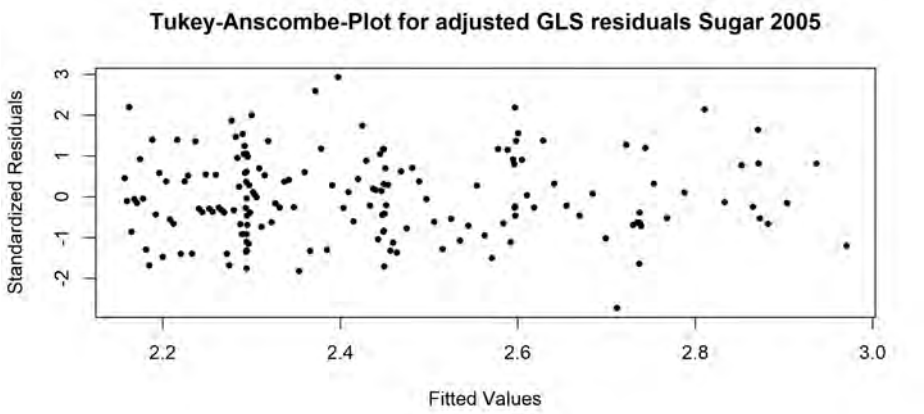

Figure 104: Tukey-Anscombe plot for the filtered residuals for the LPPLS GLSH(1,0) fit of the standardized residuals for the bubble in Sugar 2005. There is hardly any structure in the residuals.

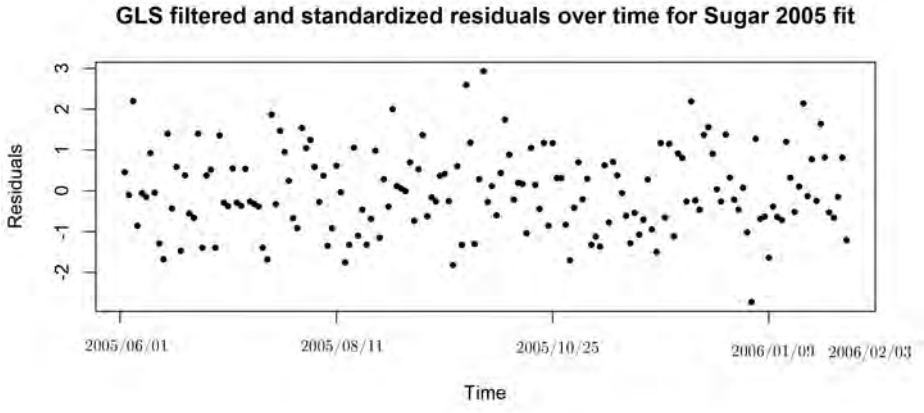

Figure 105: Standardized residuals over time plot for the LPPLS GLSH(1,0) fit for the bubble in Sugar 2005. There is hardly any structure in the residuals.

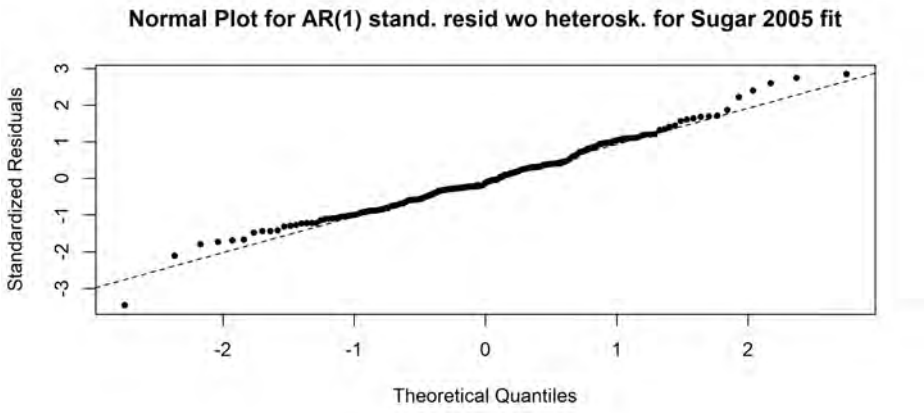

Figure 106: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit before accounting for heteroskedasticity for the bubble in Sugar 2005. We cannot reject the assumption of normality but the tails are a bit off.

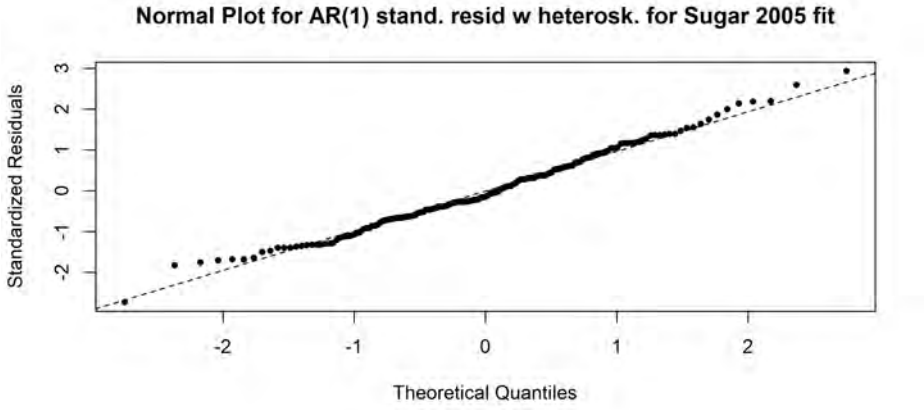

Figure 107: The quantile-quantile normal plot for the LPPLS GLSH(1,0) fit after accounting for heteroskedasticity for the bubble in Sugar 2005. The fitted heteroskedasticity is estimated by using a loess smoother on the empirical residuals after filtering with $\Delta/60$ degrees of freedom, where Δ is the number of trading days in the fitting window. The normality assumption can hardly be rejected and the fitted heteroskedasticity improves the situation.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Leave-out methods for selecting the optimal starting point of financial bubbles

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Magnusson	Sigurdur Heidar

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zurich, 28-08-2018	Sigurdur Heidar Magnusson

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.