A Time-Varying Parameter Model for Local Explosions

Francisco (F.) Blasques[1]
Siem Jan (S.J.) Koopman[1]
Marc Nientker[1]

[1] VU Amsterdam

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

A Time-Varying Parameter Model for Local Explosions

Francisco Blasques[a,b], Siem Jan Koopman[a,b,c] and Marc Nientker[a]

[a]VU Amsterdam, School of Business and Economics, The Netherlands

[b]Tinbergen Institute Amsterdam, The Netherlands

[c]CREATES, Aarhus University, Denmark

November 9, 2018

Abstract

Locally explosive behavior is observed in many economic and financial time series when bubbles are formed. We introduce a time-varying parameter model that is capable of describing this behavior in time series data. Our proposed model can be used to predict the emergence, existence and burst of bubbles. We adopt a flexible observation driven model specification that allows for different bubble shapes and behavior. We establish stationarity, ergodicity, and bounded moments of the data generated by our model. Furthermore, we obtain the consistency and asymptotic normality of the maximum likelihood estimator. Given the parameter estimates, our filter is capable of extracting the unobserved bubble process from observed data. We study finite-sample properties of our estimator through a Monte Carlo simulation study. Finally, we show that our model compares well with noncausal models in a financial application concerning the Bitcoin/US dollar exchange rate.

Keywords: bubbles, observation driven models, noncausal models, stationary, ergodic, consistency, asymptotic normality, exchange rates.
JEL codes: C22, C58, G10.

1   Introduction

Many financial and economic time series display phases of locally explosive behaviour that is followed by a burst or sharp mean-reverting dynamics. This stochastic behaviour is especially prevalent in financial asset prices, stock indices and exchange rates. The literature on rational expectations models for asset pricing typically describe the asset price process as the sum of a fundamental value process and aforementioned locally explosive process. The second process is then defined as a *speculative bubble*, see for instance Blanchard and Watson (1982), West (1987), Diba and Grossman (1988) and more. The bubble is considered to be an explosive nonstationary process and its presence is tested via unit root and cointegration tests. However, Evans (1991) noted that periodically collapsing bubbles can cause the bubble paths to look more like a stationary process, making it difficult for regular tests to detect the existence of bubbles. Using recursive testing techniques, evidence for the existence of a speculative bubble has been found by for example Phillips et al. (2011) and Homm and Breitung (2012) in the Nasdaq real price index and Phillips and Yu (2011) in the U.S. house price index, the price of crude oil and the spread between Baa and Aaa bond rates.

A different approach has been proposed by Gouriéroux and Zakoïan (2013) who describe speculative bubble dynamics using a noncausal autoregressive process of order one with Cauchy innovations. This specification is able to model speculative bubbles as in reverse time the model is a causal autoregressive process of order one with a fat tailed innovation distribution, and thus produces large spikes followed by mean reversion. From the calendar time perspective such dynamics are observed as exponential explosions followed by sudden collapse. The noncausal approach to bubble modelling has been extended to stable distributed innovations in Gouriéroux and Zakoïan (2017) and to higher order mixed causal and noncausal linear models in Fries and Zakoïan (2017). A difference with the rational expectations approach is that noncausal models work within a stationary framework, which allows for the derivation of many theoretical results. Gouriéroux and Zakoïan (2013) show that the sample autocorrelation converges to a number smaller than one in absolute value. It demonstrates that a unit root test generally rejects the unit root hypothesis and thus will be unable to identify the presence of speculative bubbles. Moreover, they indicate the possibility of calculating and predicting future bubble behaviour and show the existence of

moments. A major disadvantage of the noncausal approach is its computational challenge. Distinguishing causal and noncausal components is based on extreme value clustering, see the discussions in Fries and Zakoïan (2017). The prediction of these components depends on computational methods such as Metropolis-Hasting or sampling/importance resampling, see Gourieroux and Jasiak (2016). In addition, the models are unable to distinguish the potential speculative bubble from the fundamental value. The noncausal models allow for only one generic type of bubble baseline path for a given set of parameters.

We introduce an observation driven model with time varying parameters designed as a new approach to modelling multiple speculative bubbles. As in the literature on rational expectations, our proposed model splits the asset price into a sum of two processes. The first process represents the fundamental value and can be modelled by any contracting or mean reverting process, while the second process represents the bubble effect characterised by the typical exponential increase followed by a burst. We provide various bubble burst conditions and discuss their respective merits and shortcomings. The advantage of using such a specification is that we can filter data into its fundamental value and a potential speculative bubble. Furthermore, the sum of the two processes is very flexible due to the joint dynamics of the individual components and can describe various baseline paths for the same set of parameters. Finally, the model has a conventional observation driven specification, which implies that parameter estimation can rely on the method of maximum likelihood where the likelihood function is obtained via the prediction error decomposition. It further implies that point predictions, confidence intervals, bubble burst probabilities, bubble emergence probabilities, expected bubble life times, and more, can be derived straightforwardly.

Similar to the noncausal literature our model describes locally explosive behaviour in a strictly stationary framework. Due to earlier work in Blasques and Nientker (2017) we can immediately show that the model admits a stationary ergodic and ϕ-mixing solution under very mild conditions. Additionally, in this paper, we prove that the model as a filter also admits a stationary ergodic and mixing solution and that any initiated sample path converges to this solution. The derivations of these results are nonstandard because the filter contains a discontinuity, rendering classical contraction results such as those in Bougerol (1993) infeasible. The results are then used to obtain consistency and asymptotic normality for our maximum likelihood estimator on the parameters that enter continuously

in the likelihood. In a simulation exercise we show that other parameters are well behaved.

The remainder of the paper is structured as follows. Section 2.1 introduces our modelling framework for local explosions. In Section 3 we study probabilistic and statistical properties of the model. Evidence from simulations and a real time series are provided in Section 4. Concluding remarks are in Section 5. The proofs are presented in the Appendix.

2 Model for Local Explosions

Our model decomposes the asset price X_t into a sum of three elements

$$(2.1) \qquad\qquad X_t = \mu_t + b_t + \varepsilon_t,$$

where μ_t is the fundamental value of the asset price, b_t is the value of a potential speculative bubble and ε_t is an error term. The error ε_t an element of an independent and identically normal distributed sequence

$$(2.2) \qquad\qquad (\varepsilon_t)_{t \in \mathbb{Z}} \sim \mathrm{NID}(0, \sigma^2),$$

where σ is a strictly positive constant. The fundamental value μ_t is defined as the value of the asset price if no speculative bubbles were to exist. The main focus of this paper is on describing bubble dynamics. Hence we consider a basic observation driven updating equation for the fundamental value, that is

$$(2.3) \qquad\qquad \mu_t = \delta + \beta\mu_{t-1} + \gamma(X_{t-1} - \mu_{t-1} - b_{t-1}),$$

where δ, β and γ are fixed unknown parameters. The dynamics for the fundamental value are mean reverting if $|\beta| < 1$, but partially correct by a factor γ for the past error $\varepsilon_{t-1} = (X_{t-1} - \mu_{t-1} - b_{t-1})$. This updating equation can be interpreted as an observation driven analogue of the parameter driven local level model in Chapter 2 of Durbin and Koopman (2012) and can also be obtained when using a score updating rule for the mean as in Creal et al. (2013). Many other dynamic processes for the fundamental value can also be considered. The model specification (2.3) can be augmented with more lags of the μ and ε processes,

similar to a stationary autoregressive moving average (ARMA) process. Also, we can adopt a completely exogenous stationary process for μ_t which is potentially based on economic or financial reasoning. We will maintain the stationary framework explored in the non-causal literature. In practice, this means that one would have to add a nonstationary component when the objective is to model bubles in non-stationary time-series, such as asset prices which are typically nonstationary. Alternatively, one could allow the fundamental value to be non-stationary, such as a random walk. This is however outside the scope of this paper.

The speculative bubble process is nonnegative and defined according to the following updating equation

$$(2.4) \qquad\qquad b_t = (\omega + \alpha b_{t-1})\mathbb{1}\{\text{survival condition}\}.$$

To ensure nonnegativity of b_t we impose $\omega > 0$, while α can be any nonnegative number, but typically is thought of as a parameter that is greater than one. This implies that the bubble process satisfies an exponential increase, as is commonly observed in locally explosive time series. The bubble b_t then diverges to infinity, if not for the indicator function, which forces the bubble to collapse down to zero if the survival condition is no longer satisfied. As with the fundamental value process, many options are available for the survival condition. Let $F_t = (X_t, \mu_t, b_t)$ be the information obtained at time t. Then a general survival condition that encompasses a variety of useful model choices is given by thresholded functions

$$(2.5) \qquad\qquad \mathbb{1}\{g(F_{t-1}) < 0\},$$

where g is some real-valued function, which we will call the *survival function*. A few example choices for the survival function are given by

E1 $g(F_{t-1}) = X_{t-1} - c,$ for some $c \in \mathbb{R}$.

E2 $g(F_{t-1}) = X_{t-1} - \mu_{t-1} - c,$ for some $c \geq 0$.

E3 $g(F_{t-1}) = b_{t-1} - kX_{t-1},$ for some $k \in [0, 1]$.

E4 $g(F_{t-1}) = b_{t-1} - k(\mu_t - c),$ for some $k \geq 0$ and $c \in \mathbb{R}$.

The simplest survival function **E1** lets the bubble grow until the asset price reaches a fixed level c. This allows for various bubble sizes b_t, as X_{t-1} also depends on the fundamental value μ_{t-1} and the shock process ε_{t-1}, but does describe time series in which the asset price always drops from approximately the same critical level. To allow for varying levels one can use a survival function such as **E2**. This function allows the bubble to grow as long as the difference between the asset price and the fundamental value is not too large, which leaves flexibility for the actual critical level. Examples **E1**-**E2** have less control for the emergence rate of bubbles. If a bubble just collapsed, then X_{t-1} is equal to its fundamental value in expectation, which has dynamics that are potentially likely to immediately initiate another bubble. In example **E2** for instance, if bubbles are very large, then c will be relatively large with respect to the dynamics of the fundamental value. When the bubble collapses at time $t-1$, then $g(F_{t-1}) = \varepsilon_{t-1} - c$, which means that there is a very high probability of a new bubble being created.

To gain more control of bubble emergence dynamics one can use more involved survival functions such as example **E3**. Here a bubble collapses if it makes out more than a fraction k of the total asset price. This allows for various bubble sizes and critical levels of the asset price as X_{t-1} depends on the fundamental value and the shock process. In fact, a higher fundamental value allows for larger bubbles, a result that can be argued to be appealing as a high fundamental price can be one of the driving reasons for the existence of the bubble. Example **E3** can control for the emergence of bubbles, as $X_{t-1} < 0$ implies that the break condition is not satisfied for any possible value of b_{t-1} and thus no bubble is created. A period of negative asset value thus ensures no bubble is created during that time, hence **E3** can be used well to describe time series which contain explosive and nonexplosive windows. Finally, example **E4** captures the same effects as **E3**, but elaborates on the connection between the fundamental value and the bubble process. The bubble size and collapse and emergence times are now all directly related to the fundamental value. If the fundamental value is below the threshold c, then no bubbles are created. If the fundamental value goes above c, then a bubble is created which grows until its size is larger than k times the difference between the fundamental value and c. There are two general driving forces that cause the bubble to burst. Firstly, the fundamental value process can stay above c for an extended period of time, but as it is mean reverting while the bubble is exponentially increasing, the bubble
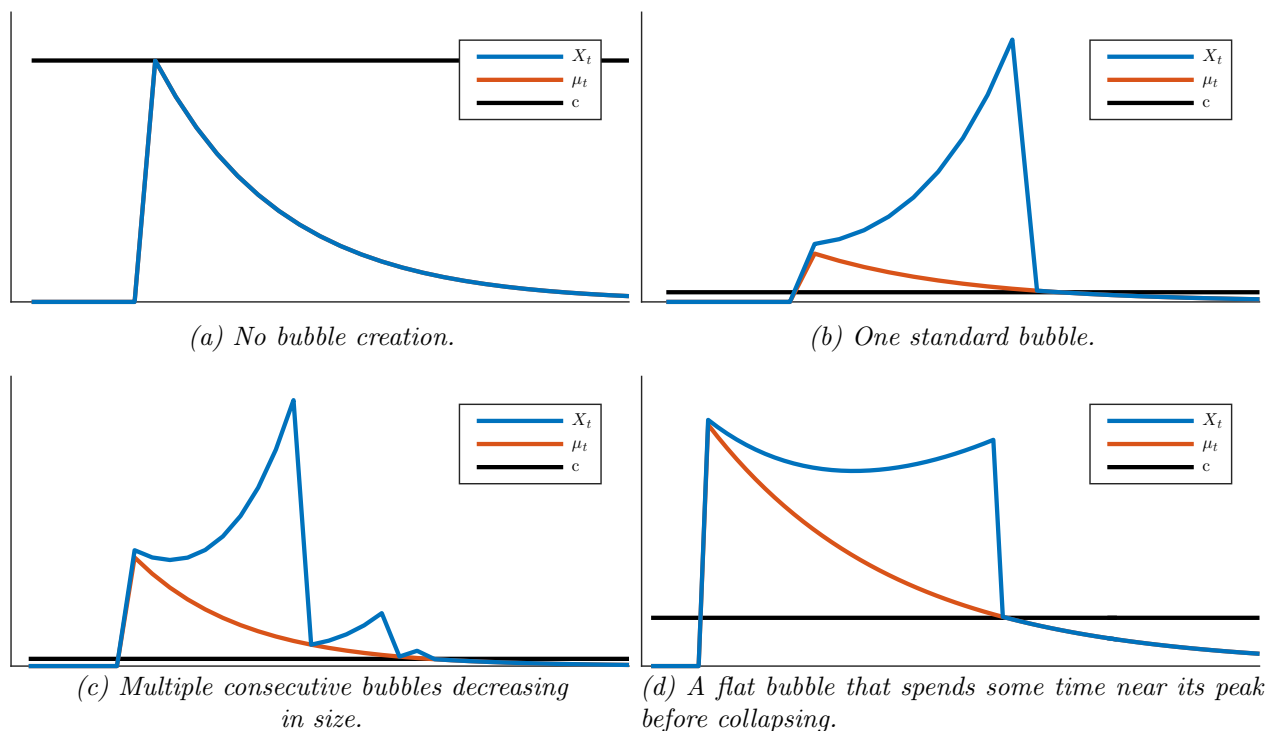
process grows much faster and thus we eventually observe that $b_{t-1} \geq k(\mu_t - c)$. Secondly, the fundamental value can fall quickly below c again, which immediately makes the bubble collapse. Combinations between these two collapse reasons are also possible, allowing for a wide variety of bubble sizes and overall asset price dynamics.

2.1   Bubble variety

The bubble process described in (2.4) might appear to be rather restricted at first sight, as conditional on the current value of the bubble the updating equation allows for only two possible values, all of which are within a countable space. However, the joint dynamics between the fundamental value and the bubble process can cause the asset price to be very flexible in describing various bubble sizes, shapes and frequencies. This is especially true for examples **E3** and **E4**, which we demonstrate in Figure 1 by examining some of the possible impulse response functions (IRFs) for the model described in equations (2.1)-(2.5) with survival condition **E4**. Figure 1(a) illustrates how a small impulse that does not push

Figure 1: *Several impulse response functions for the bubble model as described in equations* (2.1)-(2.5) *with survival condition* **E4**.

(a) No bubble creation.

(b) One standard bubble.

(c) Multiple consecutive bubbles decreasing in size.

(d) A flat bubble that spends some time near its peak before collapsing.

the fundamental process above the threshold c creates no speculative bubble. The resulting dynamics in the asset price are therefore just the mean reverting ones from the fundamental value process. In Figure 1(b) we have increased the size of the impulse, which results in a typical unique bubble characterised by its exponential increase followed by a sudden collapse. The collapse is caused by the fundamental process reverting back to its mean lower than c. If we further increase the size of the impulse as in Figure 1(c), then we obtain a similar initial scenario, but now the bubble collapses even though the fundamental value is still above the threshold, because its size is larger than $k(\mu_{t-1} - c)$. This results in another smaller bubble immediately created once the first bubble has collapsed. Finally, Figure 1(d) illustrates the effect of an impulse size that causes the mean reverting fundamental process dynamics to approximately cancel out the explosive bubble dynamics. The resulting joint dynamics for the asset price show a bubble that spends some time at its peak level before collapsing.

The different possible joint dynamics in the asset price as illustrated in Figure 1 are often encountered in financial time series. Figure 2 exhibits some time series for which evidence for the existence of a speculative bubble has been found. The bubble shapes in each time series are remarkably different. Figure 2(a) plots the monthly Nasdaq real price

Figure 2: Several time series with evidence for the existence of a speculative bubble. Panel (a) is the monthly Nasdaq real price from January 1973 to May 2005, Panel (b) is the daily Bitcoin/USD exchange rate from February 20, 2013 to July 18, 2013 and Panel (c) is the daily spread between US Baa bond rates and Aaa bond rates from January 3, 2006 to July 2, 2009.

(a) Nasdaq real price   (b) Bitcoin/USD exchange rate   (c) Bond rate spread
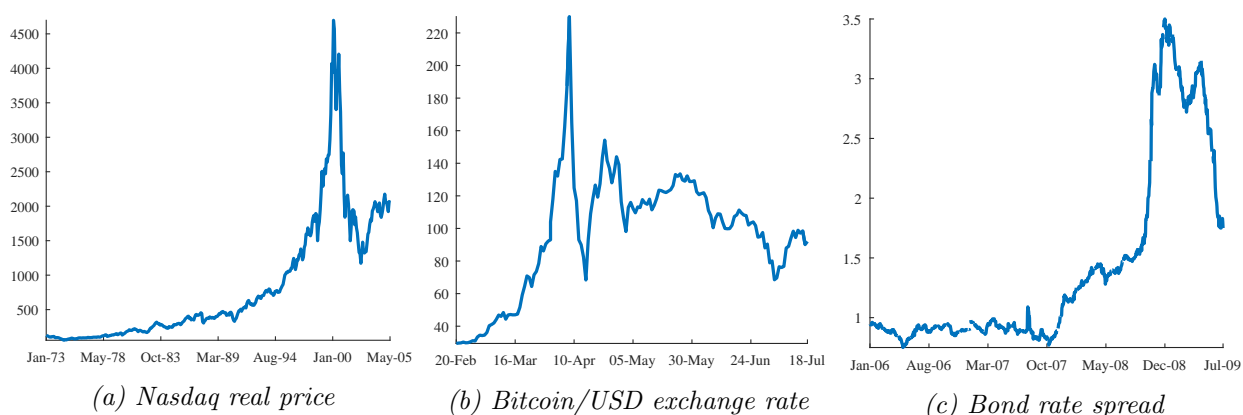
from January 1973 to May 2005, studied in Phillips et al. (2011). This time series contains a single bubble where the exponential increase is followed by an immediate burst and no new explosive behaviour, see the similarity with the impulse response function (IRF) in

Figure 1(b). Figure 2(b) depicts the daily Bitcoin/US dollar exchange rate from February 20 to July 18 in 2013, studied in Hencic and Gouriéroux (2015). This time series contains a classic bubble, which collapses on April 10. However, different to Figure 2(a), it is followed immediately by a new smaller exponential increase and downwards burst, analogous to the IRF of Figure 1(c). Figure 2(c) shows the daily spread between US Baa bond rates and Aaa bond rates from January 3, 2006 to July 2, 2009, studied in Phillips and Yu (2011). Here we observe a speculative bubble that increases exponentially, but then spends some time around its peak before collapsing, as in the IRF of Figure 1(d). All time series contain windows where no speculative bubble is apparent, as in the IRF of Figure 1(a).

3   Probabilistic and statistical analysis

In this section we study the probabilistic properties of our model as defined in equations (2.1)–(2.5). The bubble model contains several irregular components making the results in this section nonstandard. Firstly, the parameter α is allowed to be greater than one, which means that the bubble model is locally explosive on its sample space. Secondly, the updating equation for the bubble process (2.4) contains a discontinuity. These aspects imply that typical stability properties necessary for almost everywhere contraction conditions or smoothness assumptions do not hold, which means that it is not possible to employ standard stability theory results as developed in Bougerol (1993) or Meyn and Tweedie (1993). Instead we rely on previous work in Blasques and Nientker (2017) that provides stability results for resetting dynamic systems. Such a system is defined by an updating function that sometimes resets to a fixed, possibly random, value regardless of the past. These dynamics are present in the bubble process when the bubble collapses back to zero.

We split the parameter vector in two sub-vectors (θ, λ) which belong to the parameter space $\Theta \times \Lambda$. Here θ contains all the parameters that enter continuously in (μ_t, b_t) and λ contains the remaining parameters. Among the parameters $(\sigma^2, \delta, \beta, \gamma, \omega, \alpha)$, it is clear that σ^2 is always an element of θ, while the remaining parameters may be elements of θ or λ depending on the chosen survival function g. If the survival function depends on the bubble process, then (α, ω) belong to λ and if g is nonconstant in the fundamental value process, then (δ, β, γ) belong to λ. We examine the examples from Section 2 as an illustration.

9

E1 $g(F_{t-1}) = X_{t-1} - c,$ then $\theta = (\sigma^2, \delta, \beta, \gamma, \omega, \alpha)$ and $\lambda = c.$

E2 $g(F_{t-1}) = X_{t-1} - \mu_{t-1} - c,$ then $\theta = (\sigma^2, \omega, \alpha)$ and $\lambda = (\delta, \beta, \gamma, c).$

E3 $g(F_{t-1}) = b_{t-1} - kX_{t-1},$ then $\theta = (\sigma^2, \delta, \beta, \gamma)$ and $\lambda = (\omega, \alpha, k).$

E4 $g(F_{t-1}) = b_{t-1} - k(\mu_t - c),$ then $\theta = \sigma^2$ and $\lambda = (\delta, \beta, \gamma, \omega, \alpha, k).$

Deriving consistency and asymptotic normality for λ is generally difficult. Therefore we approach the problem by deriving these results for θ conditionally on a calibrated value of λ. This means that we will work with functions f in the Banach space $L^\infty(\Theta, \mathbb{R})$, we write $\|f\|_\Theta$ for the supremum norm. The parameter space Θ is assumed to be compact throughout this section.

We show in Section 3.1 that the model admits stable solutions under lenient restrictions on the parameters and the survival function. We then continue to analyse the model as a filter in Section 3.2 and show that filter paths converge to a stable solution. We derive the likelihood in Section 3.3 and provide consistency and asymptotic normality for a maximum likelihood (ML) estimator in Section 3.4. The proofs and further details are presented in Appendix A.

3.1 The model as a data generating process

This section provides results that guarantee that our model generates (strictly) stationary ergodic data with finite moments. Moreover, we show that partial solutions converge to the stationary sequence. These results will be required later on to show consistency and asymptotic normality for the ML estimator in a correctly specified model.

Data generated by our model partially adheres to very standard dynamics, as equation (2.1) holds for such data and thus (2.3) simplifies to

$$(3.1) \qquad\qquad \mu_t = \delta + \beta\mu_{t-1} + \gamma\varepsilon_{t-1}.$$

The fundamental value process therefore is an autoregressive process of order one with Gaussian errors, a specification that is well studied and known to have stable solutions. We need the following assumptions to ensure the stability results:

DGP 1. The parameter space satisfies $|\beta| < 1$.

DGP 2. Let $b \geq 0$ and $\mu, \varepsilon \in \mathbb{R}$. There exists a set $S \subset \mathbb{R} \times \mathbb{R}$ of positive Lebesque measure such that the survival function satisfies

$$\tilde{g}(\varepsilon, \mu) := \inf_{b \geq 0} g(\mu + b + \varepsilon, \mu, b) \geq 0 \quad \text{for all } (\varepsilon, \mu) \in S.$$

Assumption DGP 1 is standard in the literature on autoregressive processes. Condition DGP 2 seems complicated but essentially requires that the bubble process always has positive probability to collapse next period, regardless of its current and past values. If this were not the case, then there are scenarios in which the bubble is guaranteed to continue growing, something that can be considered unnatural. Assumption DGP 2 is usually easy to verify.

Lemma. *Assumption DGP 2 holds if $\tilde{g}$ is a continuous and surjective function.*

PROOF. The set $[0, \infty)$ contains an open subset, say O. Since $\tilde{g}$ is surjective $\tilde{g}^{-1}(O)$ is nonempty and by continuity it is open. Any nonempty open subset in Euclidean space is of positive Lebesque measure. ∎

We verify condition DGP 2 on our examples **E1**–**E4** as an illustration. If our survival function is given by **E1** then $\tilde{g}(\varepsilon, \mu) = \mu + \varepsilon - c$, if our survival function is given by **E2** then $\tilde{g}(\varepsilon, \mu) = \varepsilon - c$, if our survival function is given by **E3** then $\tilde{g}(\varepsilon, \mu) = k(\mu + \varepsilon)$ as $k \leq 1$ and finally if our survival function is given by **E4** then $\tilde{g}(\varepsilon, \mu) = -k(\delta + \beta\mu + \gamma\varepsilon - c)$. All of these functions are continuous and surjective and thus condition DGP 2 is satisfied for all our examples.

Theorem 1. *Suppose that assumptions DGP 1–2 hold. Then there exists a unique causal stationary ergodic solution $((X_t, \mu_t, b_t))_{t \in \mathbb{Z}}$ to model (2.1)–(2.5). Moreover, any other solution $((\hat{X}_t, \hat{\mu}_t, \hat{b}_t))_{t \in \mathbb{N}}$ initialised at $(\hat{X}_1, \hat{\mu}_1, \hat{b}_1)$ almost surely converges exponentially fast to the stationary ergodic one, that is*

$$\left\| (\mu_1, b_1) - (\hat{\mu}_1, \hat{b}_1) \right\|_\Theta \overset{eas}{\to} 0 \quad as \quad t \to \infty.$$

Theorem 5 establishes the existence of a unique causal solution to our model for each

choice of θ that satisfies assumptions DGP 1–2. One can simulate an arbitrary close approximation of this solution by using any initialisation of choice and discarding the first portion of the time series.

We finalize this section by providing a result on the existence of moments for the solution found in Theorem 1. Showing such existence is dependent on the chosen survival function. We will use example **E4** in our application, so we derive the result for this survival function.

Corollary 2. *Suppose that assumptions DGP 1 holds and that the survival function is given as in **E4**. Then the unique stationary ergodic solution has a uniform $n'th$ moment for all $n \in \mathbb{N}$, i.e. $\mathbb{E}\|(X_t, \mu_t, b_t)\|_\Theta^n < \infty$.*

3.2   The model as a filter

This section focusses on the model as a filter for general data $(X_t)_{t\in\mathbb{Z}}$. Such a filter will always have to be initialised at some values $\hat{\mu}_1$ and $\hat{b}_1$, as the fundamental respective bubble processes are unobserved. We impose conditions that ensure that our filtered model admits a unique stationary ergodic solution that is twice continuously differentiable, has bounded moments and that any initialised process converges to. The first set of conditions assume structure on the dependence between the (X_t). We write $\log^+(x) = \max\{0, \log x\}$ and $\mathcal{F}_s^t$ for the σ-algebra of $(X_s, \ldots, X_t)$ for any $s \leq t \in [-\infty, \infty]$.

FLT 1. The data sequence $(X_t)_{t\in\mathbb{Z}}$ is stationary ergodic and has a finite log moment, that is $\mathbb{E}\log^+ |X_t| < \infty$.

FLT 2. Each X_t is absolutely continuous with full support pdf. If $A \in \mathcal{F}_{-\infty}^{-1}$ and $B \in \mathcal{F}_0^\infty$ are events of positive probability, then $\mathbb{P}(A \text{ and } B) > 0$.

FLT 3. The conditional distributions $X_t \mid X_{t-1}, \ldots, X_{t-n}$ are absolutely continuous and of bounded density uniformly over $n \in \mathbb{N}$ and almost all possible past values with respect to Lebesque measure.

Condition FLT 1 is standard and necessary, one cannot expect to obtain stationary ergodic filter paths if the original data sequence is not so. A log moment is implied by the existence of any regular moment by Jensen's inequality. Assumption FLT 2 is less common,

but has an intuitive interpretation. It requires the sequence $(X_t)_{t\geq 0}$ to be non exclusive, that is, conditional on the past some future events are more likely than others, however, anything that was possible unconditionally can still happen. Assumption FLT 3 is a technical one that is satisfied for most reasonable distributions. We realise conditions FLT 2 and FLT 3 are unusual in the literature. We provide the following result to illustrate that many stochastic processes satisfy these conditions.

Proposition 3. *Suppose $(X_t)_{t\in\mathbb{Z}}$ is a real valued stationary ergodic solution of a Markov chain $X_t = f(X_{t-1}, \zeta_t)$. If $f(x, \cdot)$ is a continuously differentiable function for all $x \in \mathbb{R}$ with derivative bounded away from zero for almost all $x \in \mathbb{R}$ with respect to Lebesque measure, and $(\zeta_t)_{t\in\mathbb{Z}}$ is a sequence of independent, identically distributed and absolutely continuous random variables such that $f(x, \zeta_t)$ has full support for all $x \in \mathbb{R}$. Then conditions FLT 2 and FLT 3 are satisfied.*

Proposition 3 implies that typical processes such as general AR(1) given by $X_t = h(X_{t-1}) + \varepsilon_t$, or multiplicative specifications of the type $X_t = h(X_{t-1})\varepsilon_t$ usually satisfy conditions FLT 2 and FLT 3. The proposition can also be extended to multivariate processes where the data is one of the entries in the vector. This implies processes such as ARMA or GARCH satisfy our conditions.

As mentioned before, the dynamics of our model rely heavily on the survival function chosen, specifically whether g is nonconstant in any of its arguments. We provide the desired results for the most complex case in which g is nonconstant in any of its variables. We then need the following additional parameter restrictions.

FLT 4. The function g is Lipschitz with derivative bounded away from zero almost everywhere, it is monotone in its first argument, decreasing and continuous in its second argument and increasing in its third argument. Moreover, the probability $\mathbb{P}(g(X_t, \mu, 0) \geq 0)$ is positive for all $\mu \in \mathbb{R}$ and the inverse of g in its third argument is L-Lipschitz.

FLT 5. The parameters satisfy $r := |\beta - \gamma| < 1$ and the polynomial $p(x) = 1 - rx + \gamma\alpha Lx^2$ has roots outside of the unit circle.

Assumption 4 contains quite some restrictions of the survival function. It can be easily checked however that these all hold for example **E4**.

13

Theorem 4. *Suppose that assumptions FLT 1–4 hold. Then there exists a unique causal stationary ergodic solution $((\mu_t^*, b_t^*))_{t \in \mathbb{Z}}$ to model (2.3)–(2.5) that is twice continuously differentiable over Θ. Moreover, any other solution $((\hat{\mu}_t, \hat{b}_t))_{t \in \mathbb{N}}$ initialised at $(\hat{\mu}_1, \hat{b}_1)$ almost surely converges exponentially fast to the stationary ergodic one, that is,*

$$\left\| (\mu_t^*, b_t^*) - (\hat{\mu}_t, \hat{b}_t) \right\|_\Theta \overset{eas}{\to} 0 \quad as \quad t \to \infty.$$

Finally, if X_t has an n'th moment for some $n \in \mathbb{N}$, then (μ_t^*, b_t^*) has an n'th moment too.

3.3   The likelihood

As mentioned in the beginning of Section 3 we derive our asymptotic results for θ conditionally on a calibrated value of λ. The likelihood evaluated at some $\theta \in \Theta$ for a sequence $(X_1, \ldots, X_T)$ is the joint density implied by (2.1)–(2.5). The fundamental value and bubble processes are unobserved, so we choose initialised values $\hat{\mu}_1$ and $\hat{b}_1$ which deliver filtered sequences $(\hat{\mu}_t(\theta, \lambda))_{t=2}^T$ and $(\hat{b}_t(\theta, \lambda))_{t=2}^T$ according to (2.3)–(2.5). It follows that

$$X_t | X_1, \ldots, X_{t-1} = X_t | \hat{\mu}_t(\theta, \lambda), \hat{b}_t(\theta, \lambda) \sim \mathrm{N}(\hat{\mu}_t(\theta, \lambda) + \hat{b}_t(\theta, \lambda), \sigma^2)$$

and thus prediction error decomposition delivers the average log likelihood as a function $\Theta \to \mathbb{R}$ given by

$$\hat{L}_T(\theta) \propto \frac{1}{T} \sum_{t=2}^T \ell(X_t, \hat{\mu}_t(\theta, \lambda), \hat{b}_t(\theta, \lambda), \sigma^2),$$

$$\ell(X_t, \hat{\mu}_t(\theta, \lambda), \hat{b}_t(\theta, \lambda), \sigma^2) := -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X_t - \hat{\mu}_t(\theta, \lambda) - \hat{b}_t(\theta, \lambda))^2.$$

From here on out we will omit mentioning that the plug in processes depend on θ and λ to keep notation clear.

3.4 Asymptotic results

3.4.1 Consistency

The ML estimator of θ is defined as

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} \hat{L}_T(\theta).$$

We need the following conditions to obtain consistency.

CS 1. $(X_t)_{t \in \mathbb{Z}}$ is stationary and ergodic with bounded second moment: $\mathbb{E}|X_t|^2 < \infty$.

CS 2. The filter vector $((\hat{\mu}_t, \hat{b}_t))_{t \in \mathbb{N}}$ is invertible and converges to a limit process $((\mu_t^*, b_t^*))_{t \in \mathbb{Z}}$ uniformly over Θ with two uniform bounded moments. That is,

$$\left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_{\Theta} \overset{eas}{\to} 0 \quad \text{as } t \to \infty \quad \text{and} \quad \mathbb{E} \left\| (\mu_t^*, b_t^*) \right\|_{\Theta}^2 < \infty,$$

Moreover, the joint process $((X_t, \mu_t^*, b_t^*))_{t \in \mathbb{Z}}$ is strictly stationary and ergodic.

CS 3. There exists a unique maximizer θ_0 of the limit log likelihood, that is, for every $\theta \in \Theta$ that is unequal to θ_0 we have

$$\mathbb{E}\ell(X_t, \mu_t^*(\theta, \lambda), b_t^*(\theta, \lambda), \sigma^2) < \mathbb{E}\ell(X_t, \mu_t^*(\theta_0, \lambda), b_t^*(\theta_0, \lambda), \sigma_0^2).$$

The assumptions CS 1–3 are typical conditions used in the theory of M-estimators. Assumptions CS 1 and CS 2 assume stochastic properties of our model that ensure that a law of large numbers can be applied. Note that they are both implied by assumption FLT 1 – FLT 5 and Theorem 4. Assumption CS 3 ensures that the limit log likelihood is maximised at a unique point θ_0, given the fixed parameter λ. Note that the expectations exist by the moment assumptions in CS 1–2. When the model is assumed to be well specified and λ is fixed at its true value λ_0, then it is often easy to show that this assumption holds and that the parameter of interest θ_0 is the true parameter, that is, the parameter that corresponds to the data generating process for $\{X_t\}_{t \in \mathbb{Z}}$. If the model is misspecified, or λ is set at some arbitrary value $\lambda \neq \lambda_0$, then the uniqueness of the parameter of interest θ_0 is harder to establish. When the uniqueness assumption fails, set consistency can be easily established,

thus ensuring that the ML estimator converges to the limit argmin set; see Lemma 4.2 in Pötscher and Prucha (1997) for standard conditions that apply here. In this case, the limit parameter θ_0 is a 'pseudo-true parameter', i.e. a parameter that minimizes a Kullback-Leibler divergence between the true conditional density of the data and the model-implied conditional density, see Section 2.3 of White (1994).

Theorem 5 (Consistency). *If assumptions CS 1–3 hold, then $\hat{\theta}_T \overset{a.s.}{\to} \theta_0$.*

Theorem 1 establishes the a.s. convergence of the ML estimator $\hat{\theta}_T$ to the pseudo-true parameter θ_0 which is the unique maximizer of the limit log likelihood or any given value of $\lambda \in \Lambda$. In this sense, the θ_0 provides the best Kullback-Leibler approximation to the true unknown distribution of the data, for the given value of λ. Naturally, if the model is correctly specified and λ is calibrated at its true value, then θ_0 corresponds to the true parameter.

Corollary 6. *Suppose that the model is correctly specified and that λ is calibrated at its true value. If the results of Theorem 1 can be applied for all $\theta \in \Theta$, and*

$$\mathbb{E}\left\|(\mu_t, b_t)\right\|_\Theta^2 < \infty.$$

Then the ML estimator $\hat{\theta}_T$ converges a.s. to the true parameter θ_0.

3.4.2 Asymptotic normality

In what follows, we establish the asymptotic normality of the ML estimator $\hat{\theta}_T$ as $T \to \infty$. Theorem 7 focuses on the case of a well specified and correctly calibrated model, and Theorem 8 obtains asymptotic normality for a misspecified or incorrectly calibrated model where $\lambda \neq \lambda_0$. We need the following standard assumptions.

AN 1. The conditions CS 1–3 hold and θ_0 belongs to the interior of Θ.

AN 2. The limit process (μ_t^*, b_t^*) is twice continuously differentiable on Θ for all $t \in \mathbb{Z}$.

AN 3. $\{X_t\}_{t \in \mathbb{Z}}$ has four bounded moments $\mathbb{E}|X_t|^4 < \infty$.

AN 4. The Fisher information matrix is invertible.

The following assumption ensures that the filter derivatives converge almost surely and exponentially fast to limit strictly stationary and ergodic sequences. The exponential rate for the filter was established in Section 3.2. It is clear that the same argument applies to the derivative processes $\{\frac{\partial \hat{\mu}_t}{\partial \theta}\}$ and $\{\frac{\partial \hat{b}_t}{\partial \theta}\}$. In particular, we have again that $\{\frac{\partial \hat{b}_t}{\partial \theta}\}$ converges at any rate since it is reset to zero in a finite number of steps with positive probability, and $\{\frac{\partial \hat{\mu}_t}{\partial \theta}\}$ converges exponentially fast due to its autoregressive nature.

AN 5. The derivative processes are invertible at exponential rates and feature four bounded moments,

$$\left\| (\nabla^{0:2}\hat{\mu}_t, \nabla^{0:2}\hat{\beta}_t) - (\nabla^{0:2}\mu_t, \nabla^{0:2}\beta_t) \right\|_\Theta \overset{eas}{\to} 0 \quad \text{as } t \to \infty \ ,$$

$$\text{and} \quad \mathbb{E} \left\| (\nabla^{0:2}\mu_t, \nabla^{0:2}\beta_t) \right\|_\Theta^4 < \infty.$$

Theorem 7. *(Asymptotic Normality: Correct specification) Let assumptions AN 1–5 hold. Suppose further that the model is well specified and that $\lambda = \lambda_0$. Then*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{d}{\to} N(\mathbf{0}, \mathcal{I}^{-1}(\theta_0)) \quad as \ \ T \to \infty \ ,$$

where $\mathcal{I}^{-1}(\theta_0)$ denotes the inverse information matrix.

We need the following additional assumption to obtain asymptotic normality under a misspecified model.

AN 6. $\{x_t\}$ is near epoch dependent of size -1 on a ϕ-mixing sequence of size $-r/(r-1)$ for some $r > 2$.

Theorem 8. *(Asymptotic Normality: Incorrect specification/calibration) Let assumptions AN 1–6 hold. Then*

$$\sqrt{T}(\hat{\theta}_T^1 - \theta_0^1) \overset{d}{\to} N(\mathbf{0}, \Sigma(\theta_0^1, \theta^2)) \quad as \ \ T \to \infty \ ,$$

where

$$\Sigma(\theta_0^1, \theta^2) = \left(\mathbb{E}\hat{\ell}_t''(\theta_0^1, \theta^2) \right)^{-1} \left(\mathbb{E}\hat{\ell}_t'(\theta_0^1, \theta^2) \mathbb{E}\hat{\ell}_t'(\theta_0^1, \theta^2)^\top \right) \left(\mathbb{E}\hat{\ell}_t''(\theta_0^1, \theta^2) \right)^{-1}.$$

4 Illustrations

In this section we test the descriptive capability of our model and illustrate the accessability and ease of further analysis after estimation. We will use our model with survival function **E4** and estimate it on a part of the Bitcoin/US dollar exchange rate. We realise that the chosen survival function implies that most of our parameters belong to λ, the vector of parameters that enter discontinuously into the likelihood. Therefore we add a short simulation study in Section 4.1 in which we examine the distribution of our ML estimator for a representative choice of parameters. Section 4.2 contains the estimation results and further analysis.

4.1 Simulation study

We examine the distribution of the ML estimator for a given parametrization, stated in Table 1. This choice represents a medium amount of bubbles of size relative to the fundamental

σ	δ	β	γ	ω	α	k	c
1.0	0.1	0.95	0.7	0.2	1.03	7	-0.1

*Table 1: Parametrization used for simulation study.*

value process. A typical simulation for the implied model can be found in Figure 3. There

Figure 3: An example simulated path for the parametrization of the model described in Table 1.
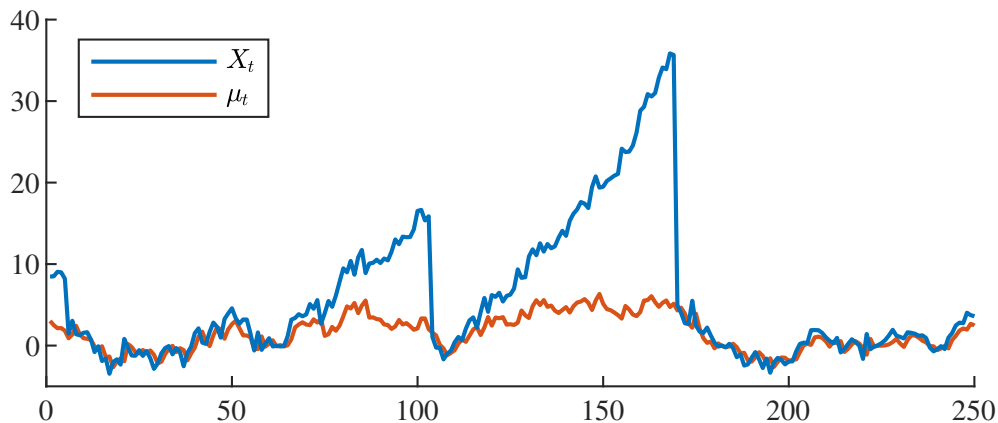
are periods of locally explosive behaviour, but also times at which no bubbles seem to form. The size of the bubbles is substantially larger than that of the fundamental value, but not so far as to render its value insignificant compared to the magnitude of the bubble process.

Figure 4: Estimated kernel densities for the ML estimator based on one thousand estimate values, each of which is based on a sample path of length one thousand.
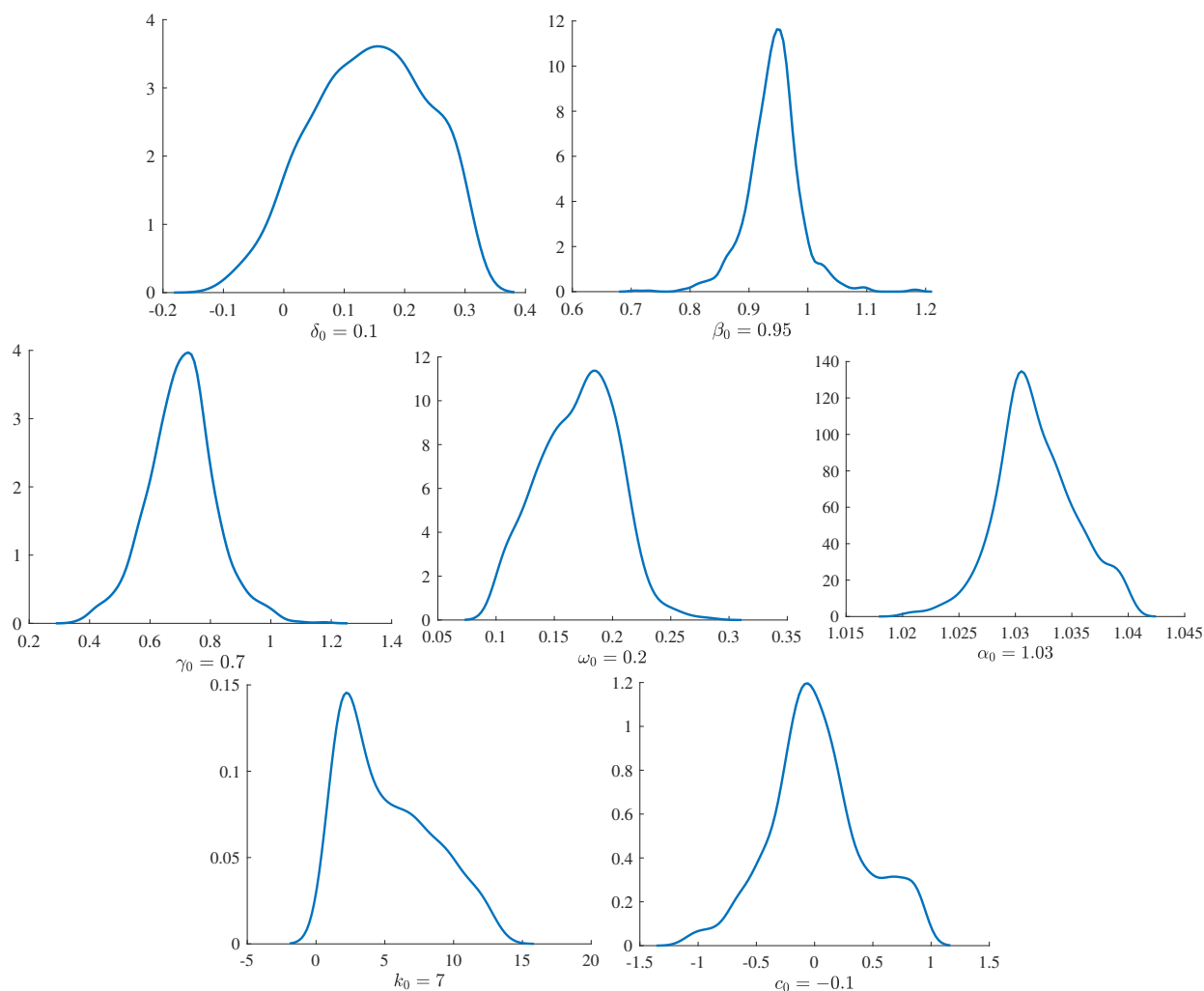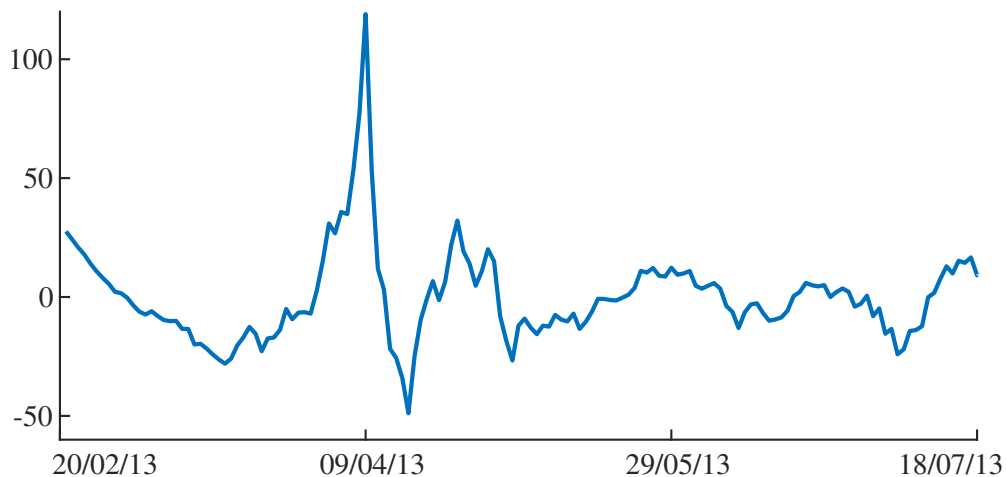
We estimate the model parameters by maximising the likelihood over an area centered around the true values. The likelihood however is discontinuous and nondifferentiable, which means that gradient based optimizing algorithms cannot be applied. Instead we implement a procedure based on the genetic algorithm in Matlab, which is inspired by natural selection observed in biological evolution. The algorithm generates a population of points and then successively selects a partially random subpopulation to be parents to the next population. We use a total of fifty generations to get near to the optimal point and then use that as a starting value for a gradient based optimizer to get quicker to the maximiser. An important observation about our procedure is that the resulting optimizing algorithm is stochastic. It implies that parameter values cannot be reproduced, however, the algorithm works sufficiently well: successive estimations on the same data produce similar results.

For the simulation we calculate one thousand estimate values, each of which is based on a sample path of length one thousand. The resulting estimated densities for the ML estimator are portrayed in Figure 4. We find that all densities, except the one for k, are close to symmetric with their peak at the true value. The estimator for k has more inaccuracy than the others, because most bubbles in this parametrization collapse due to the fundamental process dropping below the threshold c. A path of one thousand observations contains approximately ten bubbles, so therefore there is relatively little data to estimate k.

4.2   The BTC/USD exchange rate

The data set that we use is equivalent to the one studied in Hencic and Gouriéroux (2015). We take the detrended daily Bitcoin/US dollar exchange rate from February 20 to July 18 in 2013, given in Figure 5. There appears to be a big bubble that collapses on April 10, 2013. Moreover, it is potentially followed by a second smaller bubble. Afterwards it tends to behave as a standard stable and mean reverting process.

Figure 5: Detrended daily Bitcoin/USD exchange rate, taken from Hencic and Gouriéroux (2015).

σ	δ	β	γ	ω	α	k	c
2.05	-0.25	0.86	1.04	0.44	2.10	2.39	12.29

Table 2: Parameter estimates for detrended daily Bitcoin/US dollar exchange rate from February 20 to July 18, 2013.

We estimate the model parameters as discussed in the simulation study. The results are given in Table 2. The estimate of α is relatively large, which means that any potential bubble

is highly explosive. Moreover, the value of c implies that the potential smaller second bubble is mostly identified as the fundamental value moving away from its mean. These observations are substantiated when we look at the filtered time series in Figure 6, note that these are also the in sample one step ahead predictions. Our model describes only one significant bubble, which is preceded by an increase in the fundamental value. It then collapses due to the k parameter restriction on bubble size. Afterwards the fundamental value stays below the threshold c and hence the rest of the time series is filtered as an autoregressive process. Our model performs well as it predicts the burst of the bubble on April 10 correctly. It does however underestimate the additional decrease in fundamental value after the bubble burst.

Figure 6: The filtered daily Bitcoin/USD exchange rate, taken from Hencic and Gouriéroux (2015).
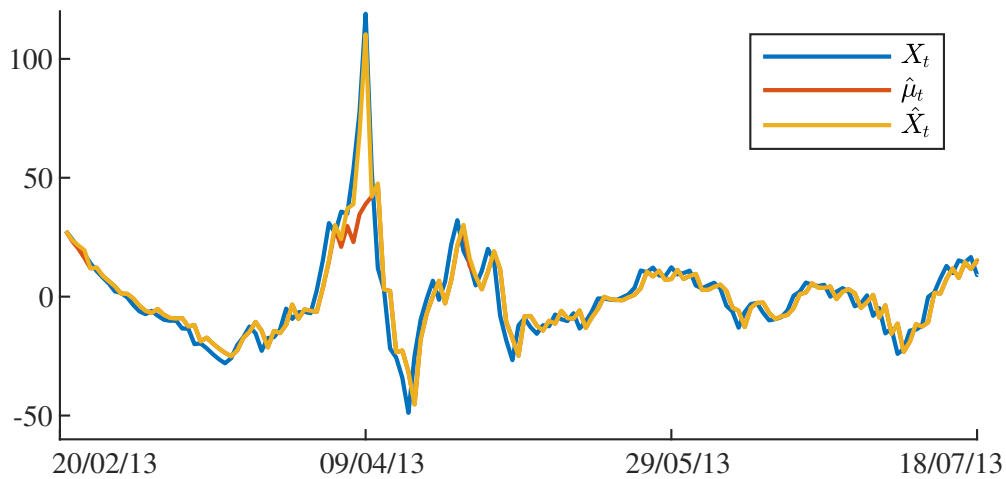

We compare our bubble model to the nested simpler model in which we set the bubble always to zero. In that case we have only four parameters left out of eight. The resulting Akaike information criteria are 646 for the full bubble model and 726 for the simpler model. Therefore we conclude that including the bubble process adds descriptive power and thus we prefer that model.

Observation driven parameter varying time series models have two main advantages. The first one is that they are easy to estimate as the likelihood is accessible through prediction error decomposition as discussed in Section 3. The second advantage is that further analysis is straightforward once the model has been estimated, as we have closed form formulas for the filtered time series. For example, we can calculate the probability that the bubble condition in the next period holds. Figure 7 plots the filtered time series and these probabilities

for some period centered around the bubble. Here we see that the probability of nonzero bubble values before the bubble start are virtually equal to zero. As the bubble starts, the probability of a nonzero bubble is almost one. The probability dips a little for April 8, as the fundamental value on April 7 has gone down a little. The fundamental value then increases on April 8 however and thus so does the probability for April 9. When we get to April 10 the probability is again almost zero as the large exponential growth of the bubble has outgrown the fundamental value process on April 9 by far too much.

Figure 7: The filtered daily Bitcoin/USD exchange rate around the bubble in the top frame, and the probability that the bubble condition holds in the next period shifted one period to the right in the second frame.
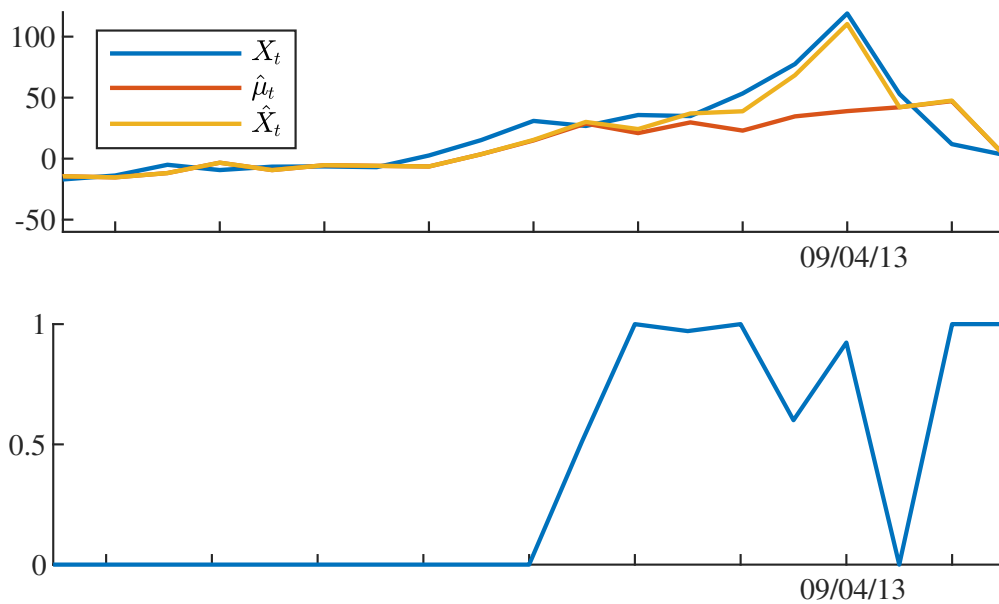
The estimated probabilities above are just an example of many possible features that can be predicted. For example, once the model parameters are estimated, one can predict expected remaining bubble life, the probability that a bubble will emerge at a given time, or the expected maximal bubble size.

5   Conclusions

We have introduced a new observation driven time varying parameter model to describe locally explosive behaviour in a stationary setting. We do so by splitting the asset price into the sum of its fundamental value and a speculative bubble. For the sake of simplicity we have

assumed an AR(1) process for the fundamental value and a collapsing AR(1) for the bubble process. However, these are not binding assumptions and many extensions and variations are possible. Of course any mean reverting stationary process for the fundamental value fits exactly in the theoretic domain presented in the paper. Dynamics can be changed by using a nonstationary process such as a random walk for the fundamental value. Most asset pricing data is nonstationary, so this would mean that one does not have to detrend the data, which makes out of sample forecasting possible. Other possibilities that could be explored are extensions to the break condition. One could for example add external stochastics allowing for more structural models that include specific financial or economic variables that can help in predicting bubble collapses. Another extension can be made by changing the sudden collapse in a more smooth exponential decrease. The dynamics of this model would then be very close to those described in the mixed causal and noncausal literature.

A   Proofs

A.1   Proof of Theorem 1

It is straightforward to verify the assumptions of Theorem 3.1 in Bougerol (1993) for the system defined in equations (2.2) and (3.1). It then immediately follows that there exists a unique causal stationary ergodic sequence $(\mu_t)_{t \in \mathbb{Z}}$ that satisfies (3.1) and that any other solution $(\hat{\mu}_t)_{t \in \mathbb{N}}$ initialised at $\hat{\mu}_1$ satisfies

$$\|\mu_t - \hat{\mu}_t\|_\Theta \overset{eas}{\to} 0 \quad \text{as} \quad t \to \infty.$$

Next, we substitute equation (2.1) into the survival function to obtain the bubble updating function

$$b_{t+1} = \phi_t(b_t), \quad \text{where} \quad \phi_t(b) = (\omega + \alpha b)\mathbb{1}\left\{g(\mu_t + b + \varepsilon_t, \mu_t, b) < 0\right\}.$$

We then check Assumption A for Theorem 2.1 in Blasques and Nientker (2017). Assumption A1 is trivial and A2 is satisfied by Krengel's Lemma: a measurable function of a stationary ergodic sequence produces a stationary ergodic sequence, see Proposition 4.3 in Krengel

(1985). For the final assumption A3 we note that $\phi_t(b) = 0$ for all $b \in [0, \infty)$, if

$$g(\mu_t + b + \varepsilon_t, \mu_t, b) \geq \inf_{b \geq 0} g(\mu_t + b + \varepsilon_t, \mu_t, b) \geq 0.$$

Therefore Assumption A3 is satisfied if

$$\mathbb{P} \left(\inf_{b \geq 0} g(\mu_t + b + \varepsilon_t, \mu_t, b) \geq 0 \right) > 0.$$

This is implied by condition DGP 2, because (2.2) implies that μ_t and ε_t are independent and both absolutely continuous on the real line. Therefore the joint random variable (μ_t, ε_t) is absolutely continuous on $\mathbb{R}^2$ and thus $\mathbb{P}((\mu_t, \varepsilon_t) \in S) > 0$. We conclude that there exists a unique causal stationary ergodic sequence $(b_t)_{t \in \mathbb{Z}}$ that satisfies the bubble updating process defined in (2.1)–(2.5) and that any other solution $(\hat{b}_t)_{t \in \mathbb{N}}$ initialised at $\hat{b}_1$ satisfies

$$\left\| b_t - \hat{b}_t \right\|_\Theta \xrightarrow{eas} 0 \quad \text{as} \quad t \to \infty.$$

The final conclusion follows again by Krengel's Lemma.

A.2 Proof of Corollary 2

The solution for the fundamental process found in Theorem 1 is given by

$$\mu_t = \sum_{i=0}^{\infty} \beta \varepsilon_{t-i-1},$$

so all moments of μ_t are finite over Θ as it is a compact space so that $|\beta|$ is bounded and the ε_t are Gaussian and thus have all finite moments. For the bubble process we have

$$b_t = (\omega + \alpha b_{t-1})\mathbb{1}\{b_{t-1} < k(\mu_t - c)\} \leq \omega + \max\{\alpha k(\mu_t - c), 0\}$$

and hence moment existence follows from those of the fundamental value process.

A.3 Proof of Proposition 3

We fix some $\zeta \in \mathbb{R}$ and let $\xi = f(x, \zeta)$. The fact that $f(x, \cdot)$ is continuously differentiable implies by the inverse function theorem that it is invertible on a neighbourhood O around ζ and that the inverse $f^{-1}(x, \cdot)$ is also continuously differentiable. Moreover, by assumption, for Lebesque almost all $x \in \mathbb{R}$ there exists an $L > 0$ such that $|f'(x, \zeta)| \geq L$ and thus $f^{-1}(x, \cdot)$ is Lipschitz as

$$\frac{d}{d\xi} f^{-1}(x, \xi) = \frac{1}{f'(x, \zeta)} \leq \frac{1}{L}.$$

The real line is separable, hence we can choose a countable number of disjunct compact neighbourhoods $\{O_k\}_{k \in \mathbb{N}}$ whose union is equal to $\mathbb{R}$ and $f(x, \cdot)$ is invertible on each neighbourhood as above. A continuously differentiable function on a compact set is absolutely continuous, which in turn implies that it has the Luzin property, that is, sets of measure zero are mapped to sets of measure zero.

We now prove that X_t is absolutely continuous. Let $E \subset \mathbb{R}$ be a set of Lebesque measure zero and let F denote the distribution function of X_t, then by independence of the ζ_t we have

$$
\begin{aligned}
\mathbb{P}(X_t \in E) &= \mathbb{P}(f(X_{t-1}, \zeta_t) \in E) \\
&= \int_{\mathbb{R}} \mathbb{P}(f(x, \zeta_t) \in E)\, F(dx) = \sum_{k=1}^{\infty} \int_{\mathbb{R}} \mathbb{P}(f(x, \zeta_t) \in E \cap O_k)\, F(dx) \\
&= \sum_{k=1}^{\infty} \int_{\mathbb{R}} \mathbb{P}(\zeta_t \in f^{-1}(x, E \cap O_k))\, F(dx) = \sum_{k=1}^{\infty} \int_{\mathbb{R}} 0\, F(dx) = 0,
\end{aligned}
$$

where we used that $E \cap O_k$ has Lebesque measure zero, $f^{-1}(x, \cdot)$ has the Luzin property on each O_k and ζ_t is absolutely continuous. The absolute continuity of the conditional distributions follows similarly as by the independence of the ζ_t and the Markov property

$$\mathbb{P}(X_t \in E \mid X_{t-1} = x_1, \ldots, X_{t-n} = x_n) = \mathbb{P}(X_t \in E \mid X_{t-1} = x_1) = \mathbb{P}(f(x_1, \zeta_t) \in E) = 0.$$

Next we show that the conditional densities are uniformly bounded. By assumption, we know that the density of ζ_t is bounded by some $B > 0$. For some $\xi \in \mathbb{R}$ and $\eta > 0$ we then

have

$$\mathbb{P}(\xi < X_t \le \xi + \eta \mid X_{t-1} = x_1, \ldots, X_{t-n} = x_n) = \mathbb{P}(\xi < f(x_1, \zeta_t) \le \xi + \eta)$$

$$= \sum_{k=1}^{\infty} \mathbb{P}(\zeta_t \in f^{-1}(x_1, (\xi, \xi + \eta] \cap O_k)) \le \frac{B}{L}\eta,$$

where we used that $f^{-1}(x, \cdot)$ is $\frac{1}{L}$-Lipschitz on each O_k and the density of ζ_t is bounded by B. Taking the limit of $\eta \to 0$ shows that the conditional densities are all bounded by $\frac{B}{L}$.

Finally we note that the full support of X_t follows directly from the fact that $f(x, \zeta_t)$ has full support for all $x \in \mathbb{R}$ and ζ_t is absolutely continuous, and show the non exclusive property in FLT 2. This follows from the Markov chain setup. Let $A_0 \subset \mathbb{R}$ be all points x such that $\mathbb{P}(A \mid X_0 = x) > 0$. Then A_0 has positive Lebesque measure as X_0 is absolutely continuous and

$$P(X_0 \in A_0) \ge \mathbb{P}(A \text{ and } X_0 \in A_0) = P(A) > 0.$$

It follows that

$$\mathbb{P}(A \text{ and } B) = \int_{\mathbb{R}} \mathbb{P}(A \text{ and } B \mid X_0 = x) \, F(dx)$$

$$= \int_{\mathbb{R}} \mathbb{P}(A \mid X_0 = x)\mathbb{P}(B \mid X_0 = x) \, F(dx) > 0,$$

where we used that $\mathbb{P}(A \mid X_0 = x)$ is greater than zero on a set of positive Lebesque measure and $\mathbb{P}(B \mid X_0 = x)$ is greater than zero for all $x \in \mathbb{R}$ as $f(x, \zeta_t)$ has full support for all $x \in \mathbb{R}$ and ζ_t is absolutely continuous.

A.4   Proof of Theorem 4

A.4.1   The existence of a stationary ergodic solution

We follow the approach used in Theorem 3.1 of Bougerol (1993) where we expand the model equations backwards and show that this converges to a stationary ergodic solution. We define the joint updating equation $(\mu_t, b_t) = \Phi_{t-1}(\mu_{t-1}, b_{t-1})$, where for $u \in \mathbb{R}$ and $v \ge 0$ we

have

$$\Phi_{t-1}(u, v) = (\phi_{t-1}(u, v), \psi_{t-1}(u, v)),$$

$$\phi_{t-1}(u, v) = \delta + ru + \gamma X_{t-1} - \gamma v,$$

$$\psi_{t-1}(u, v) = (\omega + \alpha v)\mathbb{1}\{g(X_{t-1}, u, v) < 0\}.$$

To ease notation we write $\mu_t^{(0)} = u$ and $b_t^{(0)} = v$ and then define the backward iterates recursively for $m \in \mathbb{N}$ as

$$\mu_t^{(m)} = \phi_{t-1}\left(\mu_{t-1}^{(m-1)}, b_{t-1}^{(m-1)}\right) \qquad \text{and} \qquad b_t^{(m)} = \psi_{t-1}\left(\mu_{t-1}^{(m-1)}, b_{t-1}^{(m-1)}\right).$$

The goal will be to show that $b_t^{(m)}$ is almost surely eventually constant as $m \to \infty$ and that $\lim_{m\to\infty} \mu_t^{(m)}$ exists. The stationary ergodic solution is then given by

$$\left(\lim_{m\to\infty} \mu_t^{(m)}, \lim_{m\to\infty} b_t^{(m)}\right)_{t\in\mathbb{Z}}.$$

It is stationary ergodic by Corollary 2.1.3 of Straumann and Mikosch (2006) and it is a solution since

$$\lim_{m\to\infty} \mu_t^{(m)} = \lim_{m\to\infty} \phi_{t-1}\left(\mu_{t-1}^{(m-1)}, b_{t-1}^{(m-1)}\right) = \phi_{t-1}\left(\lim_{m\to\infty} \mu_{t-1}^{(m-1)}, \lim_{m\to\infty} b_{t-1}^{(m-1)}\right),$$

where we are allowed to swap the limit in because the second argument is eventually constant and ϕ_t is continuous in its first argument. Similarly

$$\lim_{m\to\infty} b_t^{(m)} = \lim_{m\to\infty} \psi_{t-1}\left(\mu_{t-1}^{(m-1)}, b_{t-1}^{(m-1)}\right) = \psi_{t-1}\left(\lim_{m\to\infty} \mu_{t-1}^{(m-1)}, \lim_{m\to\infty} b_{t-1}^{(m-1)}\right),$$

where we are allowed to swap the limit in because the second argument is eventually constant, the random variable

$$g\left(X_{t-1}, \lim_{m\to\infty} \mu_{t-1}^{(m-1)}, \lim_{m\to\infty} b_{t-1}^{(m-1)}\right)$$

is absolutely continuous by assumption FLT 3 and the fact that g is monotone in its first argument, and finally because g is continuous in its second argument by assumption FLT 4.

Lemma. *The sequence $b_0^{(m)}$ is eventually constant as $m \to \infty$ and the limit $\lim_{m \to \infty} \mu_0^{(m)}$ converges almost surely.*

PROOF. Define

$$s_t = \limsup_{m \to \infty} \mu_t^{(m)} \qquad \text{and} \qquad i_t = \liminf_{m \to \infty} \mu_t^{(m)}.$$

The proof that the backward iterate limits above exist consists of two steps that we will show later:

(i) We show that for every $\eta > 0$ there exists an event $A_\eta \in \mathcal{F}_{-\infty}^{-1}$ of positive probability, such that conditional on A_η we have almost surely $s_0 - i_0 < \eta$ and $b_0^{(m)}$ is constant for sufficiently large m.

(ii) We show that there exists an event $B_\eta \in \mathcal{F}_{-\infty}^{\infty}$ that contains A_η, such that conditional on B_η we have $s_t - i_t \le r^t(s_0 - i_0)$ for all $t \in \mathbb{N}$. Moreover $b_t^{(m)}$ is eventually constant for all $t \in \mathbb{N}$.

Since $(X_t)_{t \in \mathbb{Z}}$ is stationary ergodic there almost surely are infinitely many $0 > -t_1 > -t_2 > \ldots$ for which the event $B\eta$ shifted by t_k to the right occurs. If it occurs for such a $-t_k$, then

$$s_0 - i_0 \le r^{t_k}(s_{-t_k} - i_{-t_k}) \le r^{t_k}\eta.$$

Taking the limit of $k \to \infty$ then delivers $s_0 = i_0$ and thus the limit $\lim_{m \to \infty} \mu_0^{(m)}$ converges almost surely. The fact that $\left(b_0^{(m)}\right)_{m \in \mathbb{N}}$ is eventually constant follows immediately from part (ii) and the same argument that the event B_η occurs for some $-t < 0$. ∎

Lemma. *Claim (i) holds.*

PROOF. We start out by showing that almost surely $s_t - i_t < \infty$. This follows by a series of upper bounds. Firstly we have

$$\mu_t^{(m)} = \delta + r\mu_{t-1}^{(m-1)} + \gamma X_{t-1} - \gamma b_{t-1}^{(m-1)} \le \delta + r\mu_{t-1}^{(m-1)} + \gamma X_{t-1}$$

28

Assumptions FLT 5 and FLT 1 together with Lemma 2.1 in Straumann and Mikosch (2006) ensure that expanding backwards and taking the limit converges, hence $s_t < \infty$. The infimum requires more work. Note that the Liptschitz condition of the inverse of g in its third argument as stated in assumption FLT 4 ensure the existence of two constants $L, K > 0$ such that

$$
\begin{aligned}
b_t^{(m)} &= \left(\omega + \alpha b_{t-1}^{(m-1)}\right) \mathbb{1}\left\{g\left(X_{t-1}, \mu_{t-1}^{(m-1)}, b_{t-1}^{(m-1)}\right) < 0\right\} \\
&= \left(\omega + \alpha b_{t-1}^{(m-1)}\right) \mathbb{1}\left\{b_{t-1}^{(m-1)} < g^{-1}\left(X_{t-1}, \mu_{t-1}^{(m-1)}, 0\right)\right\} \\
&\leq \omega + \alpha \max\left\{g^{-1}\left(X_{t-1}, \mu_{t-1}^{(m-1)}, 0\right), 0\right\} \\
&\leq \omega + \alpha \max\left\{K + L\left(\mu_{t-1}^{(m-1)} + X_{t-1}\right), 0\right\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mu_t^{(m)} &= \delta + r\mu_{t-1}^{(m-1)} + \gamma X_{t-1} - \gamma b_{t-1}^{(m-1)} \\
&\geq (\delta - \gamma(\omega + \alpha K)) + r\mu_{t-1}^{(m-1)} - \gamma\alpha L\mu_{t-2}^{(m-2)} + \gamma\left(X_{t-1} - \alpha L X_{t-2}\right).
\end{aligned}
$$

Again assumptions FLT 5 and FLT 1 together with Lemma 2.1 in Straumann and Mikosch (2006) ensure that expanding backwards converges, hence $i_t > -\infty$. We conclude that $s_t - i_t < \infty$. Note that these bounds immediately prove the moment statement in Theorem 4.

Next, we choose an $M > 0$ such that $\mathbb{P}\left(s_t - i_t < M\right) > 0$ and let $t = \left\lceil \frac{\log(\eta/m)}{\log r} \right\rceil$, where $\lceil x \rceil$ is the smallest integer larger than x. Continuity of g in its second argument, the positive probability condition in assumption FLT 4 and assumption FLT 2 guarantee that by conditioning on the past we can show for each $0 \leq v < t$ that

$$
\mathbb{P}\left(\limsup_{m\to\infty} b_{-v}^{(m)} = 0\right) \geq \mathbb{P}\left(g\left(X_{-v-1}, s_{-v-1}, 0\right) \geq 0\right) > 0.
$$

It follows by Assumption FLT 2 that there exists an event $A_\eta \in \mathcal{F}_{-\infty}^{-1}$ of positive probability such that

$$
s_t - i_t < M \qquad \text{and} \qquad \limsup_{m\to\infty} b_{-v}^{(m)} = 0 \qquad \text{for all } 0 \leq v < t.
$$

This then implies that

$$s_0 - i_0 = r(s_{-1} - i_{-1}) + \gamma \left(\limsup_{m \to \infty} b_{-1}^{(m)} - \liminf_{m \to \infty} b_{-1}^{(m-1)} \right) = r^t(s_{-t} - i_{-t}) < r^t M \leq \eta,$$

which concludes the proof of part (i). ∎

Lemma. *Claim (ii) holds.*

PROOF. The argument will be a recursive one, conditional on A_η. Suppose that $s_t - i_t < r^t \eta$ and $b_t^{(m)}$ is eventually constant, then

$$s_{t+1} - i_{t+1} = r(s_t - i_t) < r^{t+1}\eta.$$

Next we show that there exists an event such that $b_{t+1}^{(m+1)}$ is eventually constant. Note that this holds if and only if

(A.1) $$\text{sign}(g(X_t, i_t, b_t)) = \text{sign}(g(X_t, s_t, b_t)),$$

where $b_t = \lim_{m \to \infty} b_t^{(m)}$. The Lipschitz condition in assumption FLT 4 implies that there exists a $K > 0$ such that

$$|g(X_t, s_t, b_t) - g(X_t, i_t, b_t)| \leq K(s_t - i_t) < K r^t \eta.$$

Moreover, the derivative being bounded away from zero by at least some $B > 0$ and the monotonicity of g in its second argument implied by assumption FLT 4 then ensure that (A.1) follows from

$$|g(X_t, s_t, b_t)| > B K r^t \eta.$$

We conclude that statement (ii) follows if $|g(X_t, s_t, b_t)| > B K r^t \eta$ for all $t \in \mathbb{N}$.

Next we determine the probability of this event. Let $I_{2Kr^t\eta}(s_t, b_t)$ be a stochastic interval of length $K r^t \eta$ such that if $|g(X_t, s_t, b_t)| \leq B K r^t \eta$ then $X_t \in I_{2Kr^t\eta}(s_t, b_t)$. Then by

30

assumption FLT 3 there exists a $U > 0$ such that

$$
\begin{aligned}
\mathbb{P}\left(|g(X_t, s_t, b_t)| \le BKr^t\eta \,\middle|\, A_\eta\right) &\le \mathbb{P}\left(X_t \in I_{2Kr^t\eta}(s_t, b_t) \,\middle|\, A_\eta\right) \\
&= \int \mathbb{P}\left(X_t \in I_{2Kr^t\eta}(s_t, b_t) \,\middle|\, s_t, b_t, A_\eta\right) d\mathbb{P}(s_t, b_t) \\
&\le \int 2UKr^t\eta \, d\mathbb{P}(s_t, b_t) = 2UKr^t\eta.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbb{P}\left(|g(X_t, s_t, b_t)| > BKr^t\eta, \ \forall t \in \mathbb{N} \,\middle|\, A_\eta\right) &= 1 - \mathbb{P}\left(|g(X_t, s_t, b_t)| \le BKr^t\eta, \ \exists t \in \mathbb{N} \,\middle|\, A_\eta\right) \\
&\ge 1 - \sum_{t=1}^{\infty} \mathbb{P}\left(|g(X_t, s_t, b_t)| \le BKr^t\eta \,\middle|\, A_\eta\right) \\
&\ge 1 - \sum_{t=1}^{\infty} 2UKr^t\eta \ge 1 - \frac{2UK\eta r}{1 - r}.
\end{aligned}
$$

This last number can be made larger than zero by choosing η sufficiently small. ∎

A.4.2 Partial solutions and continuous differentiability

The convergence of partial solutions to the true ones is essentially almost the same as the one for the existence of a stationary ergodic solution. We can use the same bounds as in statement (i) to show that $|\mu_t^*|$ and $|\hat{\mu}_t|$ are bounded by some η with positive probability and that their respective bubble processes are zero. It then follows by the same derivation as in part (ii) that they converge with positive probability. As $(X_t)_{t \in \mathbb{Z}}$ is stationary ergodic this event happens with probability one at some point in time and thus we get the convergence.

Continuous differentiability follows by the same way as in Straumann and Mikosch (2006). The stochastic recurrence equations for the derivatives of the fundamental and bubble processes are either linear or standard resetting systems. Therefore their respective backward iterations converge to stationary ergodic solutions. This then implies the continuous differentiability by a standard analysis result.

A.5 Proof of Theorem 5

We follow the usual consistency proof for M-estimators which involves showing firstly the uniform convergence of the sample average log likelihood to the limit log likelihood and secondly the identifiable uniqueness of the parameter of interest; see e.g. Theorem 3.4 in White (1994) or Lemma 3.1 in Pötscher and Prucha (1997). To ease notation we define the following functions $\Theta \to \mathbb{R}$:

$$\hat{\ell}_t = \ell(X_t, \hat{\mu}_t(\cdot, \lambda), \hat{b}_t(\cdot, \lambda), \sigma^2),$$

$$\ell_t^* = \ell(X_t, \mu_t^*(\cdot, \lambda), b_t^*(\cdot, \lambda), \sigma^2).$$

Lemma. *The sample average log likelihood almost surely converges uniformly to the limit log likelihood, i.e.*

$$\left\| \hat{L}_T - \mathbb{E}\ell_t^* \right\|_\Theta \xrightarrow{as} 0 \quad as \quad T \to \infty.$$

PROOF. We have

$$(\text{A.2}) \qquad \left\| \hat{L}_T - \mathbb{E}\ell_t^* \right\|_\Theta = \left\| \frac{1}{T} \sum_{t=2}^T \hat{\ell}_t - \mathbb{E}\ell_t^* \right\|_\Theta \leq \frac{1}{T} \sum_{t=2}^T \left\| \hat{\ell}_t - \ell_t^* \right\|_\Theta + \left\| \frac{1}{T} \sum_{t=2}^T \ell_t^* - \mathbb{E}\ell_t^* \right\|_\Theta.$$

We will show that the two rightmost terms in (A.2) go to zero as $T \to \infty$. For the first term note that ℓ is a differentiable function, we write

$$\ell_f(\tilde{\mu}, \tilde{b}) = \left. \frac{\partial \ell(X_t, \mu, b, \sigma^2)}{\partial(\mu, b)} \right|_{(\tilde{\mu}, \tilde{b})}.$$

We then invoke the mean value theorem to obtain the existence of some $(\tilde{\mu}_t, \tilde{b}_t)$ between

$(\hat{\mu}_t, \hat{b}_t)$ and (μ_t^*, b_t^*) that satisfies

$$\left\| \hat{\ell}_t - \ell_t^* \right\|_\Theta \leq \left\| \ell_f(\tilde{\mu}_t, \tilde{b}_t) \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta$$

(A.3)
$$\leq \left\| \ell_f(\tilde{\mu}_t, \tilde{b}_t) - \ell_f(\mu_t^*, b_t^*) \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta$$

$$+ \left\| \ell_f(\mu_t^*, b_t^*) \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta$$

The function ℓ_f is linear in its arguments and thus is a K-Lipschitz function for some $K > 0$. Therefore assumption CS 2 guarantees that

$$\left\| \ell_f(\tilde{\mu}_t, \tilde{b}_t) - \ell_f(\mu_t^*, b_t^*) \right\|_\Theta \leq K \left\| (\tilde{\mu}_t, \tilde{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta$$

$$\leq K \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta \overset{eas}{\to} 0 \quad \text{eas} \quad t \to \infty \,,$$

hence (A.3) almost surely goes to zero exponentially fast by assumption CS 2 and thus we have

$$\frac{1}{T} \sum_{t=2}^{T} \left\| \ell_f(\tilde{\mu}_t, \tilde{b}_t) - \ell_f(\mu_t^*, b_t^*) \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta \overset{as}{\to} 0 \quad \text{as} \quad T \to \infty.$$

Next, note that $\left(\left\| \ell_f(\mu_t^*, b_t^*) \right\|_\Theta \right)_{t \in \mathbb{Z}}$ is a stationary sequence by assumption CS 2 and Proposition 4.3 in Krengel (1985). Therefore

$$\frac{1}{T} \sum_{t=2}^{T} \left\| \ell_f(\mu_t^*, b_t^*) \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu_t^*, b_t^*) \right\|_\Theta \overset{as}{\to} 0 \quad \text{as} \quad T \to \infty$$

if $\left\| \ell_f(\mu_t^*, b_t^*) \right\|_\Theta$ has a log moment by assumption CS 2 and Lemma 2.1 in Straumann and Mikosch (2006). Let $\log^+(x) = \max\{0, \log x\}$. The log moment follows from the fact that

$$\mathbb{E} \log^+ \left\| \ell_f(\mu_t^*, b_t^*) \right\|_\Theta = \frac{1}{\sigma^2} \mathbb{E} \log^+ \left\| \begin{array}{c} X_t - \mu_t^* - b_t^* \\ X_t - \mu_t^* - b_t^* \end{array} \right\|_\Theta ,$$

the finiteness of which is implied by the moment conditions in assumptions CS 1 and CS 2. We conclude that the first term in (A.2) converges to zero almost surely.

Finally, we discuss the second term in (A.2). We show that

$$\left\| \frac{1}{T} \sum_{t=2}^{T} \ell_t^* - \mathbb{E}\ell_t^* \right\|_\Theta \overset{as}{\to} 0 \quad \text{as} \quad T \to \infty$$

by application of the uniform law of large numbers, Theorem 2.7 in Straumann and Mikosch (2006). The law of large numbers holds since $(\ell_t^*)_{t \in \mathbb{N}}$ is strictly stationary and ergodic by assumption CS 2 and Proposition 4.3 in Krengel (1985), and because

$$\mathbb{E}\|\ell_t^*\|_\Theta = \mathbb{E}\left\| \frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X_t - \mu_t^* - \beta_t^*)^2 \right\|_\Theta$$
$$\leq \left\| \frac{1}{2}\log(2\pi\sigma^2) \right\|_\Theta + c\left\| \frac{1}{2\sigma^2} \right\|_\Theta \left(\mathbb{E}X_t^2 + \mathbb{E}\|\mu_t^*\|_\Theta^2 + \mathbb{E}\|b_t^*\|_\Theta^2 \right),$$

for some $c > 0$. This upper bound is finite by assumption CS 2, because Θ is compact and the fact that $\sigma^2 > 0$. ∎

Lemma. *The parameter θ_0 is identifiable unique on Θ.*

PROOF. The identifiable uniqueness of $\theta_0 \in \Theta$ is implied by the uniqueness assumption CS 3, the continuity of $\mathbb{E}\ell_t^*$ and the compactness of Θ, see Chapter 3 in Pötscher and Prucha (1997). The continuity of $\mathbb{E}\ell_t^*$ follows directly from the fact that the sample likelihood, which is continuous, converges uniformly to $\mathbb{E}\ell_t^*$. ∎

A.6   Proof of Corollary 6

Theorem 5 ensures that $(\mu_t^*, b_t^*)_{t \in \mathbb{Z}} = (\mu_t, b_t)_{t \in \mathbb{Z}}$, so that condition CS 2 follows. The maximiser of the limit log likelihood is equal to the minimiser of the Kullback-Leibler divergence between the true conditional density of the data and the model-implied conditional density, see for instance Section 2.3 of White (1994). Therefore it follows by the Gibbs inequality that the limit log likelihood is uniquely maximised at the true parameter θ_0 and thus condition CS 3 holds for the true parameter.

A.7 Proof of Theorem 7

This proof is identical to Section 7 of Straumann and Mikosch (2006).

A.8 Proof of Theorem 8

The desired result follows by the same argument as used above for proving asymptotic normality under correct specification, with the exception that the score is not granted to be a martingale difference sequence. However, by assumptions AN 5 and AN 6, we have that the score sequence is near epoch dependent of size -1 on a ϕ-mixing sequence of size $-r/(r-1)$ for some $r > 2$. Given the moment bounds, we can thus appeal to the central limit theorem for near epoch dependent sequences in Potscher and Prucha (1997, Theorem 10.2).

References

Blanchard, O. J. and Watson, M. W. (1982). Bubbles, rational expectations and financial markets. *NBER working paper*, w0945.

Blasques, F. and Nientker, M. H. C. (2017). A stochastic recurrence equation approach to stationarity and phi-mixing of a class of nonlinear ARCH models. *Tinbergen Institute discussion paper*, 17-072/III.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization*, 31(4):942–959.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Diba, B. T. and Grossman, H. I. (1988). Explosive rational bubbles in stock prices? *The American Economic Review*, 78(3):520–530.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press, Oxford, 2nd edition.

Evans, G. W. (1991). Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, 81(4):922–930.

Fries, S. and Zakoïan, J.-M. (2017). Mixed causal-noncausal ar processes and the modelling of explosive bubbles. *CREST working paper*.

Gourieroux, C. and Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37(3):405–430.

Gouriéroux, C. and Zakoïan, J.-M. (2013). Explosive bubble modelling by noncausal process. *CREST working paper*.

Gouriéroux, C. and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):737–756.

Hencic, A. and Gouriéroux, C. (2015). Noncausal autoregressive model in application to Bitcoin/USD exchange rates. In *Econometrics of Risk*, pages 17–40. Springer, New York.

Homm, U. and Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics*, 10(1):198–231.

Krengel, U. (1985). *Ergodic theorems*. Walter de Gruyter, Berlin.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability*. Springer, London.

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

Phillips, P. C. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer, Berlin.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, 34(5):2449–2495.

West, K. D. (1987). A specification test for speculative bubbles. *The Quarterly Journal of Economics*, 102(3):553–580.

White, H. (1994). *Estimation, inference and specification analysis*. Cambridge University Press, Cambridge.