A time-varying parameter model for local explosions
Blasques, Francisco; Koopman, Siem Jan; Nientker, Marc

published in
Journal of Econometrics
2022

DOI (link to publisher)
10.1016/j.jeconom.2021.05.008

document version
Publisher's PDF, also known as Version of record
document license
Article 25fa Dutch Copyright Act

Link to publication in VU Research Portal

citation for published version (APA)

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 15. Sep. 2023
A time-varying parameter model for local explosions

Francisco Blasques a,b, Siem Jan Koopman a,b,c,*, Marc Nientker a

a Vrije Universiteit Amsterdam, School of Business and Economics, The Netherlands
b Tinbergen Institute Amsterdam, The Netherlands
c CREATES, Aarhus University, Denmark

Abstract

Financial and economic time series can feature locally explosive behaviour when bubbles are formed. We develop a time-varying parameter model that is capable of describing this behaviour in time series data. Our proposed dynamic 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 behaviour. 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 in the model, the implied filter is capable of extracting the unobserved bubble process from the 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 existing noncausal models in a financial application concerning the Bitcoin/US dollar exchange rate.

1. Introduction

Many financial and economic time series display phases of locally explosive behaviour that are followed by a burst or sharp mean reverting dynamics. This stochastic locally explosive behaviour is especially prevalent in financial asset prices, stock indices and exchange rates. The literature on rational expectations models for asset pricing describes 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) and West (1987). The bubble is considered to be an explosive nonstationary process, which motivated (Diba and Grossman, 1988) to test for its presence 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 these unit root tests to detect the existence of bubbles. Phillips et al. (2011) (PWY) introduced a recursive test procedure that performs a right-side unit root test coupled with a sup test over subsamples of the data. Their method allows for the identification of exploding subsamples of the data and dating both the origin and collapse of bubbles. Their method has identified explosive behaviour in empirical time series such as the Nasdaq real price index, the U.S. house price index, the price of crude oil and the spread between Baa and Aaa bond rates (Phillips et al., 2011; Phillips and Yu, 2011). Other date stamping tools that can be used come...
from tests for structural breaks such as Chow-tests and model selection tests. However, Homm and Breitung (2012) find that the PWY procedure performs well in the class of recursive procedures, especially as a real-time bubble detection algorithm. Phillips et al. (2015) (PSY) find that the original recursive procedure suffers from potential inconsistency and reduced power if a time series exhibits multiple speculative bubbles, which tends to happen for long or very volatile time series. Therefore PSY suggested a generalisation that allows for more flexibility in the subsamples chosen for the combined unit-root/sup test. Their method has been widely applied to a variety of markets such as commodities (Etienne et al., 2014; Gutierrez, 2013) and real estate (Chen and Funke, 2013; Yiu and Jin, 2012). Phillips and Shi (2018) note that the PSY method has a delay bias in identifying bubble emergence and collapse dates and introduce a reverse regression method that reduces this bias. Recent empirical applications and variations of the PSY method can be found on ballooning sovereign risk (Phillips and Shi, 2019), sector trading in real time (Milunovich et al., 2019) and the U.S. regional housing market (Shi, 2017). An R implementation of the method can be found in Phillips and Shi (2020) using the psymonitor R package Caspi et al. (2018), including a bootstrap scheme that takes heteroskedasticity and multiplicity in recursive testing into account.

A different approach has been proposed by Gouriéroux and Zakoian (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 Zakoian (2017) and to higher order mixed causal and noncausal linear models in Fries and Zakoian (2017). A difference with the rational expectations approach is that noncausal models work within a stationary framework, which allows for the derivation of theoretical results such as expected bubble life times and emergence and collapse probabilities. As a motivation for their stationary framework Gouriéroux and Zakoian (2013) show that the sample autocorrelation converges to a number smaller than one in absolute value. This demonstrates that a unit root test generally rejects the unit root hypothesis and thus will be unable to identify the presence of speculative bubbles. A 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 Zakoian (2017). The prediction of these components depends on computational methods such as Metropolis–Hastings or sampling/importance resampling, see Gouriéroux and Jasjak (2016). Additionally, the models are unable to distinguish the potential speculative bubble from the fundamental value.

This paper introduces an observation driven model with time varying parameters 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. The advantage of using such a specification is that we can filter data into its fundamental value and a potential speculative bubble. This means that we can identify bubble presence, emergence and collapse as in the recursive testing procedure methods, but also explain what fraction of observed exponential increase is due to noise in the underlying fundamental value process and what fraction is represented by the speculative bubble. Similar to the noncausal literature our model describes locally explosive behaviour in a strictly stationary framework. 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. This further implies that quantities of interest such as point predictions, confidence intervals, bubble emergence and collapse probabilities, expected bubble size and life times, and more, can be derived in a relative straightforward fashion. Another advantage of our method is that the sum of the fundamental value- and bubble processes is very flexible due to the joint dynamics of the individual components. Therefore it can describe multiple kinds of bubble behaviour for a fixed set of parameters.

As a result of earlier work in Blasques and Nientker (2017) we can show that our model admits a stationary ergodic and φ-mixing solution under very mild conditions. Additionally, in this paper, we provide new results on the model as a filter by proving that this 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) and Straumann (2005) 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.

We recognise that most of the existing literature on bubble detection and date stamping in rational expectation models are cast in a random walk setting together with an explosive bubble component. This is a shortcoming in the existing noncausal literature for bubble models (Gouriéroux and Jasiak, 2016; Gouriéroux and Zakoian, 2017; Fries and Zakoian, 2017), as well as the literature on observation driven filters in general. In our application we work around this shortcoming by detrending the data before starting the analysis, much like in Hencic and Gouriéroux (2015). We then find that the results of the PSY methodology test and our approach identify complement each other very well. Our model allows us to filter the bubble component, while the test in Phillips and Shi (2020) confirms that bubbles are present at nearly the identical times at which our filter identifies a bubble component in the data.

The remainder of the paper is organised as follows. Section 2 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

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

where $\mu_t$ is the fundamental value of the asset price, $b_t$ is the value of a potential speculative bubble and $\varepsilon_t$ is an error term that satisfies

$$(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{IID}(0, \sigma^2),$$  \hfill (2.2)

where $\sigma$ is a strictly positive constant. The fundamental value $\mu_t$ 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

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

where $\delta, \beta$ and $\gamma$ are fixed unknown parameters. The dynamics for the fundamental value are mean reverting if $|\beta| < 1$, but partially correct by a factor $\gamma$ 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 $\mu$ and $\varepsilon$ processes, similar to a stationary autoregressive moving average (ARMA) process. Also, we can adopt a completely exogenous stationary process for $\mu_t$ which is potentially based on economic or financial reasoning. We will maintain the stationary framework explored in the noncausal literature. In practice, this means that one would have to add a nonstationary component when the objective is to model bubbles in non-stationary time-series such as asset prices. 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

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

To ensure nonnegativity of $b_t$ we impose $\omega > 0$, while $\alpha$ 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, we stress that many options are available for the survival condition to keep the model specification flexible. 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 threshold functions

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

where $g$ is some real-valued function, which we will call the survival function.

2.1. Choice of survival function

When choosing a survival function we have to think about three important properties: what dynamics cause a bubble to collapse, what dynamics let a bubble grow and what dynamics cause a bubble to emerge. We typically will not need very complex survival functions, because the interplay between the asset price, fundamental value process and bubble process together allows for a lot of flexibility, as we will show in Section 2.2. Some intuitive example choices for the survival function that we have tried 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 remainder of Section 2.1 will study examples E1–E4 in a detailed fashion with the purpose of helping an interested reader construct their own survival function. We summarise the discussion by noting that in the range of these four examples we obtained the best results with survival functions E3 and E4. The choice between these two depends on the amount of available data and the number of bubbles it seems to contain. The more data there is available, the more parameters one can try to include to capture extra dynamics. In the case of long time series one can also decide to use $\ell > 1$ lags by generalising to survival functions that depend on $F_{t-1}, \ldots, F_{t-\ell}$. For a final decision between specifications we suggest using information criteria.

Example **E1** is the simplest survival function that we could think of and can also be read as the survival condition $X_{t-1} < c$. This survival condition specifies that an emerged bubble in the asset price is allowed to grow as long as the
asset price stays below a fixed level $c$ and crashes once it crosses this threshold. Note that this one parameter survival function already allows for multiple bubbles of various sizes in a given time series, because $X_{t-1}^{-}$ not only depends on $b_{t-1}$ but also on the fundamental value $\mu_{t-1}$ and the shock process $\epsilon_{t-1}$. Example E1, however, does imply that the asset price always drops around the same threshold $c$. We can allow more flexibility in the bubble collapse criteria by introducing more interplay between the different processes. In Example E2 we illustrate this idea by extending example E1 to the survival condition $X_{t-1}^{-} - \mu_{t-1} < c$. The intuition here is that a bubble is allowed to grow as long as the distance between the asset price and its fundamental value stays within a ‘reasonable’ range $c$ and collapses once this threshold is passed. Note that this example does not specify a fixed level for the asset price at a collapsing moment.

We have argued that examples E1 and E2 have reasonable flexibility in their collapse and survival behaviour. However, they have less control over the third main property, bubble emergence. To understand this we focus on the dynamics at time $t$ conditional on the fact that a bubble collapsed at time $t-1$, which means that $b_{t-1} = 0$. In example E1 if a bubble just collapsed, then the expected asset price $E(X_{t-1}^{-} | b_{t-1} = 0)$ is equal to its fundamental value in expectation, which means that $g(F_{t-1}) = X_{t-1} - c$ is likely to be negative. This will typically immediately initiate a new bubble emergence, which is something that we might want to have more control over. Similarly, if bubbles in example E2 are very large, then $c$ will be relatively large with respect to the dynamics of the fundamental value. A bubble collapse at time $t-1$ implies that $g(F_{t-1}) = \epsilon_{t-1} - c$, which again is likely to be negative and thus implies a large probability of a new bubble being created.

To gain more control of bubble emergence dynamics one can opt for survival functions that can collapse regardless of the size or the dynamics of the bubble. Example E3 is an illustration of such a survival function, where a bubble is allowed to grow as long as it makes out less than a fraction $k$ of the total asset price. Similarly to the previous examples this specification allows for various bubble sizes and critical levels of the asset price as $X_{t-1}^{-}$ also 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. Different from the previous examples E3 can control for the emergence of bubbles. Note that $X_{t-1}^{-} < 0$ implies that $g(F_{t-1}) < 0$, which in turn means that the survival condition is not satisfied for any possible value of $b_{t-1}$ and thus that no bubble will emerge at time $t$. 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.

Example E4 is constructed to, similarly as in E3, have more control over bubble emergence probability, but allows for more flexibility by allowing more interplay 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. The bubble is allowed to grow as long as the bubble size stays below a multiple $k$ of the difference between the fundamental value $\mu_{t}$ and the threshold $c$. 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 so that $\mu_{t} - c > 0$. As the fundamental value is mean reverting while the bubble is exponentially increasing, the bubble process grows much faster and thus in this case we eventually observe that $b_{t-1} \geq k(\mu_{t} - c)$ causing the bubble to burst. Secondly, suppose that the fundamental value is below the threshold $\mu_{t} < c$. Then $\mu_{t} - c$ is negative, so that $g(F_{t-1}) > 0$ causes the bubble to burst regardless of the value of $b_{t-1}$. This second collapse condition also implies that no new bubble will emerge at the next time period and thus example E4 can also describe data with long periods of no explosive behaviour. Combinations between the different collapse mechanisms are also possible, allowing for a wide variety of bubble sizes and overall asset price dynamics.

2.2. Bubble variety

The bubble process described in (2.4) might appear to be rather restricted at first sight, as the bubble conditional on its previous value $b_{t-1}$ allows for only two possible values: 0 and $\omega + \alpha b_{t-1}$. Moreover, all of these are within the countable space $\{0, \omega, \omega(1+\alpha), \ldots\}$. 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 Fig. 1 by examining some of the possible impulse response functions (IRFs) for the model described in Eqs. (2.1)-(2.5) with survival condition E4.

Fig. 1(a) illustrates how a small impulse that does not push 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 Fig. 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 Fig. 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, Fig. 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 Fig. 1 are often encountered in financial time series. Fig. 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. Fig. 2(a) plots the monthly Nasdaq real price from January 1973
Fig. 1. Several impulse response functions for the bubble model as described in Eqs. (2.1)–(2.5) with survival condition $E_4$.

Fig. 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.

2.3. Shortcomings and extensions

Although examples $E_1$–$E_4$ can already capture some bubble dynamics well, as illustrated in the rest of this paper, we realise that there are some shortcomings too. In this section we discuss some of these shortcomings and propose possible solutions. The application of these extensions is outside of the scope of this paper.

First of all, the bubble process as specified in (2.4) allows for only one speed of increase for all bubbles in a considered time series. Similar to Gouriéroux and Zakoïan (2017) we can overcome this problem by allowing for an aggregation of
bubble processes. Let \((b_{t,j});_{t\in\mathbb{Z}}\) be \(J\) bubble processes as defined in (2.4) and (2.5). Then we can define an aggregate bubble process as
\[
 b_t = \sum_{j=1}^{J} \pi_j b_{t,j},
\]
\[
b_{t,j} = (\omega_j + \alpha_j b_{t-1,j}) I\{g_j(X_{t-1}, \mu_{t-1}, b_{t-1,j}) < 0\} \quad \text{for } j = 1, \ldots, J.
\]

Here the \(\pi_j\)'s represent probabilities: \(0 \leq \pi_j \leq 1\) and \(\sum_{j=1}^{J} \pi_j = 1\). In this specification each bubble process has its own rate of increase due to its own parameters \(\alpha_j\) and \(\omega_j\). Moreover the survival function \(g_j\) can have bubble specific parameters and bubble collapse and emergence behaviour that depend on its own past value \(b_{t-1,j}\). Showing that this process is stationary ergodic is similar to the approach taken in Theorem 1, although estimation becomes more involved due to the unobserved probabilities \(\pi_j\).

Another potential problem is the fact that the bubble process collapses in a single time period, something that is not the case in many empirical time series. An extension of the model that could deal with this is to specify an additional regime for the bubble process in which it decreases at an exponential rate. The bubble process would then look like
\[
b_t = \begin{cases} 
 \omega + \alpha b_{t-1} & \text{if } g(F_t-1 < 0) \leftarrow \text{increasing regime} \\
 \alpha_d b_{t-1} & \text{if } g(F_t-1 > 0) \leftarrow \text{decreasing regime} \\
 0 & \text{if } b_{t-1} < \omega \leftarrow \text{collapsing regime}
\end{cases}
\]

The interpretation here is that \(\alpha_i > 1\) so that we have exponential increase during the increasing regime, while \(\alpha_d < 1\) indicates exponential decline during the decreasing regime. The collapsing regime ensures that the bubble can cease to exist if there is a continuing period of decrease. This third regime also ensures that the theoretical results of Section 3 can be rather directly extended for this generalisation.

3. Probabilistic and statistical analysis

In this section we study the probabilistic properties of our model as defined in Eqs. (2.1)–(2.5). The bubble model contains several irregular components making the results in this section nonstandard. Firstly, the parameter \(\alpha\) 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 Straumann (2005). Markov chain theory as developed in Meyn and Tweedie (1993) is able to deal with discontinuities. However, this theory cannot be applied when we study the model as a filter, because the filter does not satisfy the Markov property as the asset price \((X_t)_{t\in\mathbb{Z}}\) is not an independent sequence. Therefore 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 \((\theta, \lambda)\) which belong to the parameter space \(\Theta \times \Lambda\). Here \(\theta\) contains all the parameters that enter continuously in \((\mu_t, b_t)\) and \(\lambda\) contains the remaining parameters. Among the parameters \((\sigma^2, \delta, \beta, \gamma, \omega, \alpha)\), it is clear that \(\sigma^2\) is always an element of \(\theta\), while the remaining parameters may be elements of \(\theta\) or \(\lambda\) depending on the chosen survival function \(g\). For a chosen function \(g\) one can quickly determine which parameters belong to \(\theta\) and which parameters belong to \(\lambda\). If the survival function depends on the past bubble process \(b_{t-1}\), then all the parameters in (2.4), that is \((\alpha, \omega)\), belong to \(\lambda\). If \(g\) is nonconstant in the fundamental value process \(\mu_{t-1}\), then all the parameters in (2.3), that is \((\delta, \beta, \gamma)\), belong to \(\lambda\). Any parameters that are inside of \(g\) belong to \(\lambda\). We examine the examples from Section 2 as an illustration.

\begin{align*}
E1 & \quad g(F_{t-1}) = X_{t-1} - c, \quad \text{then } \theta = (\sigma^2, \delta, \beta, \gamma, \omega, \alpha) \text{ and } \lambda = c. \\
E2 & \quad g(F_{t-1}) = X_{t-1} - \mu_{t-1} - c, \quad \text{then } \theta = (\sigma^2, \omega, \alpha) \text{ and } \lambda = (\delta, \beta, \gamma, c). \\
E3 & \quad g(F_{t-1}) = b_{t-1} - kX_{t-1}, \quad \text{then } \theta = (\sigma^2, \delta, \beta, \gamma) \text{ and } \lambda = (\omega, \alpha, k). \\
E4 & \quad g(F_{t-1}) = b_{t-1} - k(\mu_{t-1} - c), \quad \text{then } \theta = \sigma^2 \text{ and } \lambda = (\delta, \beta, \gamma, \omega, \alpha, k).
\end{align*}

Deriving consistency and asymptotic normality for \(\lambda\) is generally difficult. Therefore we approach the problem by deriving these results for \(\theta\) conditionally on a calibrated value of \(\lambda\). This means that we will work with functions \(f\) in the Banach space \(L^\infty(\Theta, \mathbb{R})\), where we write \(\|f\|_{\Theta}\) for the supremum norm. The parameter space \(\Theta\) 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 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. All the proofs can be found in the Appendix.
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 Eq. (2.1) holds for such data and thus (2.3) simplifies to

$$
\mu_t = \delta + \beta \mu_{t-1} + \gamma \epsilon_{t-1}.
$$

(3.1)
The fundamental value process therefore is an autoregressive process of order one, a specification that is well studied and known to have stable solutions. We write $\log^+(x) = \max(0, \log x)$ and need the following assumptions to ensure the stability results:

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

**DGP 2.** The error $\epsilon_t$ is a continuous random variable with full support and has a finite log moment $\mathbb{E} \log^+ |\epsilon_t| < \infty$.

**DGP 3.** Let $b \geq 0$ and $\mu, \epsilon \in \mathbb{R}$. There exists a set $S \subset \mathbb{R} \times \mathbb{R}$ of positive Lebesgue measure such that the survival function satisfies

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

Assumption DGP 1 is standard in the literature on autoregressive processes. The moment assumption in DGP 2 is also standard. The full support part and Assumption DGP 3 seem complicated but essentially require 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 3 is usually easy to verify.

**Lemma.** Assumption **DGP 3** 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 Lebesgue measure. ■

We verify condition DGP 3 on our examples **E1-E4** as an illustration. If our survival function is given by **E1** then $\tilde{g}(\epsilon, \mu) = \mu + \epsilon - c$, if our survival function is given by **E2** then $\tilde{g}(\epsilon, \mu) = \epsilon - c$, if our survival function is given by **E3** then $\tilde{g}(\epsilon, \mu) = k(\mu + \epsilon)$ as $k \leq 1$ and finally if our survival function is given by **E4** then $\tilde{g}(\epsilon, \mu) = -k(\delta + \beta \mu + \gamma \epsilon - c)$. All of these functions are continuous and surjective and thus condition DGP 3 is satisfied for all our examples. The following theorem now quickly follows from previous work in Blasques and Nientker (2017).

**Theorem 1.** Suppose that assumptions **DGP 1–3** 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

$$
\| (\mu_t, b_t) - (\hat{\mu}_t, \hat{b}_t) \|_{\infty} \xrightarrow{\text{e.p.}} 0 \quad \text{as} \quad t \to \infty.
$$

Theorem 5 establishes the existence of a unique causal solution to our model for each choice of $\theta$ that satisfies assumptions DGP 1–3. 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 finalise 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 1–3** hold, that the survival function is given as in **E4** and that $\epsilon_t$ has an $n$th moment $\mathbb{E}|\epsilon_t|^n < \infty$. Then the unique stationary ergodic solution has a uniform $n$th moment, i.e. $\mathbb{E}\| (X_t, \mu_t, b_t) \|_\infty^n < \infty$.

3.2. The model as a filter

This section focuses 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 $\mu_0$ and $b_0$, 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 front set of conditions assumes structure on the dependence between the $(X_t)$. We write $\mathcal{F}_s^t$ for the $\sigma$-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 Lebesgue 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.

Proposition 3. Suppose \((X_t)_{t \in \mathbb{Z}}\) is a real valued stationary ergodic solution of a Markov chain \(X_t = f(X_{t-1}, \xi_t)\). 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 Lebesgue measure, and \((\xi_t)_{t \in \mathbb{Z}}\) is a sequence of independent, identically distributed and absolutely continuous random variables such that \(f(x, \xi_t)\) 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 \(p(x) = 1 - rx + \gamma ax Lx^2\) has roots outside of the unit circle.

Assumption 4 contains quite some restrictions on the survival function. It can be easily checked however that these all hold for example E4. We are now ready to prove our main result for the model as a filter.

Theorem 4. Suppose that assumptions FLT 1–5 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 \(\Theta\). Moreover, any other solution \((\hat{\mu}_t, \hat{b}_t)_{t \in \mathbb{N}}\) almost surely converges exponentially fast to the stationary ergodic one, that is,
\[
\left\| (\mu^*_t, b^*_t) - (\hat{\mu}_t, \hat{b}_t) \right\|_{\Theta} \xrightarrow{\text{a.s.}} 0 \text{ as } t \to \infty.
\]

Finally, if \(X_t\) has an \(n\)'th moment for some \(n \in \mathbb{N}\), then \(\mathbb{E}\left\| (\mu^*_t, b^*_t) \right\|^n_{\Theta} < \infty\) too.

3.3. The likelihood

To define the likelihood for our model we will assume that our errors have a Gaussian distribution
\[
\varepsilon_t \sim N(0, \sigma^2).
\] (3.2)

We recognise that many other distributions such as the \(t\) or \(\alpha\)-stable distributions can be chosen instead of the Gaussian. However, we need to make some distributional assumptions in order to define a likelihood. If normality fails, then the results could potentially be interpreted as quasi-ML results, although that would require additional conditions.

As mentioned in the beginning of Section 3 we derive our asymptotic results for \(\theta\) conditionally on a calibrated value of \(\lambda\). 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) and the Gaussian assumption. The fundamental value and bubble processes are unobserved, so we choose initialised values \(\hat{\mu}_1\) and \(\hat{b}_1\) which deliver filtered sequences \((\hat{\mu}_i(\theta, \lambda))_{i=2}^{T}\) and \((\hat{b}_i(\theta, \lambda))_{i=2}^{T}\) according to (2.3)–(2.5). It follows that the distribution of \(X_t\) conditional on its past is given by
\[
X_t \mid X_1, \ldots, X_{t-1} = X_t \mid \hat{\mu}_i(\theta, \lambda), \hat{b}_i(\theta, \lambda) \sim N(\hat{\mu}_i(\theta, \lambda) + \hat{b}_i(\theta, \lambda), \sigma^2)
\]
and thus prediction error decomposition delivers the average log likelihood as a function $\Theta \rightarrow \mathbb{R}$ given by
\[
\hat{L}_T(\theta) \propto \frac{1}{T} \sum_{t=2}^T \ell(X_t, \hat{\mu}_t(\theta, \lambda), \hat{b}_t(\theta, \lambda), \sigma^2),
\]
\[
\ell(X_t, \hat{\mu}_t(\theta, \lambda), \hat{b}_t(\theta, \lambda), \sigma^2) := -\frac{1}{2} \log(2\pi \sigma^2) - \frac{1}{2\sigma^2} (X_t - \hat{\mu}_t(\theta, \lambda) - \hat{b}_t(\theta, \lambda))^2.
\]

To ease notation in the following text we define the functions $\Theta \rightarrow \mathbb{R}$:
\[
\hat{\epsilon}_t = \ell(X_t, \hat{\mu}_t(\cdot, \lambda), \hat{b}_t(\cdot, \lambda), \sigma^2), \quad \epsilon^*_t = \ell(X_t, \mu^*_t(\cdot, \lambda), b^*_t(\cdot, \lambda), \sigma^2).
\]
Hereafter, we stress no longer that $\hat{\mu}_t$, $\hat{b}_t$, $\mu^*_t$ and $b^*_t$ depend on $\theta$ and $\lambda$, as long as it causes no confusion.

### 3.4. Asymptotic results

We will provide sufficient conditions for the consistency and asymptotic normality of the ML estimator. Each result is stated once under an assumption of correct specification, i.e. $(X_t)_{t\in\mathbb{Z}}$ satisfies (2.1)–(2.5) plus (3.2), and a misspecified model assumption.

#### 3.4.1. Consistency

The ML estimator of $\theta$ 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 $((\hat{\mu}_t, \hat{b}_t))_{t\in\mathbb{Z}}$ converges to a limit process $((\mu^*_t, b^*_t))_{t\in\mathbb{Z}}$ uniformly over $\Theta$ with two uniform bounded moments. That is,
\[
\left\| (\hat{\mu}_t, \hat{b}_t) - (\mu^*_t, b^*_t) \right\|_{\Theta} \xrightarrow{\text{ess}} 0 \quad \text{as } t \rightarrow \infty \quad \text{and} \quad \mathbb{E} \left\| (\mu^*_t, b^*_t) \right\|_{\Theta}^2 < \infty.
\]
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 maximiser $\theta_0$ of the limit log likelihood, that is, for every $\theta \in \Theta$ that is unequal to $\theta_0$ we have
\[
\mathbb{E} \ell(X_t, \mu^*_t(\theta, \lambda), b^*_t(\theta, \lambda), \sigma^2) < \mathbb{E} \ell(X_t, \mu^*_t(\theta_0, \lambda), b^*_t(\theta_0, \lambda), \sigma^2_0).
\]

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. The invertibility in **CS 2** refers to the fact that Theorem 4 can be used to show that $(\mu^*_t, b^*_t)$ can be written as a function of past observations $X_{t-1}, X_{t-2}, \ldots$ Therefore $\epsilon_t = X_t - \mu^*_t - b^*_t$ can be written as a function of past and present observations $X_t, X_{t-1}, X_{t-2}, \ldots$ which is similar to the notion of invertibility in ARMA models, see Straumann and Mikosch (2006, Section 3.2) for a detailed derivation of these results. Assumption **CS 3** ensures that the limit log likelihood is maximised at a unique point $\theta_0$, given the fixed parameter $\lambda$. Note that the expectations exist by the moment assumptions in **CS 1–2**. When the model is assumed to be well specified and $\lambda$ is fixed at its true value $\lambda_0$, then we will show in Corollary 6 that this assumption holds for survival function $E4$ under a standard identifiability assumption. Moreover, the parameter of interest $\theta_0$ 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 $\lambda$ is set at some arbitrary value $\lambda \neq \lambda_0$, then the uniqueness of the parameter of interest $\theta_0$ is harder to establish.\(^1\) In this case, the limit parameter $\theta_0$ is a ‘pseudo-true parameter’, i.e. a parameter that minimises a Kullback–Leibler divergence between the true conditional density of the data and the model-implied conditional density, see White (1994, Section 2.3).

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

Theorem 1 establishes the a.s. convergence of the ML estimator $\hat{\theta}_T$ to the pseudo-true parameter $\theta_0$ which is the unique maximiser of the limit log likelihood for any given value of $\lambda \in \Lambda$. In this sense $\theta_0$ provides the best Kullback–Leibler approximation to the true unknown distribution of the data, for the given value of $\lambda$. If the model is correctly specified and $\lambda$ is calibrated at its true value, then $\theta_0$ corresponds to the true parameter under an additional standard identifiability assumption.

\(^1\) 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.
**Corollary 6.** Suppose that the model is correctly specified, i.e. \((X_t)_{t \in \mathbb{Z}}\) satisfies (2.1)–(2.5) plus (3.2), and that \(\lambda\) is calibrated at its true value. Assume moreover that the survival function \(g\) is given by example E4, assumptions DGP 1–2 and assumption FLT 5 hold and that the following identifiability condition holds:

\[
\ell^*_t(\theta) = \ell^*_1(\theta) \quad \text{a.s. if and only if} \quad \theta = \theta_0.
\]

Then the ML estimator \(\hat{\theta}_T\) converges a.s. to the true parameter \(\theta_0\).

**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 \(\theta_0\) belongs to the interior of \(\Theta\).

**AN 2.** The limit process \((\mu^*_t, b^*_t)\) is twice continuously differentiable on \(\Theta\) 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 \(\{ \frac{\partial \mu^*_t}{\partial \theta} \}\) and \(\{ \frac{\partial b^*_t}{\partial \theta} \}\). In particular, we have again that \(\{ \frac{\partial \mu^*_t}{\partial \theta} \}\) converges at any rate since it is reset to zero in a finite number of steps with positive probability, and \(\{ \frac{\partial b^*_t}{\partial \theta} \}\) converges exponentially fast due to its autoregressive nature.

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

\[
\left\| (\nabla^{0,2}_t \mu^*_t, \nabla^{0,2}_t b^*_t) - (\nabla^{0,2}_0 \mu^*_t, \nabla^{0,2}_0 b^*_t) \right\|_{\Theta} \overset{a.s.}{\to} 0 \quad \text{as} \quad T \to \infty,
\]

and

\[
\mathbb{E} \left\| (\nabla^{0,2}_t \mu^*_t, \nabla^{0,2}_t b^*_t) \right\|^4_{\Theta} < \infty.
\]

**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

\[
\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(\mathbf{0}, I^{-1}(\theta_0)) \quad \text{as} \quad T \to \infty,
\]

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

We further impose a limited memory assumption on the data to obtain the asymptotic normality of the MLE under misspecification. In particular, we assume that the data is near epoch dependent on a mixing sequence of appropriate size. The sizes are selected to ensure that the score satisfies a central limit theorem; see e.g. Chapter 6 of Pötscher and Prucha (1997) or chapters 17 and 24 of Davidson (1994). As noted in Pötscher and Prucha (1997), unlike other fading memory concepts (such as mixing), the concept of near epoch dependence is well suited for nonlinear dynamic models. Additionally, it allows us to obtain enough fading memory for a central limit theorem to hold, even if the score fails to be a martingale difference sequence due to the potential dynamic misspecification of the model.

**AN 6.** \((x_t)\) is near epoch dependent of size \(-1\) on a \(\phi\)-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

\[
\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(\mathbf{0}, \Sigma(\theta_0, \lambda)) \quad \text{as} \quad T \to \infty,
\]

where

\[
\Sigma(\theta_0, \lambda) = \left(\mathbb{E} \hat{L}''_t(\theta_0, \lambda)\right)^{-1}\left(\mathbb{E} \hat{L}'_t(\theta_0, \lambda)\mathbb{E} \hat{L}'_t(\theta_0, \lambda)\right)^{-1}\left(\mathbb{E} \hat{L}'_t(\theta_0, \lambda)\right)^{-1}.
\]

**4. Illustrations**

In this section we test the descriptive capability of our model and illustrate the accessibility 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. In keeping with the theory of Section 3, we treat the parameter vector \(\lambda\), which enters discontinuously...
Table 1  
Parametrisation used for simulation study.

<table>
<thead>
<tr>
<th>σ</th>
<th>δ</th>
<th>β</th>
<th>γ</th>
<th>ω</th>
<th>α</th>
<th>k</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.1</td>
<td>0.95</td>
<td>0.7</td>
<td>0.2</td>
<td>1.03</td>
<td>7</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

Table 2  
Parametrisation used for simulation study.

<table>
<thead>
<tr>
<th>σ</th>
<th>δ</th>
<th>β</th>
<th>γ</th>
<th>ω</th>
<th>α</th>
<th>k</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.05</td>
<td>-0.25</td>
<td>0.86</td>
<td>1.04</td>
<td>0.44</td>
<td>2.10</td>
<td>2.39</td>
<td>12.29</td>
</tr>
</tbody>
</table>

Fig. 3. An example simulated path for the parametrisation of the model described in Table 1.

into the likelihood function, as fixed. For each given $\lambda$, we effectively have a model parameterised by $\theta$, and we estimate the vector $\theta$ by maximum likelihood to obtain $\hat{\theta}$. Finally, we use standard model selection procedures to select the best possible model; i.e. the best possible pair $(\hat{\theta}, \hat{\lambda})$. We label the resulting pair $(\hat{\theta}, \hat{\lambda})$. Section 3 covers the MLE theory even when $\lambda$ is set incorrectly, since we allow for model misspecification in Theorems 5 and 8.

We use the simulation study in Section 4.1 to examine the distribution of the resulting $(\hat{\theta}, \hat{\lambda})$ using 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 estimator behind $\hat{\lambda}$ for a given parametrisation, stated in Table 1. This choice represents a medium amount of bubbles of size relative to the fundamental value process. A typical simulation for the implied model can be found in Fig. 3. Note that there are windows 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.

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 model comparison estimator are portrayed in Fig. 4. We observe 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 parametrisation 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

In our application we study the daily Bitcoin/US dollar exchange rate from February 20 to July 18 in 2013. We note that this is a nonstationary time series, which motivates us to detrend the time series. Our approach is equivalent to the one used in Hencic and Gouriéroux (2015) and the resulting time series can be found in Fig. 5. There appears to be a big bubble that collapses on April 9, 2013. Moreover, it is potentially followed by a second smaller bubble. Afterwards it tends to behave as a standard stable and mean reverting process. Hencic and Gouriéroux (2015) have found evidence for the presence of a speculative bubble in this time series. In this paper we perform a PSY type recursive testing procedure to review this statement. We performed the bootstrap test from Phillips and Shi (2020) using the psymonitor package in R. The bubble identification results can be found in Fig. 6. We can conclude that the bootstrap test confirms the findings of Hencic and Gouriéroux (2015) by identifying the presence of a speculative bubble that emerges on April 3, 2013 and collapses on April 9, 2013. We find it hard to explain the additional identification of a single day bubble on May 28, 2013. Nevertheless the presence of the bubble in April motivates us to estimate our model. The results are given in Table 2.
The estimate of $\alpha$ 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 Fig. 7, 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 fits the in sample burst of the bubble on April 9 well. It does however underestimate the additional decrease in fundamental value after the bubble burst.

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. Fig. 8 plots the filtered time series and these probabilities for some period centred around the bubble.

Here we see that the probability of nonzero bubble values before the bubble start is 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.

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 as developed in this study. 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 such a model would then be very close to those described in the mixed causal and noncausal literature.

Appendix. 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 Eqs. (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

\[
\|\mu_t - \hat{\mu}_t\|_{\mathcal{F}} \xrightarrow{a.s.} 0 \quad \text{as} \quad t \to \infty.
\]

This unique sequence \((\mu_t)_{t \in \mathbb{Z}}\) is given by a geometric sum of past independent errors:

\[
\mu_t = \sum_{i=0}^{\infty} \beta \epsilon_{t-i-1}.
\]

Therefore DGP 2 implies that \(\mu_t\) is a continuous random variable with full support.

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

\[
b_{t+1} = \phi_t(b_t), \quad \text{where} \quad \phi_t(b) = (\omega + \alpha b) \mathbb{I}_{\{g(\mu_t + b + \epsilon_t, \mu_t, b) < 0\}}.
\]

We then check Assumption A for Theorem 2.1 in Blasques and Nientker (2017). Assumption A1 is trivial and A2 is satisfied by Krenchel’s Lemma: a measurable function of a stationary ergodic sequence produces a stationary ergodic sequence, see Proposition 4.3 in Krenchel (1985). For the final assumption A3 we note that \(\phi_t(b) = 0\) for all \(b \in [0, \infty)\), if

\[
g(\mu_t + b + \epsilon_t, \mu_t, b) \geq \inf_{b \leq 0} g(\mu_t + b + \epsilon_t, \mu_t, b) \geq 0.
\]
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 3, because Assumption DGP 2 implies that \( \mu_t \) and \( \varepsilon_t \) are independent and both absolutely continuous on the real line. Therefore the joint random variable \( (\mu_t, \varepsilon_t) \) 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 \( (\hat{b}_t)_{t \in \mathbb{Z}} \) that satisfies the bubble updating process defined in (2.1)-(2.5) and that any other solution \( (\tilde{b}_t)_{t \in \mathbb{Z}} \) initialised at \( \tilde{b}_1 \) satisfies
\[
\left\| b_t - \hat{b}_t \right\|_{\mathbb{E}} \to 0 \quad \text{as} \quad t \to \infty.
\]
The final conclusion follows again by Krenkel’s Lemma.

A.2. Proof of Corollary 2

The solution for the fundamental process found in Theorem 1 is given by
\[
\mu_t = \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-i-1},
\]
which is well known to have an \( n' \)th moment over \( \Theta \) as it is a compact space so that \( |\beta| \) is bounded and the \( \varepsilon_t \) have an \( n' \)th moment. For the bubble process we have
\[
\tilde{b}_t = (\omega + \alpha b_{t-1})1[b_{t-1} < k(\mu_t - c)] \leq \omega + \max(\alpha k(\mu_t - c), 0)
\]
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 \( \zeta \) and that the inverse \( f^{-1}(x, \cdot) \) is also continuously differentiable. Moreover, by assumption, for Lebesgue 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 Lebesgue measures zero and let \( F \) denote the distribution function of \( X_t \), then by independence of the \( \zeta_t \) we have
\[
\mathbb{P}(X_t \in E) = \mathbb{P}(f(X_{t-1}, \zeta_t) \in E)
\]
\[
= \int_{\mathbb{R}} \mathbb{P}(f(x, \zeta_t) \in E) F(dx) = \sum_{k=1}^{\infty} \int_{\mathbb{R}} \mathbb{P}(f(x, \zeta_t) \in E \cap O_k) F(dx)
\]
\[
= \sum_{k=1}^{\infty} \int_{\mathbb{R}} \mathbb{P}(\zeta_t \in f^{-1}(x, E \cap O_k)) F(dx) = \sum_{k=1}^{\infty} \int_{\mathbb{R}} 0 F(dx) = 0,
\]
where we used that \( E \cap O_k \) has Lebesgue measure zero, \( f^{-1}(x, \cdot) \) has the Luzin property on each \( O_k \) and \( \zeta_t \) is absolutely continuous. The absolute continuity of the conditional distributions follows similarly as by the independence of the \( \zeta_t \) 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 \( \zeta_t \) is bounded by some \( B > 0 \). For some \( \xi \in \mathbb{R} \) and \( \eta > 0 \) we then have
\[
\mathbb{P}(\xi < X_t \leq \xi + \eta \mid X_{t-1} = x_1, \ldots, X_{t-n} = x_n) = \mathbb{P}(\xi < f(x_1, \zeta_t) \leq \xi + \eta)
\]
\[
= \sum_{k=1}^{\infty} \mathbb{P}(\zeta_t \in f^{-1}(x_1, (\xi, \xi + \eta) \cap O_k)) \leq \frac{B}{L} \eta,
\]
79
where we used that $f^{-1}(x, \cdot)$ is $\frac{1}{\varepsilon}$-Lipschitz on each $O_k$ and the density of $\zeta_t$ is bounded by $B$. Taking the limit of $\eta \to 0$ shows that the conditional densities are all bounded by $\frac{B}{\varepsilon}$.

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 $\zeta_t$ is absolutely continuous, and show the non exclusive property in FLT 2. This follows from the Markov chain setup. Let $A_0 \subseteq \mathbb{R}$ be all points $x$ such that $P(A \mid X_0 = x) > 0$. Then $A_0$ has positive Lebesgue measure as $X_0$ is absolutely continuous and

$$P(X_0 \in A_0) \geq P(A) > 0.$$ 

It follows that

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

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

where we used that $P(A \mid X_0 = x)$ is greater than zero on a set of positive Lebesgue measures and $P(B \mid X_0 = x)$ 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 $\zeta_t$ 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 $v \geq 0$ we have

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

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

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

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

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

and

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

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 2.13 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 $\phi_t$ 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 $\lim_{m \to \infty} b_0^{(m)}$ converges almost surely.

#### Proof

Define

$$s_t = \limsup_{m \to \infty} \mu_t^{(m)}$$

and

$$i_t = \liminf_{m \to \infty} \mu_t^{(m)}.$$ 

The proof that these backward iterate limits exist consists of two steps that we show below:

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

80
(ii) We show that there exists an event $B_\eta \in \mathcal{F}_{-\infty}^\infty$ that contains $A_\eta$, such that conditional on $B_\eta$ we have $s_t - i_t \leq r'(s_0 - i_0)$ 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 $i_t$ to the right occurs. If it occurs for such an $-i_t$, then

$$s_0 - i_0 \leq r^k(s_{-i_t} - i_{-i_t}) \leq r^k\eta.$$  

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 $(b_0^{(m)})_{m \in \mathbb{N}}$ is eventually constant follows immediately from part (ii) and the same argument that the event $B_\eta$ occurs for some $-i < 0$.  

**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

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

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 Lipschitz 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

$$b_{t-1}^{(m)} = (\omega + \alpha b_{t-1}^{(m-1)}) \gamma g^{-1}(X_{t-1}, \mu_{t-1}, t_{t-1}) < 0$$  

$$\frac{\omega + \alpha b_{t-1}^{(m-1)}}{\gamma g^{-1}(X_{t-1}, \mu_{t-1}, t_{t-1})} \leq \frac{\omega + \alpha}{\gamma} \max\left\{K + L(\mu_{t-1}^{(m-1)} + X_{t-1}), 0\right\}.  

It follows that

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

$$\geq (\delta - \gamma(\omega + \alpha K)) + r\mu_{t-1}^{(m-1)} - \gamma \alpha L \mu_{t-1}^{(m-2)} + \gamma (X_{t-1} - \alpha L X_{t-1}).$$  

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

Next, we choose an $M > 0$ such that $\mathbb{P}(s_t - i_t < M) > 0$ and let $t = \left\lceil \frac{\log u/m}{\log r}\right\rceil$, 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

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

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

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

This then implies that

$$s_0 - i_0 = r(s_t - i_{-1}) + \gamma \limsup_{m \to \infty} b_{v}^{(m)} = \liminf_{m \to \infty} b_{v}^{(m-1)} = r'(s_t - i_t) < r'M \leq \eta,$$  

which concludes the proof of part (i).  

**Lemma.** Claim (ii) holds.

**Proof.** The argument will be a recursive one, conditional on $A_\eta$. 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

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

(A.1)
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

$$|g(X_t, s_t, b_T) - g(X_r, i_r, b_T)| \leq K(s_t - i_r) < K r^T \eta.$$ 

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 $I_{2K r^T \eta}(s_t, b_T)$ 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 $X_t \in I_{2K r^T \eta}(s_t, b_T)$. Then by assumption FLT 3 there exists a $U > 0$ such that

$$\mathbb{P}(|g(X_t, s_t, b_T)| \leq B K r^T \eta | A_\eta) \leq \mathbb{P}(X_t \in I_{2K r^T \eta}(s_t, b_T) | A_\eta) = \int \mathbb{P}(X_t \in I_{2K r^T \eta}(s_t, b_T) | s_t, b_T, A_\eta) d\mathbb{P}(s_t, b_T) \leq \int 2 U K r^T \eta d\mathbb{P}(s_t, b_T) = 2 U K r^T \eta.$$ 

It follows that

$$\mathbb{P}(|g(X_t, s_t, b_T)| > B K r^T \eta, \forall t \in \mathbb{N} | A_\eta) = 1 - \mathbb{P}(|g(X_t, s_t, b_T)| \leq B K r^T \eta, \exists t \in \mathbb{N} | A_\eta) \geq 1 - \sum_{t=1}^{\infty} \mathbb{P}(|g(X_t, s_t, b_T)| \leq B K r^T \eta | A_\eta) \geq 1 - \sum_{t=1}^{\infty} 2 U K r^T \eta \geq 1 - \frac{2 U K r^T \eta}{1 - r}.$$ 

This last number can be made larger than zero by choosing $\eta$ sufficiently small. \hfill \blacksquare

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 $|\hat{\mu}_T|$ and $|\hat{\mu}_T|$ are bounded by some $\eta$ 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 it follows that 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).

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

$$\left\| \hat{\ell}_T - \ell^*_\ell \right\|_\vartheta \overset{a.s.}{\to} 0 \quad \text{as} \quad T \to \infty.$$ 

Proof. We have

$$\left\| \hat{\ell}_T - \ell^*_\ell \right\|_\vartheta = \left\| \frac{1}{T} \sum_{t=2}^{T} \hat{\ell}_t - \ell^*_\ell \right\|_\vartheta \leq \frac{1}{T} \sum_{t=2}^{T} \left\| \hat{\ell}_t - \ell^*_\ell \right\|_\vartheta + \frac{1}{T} \sum_{t=2}^{T} \ell^*_\ell - \ell^*_\ell \right\|_\vartheta \overset{(A.2)}{\to} 0.$$ 

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

$$\ell^*_\ell(\hat{\mu}, \hat{\sigma}) = \frac{\partial \ell(X_t, \mu, b, \sigma^2)}{\partial (\mu, b)} \Big|_{(\hat{\mu}, \hat{\sigma})}.$$
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 \((\mu^*_t, b^*_t)\) that satisfies
\[
\left\| \tilde{\ell}_t - \ell^*_t \right\|_\Theta \leq \left\| \ell_{(\hat{\mu}_t, \hat{b}_t)} - \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta \\
\leq \left\| \ell_{(\tilde{\mu}_t, \tilde{b}_t)} - \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta + \left\| \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta \left( \left\| \tilde{\mu}_t - \mu^*_t, \tilde{b}_t - b^*_t \right\|_\Theta \right)
\] (A.3)

The function \(\ell_t\) is linear in its arguments and thus is a K-Lipschitz function for some \(K > 0\). Therefore assumption CS 2 guarantees that
\[
\left\| \ell_{(\tilde{\mu}_t, \tilde{b}_t)} - \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta \leq K \left\| (\tilde{\mu}_t, \tilde{b}_t) - (\mu^*_t, b^*_t) \right\|_\Theta \\
\leq K \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu^*_t, b^*_t) \right\|_\Theta \xrightarrow{\text{a.s.}} 0 \quad \text{as } t \to \infty.
\]
hence (A.3) almost surely goes to zero exponentially fast by assumption CS 2 and thus we have
\[
\frac{1}{T} \sum_{t=2}^{T} \left\| \frac{1}{T} \sum_{\ell(t) = 2}^{T} \ell_{\ell(t)}(\mu^*_t, b^*_t) \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu^*_t, b^*_t) \right\|_\Theta \xrightarrow{\text{a.s.}} 0 \quad \text{as } T \to \infty.
\]

Next, note that \(\left( \left\| \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta \right)\) is a stationary sequence by assumption CS 2 and Proposition 4.3 in Krenkel (1985). Therefore
\[
\frac{1}{T} \sum_{t=2}^{T} \left\| \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta \left\| (\hat{\mu}_t, \hat{b}_t) - (\mu^*_t, b^*_t) \right\|_\Theta \xrightarrow{\text{a.s.}} 0 \quad \text{as } T \to \infty
\]
if \(\left\| \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta\) 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
\[
\mathbb{E} \log^+ \left\| \ell_{(\mu^*_t, b^*_t)} \right\|_\Theta = \frac{1}{\sigma^2} \mathbb{E} \log^+ \left\| X_t - \mu^*_t - b^*_t \right\|_\Theta
\]
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
\[
\left\| \frac{1}{T} \sum_{t=2}^{T} \ell^*_t - \mathbb{E} \ell^*_t \right\|_\Theta \xrightarrow{\text{a.s.}} 0 \quad \text{as } T \to \infty
\]
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 Krenkel (1985), and because
\[
\mathbb{E} \left\| \ell^*_t \right\|_\Theta = \mathbb{E} \left\| \frac{1}{2} \log(2\pi \sigma^2) + \frac{1}{2\sigma^2}(X_t - \mu^*_t - b^*_t)^2 \right\|_\Theta \\
\leq \left\| \frac{1}{2} \log(2\pi \sigma^2) \right\|_\Theta + c \left\| \frac{1}{2\sigma^2} \right\|_\Theta \left( \mathbb{E} X_t^2 + \mathbb{E} \left\| \mu^*_t \right\|^2_\Theta + \mathbb{E} \left\| b^*_t \right\|^2_\Theta \right),
\]
for some \(c > 0\). This upper bound is finite by CS 2, since \(\Theta\) is compact and \(\sigma^2 > 0\). \(\blacksquare\)

**Lemma.** The parameter \(\theta_0\) is identifiable unique on \(\Theta\).

**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 \(\Theta\), 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 \(\mathbb{E} \ell^*_t\). \(\blacksquare\)

**A.6. Proof of Corollary 6**

The assumption that \(g\) is given by example E4 implies assumption DGP 3, which together with Assumptions DGP 1–2 implies that we can use Theorem 1 and Corollary 2. These results combined with the fact that the Gaussian specification (3.2) implies that the error \(\varepsilon_t\) has a finite second moment, then immediately delivers assumptions FLT 1 and CS 1.

Models (2.1)–(2.5) plus assumption DGP 2 directly allow for the multivariate application of Proposition 3 as \(\varepsilon_t\) enters just linearly. Therefore assumptions FLT 1–3 hold, while FLT 4 follows by our choice of survival function and FLT 5 follows by assumption. We conclude that we can apply Theorem 4 so that assumption CS 2 holds.
It remains to show that assumption CS 3 holds for the true parameter \( \theta_0 \). Blasques et al. (2018) show in the proof of their Theorem 4.1 that this is implied if

1. \( E[\ell_t^*(\theta_0)] < \infty \).
2. \( \|\ell_t^*\|_\alpha < \infty \).
3. \( \ell_t^*(\theta_0) = \ell_t^*(\theta) \) a.s. if and only if \( \theta = \theta_0 \).

The first point here follows because Theorems 1 and 4 imply that \( (\mu_t^*(\theta_0), b_t^*(\theta_0)) = (\mu_t, b_t) \) and so

\[
E[\ell_t^*(\theta_0)] = E\left( -\frac{1}{2} \log(2\pi \sigma^2) - \frac{1}{2\sigma^2} (X_t - \mu_t - b_t)^2 \right) = E[\ell_t] < \infty
\]

by the application of Corollary 2 and model specification (3.2). The second point follows from direct application of Theorem 4.

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 \( \phi \)-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