TI 2020-004/III
Tinbergen Institute Discussion Paper

Beta observation-driven models with exogenous regressors: a joint analysis of realized correlation and leverage effects

Paolo Gorgi[1]
Siem Jan Koopman[1]

[1] Vrije Universiteit Amsterdam, Tinbergen Institute, The Netherlands

Beta observation-driven models with exogenous regressors: a joint analysis of realized correlation and leverage effects

P. Gorgi[a,b] and S. J. Koopman[a,b,c]

[a]Vrije Universiteit Amsterdam, The Netherlands
[b]Tinbergen Institute, The Netherlands
[c]Aarhus University, Denmark

January 20, 2020

Abstract

We consider a general class of observation-driven models with exogenous regressors for double bounded data that are based on the beta distribution. We obtain a stationary and ergodic beta observation-driven process subject to a contraction condition on the stochastic dynamic model equation. We derive conditions for strong consistency and asymptotic normality of the maximum likelihood estimator. The general results are used to study the properties of a beta autoregressive process with threshold effects and to establish the asymptotic properties of the maximum likelihood estimator. We employ the threshold autoregressive model with leverage effects to analyze realized correlations for several sets of stock returns. We find that the impact of past values of realized correlation on future values is at least 10% higher when stock returns are negative rather than positive. This finding supports the conjecture that correlation between stock returns tends to be higher when stock prices are falling, and lower when there is a surge in stock prices. Finally, we conduct an out-of-sample study that shows that our model with leverage effects can enhance the accuracy of point and density forecasts of realized correlations.

Key words: Double bounded time series, financial econometrics, leverage effects, observation-driven models, realized correlation.

JEL classification: C32, C52, C58.

1 Introduction

There is an increasing interest in studying the dynamic statistical properties of double bounded time series data. A few typical examples of bounded variables are correlation measures, bounded between -1 and 1, market share variables, bounded between 0 and 100 (Ghosh et al., 1984), wind directions, bounded between 0 and 360 (Martín et al., 1999), and ages of first-time mothers, bounded between 15 and 45 (Falster et al., 2018). Time series of double bounded variables can be treated after the result of a transformation. For example, the logistic transformation of a variable leads to a double bounded variable between 0 and 1. However, data transformations introduce nonlinear effects in time series which need to be treated when conducting a statistical analysis of the (transformed) time series data.

A direct modeling approach for a double bounded variable is to consider a beta distribution. In the case of a time series variable, the dynamics in the beta distribution can be introduced by having a time-varying mean parameter. Earlier contributions of a time series analyses based on the beta distribution are as follows. Rocha and Cribari-Neto (2009) introduce a class of ARMA models based on the beta distribution. Casarin et al. (2012) consider a Bayesian approach for estimation and model section of beta autoregressive models. Guolo and Varin (2014) propose a beta time series model to analyze the dataset of Google Flu Trends. A convenient approach to specify time-variation in the mean of the beta distribution is to consider an observation-driven equation. In this way, the dynamic mean is a function of lagged values of the series and, possibly, exogenous regressors. This is in contrast with parameter-driven model specifications where the dynamic component of the model is specified through an unobserved autoregressive process. A key advantage of an observation-driven specification is that the practical implementation of the model is simpler since the likelihood function is in closed form and therefore estimation can be easily performed by the method of Maximum Likelihood (ML).

We consider a general class of observation-driven models based on the beta distribution that allows the inclusion of exogenous regressors. We show stationarity and ergodicity of the class of processes under a contraction condition on the dynamic equation of the model. We argue that a Markov chain approach for martingale difference sequences as considered in Zheng et al. (2015) is not applicable in this framework. The Markov chain property for our class of beta processes is lost due to the presence of regressors. The regressors are only assumed to follow a stationary and ergodic process without imposing any specification. Results on stationarity of observation-driven models with exogenous regressors can be found in Agosto et al. (2016) for Poisson autoregressions and Aknouche and Francq (2018) for count and duration models. We also investigate the asymptotic properties of the ML estimator by deriving conditions for strong consistency and asymptotic normality. Although ML is widely used for the estimation of beta autoregressive models as in Rocha and Cribari-Neto (2009) and Guolo and Varin (2014), no formal results are available in literature on its asymptotic properties. Our theoretical framework is applied to study the properties of a beta observation-driven model with threshold effects. This model forms the basis for our empirical analysis on realized correlations and leverage effects.

The measurement of dependence between financial assets plays a key role in financial risk management and portfolio optimization. Realized correlation has become an important measure of dependence that exploits the information content of high-frequency financial transaction data. By construction, the realized correlation takes values in the interval between -1 and 1. Therefore, our modeling framework is particularly suited to model realized correlation as it properly accounts for the double bounded nature of the data. Particular attention in the empirical analysis is given to the leverage effects in correlation time series. There is a vast literature on leverage effects in volatility and covariation. Most literature discusses leverage effects in volatility, where it is well established that negative returns tend to have a larger impact on the volatility than positive returns. The standard approach is to consider the generalized autoregressive conditional heteroskedasticity (GARCH) model (Engle, 1982; Bollerslev, 1986) and introduce leverage in the specification of the conditional variance; see Glosten et al. (1993) and Nelson (1991) for further discussions. We refer to Rodríguez and Ruiz (2012) for a review on GARCH models with leverage effects. On the other hand, studies on leverage effects in correlation are typically based on multivariate GARCH models and, in particular, dynamic conditional correlation (DCC) models; see Cappiello et al. (2006) and Audrino and Trojani (2011). Recently, Bollerslev et al. (2018) introduce leverage effects in multivariate GARCH models by using realized semi-covariances.

In the empirical part of this paper we study leverage effects in realized correlation directly instead of using a multivariate GARCH or DCC model. An important benefit of this approach is its robustness to model misspecification since the realized correlation is a direct measure of correlation. On the contrary, the estimated dynamic properties of the conditional correlation in a DCC model also rely on the correct specification of the conditional variance of the assets. We base our empirical study on a beta observation-driven model with threshold effects to describe leverage in correlation. We consider two specifications of the leverage: the first is based on the returns of the assets themselves and the second is based on the market returns of the S&P500 index. We analyze realized correlation series of ten pairs of stocks over a time span of ten years. The results indicate that there is a highly significant leverage effect in correlation of about 10%. This finding implies that the impact of lagged returns on future correlation is 10% higher when we are in a period of overall negative returns compared to a period with positive returns. This result appears to be consistent across the different pairs of stocks and robust with respect to the specification of the leverage effect. Additionally, the empirical analysis suggests that leverage effects are better explained by market returns than the returns of the individual stocks. This finding is confirmed by the in-sample fit of the models as well as the out-of-sample forecasting accuracy. More specifically, the model with leverage effects determined by the S&P500 produces the most accurate point and density forecasts.

The remainder of the paper is organized in the following way. Section 2 presents the general specification of the model and derives its stochastic properties. Section 3 discusses asymptotic properties of the maximum likelihood estimator for the vector of unknown parameters in the model. Section 4 introduces the threshold specification as a special case of the general model. Section 5 presents a Monte Carlo simulation study to evaluate the small sample properties of the maximum likelihood estimator. Section 6 presents the empirical application to analyze leverage effects in realized correlation. Section 7 presents concluding remarks. The proofs and technical derivations are presented in the Appendix.

2 Beta observation-driven models with exogenous regressors

Assume we are modeling time series of continuous random variables that take values in an interval of the form $[a, b]$, where $a, b \in \mathbb{R}$ and $b > a$. The upper and lower bounds of the interval a and b are known, *a priori*. For example, time series of proportions have $a = 0$ and $b = 1$, of realized correlation series have $a = -1$ and $b = 1$, and of directions have $a = 0$ and $b = 360$. However, from a modeling perspective and since the bounds are known, we can focus on time series of values within the unit interval: any time series variable can be transformed such that it take values in $[0, 1]$. We consider $\{y_t\}_{t\in\mathbb{Z}}$ as the time series of interest. We further assume that n regressor variables are available, we have $\{x_t\}_{t\in\mathbb{Z}}$ where x_t is an n-variate random vector. The specification of the beta observation-driven model with explanatory variables is given by

$$y_t \,\big|\, \mathcal{F}_{t-1} \sim \mathcal{B}eta\big(\phi \cdot \mu_t \,,\, \phi\,(1 - \mu_t)\big), \tag{1}$$

where $\mathcal{F}_t$ is the σ-field generated by $\{(y_j, x_j^\top) : j \leq t\}$, ϕ is a precision parameter and μ_t is the conditional mean of the beta distribution. To address the possible serial dependence in the time series y_t, we consider the dynamic specification for μ_t given by

$$\mu_{t+1} = g_\theta(\mu_t, y_t, x_t), \tag{2}$$

where $g_\theta : [0, 1]^2 \times \mathbb{R}^n \mapsto (0, 1)$ is a parametric function indexed by a k-dimensional parameter vector $\theta \in \mathbb{R}^k$. It is the dependence of μ_t on y_{t-1} in (2) that classifies the model as observation-driven; see the exposition in Cox (1981). We have implicitly assumed that μ_t is $\mathcal{F}_{t-1}$-measurable. Below we shall introduce conditions on g_θ to ensure that the process is stationary and ergodic, and to ensure that μ_t is $\mathcal{F}_{t-1}$-measurable. The above specification entails that μ_t is the conditional expectation of y_t, that is, $\mu_t = \mathbb{E}(y_t|\mathcal{F}_{t-1})$. More generally, the conditional density function of y_t is given by

$$p(y_t|\mu_t; \phi) = \frac{y_t^{\phi\mu_t - 1}(1 - y_t)^{\phi(1-\mu_t)-1}}{B\big(\phi\mu_t, \phi(1 - \mu_t)\big)},$$

where $B(\cdot, \cdot)$ denotes the beta function.

Next we derive stationarity conditions which form the basis towards consistent inference for the model. For simplicity of exposition, we rewrite the process in equations (1) and (2) as follows

$$y_t = \mathcal{B}(u_t; \mu_t, \phi), \qquad \mu_{t+1} = g_\theta(\mu_t, y_t, x_t), \tag{3}$$

where $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed (iid) uniform random variables in the unit interval, $u_t \sim \mathcal{U}(0, 1)$, and $\mathcal{B}(\cdot; \mu_t, \phi)$ is the inverse of the cumulative distribution function of a beta distribution with mean μ_t and precision parameter ϕ. We impose some assumptions on the regressors and a contraction condition on the updating function g_θ. More specifically, we consider the following conditions:

S1: The sequence $\{(u_t, x_t^\top)\}_{t\in\mathbb{Z}}$ is stationary and ergodic, and random variable u_t is independent of $\{(u_j, x_j^\top) : j \le t-1\}$.

S2: The updating function g_θ satisfies the following contraction condition with probability one

$$|g_\theta(\mu, y, x_t) - g_\theta(\mu^*, y^*, x_t)| \le a_\mu|\mu - \mu^*| + b_\mu|y - y^*|, \tag{4}$$

for any $(\mu, \mu^*, y, y^*) \in [0,1]^4$, where a_μ, b_μ are positive constants such that $a_\mu + b_\mu < 1$.

Condition **S1** imposes stationarity and ergodicity of x_t together with some assumptions on how x_t relates to u_t. In particular, variables u_t and x_t are imposed to be jointly stationary and ergodic. It implies that u_t and x_t are allowed to be dependent random variables. As a result, variable of interest y_t is not necessarily independent of x_t conditional on $\mathcal{F}_{t-1}$, and hence contemporaneous dependence is allowed. A further assumption is that u_t is independent of the past. This condition guarantees that y_t conditional on the past has a beta distribution with mean μ_t and precision parameter ϕ. This result is needed to ensure the equivalence between the representation of the process given by equations (1), (2), and equation (3). Condition **S2** imposes a contraction condition on g_θ by requiring the Lipschitz coefficient $a_\mu + b_\mu$ to be smaller than one. The next results delivers the stationarity and ergodicity of the observation-driven beta process under **S1** and **S2**.

Theorem 2.1. *Let **S1** and **S2** hold. Then, the process defined by equation (3) admits a stationary and ergodic solution $\{(y_t, \mu_t, x_t)\}_{t\in\mathbb{Z}}$. Furthermore, μ_t is $\mathcal{F}_{t-1}$-measurable.*

In the next sections, Theorem 2.1 shall be employed to derive the asymptotic theory of the maximum likelihood estimator and to study the stochastic properties of a threshold model.

3 Maximum likelihood estimation

In this section, we discuss parameter estimation by the method of Maximum Likelihood (ML) for the model presented in Section 2. The parameters of the model are collected in the $(k+1)\times 1$ vector

$\kappa = (\theta^\top, \phi)^\top$. We assume that we observe a sample of T realisations of y_t and x_t generated by their respective processes, with true parameter value $\kappa_0 = (\theta_0^\top, \phi_0)^\top$, and we denote the sample by $\{(y_t, x_t^\top)\}_{t=1}^T$. The first step to derive the likelihood function is to recover the time-varying mean from the observed data through the following recursive equation

$$\hat{\mu}_{t+1}(\theta) = g_\theta(\hat{\mu}_t(\theta), y_t, x_t), \qquad t = 1, \dots, T, \tag{5}$$

where the recursion is initialized at a fixed point $\hat{\mu}_1(\theta) \in (0, 1)$. We refer to $\hat{\mu}_t(\theta)$ as the filtered parameter. The contribution to the log-likelihood of the t-th observation is

$$\hat{l}_t(\kappa) = \log p(y_t | \hat{\mu}_t(\theta); \phi).$$

Finally, the ML estimator is defined as the maximizer of the log-likelihood function

$$\hat{\kappa}_T = \arg\sup_{\kappa \in K} \hat{L}_T(\kappa), \quad \text{with} \quad \hat{L}_T(\kappa) = \frac{1}{T} \sum_{t=1}^{T} \hat{l}_t(\kappa),$$

where $K = \Theta \times \Phi$, and $\Theta \subset \mathbb{R}^k$ and $\Phi \subset (0, \infty)$ are compact parameter sets.

To derive the strong consistency and asymptotic normality of the ML estimator, we consider the following conditions.

A1: The observed sample $\{(y_t, x_t^\top)\}_{t=1}^T$ follows the model's equations in (3) at $\kappa = \kappa_0$ and conditions **S1** and **S2** are satisfied.

A2: The function $(\theta, \mu) \mapsto g_\theta(\mu, y, x)$ is continuous for any $y \in [0, 1]$ and any $x \in \mathbb{R}^n$.

A3: The contraction condition in (4) is satisfied for any $\theta \in \Theta$ and $\theta_0 \in \Theta$.

A4: There is a $\bar{c} > 0$ such that $g_\theta(\mu, y, x) \in [\bar{c}, 1 - \bar{c}]$ for any $(\theta, \mu, y, x) \in \Theta \times [0, 1]^2 \times \mathbb{R}^n$.

A5: For any $\theta_1, \theta_2 \in \Theta$ and $\mu \in (0, 1)$, the equality $g_{\theta_1}(\mu, y_t, x_t) = g_{\theta_2}(\mu, y_t, x_t)$ holds true with probability one if and only if $\theta_1 = \theta_2$.

A6: The function $(\theta, \mu) \mapsto g_\theta(\mu, y, x)$ is three times continuously differentiable with uniformly bounded derivatives.

A7: The elements of vector $\partial g_\theta(\mu_t, y_t, x_t) \,/\, \partial\theta\big|_{\theta=\theta_0}$ are linearly independent random variables.

Condition **A1** ensures that the observed sample of data follows a stationary and ergodic process. Condition **A2** imposes continuity of the updating function, which in turn ensures that $\theta \mapsto \hat{\mu}_t(\theta)$ is continuous for any t. This smoothness condition is sufficient for consistency, however, the additional differentiability condition in **A6** is needed to ensure asymptotic normality. Condition **A3** ensures that $\hat{\mu}_t(\cdot)$ converges to a stationary and ergodic $\mathcal{F}_{t-1}$-measurable function $\tilde{\mu}_t(\cdot)$ as $t \to \infty$. This is typically referred in the literature as invertibility (Straumann and Mikosch, 2006; Blasques et al., 2018). Condition **A4** sets bounds on the updating function. Bound conditions are standard in the literature when the time-varying parameter takes values on a bounded set. For instance, lower bounds are considered in the estimation of conditional heteroscedastic models (Straumann and Mikosch, 2006) and of positive-valued observation-driven models (Davis and Liu, 2016). Finally, conditions **A5** and **A7** ensure identifiability and positive definiteness of the asymptotic covariance matrix of the ML estimator, respectively.

To establish the asymptotic properties of the ML estimator, we first derive the stochastic limit properties of the filtered parameter defined in (5). The next result ensures the uniform convergence over Θ of the filtered parameter $\hat{\mu}_t(\cdot)$ to an $\mathcal{F}_{t-1}$-measurable stationary and ergodic limit $\tilde{\mu}_t(\cdot)$ such that $\tilde{\mu}_t(\theta_0) = \mu_t$. The rate of convergence is shown to be exponentially fast. In particular, a sequence of random variables $\{\hat{\eta}_t\}_{t\in\mathbb{N}}$ is said to converge exponentially almost surely (e.a.s.) to another sequence $\{\tilde{\eta}_t\}_{t\in\mathbb{N}}$ if there is a constant $c > 1$ such that $c^t|\hat{\eta}_t - \tilde{\eta}_t| \xrightarrow{a.s.} 0$ as $t \to \infty$. We denote with $\|\cdot\|_\Theta$ the supremum norm. For a given function, $f : \Theta \mapsto \mathbb{R}$, the supremum norm is defined as $\|f\|_\Theta = \sup_{\theta\in\Theta} |f(\theta)|$.

Proposition 3.1. *Let **A1**-**A3** hold. Then $\{\hat{\mu}_t(\cdot)\}_{t\in\mathbb{N}}$ converses e.a.s. and uniformly over Θ to a unique stationary and ergodic sequence $\{\tilde{\mu}_t(\cdot)\}_{t\in\mathbb{Z}}$,*

$$\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta \xrightarrow{e.a.s.} 0, \quad t \to \infty,$$

for any initialization $\hat{\mu}_1(\theta) \in (0, 1)$.

The convergence result provided in Proposition 3.1 is useful as it ensures that the log-likelihood contribution $\hat{l}_t(\cdot)$ converges to a stationary and ergodic function $l_t(\cdot)$, given by

$$l_t(\kappa) = \log p(y_t|\tilde{\mu}_t(\theta); \phi).$$

The next results deliver the strong consistency and the asymptotic normality of the ML estimator.

Theorem 3.1. *Let **A1**-**A5** hold and $\kappa_0 \in K$. Then, the ML estimator is strongly consistent, that is*

$$\hat{\kappa}_T \xrightarrow{a.s.} \kappa_0, \quad as \quad T \to \infty.$$

*Assume further that the additional conditions **A6** and **A7** hold and $\kappa_0 \in int(K)$. Then, the ML estimator is asymptotically normally distributed*

$$\sqrt{T}(\hat{\kappa}_T - \kappa_0) \xrightarrow{d} N(0, \Omega), \quad as \quad T \to \infty, \quad where \quad \Omega = -\mathbb{E}\left(\frac{\partial^2 l_t(\kappa_0)}{\partial \kappa \partial \kappa^\top}\right)^{-1},$$

and Ω is positive definite.

Theorem 3.1 will be employed in the next section to derive the asymptotic properties of the ML estimator for a beta observation-driven process with threshold effects.

4 Threshold beta autoregressive model

To illustrate our general dynamic framework for beta observation-driven processes, as discussed in Sections 2 and 3, we consider a specific functional form for the updating function $g_\theta(\cdot)$. Specifically, we introduce a threshold specification for the conditional expectation of the beta observation-driven model. We allow the impact of y_t on μ_{t+1} to differ depending on the state of a threshold vector of regressors x_t. The threshold model will be employed in the simulation study of Section 5 where the asymptotic properties of the ML estimator are confronted with its small sample properties. Furthermore, the threshold beta regression model is adopted in Section 6 for the purpose of modeling time-varying realized correlation of pairs of financial assets.

The standard threshold regression model, including its inference and asymptotic properties, are well developed by Chan (1993) and Hansen (2000). The threshold autoregression model is introduced by Tong and Lim (1980) and explored in detail by Tong (1983). A related class of models is the widely explored smooth transition autoregressive model of Teräsvirta (1994) and Teräsvirta et al. (2010). We regard the model below as an adaptation of the threshold autoregressive models for bounded time series; when we replace the beta distribution $\mathcal{B}eta\big(\phi \cdot \mu_t\,,\, \phi\,(1-\mu_t)\big)$ in (1) by the normal distribution $\mathcal{N}\big(\mu_t\,,\,\sigma^2\big)$, with mean $\mu_t \in \mathbb{R}$ and variance $\sigma^2 > 0$, we can obtain the threshold autoregressive model of Tong and Lim (1980) as a special case.

Before we formally define the model, we start by considering a collection of disjoint sets, $A_1, \ldots, A_J$, which are subsets of the sample space of x_t, that is $A_j \subset \mathbb{R}^k$, $j = 1, \ldots, J$. Next, we define the following indicator variables

$$I(A_j, x_t) = \begin{cases} 1 & \text{if} \quad x_t \in A_j, \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1, \ldots, J$ and $t \in \mathbb{Z}$. Since $\{A_1, \ldots, A_J\}$ is not necessarily a cover of $\mathbb{R}^k$, we can have $\sum_{j=1}^{J} I(A_j, x_t) = 0$ with positive probability. In general, we have $\sum_{j=1}^{J} I(A_j, x_t) \in \{0, 1\}$ with probability one, given that the sets are disjoint. The specification of μ_t with threshold effects is

$$\mu_{t+1} = \delta + \beta\mu_t + \Big(\gamma + \sum_{j=1}^{J} \gamma_j I(A_j, x_t)\Big) y_t, \tag{6}$$

where $\delta > 0$, $\beta \geq 0$, $\gamma > 0$, and $\gamma_j \geq -\gamma$, for $j = 1, \ldots, J$.

To ensure that $\mu_t \in (0, 1)$ with probability one, we impose the following restriction on the parameters

$$\delta + \beta + \gamma + \max\{0, \gamma_1, \ldots, \gamma_J\} < 1. \tag{7}$$

As shown in the next result, it turns out that the double bound restriction in (7) is sufficient to ensure the stationarity and ergodicity of the process. More specifically, we have that (7) implies $\beta + \gamma + \max\{0, \gamma_1, \ldots, \gamma_J\} < 1$ which makes the contraction condition in (4) to be satisfied.

Theorem 4.1. *Assume that the vector of regressors x_t satisfies condition **S1**. Then, the threshold beta autoregressive process specified by equations (1) and (6) admits a stationary and ergodic solution.*

Theorem 4.1 shows that, under the stationary assumption of the regressors in **S1**, the threshold beta observation-driven process always admits a stationary and ergodic solution.

Next, we employ Theorem 4.2 to derive the consistency and asymptotic normality of the ML estimator for the threshold model. We obtain the following result.

Theorem 4.2. *Assume the following conditions hold:*

(i) The parameter set K is a compact set such that $\delta > 0$, $\beta \geq 0$, $\gamma > 0$, $\gamma_j \geq -\gamma$, $j = 1, \ldots, J$, and (7) is satisfied for any $\kappa \in K$. Furthermore, $\kappa_0 \in K$.

*(ii) The vector of regressors x_t satisfies condition **S1**. Furthermore, $x_t \in A_j$ with positive probability for $j = 1, \ldots, J$ and $x_t \notin \bigcup_{j=1}^{J} A_j$ with positive probability.*

Then, the ML estimator is strongly consistent. Assume further that $\kappa_0 \in int(K)$. Then, the ML estimator is asymptotically normally distributed

$$\sqrt{T}(\hat{\kappa}_T - \kappa_0) \xrightarrow{d} N(0, \Omega), \quad \text{as} \quad T \to \infty.$$

The proof of Theorem 4.2 is obtained by verifying that **A1**-**A7** are satisfied. Condition (i) in Theorem 4.2 is imposed to ensure that $\hat{\mu}_t(\theta) \in (0,1)$ for any $\theta \in \Theta$. Condition (ii) is an identifiability condition that makes **A5** hold. It is clear that if the event $x_t \in A_j$ occurs with probability zero, then γ_j is not identified. Similarly, if the probability of $x_t \notin \bigcup_{j=1}^{J} A_j$ is zero, then any re-parameterization of the form $\gamma + c_\gamma$ and $\gamma_j - c_\gamma$, $j = 1, \ldots, J$, will lead to the same process for any $c_\gamma \in \mathbb{R}$.

5 Simulation study

We conduct a Monte Carlo simulation study to investigate the small sample properties of the ML estimator. For this purpose, we consider the threshold beta observation-driven model as the data generating process. Our specific model for this study is given by

$$y_t = \mathcal{B}(u_t; \mu_t, \phi), \qquad \mu_{t+1} = \delta + \beta\mu_t + \big(\gamma + \gamma_1 I(A_1, x_t)\big)y_t, \tag{8}$$

where the inverse of the cumulative beta distribution function is denoted by $\mathcal{B}$, with mean μ_t and precision ϕ, as discussed in more detail below (3), the sequence of random variables $\{u_t\}_{t\in\mathbb{Z}}$ is iid with $u_t \sim \mathcal{U}(0,1)$, indicator function $I(A_1, x_t)$ is based on the set $A_1 = (-\infty, 0)$ and on the zero-mean univariate regressor x_t following the first-order autoregressive moving average process given by

$$x_t = bx_{t-1} + \varepsilon_t + a\varepsilon_{t-1},$$

with autoregressive coefficient b, moving average coefficient a, and where the sequence of random normal variables $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is iid with $\varepsilon_t \sim \mathcal{N}(0,1)$. The theoretical results allow the random variables ε_t and u_t to be dependent of each other. We introduce this dependence in the model (8) by means of a Gaussian copula function, see McNeil et al. (2015, Chapter 7) for a textbook treatment. In particular, we define $v_t = (v_{1,t}, v_{2,t})^\top$ as an iid sqeunce from a bivariate Gaussian copula with correlation parameter ρ. From this sequence we obtain the two model noise variables by having $u_t = v_{1,t}$ and $\varepsilon_t = \Phi^{-1}(v_{2,t})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. As a result, we have u_t coming from a uniform marginal distribution and ε_t coming from a standard normal marginal distribution. The dependence between ε_t and u_t is determined by the correlation parameter ρ of the copula; its functional form is nonlinear.

Table 1 reports the summary results of the simulation experiment. The asymptotic standard error for each ML estimate is computed from the so-called plug-in estimate of the asymptotic covariance matrix of the ML estimator as given by

$$\hat{\Omega} = -\left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \hat{l}_t(\hat{\kappa}_T)}{\partial\kappa\partial\kappa^\top} \right)^{-1}. \tag{9}$$

The results in Table 1 show that the estimation bias is negligible for all parameters and for different sample sizes. This finding is confirmed by fact that the root mean squared error coincides with the standard error of the parameter estimates. Hence the (squared) bias contribution to the MSE can be considered as negligible. Furthermore, the estimates of the standard errors obtained from the asymptotic distribution of the ML estimator can be considered as accurate, even for small sample sizes. This finding follows from the fact that the sample mean of the estimated asymptotic standard errors for the 1000 parameter estimates is almost equivalent to the standard error obtained from the

Table 1: Summary statistics of a simulation experiment.
We have simulated 1000 time series for y_t, with sample size $T = \{500, 1000, 2500\}$, from model (8) with parameter values $\delta = 0.009$, $\beta = 0.85$, $\gamma = 0.14$, $\gamma_1 = -0.03$, $\phi = 25.0$, $b = 0.8$, $a = 0.2$, and $\rho = 0.75$. For each simulated time series, the parameters are estimated by the method of ML as discussed in Section 3, except parameters b, a and ϕ which are only used to simulate the regressor and are therefore not estimated. From the 1000 ML estimates of each parameter, we report the sample mean (Mean), root mean squared error (RMSE), sample standard error (SE), and the sample mean of the asymptotic standard error (ASE).

		$\delta = 0.009$	$\beta = 0.85$	$\gamma = 0.14$	$\gamma_1 = -0.03$	$\phi = 25.0$
$T = 500$	Mean	0.011	0.844	0.141	-0.029	25.157
	RMSE	0.004	0.030	0.027	0.008	1.468
	SE	0.004	0.029	0.027	0.008	1.460
	ASE	0.003	0.029	0.026	0.007	1.568
$T = 1000$	Mean	0.010	0.846	0.141	-0.029	25.083
	RMSE	0.002	0.021	0.020	0.005	1.057
	SE	0.002	0.021	0.019	0.005	1.055
	ASE	0.002	0.020	0.018	0.005	1.103
$T = 2500$	Mean	0.009	0.848	0.140	-0.030	25.050
	RMSE	0.001	0.012	0.011	0.003	0.669
	SE	0.001	0.012	0.011	0.003	0.668
	ASE	0.001	0.012	0.011	0.003	0.696

sample variance of the same 1000 parameter estimates. Finally, the results in Table 1 are obtained for a correlation parameter of the copula that is set equal to $\rho = 0.75$. However, after some ample experimentations with different correlation values, we have concluded that the value for ρ has a negligible effect on the presented simulation results. This finding is coherent with the theoretical results in Sections 2 and 3; they indicate that the asymptotic properties hold irrespective of the contemporaneous dependence between y_t and x_t.

6 Modeling and forecasting realized correlation with leverage

We employ the threshold beta observation-driven model of Section 4 to analyze realized correlation between pairs of financial assets. Realized measures of volatility and dependence are obtained from high-frequency intra-daily (minutes, seconds, ticks) financial returns; see the seminal work of Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002). Much emphasis has been given to the modeling and forecasting of realized volatility, and its incorporation in financial risk analyses. For this same purpose, and in particular for financial hedging analyses, econometric treatments for realized correlation have been developed; see, for example, Audrino and Corsi (2010) and Aslanidis and Christiansen (2012). In this section we propose an alternative approach to the modeling and forecasting of realized correlation. We explicitly account for leverage effects in the dynamic process for realized correlation. The study of leverage effects in realized correlation is a novel development as leverage effects in correlation are typically studied through multivariate GARCH models.

6.1 Model for realized correlation with leverage

Let y_t denote the realized correlation at time t between some pair, i and j, of financial asset returns, where the time index t typically refers to a trading day. We consider the dynamic model (3) for y_t where the conditional expectation μ_t is subject to a threshold function that effectively represents the leverage effect. We introduce the leverage effect by considering two different specifications of μ_t that are based on the threshold specification in (6).

In the first specification, the leverage effect is determined by the daily returns of the pair of assets i and j, which we denote with $r_{i,t}$ and $r_{j,t}$, respectively. The realized correlation y_t is for the same pair of assets but is computed using high-frequency intra-daily returns; see Section 6.2 below. The model for y_t is given by $y_t = \mathcal{B}(u_t; \mu_t, \phi)$ as in (3). The updating function for the conditional expectation μ_t is given by

$$\mu_{t+1} = \delta + \beta\mu_t + (\gamma + \gamma_+ I_t^+ + \gamma_- I_t^-)y_t, \tag{10}$$

where parameters δ, β, γ, γ_+ and γ_- are placed in the unknown parameter vector θ, and the indicator functions are specified by

$$I_t^+ = \begin{cases} 1 & \text{if} \quad r_{i,t} > 0, r_{j,t} > 0, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad I_t^- = \begin{cases} 1 & \text{if} \quad r_{i,t} < 0, r_{j,t} < 0, \\ 0 & \text{otherwise.} \end{cases}$$

In this specification the default impact of y_t on μ_{t+1} is measured by γ, the impact equals $\gamma + \gamma_+$ when both daily returns are positive and equals $\gamma + \gamma_-$ when both daily returns are negative. In empirical studies, we expect $\gamma_- > \gamma_+$ and $\gamma_- > 0$ which defines implicitly the leverage effect. Hence, we anticipate that the impact of past values of realized correlation on μ_t is higher when the daily returns of the two assets are both negative compared to when they are both positive or have opposite signs. Finally, we note that contemporaneous dependence between u_t and the pair of returns $(r_{i,t}, r_{j,t})$ may be present given that they are based on the same equities. This dependence is accommodated in our framework as discussed in the theory section and in the simulation study.

In the second specification, we let the leverage effect be determined by the daily returns of the S&P500 market index, which we denote with r_t^{sp}. The conditional mean μ_t is specified as

$$\mu_{t+1} = \delta + \beta\mu_t + (\gamma + \gamma_-^{\text{sp}} I_t^{\text{sp}})y_t, \tag{11}$$

$$\text{where} \quad I_t^{\text{sp}} = \begin{cases} 1 & \text{if} \quad r_t^{\text{sp}} < 0, \\ 0 & \text{otherwise.} \end{cases}$$

In this specification, the impact of y_t on μ_{t+1} is equal to $\gamma + \gamma_-^{\text{sp}}$ when the daily return of the market is negative and γ when it is non-negative. Other variables or other sets of variables can be considered in this specification. Also, the two specifications can be combined.

6.2 Data: daily realized correlation and daily returns

The data set consists of realized correlation series for 10 randomly selected pairs of equities from 15 equities with ticker symbols AA, AXP, BA, CAT, GE, HD, HON, IBM, JPM, KO, MCD, PFE, PG, WMT, and XOM; all these equities are included in the Dow Jones Industrial Average. The pairs considered are the same as in Table 1 of Gorgi et al. (2019). The sample size of the data is $T = 2515$ trading days; from 1 January 2001 to 31 December 2010. The realized correlation on trading day t for a pair of equities is obtained from the realized covariance matrix of the pair by standardizing the realized covariance with the square roots of the realized variances. Realized covariance matrices are constructed from transaction prices from consolidated trades in the Trade and Quote (TAQ) database through the Wharton Research Data Services (WRDS) system. In constructing the realized measures, a standard cleaning procedure is carried out which includes the exclusion of overnight returns. The realized measures are based on an overall sample frequency of five minutes. We have adopted a kernel function that is based on the refresh sampling scheme of Barndorff-Nielsen et al. (2011) where the irregular sampling time interval ends when at least one realization is recorded for all assets. By shifting the starting time by one-second increments, 300 different estimates in a five-minute interval are obtained; the average gives the subsampled realized covariance matrix. We refer to Gorgi et al. (2019) for further details on the construction of this data set. The daily asset returns and the S&P500 returns are based on open-to-close returns.

Figure 1 displays the daily time series of realized correlations between the equities Alcoa (AA) and Caterpillar (CAT), together with the estimated μ_t (red line) based on the default autoregressive beta model (3) without leverage, that is $\gamma_+ = \gamma_- = 0$ in (10). The realized correlations are strongly varying over time and show strong temporal dependence. We further learn that the correlations sharply increases during the financial crisis of 2008. These empirical features can also be observed for realized correlations between other pairs of equities. These findings highlight the empirical fact that correlations tend to be higher during periods of which many negative returns are observed. It is also a motivation to consider the leverage effect (10) in our model when modeling and forecasting realized correlations.

Table 2: Parameter estimation results.

We present the maximum likelihood estimates of the parameters in our realized correlation model with leverage, for ten different pairs of assets. The standard errors of the estimates are in brackets. The columnn "log-lik" reports the maximized loglikelihood value. The last column presents the p-value of the likelihood ratio test for the null hypothesis of no leverage effects ($H_0 : \gamma_+ = \gamma_- = 0$ or $\gamma_-^{\text{sp}} = 0$).

	δ	β	γ	γ_+	γ_-	γ_-^{sp}	ϕ	log-lik	p-value
AA/CAT	0.013 (0.003)	0.820 (0.014)	0.160 (0.012)	-	-	-	39.796 (1.110)	3059.43	-
	0.013 (0.003)	0.840 (0.009)	0.139 (0.009)	-0.004 (0.003)	0.010 (0.003)	-	40.241 (1.138)	3073.00	1.3×10^{-6}
	0.012 (0.002)	0.842 (0.012)	0.135 (0.011)	-	-	0.013 (0.002)	40.261 (1.124)	3074.48	4.1×10^{-8}
AXP/PFE	0.026 (0.005)	0.772 (0.020)	0.188 (0.015)	-	-	-	45.030 (1.257)	3127.88	-
	0.026 (0.004)	0.779 (0.014)	0.178 (0.012)	-0.005 (0.003)	0.009 (0.003)	-	45.348 (1.274)	3136.69	1.5×10^{-4}
	0.027 (0.005)	0.793 (0.019)	0.155 (0.014)	-	-	0.019 (0.003)	45.822 (1.280)	3149.83	3.4×10^{-11}
AXP/WMT	0.013 (0.003)	0.847 (0.012)	0.134 (0.010)	-	-	-	44.144 (1.233)	3151.36	-
	0.013 (0.003)	0.853 (0.008)	0.127 (0.008)	-0.003 (0.003)	0.008 (0.003)	-	44.493 (1.261)	3161.09	5.9×10^{-5}
	0.015 (0.003)	0.862 (0.011)	0.109 (0.009)	-	-	0.015 (0.002)	44.880 (1.254)	3172.33	9.4×10^{-11}
BA/HON	0.011 (0.003)	0.820 (0.013)	0.164 (0.012)	-	-	-	37.408 (1.044)	3057.62	-
	0.012 (0.003)	0.825 (0.010)	0.155 (0.010)	0.002 (0.003)	0.005 (0.003)	-	37.491 (1.057)	3059.80	0.112
	0.011 (0.002)	0.830 (0.012)	0.149 (0.011)	-	-	0.011 (0.002)	37.713 (1.052)	3068.30	3.8×10^{-6}
CAT/KO	0.017 (0.004)	0.843 (0.013)	0.130 (0.011)	-	-	-	40.830 (1.139)	2997.71	-
	0.018 (0.003)	0.854 (0.008)	0.112 (0.008)	0.004 (0.003)	0.013 (0.003)	-	41.200 (1.165)	3008.78	1.5×10^{-5}
	0.019 (0.004)	0.855 (0.013)	0.110 (0.010)	-	-	0.013 (0.003)	41.257 (1.149)	3008.72	2.7×10^{-6}
GE/PFE	0.026 (0.005)	0.740 (0.019)	0.221 (0.015)	-	-	-	48.672 (1.360)	3272.49	-
	0.028 (0.004)	0.748 (0.014)	0.208 (0.013)	0.000 (0.003)	0.010 (0.003)	-	48.953 (1.376)	3279.57	8.4×10^{-4}
	0.028 (0.005)	0.763 (0.018)	0.186 (0.014)	-	-	0.019 (0.003)	49.567 (1.386)	3295.46	1.2×10^{-11}
HD/JPM	0.022 (0.005)	0.812 (0.016)	0.156 (0.012)	-	-	-	48.939 (1.368)	3281.55	-
	0.023 (0.004)	0.821 (0.010)	0.141 (0.010)	-0.002 (0.003)	0.011 (0.003)	-	49.408 (1.397)	3293.41	7.1×10^{-6}
	0.024 (0.004)	0.825 (0.014)	0.132 (0.011)	-	-	0.016 (0.002)	49.822 (1.393)	3304.11	1.9×10^{-11}
IBM/PG	0.037 (0.007)	0.747 (0.023)	0.198 (0.016)	-	-	-	47.708 (1.333)	3234.79	-
	0.041 (0.005)	0.751 (0.016)	0.185 (0.013)	-0.007 (0.003)	0.015 (0.003)	-	48.599 (1.366)	3258.01	8.1×10^{-11}
	0.034 (0.006)	0.778 (0.019)	0.161 (0.014)	-	-	0.020 (0.003)	48.719 (1.362)	3261.25	3.5×10^{-13}
JPM/XOM	0.021 (0.004)	0.743 (0.017)	0.225 (0.014)	-	-	-	51.358 (1.436)	3333.40	-
	0.022 (0.004)	0.750 (0.013)	0.214 (0.012)	-0.005 (0.003)	0.010 (0.003)	-	51.785 (1.456)	3344.04	2.3×10^{-5}
	0.022 (0.004)	0.757 (0.016)	0.202 (0.013)	-	-	0.015 (0.003)	52.011 (1.455)	3349.36	1.6×10^{-8}
MCD/PG	0.032 (0.006)	0.774 (0.020)	0.176 (0.014)	-	-	-	42.214 (1.178)	3034.41	-
	0.035 (0.005)	0.783 (0.014)	0.158 (0.012)	0.001 (0.003)	0.014 (0.004)	-	42.513 (1.194)	3043.25	1.4×10^{-4}
	0.031 (0.006)	0.794 (0.019)	0.150 (0.013)	-	-	0.017 (0.003)	42.777 (1.194)	3051.13	7.4×10^{-9}

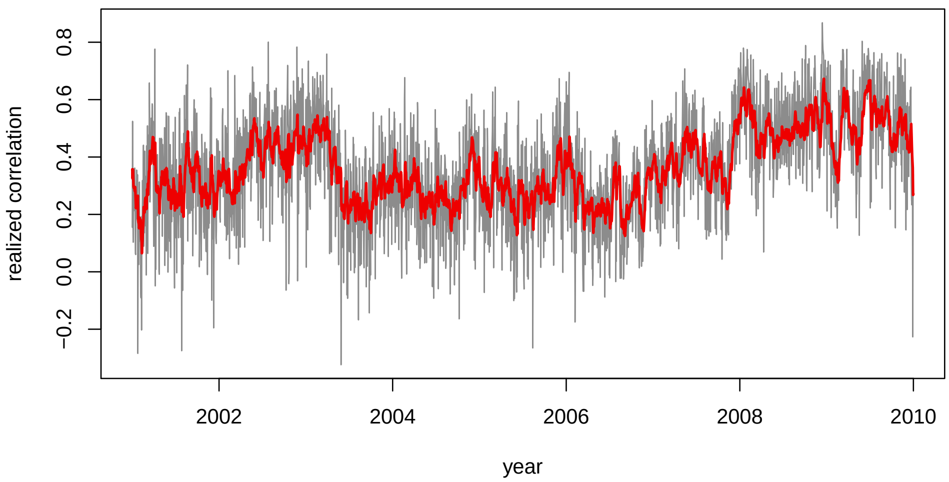

Figure 1: In-sample fit for daily realized correlation between Alcoa and Caterpillar.
The gray line represents the daily time series of realized correlations between Alcoa (AA) and Caterpillar (CAT) asset returns, from January 2001 to December 2010. The red line represents the estimated μ_t from our default beta autoregressive model, *without* leverage effects.

6.3 Estimation results

Table 2 presents the parameter estimates obtained from the maximum likelihood method applied to model (3) with conditional mean μ_t that possibly includes leverage effects as specified by (10) or (11). The parameter estimates are reported together with their asymptotic standard error based on $\hat{\Omega}$ in (9). From the estimation results we learn that the model with the leverage effect implied by the S&P500 specification (11) provides the most significant fit for all correlation series. This is most easily verified from the p-values of the likelihood ratio test for the null hypothesis of $\gamma_-^{\text{sp}} = 0$. The estimate of the leverage parameter γ_-^{sp} is in all cases positive, from which we can conclude that the impact of the realized correlation on the S&P500 return is larger when it is negative. When the leverage effect is determined by the returns of both assets, as in the specification (10), the estimate of the leverage parameter γ_- is highly significant for all correlation series, with the only exception of BA/HON. Furthermore, the actual estimates for γ_- are positive in all cases, which is in accordance to the leverage effect. The estimates for γ_+ are in various cases negative but in all cases the corresponding standard errors imply that the estimates are not significantly different from zero (except for IBM/PG). However, the overall leverage effect is present in all cases as the p-values of the likelihood ratio test for $H_0 : \gamma_+ = \gamma_- = 0$ are very small (except for BA/HON).

The estimation results indicate that the size of the leverage effect is on average 10%. It implies that overall the impact of y_t on μ_{t+1} is about 10% larger in periods with negative returns compared to those with positive returns. Furthermore, the results suggest that the leverage effect is better described by the S&P500 returns than the returns of the individual stocks. This final conclusion becomes apparent when comparing the fit of the models using the Akaike information criteria (or others): the maximized log-likelihood value for the model with the S&P500 leverage effect is always the highest and only needs one additional parameter.

6.4 Out-of-sample results

To complete the empirical illustration for our threshold beta observation-driven model, we conduct a forecasting exercise and assess whether the leverage effect improves the out-of-sample performance. We consider both point and density forecasts. The accuracy of point forecasts is evaluated by the root mean squared error (RMSE) of one-step ahead forecasts while the accuracy of density forecasts is evaluated by the log-score criterion of Geweke and Amisano (2011). We obtain one-step-ahead forecasts for realized correlations of 2010, which means that forecasts for 261 trading days are produced. A rolling window estimation approach is considered, where the models are re-estimated for each forecast.

Table 3 reports the results for the three different model specifications: the default model with no leverage effects, model with leverage effects as specified in (10), and model with leverage effects based on S&P500 returns as specified in (11). The results confirm the relevance of the leverage effect in forecasting realized correlation. In particular, it shows also relevance of leverage for out-of-sample analyses. The model with the leverage effect based on S&P500 returns has also the best out-of-sample performance both in terms of point and density forecast precisions for all correlation series (except for JPM/XOM). We also learn from these results that the default model (without leverage effects) shows the worst out-of-sample performance in almost all cases.

7 Conclusion

We have studied a class of observation-driven models for double bounded data with exogenous regressors. We have derived conditions for stationarity and ergodicity of the dynamic process and formally study the asymptotic properties of the maximum likelihood estimator of the parameter

Table 3: Out-of-sample results.

We present the relative root mean squared error ("RMSE") and the log-score criterion ("log-score") for 261 one-step-ahead forecasts in 2010, which are obtained from the three different beta autoregressive model specifications: without leverage ("default"), with leverage ("leverage") as in (10), with leverage based on the S&P500 index ("S&P500") as in (11); default model is the benchmark to calculate the relative RMSE.

	RMSE			log-score		
	default	leverage	S&P500	default	leverage	S&P500
AA/CAT	1.00	**0.99**	**0.99**	1.28	**1.29**	**1.29**
AXP/PFE	1.00	0.99	**0.98**	1.14	1.15	**1.16**
AXP/WMT	1.00	1.00	**0.99**	1.19	1.19	**1.20**
BA/HON	1.00	1.01	**0.99**	1.29	1.28	**1.30**
CAT/KO	1.00	1.00	**0.99**	1.23	1.23	**1.25**
GE/PFE	1.00	0.99	**0.98**	1.12	1.13	**1.15**
HD/JPM	1.00	0.99	**0.98**	1.22	**1.23**	**1.23**
IBM/PG	1.00	1.02	**0.99**	1.21	1.18	**1.22**
JPM/XOM	1.00	**0.98**	0.99	1.39	**1.41**	1.40
MCD/PG	1.00	1.00	**0.99**	1.15	**1.16**	**1.16**

vector. We have applied the theoretical results to study the properties of a threshold autoregressive model. The finite-sample properties of the maximum likelihood estimator are assessed in a Monte Carlo study and they compare well with the corresponding asymptotic properties. Further we have shown that the threshold specification is well suited to introduce a leverage effect in a beta autoregressive model for daily realized correlation time series. In an empirical application using realized correlations of pairs of asset returns, we have shown that the beta autoregressive model with leverage performs convincingly well, both in-sample and out-of-sample. Our approach to the modeling and forecasting of double bounded time series can be explored further in other empirical studies. Extensions towards multivariate bounded data and towards other updating functions for the conditional mean, or other higher order moments, may provide interesting directions for further research.

References

Agosto, A., Cavaliere, G., Kristensen, D., and Rahbek, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *Journal of Empirical Finance*, 38:640–663.

Aknouche, A. and Francq, C. (2018). Count and duration time series with equal conditional stochastic and mean orders. *MPRA paper 90838*.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55.

Aslanidis, N. and Christiansen, C. (2012). Smooth transition patterns in the realized stock–bond correlation. *Journal of Empirical Finance*, 19(4):454–464.

Audrino, F. and Corsi, F. (2010). Modeling tick-by-tick realized correlations. *Computational Statistics & Data Analysis*, 54(11):2372–2382.

Audrino, F. and Trojani, F. (2011). A general multivariate threshold GARCH model with dynamic conditional correlations. *Journal of Business & Economic Statistics*, 29(1):138–149.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Subsampling realised kernels. *Journal of Econometrics*, 160:204–219.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64:253–280.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley, New York, 2nd edition.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019–1052.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Multivariate leverage effects and realized semicovariance GARCH models. *Available at SSRN 3164361*.

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

Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, 4(4):537–572.

Casarin, R., Dalla Valle, L., and Leisen, F. (2012). Bayesian model selection for beta autoregressive processes. *Bayesian Analysis*, 7(2):385–410.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*, 21:520–533.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, 8:93–115.

Davis, R. A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, pages 1673–1707.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Falster, K., Hanly, M., Banks, E., Lynch, J., Chambers, G., Brownell, M., Eades, S., and Jorm, L. (2018). Maternal age and offspring developmental vulnerability at age five: A population-based cohort study of australian children. *PLOS Medicine*, 15(4):1–20.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.

Ghosh, A., Neslin, S., and Shoemaker, R. (1984). A comparison of market share models and estimation procedures. *Journal of Marketing Research*, 21(2):202–210.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Gorgi, P. (2019). BNB autoregressions for modeling integer-valued time series with extreme observations. *arXiv preprint arXiv:1909.02929*.

Gorgi, P., Hansen, P., Janus, P., and Koopman, S. (2019). Realized Wishart-GARCH: a score-driven multi-asset volatility model. *Journal of Financial Econometrics*, 17(1):1–32.

Guolo, A. and Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada Google® Flu Trends. *The Annals of Applied Statistics*, 8(1):74–88.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68:575–603.

Martín, M., Cremades, L. V., and Santabárbara, J. (1999). Analysis and modelling of time series of surface wind speed and direction. *International Journal of Climatology*, 19:197–209.

McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, 2nd edition.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(2):659–680.

Rocha, A. V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models. *Test*, 18(3):529.

Rodríguez, M. J. and Ruiz, E. (2012). Revisiting several popular GARCH models with leverage effect: Differences and similarities. *Journal of Financial Econometrics*, 10(4):637–668.

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.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer-Verlag, New York.

Tong, H. and Lim, K. S. (1980). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient (with discussion). *Journal of the Royal Statistical Society, Series B*, 42(3):245–292.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601.

Zheng, T., Xiao, H., and Chen, R. (2015). Generalized ARMA models with martingale difference errors. *Journal of Econometrics*, 189(2):492–506.

A Appendix

Proof of Theorem 2.1. For notational convenience, we express the stochastic equation in (2) by using the following shorthand notation

$$\mu_{t+1} = g_{\theta,t}(\mu_t),$$

where $g_{\theta,t}(\mu) = g_\theta(\mu, \mathcal{B}(u_t; \mu_t, \phi), x_t)$ for any $\mu \in [0,1]$. Next, for any $t \in \mathbb{Z}$, we define the sequence $\{\mu_t^{(k)}\}_k$ as follows

$$\mu_t^{(k)}(\bar{\mu}) = g_{\theta,t-1}\big(\mu_{t-1}^{(k-1)}(\bar{\mu})\big),$$

where $\mu_t^{(k)} = \bar{\mu} \in [0,1]$ for $k \leq 0$. The process $\mu_t^{(k)}$ can also be expressed as the k-th iteration of the random function $g_{\theta,t}$ starting from the fixed point $\bar{\mu}$, that is,

$$\mu_t^{(k)}(\bar{\mu}) = g_{\theta,t-1} \circ g_{\theta,t-2} \circ \cdots \circ g_{\theta,t-k}(\bar{\mu}).$$

From the above formulation, we immediately obtain that $\mu_t^{(k)}(\bar{\mu})$ is $\tilde{\mathcal{F}}_{t-1}^k$-measurable, where $\tilde{\mathcal{F}}_{t-1}^k$ denotes the σ-field generated by $\{(u_j, x_j^\top) : t-k \leq j \leq t-1\}$. Next, we note that if the limit $\lim_{k\to\infty} \mu_t^{(k)}(\bar{\mu})$ exists almost surely, than equation (2) admits a solution and there exists a measurable function g_θ^∞ such that

$$\mu_t = \lim_{k\to\infty} \mu_t^{(k)}(\bar{\mu}) = g_\theta^\infty(u_{t-1}, u_{t-2}, \dots, x_{t-1}, x_{t-2}, \dots).$$

Therefore, given the stationarity and ergodicity of $\{(u_t, x_t^\top)\}_{t\in\mathbb{Z}}$ ensured by **S2**, we obtain that $\{\mu_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic. More in general, we can conclude that $\{(y_t, \mu_t, x_t)\}_{t\in\mathbb{Z}}$ is stationary and ergodic. Below we show that $\lim_{k\to\infty} \mu_t^{(k)}(\bar{\mu})$ exists almost surely. The proof follows a similar argument as the proof of Proposition 3.1 of Aknouche and Francq (2018).

We note that u_t is independent of $\mu_t^{(k+1)}$ and $\mu_t^{(k)}$ by **S2** since $\mu_t^{(k+1)}$ and $\mu_t^{(k)}$ are $\tilde{\mathcal{F}}_{t-1}^{k+1}$-measurable. Furthermore, accounting that $\mathbb{E}\big(\mathcal{B}(u_t; \mu, \phi)\big) = \mu$, by the stochastic ordering of Lemma A.1 we have that $\mathbb{E}|\mathcal{B}(u_t; \bar{\mu}_1, \phi) - \mathcal{B}(u_t; \bar{\mu}_2, \phi)| = |\bar{\mu}_1 - \bar{\mu}_2|$ for any $\bar{\mu}_1, \bar{\mu}_2 \in (0,1)$. Therefore, we obtain that

$$\begin{aligned}\mathbb{E}\big|\mathcal{B}(u_t; \mu_t^{(k+1)}, \phi) - \mathcal{B}(u_t; \mu_t^{(k)}, \phi)\big| &= \mathbb{E}\Big(\mathbb{E}\Big(\big|\mathcal{B}(u_t; \mu_t^{(k+1)}, \phi) - \mathcal{B}(u_t; \mu_t^{(k)}, \phi)\big|\Big|\tilde{\mathcal{F}}_{t-1}^{k+1}\Big)\Big) \\ &= \mathbb{E}\big|\mu_t^{(k+1)} - \mu_t^{(k)}\big|.\end{aligned}$$

Then, by repeated use of the above equality and the contraction condition in (4), we obtain that

$$
\begin{aligned}
\mathbb{E}\big|\mu_t^{(k+1)}(\bar{\mu}) - \mu_t^{(k)}(\bar{\mu})\big| &\leq a_\mu \mathbb{E}\big|\mu_{t-1}^{(k)}(\bar{\mu}) - \mu_{t-1}^{(k-1)}(\bar{\mu})\big| + b_\mu \mathbb{E}\big|\mathcal{B}(u_{t-1}; \mu_{t-1}^{(k)}, \phi) - \mathcal{B}(u_{t-1}; \mu_{t-1}^{(k-1)}, \phi)\big| \\
&\leq (a_\mu + b_\mu)\mathbb{E}\big|\mu_{t-1}^{(k)}(\bar{\mu}) - \mu_{t-1}^{(k-1)}(\bar{\mu})\big| \\
&\leq (a_\mu + b_\mu)^k \mathbb{E}\big|\mu_{t-k}^{(1)}(\bar{\mu}) - \bar{\mu}\big| \\
&\leq (a_\mu + b_\mu)^k C_\mu,
\end{aligned}
$$

where $C_\mu = \mathbb{E}\big|g_\theta(\bar{\mu}, \mathcal{B}(u_0; \bar{\mu}, \phi), x_0)\big| + \bar{\mu} < \infty$ and $a_\mu + b_\mu < 1$. This implies that the sequence $\{\mu_t^{(k)}\}_k$ converges in L^1-norm at an exponential rate and hence almost surely.

Finally, we conclude the proof of the theorem by showing that that μ_t is $\mathcal{F}_{t-1}$ measurable. We rewrite the stochastic equation of μ_t as

$$\mu_{t+1} = \tilde{g}_{\theta,t}(\mu_t),$$

where $\tilde{g}_{\theta,t}(\mu) = g_\theta(\mu, y_t, x_t)$ and $\{y_t\}_{t\in\mathbb{Z}}$ is the stationary and ergodic solution of the model's equations. We define the sequence $\{\tilde{\mu}_t^{(k)}\}_k$ as follows

$$\tilde{\mu}_t^{(k)}(\bar{\mu}) = \tilde{g}_{\theta,t-1}\big(\tilde{\mu}_{t-1}^{(k-1)}(\bar{\mu})\big),$$

where $\tilde{\mu}_t^{(k)} = \bar{\mu} \in [0,1]$ for $k \leq 0$. The above equation entails that $\tilde{\mu}_t^{(k)}$ is $\mathcal{F}_{t-1}^k$-measurable, where $\mathcal{F}_{t-1}^k$ is the σ-field generated by $\{(y_j, x_j^\top) : t-k \leq j \leq t-1\}$. Similarly as before, we show that $\mu_t = \lim_{k\to\infty} \tilde{\mu}_t^{(k)}(\bar{\mu})$ exists almost surely and therefore μ_t is $\mathcal{F}_{t-1}$-measurable. By the contraction condition in (4), we obtain

$$
\begin{aligned}
\mathbb{E}\big|\tilde{\mu}_t^{(k+1)}(\bar{\mu}) - \tilde{\mu}_t^{(k)}(\bar{\mu})\big| &\leq a_\mu \mathbb{E}\big|\tilde{\mu}_{t-1}^{(k)}(\bar{\mu}) - \tilde{\mu}_{t-1}^{(k-1)}(\bar{\mu})\big| \\
&\leq a_\mu^k \mathbb{E}\big|\tilde{\mu}_{t-k}^{(1)}(\bar{\mu}) - \bar{\mu}\big| \\
&\leq a_\mu^k \tilde{C}_\mu,
\end{aligned}
$$

where $\tilde{C}_\mu = \mathbb{E}\big|\tilde{g}_\theta(\bar{\mu}, y_0, x_0)\big| + \bar{\mu} < \infty$. This implies almost sure convergence and concludes the proof of the theorem. □

Proof of Proposition 3.1. The proof is obtained by an application of Theorem 3.1 of Bougerol (1993) in the space of continuous $\mathbb{C}(\Theta, \mathbb{R})$ equipped with the uniform norm $\|\cdot\|_\Theta$ as considered in Proposition 3.12 Straumann and Mikosch (2006). In particular, the stochastic equation

$$\hat{\mu}_{t+1}(\theta) = g_\theta(\hat{\mu}_t(\theta), y_t, x_t),$$

defines a sequence of functions $\{\hat{\mu}_t(\cdot)\}_{t\in\mathbb{N}}$ in $\mathbb{C}(\Theta,\mathbb{R})$ given the continuity condition in **A2**. We show that conditions C1 and C2 in Theorem 3.1 of Bougerol (1993) are satisfied. Condition C1 holds immediately since g_θ is a bounded function and condition C2 is holds since

$$\frac{|g_\theta(\mu,y_t,x_t)-g_\theta(\mu^*,y_t,x_t)|}{|\mu-\mu^*|}\leq a_\mu<1$$

with probability one for any $\theta\in\Theta$ by **A3**. □

Proof of Theorem 3.1. *Proof of consistency:*
By standard arguments that go back to Wald (1949), given the compactness of the parameter space K, the consistency result follows by showing that

(a) The average log-likelihood converges uniformly to a function $L(\cdot)$, i.e. $\|\hat{L}_T-L\|_K\xrightarrow{a.s.}0$.

(b) The true parameter value κ_0 is the unique maximizer of $L(\kappa)$ in K, i.e. $L(\kappa_0)>L(\kappa)\ \ \forall\ \kappa\in K,\kappa\neq\kappa_0$.

First, we show that (a) holds. An application of the triangle inequality yields

$$|\hat{L}_T(\kappa)-L(\kappa)|\leq|\hat{L}_T(\kappa)-L_T(\kappa)|+|L_T(\kappa)-L(\kappa)|,$$

where $L_T(\kappa)=\frac{1}{T}\sum_{t=1}^T l_t(\kappa)$. Next we show that both terms on the right hand side of the above inequality converge uniformly to zero almost surely. As concerns the first term, **A4** implies that $\tilde{\mu}_t(\theta)\in[\bar{c},1-\bar{c}]$ with probability one for any $\theta\in\Theta$. An application of the mean value theorem yields

$$|\hat{l}_t(\kappa)-l_t(\kappa)|\leq\sup_{\mu\in[\bar{c},1-\bar{c}]}\left|\frac{\partial\log p(y_t|\mu,\phi)}{\partial\mu}\right||\hat{\mu}_t(\theta)-\tilde{\mu}_t(\theta)|.$$

Furthermore, we note that

$$\frac{\partial\log p(y_t|\mu,\phi)}{\partial\mu}=-\phi\psi(\phi\mu)+\psi(\phi(1-\mu))+\phi\log(y_t)-\phi\log(1-y_t),$$

where $\psi(\cdot)$ denotes the digamma function. Therefore, given the continuity of the digamma function in $(0,+\infty)$ and the compactness of Φ, we obtain that

$$\|\hat{l}_t-l_t\|_K\leq\big(c_1-c_2\log(y_t)-c_3\log(1-y_t)\big)\|\hat{\mu}_t-\tilde{\mu}_t\|_\Theta,$$

for some positive constants c_1, c_2 and c_3. Finally, from the properties of the beta distribution, we obtain that

$$\begin{aligned}\mathbb{E}\log(y_t) &= -\mathbb{E}(\mathbb{E}(\log(y_t)|\mathcal{F}_{t-1}))\\ &= \mathbb{E}(\psi(\phi_0) - \psi(\mu_t\phi_0))\\ &\leq \psi(\phi_0) - \psi(\bar{c}\phi_0) < \infty,\end{aligned}$$

and we note that $\mathbb{E}\log(1-y_t) < \infty$ can be shown using the same argument. Therefore, $\|\hat{l}_t - l_t\|_K \xrightarrow{a.s.} 0$ follows by an application of Lemma 2.1 of Straumann and Mikosch (2006) since $\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta \xrightarrow{e.a.s.} 0$ by Proposition 3.1. As a result,

$$\|\hat{L}_T - L_T\|_K \leq \frac{1}{T}\sum_{t=1}^{T}\|\hat{l}_t - l_t\|_K \xrightarrow{a.s.} 0.$$

As concerns the second term, we show that $\|L_T - L\|_K \xrightarrow{a.s.} 0$ by an application of the ergodic theorem of Rao (1962). Given that $\{l_t(\cdot)\}$ is a stationary and ergodic sequence of continuous functions and K is compact, the ergodic theorem follows if $\mathbb{E}\|l_t\|_K < \infty$. We obtain that

$$\begin{aligned}|l_t(\kappa)| &\leq |\log B\big(\phi\tilde{\mu}_t(\theta), \phi(1-\tilde{\mu}_t(\theta))\big)| - \phi\tilde{\mu}_t(\theta)\log(y_t) - \phi(1-\tilde{\mu}_t(\theta))\log(1-y_t)\\ &\leq C - \phi\log(y_t) - \phi\log(1-y_t),\end{aligned}$$

where $\sup_{\phi\in\Phi}|\log B\big(\phi\tilde{\mu}_t(\theta), \phi(1-\tilde{\mu}_t(\theta))\big)| \leq C$ a.s. for some positive constant C since $\tilde{\mu}_t(\theta) \in [\bar{c}, 1-\bar{c}]$ with probability one for any $\theta \in \Theta$ and $B(z_1, z_2)$ is continuous for $z_1, z_2 > 0$. Therefore, we have that $\mathbb{E}\|l_t\|_K < \infty$ since $\mathbb{E}\log(y_t) < \infty$ and $\mathbb{E}\log(1-y_t) < \infty$ as shown before. This concludes the proof of (a).

As concerns (b), following the same argument as in the proof of Theorem 4.1 of Gorgi (2019), we have that (b) is satisfied if we can show that $\tilde{\mu}_t(\theta_1) = \tilde{\mu}_t(\theta_2)$ a.s. if and only if $\theta_1 = \theta_2$ for any $\theta_1, \theta_2 \in \Theta$. The if part of the statement holds trivially. Below we show that the only if part holds by contradiction. Assume that $\theta_1 \neq \theta_2$ and $\tilde{\mu}_t(\theta_1) = \tilde{\mu}_t(\theta_2)$ a.s., then we have that $\tilde{\mu}_t(\theta_1) = \tilde{\mu}_t(\theta_2)$ a.s. for any t since $\tilde{\mu}_t$ is stationary. Therefore, we can assume that $\tilde{\mu}_t(\theta_1) = \tilde{\mu}_t(\theta_2) = \mu$ and it must hold that $g_{\theta_1}(\mu, y_t, x_t) = g_{\theta_2}(\mu, y_t, x_t)$ with probability one. This contradicts **A5**.

Proof of asymptotic normality:

To prove asymptotic normality, we follow a similar argument as in the proof in Section 7 of Straumann and Mikosch (2006). In particular, first, we derive the asymptotic distribution of the

ML estimator $\tilde{\kappa}_T$ based on the limit likelihood L_T, which is defined as

$$\tilde{\kappa}_T = \arg\sup_{\kappa \in K} L_T(\kappa).$$

Then, we show that $\hat{\kappa}_T$ and $\tilde{\kappa}_T$ have the same asymptotic distribution.

The uniform convergence results in Lemma A.2 imply that $\tilde{\mu}_t$ is twice continously differentiable in Θ with first and second derivatives given by $\tilde{\mu}_t'$ and $\tilde{\mu}_t''$, respectively. Therefore, we have that L_T is twice continuously differentiable in K. A Taylor expansion around κ_0 yields

$$L_T'(\tilde{\kappa}_T) = L_T'(\kappa_0) + L_T''(\eta_T)(\tilde{\kappa}_T - \kappa_0),$$

where η_T is a point between $\tilde{\kappa}_T$ and κ_0. By definition, $\tilde{\kappa}_T$ is the maximizer of $L_T(\kappa)$. Therefore, we have that $L_T'(\tilde{\kappa}_T) = 0$ for large enough T since $\tilde{\kappa}_T \xrightarrow{a.s.} \kappa_0$ and $\kappa_0 \in \text{int}(K)$. As a result, the following equation holds true

$$\sqrt{T} L_T'(\kappa_0) = -L_T''(\eta_T)\sqrt{T}(\tilde{\kappa}_T - \kappa_0).$$

By Lemma A.3 together with an application of the ergodic theorem of Rao (1962), we obtain that $-L_T''(\eta_T) \xrightarrow{a.s.} -\mathbb{E}[l_t''(\kappa_0)]$. Furthermore, Lemma A.5 ensures that $-\mathbb{E}[l_t''(\kappa_0)]$ is positive definite and Lemma A.4 shows that $\sqrt{T} L_T'(\kappa_0) \xrightarrow{d} N(0, \Omega^{-1})$. Therefore, we get that

$$\sqrt{T}(\tilde{\kappa}_T - \kappa_0) = \Omega\sqrt{T} L_T'(\kappa_0) + o_p(1),$$

where $\Omega = -\mathbb{E}[l_t''(\kappa_0)]^{-1}$. This implies that $\sqrt{T}(\tilde{\kappa}_T - \kappa_0) \xrightarrow{d} N(0, \Omega)$ as $T \to \infty$.

We conclude the proof by showing that $\hat{\kappa}_T$ and $\tilde{\kappa}_T$ have the same asymptotic distribution. A Taylor expansion yields

$$L_T'(\hat{\kappa}_T) = L_T'(\tilde{\kappa}_T) + L_T''(\tilde{\eta}_T)(\hat{\kappa}_T - \tilde{\kappa}_T),$$

where $\tilde{\eta}_T$ is a point between $\tilde{\kappa}_T$ and $\hat{\kappa}_T$. Furthermore, we note that $\hat{L}_T'(\hat{\kappa}_T) = 0$ and $L_T'(\tilde{\kappa}_T) = 0$ for large enough T by definition. Therefore, we have that

$$\sqrt{T}\big(L_T'(\hat{\kappa}_T) - \hat{L}_T'(\hat{\kappa}_T)\big) = L_T''(\tilde{\eta}_T)\sqrt{T}(\hat{\kappa}_T - \tilde{\kappa}_T).$$

The left hand side of the above equation goes to zero almost surely as $T \to \infty$ by an application of Lemma A.6. Furthermore, Lemma A.3 ensures that $L_T''(\tilde{\eta}_T) \xrightarrow{a.s.} \mathbb{E}[l_t''(\kappa_0)]$. Therefore, we obtain that $\sqrt{T}(\hat{\kappa}_T - \tilde{\kappa}_T) \xrightarrow{a.s.} 0$, which implies that $\hat{\kappa}_T$ and $\tilde{\kappa}_T$ have the same asymptotic distribution. □

Proof of Theorem 4.1. By an application of Theorem 2.1, we have that the result holds true if the contraction condition in (4) is satisfied. From the expression of g_θ in 6, we obtain

$$\begin{aligned}|g_\theta(\mu, y, x_t) - g_\theta(\mu^*, y^*, x_t)| &\leq \beta|\mu - \mu^*| + \left|\gamma + \sum_{j=1}^{J} I(A_j, x_t)\gamma_j\right||y - y^*| \\ &\leq \beta|\mu - \mu^*| + (\gamma + \max\{0, \gamma_1, \ldots, \gamma_J\})|y - y^*|,\end{aligned}$$

where the second equality follows since $\sum_{j=1}^{J} I(A_j, x_t) \in \{0, 1\}$ with probability one. This concludes the proof as $\beta + \gamma + \max\{0, \gamma_1, \ldots, \gamma_J\} < 1$ is implied by the double bound condition in (7). □

Proof of Theorem 4.2. We show that conditions **A1**-**A7** hold. Condition **A1** is implied by (i) and (ii). In particular, (i) imposes that θ_0 satisfies (7), which ensures that the contraction in **S2** holds as shown in the proof of Theorem 4.1. Instead, (ii) directly imposes that the regressors satisfy **S1**. The continuity condition **A2** trivially holds from the expression of the stochastic equation given in (6). Condition **A3** holds since the parameter restrictions in (i) impose that (7) holds and therefore the contraction holds for any $\theta \in \Theta$. Condition **A4** holds since $g_\theta(\mu, y, x)$ is bounded from below by $\delta > 0$ and from above by $\delta + \beta + \gamma + \max\{0, \gamma_1, \ldots, \gamma_J\} < 1$. Condition **A6** trivially holds from the expression of $g_\theta(\mu, y, x)$. Finally, first we prove that **A7** holds and then we show that **A7** implies **A5**. From equation (6), we obtain that

$$\frac{\partial g_\theta}{\partial \theta}(\mu_t, y_t, x_t)\big|_{\theta=\theta_0} = \begin{bmatrix} 1 \\ \mu_t \\ y_t \\ I(A_1, x_t)y_t \\ \vdots \\ I(A_J, x_t)y_t \end{bmatrix}.$$

Therefore, we show that

$$c_1 + c_2\mu_t + c_3 y_t + \sum_{j=1}^{J} c_{3+j} I(A_j, x_t) y_t = 0 \qquad \text{a.s.} \tag{12}$$

only if $\mathrm{c} = 0$, where $\mathrm{c} = (c_1, \ldots, c_{J+3})^\top$. We note that $\mathrm{c} \neq 0$ can occur in one of the following three ways: (a) $(c_1, c_2)^\top \neq 0$ and $(c_3, \ldots, c_{J+3})^\top = 0$, (b) $(c_1, c_2)^\top = 0$ and $(c_3, \ldots, c_{J+3})^\top \neq 0$, and (c) $(c_1, c_2)^\top \neq 0$ and $(c_3, \ldots, c_{J+3})^\top \neq 0$. Below we show that (12) does not hold in each of these three cases. In the case (a) is true, we have that (12) holds only if $c_1 + c_2\mu_t = 0$ a.s. However, this is not true for $(c_1, c_2)^\top \neq 0$ since μ_t is non-degenerate with positive probability. In the case (b) is true, we have that (12) holds only if

$$\Big(c_3 + \sum_{j=1}^{J} c_{3+j} I(A_j, x_t)\Big) y_t = 0 \qquad \text{a.s.},$$

and since $y_t = 0$ with probability zero the above equation is equivalent to

$$c_3 + \sum_{j=1}^{J} c_{3+j} I(A_j, x_t) = 0 \qquad \text{a.s.}$$

Furthermore, given that the sets $A_1, \ldots, A_J$ are disjoint, $c_3 + \sum_{j=1}^{J} c_{3+j} I(A_j, x_t)$ takes values in the set $\{c_3, (c_3 + c_4), \ldots, (c_3 + c_{3+J})\}$ with probability one. Furthermore, condition (ii) ensures that each element of the set $\{c_3, (c_3 + c_4), \ldots, (c_3 + c_{3+J})\}$ occurs with positive probability. This obviously implies that $c_3 + \sum_{j=1}^{J} c_{3+j} I(A_j, x_t) \neq 0$ with positive probability when (b) is true. In the case (c) is true, we have that (12) holds only if

$$c_1 + c_2\mu_t = \Big(c_3 + \sum_{j=1}^{J} c_{3+j} I(A_j, x_t)\Big) y_t \qquad \text{a.s.}$$

However, the above equality does not hold with probability one since the left hand side is $\mathcal{F}_{t-1}$-measurable and instead the right hand side is not measurable with respect to $\mathcal{F}_{t-1}$. This shows that (12) holds only if c $= 0$. Finally, we note that (12) implies that **A5** holds since we can write

$$g_\theta(\mu_t, y_t, x_t) - g_{\theta^*}(\mu_t, y_t, x_t) = \delta - \delta^* + (\beta - \beta^*)\mu_t + \Big((\gamma - \gamma^*) + \sum_{j=1}^{J} (\gamma_j - \gamma_j^*) I(A_j, x_t)\Big) y_t,$$

and therefore $g_\theta(\mu_t, y_t, x_t) = g_{\theta^*}(\mu_t, y_t, x_t)$ a.s. only if $\theta^* = \theta$. □

A.1 Lemmas

Lemma A.1. *Let* $X \sim \mathcal{B}eta\big(\phi\mu_1, \phi(1-\mu_1)\big)$ *and* $Y \sim \mathcal{B}eta\big(\phi\mu_2, \phi(1-\mu_2)\big)$ *with* $\mu_1 \geq \mu_2$. *Then* X *is stochastically greater than* Y, $X \geq_{st} Y$,

$$F_{\mu_1}(z) \leq F_{\mu_2}(z), \quad \textit{for any } z \in \mathbb{N},$$

where F_{μ_1} *and* F_{μ_2} *denote the cumulative distribution functions of* X *and* Y*, respectively.*

Proof. We prove the lemma by showing that X is greater than Y in likelihood ratio order $X \geq_{lr} Y$. Note that likelihood ratio order $X \geq_{lr} Y$ implies stochastic order $X \geq_{st} Y$. The ratio between the density function of X and Y is

$$\text{lr}(z) = \frac{f_{\mu_1}(z)}{f_{\mu_2}(z)} \propto \left(\frac{z}{1-z}\right)^{\phi(\mu_1-\mu_2)}.$$

Therefore, given that $\phi > 0$ and $\mu_1 \geq \mu_2$, we obtain that the likelihood ratio $\text{lr}(z)$ is an increasing function in the interval $(0, 1)$ and hence $X \geq_{lr} Y$. □

Lemma A.2. *Let **A1**-**A7** hold. Then,*

$$\|\hat{\mu}'_t - \tilde{\mu}'_t\|_\Theta \xrightarrow{e.a.s.} 0 \quad \textit{and} \quad \|\hat{\mu}''_t - \tilde{\mu}''_t\|_\Theta \xrightarrow{e.a.s.} 0 \quad \textit{as} \quad t \to \infty,$$

where $\tilde{\mu}'_t(\theta) = \frac{\partial \tilde{\mu}_t(\theta)}{\partial \theta}$ *and* $\tilde{\mu}''_t(\theta) = \frac{\partial^2 \tilde{\mu}_t(\theta)}{\partial \theta \partial \theta^\top}$ *are stationary and ergodic. Furthermore, there is a constant* $C_\mu > 0$ *such that* $\|\hat{\mu}'_t\|_\Theta \leq C_\mu$ *and* $\|\hat{\mu}''_t\|_\Theta \leq C_\mu$ *with probability one. The norm* $\|\cdot\|$ *denotes the* L^1*-norm when applied to a vector and the operator norm induced by the* L^1*-norm when applied to a matrix.*

Proof. The convergence results $\|\hat{\mu}'_t - \tilde{\mu}'_t\|_\Theta \xrightarrow{e.a.s.} 0$ and $\|\hat{\mu}''_t - \tilde{\mu}''_t\|_\Theta \xrightarrow{e.a.s.} 0$ are obtained by showing that the conditions S.1-S.3 of Theorem 2.10 of Straumann and Mikosch (2006) are satisfied.

The derivative process $\tilde{\mu}'_t(\theta)$ can be expressed through the following stochastic equation

$$\tilde{\mu}'_{t+1}(\theta) = g^\theta_t(\theta) + \, g^\mu_t(\theta)\, \tilde{\mu}'_t(\theta),$$

where $g^\mu_t(\theta) = \frac{\partial g_\theta(\mu, y_t, x_t)}{\partial \mu}\big|_{\mu=\tilde{\mu}_t(\theta)}$ and $g^\theta_t(\theta) = \frac{\partial g_\theta(\mu, y_t, x_t)}{\partial \theta}\big|_{\mu=\tilde{\mu}_t(\theta)}$. We note that conditions S.1 and S.2 are immediately satisfied because $\|g^\mu_t\|_\Theta \leq a_\mu < 1$ a.s. by **A2** together with the differentiability conditions in **A6**, and $\|g^\theta_t\|_\Theta$ is bounded by a constant by **A6**. Next we show that S.3 is satisfied. This is the equivalent of showing $\|\hat{g}^\theta_t - g^\theta_t\|_\Theta \xrightarrow{e.a.s.} 0$ and $\|\hat{g}^\mu_t - g^\mu_t\|_\Theta \xrightarrow{e.a.s.} 0$, where $\hat{g}_t^{\mu}(\theta) = \frac{\partial g_\theta(\mu, y_t, x_t)}{\partial \mu}\big|_{\mu=\hat{\mu}_t(\theta)}$ and $\hat{g}_t^{\theta}(\theta) = \frac{\partial g_\theta(\mu, y_t, x_t)}{\partial \theta}\big|_{\mu=\hat{\mu}_t(\theta)}$. By an application of the mean value theorem together with the smoothness conditions in **A6**, there is a constants $c > 0$ such that

$$\|\hat{g}_t^{\theta} - g_t^{\theta}\|_\Theta \leq c\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta \quad \text{and} \quad \|\hat{g}_t^{\mu} - g_t^{\mu}\|_\Theta \leq c\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta.$$

As a result, S.3 is satisfied since $\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta \xrightarrow{e.a.s.} 0$ by Proposition 3.1 and therefore we conclude that $\|\hat{\mu}_t' - \tilde{\mu}_t'\|_\Theta \xrightarrow{e.a.s.} 0$ holds true. Finally, we note that the expression of the stochastic equation of $\tilde{\mu}_t'(\theta)$ together with $\|g_t^{\mu}\|_\Theta$ and $\|g_t^{\theta}\|_\Theta$ being bounded by some constant imply that $\|\tilde{\mu}_t'\|_\Theta$ is bounded with probability one.

Next, we note that the derivative process $\tilde{\mu}_t''(\theta)$ satisfies the following stochastic equation

$$\tilde{\mu}_{t+1}''(\theta) = m_t(\theta) + g_t^{\mu}(\theta)\, \tilde{\mu}_t''(\theta),$$

where

$$m_t(\theta) = g_t^{\theta\theta}(\theta) + g_t^{\theta\mu}(\theta)\, \tilde{\mu}_t'(\theta)^\top + \tilde{\mu}_t'(\theta)\, g_t^{\theta\mu}(\theta)^\top + g_t^{\mu\mu}(\theta)\, \tilde{\mu}_t'(\theta)\tilde{\mu}_t'(\theta)^\top,$$

with $g_t^{\theta\theta}(\theta) = \frac{\partial^2 g_\theta(\mu, y_t, x_t)}{\partial\theta\partial\theta^\top}\big|_{\mu=\tilde{\mu}_t(\theta)}$, $g_t^{\theta\mu}(\theta) = \frac{\partial^2 g_\theta(\mu, y_t, x_t)}{\partial\theta\partial\mu}\big|_{\mu=\tilde{\mu}_t(\theta)}$ and $g_t^{\mu\mu}(\theta) = \frac{\partial^2 g_\theta(\mu, y_t, x_t)}{\partial\mu^2}\big|_{\mu=\tilde{\mu}_t(\theta)}$. First, we obtain that S.1 is satisfied since $\|m_t\|_\Theta$ is bounded by a constant with probability one. In particular, $\|g_t^{\mu\mu}\|_\Theta$, $\|g_t^{\theta\mu}\|_\Theta$ and $\|g_t^{\theta\theta}\|_\Theta$ are bounded by a constant with probability one by **A6**, and $\|\tilde{\mu}_{t+1}'\|_\Theta$ is also bounded by a constant as shown before. Second, S.2 is satisfied since $\|g_t^{\mu}\|_\Theta \leq a_\mu < 1$ a.s. by **A2**. Finally, we have that S.3 is satisfied if

$$\|\hat{g}_t^{\theta\theta} - g_t^{\theta\theta}\|_\Theta \xrightarrow{e.a.s.} 0, \quad \|\hat{g}_t^{\theta\mu}\, \hat{\mu}_t'^\top - g_t^{\theta\mu}\, \tilde{\mu}_t'^\top\|_\Theta \xrightarrow{e.a.s.} 0, \quad \text{and}$$
$$\|\hat{g}_t^{\mu\mu}\, \hat{\mu}_t'\hat{\mu}_t'^\top - g_t^{\mu\mu}\, \tilde{\mu}_t'\tilde{\mu}_t'^\top\|_\Theta \xrightarrow{e.a.s.} 0, \quad \text{as} \quad t \to \infty,$$

where $\hat{g}_t^{\theta\theta}(\theta) = \frac{\partial^2 g_\theta(\mu, y_t, x_t)}{\partial\theta\partial\theta^\top}\big|_{\mu=\hat{\mu}_t(\theta)}$, $\hat{g}_t^{\theta\mu}(\theta) = \frac{\partial^2 g_\theta(\mu, y_t, x_t)}{\partial\theta\partial\mu}\big|_{\mu=\hat{\mu}_t(\theta)}$ and $\hat{g}_t^{\mu\mu}(\theta) = \frac{\partial^2 g_\theta(\mu, y_t, x_t)}{\partial\mu^2}\big|_{\mu=\hat{\mu}_t(\theta)}$. Note that $\|\hat{g}_t^{\mu} - g_t^{\mu}\|_\Theta \xrightarrow{e.a.s.} 0$ holds true as shown before. By **A6** and the mean value theorem, we obtain that there is a constant $c > 0$ such that $\|\hat{g}_t^{\theta\theta} - g_t^{\theta\theta}\|_\Theta \leq c\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta$. Therefore, $\|\hat{g}_t^{\theta\theta} - g_t^{\theta\theta}\|_\Theta \xrightarrow{e.a.s.} 0$ by an application of Proposition 3.1. As concerns $\|\hat{g}_t^{\theta\mu}\, \hat{\mu}_t'^\top - g_t^{\theta\mu}\, \tilde{\mu}_t'^\top\|_\Theta \xrightarrow{e.a.s.} 0$, by condition **A6** and the mean value theorem, we obtain that there is some constant $c > 0$ such that

$$\begin{aligned}
\|\hat{g}_t^{\mu}\, \hat{\mu}_t'^\top - g_t^{\mu}\, \tilde{\mu}_t'^\top\|_\Theta &\leq \|g_t^{\mu}(\hat{\mu}_t'^\top - \tilde{\mu}_t'^\top)\|_\Theta + \|(\hat{g}_t^{\theta\mu} - g_t^{\theta\mu})\, \hat{\mu}_t'^\top\|_\Theta \\
&\leq \|g_t^{\mu}\|_\Theta\|\hat{\mu}_t' - \tilde{\mu}_t'\|_\Theta + \|\hat{\mu}_t'\|_\Theta\|\hat{g}_t^{\theta\mu} - g_t^{\theta\mu}\|_\Theta \\
&\leq c\|\hat{\mu}_t' - \tilde{\mu}_t'\|_\Theta + c\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta.
\end{aligned}$$

Therefore, the result follows since $\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta$ and $\|\hat{\mu}'_t - \tilde{\mu}'_t\|_\Theta$ converge to zero e.a.s. As concerns $\|\hat{g}^{\mu\mu}_t \ \hat{\mu}'_t\hat{\mu}'^{\top}_t - g^{\mu\mu}_t \ \tilde{\mu}'_t\tilde{\mu}'^{\top}_t\|_\Theta \xrightarrow{e.a.s.} 0$, similarly as before, we obtain that

$$\begin{aligned}\|\hat{g}^{\mu\mu}_t \ \hat{\mu}'_t\hat{\mu}'^{\top}_t - g^{\mu\mu}_t \ \tilde{\mu}'_t\tilde{\mu}'^{\top}_t\|_\Theta &\le \|g^{\mu\mu}_t \ (\hat{\mu}'_t\hat{\mu}'^{\top}_t - \tilde{\mu}'_t\tilde{\mu}'^{\top}_t)\|_\Theta + \|(\hat{g}^{\mu\mu}_t - g^{\mu\mu}_t)\hat{\mu}'_t\hat{\mu}'^{\top}_t\|_\Theta \\ &\le \|g^{\mu\mu}_t\|_\Theta \ \|\hat{\mu}'_t\hat{\mu}'^{\top}_t - \tilde{\mu}'_t\tilde{\mu}'^{\top}_t\|_\Theta + \|\hat{\mu}'_t\|^2_\Theta\|\hat{g}^{\mu\mu}_t - g^{\mu\mu}_t\|_\Theta \\ &\le c\|\hat{\mu}'_t - \tilde{\mu}'_t\|_\Theta + c\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta.\end{aligned}$$

The result then follows since $\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta$ and $\|\hat{\mu}'_t - \tilde{\mu}'_t\|_\Theta$ converge to zero e.a.s. Finally, we note that $\|g^{\mu}_t\|_\Theta$ and $\|m_t\|_\Theta$ are bounded by some constant and therefore, given the expression of the stochastic equation of $\tilde{\mu}''_t$, we obtain that $\|\tilde{\mu}''_t\|_\Theta$ is bounded with probability one. This concludes the proof of the theorem. □

Lemma A.3. *Let **A1**-**A7** hold, then the second derivative of the likelihood function has a uniformly bounded moment, i.e.* $\mathbb{E}\|l''_t\|_K < \infty$.

Proof. The second derivatives of the likelihood is

$$l''_t(\kappa) = \begin{bmatrix} l^{\phi\phi}_t(\kappa) & l^{\phi\mu}_t(\kappa) \ \tilde{\mu}'_t(\theta)^\top \\ \tilde{\mu}'_t(\theta) \ l^{\phi\mu}_t(\kappa)^\top & l^{\mu\mu}_t(\kappa) \ \tilde{\mu}'_t(\theta) \ \tilde{\mu}'_t(\theta)^\top + l^{\mu}_t(\kappa)\tilde{\mu}''_t(\theta)\end{bmatrix},$$

where $l^{\mu}_t(\kappa) = \frac{\partial \log p(y_t|\mu;\phi)}{\partial\mu}\big|_{\mu=\tilde{\mu}_t(\theta)}$, $l^{\phi\phi}_t(\kappa) = \frac{\partial^2 l_t(\kappa)}{\partial\phi^2}$, $l^{\phi\mu}_t(\kappa) = \frac{\partial^2 \log p(y_t|\mu;\phi)}{\partial\phi\partial\mu}\big|_{\mu=\tilde{\mu}_t(\theta)}$, and $l^{\mu\mu}_t(\kappa) = \frac{\partial^2 \log p(y_t|\mu;\phi)}{\partial\mu^2}\big|_{\mu=\tilde{\mu}_t(\theta)}$. By Lemma A.2 and the submultiplicativity of the norm $\|\cdot\|_\Theta$, we obtain that there is a constant $c > 0$ such that

$$\begin{aligned}\mathbb{E}\|l''_t\|_K \le \ &\mathbb{E}\|l^{\phi\phi}_t\|_K + 2\,\mathbb{E}\|l^{\phi\mu}_t \ \tilde{\mu}'^{\top}_t\|_K + \mathbb{E}\|l^{\mu\mu}_t \ \tilde{\mu}'_t \ \tilde{\mu}'^{\top}_t\|_K + \mathbb{E}\|l^{\mu}_t \ \tilde{\mu}''_t\|_K \\ \le \ &\mathbb{E}\|l^{\phi\phi}_t\|_K + 2c\,\mathbb{E}\|l^{\phi\mu}_t\|_K + c\,\mathbb{E}\|l^{\mu\mu}_t \ \|_K + c\,\mathbb{E}\|l^{\mu}_t\|_K.\end{aligned}$$

Therefore, $\mathbb{E}\|l''_t\|_K < \infty$ holds true if $\mathbb{E}\|l^{\phi\phi}_t\|_K < \infty$, $\mathbb{E}\|l^{\phi\mu}_t\|_K < \infty$, $\mathbb{E}\|l^{\mu\mu}_t\|_K < \infty$ and $\mathbb{E}\|l^{\mu}_t\|_K < \infty$. We note that, up to some additive constant, the expression of the likelihood function is

$$\begin{aligned}l_t(\kappa) = &\log\Gamma(\phi) - \log\Gamma(\phi\tilde{\mu}_t(\theta)) - \log\Gamma(\phi(1-\tilde{\mu}_t(\theta))) \\ &+ \phi\tilde{\mu}_t(\theta)\log(y_t) + \phi(1-\tilde{\mu}_t(\theta))\log(1-y_t),\end{aligned}$$

where $\Gamma(\cdot)$ denotes the gamma function. Furthermore, $\tilde{\mu}_t(\theta) \in [\bar{c}, 1-\bar{c}]$ for any $\theta \in \Theta$ by **A1** with $\bar{c} > 0$. Therefore, since the log-gamma function is twice continuously differentiable in the set of positive real numbers we obtain that first and second partial derivatives of the first three terms in the expression of $l_t(\kappa)$ are uniformly bounded with probability one. Similarly, we obtain that first and second partial derivatives of the last three terms in the expression of $l_t(\kappa)$ have a uniformly bounded first moment since $\mathbb{E}|\log(y_t)| < \infty$ and $\mathbb{E}|\log(1-y_t)| < \infty$ as shown in the proof of Theorem 2.1. □

Lemma A.4. *Let **A1**-**A7** hold, then the following asymptotic result holds true*

$$\sqrt{T}L_T'(\kappa_0) \xrightarrow{d} N\big(0, \mathbb{E}[l_t'(\kappa_0)l_t'(\kappa_0)^\top]\big), \quad T \to \infty.$$

Proof. The first derivative of $l_t(\kappa)$ evaluated at $\kappa = \kappa_0$ can be expressed as

$$l_t'(\kappa_0) = \begin{bmatrix} l_t^{\phi}(\kappa_0) \\ l_t^{\mu}(\kappa_0)\tilde{\mu}_t'(\theta_0) \end{bmatrix},$$

where $l_t^{\mu}(\kappa_0) = \frac{\partial \log p(y_t|\mu;\phi_0)}{\partial \mu}\big|_{\mu=\tilde{\mu}_t(\theta_0)}$ and $l_t^{\phi}(\kappa_0) = \frac{\partial l_t(\kappa_0)}{\partial \phi}$. First, we obtain that $l_t'(\kappa_0)$ has a finite second moment by showing that $\mathbb{E}\|l_t'(\kappa_0)\|^2 < \infty$. The proof follows as in the proof of Lemma A.3 by noticing the first and second derivatives of the log-likelihood are uniformly bounded by a linear combination of $\log(y_t)$ and $\log(1-y_t)$. Therefore, $\mathbb{E}\|l_t'(\kappa_0)\|^2 < \infty$ holds if $\mathbb{E}\log^2(y_t) < \infty$ and $\mathbb{E}\log^2(1-y_t) < \infty$. We can show that $\log(y_t)$ has a finite second moment by relying on the properties of the beta distribution. In particular, $\mathbb{V}ar(\log(y_t)|\mathcal{F}_{t-1}) = \psi_1(\phi_0\mu_t) - \psi_1(\phi_0)$, where $\psi_1(\cdot)$ denotes the trigamma function. Therefore, $\psi_1(\phi_0\mu_t)$ is bounded by a constant with probability one since $\mu_t \in [\bar{c}, 1-\bar{c}]$ a.s. and the trigamma function is continuous in $\mathbb{R}^+$. This implies that the unconditional second moment is finite. Finiteness of $\mathbb{E}\log^2(1-y_t)$ can be obtain by an application of the same argument. Next, we obtain that $\mathbb{E}[l_t'(\kappa_0)|\mathcal{F}_{t-1}] = 0$. We note that $\tilde{\mu}_t'(\theta_0)$ is $\mathcal{F}_{t-1}$-measurable and therefore

$$\mathbb{E}[l_t'(\kappa_0)|\mathcal{F}_{t-1}] = \begin{bmatrix} \mathbb{E}[l_t^{\phi}(\kappa_0)|\mathcal{F}_{t-1}] \\ \mathbb{E}[l_t^{\mu}(\kappa_0)|\mathcal{F}_{t-1}]\ \tilde{\mu}_t'(\theta_0) \end{bmatrix}.$$

The result is obtained noticing that $\mathbb{E}[l_t^{\mu}(\kappa_0)|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[l_t^{\phi}(\kappa_0)|\mathcal{F}_{t-1}] = 0$ since $l_t^{\mu}(\kappa_0)$ and $l_t^{\phi}(\kappa_0)$ are conditional scores of the beta log-density evaluated at the true parameter vector (μ_t, ϕ_0).

The above results imply that $\{l_t'(\kappa_0)\}_{t\in\mathbb{N}}$ is a stationary and ergodic martingale difference sequence with finite second moment. Therefore, we conclude that

$$\sqrt{T}L_T'(\theta_0) \xrightarrow{d} N\left(0, \mathbb{E}[l_t'(\kappa_0)l_t'(\kappa_0)^\top]\right), \quad \text{as} \quad T \to \infty,$$

by an application of the Central Limit theorem for stationary and ergodic martingale difference sequences, see Billingsley (1999). $\square$

Lemma A.5. *Let **A1**-**A7** hold, then the Fisher information matrix is positive definite, i.e.* $-\mathbb{E}[l_t''(\kappa_0)] = \mathbb{E}[l_t'(\kappa_0)l_t'(\kappa_0)^\top] > 0$.

Proof. We note that it is obvious that $\mathbb{E}[l_t'(\kappa_0)l_t'(\kappa_0)^\top]$ is positive semi-definite. Therefore, we only need to show that $\mathrm{w}^\top l_t'(\kappa_0) = 0$ a.s. only if $\mathrm{w} = 0$ to rule out the possibility that $\mathbb{E}[l_t'(\kappa_0)l_t'(\kappa_0)^\top]$ is singular. Consider the partition $\mathrm{w} = (\mathrm{w}_1, \mathrm{w}_{-1}^\top)^\top$, where $\mathrm{w}_1 \in \mathbb{R}$ and $\mathrm{w}_{-1} \in \mathbb{R}^k$. We have that

$$\mathrm{w}^\top l_t'(\kappa_0) = \mathrm{w}_1 \frac{\partial l_t(\kappa_0)}{\partial \phi} + \mathrm{w}_{-1}^\top \frac{\partial l_t(\kappa_0)}{\partial \theta}.$$

Below, we show that $\mathrm{w}^\top l_t'(\kappa_0) = 0$ a.s. implies $\mathrm{w} = 0$. Having $\mathrm{w} \neq 0$ can occur in one of the following ways: (i) $\mathrm{w}_1 \neq 0$ and $\mathrm{w}_{-1} = 0$, (ii) $\mathrm{w}_1 = 0$ and $\mathrm{w}_{-1} \neq 0$, and (iii) $\mathrm{w}_1 \neq 0$ and $\mathrm{w}_{-1} \neq 0$. In the case (i) holds, we have that $\mathrm{w}^\top l_t'(\kappa_0) = 0$ a.s. holds only if $\frac{\partial l_t(\kappa_0)}{\partial \phi} = 0$ with probability one. However, it is trivial to see that $\frac{\partial l_t(\kappa_0)}{\partial \phi} \neq 0$ with positive probability. As concerns (ii), we obtain that $\mathrm{w}^\top l_t'(\kappa_0) = 0$ a.s. is satisfied only if

$$\mathrm{w}_{-1}^\top \frac{\partial l_t(\kappa_0)}{\partial \theta} = l_t^\mu(\kappa_0)\, \mathrm{w}_{-1}^\top \frac{\partial \tilde{\mu}_t(\theta_0)}{\partial \theta} = 0 \quad \text{a.s.},$$

where $l_t^\mu(\kappa_0) = \frac{\partial \log p(y_t|\mu;\phi_0)}{\partial \mu}\big|_{\mu=\tilde{\mu}_t(\theta_0)}$. Given that $l_t^\mu(\kappa_0) \neq 0$ with probability one, the above equality would imply that $\mathrm{w}_{-1}^\top \frac{\partial \tilde{\mu}_t(\theta_0)}{\partial \theta} = 0$ a.s. However, this possibility is ruled out by **A7**. Finally, if (iii) holds, then $\mathrm{w}^\top l_t'(\kappa_0) = 0$ a.s. is satisfied only if

$$\mathrm{w}_1 \frac{\partial l_t(\kappa_0)}{\partial \phi} \left(l_t^\mu(\kappa_0)\right)^{-1} = -\mathrm{w}_{-1}^\top \frac{\partial \tilde{\mu}_t(\theta_0)}{\partial \theta} \quad \text{a.s.}$$

However, this equation does not hold because $\frac{\partial \tilde{\mu}_t}{\partial \theta}$ is $\mathcal{F}_{t-1}$-measurable and, instead, the term on the left hand side is not $\mathcal{F}_{t-1}$-measurable as it depends on y_t.

Finally, the Fisher information equality $-\mathbb{E}[l_t''(\kappa_0)] = \mathbb{E}[l_t'(\kappa_0)l_t'(\kappa_0)^\top]$ holds true. In particular, $l_t(\kappa_0)$ is the true conditional log-density evaluated at y_t and the likelihood function is twice continuously differentiable with a uniformly bounded moment. This enables interchanging integration and differentiation, which entails the desired result by following standard arguments. $\square$

Lemma A.6. *Let **A1**-**A7** hold, then*

$$\sqrt{T}\,\|\hat{L}'_T - L'_T\|_K \xrightarrow{a.s.} 0, \quad \textit{as} \quad T \to \infty.$$

Proof. From the expression of l'_t given in the proof of Lemma A.4, the submultiplicativity of matrix norms and the mean value theorem, we obtain

$$\begin{aligned}
\|\hat{l}'_t - l'_t\|_K &\leq \|\hat{l}^{\phi}_t - l^{\phi}_t\|_K + \|\hat{l}^{\mu}_t\,\hat{\mu}'_t - l^{\mu}_t\,\tilde{\mu}'_t\|_K \\
&\leq \|\hat{l}^{\phi}_t - l^{\phi}_t\|_K + \|l^{\mu}_t(\hat{\mu}'_t - \tilde{\mu}'_t)\|_K + \|(\hat{l}^{\mu}_t - l^{\mu}_t)\hat{\mu}'_t\|_K \\
&\leq \|v^{\phi\mu}_t\|_\Phi\,\|\hat{\mu}_t - \tilde{\mu}_t\|_K + \|l^{\mu}_t\|_K\,\|\hat{\mu}'_t - \tilde{\mu}'_t\|_\Theta + \|\hat{\mu}'_t\|_\Theta\,\|v^{\mu\mu}_t\|_\Phi\,\|\hat{\mu}_t - \tilde{\mu}_t\|_\Theta,
\end{aligned}$$

where $v^{\phi\mu}_t(\phi) = \sup_{\mu\in[\bar{c},(1-\bar{c})]}\left|\frac{\partial^2 \log p(y_t|\mu;\phi)}{\partial\phi\partial\mu}\right|$ and $v^{\mu\mu}_t(\phi) = \sup_{\mu\in[\bar{c},(1-\bar{c})]}\left|\frac{\partial^2 \log p(y_t|\mu;\phi)}{\partial\mu^2}\right|$. Next, we note that $\mathbb{E}\|l^{\mu}_t\|_K < \infty$, $\mathbb{E}\|v^{\phi\mu}_t\|_\Phi < \infty$ and $\mathbb{E}\|v^{\mu\mu}_t\|_\Phi < \infty$ since l^{μ}_t, $v^{\phi\mu}_t$ and $v^{\mu\mu}_t$ are bounded by linear combinations of $\log(y_t)$ and $\log(1-y_t)$, which have a bounded moment. Furthermore, $\|\hat{\mu}'_t\|_\Theta$ is bounded by a constant with probability one by Lemma A.2. Therefore, $\|\hat{l}'_t - l'_t\|_K \xrightarrow{e.a.s.} 0$ follows by an application of Lemma 2.1 of Straumann and Mikosch (2006) since $\|\hat{\mu}_t - \tilde{\mu}_t\|_K \xrightarrow{e.a.s.} 0$ and $\|\hat{\mu}'_t - \tilde{\mu}'_t\|_K \xrightarrow{e.a.s.} 0$ by Proposition 3.1 and Lemma A.2, respectively. Finally, we conclude the proof noticing that

$$\lim_{T\to\infty} T\,\|\hat{L}'_T - L'_T\|_K \leq \sum_{i=1}^{T}\|\hat{l}'_t - l'_t\|_K < \infty \quad \text{a.s.}$$

Therefore, $\sqrt{T}\,\|\hat{L}'_T - L'_T\|_K \xrightarrow{a.s.} 0$ holds true. □