Asymptotic Theory for Beta-t-GARCH

Ryoko Ito*

Nuffield College and the Department of Economics, Oxford University

Cambridge Working Papers in Economics CWPE1607

This version: May 24, 2016

ABSTRACT

The dynamic conditional score (DCS) models with variants of Student's t innovation are gaining popularity in volatility modeling, and studies have found that they outperform GARCH-type models of comparable specifications. DCS is typically estimated by the method of maximum likelihood, but there is so far limited asymptotic theories for justifying the use of this estimator for non-Gaussian distributions. This paper develops asymptotic theory for Beta-t-GARCH, which is DCS with Student's t innovation and the benchmark volatility model of this class. We establish the necessary and sufficient condition for strict stationarity of the first-order Beta-t-GARCH using one simple moment equation, and show that its MLE is consistent and asymptotically normal under this condition. The results of this paper theoretically justify applying DCS with Student's t innovation to heavy-tailed data with a high degree of kurtosis, and performing standard statistical inference for model selection using the estimator. Since GARCH is Beta-t-GARCH with infinite degrees of freedom, our results imply that Beta-t-GARCH can capture the size of the tail or the degree of kurtosis that is too large for GARCH.

KEYWORDS: robustness; score; consistency; asymptotic normality.

JEL Classification: C22, C58

*Email: ryoko.ito@economics.ox.ac.uk. I would like to thank Andrew Harvey for his helpful comments and guiding my research. I would also like to thank Francisco Blasques, Axel Gottfries, Søren Johansen, Oliver Linton, Andre Lucas, Vladimirs Murevics, Bent Nielsen, Rasmus Pedersen, Anders Rahbek, Mark Salmon, and Stephen Thiele for providing thoughtful comments. Finally, I would like to thank the Institute for New Economic Thinking at the Oxford Martin School, the Cambridge Trust, the International Monetary Fund, the Keynes Fund, and the Stevenson Fund (of the Faculty of Economics at Cambridge University) for providing various forms of funding.

1. INTRODUCTION

Asset price volatility is a key ingredient in many aspects of finance, including risk management, security pricing, and trading derivatives that are written on volatility. It is also monitored as an indicator of vulnerability in financial markets and used in assessing the portfolio risks of financial institutions by policymakers. The generalized autoregressive heteroscedasticity (GARCH) model introduced by Engle (1982) and Bollerslev (1986) is perhaps the most popular mode of forecasting volatility. Its empirical success stems from its simplicity, practicality, and intuitive structure. The dynamic equation in GARCH is a linear function of squared observations and analogous to the sample variance formula, which is an efficient estimator for Gaussian data. Studies have found that the widely-documented non-Gaussian, particularly heavy-tailed, features of financial returns determine the efficiency, robustness, and asymptotic normality of (quasi-)maximum likelihood estimators ((Q)MLE) in GARCH. An enormous GARCH literature dedicated to developing robust-modifications of it have highlighted the need for a simple, practical, and unified approach to modeling heavy-tailed data. The dynamic conditional score (DCS) model[1] developed by Creal et al. (2011, 2013) and Harvey (2013) takes a step in the direction called for above.

DCS is a new class of simple and elegant observation-driven model that encompasses as its special or limiting cases popular existing models including GARCH, the autoregressive conditional duration (ACD) or intensity (ACI) models by Engle and Russell (1998) and Russell (1998), and Poisson count models with dynamic mean (see Davis et al. (2005)). DCS models with variants of Student's t innovation are gaining popularity in volatility forecasting and contrast with GARCH models of comparable specifications.[2] DCS is also extended to time-varying copula functions, non-negative distributions, and multivariate distributions.[3] These studies find that DCS captures heavy-tails well and outperforms existing, particularly GARCH-type, forecasting methodologies in respective applications. Successful applications of DCS abound: they include high-frequency trade volume prediction, inflation forecasting in macroeconomics, forecasting value-at-risk, modeling credit or sovereign-default

[1]DCS is also called the generalized autoregressive score (GAS) model.

[2]See, for instance, Harvey and Chakravarty (2008), Harvey and Sucarrat (2014), Janus et al. (2014), Harvey and Lange (2015), Gao and Zhou (2016), and Lucas and Zhang (2016).

[3]See, for instance, Creal et al. (2011), Ito (2013, 2016), Avdulaj and Barunik (2015), and Salvatierra and Patton (2015).

risk, modeling mixed-measurement and mixed-frequency panel data, and dynamic location modeling.[4]

In these applications, DCS is typically estimated by the method of maximum likelihood, but there is so far limited asymptotic theories for justifying the use of this estimator for non-Gaussian distributions. In this paper, we develop asymptotic theory for Beta-t-GARCH, which is a version of DCS with Student's t innovation and a benchmark volatility model in DCS capable of capturing heavy-tails in asset returns. In particular, we derive the necessary and sufficient condition for strict stationarity of the first-order Beta-t-GARCH, and show that its MLE is consistent and asymptotically normal under this condition. We show that the asymptotic normality does not require the fourth moment of the error distribution to be finite. The only restriction instead is a finite second moment since Beta-t-GARCH is a volatility model.[5] The results of this paper theoretically justify applying Beta-t-GARCH to heavy-tailed data with a high degree of kurtosis, and performing standard statistical inference for model selection using the estimator. Since GARCH is Beta-t-GARCH with infinite degrees of freedom, our results also imply that Beta-t-GARCH can capture the size of the tail or the degree of kurtosis that is too large for GARCH.

QMLE is perhaps the most popular mode of estimation for GARCH when the data exhibits non-Gaussian features. The properties of this estimator are studied by Lumsdaine (1996), Lee and Hansen (1994), Boussama (2000), Berkes et al. (2003), Francq and Zakoïan (2004), Straumann and Mikosch (2006), and Jensen and Rahbek (2004), among others. Also see Francq and Zakoïan (2010, Ch. 7 and 9) for a review. The asymptotic normality of the estimator fails (for both the strictly stationary and the nonstationary cases) when the fourth moment of the error distribution is not finite. Moreover, the efficiency of the estimator is determined by how far the data is from normality. These are relevant issues in high-frequency finance as emphasized above, since it primarily deals with heavy-tailed data with a high degree of kurtosis. See, for instance, Bollerslev and Wooldridge (1992), Caviano and Harvey (2013a, 2013b), and Ibragimov et al. (2013).

The dynamic equation of GARCH is sensitive to large-sized observations since it is a linear function of past squared observations. Its robust modifications typically take approaches classified as Winsorising or trimming. However, many

[4]See, for instance, Creal et al. (2014), Harvey and Luati (2014), Lucas et al. (2014), and Caviano and Harvey (2014), as well as the above references.

[5]Even this moment assumption may be relaxed if we reformulate the Student's t likelihood to model dynamic scale instead of volatility.

robust-GARCH models are still found to lack robustness against isolated additive outliers. See, for instance, Li et al. (2010), Park (2002), and Muler and Yohai (2008). To deal with non-Gaussianity, some attention in this literature has shifted to nonparametric procedures or the use of non-Gaussian likelihood in (Q)MLE. See, for instance, Hall and Yao (2003), Francq et al. (2011), and Fan et al. (2014). The asymptotic properties of MLE in GARCH with Student's t is shown by Berkes and Horváth (2004), Straumann (2005, Ch. 6), and Pedersen and Rahbek (2016),[6] but its dynamic equation is still sensitive to large-sized observations.

In contrast to the dynamic equation in question, a novel feature of DCS is that the score function of the error distribution drives its dynamic equation. Thus, the choice of likelihood influences the sensitivity of time-varying parameters to large-sized observations. For instance, the score of Student's t dampens the effect of large-sized observations when the distribution is heavy-tailed. This is also how applying the model to non-Gaussian distributions leads to unified formulations of different observation-driven models including, and not limited to, ACI, ACD, and Poisson count models as mentioned above.

Harvey (2013) studies the asymptotic properties of MLE in weakly stationary DCS with the exponential link function and a selection of well-known heavy-tailed distributions. This version of DCS can be compared with EGARCH of Nelson (1991). Blasques et al. (2014) derive a list of sufficient conditions of the error distribution for MLE in DCS to be consistent and asymptotically normal. In contrast, we derive the necessary and sufficient condition of the parameter space for which Beta-t-GARCH(1,1) is strictly stationary. Thus, we identify the parameter space for Beta-t-GARCH(1,1) that is larger than implied in the results by Blasques et al. (2014). The condition we derive is given by one simple and explicit moment equation.

We impose very mild assumptions on the parameter space: the first moment (or the location parameter, denoted by γ) of observations is finite, the dynamic parameters are bounded and strictly positive (so that volatility is positive), and the second moment of Student's t is finite. To derive asymptotic normality, we also assume that the persistence parameter on the lagged conditional variance (denoted by β) is less than one.[7] This upper-bound is also a standard

[6]The asymptotic normality does not require the fourth moment of the error distribution is finite.

[7]See Lemmas 15-16 in Appendix D.

assumption in the GARCH literature.[8] In the strictly stationary case, apart from the initial volatility parameter (ω), consistency and asymptotic normality are established for all other parameters. In the nonstationary case, we conjecture from simulation results that the asymptotic results can be established for three parameters: β and the coefficients on the lagged score (α and the degrees of freedom parameter, ν). We think that the asymptotic results do not hold for γ and the intercept parameter (δ) in the nonstationary case since the second derivatives of the log-likelihood with respect to these parameters collapse to zero asymptotically (see Lemma 11). These parameters are not identified when the process loses ergodicity and stationarity.[9] These findings of this paper are reinforced by our simulation results in Section 5.

2. THE MODEL

The first-order Beta-t-GARCH is given by

$$\begin{aligned} y_t &= \gamma_0 + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_{0t}} z_t, \quad b_{0t} = \frac{z_t^2}{z_t^2 + (\nu_0 - 2)}, \\ h_{0t} &= \delta_0 + \beta_0 h_{0t-1} + \alpha_0 (\nu_0 + 1) h_{0t-1} b_{0t-1}, \end{aligned} \tag{1}$$

for $t \in \mathbb{N}_{>0}$. Each z_t is assumed to be an independently and identically distributed (i.i.d.) Student's t random variable with the degrees of freedom parameter, ν_0, and the first two moments, $\mathbb{E}[z_t] = 0$, and $\mathbb{E}[z_t^2] = 1$. Beta-t-GARCH encompasses GARCH as its limiting case for when $\nu_0 \to \infty$. This is because Student's t becomes standard normal and $(\nu_0 + 1) b_{0t} \to \varepsilon_t^2 / h_{0t} = z_t^2$ as $\nu_0 \to \infty$. Then the dynamic equation becomes $h_{0t} = \delta_0 + \beta_0 h_{0t-1} + \alpha_0 \varepsilon_{t-1}^2$.

This model is nested in the first-order DCS, which is

$$\begin{aligned} y_t &= \gamma_0 + \varepsilon_t, \quad \varepsilon_t = s_{0t} z_t, \quad s_{0t} = \mathsf{link}(h_{0t}), \\ h_{0t} &= \delta_0 + \phi_0 h_{0t-1} + \kappa_0 u_{t-1} \end{aligned} \tag{2}$$

for $t = 1, \ldots, n$. $\mathsf{link}(\cdot)$ denotes some link function with the canonical link parameter, h_{0t}. z_t is some i.i.d. centered standard random variable, and its

[8]See, for instance, Lee and Hansen (1994) and Berkes et al. (2003). The consistency of global QMLE in Lee and Hansen (1994) also requires $\alpha + \beta < 1$, where α and β are the coefficients on the lagged squared observation and the lagged conditional variance, respectively. The (strong) consistency of QMLE in GARCH(p, q) requires that the coefficients on the lags of conditional variance sum to less than one (i.e. $\sum_{j=1}^{p} \beta_j < 1$), which is implied by the strict stationarity assumption (Corollary 2.3 of Bougerol and Picard (1992)).

[9]This compares with the results by Jensen and Rahbek (2004), which require γ, δ and ω to be fixed and known. The QMLE of δ is found to be inconsistent in nonstationary GARCH$(1, 1)$. (See the discussions in Jensen and Rahbek (2004) and Francq and Zakoïan (2010, p. 180)).

distribution characterizes the conditional distribution, $F_y(\cdot)$, with the density, $f_y(\cdot)$, of an observation y_t. u_t is the score of $f_y(\cdot)$ standardized by the conditional Fisher information quantity:

$$u_t = \frac{\partial \log f_y(y_t)}{\partial h_{0t}} \Bigg/ \mathbb{E}\left[-\frac{\partial^2 \log f_y(y_t)}{\partial h_{0t}^2} \middle| \mathcal{F}_{t-1} \right],$$

where $\mathcal{F}_{t-1}$ denotes the set of information available at time $t-1$.

Beta-t-GARCH is DCS with the centered standard Student's t distribution and the square-root link function, in which case

$$u_t = \frac{h_{0t}}{2}\left(\frac{(\nu_0+1)\varepsilon_t^2}{(\nu_0-2)h_{0t}+\varepsilon_t^2} - 1 \right) = \frac{h_{0t}}{2}\left((\nu_0+1)b_{0t}-1\right).$$

This gives (1) by setting $\phi_0 = \alpha_0 + \beta_0$ and $\alpha_0 = \kappa_0/2$. A higher-order DCS model includes additional lags of h_{0t} and u_t in the dynamic equation of (2). If no lags of h_{0t} are included, the model becomes analogous to the ARCH model of Engle (1982).

3. STATIONARITY AND ERGODICITY

Our analysis is conditional on the initial value $y_0 \in \mathbb{R}$. The initial value of the conditional variance, h_{0t}, is parameterized by $h_{00} = \omega_0 \in \mathbb{R}_{>0}$, where $\mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$. The vector of true parameters are denoted by $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top \in \Theta$, where

$$\begin{aligned}\Theta = \{\theta \in \mathbb{R}^6 : 2 < \nu_l \le \nu \le \nu_u < \infty, 0 < \alpha_l \le \alpha \le \alpha_u < \infty, 0 < \beta_l \le \beta \le \beta_u < \infty,\\ 0 < \delta_l \le \delta \le \delta_u < \infty, -\infty < \gamma_l \le \gamma \le \gamma_u < \infty, 0 < \omega_l \le \omega \le \omega_u < \infty\}.\end{aligned}$$

Assuming that θ_0 is unknown, we estimate the model,

$$\begin{aligned} &y_t = \gamma + e_t, \quad h_t(\theta) = \delta + \beta h_{t-1}(\theta) + \alpha(\nu+1)h_{t-1}(\theta)b_{t-1}(\theta), \\ &b_t(\theta) = \frac{e_t^2}{e_t^2 + (\nu-2)h_t(\theta)}, \end{aligned} \tag{3}$$

for $\theta \in \Theta$, where $t \in \mathbb{N}_{\ge 0}$ and $\theta = (\nu, \alpha, \beta, \delta, \gamma, \omega)^\top \in \Theta$. At $\theta = \theta_0$, we have $h_t(\theta_0) = h_{0t}$ and $b_t(\theta_0) = b_{0t}$. Furthermore, we split Θ into two regions;

$$\begin{aligned}\Theta_L &\equiv \{\theta_0 \in \Theta : \mathbb{E}\left[\ln(\beta_0 + \alpha_0(\nu_0+1)b_{0t})\right] < 0\},\\ \Theta_U &\equiv \{\theta_0 \in \Theta : \mathbb{E}\left[\ln(\beta_0 + \alpha_0(\nu_0+1)b_{0t})\right] \ge 0\}.\end{aligned}$$

THEOREM 1. *If $\theta_0 \in \Theta_L$, $(h_{0t})_{t\in\mathbb{N}_{>0}}$ is strictly stationary and ergodic with a well-defined probability measure μ_∞ on (δ_0, ∞). If $\theta_0 \in \Theta_U$, $(h_{0t})_{t\in\mathbb{N}_{>0}}$ is divergent almost surely (a.s.) and its reciprocal converges to zero a.s. as well as in L^p for any $p \ge 1$ as $t \to \infty$.*

When $\nu_0 \to \infty$, Nelson (1990) shows that $\mathbb{E}[\ln(\beta_0 + \alpha_0 z_t^2)] < 0$ is the necessary and sufficient condition for the true GARCH$(1,1)$ process to be strictly stationarity. Theorem 1 is consistent with this since $(\nu_0 + 1)b_{0t}$ becomes z_t^2 when ν_0 is large.

Note that, if z_t is centered standard Student's t, b_{0t} is a beta random variable with shape parameters $(1/2, \nu_0/2)$ (denoted by Beta$(1/2, \nu_0/2)$). These distributions are defined in Appendix A. Using this property, it may be possible to rewrite $\mathbb{E}\left[\ln(\beta_0 + \alpha_0(\nu_0+1)b_{0t})\right]$ as an explicit function of β_0, α_0, and ν_0. Then we know the exact values of these parameters for which h_{0t} is strictly stationary.

In what follows, the i-th element of $\theta = (\nu, \alpha, \beta, \delta, \gamma, \omega)^\top \in \Theta$ may be denoted by θ_i for $i = 1, 2, \ldots, 6$, so that $\theta_1 \equiv \nu$ and so on. Define

$$h_{\theta_i t}(\theta) \equiv \frac{\partial h_t(\theta)}{\partial \theta_i}\frac{1}{h_t(\theta)}, \quad h_{\theta_i\theta_j t}(\theta) \equiv \frac{\partial^2 h_t(\theta)}{\partial \theta_i \partial \theta_j}\frac{1}{h_t(\theta)}, \quad h_{\theta_i\theta_j\theta_k t}(\theta) \equiv \frac{\partial^3 h_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}\frac{1}{h_t(\theta)}$$

for $i, j, k = 1, 2, \ldots, 6$, so that $h_{\beta t}(\theta) = h_t(\theta)^{-1}(\partial h_t(\theta)/\partial \beta)$ and so on. The analytic expressions for the derivatives of the log-likelihood, as well as $h_{\theta_i t}(\theta)$, $h_{\theta_i\theta_j t}(\theta)$, and $h_{\theta_i\theta_j\theta_k t}(\theta)$ for $i, j, k = 1, 2, \ldots, 6$ are given in Appendix B.1. We also set $\prod_{j=1}^{0} \cdot = 1$ for notational convenience.

Next, we consider the stationarity property of the log-likelihood function. Given a finite sequence of observations $(y_t)_{t=1}^n$ for some $n \in \mathbb{N}_{>0}$, the log-likelihood for the Beta-t-GARCH model is

$$L_n(\theta) = n^{-1}\sum_{t=1}^{n} l_t(\theta),$$

where

$$l_t(\theta) \equiv \ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \frac{1}{2}\ln(\nu-2) - \frac{1}{2}\ln(\pi) - \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\ln\left(h_t(\theta)\right) - \frac{\nu+1}{2}\ln\left(1 + \frac{e_t^2}{(\nu-2)h_t(\theta)}\right).$$

Theorem 2 establishes the stationarity and ergodicity properties of the log-likelihood function and its first two derivatives with respect to θ evaluated at $\theta = \theta_0 \in \Theta_L$.

THEOREM 2. *If $\theta_0 \in \Theta_L$, $(l_t(\theta_0))_{t\in\mathbb{N}}$ and its first two derivatives of $(l_t(\theta))_{t\in\mathbb{N}}$ with respect to θ evaluated at $\theta = \theta_0$, denoted by $(\nabla_\theta l_t(\theta_0))_{t\in\mathbb{N}}$ and $(\nabla_\theta^2 l_t(\theta_0))_{t\in\mathbb{N}}$, are strictly stationary and ergodic.*

4. CONSISTENCY AND ASYMPTOTIC NORMALITY

We assume that the true initial value of the volatility process is known (i.e. $h_0(\theta) \equiv \omega = \omega_0 \equiv h_{0t}$) throughout Section 4. Thus, throughout Section 4, we reduce the dimension of the parameter space to

$$\theta = (\nu, \alpha, \beta, \delta, \gamma)^\top \in \Theta \subset \mathbb{R}^5,$$

where the dimension of Θ is adjusted accordingly. We write $L_n(\theta) \equiv L_n(\delta, \alpha, \beta, \gamma, \nu, \omega_0)$, and likewise for the single log-likelihood function, $l_t(\cdot)$, and $h_t(\cdot)$.

Theorem 3 states that the MLE of θ_0 is consistent and asymptotically normal when $\theta_0 \in \Theta_L$, $\omega = \omega_0$, and $z_t \sim$ i.i.d. Student's t_{ν_0} with $\nu_0 > 2$.

Define $Q(\theta_0) \equiv \mathbb{E}[\nabla^2_\theta l_t(\theta_0)]$ and $R(\theta_0) \equiv \mathbb{E}[\nabla_\theta l_t(\theta_0) \nabla_\theta l_t(\theta_0)^\top]$. The existence of these moments are by Lemmas 6 and 9. These notations mean that the derivatives of the log-likelihood function are taken with respect to the free parameters only, i.e.

$$\nabla_\theta l_t(\theta_0) = (\partial l_t(\theta_0)/\partial\nu, \partial l_t(\theta_0)/\partial\alpha, \partial l_t(\theta_0)/\partial\beta, \partial l_t(\theta_0)/\partial\delta, \partial l_t(\theta_0)/\partial\gamma).$$

We denote the convergence in probability by $\xrightarrow{P}$ and in distribution by $\xrightarrow{\mathcal{D}}$. We use $\|\cdot\|_p$ for $p \geq 1$ to denote the L^p-norm on $(\Omega, \mathcal{F}, \mathbb{P})$.

THEOREM 3 (Consistency and Asymptotic Normality of MLE). *Suppose θ_0 is an interior point of Θ. Assume that $\theta_0 \in \Theta_L$. Assume also that $z_t \sim$ i.i.d. Student's t_{ν_0} with $\nu_0 > 2$. Then, with probability tending to one, there exists a unique maximum point $\widehat{\theta}_n$ of $L_n(\theta)$ such that $\widehat{\theta}_n \xrightarrow{\mathcal{P}} \theta_0$ and*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, V(\theta_0))$$

as $n \to \infty$, where $V(\theta_0) \equiv Q(\theta_0)^{-1} R(\theta_0) Q(\theta_0)^{-1} = Q(\theta_0)^{-1}$.

Our asymptotic results are derived using Lemma 1 of Jensen and Rahbek (2004). A notable feature of this lemma is the condition (A.3), which requires the third derivative of log-likelihood to be bounded in some neighborhood of true parameter values by a process that is convergent in probability. In contrast, Lumsdaine (1996) and Lee and Hansen (1994) apply the uniform convergence results of functionals by Andrews (1987 or 1992), and the convergence results, described in Amemiya (1985), for the maximizer of a function defined over a compact parameter space. The use of the results in Amemiya (1985) in establishing asymptotic normality require the authors to prove that the third-derivative of log-likelihood is bounded in L^1 over some admissible parameter region. We found that the parameters for the unconditional mean of observations (γ_0) and the degrees of freedom (ν_0) were the most difficult to handle.

In this paper, we do not show that the above results are asymptotically independent of the value of ω in the strictly stationary case. We will consider this in our future research.

4.1. Remark on QMLE

In the GARCH literature, QMLE using Student's t likelihood function has been proposed to deal with non-Gaussianity in the data. See, for instance, Fan et al. (2014) and Francq et al. (2011). Non-Gaussian QMLE in GARCH is possible partly because the score is a martingale difference as long as the second moment of the error distribution exists. We do not think that non-Gaussian QMLE in Beta-t-GARCH using Student's t likelihood is generally possible. This is because there is no guarantee that there exists a value of ν_0 for which the score variable, u_t, is a martingale difference when z_t is not Student's t.

5. Simulation results

We simulate the asymptotic distribution of MLE and check its large-sample behavior. We generate $K = 500$ sets of data from the model, (1) and compute MLE at each simulation. The sample length at each simulation is up to $n = 5,000$.

The simulation results in Table 1 suggest that biases and the size of errors generally decrease as sample size increases, suggesting consistency.[10] The 95% coverage probabilities seem to validate standard statistical inference for model selection using this estimator at sample size as large as $n = 5,000$. The Kolmogorov-Smirnov (KS) statistics testing the null of Gausianity of MLE at $n = 5,000$ are outside the rejection region at the 5% level for all parameters except for ν in the stationary case 1. But this result for ν is largely due to the fact that ν is difficult to estimate when it is large, and appears inconsequential given the farily close coverage probability.

[10] The median bias and absolute deviation quantities are more reliable than the mean bias or squared error quantities, as we do not know the existence of these moments in small sample.

Following Theorem 1, we check stationarity of generated series by computing the sample counterpart of $\mathbb{E}[\beta_0 + \alpha_0(\nu_0)b_t(\theta_0)]$ at each simulation. If the resulting 500 sample means for a given θ_0 are comfortably in the negative (positive) region, that θ_0 gives a stationary (nonstationary) case. Figure 1 shows that that the first two cases of θ_0 in Table 1 comfortably give stationary cases (i.e. $\theta_0 \in \Theta_L$), and the last case of θ_0 in the same table comfortably gives a nonstationary case (i.e. $\theta_0 \in \Theta_U$).

Thus, the simulation results support Theorem 3. When $\theta_0 \in \Theta_U$, the simulation results suggest that MLE for ν, α, and β are consistent and asymptotically normal. We found that these results do not hold for γ and δ when the process is nonstationary since the second derivatives of the log-likelihood with respect to these parameters collapse to zero asymptotically (see Lemma 11). These parameters are not identified when the process loses ergodicity and stationarity.

The selected three cases of θ_0 are roughly in line with parameter estimates we would obtain in empirical applications. For instance, if we fit the first-order Beta-t-GARCH model to the daily returns (computed using the daily closing level) of the Dow Jones Industrial Average index between 1 October 1975 and 5 July 2011, we obtain $\widehat{\theta} = (\widehat{\nu}, \widehat{\alpha}, \widehat{\beta}, \widehat{\delta}, \widehat{\gamma}, \widehat{\omega})^\top = (5.96, 0.13, 0.83, 0.04, 0.05, 6.2)^\top$. The parameter values are chosen also to ensure that the sample counterpart of $\mathbb{E}[\beta_0 + \alpha_0(\nu_0)b_t(\theta_0)]$ is comfortably and always positive or always negative at each simulation, so that we can be farily certain about the stationarity of generated data.

(a) Stationary case 1 ($\omega = 1$)

	True value	Mean bias				MSE				KS test (p-val.)
n		500	1000	2000	5000	500	1000	2000	5000	5000
ν	10	0.9	0.6	0.7	0.4	12.5	9.0	8.7	2.2	0.0
α	0.13	0.003	0.001	0.000	0.000	0.004	0.002	0.001	0.000	0.735
β	0.7	-0.064	-0.024	-0.015	-0.004	0.065	0.017	0.007	0.002	0.172
δ	1	0.34	0.13	0.08	0.02	1.56	0.37	0.15	0.04	0.616
γ	0	-0.002	0.001	-0.001	-0.001	0.009	0.005	0.003	0.001	0.648

	Median bias				Med. abs. dev.				95% cov. prob.			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
ν	0.2	-0.1	0.1	0.2	2.2	1.7	1.3	0.8	0.97	0.93	0.92	0.93
α	-0.001	0.001	0.001	0.000	0.042	0.029	0.019	0.013	0.93	0.94	0.95	0.95
β	-0.019	-0.010	-0.007	-0.005	0.098	0.066	0.051	0.031	0.89	0.89	0.91	0.95
δ	0.07	0.04	0.04	0.01	0.41	0.27	0.21	0.13	0.90	0.91	0.92	0.95
γ	-0.001	-0.003	0.001	-0.001	0.065	0.049	0.034	0.021	0.97	0.94	0.95	0.95

(b) Stationary case 2 ($\omega = 1$)

	True value	Mean bias				MSE				KS test (p-val)
n		500	1000	2000	5000	500	1000	2000	5000	5000
ν	5	0.7	0.3	0.1	0.0	3.8	0.9	0.4	0.1	0.5
α	0.12	-0.001	0.001	0.002	0.001	0.005	0.002	0.001	0.000	0.426
β	0.7	-0.073	-0.034	-0.021	-0.008	0.082	0.028	0.011	0.004	0.158
δ	2	0.78	0.37	0.22	0.08	7.72	2.76	0.86	0.28	0.107
γ	0	-0.001	0.000	0.000	-0.001	0.014	0.008	0.004	0.002	0.836

	Median bias				Med. abs. dev.				95% cov. prob.			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
ν	0.2	0.2	0.1	0.0	0.9	0.5	0.4	0.2	0.89	0.93	0.94	0.94
α	-0.003	-0.002	-0.002	-0.001	0.048	0.031	0.020	0.013	0.92	0.92	0.95	0.94
β	-0.023	-0.014	-0.006	-0.003	0.115	0.086	0.062	0.037	0.88	0.91	0.93	0.93
δ	0.07	0.10	0.08	0.00	0.91	0.73	0.50	0.32	0.87	0.90	0.93	0.93
γ	0.001	0.000	0.000	0.000	0.081	0.061	0.043	0.028	0.97	0.94	0.95	0.95

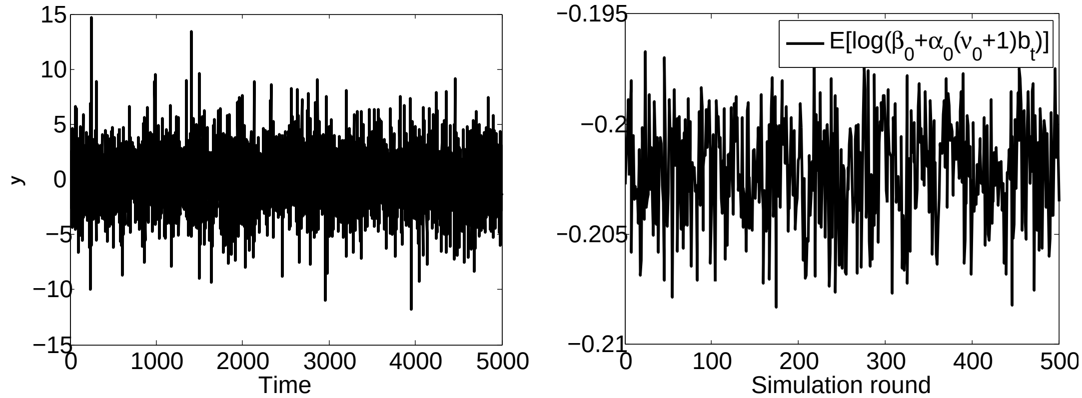
(c) Nonstationary case ($\omega = 1, \delta = 1, \gamma = 0$)

	True value	Mean bias				MSE				KS test (p-val)
n		500	1000	2000	5000	500	1000	2000	5000	5000
ν	10	0.8	0.6	0.6	0.3	12.4	8.6	6.6	1.9	0.1
α	0.18	-0.009	-0.004	-0.003	-0.001	0.002	0.001	0.000	0.000	0.641
β	0.86	0.008	0.004	0.003	0.001	0.001	0.000	0.000	0.000	0.250

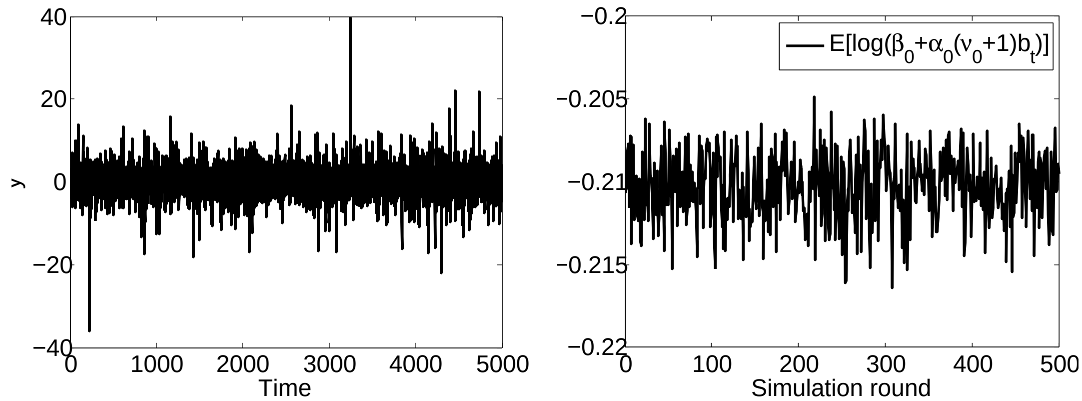
	Median bias				Med. abs. dev.				95% cov. prob.			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
ν	0.2	0.0	0.0	0.1	2.2	1.7	1.2	0.8	0.95	0.93	0.93	0.94
α	-0.008	-0.005	-0.004	-0.001	0.025	0.019	0.013	0.008	0.95	0.95	0.94	0.96
β	0.009	0.004	0.003	0.001	0.020	0.015	0.011	0.006	0.95	0.95	0.95	0.96

Table 1 Selected statistics from the simulated asymptotic distribution of MLE. The sample length of $n = 5,000$ is simulated $K = 500$ times and the MLE is computed at each simulation to simulate its asymptotic distribution. The KS statistics test the null of Gaussianity as in Theorem 3 with $n = 5000$. The in-sample computation of the information quantity takes the sample mean of the outer-product of the first derivative of the log-likelihood.

(a) Stationary case 1: $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (10, 0.13, 0.7, 1, 0, 1)^\top$



(b) Stationary case 2: $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (5, 0.12, 0.7, 2, 0, 1)^\top$



(c) Nonstationary case: $\theta_0 = (\nu_0, \alpha_0, \beta_0, \delta_0, \gamma_0, \omega_0)^\top = (10, 0.18, 0.86, 1, 0, 1)^\top$

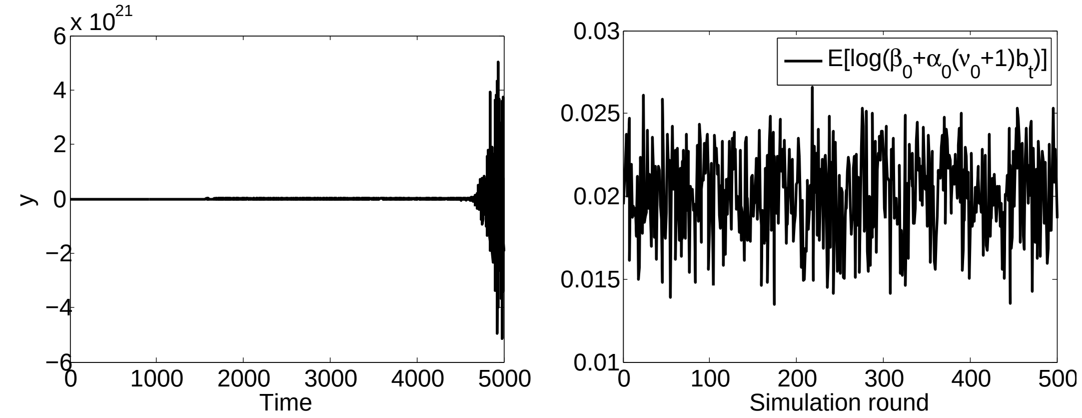


Figure 1 The time series plot of $(y_t)_{t=1}^n$ (left) and $n^{-1}\sum_t \log(\beta_0+\alpha_0(\nu_0+1)b_{0t})$ at each of the $K = 500$ simulations (right) when $n = 5{,}000$. We should have $\theta_0 \in \Theta_L$ for the top two cases and $\theta_0 \in \Theta_U$ in the last case, since this sample mean quantity is always comfortably in the negative or positive region.

REFERENCES

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

Andrews, D. W. (1987), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

— (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

Avdulaj, K. and Barunik, J. (2015), "Are Benefits From Oil-Stocks Diversification Gone? New Evidence From a Dynamic Copula and High Frequency Data," *Energy Economics*, 51, 31–44.

Berkes, I. and Horváth, L. (2004), "The Efficiency of the Estimator of the Parameters in GARCH Processes," *The Annals of Statistics*, 32, 633–655.

Berkes, I., Horváth, L., and Kokoszka, P. (2003), "Asymptotics for GARCH Squared Residual Correlations," *Econometric Theory*, 19, 515–540.

Billingsley, P. (1986), *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, Wiley: New York, 2nd ed.

Blasques, F., Koopman, S. J., and Lucas, A. (2014), "Maximum Likelihood Estimation for Generalized Autoregressive Score Models," Tinbergen Institute Discussion Paper TI2014-029/III, Tinbergen Institute.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T. and Wooldridge, J. M. (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time-Varying Covariances," *Econometric Reviews*, 11, 143–172.

Bougerol, P. and Picard, N. (1992), "Stationarity of GARCH Processes and of Some Nonnegative Time Series," *Journal of Econometrics*, 52, 115–127.

Boussama, F. (2000), "Normalité Asymptotique de l'Estimateur du Pseudo-Maximum de Vraisemblance d'un Modéle GARCH," *Comptes Rendus de l'Académie de Sciences Paris*, 331, 81–84.

Brandt, A. (1986), "The Stochastic Equation Yn+1=AnYn+Bn with Stationary Coefficients," *Advances in Applied Probability*, 18, 211–220.

Caviano, M. and Harvey, A. C. (2013a), "Time Series Models with an EGB2 Conditional Distribution," Cambridge Working Papers in Economics CWPE1325, University of Cambridge.

— (2013b), "Two EGARCH Models and One Fat Tail," Cambridge Working Papers in Economics CWPE1326, University of Cambridge.

— (2014), "Time-Series Models With an EGB2 Conditional Distribution," *Journal of Time Series Analysis*, 35, 558–571.

Creal, D. D., Koopman, S. J., and Lucas, A. (2011), "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations," *Journal of Business and Economic Statistics*, 29, 552–563.

— (2013), "Generalized Autoregressive Score Models with Applications," *Journal of Applied Econometrics*, 28, 777–795.

Creal, D. D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014), "Observation-Driven Mixed-Measurement Dynamic Factor Models With An Application To Credit Risk," *The Review of Economics and Statistics*, 96, 898–915.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2005), "Maximum Likelihood Estimation for an Observation Driven Model for Poisson Counts," *Methodology and Computing in Applied Probability*, 7, 149–159.

Engle, R. F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Engle, R. F. and Russell, J. R. (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Fan, J., Qi, L., and Xiu, D. (2014), "Quasi-Maximum Likelihood Estimation of GARCH Models With Heavy-Tailed Likelihoods," *Journal of Business and Economic Statistics*, 32, 178–191.

Francq, C., Lepage, G., and Zakoïan, J.-M. (2011), "Tow-Stage Non Gaussian QML Estimation of GARCH models and Testing the Efficiency of the Gaussian QMLE," *Journal of Econo*, 165, 246–257.

Francq, C. and Zakoïan, J.-M. (2004), "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637.

— (2010), *GARCH Models: Structure, Statistical Inference and Financial Applications*, Wiley.

Gao, C.-T. and Zhou, X.-H. (2016), "Forecasting VaR and ES Using Dynamic Conditional Score Models and Skew Student Distribution," *Economic Modelling*, 53, 216–223.

Hall, P. and Yao, Q. (2003), "Inference in ARCH and GARCH Models with Heavy-Tailed Errors," *Econometrica*, 71, 285–317.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Harvey, A. C. (2013), *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Econometric Society Monograph, Cambridge University Press.

Harvey, A. C. and Chakravarty, T. (2008), "Beta-t-(E)GARCH," Cambridge Working Papers in Economics CWPE0840, University of Cambridge.

Harvey, A. C. and Lange, R.-J. (2015), "Volatility Modeling with a Generalized t-Distribution," Cambridge Working Papers in Economics CWPE1517, University of Cambridge.

Harvey, A. C. and Luati, A. (2014), "Filtering With Heavy Tails," *Journal of the American Statistical Association*, 109, 1112–1122.

Harvey, A. C. and Sucarrat, G. (2014), "EGARCH Models With Fat Tails, Skewness and Leverage," *Computational Statistics and Data Analysis*, 76, 320–338.

Ibragimov, M., Ibragimov, R., and Kattuman, P. (2013), "Emerging Markets and Heavy Tails," *Journal of Banking and Finance*, 37, 2546–2559.

Ito, R. (2013), "Modeling Dynamic Diurnal Patterns in High Frequency Financial Data," Cambridge Working Papers in Economics CWPE1315, University of Cambridge.

— (2016), "Spline-DCS for Forecasting Trade Volume in High-Frequency Financial Data," Cambridge Working Papers in Economics CWPE1606, University of Cambridge.

Janus, P., Koopman, S. J., and Lucas, A. (2014), "Long Memory Dynamics For Multivariate Dependence Under Heavy Tails," *Journal of Empirical Finance*, 29, 187–206.

Jensen, S. T. and Rahbek, A. (2004), "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

Lee, S.-W. and Hansen, B. E. (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Li, J., Kao, C., and Zhang, W. D. (2010), "Bounded Influence Estimator for GARCH Models: Evidence from Foreign Exchange Rates," *Applied Economics*, 42, 1437–1445.

Lucas, A., Schwaab, B., and Zhang, X. (2014), "Conditional Euro Area Sovereign Default Risk," *Journal of Business and Economic Statistics*, 32, 271–284.

Lucas, A. and Zhang, X. (2016), "Score-Driven Exponentially Weighted Moving Averages and Value-At-Risk Forecasting," *International Journal of Forecasting*, 32, 293–302.

Lumsdaine, R. L. (1996), "Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575–96.

Muler, N. and Yohai, V. J. (2008), "Robust Estimates for GARCH Models,"

Journal of Statistical Planning and Inference, 138, 2918–2940.
Nelson, D. B. (1990), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.
— (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.
Park, B. (2002), "An Outlier Robust GARCH Model and Forecasting Volatility of Exchange Rate Returns," *Journal of Forecasting*, 21, 381–393.
Pedersen, R. S. and Rahbek, A. (2016), "Nonstationary GARCH with t-distributed innovations," *Economics Letters*, 138, 19–21.
Royden, H. L. (1988), *Real Analysis*, Macmillan: New York, 3rd ed.
Russell, J. R. (1998), "Econometric Modelling of Multivariate Irregularly-Spaced High Frequency Data," Tech. rep., Graduate School of Business, University of Chicago.
Salvatierra, I. D. L. and Patton, A. J. (2015), "Dynamic Copula Models and High Frequency Data," *Journal of Empirical Finance*, 30, 120–135.
Stout, W. F. (1974), *Almost Sure Convergence*, Probability and mathematical statistics, Academic Press: New York.
Straumann, D. (2005), "Estimation in Conditionally Heteroscedastic Time Series Models," in *Lecture Notes in Statistics*, Springer, Heidelberg.
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, 2449–2495.
Varadhan, S. R. S. (2001), *Probability Theory*, Courant Lecture Notes, Courant Institute of Mathematical Sciences.

APPENDIX A: Student's t-distribution

The standard Student's t-distribution has the probability density function (pdf),

$$f(x;\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}}\left(1+\frac{x^2}{(\nu-2)}\right)^{-(\nu+1)/2}, \quad x \in \mathbb{R},\ \nu > 2,\ h > 0,$$

where $\nu > 0$ is the degrees of freedom and $\Gamma(\cdot)$ is the gamma function. The mean is 0 and variance is 1.

If a random variable Y follows the standard Student's t-distribution after it is standardized by a scaling parameter $h > 0$, the pdf of Y denoted by $f_Y : \mathbb{R}_{>0} \to \mathbb{R}$ is $f_Y(y; h, \nu) = f(y/h; \nu)/h$ for $y \in \mathbb{R}$. Since the variance of the standard Student's t-distribution is normalized, the variance of Y is determined by h and not by ν. For a set of i.i.d. observations $y_1, \ldots, y_N$ where each follows the non-standardized Student's t, the log-likelihood function of a single observation y_t can be written as:

$$\log f_Y(y_t) = \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\log((\nu-2)\pi h) - \frac{\nu+1}{2}\log\left(1+\frac{y_t^2}{(\nu-2)h}\right).$$

The *score* of f_Y (i.e. the first derivative with respect to h) computed at y_t and standardized by the Fisher information, h^{-2}, is

$$\frac{\partial \log f_Y(y_t)}{\partial h} = \frac{h}{2}\left(\frac{(\nu+1)y_t^2}{(\nu-2)h+y_t^2} - 1\right) = \frac{h}{2}\left((\nu+1)b_t(\nu) - 1\right) \qquad \text{(A.1)}$$

where we used the notation $b_t(\nu) \equiv y_t^2/((\nu-2)h + y_t^2)$. It is easy to check that the mean of (A.1) is zero. By the properties of Student's t, $b_t(\nu)$ follows the beta distribution with parameters $(1/2, \nu/2)$. The beta distribution with parameters (α, β) characterized by the pdf is

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1],\ \alpha,\beta > 0.$$

where $B(\cdot,\cdot)$ is the beta function. We denote this distribution by $\mathrm{Beta}(\alpha,\beta)$.

APPENDIX B: Functions and Equations

B.1. Derivatives of $l_t(\theta)$

The first three derivatives of $l_t(\theta)$ with respect to β are

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \beta} &= \frac{1}{2}h_{\beta t}(\theta)\left[(\nu+1)b_t(\theta) - 1\right], \\
\frac{\partial^2 l_t(\theta)}{\partial \beta^2} &= \frac{1}{2}h_{\beta t}(\theta)^2\left[(\nu+1)b_t(\theta)(b_t(\theta)-2) + 1\right] \qquad \text{(B.1)} \\
&\quad + \frac{1}{2}h_{\beta\beta t}(\theta)\left[(\nu+1)b_t(\theta) - 1\right], \\
\frac{\partial^3 l_t(\theta)}{\partial \beta^3} &= (\nu+1)b_t(\theta)(1-b_t(\theta))\left[h_{\beta t}(\theta)^3 - \frac{3}{2}h_{\beta t}(\theta)h_{\beta\beta t}(\theta)\right] \qquad \text{(B.2)} \\
&\quad + (\nu+1)h_{\beta t}(\theta)^3 b_t(\theta)(1-b_t(\theta))^2 \\
&\quad - \frac{1}{2}\left(3h_{\beta t}(\theta)h_{\beta\beta t}(\theta) - 2h_{\beta t}(\theta)^3 - h_{\beta\beta\beta t}(\theta)\right)\left[(\nu+1)b_t(\theta) - 1\right].
\end{aligned}$$

Using the fact that

$$\frac{\partial b_t(\theta)}{\partial \beta} = -b_t(\theta)(1-b_t(\theta))h_{\beta t}(\theta),$$
$$\frac{\partial^2 b_t(\theta)}{\partial \beta^2} = 2b_t(\theta)\left(1-b_t(\theta)\right)^2 h_{\beta t}(\theta)^2 - b_t(\theta)\left(1-b_t(\theta)\right) h_{\beta\beta t}(\theta),$$

recursive substitution gives

$$\begin{aligned}
h_{\beta t}(\theta) &= \sum_{k=1}^{t} \frac{\widehat{h}_{\beta t-k}(\theta)}{\beta+\alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)(\beta+\alpha(\nu+1)b_{t-j}(\theta)^2)}{h_{t-j+1}(\theta)}, \\
h_{\beta\beta t}(\theta) &= \sum_{k=1}^{t-1} \frac{\widehat{h}_{\beta\beta t-k}(\theta)}{\beta+\alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)(\beta+\alpha(\nu+1)b_{t-j}(\theta)^2)}{h_{t-j+1}(\theta)}, \\
h_{\beta\beta\beta t}(\theta) &= \sum_{k=1}^{t-1} \frac{\widehat{h}_{\beta\beta\beta t-k}(\theta)}{\beta+\alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)\left(\beta+\alpha(\nu+1)b_{t-j}(\theta)^2\right)}{h_{t-j+1}(\theta)},
\end{aligned}$$

where

$$\begin{aligned}
\widehat{h}_{\beta t}(\theta) &= 1 \\
\widehat{h}_{\beta\beta t}(\theta) &= 2h_{\beta t}(\theta)\left[1-\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\beta t}(\theta)\right], \\
\widehat{h}_{\beta\beta\beta t}(\theta) &= 3h_{\beta\beta t}(\theta) - 6\alpha(\nu+1)b_t(\theta)^2\left(1-b_t(\theta)\right)^2 h_{\beta t}(\theta)^3 \\
&\quad + 2\alpha(\nu+1)b_t(\theta)^2\left(1-b_t(\theta)\right) h_{\beta t}(\theta)h_{\beta\beta t}(\theta).
\end{aligned}$$

With respect to other parameters, we have

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \delta} &= \frac{1}{2}h_{\delta t}(\theta)\left((\nu+1)b_t(\theta)-1\right), \\
\frac{\partial l_t(\theta)}{\partial \alpha} &= \frac{1}{2}h_{\alpha t}(\theta)\left((\nu+1)b_t(\theta)-1\right), \\
\frac{\partial l_t(\theta)}{\partial \gamma} &= \frac{1}{2}\left((\nu+1)b_t(\theta)-1\right)h_{\gamma t}(\theta) + (\nu+1)\frac{e_t}{e_t^2+(\nu-2)h_t(\theta)}, \\
\frac{\partial l_t(\theta)}{\partial \nu} &= \frac{1}{2}\left(\psi_0\left(\frac{\nu+1}{2}\right)-\psi_0\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}\ln\left(1+\frac{e_t^2}{(\nu-2)h_t(\theta)}\right) \\
&\quad + \frac{(\nu+1)b_t(\theta)-1}{2(\nu-2)} + \frac{(\nu+1)b_t(\theta)-1}{2}h_{\nu t}(\theta),
\end{aligned}$$

where $\psi_0(\cdot)$ is the digamma function. By recursion, we have

$$h_{\theta_i t}(\theta) = \sum_{k=1}^{t} \frac{\widehat{h}_{\theta_i t-k}(\theta)}{\beta+\alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{\left(\beta+\alpha(\nu+1)b_{t-j}(\theta)^2\right)h_{t-j}(\theta)}{h_{t-j+1}(\theta)}, \qquad \text{(B.3)}$$

for $i = 1, 2, \ldots, 6$, where

$$\begin{aligned}
\widehat{h}_{\delta t}(\theta) &= 1/h_t(\theta) \\
\widehat{h}_{\alpha t}(\theta) &= (\nu+1)b_t(\theta) \\
\widehat{h}_{\gamma t}(\theta) &= -2\alpha(\nu+1)(1-b_t(\theta))\frac{e_t}{e_t^2+(\nu-2)h_t(\theta)} \qquad \text{(B.4)}
\end{aligned}$$

$$\widehat{h}_{\nu t}(\theta) = \alpha b_t(\theta) - \alpha(\nu+1)b_t(\theta)(1-b_t(\theta))(\nu-2)^{-1}. \tag{B.5}$$

The diagonal elements of $\nabla^2 l_t(\theta)$ are

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\delta^2} &= -\frac{1}{2}h_{\delta t}(\theta)^2((\nu+1)b_t(\theta)-1) - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\delta t}(\theta)^2 \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\delta\delta t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\alpha^2} &= -\frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\alpha t}(\theta)^2 - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\alpha t}(\theta)^2 \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\alpha\alpha t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\gamma^2} &= -2(\nu+1)(1-b_t(\theta))h_{\gamma t}(\theta)\frac{e_t}{e_t^2+(\nu-2)h_t(\theta)} \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta)^2 + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\gamma\gamma t}(\theta) \\
&\quad - \frac{(\nu+1)(1-2b_t(\theta))}{(\nu-2)h_t(\theta)+e_t^2} \\
\frac{\partial^2 l_t(\theta)}{\partial\nu^2} &= 2\psi_1(\nu) + \frac{1}{2(\nu-2)^2} + \frac{1}{2}h_{\nu t}(\theta)^2 - \frac{1}{2}h_{\nu\nu t}(\theta) + b_t(\theta)h_{\nu t}(\theta) \\
&\quad - \frac{\nu+1}{\nu-2}b_t(\theta)(1-b_t(\theta))h_{\nu t}(\theta) - \frac{\nu+1}{2}b_t(\theta)(1-b_t(\theta))h_{\nu t}(\theta)^2 \\
&\quad + \frac{1}{2(\nu-2)^2}b_t(\theta)((\nu+1)b_t(\theta)+(\nu-5)).
\end{aligned}$$

For the cross derivatives, with $i,j = 2,3,4$, (i.e. $\theta_i = \alpha$ or β or δ, and $\theta_j = \alpha$ or β or δ), we have

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j} &= -\frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta)h_{\theta_j t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i t}(\theta)h_{\theta_j t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i\theta_j t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\gamma} &= -\frac{1}{2}(\nu+1)(1-b_t(\theta))h_{\theta_i t}(\theta)\frac{2e_t}{e_t^2+(\nu-2)h_t(\theta)} \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta)h_{\gamma t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i t}(\theta)h_{\gamma t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i\gamma t}(\theta) \\
\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\nu} &= \frac{1}{2}b_t(\theta)h_{\theta_i t}(\theta) - \frac{\nu+1}{2(\nu-2)}b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta) \\
&\quad - \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\theta_i t}(\theta)h_{\nu t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\theta_i t}h_{\nu t}(\theta) \\
&\quad + \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\nu\theta_i t}(\theta)
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\nu\partial\gamma} &= \frac{1}{2}b_t(\theta)h_{\gamma t}(\theta) - \frac{\nu+1}{2(\nu-2)}b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta) \\
&- \frac{1}{2}(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\gamma t}(\theta)h_{\nu t}(\theta) - \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\gamma t}(\theta)h_{\nu t}(\theta) \\
&+ \frac{1}{2}((\nu+1)b_t(\theta)-1)h_{\nu\gamma t}(\theta) + \frac{1}{\nu-2}\frac{b_t(\theta)}{e_t}((\nu+1)b_t(\theta)-3) \\
&- (\nu+1)\frac{b_t(\theta)}{e_t}(1-b_t(\theta))h_{\nu t}(\theta).
\end{aligned}
$$

By recursion, we have

$$
h_{\theta_i\theta_j t}(\theta) = \sum_{k=1}^{t} \frac{\widehat{h}_{\theta_i\theta_j t-k}(\theta)}{\beta+\alpha(\nu+1)b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{(\beta+\alpha(\nu+1)b_{t-j}(\theta)^2)h_{t-j}(\theta)}{h_{t-j+1}(\theta)} \tag{B.6}
$$

for all t and $i, j = 1, 2, \ldots, 6$, where

$$
\begin{aligned}
\widehat{h}_{\alpha\alpha t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)^2 \\
\widehat{h}_{\delta\delta t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)^2, \\
\widehat{h}_{\gamma\gamma t}(\theta) &= -\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))\left[-\frac{2}{e_t^2} + 2b_t(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right)^2\right], \\
\widehat{h}_{\nu\nu t}(\theta) &= \alpha(\nu-2)^{-1}b_t(\theta)h_{\nu t}(\theta)[4(\nu+1)b_t(\theta)^2 - 2(\nu-4)b_t(\theta) - (\nu-2)] \\
&- 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta)^2)h_{\nu t}(\theta)^2 \\
&+ 2\alpha(\nu-2)^{-2}b_t(\theta)(1-b_t(\theta))(3-(\nu+1)b_t(\theta)), \\
\widehat{h}_{\delta\alpha t}(\theta) &= (\nu+1)b_t(\theta)^2 h_{\delta t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)h_{\alpha t}(\theta), \\
\widehat{h}_{\delta\beta t}(\theta) &= h_{\delta t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)h_{\beta t}(\theta), \\
\widehat{h}_{\delta\gamma t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right), \\
\widehat{h}_{\delta\nu t}(\theta) &= \alpha b_t(\theta)^2 h_{\delta t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\delta t}(\theta)((\nu-2)^{-1} + h_{\nu t}(\theta)), \\
\widehat{h}_{\alpha\beta t}(\theta) &= (\nu+1)b_t(\theta)^2 h_{\beta t}(\theta) + h_{\alpha t}(\theta) - 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)h_{\beta t}(\theta), \\
\widehat{h}_{\alpha\gamma t}(\theta) &= -(\nu+1)b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right) + (\nu+1)b_t(\theta)h_{\gamma t}(\theta) \\
&- 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right), \\
\widehat{h}_{\alpha\nu t}(\theta) &= b_t(\theta) - (\nu+1)b_t(\theta)(1-b_t(\theta))((\nu-2)^{-1} + h_{\nu t}(\theta)) \\
&+ (\nu+1)b_t(\theta)h_{\nu t}(\theta) + \alpha b_t(\theta)^2 h_{\alpha t}(\theta) \\
&- 2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))h_{\alpha t}(\theta)((\nu-2)^{-1} + h_{\nu t}(\theta)), \\
\widehat{h}_{\beta\gamma t}(\theta) &= h_{\gamma t}(\theta) - 2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))h_{\beta t}(\theta)\left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right)
\end{aligned}
$$

$$\begin{aligned}
\widehat{h}_{\beta\nu t}(\theta) &= h_{\nu t}(\theta) + \alpha b_t(\theta)^2 h_{\beta t}(\theta) \\
&\quad - 2\alpha(\nu+1) b_t(\theta)(1-b_t(\theta)) h_{\beta t}(\theta)((\nu-2)^{-1} + h_{\nu t}(\theta)), \\
\widehat{h}_{\gamma\nu t}(\theta) &= -\alpha b_t(\theta)(1-b_t(\theta)) \left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right) + \alpha b_t(\theta) h_{\gamma t}(\theta) \\
&\quad - \alpha(\nu+1)(\nu-2)^{-1} b_t(\theta)(1-b_t(\theta)) h_{\gamma t}(\theta) \\
&\quad + \alpha(\nu+1)(\nu-2)^{-1}(1-2b_t(\theta)) b_t(\theta)(1-b_t(\theta)) \left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right) \\
&\quad - 2\alpha(\nu+1) b_t(\theta)^2 (1-b_t(\theta)) h_{\nu t}(\theta) \left(\frac{2}{e_t} + h_{\gamma t}(\theta)\right).
\end{aligned}$$

B.2. Definition of $\widetilde{u}_{\theta_i\theta_j t}(\theta_0)$

$$
\begin{aligned}
\widetilde{u}_{\alpha\alpha t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)^2, \\
\widetilde{u}_{\delta\delta t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)^2, \\
\widetilde{u}_{\beta\beta t}(\theta_0) &= 2u^*_{\beta t}(\theta_0)\left(1+\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)\right), \\
\widetilde{u}_{\gamma\gamma t}(\theta_0) &= \alpha_u(\nu_u+1)\left[\frac{1}{(\nu_l-2)\delta_l}+2\left(\frac{2}{(\nu_l-2)\delta_l}+u^*_{\gamma t}(\theta_0)\right)^2\right], \\
\widetilde{u}_{\nu\nu t}(\theta_0) &= \alpha_u(\nu_l-2)^{-1}u^*_{\nu t}(\theta_0)[4(\nu_u+1)+2(\nu_u-4)+(\nu_u-2)] \\
&\quad + 2\alpha_u(\nu_u+1)u^*_{\nu t}(\theta_0)^2+2\alpha_u(\nu_l-2)^{-2}(3+(\nu_u+1)), \\
\widetilde{u}_{\delta\alpha t}(\theta_0) &= (\nu_u+1)u^*_{\delta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)u^*_{\alpha t}(\theta_0), \\
\widetilde{u}_{\delta\beta t}(\theta_0) &= u^*_{\delta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)u^*_{\beta t}(\theta_0), \\
\widetilde{u}_{\delta\gamma t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right), \\
\widetilde{u}_{\delta\nu t}(\theta_0) &= \alpha_u u^*_{\delta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\delta t}(\theta_0)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0)), \\
\widetilde{u}_{\alpha\beta t}(\theta_0) &= (\nu_u+1)u^*_{\beta t}(\theta_0)+u^*_{\alpha t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)u^*_{\beta t}(\theta_0), \\
\widetilde{u}_{\alpha\gamma t}(\theta_0) &= (\nu_u+1)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right)+(\nu_u+1)u^*_{\gamma t}(\theta_0) \\
&\quad + 2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right), \\
\widetilde{u}_{\alpha\nu t}(\theta_0) &= 1+(\nu_u+1)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0))+(\nu_u+1)u^*_{\nu t}(\theta_0)+\alpha_u u^*_{\alpha t}(\theta_0) \\
&\quad + 2\alpha_u(\nu_u+1)u^*_{\alpha t}(\theta_0)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0)), \\
\widetilde{u}_{\beta\gamma t}(\theta_0) &= 2\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right) \\
&\quad + u^*_{\gamma t}(\theta_0), \\
\widetilde{u}_{\beta\nu t}(\theta_0) &= u^*_{\nu t}(\theta_0)+\alpha_u u^*_{\beta t}(\theta_0)+2\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)((\nu_l-2)^{-1}+u^*_{\nu t}(\theta_0)), \\
\widetilde{u}_{\gamma\nu t}(\theta_0) &= \alpha_u\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right)+\alpha_u u^*_{\gamma t}(\theta_0) \\
&\quad + \alpha_u(\nu_u+1)(\nu_u-2)^{-1}u^*_{\gamma t}(\theta_0) \\
&\quad + \alpha_u(\nu_u+1)(\nu_l-2)^{-1}\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right) \\
&\quad + 2\alpha_u(\nu_u+1)u^*_{\nu t}(\theta_0)\left(2\max\{1,((\nu_l-2)\delta_l)^{-1}\}+u^*_{\gamma t}(\theta_0)\right).
\end{aligned}
$$

B.3. Definition of $\widehat{u}_{\theta_i\theta_j t}(\theta)$

$$
\begin{aligned}
\widehat{u}_{\delta\delta t}(\theta) &= 0,\\
\widehat{u}_{\alpha\alpha t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)^2\\
\widehat{u}_{\beta\beta t}(\theta) &= 2u_{\beta t}(\theta)\left[1-\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\beta t}(\theta)\right],\\
\widehat{u}_{\gamma\gamma t}(\theta) &= 0,\\
\widehat{u}_{\nu\nu t}(\theta) &= \alpha(\nu-2)^{-1}b_t(\theta)u_{\nu t}(\theta)[4(\nu+1)b_t(\theta)^2-2(\nu-4)b_t(\theta)-(\nu-2)]\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\nu t}(\theta)^2\\
&\quad +2\alpha(\nu-2)^{-2}b_t(\theta)(1-b_t(\theta))(3-(\nu+1)b_t(\theta)),\\
\widehat{u}_{\delta\alpha t}(\theta) &= (\nu+1)b_t(\theta)^2u_{\delta t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)u_{\alpha t}(\theta),\\
\widehat{u}_{\delta\beta t}(\theta) &= u_{\delta t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)u_{\beta t}(\theta),\\
\widehat{u}_{\delta\gamma t}(\theta) &= -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right),\\
\widehat{u}_{\delta\nu t}(\theta) &= \alpha b_t(\theta)^2u_{\delta t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\delta t}(\theta)((\nu-2)^{-1}+u_{\nu t}(\theta)),\\
\widehat{u}_{\alpha\beta t}(\theta) &= (\nu+1)b_t(\theta)^2u_{\beta t}(\theta)+u_{\alpha t}(\theta)-2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)u_{\beta t}(\theta),\\
\widehat{u}_{\alpha\gamma t}(\theta) &= -(\nu+1)b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)+(\nu+1)b_t(\theta)u_{\gamma t}(\theta)\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right),\\
\widehat{u}_{\alpha\nu t}(\theta) &= b_t(\theta)-(\nu+1)b_t(\theta)(1-b_t(\theta))((\nu-2)^{-1}+u_{\nu t}(\theta))\\
&\quad +(\nu+1)b_t(\theta)u_{\nu t}(\theta)+\alpha b_t(\theta)^2u_{\alpha t}(\theta)\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\alpha t}(\theta)((\nu-2)^{-1}+u_{\nu t}(\theta)),\\
\widehat{u}_{\beta\gamma t}(\theta) &= u_{\gamma t}(\theta)-2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))u_{\beta t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)\\
\widehat{u}_{\beta\nu t}(\theta) &= u_{\nu t}(\theta)+\alpha b_t(\theta)^2u_{\beta t}(\theta)\\
&\quad -2\alpha(\nu+1)b_t(\theta)(1-b_t(\theta))u_{\beta t}(\theta)((\nu-2)^{-1}+u_{\nu t}(\theta)),\\
\widehat{u}_{\gamma\nu t}(\theta) &= -\alpha b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)+\alpha b_t(\theta)u_{\gamma t}(\theta)\\
&\quad -\alpha(\nu+1)(\nu-2)^{-1}b_t(\theta)(1-b_t(\theta))u_{\gamma t}(\theta)\\
&\quad +\alpha(\nu+1)(\nu-2)^{-1}(1-2b_t(\theta))b_t(\theta)(1-b_t(\theta))\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right)\\
&\quad -2\alpha(\nu+1)b_t(\theta)^2(1-b_t(\theta))u_{\nu t}(\theta)\left(\frac{2}{e_t}+u_{\gamma t}(\theta)\right).
\end{aligned}
$$

APPENDIX C: Theorem proofs

Proof for Theorem 1. The proof of this lemma is analogous to Theorem 1 of Nelson (1990). By recursion, we have

$$h_{0t} = \delta_0 \left(1 + \sum_{k=1}^{t-1} \prod_{j=1}^{k} (\beta_0 + \alpha_0(\nu_0 + 1)b_{0t-j}) \right) + \omega_0 \prod_{j=1}^{t} (\beta_0 + \alpha_0(\nu_0 + 1)b_{0t-j}) . \tag{C.1}$$

Clearly, $\delta_0 < h_{0t}$ a.s. for all $t \in \mathbb{N}_{>0}$ and any $\theta_0 \in \Theta$. Moreover, (C.1) is absolutely convergent almost surely as $t \to \infty$ if $\mathbb{E}[\ln(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t})] < 0$ (i.e. $\theta_0 \in \Theta_L$), and otherwise it is divergent [Stout (1974, p. 332) or Theorem 1 of Brandt (1986)].

Thus h_{0t} for each $t \in \mathbb{N}$ and $\lim_{t\to\infty} h_{0t}$ are measurable if $\theta_0 \in \Theta_L$. Since b_{0t} is strictly stationary and ergodic for all $\theta_0 \in \Theta$ and h_{0t} is a measurable function of $(b_{0t}, b_{0t-1}, \dots)$ if $\theta_0 \in \Theta_L$, h_{0t} is strictly stationary and ergodic by Theorem 3.5.8 of Stout (1974).

The L^p convergence of $(1/h_{0t})_{t\in\mathbb{N}}$ to zero when $\theta_0 \in \Theta_U$ is by dominated convergence, since $0 < 1/h_{0t} \leq 1/\delta_l < \infty$ a.s. for all t and $\theta_0 \in \Theta$. ■

Proof for Theorem 2. By Theorem 1 and Lemma 3, we know that $(h_{0t})_{t\in\mathbb{N}}$, $(h_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$, and $(h_{\theta_i\theta_j t}(\theta_0))_{t\in\mathbb{N}}$ are strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i, j = 1, \dots, 6$. $(b_{0t})_{t\in\mathbb{N}}$ is i.i.d., and so it is also strictly stationary and ergodic. Thus, the desired property holds by Theorem 13.3 of Billingsley (1986) and Theorem 3.5.8 of Stout (1974). ■

Proof for Theorem 3. By Lemmas 16, 8, and 10, the necessary conditions (A.1)-(A.3) in Lemma 1 of Jensen and Rahbek (2004) are satisfied. ■

APPENDIX D: Lemmas

Throughout the following analysis, note that e_t is a function of γ because

$$e_t = \varepsilon_t + (\gamma_0 - \gamma) = \varepsilon_t + g$$

with $g \equiv \gamma_0 - \gamma$. We repeatedly use Lemma 1 to bound several quantities in the subsequent lemmas and obtain the convergence results of Sections 4.

LEMMA 1. *For all $\theta_0, \theta \in \Theta$ and $t \in \mathbb{N}$, we have*

$$\frac{|e_t|}{e_t^2 + (\nu - 2)h_t(\theta)} \leq \begin{cases} 1 & \text{if} \quad |e_t| \geq 1, \\ ((\nu-2)\delta)^{-1} < \infty & \text{if} \quad |e_t| < 1, \end{cases} \tag{D.1}$$

a.s. If we take the L^p-norm of the LHS quantity, it has the following simple upper-bound;

$$\left\| \frac{e_t}{e_t^2 + (\nu - 2)h_t(\theta)} \right\|_p \leq \sqrt{\frac{1}{(\nu-2)\delta}} < \infty \tag{D.2}$$

for any $p \geq 1$, $t \in \mathbb{N}$, and $\theta \in \Theta$. Moreover, at $\theta = \theta_0 \in \Theta_U$, the quantity on the LHS of (D.2) tends to zero as $t \to \infty$ for any $p \geq 1$.

Proof. For all $\theta_0, \theta \in \Theta$ and t, we have

$$\frac{|e_t|}{e_t^2 + (\nu - 2)h_t(\theta)} \leq \frac{e_t^2}{e_t^2 + (\nu - 2)h_t(\theta)} \leq 1,$$

if $|e_t| \geq 1$, and

$$\frac{|e_t|}{e_t^2 + (\nu - 2)h_t(\theta)} \leq \frac{1}{(\nu - 2)h_t(\theta)} \leq \frac{1}{(\nu-2)\delta} < \infty,$$

if $|e_t| < 1$ a.s. This gives (D.1). As the L^p-norms are increasing in p, we have

$$\begin{aligned}\left(\mathbb{E}\left[\left(\frac{e_t}{e_t^2 + (\nu - 2)h_t(\theta)}\right)^p\right]\right)^{1/p} &\leq \sqrt{\frac{1}{(\nu-2)\delta}} \left(\mathbb{E}\left[\left(\frac{e_t^2}{e_t^2 + (\nu - 2)h_t(\theta)}\right)^p\right]\right)^{1/(2p)} \\ &\leq \sqrt{\frac{1}{(\nu-2)\delta}} < \infty,\end{aligned}$$

for all t, $p \geq 1$, and $\theta_0, \theta \in \Theta$. This shows (D.2). Finally, using the property that $\|XY\|_p \leq \|X\|_{2p}\|Y\|_{2p}$ for any random variables X and Y, we obtain for any $\theta = \theta_0 \in \Theta_U$,

$$\begin{aligned}\left\| \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_p &\leq \left\| \frac{z_t}{z_t^2 + (\nu_0 - 2)} \right\|_{2p} \left\| \frac{1}{\sqrt{h_{0t}}} \right\|_{2p} \\ &= \left\| \frac{z_t^2}{(z_t^2 + (\nu_0 - 2))^2} \right\|_p^{1/2} \left\| \frac{1}{h_{0t}} \right\|_p^{1/2} \leq \sqrt{\frac{1}{\nu_0 - 2}} \left\| \frac{1}{h_{0t}} \right\|_p^{1/2}\end{aligned}$$

and the last quantity on the RHS tends to zero as $t \to \infty$ for any $p \geq 1$ by Theorem 1. ■

The following lemma is used to show that several quantities, especially the derivatives of the log-likelihood, in the subsequent lemmas are bounded in the L^p-norm.

LEMMA 2. *For any $\theta_0 \in \Theta$ and $t \in \mathbb{N}$,* *(i)* $h_{\theta_i t}(\theta_0)$ *is bounded in* L^p *for any* $p \geq 1$ *and* $i = 1, \ldots, 5$. *For instance, we write* $\|h_{\beta t}(\theta_0)\|_p \leq H_p(\theta_0) < \infty$ *for some* $H_p(\theta_0) > 0$.

(ii) $h_{\theta_i \theta_j t}(\theta_0)$ *is bounded in* L^p *for any* $p \geq 1$ *and* $i, j = 1, \ldots, 5$. *For instance, we write* $\|h_{\beta\beta t}(\theta_0)\|_p \leq H_p^{\dagger}(\theta_0) < \infty$ *for some* $H_p^{\dagger}(\theta_0) > 0$.

(iii) $h_{\theta_i \theta_j \theta_k t}(\theta_0)$ *is bounded in* L^p *for any* $p \geq 1$ *and* $i, j, k = 1, \ldots, 5$. *For instance, we write* $\|h_{\beta\beta\beta t}(\theta_0)\|_p \leq H_p^{\ddagger}(\theta_0) < \infty$ *for some* $H_p^{\ddagger}(\theta_0) > 0$.

Proof. (i) For all t and $\theta_0 \in \Theta$, we have

$$0 < \frac{h_{0t}(\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2)}{h_{0t+1}} < \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} < 1 \tag{D.3}$$

a.s. because $b_{0t} \in (0, 1)$ a.s. and it is a non-degenerate continuous random variable for each t and $\theta_0 \in \Theta$. Define

$$\left\| \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} \right\|_p \equiv D_p(\theta_0) \in (0, 1)$$

for each $p \geq 1$, t, and $\theta_0 \in \Theta$. We have $D_p(\theta_0) \in (0, 1)$ for each t by (D.3). Note that, for any $k \in \mathbb{N}_{>0}$, $p \geq 1$, and $\theta_0 \in \Theta$, we have

$$\left\| \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} \right\|_p = \prod_{j=1}^{k} \left\| \frac{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1)b_{0t}} \right\|_p = D_p(\theta_0)^k$$

by the i.i.d. property of $(b_{0t})_{t \in \mathbb{N}}$. Then (B.3) implies that

$$\|h_{\theta_i t}(\theta_0)\|_p \leq \beta_l^{-1} \sum_{k=1}^{t} \left\| \widehat{h}_{\theta_i t-k}(\theta_0) \right\|_{2p} D_{2p}(\theta_0)^k$$

for all t, $\theta_0 \in \Theta$, $p \geq 1$, and $i = 1, \ldots, 5$ by the Minkowski inequality and the property that $\|XY\|_p \leq \|X\|_{2p}\|Y\|_{2p}$ for any random variables X and Y. Since $D_p(\theta_0) \in (0, 1)$ for any $p \geq 1$ and $\theta_0 \in \Theta$, $\|h_{\theta_i t}(\theta_0)\|_p$ is bounded in L^p for any $p \geq 1$, t and $\theta \in \Theta$ if so is $\widehat{h}_{\theta_i t-k}(\theta_0)$. For $i = 3$, since $\widehat{h}_{\beta t}(\theta_0) = 1$, we obtain

$$\|h_{\beta t}(\theta_0)\|_p \leq \beta_l^{-1} \sum_{k=1}^{t} D_{2p}(\theta_0)^k \leq \frac{D_{2p}(\theta_0)}{\beta_l(1 - D_{2p}(\theta_0))} \equiv H_p(\theta_0) < \infty. \tag{D.4}$$

for all t, $p \geq 1$, and $\theta_0 \in \Theta$. For $i = 1, 2, 4$, we have

$$\left| \widehat{h}_{\nu t}(\theta_0) \right| \leq \alpha_0 + \alpha_0(\nu_0 + 1)(\nu_0 - 2)^{-1} < \infty,$$
$$\left| \widehat{h}_{\alpha t}(\theta_0) \right| \leq (\nu_0 + 1) < \infty, \quad \left| \widehat{h}_{\delta t}(\theta_0) \right| \leq \delta_0^{-1} < \infty,$$

for all t and $\theta_0 \in \Theta$. Thus, we have

$$\|h_{\nu t}(\theta_0)\|_p \leq (\alpha_u + \alpha_u(\nu_u + 1)(\nu_l - 2)^{-1})H_p(\theta_0) < \infty, \tag{D.5}$$
$$\|h_{\alpha t}(\theta_0)\|_p \leq (\nu_u + 1)H_p(\theta_0) < \infty,$$

$$\|h_{\delta t}(\theta_0)\|_p \leq \frac{D_{2p}(\theta_0)}{\beta_l \delta_l (1 - D_{2p}(\theta_0))} = H_p(\theta_0)/\delta_l < \infty,$$

for all t and $\theta_0 \in \Theta$. For $i = 5$, we have

$$\left\|\widehat{h}_{\gamma t}(\theta_0)\right\|_p \leq \frac{2\alpha_0(\nu_0 + 1)}{\sqrt{(\nu_0 - 2)\delta_0}} < \infty$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$ by Lemma 1. Then we have

$$\|h_{\gamma t}(\theta_0)\|_p \leq \frac{2\alpha_u(\nu_u + 1)}{\sqrt{\delta_l(\nu_l - 2)}} H_p(\theta_0) < \infty$$

for any t, $p \geq 1$ and $\theta_0 \in \Theta$.

(ii) By (B.6), we have

$$\left\|h_{\theta_i\theta_j t}(\theta_0)\right\|_p \leq \beta_l^{-1} \sum_{k=1}^{t} \left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)\right\|_{2p} D_{2p}(\theta_0)^k.$$

for all t, $p \geq 1$, $i, j = 1, \ldots, 5$, and $\theta_0 \in \Theta$ by the Minkowski and Hölder inequalities. Thus, $\left\|h_{\theta_i\theta_j t}(\theta_0)\right\|_p$ is bounded for any $p \geq 1$, t, and $\theta_0 \in \Theta$ if so is $\left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)\right\|_p$. For instance, we have

$$\left\|\widehat{h}_{\beta\beta t}(\theta_0)\right\|_p \leq 2H_{2p}(\theta_0)\left(1 + \alpha_u(\nu_u + 1)H_{2p}(\theta_0)\right) < \infty$$

for all t and $\theta_0 \in \Theta$ by Lemma 2 (i). Then we have

$$\begin{aligned}\|h_{\beta\beta t}(\theta_0)\|_p &\leq \frac{2H_{2p}(\theta_0)\left(1 + \alpha_u(\nu_u + 1)H_{2p}(\theta_0)\right)}{\beta_l} \sum_{k=1}^{t} D_{2p}(\theta_0)^k \\ &\leq 2H_{2p}(\theta_0)H_p(\theta_0)\left(1 + \alpha_u(\nu_u + 1)H_{2p}(\theta_0)\right) \equiv H_p^{\dagger}(\theta_0) < \infty\end{aligned}$$

for all t and $\theta_0 \in \Theta$. Similarly, it is easy to establish that $\left\|h_{\theta_i\theta_j t}(\theta_0)\right\|_p < \infty$ for all t, $p \geq 1$, $\theta_0 \in \Theta$, and $i, j = 1, 2, 3, 4$ by Lemma 2 (i). For $i = 5$ (or $\theta_i = \gamma$) and $j = 1, \ldots, 5$, first note that

$$\begin{aligned}\left|\frac{b_{0t}}{\varepsilon_t^2}\right| &= \frac{1}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \leq \frac{1}{(\nu_0 - 2)\delta_0} < \infty, \\ \left\|b_{0t}\left(\frac{2}{\varepsilon_t} + h_{\gamma t}(\theta_0)\right)\right\|_p &\leq 2\left\|\frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}}\right\|_p + \|h_{\gamma t}(\theta_0)\|_p < \infty,\end{aligned} \tag{D.6}$$

for all t, $p \geq 1$, and $\theta \in \Theta$ by the Minkowski inequality, Lemma 1, and Lemma 2 (i). Using these properties, it is easy to establish that $\left\|\widehat{h}_{\theta_i\theta_j t}(\theta_0)\right\|_p < \infty$ for any $p \geq 1$, t, and $\theta_0 \in \Theta$ when $i = 5$ and $j = 1, \ldots, 5$.

(iii) Derivations analogous to the above show that the desired property holds for $\left\|h_{\theta_i\theta_j\theta_k t}(\theta_0)\right\|_p$ with $i, j, k = 1, \ldots, 5$ for any t, $p \geq 1$, and $\theta_0 \in \Theta$. For instance, for the case of $i = j = k = 3$, we have

$$\left\|\widehat{h}_{\beta\beta\beta t}(\theta_0)\right\|_p \leq 3H_p^{\dagger}(\theta_0) + 6\alpha_u(\nu_u + 1)H_{3p}(\theta_0)^3 + 2\alpha_u(\nu_u + 1)H_{2p}(\theta_0)H_{2p}^{\dagger}(\theta_0) < \infty$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$ by the Minkowski and Hölder inequalities and Lemma 2 (i)(ii). Thus we have

$$\begin{aligned}
&\|h_{\beta\beta\beta t}(\theta_0)\|_p \\
&\quad \leq \left(3H_p^{\dagger}(\theta_0) + 6\alpha_u(\nu_u+1)H_{3p}(\theta_0)^3 + 2\alpha_u(\nu_u+1)H_{2p}(\theta_0)H_{2p}^{\dagger}(\theta_0)\right) H_p(\theta_0) \\
&\quad \equiv H_p^{\ddagger}(\theta_0) < \infty
\end{aligned}$$

for any $p \geq 1$, t, and $\theta_0 \in \Theta$. ■

The following lemma is used to show the strict stationarity and ergodicity of $\nabla l_t(\theta_0)$ in Theorem 2.

LEMMA 3. *If $\theta_0 \in \Theta_L$, $(h_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ and $\left(h_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$ are strictly stationary and ergodic for $i, j = 1, \ldots, 5$.*

Proof. Note that $0 < (\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2)h_{0t}/h_{0t+1} < 1$ a.s. for all $t \in \mathbb{N}$ and $\theta_0 \in \Theta$, and the middle term is strictly stationary and ergodic if $\theta_0 \in \Theta_L$. Then we have

$$\mathbb{E}\left[\log\left(\frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2 h_{0t})}{h_{0t+1}}\right)\right] < 0$$

for all $\theta_0 \in \Theta_L$. By Lemma 1 and Theorem 1, $\left(\widehat{h}_{\theta_i t}(\theta_0)\right)_{t\in\mathbb{N}}$ is strictly stationary and ergodic and bounded by some fixed real number a.s. for all $\theta_0 \in \Theta_L$ and $i = 1, \ldots, 5$. Then we have

$$\mathbb{E}\left[\max\left\{0, \log\left(\frac{\widehat{h}_{\theta_i t}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2}\right)\right\}\right] < \infty.$$

Thus $h_{\theta_i t}(\theta_0)$ is convergent absolutely a.s. for all $t \in \mathbb{N}$, $i = 1, \ldots, 5$, and $\theta_0 \in \Theta_L$ by Theorem 1 of Brandt (1986). Then $h_{\theta_i t}(\theta_0)$ is measurable for all $t \in \mathbb{N}$, $i = 1, \ldots, 5$, and $\theta_0 \in \Theta_L$. Hence $(h_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i = 1, \ldots, 5$ by Theorem 3.5.8 of Stout (1974).

Then $\left(\widehat{h}_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i, j = 1, \ldots, 5$. Moreover, using the properties that $b_{0t} \in (0,1)$ a.s. for all $t \in \mathbb{N}$ and $\max\{0, \log|X|\} \leq |X|$ for any real-valued random variable X, we obtain

$$\mathbb{E}\left[\max\left\{0, \log\left(\frac{\widehat{h}_{\theta_i\theta_j t}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t}^2}\right)\right\}\right] \leq \beta_0^{-1}\mathbb{E}\left[\left|\widehat{h}_{\theta_i\theta_j t}(\theta_0)\right|\right] < \infty$$

for all $t \in \mathbb{N}$, $i, j = 1, \ldots, 5$, and $\theta_0 \in \Theta_L$ by Lemma 2 (i). Then $\left(h_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ and $i, j = 1, \ldots, 5$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974). ■

Lemma 4 is used to show Lemmas 5 and 9. Note that the condition, $1 \le p \le 4$, in Lemma 4 may be relaxed to any $p \ge 1$ if one uses the properties of the beta distribution to express the quantity,

$$\left\| \ln\left(1 + \frac{\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right) \right\|_p = \left\| \ln\left(1 + \frac{z_t^2}{(\nu_0 - 2)}\right) \right\|_p = \|\ln(1 - b_{0t})\|_p \,,$$

in terms of polygamma functions. This quantity should be finite for any ν_0 in our parameter space.

LEMMA 4. *For $1 \le p \le 4$, $t \in \mathbb{N}_{>0}$ and $\theta_0 \in \Theta$, we have*

$$\left\| \ln\left(1 + \frac{\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right) \right\|_p < \infty. \tag{D.7}$$

Proof. As the L^p-norm is increasing in $p \ge 1$, it is enough to show (D.7) for $p = 4$. Using the property that $(\ln(1+x))^4 < 5x$ for all $x > 0$, we have

$$\mathbb{E}\left[\left(\ln\left(1 + \frac{\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right)\right)^4\right] < \mathbb{E}\left[\frac{5\varepsilon_t^2}{(\nu_0 - 2)h_{0t}}\right] \le \frac{5}{(\nu_l - 2)} < \infty.$$

■

The preceding results can be used to establish that the elements of $\nabla l_t(\theta_0)$ and $\nabla^2 l_t(\theta_0)$ are bounded in L^p for some p.

LEMMA 5. $\|\partial l_t(\theta_0)/\partial\theta_i\|_p < \infty$ *for $1 \le p \le 4$, $i = 1, \ldots, 5$, $t \in \mathbb{N}_{>0}$, and any $\theta_0 \in \Theta$.*

Proof. For the derivative with respect to β, we have

$$\left\| \frac{\partial l_t(\theta_0)}{\partial \beta} \right\|_p \le \frac{\nu_u + 2}{2} \|h_{\beta t}(\theta_0)\|_p \le \frac{\nu_u + 2}{2} H_p(\theta_0) < \infty$$

for any $p \ge 1$, t, and $\theta_0 \in \Theta$ by Lemma 2 (i). Similar derivations show that $\partial l_t(\theta_0)/\partial\delta$ and $\partial l_t(\theta_0)/\partial\alpha$ are bounded in L^p for all t, $p \ge 1$, and $\theta_0 \in \Theta$ by Lemma 2 (i). $\partial l_t(\theta_0)/\partial\gamma$ is bounded in L^p for all $p \ge 1$, t, and $\theta_0 \in \Theta$ by Lemma 1 and Lemma 2 (i). Finally, $\partial l_t(\theta_0)/\partial\nu$ is bounded in L^4 for all t and $\theta_0 \in \Theta$ by Lemma 2 (i) and Lemma 4. ■

LEMMA 6. $\|\partial^2 l_t(\theta_0)/\partial\theta_i\partial\theta_j\|_p < \infty$ *for all $p \ge 1$, $i, j = 1, \ldots, 5$, $t \in \mathbb{N}_{>0}$, and $\theta_0 \in \Theta$.*

Proof. This can be established by the Minkowski and Hölder inequalities, Lemma 1 and Lemma 2 (i)(ii). For instance, consider $\partial^2 l_t(\theta_0)/\partial\beta^2$. We obtain from (B.1) that

$$\left\| \frac{\partial^2 l_t(\theta_0)}{\partial \beta^2} \right\|_p \leq \frac{2\nu_u + 3}{2} \left\| h_{\beta t}(\theta_0) \right\|_{2p}^2 + \frac{\nu_u + 2}{2} \left\| h_{\beta\beta t}(\theta_0) \right\|_p$$

and the RHS is bounded for all t, $p \geq 1$, and $\theta_0 \in \Theta$ by Lemma 2 (i)(ii). Similar derivations show the desired property for other second derivatives. ■

In order to establish the asymptotic properties of $\nabla L_n(\theta)$ and $\nabla^2 L_n(\theta)$, we define the following new processes

$$u_{\theta_i t}(\theta) = \sum_{k=1}^{t} \frac{\widehat{u}_{\theta_i t-k}(\theta)}{\beta + \alpha(\nu+1) b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{\beta + \alpha(\nu+1) b_{t-j}(\theta)^2}{\beta + \alpha(\nu+1) b_{t-j}(\theta)}$$

for $i = 1, \ldots, 5$, where $\widehat{u}_{\theta_i t}(\theta)$ is set to be the limit of $\widehat{h}_{\theta_i t}(\theta)$ when $\theta = \theta_0 \in \Theta_U$ and $t \to \infty$. Thus, we define

$$\widehat{u}_{\nu t}(\theta) = \alpha b_t(\theta) - \alpha(\nu+1)(\nu-2)^{-1} b_t(\theta)(1 - b_t(\theta)),$$
$$\widehat{u}_{\alpha t}(\theta) = (\nu+1) b_t(\theta), \quad \widehat{u}_{\beta t}(\theta) = 1, \quad \widehat{u}_{\delta t}(\theta) = 0, \quad \widehat{u}_{\gamma t}(\theta) = 0.$$

Furthermore, define

$$u_{\theta_i \theta_j t}(\theta) = \sum_{k=1}^{t} \frac{\widehat{u}_{\theta_i \theta_j t-k}(\theta)}{\beta + \alpha(\nu+1) b_{t-k}(\theta)^2} \prod_{j=1}^{k} \frac{\beta + \alpha(\nu+1) b_{t-j}(\theta)^2}{\beta + \alpha(\nu+1) b_{t-j}(\theta)}$$

for $i, j = 1, \ldots, 5$, where $\widehat{u}_{\theta_i \theta_j t}(\theta)$ is set to be the limit of $\widehat{h}_{\theta_i \theta_j t}(\theta)$ when $\theta = \theta_0 \in \Theta_U$ and $t \to \infty$. They are defined in Appendix B.3. The following lemma establishes some of the useful properties of these processes.

LEMMA 7. *The processes, $(u_{\theta_i t}(\theta))_{t \in \mathbb{N}}$ and $\left(u_{\theta_i \theta_j t}(\theta)\right)_{t \in \mathbb{N}}$, satisfy the following properties.*

(i) $(u_{\theta_i t}(\theta_0))_{t \in \mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta$ and $i = 1, \ldots, 5$.

(ii) $u_{\theta_i t}(\theta_0)$ is bounded in L^p for all $p \geq 1$, $t \in \mathbb{N}$, $\theta_0 \in \Theta$, and $i = 1, \ldots, 5$.

(iii) $0 \leq h_{\theta_i t}(\theta_0) \leq u_{\theta_i t}(\theta_0)$ for all $\theta_0 \in \Theta$ and $i = 2, 3$ (i.e. $\theta_i = \alpha$ or β).

(iv) Define

$$y_{t-k}^*(\theta) \equiv \prod_{j=1}^{k} \frac{\beta + \alpha(\nu+1) b_{t-j}(\theta)^2}{\beta + \alpha(\nu+1) b_{t-j}(\theta)} - \prod_{j=1}^{k} \frac{(\beta + \alpha(\nu+1) b_{t-j}(\theta)^2) h_{t-j}(\theta)}{h_{t-j+1}(\theta)}$$

for $k, t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Then $\|y_{t-k}^*(\theta_0)\|_p \to 0$ as $t \to \infty$ for any $p \geq 1$, $k \in \mathbb{N}$, and $\theta_0 \in \Theta_U$.

(v) $h_{\theta_i t}(\theta_0) - u_{\theta_i t}(\theta_0)$ *is bounded in* L^p *for all* $p \geq 1$, $t \in \mathbb{N}$, $\theta_0 \in \Theta$ *and* $i = 1, \ldots, 5$.

(vi) *We have*

$$\|h_{\theta_i t}(\theta_0) - u_{\theta_i t}(\theta_0)\|_p \to 0 \tag{D.8}$$

as $t \to \infty$ *for any* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i = 1, \ldots, 5$.

(vii) *We have*

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \left(h_{\theta_i t}(\theta_0) - u_{\theta_i t}(\theta_0) \right) \right\|_p \to 0, \tag{D.9}$$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \left(h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0) \right) \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2) h_{0t}} \right\|_p \to 0, \tag{D.10}$$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \left(h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \right) \right\|_p \to 0, \tag{D.11}$$

as $n \to \infty$ *for all* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i, j = 1, \ldots, 5$.

(viii) $u_{\theta_i \theta_j t}(\theta_0)$ *is strictly stationary ergodic for all* $\theta_0 \in \Theta$ *and* $i, j = 1, \ldots, 5$.

(ix) $u_{\theta_i \theta_j t}(\theta_0)$ *is bounded in* L^p *for any* $p \geq 1$, $t \in \mathbb{N}$, $\theta_0 \in \Theta$, *and* $i, j = 1, \ldots, 5$.

(x) $\left\| h_{\theta_i \theta_j t}(\theta_0) - u_{\theta_i \theta_j t}(\theta_0) \right\|_p < \infty$ *for all* $t \in \mathbb{N}$, $p \geq 1$, $\theta_0 \in \Theta$, *and* $i, j = 1, \ldots, 5$.

(xi) $\left\| h_{\theta_i \theta_j t}(\theta_0) - u_{\theta_i \theta_j t}(\theta_0) \right\|_p \to 0$ *as* $t \to \infty$ *for all* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i, j = 1, \ldots, 5$.

(xii) $\left\| n^{-1} \sum_{t=1}^{n} \left(h_{\theta_i \theta_j t}(\theta_0) - u_{\theta_i \theta_j t}(\theta_0) \right) \right\|_p \to 0$ *as* $n \to \infty$ *for all* $p \geq 1$, $\theta_0 \in \Theta_U$, *and* $i, j = 1, \ldots, 5$.

Proof. (i) (viii) By (D.3), we have

$$\mathbb{E}\left[\ln \left(\frac{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t}^2}{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t}} \right) \right] < 0.$$

Moreover, using the property that $\ln(x) \leq x - 1$ for all $x > 0$, we have

$$\mathbb{E}\left[\max\{0, \ln |\widetilde{u}_{\theta_i}|\} \right] \leq \mathbb{E}\left[\max\{0, |\widetilde{u}_{\theta_i}|\} \right] = \mathbb{E}\left[|\widetilde{u}_{\theta_i}| \right] < \infty$$

for $i = 1, \ldots, 5$. Then $(u_{\theta_i t}(\theta_0))_{t \in \mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta$ by Theorem 1 of Brandt (1986). Likewise, we can show that $\mathbb{E}\left[\max\{0, \ln |\widetilde{u}_{\theta_i \theta_j t}(\theta_0)|\} \right] < \infty$ and $\mathbb{E}\left[\max\{0, \ln |\widetilde{u}_{\theta_i \theta_j \theta_m t}(\theta_0)|\} \right] < \infty$ for all t, $\theta_0 \in \Theta$, and $i, j, m = 1, \ldots, 5$. Then we can deduce that $(u^*_{\theta_i \theta_j t}(\theta_0))_{t \in \mathbb{N}}$ and

$(u^*_{\theta_i\theta_j\theta_m t}(\theta_0))_{t\in\mathbb{N}}$ are strictly stationary and ergodic for any $\theta_0 \in \Theta$ and $i, j, m = 1, \ldots, 5$ by Theorem 1 of Brandt (1986).

(ii) The proof for $i = 4, 5$ (i.e. $\theta_i = \delta$ or γ) is trivial as $\widehat{u}_{\delta t}(\theta) = \widehat{u}_{\gamma t}(\theta) = 0$ for all t and $\theta \in \Theta$. Recalling (D.3) and $D_p(\theta_0)$ defined in in Lemma 2 (i), we obtain

$$\begin{aligned}
\|u_{\alpha t}(\theta_0)\|_p &\leq \frac{\nu_u + 1}{\beta_l} \sum_{k=1}^{\infty} D_p(\theta_0)^k = \frac{(\nu_u + 1) D_p(\theta_0)}{\beta_l(1 - D_p(\theta_0))} < \infty, \\
\|u_{\beta t}(\theta_0)\|_p &\leq \frac{1}{\beta_l} \sum_{k=1}^{\infty} D_p(\theta_0)^k = \frac{D_p(\theta_0)}{\beta_l(1 - D_p(\theta_0))} < \infty, \\
\|u_{\nu t}(\theta_0)\|_p &\leq \frac{\alpha_u + \alpha_u(\nu_u + 1)(\nu_l - 2)^{-1}}{\beta_l} \sum_{k=1}^{\infty} D_p(\theta_0)^k \\
&\leq \frac{(\alpha_u + \alpha_u(\nu_u + 1)(\nu_l - 2)^{-1}) D_p(\theta_0)}{\beta_l(1 - D_p(\theta_0))} < \infty.
\end{aligned} \tag{D.12}$$

for all t, $p \geq 1$, and $\theta_0 \in \Theta$.

(iii) This is by (D.3) and the fact that $\widehat{u}_{\theta_i t}(\theta_0) = \widehat{h}_{\theta_i t}(\theta_0)$ for all t, $\theta_0 \in \Theta$, and $i = 2, 3$ (i.e. $\theta_i = \alpha$ or β).

(iv) By Theorem 1 and (D.3), we have

$$0 < \frac{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-j}} - \frac{(\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-j}^2) h_{0t-j}}{h_{0t-j+1}} \to 0$$

a.s. as $t \to \infty$ for any $j \in \mathbb{N}$ and $\theta_0 \in \Theta_U$. Thus $0 \leq y^*_{t-k}(\theta_0) \to 0$ a.s. as $t \to \infty$ for any $k \in \mathbb{N}$ and all $\theta_0 \in \Theta_U$. Moreover, $y^*_t(\theta)^p \leq 1$ for any $p \geq 1$, $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Thus $\|y^*_{t-k}(\theta_0)\|_p \to 0$ as $t \to \infty$ for any $k \in \mathbb{N}$, $p \geq 1$, and $\theta_0 \in \Theta_U$ by dominated convergence.

(v) This is by the Minkowski inequality, Lemma 2 (i), and Lemma 7 (ii).

(vi) We prove (D.8) for $p = 1$ first. For any $t_0 < t$ and $\theta_0 \in \Theta_U$, we have

$$\begin{aligned}
&\mathbb{E}\left[|u_{\alpha t}(\theta_0) - h_{\alpha t}(\theta_0)|\right] \\
&\quad = \mathbb{E}\left[\left|\sum_{k=1}^{\infty} \frac{(\nu_0 + 1) b_{0t-k}}{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-j}} \right.\right. \\
&\qquad \left.\left. - \sum_{k=1}^{t} \frac{(\nu_0 + 1) b_{0t-k}}{\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-k}^2} \prod_{j=1}^{k} \frac{h_{0t-j}(\beta_0 + \alpha_0(\nu_0 + 1) b_{0t-j}^2)}{h_{0t-j+1}} \right|\right]
\end{aligned}$$

$$\leq \sum_{k=1}^{t_0} \mathbb{E}\left[\left|\frac{(\nu_0+1)b_{0t-k}}{\beta_0+\alpha_0(\nu_0+1)b_{0t-k}^2} y_{t-k}^*(\theta_0)\right|\right]$$
$$+ \sum_{k=t_0+1}^{\infty} \mathbb{E}\left[\left|\frac{(\nu_0+1)b_{0t-k}}{\beta_0+\alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0+\alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0+\alpha_0(\nu_0+1)b_{0t-j}}\right|\right]$$
$$\leq \frac{\nu_u+1}{\beta_l} \sum_{k=1}^{t_0} \mathbb{E}\left[\left|y_{t-k}^*(\theta_0)\right|\right] + \frac{\nu_u+1}{\beta_l} \frac{D_1(\theta_0)^{t_0+1}}{1-D_1(\theta_0)}$$

by the triangle inequality and (D.3). The first term in the final line tends to zero as $t \to \infty$ for any $t_0 < t$ and $\theta_0 \in \Theta_U$ by Lemma 7 (iv). As the choice of $t_0 < t$ was arbitrary and $D_1(\theta_0) \in (0,1)$ for any $\theta_0 \in \Theta$, the second term in the final line tends to zero as $t_0 \to \infty$. Thus $\|u_{\alpha t}(\theta_0) - h_{\alpha t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ for all $\theta_0 \in \Theta_U$. Analogous derivations show that $\|u_{\beta t}(\theta_0) - h_{\beta t}(\theta_0)\|_1 \to 0$ and $\|u_{\nu t}(\theta_0) - h_{\nu t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ for all $\theta_0 \in \Theta_U$. For $i = 4$ (i.e. $\theta_i = \delta$), note that, for any $k \in \mathbb{N}$, we have

$$0 \leq \left(\frac{1}{(\beta_0+\alpha_0(\nu_0+1)b_{0t-k}^2)h_{0t-k}}\right)^p \to 0$$

a.s. as $t \to \infty$ for any $p \geq 1$ and $\theta_0 \in \Theta_U$ by Theorem 1. The term in the middle is also bounded above by $1/(\beta_l \delta_l)^p$ a.s. Thus, by dominated convergence,

$$\left\|\frac{1}{(\beta_0+\alpha_0(\nu_0+1)b_{0t-k}^2)h_{0t-k}}\right\|_p \to 0 \tag{D.13}$$

as $t \to \infty$ for any $p \geq 1$, $k \in \mathbb{N}$, and all $\theta_0 \in \Theta_U$. Then, for any arbitrary $t_0 < t$,

$$0 \leq \|h_{\delta t}(\theta_0) - u_{\delta t}(\theta_0)\|_1 \leq \sum_{k=1}^{t_0} \left\|\frac{1}{(\beta_0+\alpha_0(\nu_0+1)b_{0t-k}^2)h_{0t-k}}\right\|_2 D_2(\theta_0)^k$$
$$+ \frac{D_2(\theta_0)^{t_0+1}}{\beta_0\delta_0(1-D_2(\theta_0))}.$$

By (D.13), the first term on the RHS tends to zero as $t \to \infty$ for any $t_0 < t$ and $\theta_0 \in \Theta_U$. The second term on the RHS also tends to zero as $t_0 \to \infty$. Hence $\|h_{\delta t}(\theta_0) - u_{\delta t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ for all $\theta_0 \in \Theta_U$. For $i = 5$ (i.e. $\theta_i = \gamma$), note that

$$0 \leq \mathbb{E}\left[|\widehat{h}_{\gamma t-k}(\theta_0)|\right] \leq 2\alpha_0(\nu_0+1)\mathbb{E}\left[\frac{|\varepsilon_t|}{\varepsilon_t^2+(\nu_0-2)h_{0t}}\right] \to 0$$

as $t \to \infty$ for all $\theta_0 \in \Theta_U$ by Lemma 1. Moreover, $\widehat{h}_{\gamma t}(\theta_0)$ is bounded in L^p for any $p \geq 1$, $\theta_0 \in \Theta$, and t by Lemma 2 (i). Thus, we can show that $\|h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)\|_1 \to 0$ as $t \to \infty$ by derivations similar to the $i = 4$ (i.e. $\theta_i = \delta$) case.

By these properties of L^1 convergence to zero and the uniform integrability established in Lemma 7 (v), we have (D.8) for any $p \geq 1$, $\theta_0 \in \Theta_U$, and $i = 1, \ldots, 5$.

(vii) (D.9) is by the Minkowski inequality and Lemma 7 (vi). We also have

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{t=1}^{n} (h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)) \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_p \\
&\qquad \leq \frac{1}{n} \sum_{t=1}^{n} \|h_{\gamma t}(\theta_0) - u_{\gamma t}(\theta_0)\|_{2p} \left\| \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_{2p} \to 0,
\end{aligned}$$

as $n \to \infty$ for any $\theta_0 \in \Theta_U$ by Lemma 7 (vi) and Lemma 1. This establishes (D.10). Next we show (D.11). For any $p \geq 1$, $\theta_0 \in \Theta_U$, and $i, j = 1, \ldots, 5$, we have

$$\begin{aligned}
&\left\| u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) - h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) \right\|_p \\
&\leq \left\| u_{\theta_i t}(\theta_0) \left(u_{\theta_j t}(\theta_0) - h_{\theta_j t}(\theta_0) \right) \right\|_p + \left\| h_{\theta_j t}(\theta_0) (u_{\theta_i t}(\theta_0) - h_{\theta_i t}(\theta_0)) \right\|_p \\
&\leq \|u_{\theta_i t}(\theta_0)\|_{2p} \left\| u_{\theta_j t}(\theta_0) - h_{\theta_j t}(\theta_0) \right\|_{2p} + \left\| h_{\theta_j t}(\theta_0) \right\|_{2p} \|u_{\theta_i t}(\theta_0) - h_{\theta_i t}(\theta_0)\|_{2p} \\
&\to 0
\end{aligned} \tag{D.14}$$

as $t \to \infty$ by Lemma 2 (i) and Lemma 7 (ii)(vi). Thus we obtain

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{t=1}^{n} \left(h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \right) \right\|_p \\
&\qquad \leq \frac{1}{n} \sum_{t=1}^{n} \left\| \left(h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \right) \right\|_p \to 0
\end{aligned}$$

as $n \to \infty$ for all $\theta_0 \in \Theta_U$, $p \geq 1$, and $i, j = 1, \ldots, 5$. This shows (D.11).

(ix) The proof is analogous to Lemma 2 (i)(ii), and thus omitted.

(x) This is by the Minkowski inequality, Lemma 2 (ii), and Lemma 7 (ix).

(xi) For any $t_0 < t$, we have

$$
\begin{aligned}
&\left\|h_{\theta_i\theta_j t}(\theta_0) - u_{\theta_i\theta_j t}(\theta_0)\right\|_p \\
&\quad \leq \sum_{k=1}^{t_0} \left\| \frac{\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} y_{t-k}^*(\theta_0) \right\|_p \\
&\qquad + \sum_{k=1}^{t_0} \left\| \frac{\widehat{h}_{\theta_i\theta_j t-k}(\theta_0) - \widehat{u}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \right\|_p \\
&\qquad + \sum_{k=t_0+1}^{t} \left\| \frac{\widehat{u}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-k}^2} \prod_{j=1}^{k} \frac{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}^2}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \right\|_p \qquad \text{(D.15)} \\
&\quad \leq \beta_l^{-1} \sum_{k=1}^{t_0} \left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0)\right\|_{2p} \left\|y_{t-k}^*(\theta_0)\right\|_{2p} \\
&\qquad + \beta_l^{-1} \sum_{k=1}^{t_0} \left\|\widehat{h}_{\theta_i\theta_j t-k}(\theta_0) - \widehat{u}_{\theta_i\theta_j t-k}(\theta_l)\right\|_{2p} D_{2p}(\theta_0)^k \\
&\qquad + \beta_0^{-1} \sup_{i,j} \left\|\widehat{u}_{\theta_i\theta_j t-k}(\theta_0)\right\|_{2p} \frac{D_{2p}(\theta_0)^{t_0+1}}{1 - D_{2p}(\theta_0)}
\end{aligned}
$$

for all t and $\theta_0 \in \Theta$. The second inequality used the property that $\widehat{u}_{\theta_i\theta_j t}(\theta_0)$ is bounded in L^p and strictly stationary for all $p \geq 1$, $\theta_0 \in \Theta$, and $i, j = 1, \ldots, 5$. Note that we have

$$\left\|\widehat{h}_{\theta_i\theta_j t}(\theta_0) - \widehat{u}_{\theta_i\theta_j t}(\theta_0)\right\|_p \to 0 \tag{D.16}$$

as $t \to \infty$ for any $p \geq 1$, $\theta_0 \in \Theta_U$, and $i, j = 1, \ldots, 5$ by (D.14), Lemma 7 (vi), and Lemma 1. Then, by (D.16), Lemma 2 (ii), Lemma 7 (iv)(ix), and the property that $D_p(\theta_0) \in (0, 1)$ for all $p \geq 1$ and $\theta_0 \in \Theta$, the terms after the second inequality of (D.15) tends to zero as $t \to \infty$ and $t_0 \to \infty$ (since the choice of $t_0 < t$ was arbitrary) for any $\theta_0 \in \Theta_U$, $p \geq 1$, and $i, j = 1, \ldots, 5$.

(xii) This is by Lemma 7 (xi) and derivations analogous to the proof for (D.9) in Lemma 7 (vii). ■

Lemma 7 is used in the following lemma, which is used to show the asymptotic property of $\nabla_\theta L_n(\theta_0)$ and $\nabla_{\theta^*} L_n(\theta_0^*)$.

LEMMA 8. *Assume that* $\theta_0 \in \Theta_L$. *Then*

$$\frac{1}{n}\sum_{t=1}^{n} \nabla_\theta^2 l_t(\theta_0) \xrightarrow{P} \mathbb{E}\left[\nabla_\theta^2 l_t(\theta_0)\right] \equiv Q(\theta_0), \tag{D.17}$$

where $Q(\theta_0)$ is a constant symmetric matrix given θ_0. Moreover, if $\theta_0 \in \Theta_U$, then

$$\frac{1}{n}\sum_{t=1}^{n} \nabla^2_{\theta^*} l_t(\theta_0) \xrightarrow{P} \mathbb{E}\left[\nabla^2_{\theta^*} l_t(\theta_0)\right] \equiv Q^*(\theta_0) \tag{D.18}$$

for all $\theta_0 \in \Theta_U$, where $Q^*(\theta_0)$ is a constant symmetric matrix given θ_0.

Proof. For all $\theta \in \Theta_L$, $\nabla^2_\theta l_t(\theta_0)$ is strictly stationary and ergodic by Theorem 2. Moreover, $\mathbb{E}\left[|\partial^2 l_t(\theta_0)/\partial\theta_i\partial\theta_j|\right] < \infty$ for all $\theta_0 \in \Theta$ and $i, j = 1, \ldots, 5$ by Lemma 6. Thus (D.17) holds for all $\theta_0 \in \Theta_L$ by replacing the almost sure convergence of Theorem 3.5.7 of Stout (1974) by convergence in probability. For $\theta_0 \in \Theta_U$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\beta^2} &= \frac{1}{2n}\sum_{t=1}^{n}\left[(\nu_0+1)b_{0t}(b_{0t}-2)+1\right]\left(h_{\beta t}(\theta_0)^2 - u_{\beta t}(\theta_0)^2\right)\\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)\left(h_{\beta\beta t}(\theta_0) - u_{\beta\beta t}(\theta_0)\right)\\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left[(\nu_0+1)b_{0t}(b_{0t}-2)+1\right]u_{\beta t}(\theta_0)^2\\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)u_{\beta\beta t}(\theta_0)\\
&\xrightarrow{P} \frac{1}{2}\mathbb{E}\left[(\nu_0+1)b_{0t}(b_{0t}-2)+1\right]\mathbb{E}\left[u_{\beta t}(\theta_0)^2\right]\\
&= \frac{1}{2}\left(\frac{3}{\nu_0+3}-1\right)\mathbb{E}\left[u_{\beta t}(\theta_0)^2\right] < 0
\end{aligned}$$

as $n \to \infty$ by Lemma 7 (i)(vii)(viii)(xi)(xii) and the property that b_{0t} for each $t \in \mathbb{N}$ is i.i.d. with the distribution Beta$(1/2, \nu_0/2)$. Similarly, we can use the results of Lemma 7 to establish the desired convergence of other diagonal and off-diagonal elements of (D.18). ■

The following lemma verifies that the limiting distribution of $\nabla_\theta L_n(\theta_0)$ and $\nabla_{\theta^*} L_n(\theta_0)$ in Lemma 10 have well-defined variances.

LEMMA 9. $\mathbb{E}\left[|(\partial l_t(\theta_0)/\partial\theta_i)(\partial l_t(\theta_0)/\partial\theta_j)|\right] < \infty$ *for all $t \in \mathbb{N}$, $i, j = 1, \ldots, 5$, and $\theta_0 \in \Theta$.*

Proof. Using the property that $\|X^2\|_1 = \|X\|_2^2$ for any random variable X, we know that $\mathbb{E}\left[|\partial l_t(\theta_0)/\partial\theta_i|^2\right]$ is bounded for all $\theta_0 \in \Theta$ and $i = 1, \ldots, 5$ by Lemma 5. For $i \neq j$, $(\partial l_t(\theta_0)/\partial\theta_i)(\partial l_t(\theta_0)/\partial\theta_j)$ are also bounded in L^1 since

$$\left\|\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\right\|_1 \leq \left\|\frac{\partial l_t(\theta_0)}{\partial\theta_i}\right\|_2 \left\|\frac{\partial l_t(\theta_0)}{\partial\theta_j}\right\|_2 < \infty$$

for all $i, j = 1, \ldots, 5$ and $i \neq j$ by the Hölder inequality and Lemma 5. ■

The following proposition is used to establish Lemma 10.

PROPOSITION 1. Let $(X_t)_{t=1}^{\infty}$ be a sequence of random variables satisfying $n^{-1}\sum_{t=1}^{n} X_t \overset{P}{\to} 0$ as $n \to \infty$. If $n^{-1}\sum_{t=1}^{n} \mathbb{E}[X_t]$ converges as $n \to \infty$, then its limit is zero.

Proof. We can find a subsequence $(X_{t_k})_{k=1}^{\infty}$ such that $n^{-1}\sum_{k=1}^{n} X_{t_k} \to 0$ a.s. as $n \to \infty$. Then $n^{-1}\sum_{k=1}^{n} \mathbb{E}[X_{t_k}] \to 0$ as $n \to \infty$, and since $n^{-1}\sum_{t=1}^{n} \mathbb{E}[X_t]$ is convergent, its limit must be zero. ■

We are now ready to show that the asymptotic distribution of $\nabla_\theta L_n(\theta_0)$ and $\nabla_{\theta^*} L_n(\theta_0)$ are normal with a well-defined covariance matrices.

LEMMA 10. *For all* $\theta_0 \in \Theta_L$,

$$R(\theta_0)^{-1/2}\sqrt{n}\nabla_\theta L_n(\theta_0) \overset{\mathcal{D}}{\to} N(0,1) \quad \text{as} \quad n \to \infty, \tag{D.19}$$

where $R(\theta_0) \equiv \mathbb{E}[\nabla_\theta l_t(\theta_0)\nabla_\theta l_t(\theta_0)^\top]$. *Moreover, for all* $\theta_0 \in \Theta_U$,

$$R^*(\theta_0)^{-1/2}\sqrt{n}\nabla_{\theta^*} L_n(\theta_0) \overset{\mathcal{D}}{\to} N(0,1) \quad \text{as} \quad n \to \infty, \tag{D.20}$$

where $R^*(\theta_0) \equiv \mathbb{E}[\nabla_{\theta^*} l_t(\theta_0)\nabla_{\theta^*} l_t(\theta_0)^\top]$.

Proof. We first verify that $(\nabla_\theta l_t(\theta_0))_{t\in\mathbb{N}}$ and $(\nabla_{\theta^*} l_t(\theta_0))_{t\in\mathbb{N}}$ are martingale difference sequences (MD). Since $(b_{0t})_{t\in\mathbb{N}}$ is i.i.d. with the distribution, Beta$(1/2, \nu_0/2)$, for each t, we have

$$\mathbb{E}\left[\left.\frac{\partial l_t(\theta_0)}{\partial \delta}\right| \mathcal{F}_{t-1}\right] = \mathbb{E}\left[\left.\frac{\partial l_t(\theta_0)}{\partial \alpha}\right| \mathcal{F}_{t-1}\right] = \left[\left.\frac{\partial l_t(\theta_0)}{\partial \beta}\right| \mathcal{F}_{t-1}\right] = 0,$$

for all t. Moreover, we have

$$\mathbb{E}\left[\left.\ln\left(1 + \frac{z_t^2}{(\nu_0 - 2)}\right)\right| \mathcal{F}_{t-1}\right] = -\mathbb{E}\left[\ln(1 - b_{0t})|\, \mathcal{F}_{t-1}\right] = \psi_0\left(\frac{\nu_0 + 1}{2}\right) - \psi_0\left(\frac{\nu_0}{2}\right),$$

for all t by the properties of the beta distribution. Thus $\mathbb{E}\left[\partial l_t(\theta_0)/\partial\nu|\mathcal{F}_{t-1}\right] = 0$. Finally, we have $\mathbb{E}\left[\partial l_t(\theta_0)/\partial\gamma|\, \mathcal{F}_{t-1}\right] = 0$ if

$$\mathbb{E}\left[\left.\frac{z_t}{z_t^2 + (\nu_0 - 2)}\right| \mathcal{F}_{t-1}\right] = 0. \tag{D.21}$$

Computing the integral directly, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\frac{z_t}{z_t^2+(\nu_0-2)}\middle|\mathcal{F}_{t-1}\right] &\propto \int_{-\infty}^{\infty} \frac{z}{z^2+(\nu_0-2)}\left(1+\frac{z^2}{\nu_0-2}\right)^{-\frac{\nu_0+1}{2}} dz \\
&= \int_{-\pi/2}^{\pi/2} \frac{\tan x}{\tan^2 x+1}\left(1+\tan^2 x\right)^{-\frac{\nu_0+1}{2}} \sec^2 x \, dx \\
&\propto \int_{-\pi/2}^{\pi/2} \sin x \, (\cos x)^{\nu_0} \, dx \\
&= 0,
\end{aligned}
$$

for all $t \in \mathbb{N}$, where the second line is by the change of variable, $z/\sqrt{\nu_0-2} = \tan x$, the third line is by basic trigonometric identities, and the last line is by the fact that the integrand is an odd function for any $\nu_0 \in \mathbb{R}$. Thus (D.21) holds. Then, since $\nabla_\theta l_t(\theta_0)$ and $\nabla_{\theta^*} l_t(\theta_0)$ are integrable for all $\theta_0 \in \Theta_L$ and $\theta_0 \in \Theta_U$, respectively, by Lemma 5, $(\nabla l_t(\theta_0))_{t\in\mathbb{N}}$ and $(\nabla_{\theta^*} l_t(\theta_0))_{t\in\mathbb{N}}$ are MDs.

If $\theta_0 \in \Theta_L$, $\nabla_\theta l_t(\theta_0)$ is a strictly stationary and ergodic martingale difference with finite unconditional second moment by Theorem 2 and Lemma 9. Thus (D.19) holds at $\theta = \theta_0$ by the central limit theorem for stationary ergodic martingales [Theorem 6.11 of Varadhan (2001, p.144)].

For $\theta_0 \in \Theta_U$, we aim to show that

$$
\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\nabla_{\theta^*} l_t(\theta_0) \nabla_{\theta^*} l_t(\theta_0)^\top\right] \to R^*(\theta_0), \tag{D.22}
$$

$$
\mathbb{E}\left[\frac{\partial l_t(\theta_0)}{\partial \theta_i}\frac{\partial l_t(\theta_0)}{\partial \theta_j}\frac{\partial l_t(\theta_0)}{\partial \theta_k}\right] < \infty \text{ for all } t, \text{ and} \tag{D.23}
$$

$$
\frac{1}{n}\sum_{t=1}^{n} \nabla_{\theta^*} l_t(\theta_0) \nabla_{\theta^*} l_t(\theta_0)^\top \xrightarrow{P} R^*(\theta_0), \tag{D.24}
$$

where $i, j, k = 1, 2, 3$, convergence is as $n \to \infty$, and $R^*(\theta_0)$ is a deterministic and finite positive definite matrix for each $\theta_0 \in \Theta_U$. (D.22)-(D.24) imply that (D.20) holds for $\theta_0 \in \Theta_U$ by Proposition 7.9 of Hamilton (1994, p.194). (Note that the proofs for (D.22) and (D.23) presented below holds for all $\theta_0 \in \Theta$, whereas the proof for (D.24) holds only for $\theta_0 \in \Theta_U$.)

By the integrability of $\nabla_{\theta^*} l_t(\theta_0) \nabla_{\theta^*} l_t(\theta_0)^\top$ shown in Lemma 9, $\mathbb{E}\left[\nabla_{\theta^*} l_t(\theta_0) \nabla_{\theta^*} l_t(\theta_0)^\top\right] \equiv R_t^*(\theta_0)$ is a finite positive definite matrix for each $t \in \mathbb{N}$ and $\theta_0 \in \Theta$. Since $(R_t^*(\theta_0))_{t\in\mathbb{N}}$ is a deterministic sequence of real matrices, its sample average, $n^{-1}\sum_{t=1}^{n} R_t^*(\theta_0)$, converges to some constant positive definite matrix $R^*(\theta_0)$ as $n \to \infty$. (This convergence is verified easily by considering a special case of the law of large numbers where the sequence of i.i.d. random variables are replaced by a deterministic sequence.) Thus (D.22) holds for all

$\theta_0 \in \Theta$.

(D.23) holds if

$$\left\| \frac{\partial l_t(\theta_0)}{\partial \theta_i} \right\|_3 \left\| \frac{\partial l_t(\theta_0)}{\partial \theta_j} \right\|_3 \left\| \frac{\partial l_t(\theta_0)}{\partial \theta_k} \right\|_3 < \infty \tag{D.25}$$

for $i, j, k = 1, 2, 3$ and all $t \in \mathbb{N}$, since

$$|\mathbb{E}[XYZ]| \leq \|XYZ\|_1 \leq \|X\|_3 \|Y\|_3 \|Z\|_3$$

for any random variables X, Y, and Z. (D.25) holds if $\nabla_{\theta^*} l_t(\theta_0)$ is bounded in L^3. By Lemma 5, (D.25) holds for all $\theta_0 \in \Theta$. Thus, we have (D.23) for all $\theta_0 \in \Theta$.

Finally, we show (D.24) for $\theta_0 \in \Theta_U$. If we can show that $n^{-1} \sum_{t=1}^n \nabla_{\theta^*} l_t(\theta_0) \nabla_{\theta^*} l_t(\theta_0)^\top$ converges in probability to *some* constant positive definite matrix as $n \to \infty$, then the limiting quantity must be the same as the RHS of (D.22) by Proposition 1. Thus there is no need to verify that the limit of (D.22) and (D.24) are the same. We deal with the diagonal elements first. For all $\theta_0 \in \Theta_U$ and $i = 2, 3$, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial l_t(\theta_0)}{\partial \theta_i} \right)^2 = \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 (h_{\theta_i t}(\theta_0)^2 - u_{\theta_i t}(\theta_0)^2) \\ + \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 u_{\theta_i t}(\theta_0)^2, \end{aligned}$$

where the RHS converges in probability to $((\nu_0 + 1)^2/4) \mathrm{Var}(b_{0t}) \mathbb{E}[u_{it}(\theta_0)^2] < \infty$ by Lemma 7 (i)(ii)(vii). (Note also that b_{0t} and $u_{it}(\theta_0)$ are independent for all t.)

Next, we have

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \left(\frac{\partial l_t(\theta_0)}{\partial \nu} \right)^2 \\ &= \frac{1}{n} \sum_{t=1}^n \left[\frac{(\nu_0 + 1) b_{0t} - 1}{2(\nu_0 - 2)} - \frac{1}{2} \ln \left(1 + \frac{z_t^2}{\nu_0 - 2} \right) + \frac{1}{2} \left(\psi_0 \left(\frac{\nu_0 + 1}{2} \right) - \psi_0 \left(\frac{\nu_0}{2} \right) \right) \right]^2 \\ &+ \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 \left(h_{\nu t}(\theta_0)^2 - u_{\nu t}(\theta_0)^2 \right) \\ &+ \frac{2}{n} \sum_{t=1}^n \left\{ \left[\frac{(\nu_0 + 1) b_{0t} - 1}{2(\nu_0 - 2)} - \frac{1}{2} \ln \left(1 + \frac{z_t^2}{\nu_0 - 2} \right) + \frac{1}{2} \left(\psi_0 \left(\frac{\nu_0 + 1}{2} \right) - \psi_0 \left(\frac{\nu_0}{2} \right) \right) \right] \right. \\ &\qquad \left. \times \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right) (h_{\nu t}(\theta_0) - u_{\nu t}(\theta_0)) \right\} \\ &+ \frac{1}{n} \sum_{t=1}^n \left(\frac{(\nu_0 + 1) b_{0t} - 1}{2} \right)^2 u_{\nu t}(\theta_0)^2 \end{aligned}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left\{\left[\frac{(\nu_0+1)b_{0t}-1}{2(\nu_0-2)}-\frac{1}{2}\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{1}{2}\left(\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)\right)\right]\right.$$
$$\left.\times\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)u_{\nu t}(\theta_0)\right\}$$

The summand of the first, fourth, and fifth terms are stationary and ergodic by Lemma 7 (i). By (D.9) and (D.11) of Lemma 7 (vii), and by the properties that $u_{\nu t}(\theta_0)$ and $h_{\nu t}(\theta_0)$ are independent of b_{0t} and z_t, the second and third terms converge to zero in L^1 for all $\theta_0 \in \Theta_U$. Then we obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta_0)}{\partial \nu}\right)^2 \xrightarrow{P} & \operatorname{Var}\left(\frac{(\nu_0+1)b_{0t}}{2(\nu_0-2)}-\frac{1}{2}\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)\right)\\ &+\frac{(\nu_0+1)^2}{4}\operatorname{Var}(b_{0t})\mathbb{E}[u_{\nu t}(\theta_0)^2]\\ &+\frac{\nu_0+1}{2}\operatorname{Cov}\left(\frac{(\nu_0+1)b_{0t}}{\nu_0-2}-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right), b_{0t}\right)\mathbb{E}[u_{\nu t}(\theta_0)]\\ &<\infty\end{aligned}$$

as $n\to\infty$ for all $\theta_0\in\Theta_U$.

Next, we consider the off-diagonal elements of (D.24). We have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial \theta_i}\frac{\partial l_t(\theta_0)}{\partial \theta_j} &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2\left(h_{\theta_i t}(\theta_0)h_{\theta_j t}(\theta_0)-u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)\right)\\ &\quad+\frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2 u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)\\ &\xrightarrow{P}\frac{(\nu_0+1)^2}{4}\operatorname{Var}(b_{0t})\mathbb{E}[u_{\theta_i t}(\theta_0)u_{\theta_j t}(\theta_0)]<\infty\end{aligned}$$

for all $\theta_0\in\Theta_U$, $i,j=2,3$, and $i\neq j$ by Lemma 7 (i)(ii)(vii). Similarly, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial \beta}\frac{\partial l_t(\theta_0)}{\partial \nu} &= \frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2\left(h_{\beta t}(\theta_0)h_{\nu t}(\theta_0)-u_{\beta t}(\theta_0)u_{\nu t}(\theta_0)\right)\\ &+\frac{1}{4n}\sum_{t=1}^{n}\left\{\left(\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right)\right.\\ &\qquad\left.\times\left((\nu_0+1)b_{0t}-1\right)\left(h_{\beta t}(\theta_0)-u_{\beta t}(\theta_0)\right)\right\}\\ &+\frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2 u_{\beta t}(\theta_0)u_{\nu t}(\theta_0)\end{aligned}$$

$$+\frac{1}{4n}\sum_{t=1}^{n}\left\{\left(\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right)\right.$$
$$\left.\times\left((\nu_0+1)b_{0t}-1\right)u_{\beta t}(\theta_0)\right\}$$
$$\xrightarrow{P}\frac{(\nu_0+1)^2}{4}\mathrm{Var}(b_{0t})\mathbb{E}\left[u_{\beta t}(\theta_0)u_{\nu t}(\theta_0)\right]$$
$$+\frac{\nu_0+1}{4}\mathrm{Cov}\left(\frac{(\nu_0+1)b_{0t}}{\nu_0-2}-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right),b_{0t}\right)\mathbb{E}\left[u_{\beta t}(\theta_0)\right]<\infty.$$

for all $\theta_0\in\Theta_U$ by Lemma 7 (i)(ii)(vii). Analogous derivations show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\alpha}\frac{\partial l_t(\theta_0)}{\partial\nu}\xrightarrow{P}\frac{(\nu_0+1)^2}{4}\mathrm{Var}(b_{0t})\mathbb{E}\left[u_{\alpha t}(\theta_0)u_{\nu t}(\theta_0)\right]$$
$$+\frac{\nu_0+1}{4}\mathrm{Cov}\left(\frac{(\nu_0+1)b_{0t}}{\nu_0-2}-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right),b_{0t}\right)\mathbb{E}\left[u_{\alpha t}(\theta_0)\right].$$

for all $\theta_0\in\Theta_U$. This completes the proof of (D.24) for $\theta_0\in\Theta_U$. ■

In the next lemma, we show that, if $\theta_0\in\Theta_U$, the joint log-likelihood function is asymptotically flat in the δ and γ dimensions, so that the consistency and asymptotic normality of MLE do not hold for these parameters when $\theta_0\in\Theta_U$.

LEMMA 11. *For $i=4,5$ and $j=1,\ldots,5$, we have*

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\xrightarrow{P}0.$$

when $\theta_0\in\Theta_U$.

Proof. For all $\theta_0\in\Theta_U$ and $i=4$ (i.e. $\theta_i=\delta$), we have

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta_0)}{\partial\delta}\right)^2=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2\left(h_{\delta t}(\theta_0)^2-u_{\delta t}(\theta_0)^2\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n}\left(\frac{(\nu_0+1)b_{0t}-1}{2}\right)^2u_{\delta t}(\theta_0)^2,$$

where the RHS converges in probability to $((\nu_0+1)^2/4)\mathrm{Var}(b_{0t})\mathbb{E}[u_{\delta t}(\theta_0)^2]=0$ by Lemma 7 (i)(ii)(vii). Next, we have

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta_0)}{\partial\gamma}\right)^2=\frac{1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2\left(h_{\gamma t}(\theta_0)^2-u_{\gamma t}(\theta_0)^2\right)$$
$$+\frac{\nu_0+1}{n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)\left(h_{\gamma t}(\theta_0)-u_{\gamma t}(\theta_0)\right)\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}$$
$$+\frac{(\nu_0+1)^2}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{(\varepsilon_t^2+(\nu_0-2)h_{0t})^2}.$$

The first two terms on the RHS converges in probability to zero by (D.10) and (D.11) of Lemma 7 (vii). The third term converges in L^1 to zero because

$$0 \leq \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{(\varepsilon_t^2 + (\nu_0 - 2)h_{0t})^2} \right\|_1 \leq \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}} \right\|_2^2 \to 0$$

as $n \to \infty$ for all $\theta_0 \in \Theta_U$ by Lemma 1. Thus $n^{-1} \sum_{t=1}^{n} (\partial l_t(\theta_0)/\partial \gamma)^2 \xrightarrow{P} 0$ for all $\theta_0 \in \Theta_U$. We also have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta_i} \frac{\partial l_t(\theta_0)}{\partial \theta_j} &= \frac{1}{n} \sum_{t=1}^{n} \left(\frac{(\nu_0 + 1)b_{0t} - 1}{2} \right)^2 \left(h_{\theta_i t}(\theta_0) h_{\theta_j t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^{n} \left(\frac{(\nu_0 + 1)b_{0t} - 1}{2} \right)^2 u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0) \\ &\xrightarrow{P} \frac{(\nu_0 + 1)^2}{4} \mathrm{Var}(b_{0t}) \mathbb{E}[u_{\theta_i t}(\theta_0) u_{\theta_j t}(\theta_0)] = 0 \end{aligned}$$

for all $\theta_0 \in \Theta_U$, $i = 4$, and $j = 2, 3$ by Lemma 7 (i)(ii)(vii). Analogous derivations show that

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \delta} \frac{\partial l_t(\theta_0)}{\partial \nu} \xrightarrow{P} 0$$

for all $\theta_0 \in \Theta_U$. Next, for $i = 2, 3, 4$, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta_i} \frac{\partial l_t(\theta_0)}{\partial \gamma} &= \frac{1}{4n} \sum_{t=1}^{n} \left((\nu_0 + 1)b_{0t} - 1 \right)^2 \left(h_{\theta_i t}(\theta_0) h_{\gamma t}(\theta_0) - u_{\theta_i t}(\theta_0) u_{\gamma t}(\theta_0) \right) \\ &\quad + \frac{\nu_0 + 1}{2n} \sum_{t=1}^{n} \left((\nu_0 + 1)b_{0t} - 1 \right) h_{\theta_i t}(\theta_0) \frac{\varepsilon_t}{\varepsilon_t^2 + (\nu_0 - 2)h_{0t}}. \end{aligned}$$

By (D.11) of Lemma 7 (vii), the first term converges in L^1 to zero for all $\theta_0 \in \Theta_U$. The second term also converges in L^1 to zero for all $\theta_0 \in \Theta_U$ by the Minkowski and Hölder inequalities, Lemma 2 (i), and Lemma 1. Thus $n^{-1} \sum_{t=1}^{n} (\partial l_t(\theta_0)/\partial \theta_i)(\partial l_t(\theta_0)/\partial \gamma) \xrightarrow{P} 0$ for all $\theta_0 \in \Theta_U$ and $i = 2, 3, 4$. Finally, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\nu}\frac{\partial l_t(\theta_0)}{\partial\gamma}\\
&\quad=\frac{1}{4n}\sum_{t=1}^{n}\left\{\left[\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right]\right.\\
&\qquad\qquad\left.\times\left((\nu_0+1)b_{0t}-1\right)\left(h_{\gamma t}(\theta_0)-u_{\gamma t}(\theta_0)\right)\right\}\\
&\qquad+\frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2\left(h_{\gamma t}(\theta_0)h_{\nu t}(\theta_0)-u_{\gamma t}(\theta_0)u_{\nu t}(\theta_0)\right)\\
&\qquad+\frac{\nu_0+1}{2n}\sum_{t=1}^{n}\left\{\left[\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right]\right.\\
&\qquad\qquad\left.\times\left(\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}\right)\right\}\\
&\qquad+\frac{\nu_0+1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)\left(h_{\nu t}(\theta_0)-u_{\nu t}(\theta_0)\right)\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}\\
&\qquad+\frac{1}{4n}\sum_{t=1}^{n}\left[\psi_0\left(\frac{\nu_0+1}{2}\right)-\psi_0\left(\frac{\nu_0}{2}\right)-\ln\left(1+\frac{z_t^2}{\nu_0-2}\right)+\frac{(\nu_0+1)b_{0t}-1}{\nu_0-2}\right]\times\\
&\qquad\quad\times\left((\nu_0+1)b_{0t}-1\right)u_{\gamma t}(\theta_0)\\
&\qquad+\frac{1}{4n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)^2u_{\gamma t}(\theta_0)u_{\nu t}(\theta_0)\\
&\qquad+\frac{\nu_0+1}{2n}\sum_{t=1}^{n}\left((\nu_0+1)b_{0t}-1\right)u_{\nu t}(\theta_0)\frac{\varepsilon_t}{\varepsilon_t^2+(\nu_0-2)h_{0t}}.
\end{aligned}
$$

Note that $u_{\gamma t}(\theta_0)=0$ for all $t\in\mathbb{N}$ and $\theta_0\in\Theta$ by definition. By (D.9) and (D.11) of Lemma 7 (vii), the first, second, and fourth terms converge in L^1 to zero. The third and seventh terms converge in L^1 to zero by the Minkowski and Hölder inequalities, Lemma 1, and Lemma 7 (ii). Thus we have shown that $n^{-1}\sum_{t=1}^{n}(\partial l_t(\theta_0)/\partial\nu)(\partial l_t(\theta_0)/\partial\gamma)\xrightarrow{P}0$ for all $\theta_0\in\Theta_U$.

■

Next, we aim to show that the elements of $\nabla^3_{\theta}L_n(\theta)$ and $\nabla^3_{\theta^*}L_n(\theta^*)$ are bounded by some stationary processes for all $n\in\mathbb{N}_{>0}$ and $\theta_0,\theta\in\Theta$. For this purpose, we introduce the stationary process $w_t(\theta_0)$ defined by

$$\begin{aligned} w_t(\theta_0) = 1 &+ \frac{\delta_u - \delta_0}{\delta_0}\frac{1}{1-\beta_u} + \frac{\beta_u - \beta_0}{\beta_l}\sum_{k=1}^{t}\prod_{j=1}^{k}\frac{\beta_u}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \\ &+ \frac{\alpha_u(\nu_u+1)}{\beta_l(\nu_l-2)}\sum_{k=1}^{t}\frac{1}{b_{0t-k}}\left(|z_{t-k}| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2\prod_{j=1}^{k}\frac{\beta_u}{\beta_0+\alpha_0(\nu_0+1)b_{0t-k}}, \end{aligned} \tag{D.26}$$

for any $0 < \beta_l \leq \beta \leq \beta_u < 1$ and $\theta_0 \in \Theta$, where $\bar{g} \equiv \max\{\gamma_u - \gamma_0, \gamma_0 - \gamma_l\}$.

LEMMA 12. *There exists $\beta_u \in (0,1)$ such that $w_t(\theta_0)$ is strictly stationary and ergodic for each $p \geq 1$, $t \in \mathbb{N}$, $0 < \beta_l \leq \beta \leq \beta_u < 1$, and $\theta_0 \in \Theta$.*

Proof. First, note that

$$\begin{aligned} &\mathbb{E}\left[\max\left\{0, \log\left(\frac{1}{b_{0t}}\left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2\right)\right\}\right] \\ &\leq \mathbb{E}\left[\max\left\{0, -\log\left(b_{0t}\right)\right\}\right] + \mathbb{E}\left[\max\left\{0, 2\log\left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)\right\}\right] \\ &< \infty \end{aligned}$$

for all $t \in \mathbb{N}$ by the property of the beta and Student's t distributions. Moreover,

$$\mathbb{E}\left[\log\left(\frac{\beta}{\beta_0+\alpha_0(\nu_0+1)b_{0t}}\right)\right] \leq \mathbb{E}\left[\frac{\beta_u}{\beta_0+\alpha_0(\nu_0+1)b_{0t}}\right] < 1$$

for some $\beta_u \in (0,1)$ as b_{0t} is non-degenerate for all $t \in \mathbb{N}$. Thus, the proof is complete by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974).

■

In order to show that the elements of $\nabla^3_\theta L_n(\theta)$ and $\nabla^3_{\theta^*} L_n(\theta^*)$ are bounded by some stationary processes for all $n \in \mathbb{N}_{>0}$ and $\theta_0, \theta \in \Theta$, we show in Lemma 15 that $h_{\theta_i}(\theta)$, $h_{\theta_i\theta_j}(\theta)$, and $h_{\theta_i\theta_j\theta_k}(\theta)$ are bounded by some stationary processes for all t, $\theta, \theta_0 \in \Theta$, and $i, j, k = 1, \ldots, 5$. In order to show Lemma 15, we use the properties of $h_t(\theta)$ and h_{0t} shown in Lemmas 13 and 14. Lemma 14 is the only place where we use the unit upper-bound on β.

LEMMA 13. *For all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$, we have*

$$\begin{aligned} h_t(\theta) = h_{0t} &+ (\delta - \delta_0)\sum_{k=0}^{t-1}\beta^k + (\beta - \beta_0)\sum_{k=1}^{t}\beta^{k-1}h_{0t-k} \\ &+ \sum_{k=1}^{t}\beta^k\Big[\alpha(\nu+1)b_{t-k}(\theta)h_{t-k}(\theta) - \alpha_0(\nu_0+1)b_{0t-k}h_{0t-k}\Big], \end{aligned} \tag{D.27}$$

$$
\begin{aligned}
h_{0t} &= h_t(\theta) + (\delta_0 - \delta) \sum_{k=0}^{t-1} \beta_0^k + (\beta_0 - \beta) \sum_{k=1}^{t} \beta_0^{k-1} h_{t-k}(\theta) \\
&\quad + \sum_{k=1}^{t} \beta_0^k \Big[\alpha_0(\nu_0 + 1) b_{0t-k} h_{0t-k} - \alpha(\nu+1) b_{t-k}(\theta) h_{t-k}(\theta) \Big].
\end{aligned} \tag{D.28}
$$

Proof. Since $\delta_0 = h_{0t} - \beta_0 h_{0t-1} - \alpha_0(\nu_0 + 1) b_{0t-1} h_{0t-1}$, adding and subtracting δ_0 in the equation for $h_t(\theta)$ give

$$
\begin{aligned}
h_t(\theta) = \delta - \delta_0 + \beta h_{t-1}(\theta) + h_{0t} - \beta_0 h_{0t-1} - \alpha_0(\nu_0 + 1) b_{0t-1} h_{0t-1} \\
+ \alpha(\nu + 1) b_{t-1}(\theta) h_{t-1}(\theta).
\end{aligned}
$$

Then (D.27) follows by noting that $h_0(\theta) = h_{00} = \omega_0$. Similarly,

$$
\begin{aligned}
h_{0t} = \delta_0 - \delta + \beta_0 h_{0t-1} + h_t(\theta) - \beta h_{t-1}(\theta) - \alpha(\nu + 1) b_{t-1}(\theta) h_{t-1}(\theta) \\
+ \alpha_0(\nu_0 + 1) b_{0t-1} h_{0t-1}.
\end{aligned}
$$

Then (D.28) follows. ■

LEMMA 14. *For all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$, we have*

(i)

$$
0 \le \frac{b_t(\theta) h_t(\theta)}{h_{0t}} \le \frac{1}{\nu - 2} \left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}} \right)^2.
$$

(ii) Define $q_t \equiv q_t(\theta_0) \equiv 1\!\!1_{\{z_t \ge g_u\}}(z_t + g_u/\sqrt{\delta_0})^2 + 1\!\!1_{\{z_t \le g_l\}}(z_t + g_l/\sqrt{\delta_0})^2$, where $g_u \equiv \gamma_0 - \gamma_l$ and $g_l \equiv \gamma_0 - \gamma_u$. Then

$$
0 \le \frac{q_t(\theta_0)}{q_t(\theta_0) + (\nu_u - 2) h_t(\theta)/h_{0t}} \le b_t(\theta) \le \frac{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2}{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2 + (\nu_l - 2) h_t(\theta)/h_{0t}} \le 1
$$

a.s. for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$.

(iii)

$$
0 < h_t(\theta)/h_{0t} \le w_t(\theta_0) \tag{D.29}
$$

a.s. for some strictly stationary process $w_t(\theta_0)$ for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Moreover,

$$
0 < x_t(\theta_0) \le h_t(\theta)/h_{0t} \tag{D.30}
$$

a.s. for some strictly stationary process $x_t(\theta_0)$ for all $t \in \mathbb{N}$, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$.

(iv) $0 \le \underline{b}_t(\theta_0) \le b_t(\theta)$, where $\underline{b}_t(\theta_0)$ is some strictly stationary process, for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$. Moreover, $0 \le b_t(\theta) \le \bar{b}_t(\theta_0) \le 1$, where $\bar{b}_t(\theta_0)$ is some strictly stationary processes, for all $t \in \mathbb{N}$, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$.

Proof. (i)-(ii) Note that we have

$$q_t \le \left(z_t + \frac{g}{\sqrt{h_{0t}}}\right)^2 \le \left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2. \tag{D.31}$$

Then, for all $t \in \mathbb{N}$ and $\theta, \theta_0 \in \Theta$, we have a.s.

$$\begin{aligned}
\frac{b_t(\theta)h_t(\theta)}{h_{0t}} &= \frac{\left(z_t + g/\sqrt{h_{0t}}\right)^2 h_t(\theta)/h_{0t}}{\left(z_t + g/\sqrt{h_{0t}}\right)^2 + (\nu-2)h_t(\theta)/h_{0t}} \\
&\le \frac{1}{\nu-2}\left(z_t + \frac{g}{\sqrt{h_{0t}}}\right)^2 \le \frac{1}{\nu_l - 2}\left(|z_t| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2.
\end{aligned}$$

Likewise, (ii) also follows from (D.31).

(iii) First, we show (D.29). By Lemma 13, we obtain, for all $\theta, \theta_0 \in \Theta$ and $t \in \mathbb{N}$,

$$\begin{aligned}
\frac{h_t(\theta)}{h_{0t}} &= 1 + \frac{\delta - \delta_0}{h_{0t}} \sum_{k=0}^{t-1} \beta^k + \frac{\beta - \beta_0}{\beta} \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta h_{0t-j}}{h_{0t-j+1}} \\
&\quad + \frac{\alpha_0(\nu_0+1)}{\beta} \sum_{k=1}^{t} \beta^k \frac{b_{0t-k}h_{0t-k}}{h_{0t}} \left[\frac{\alpha(\nu+1)b_{t-k}(\theta)h_{t-k}(\theta)}{\alpha_0(\nu_0+1)b_{0t-k}h_{0t-k}} - 1\right] \\
&\le 1 + \frac{\delta_u - \delta_0}{\delta_0}\frac{1}{1-\beta_u} + \frac{\beta_u - \beta_0}{\beta_l} \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_u}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \\
&\quad + \frac{\alpha_u(\nu_u+1)}{\beta_l(\nu_l-2)} \sum_{k=1}^{t} \frac{1}{b_{0t-k}} \left(|z_{t-k}| + \frac{\bar{g}}{\sqrt{\delta_0}}\right)^2 \prod_{j=1}^{k} \frac{\beta_u}{\beta_0 + \alpha_0(\nu_0+1)b_{0t-j}} \\
&= w_t(\theta_0),
\end{aligned}$$

where the inequality in the middle is by Lemma 14 (i).

Next, we show (D.30). By Lemma 14 (ii) and the just derived inequality, we have

$$b_t(\theta) \ge \frac{q_t(\theta_0)}{q_t(\theta_0) + (\nu_u - 2)w_t(\theta_0)} \equiv \underline{b}_t(\theta_0) \ge 0,$$

where the process $\underline{b}_t(\theta_0)$ is in terms of the i.i.d. process z_t for any $\theta_0 \in \Theta$ and $t \in \mathbb{N}$. $\underline{b}_t(\theta_0)$ is strictly stationary and ergodic by Lemma 12 and Theorem 3.5.8 of Stout (1974) [also see the relevant results in Royden (1988, p.66-68)]. Note that, by (D.28) of Lemma 13,

$$\begin{aligned}
\frac{h_{0t}}{h_t(\theta)} &= 1 + \frac{\delta_0 - \delta}{h_t(\theta)} \sum_{k=0}^{t-1} \beta_0^k + \frac{\beta_0 - \beta}{\beta_0} \sum_{k=1}^{t} \beta_0^k \frac{h_{t-k}(\theta)}{h_t(\theta)} \\
&\quad + \sum_{k=1}^{t} \frac{\beta_0^k}{h_t(\theta)} \left[\alpha_0(\nu_0+1)b_{0t-k}h_{0t-k} - \alpha(\nu+1)b_{t-k}(\theta)h_{t-k}(\theta)\right].
\end{aligned}$$

Since $a_{t-k}/a_t = \prod_{j=1}^{k} a_{t-j}/a_{t-j+1}$ for any sequence $(a_t)_{t\in\mathbb{N}}$ and $0 < k < t$, we get

$$
\begin{aligned}
0 &\leq \frac{h_{0t}}{h_t(\theta)} \\
&= \left| 1 + \frac{\delta_0 - \delta}{h_t(\theta)} \sum_{k=0}^{t-1} \beta_0^k + \frac{\beta_0 - \beta}{\beta_0} \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_0 h_{t-j}(\theta)}{h_{t-j+1}(\theta)} - \alpha(\nu+1) \sum_{k=1}^{t} b_{t-k}(\theta) \prod_{j=1}^{k} \frac{\beta_0 h_{t-j}(\theta)}{h_{t-j+1}(\theta)} \right| \\
&\quad \div \left| 1 - \alpha_0(\nu_0+1) \sum_{k=1}^{t} b_{0t-k} \prod_{j=1}^{k} \frac{\beta_0 h_{0t-j}}{h_{0t-j+1}} \right| \qquad \text{(D.32)}
\end{aligned}
$$

The numerator of (D.32) is bounded above by

$$
\begin{aligned}
1 &+ \frac{\delta_u - \delta_l}{\delta_l(1-\beta_u)} + \frac{\beta_u - \beta_l}{\beta_0} \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_0}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)} \\
&+ \alpha_u(\nu_u+1) \sum_{k=1}^{t} \prod_{j=1}^{k} \frac{\beta_0}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)}.
\end{aligned}
$$

Since $\underline{b}_t(\theta_0)$ is non-degenerate, strictly stationary and ergodic, there exists $\beta_l \in (0,1)$ such that this quantity is strictly stationary and ergodic by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974). In the denominator of (D.32), we have

$$
\sum_{k=1}^{t} b_{0t-k} \prod_{j=1}^{k} \frac{\beta_0 h_{0t-j}}{h_{0t-j+1}} = \sum_{k=1}^{t} b_{0t-k} \prod_{j=1}^{k} \frac{\beta_0}{\delta_0/h_{0t-j} + \beta_0 + \alpha_0(\nu_0+1) b_{0t-j}} \qquad \text{(D.33)}
$$

Since $\max\{0, \log|X|\} \leq |X|$ for any real valued random variable X, we have $\mathbb{E}[\max\{0, |b_{0t}|\}] \leq \mathbb{E}[|b_{0t}|] < 1$ for all $t \in \mathbb{N}$ and $\theta_0 \in \Theta$. Moreover,

$$
\mathbb{E}\left[\log\left|\frac{\beta_0}{\delta_0/h_{0t} + \beta_0 + \alpha_0(\nu_0+1) b_{0t}}\right|\right] < 0
$$

for all $t \in \mathbb{N}$ and $\theta_0 \in \Theta$. Thus, the denominator of (D.32) is strictly stationary and ergodic if $\theta_0 \in \Theta_L$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974).[11] Thus, we have found a strictly stationary process $x_t(\theta_0)$ such that $0 < x_t(\theta_0) \leq h_t(\theta)/h_{0t}$ for all $t \in \mathbb{N}$, $\theta \in \Theta$ and $\theta_0 \in \Theta_L$.

(iv) By Lemma 14 (iii), we obtained $b_t(\theta) \geq \underline{b}_t(\theta_0) \geq 0$. Moreover, by Lemma

[11] This is the only place that limits us from proving the consistency and asymptotic normality results for the nonstationary case (i.e. when $\theta_0 \in \Theta_U$). In order to show the asymptotic properties for $\theta_0 \in \Theta_U$, we would need to find a strictly stationary and ergodic process that bounds the denominator of (D.32) from below. We find showing this difficult when the RHS of (D.33) is greater than or equal to one.

14 (ii)(iii), we have

$$b_t(\theta) \leq \frac{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2}{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2 + (\nu_l - 2)h_t(\theta)/h_{0t}}$$
$$\leq \frac{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2}{\left(|z_t| + \bar{g}/\sqrt{\delta_0}\right)^2 + (\nu_l - 2)x_t(\theta_0)} \equiv \bar{b}_t(\theta_0),$$

where, for all $\theta \in \Theta$ and $\theta_0 \in \Theta_L$, $\bar{b}_t(\theta) \in [0,1]$ and $(\bar{b}_t(\theta))_{t\in\mathbb{N}}$ is strictly stationary and ergodic by Theorem 3.5.8 of Stout (1974) and Royden (1988, p.66-68). ■

Finally, in order to show Lemma 15, define the following process.

$$u^*_{\theta_i t}(\theta_0) \equiv \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i}}{\beta_l} \prod_{j=1}^{k} m_{t-j}(\theta_0)$$

for $i = 1, \ldots, 5$, where

$$m_t(\theta_0) \equiv \max\left\{ \frac{\beta_u + \alpha_u(\nu_u+1)\underline{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)}, \frac{\beta_u + \alpha_u(\nu_u+1)\bar{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)\bar{b}_t(\theta_0)} \right\}$$

is strictly stationary and ergodic by Theorem 3.5.8 of Stout (1974) and Royden (1988, p.66-68). $\widetilde{u}_{\theta_i}$ bounds $\widehat{h}_{\theta_i t}(\theta)$ for all t, $\theta \in \Theta$, and $i = 1, \ldots, 5$. We set

$$\widetilde{u}_\delta = 1/\delta_l, \qquad \widetilde{u}_\alpha = \nu_u + 1, \qquad \widetilde{u}_\beta = 1,$$
$$\widetilde{u}_\gamma = 2\alpha_u(\nu_u+1)\max\{1, ((\nu_l - 2)\delta_l)^{-1}\},$$
$$\widetilde{u}_\nu = \alpha_u + \alpha_u(\nu_u+1)(\nu_l-2)^{-1}.$$

Moreover, define the following process;

$$u^*_{\theta_i\theta_j t}(\theta_0) \equiv \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i\theta_j t-k}(\theta_0)}{\beta_l} \prod_{j=1}^{k} m_{t-j}(\theta_0)$$

for $i, j = 1, \ldots, 5$, where $\widetilde{u}_{\theta_i\theta_j t}(\theta_0)$ bounds $\widehat{h}_{\theta_i\theta_j t}(\theta)$ for any t, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$. They are defined in Appendix B.2. Similarly, we define

$$u^*_{\theta_i\theta_j\theta_m t}(\theta_0) \equiv \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i\theta_j\theta_m t-k}(\theta_0)}{\beta_l} \prod_{j=1}^{k} m_{t-j}(\theta_0)$$

for $i, j, m = 1, \ldots, 5$, where $\widetilde{u}_{\theta_i\theta_j\theta_m t}(\theta_0)$ bounds $\widehat{h}_{\theta_i\theta_j\theta_m t}(\theta)$ for any t, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$. For instance, for $i = j = m = 3$ (i.e. $\theta_i = \theta_j = \theta_m = \beta$), we set

$$\widetilde{u}_{\beta\beta\beta t}(\theta_0) = 3u^*_{\beta\beta t}(\theta_0) + 6\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)^3 + 2\alpha_u(\nu_u+1)u^*_{\beta t}(\theta_0)u^*_{\beta\beta t}(\theta_0).$$

Lemma 15 establishes some of the useful properties of these processes.

LEMMA 15. *For all $\theta \in \Theta$, $\theta_0 \in \Theta_L$, and $i, j, m = 1, \ldots, 5$;*

(i) $|h_{\theta_i t}(\theta)| \leq u^*_{\theta_i t}(\theta_0)$ for all $t \in \mathbb{N}$.

(ii) $|h_{\theta_i\theta_j t}(\theta)| \leq u^*_{\theta_i\theta_j t}(\theta_0)$ *for all* $t \in \mathbb{N}$.

(iii) $|h_{\theta_i\theta_j\theta_m t}(\theta)| \leq u^*_{\theta_i\theta_j\theta_m t}(\theta_0)$ *for all* $t \in \mathbb{N}$.

(iv) $\left(u^*_{\theta_i t}(\theta_0)\right)_{t\in\mathbb{N}}$, $\left(u^*_{\theta_i\theta_j t}(\theta_0)\right)_{t\in\mathbb{N}}$, *and* $\left(u^*_{\theta_i\theta_j\theta_m t}(\theta_0)\right)_{t\in\mathbb{N}}$ *are strictly stationary and ergodic.*

Proof.

(i) It is easy to show that $\left|\widehat{h}_{\theta_i t}(\theta)\right| < \widetilde{u}_{\theta_i}$ for all t and $i = 1, \ldots, 5$. Note that $|\widehat{h}_{\gamma t}(\theta)| < \widetilde{u}_\gamma$ can be verified by the condition, (D.1), of Lemma 1. Then, by the condition, (D.3), of Lemma 2 (i), we obtain

$$\begin{aligned}
|h_{\theta_i t}(\theta)| &\leq \sum_{k=1}^{t} \frac{|\widehat{h}_{\theta_i t}(\theta)|}{\beta_l} \prod_{j=1}^{k} \frac{h_{t-j}(\theta)(\beta + \alpha(\nu+1)b_{t-j}(\theta)^2)}{h_{t-j+1}(\theta)} \\
&\leq \sum_{k=1}^{t} \frac{\widetilde{u}_{\theta_i}}{\beta_l} \prod_{j=1}^{k} \frac{\beta_u + \alpha_u(\nu_u+1)b_{t-j}(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)b_{t-j}(\theta_0)} \\
&\leq u^*_{\theta_i t}(\theta_0),
\end{aligned}$$

where the last inequality used the fact that

$$\begin{aligned}
\frac{\beta + \alpha(\nu+1)b_t(\theta)^2}{\beta + \alpha(\nu+1)b_t(\theta)} &\leq \max\left\{\frac{\beta_u + \alpha_u(\nu_u+1)\underline{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)\underline{b}_t(\theta_0)}, \frac{\beta_u + \alpha_u(\nu_u+1)\bar{b}_t(\theta_0)^2}{\beta_l + \alpha_l(\nu_l+1)\bar{b}_t(\theta_0)}\right\} \\
&\equiv m_t(\theta)
\end{aligned}$$

for all t, $\theta \in \Theta$, $\theta_0 \in \Theta_L$, and $i = 1, \ldots, 5$ by Lemma 14 (iv).

(ii) Derivations analogous to Lemma 15 (i) show that $|h_{\theta_i\theta_j t}(\theta)| \leq u^*_{\theta_i\theta_j t}(\theta_0)$ for all t, $\theta \in \Theta$, $\theta_0 \in \Theta_L$, and $i, j = 1, \ldots, 5$. Note that we can verify $|\widehat{h}_{\theta_i\theta_j t}(\theta)| < \widetilde{u}_{\theta_i\theta_j t}(\theta_0)$ whenever $i = 4$ or $j = 4$ (i.e. $\theta_i = \theta_4 = \gamma$ or $\theta_j = \gamma$) by the condition, (D.1), of Lemma 1 and by noting that

$$\left|\frac{b_t(\theta)}{e_t}\right| = \frac{|e_t|}{e_t^2 + (\nu-2)h_t(\theta)},$$

which is bounded a.s. by Lemma 1.

(iii) This proof is analogous to the proofs for Lemma 15 (i)(ii).

(iv) $(m_t(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary, and we can find $(\beta_u, \beta_l, \alpha_u, \alpha_l, \nu_u, \nu_l)$ such that $\mathbb{E}\left[\ln\left(m_t(\theta_0)\right)\right] < 0$ for all $\theta_0 \in \Theta_L$ since $\bar{b}(\theta_0) \in (0,1)$ and $\underline{b}(\theta_0) \in (0,1)$ are non-degenerate. Moreover, using the property that $\ln(x) \leq x - 1$ for all $x > 0$, we have

$$\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i}|\}\right] \leq \mathbb{E}\left[\max\{0, |\widetilde{u}_{\theta_i}|\}\right] = \mathbb{E}\left[|\widetilde{u}_{\theta_i}|\right] < \infty$$

for $i = 1, \ldots, 5$. Then $(u_{\theta_i t}(\theta_0))_{t\in\mathbb{N}}$ is strictly stationary and ergodic for all $\theta_0 \in \Theta_L$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974).

Likewise, we can show that $\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i\theta_j t}(\theta_0)|\}\right] < \infty$ and $\mathbb{E}\left[\max\{0, \ln|\widetilde{u}_{\theta_i\theta_j\theta_m t}(\theta_0)|\}\right] < \infty$ for all t, $\theta_0 \in \Theta$, and $i, j, m = 1, \ldots, 5$. Then we can deduce that $(u^*_{\theta_i\theta_j t}(\theta_0))_{t\in\mathbb{N}}$ and $(u^*_{\theta_i\theta_j\theta_m t}(\theta_0))_{t\in\mathbb{N}}$ are strictly stationary and ergodic for any $\theta_0 \in \Theta_L$ and $i, j, m = 1, \ldots, 5$ by Theorem 1 of Brandt (1986) and Theorem 3.5.8 of Stout (1974). ■

We are now ready to show that the elements of $\nabla^3_\theta L_n(\theta)$ are bounded by some stationary and ergodic sequence for all $n \in \mathbb{N}_{>0}$, $\theta \in \Theta$, and $\theta_0 \in \Theta_L$.

LEMMA 16. *For any $\theta_0 \in \Theta_L$ and $n \in \mathbb{N}$, we have*

$$\max_{i,j,m=1,\ldots,5} \sup_{\theta\in\Theta} \left|\partial^3 L_n(\theta)/\partial\theta_i\partial\theta_j\partial\theta_m\right| \leq c_n,$$

where $0 \leq c_n \xrightarrow{P} c$ as $n \to \infty$ and $0 < c < \infty$.

Proof. For the third derivative with respect to β, from (B.2), we have

$$\begin{aligned}
\left|\partial^3 L_n(\theta)/\partial\beta^3\right| &= \frac{1}{n}\left|\sum_{t=1}^{n} \frac{\partial^3 l_t(\theta)}{\partial\beta^3}\right| \\
&\leq \frac{1}{n}\sum_{t=1}^{n}(\nu+1)b_t(\theta)(1-b_t(\theta))\left[\left|h_{\beta t}(\theta)^3\right| + \frac{3}{2}\left|h_{\beta t}(\theta)h_{\beta\beta t}(\theta)\right|\right] \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}(\nu+1)\left|h_{\beta t}(\theta)^3\right| b_t(\theta)(1-b_t(\theta))^2 \\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\left(3\left|h_{\beta t}(\theta)h_{\beta\beta t}(\theta)\right| + 2\left|h_{\beta t}(\theta)^3\right| + \left|h_{\beta\beta\beta t}(\theta)\right|\right)\left[(\nu+1)b_t(\theta) - 1\right] \\
&\leq \frac{\nu_u+1}{n}\sum_{t=1}^{n}\left(2u^*_{\beta t}(\theta_0)^3 + \frac{3}{2}u^*_{\beta t}(\theta_0)u^*_{\beta\beta t}(\theta_0)\right) \\
&\quad + \frac{\nu_u+2}{2n}\sum_{t=1}^{n}\left(3u^*_{\beta t}(\theta_0)u^*_{\beta\beta t}(\theta_0) + 2u^*_{\beta t}(\theta_0)^3 + u^*_{\beta\beta\beta t}(\theta_0)\right) \\
&\xrightarrow{P} (\nu_u+1)\left(2\mathbb{E}\left[u^*_{\beta 1}(\theta_0)^3\right] + \frac{3}{2}\mathbb{E}\left[u^*_{\beta 1}(\theta_0)u^*_{\beta\beta 1}(\theta_0)\right]\right) \\
&\quad + \frac{\nu_u+2}{2}\left(3\mathbb{E}\left[u^*_{\beta 1}(\theta_0)u^*_{\beta\beta 1}(\theta_0)\right] + 2\mathbb{E}\left[u^*_{\beta 1}(\theta_0)^3\right] + \mathbb{E}\left[u^*_{\beta\beta\beta 1}(\theta_0)\right]\right) \\
&\in (0,\infty)
\end{aligned}$$

as $n \to \infty$ by Lemma 15(i)-(iv) and Theorem 3.5.7 of Stout (1974). Straightforward differentiation shows that the desired inequality holds for other third derivatives by Lemma 15(i)-(iv) and Lemma 1. ■