Asymptotic properties of the QML estimator in a smooth transition GARCH(1,1) model

by

Kerem Tuzcuoğlu

A Thesis Submitted to the
Graduate School of Social Sciences
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts
in
Economics

Koc University

July 2010

Acknowledgements

I would like to express my deepest respect and most sincere gratitude to my supervisor Mika Meitz for his expertise, patience, kindness, support and encouragement, both in the fulfillment of this thesis and in developing my academic interests generally. He met with me for countless hours for the discussion of a subject, and even more hours for commenting on the drafts that lead carefully towards the creation of the final level of this thesis. Finally, I am heartily thankful to Mr. Meitz for his continuous guidance in my PhD application periods.

Besides my supervisor, I would like to thank Kamil Yılmaz for helping me in deciding the university for my further studies, and for being a member in my thesis committee. My sincere thanks also go to Atakan Yalçın for sparing his valuable time to my thesis defense.

I extend my gratitude to TÜBİTAK (The Scientific and Technological Research Council of Turkey) for providing me the scholarship.

My special thanks go to Sinem for her support, for easing my life during the difficult times, and especially for all the fun we have had in the last two years.

Lastly, it is an honor for me to have such wonderful parents, who have always believed in and encouraged me in every field of life and tried to help whenever I needed.

Abstract

This thesis examines asymptotic properties of the quasi maximum likelihood (QML) estimator in a specific nonlinear generalized autoregressive conditionally heteroskedastic (GARCH) process. The conditional mean is set to zero. The nonlinearity is established via smooth transition mechanism in the conditional variance, where the distribution function of a logistic distribution is used for the smooth transition function. Strong consistency and asymptotic normality of the QML estimator is proved in this smooth transition GARCH$(1,1)$ model. For most of the analysis, we follow the work done in Meitz and Saikkonen (2008c), where asymptotic properties of the QML estimator are studied in nonlinear AR-GARCH models.

Key Words: Smooth transition, GARCH, nonlinear financial econometrics, strong consistency, asymptotic normality, quasi maximum likelihood estimation.

Özet

Bu tez, doğrusal olmayan belirgin bir genelleştirilmiş kendiyle bağlaşımlı koşullu farklı yayılım sürecinde sözde ençok olabilirlik tahmincisinin yanaşık özelliklerini incelemektedir. Koşullu ortalama sıfır olarak alınmıştır. Doğrusal dışılık, koşullu değişirlikteki yumuşak geçiş mekanizmasıyla sağlandı. Yumuşak geçiş işlevi için ise lojistik dağılımın dağılım işlevi kullanıldı. Ençok olabilirlik tahmincisinin güçlü tutarlılığı ve yanaşık normalliği, bahsi geçen yumuşak geçişli genelleştirilmiş kendiyle bağlaşımlı koşullu farklı yayılım modelinde ispatlandı. Tezdeki çözümlemelerin çoğunda Meitz ve Saikkonen (2008c) makalesindekilere benzer çözümlemeler kullanıldı. Adı geçen makalede, sözde ençok olabilirlik tahmincisinin yanaşık özellikleri, koşullu ortalamasında doğrusal olmayan kendiyle bağlaşımlı, koşullu değişirliğinde ise doğrusal olmayan genelleştirilmiş kendiyle bağlaşımlı koşullu farklı yayılım süreci olan modeller içinde incelenmiştir.

Anahtar Kelimeler: Yumuşak geçiş, genelleştirilmiş kendiyle bağlaşımlı koşullu farklı yayılım, doğrusal olmayan finansal ekonometri, güçlü tutarlılık, yanaşık normallik, sözde ençok olabilirlik tahmini

To my parents

for their endless love and support in my life...

Contents

1 Introduction and Literature Review **1**

1.1 Linear Generalizations of ARCH/GARCH Models 3

1.2 Nonlinear Generalizations 4

1.3 Families of GARCH Models 5

1.4 Theoretical Results 5

1.5 The Topic of this Thesis 7

2 The Model **9**

2.1 Assumptions and Results for Strong Consistency 10

3 Strong Consistency of the QML estimator **14**

4 Asymptotic Normality of the QML estimator **16**

4.1 Assumptions 16

4.2 Results and the Main Theorem for Asymptotic Normality 17

5 Conclusion **21**

Appendices **22**

Appendix A: Properties of the Function G 22

Appendix B: Auxiliary Lemmas 26

Appendix C: Derivation of Results 1-5 27

Appendix D: Derivation of Results 6-10 30

Appendix E: Proofs of the Main Theorems 38

The Proof of Theorem 1 38

The Proof of Theorem 2 46

References **58**

1 Introduction and Literature Review

Risk and uncertainty in financial markets are of interest to everybody from individual investors to hedge fund managers. They are measured in various ways, but the main measure of risk and uncertainty is volatility, which is an essential character of financial markets. Volatility is present not only in stock market asset prices, but also in fixed income security returns, such as treasury bonds, due to changes in interest rates. One important feature of volatility is it being time varying. If we take a look at the volatility of returns of financial assets, then we get different volatilities at different points in time. Because volatility can be inferred through variances, modelling the changing variance, i.e. heteroskedasticity, is at the heart of financial analyses. Models used for analyzing the time varying conditional variance have been used since Engle first introduced the autoregressive conditionally heteroskedastic (ARCH) processes in 1982. In the past three decades, Engle's model became very popular and has been used in modelling financial time series.

Suppose one is interested in modelling the univariate time series y_t, which is often assumed to consist of the returns of a financial asset. The ARCH process allows the conditional variance of y_t to change over time as a function of past y_t's although the unconditional variance of y_t is constant. The ARCH process proposed by Engle (1982) is defined as

$$y_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2),$$

where $\mathcal{F}_{t-1}$ is the information set available at time $t-1$, and the conditional variance is an explicit function of lagged y_t's

$$\sigma_t^2 = f(y_{t-1},\ y_{t-2}, \dots,\ y_{t-p}; \theta).$$

The order of the ARCH process is p, and θ is a vector of unknown parameters. In the simplest and typical case, the conditional variance is described as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2,$$

where $\omega > 0$ and $\alpha_i \geq 0$ for $i = 1, 2, \dots, p$. Additionally, the log-likelihood function for the ARCH model is found to be

$$\log L(y_T,\ y_{T-1}, \ldots,\ y_1; \theta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\log(\sigma_t^2) + \frac{y_t^2}{\sigma_t^2}\right].$$

Engle (1982) proved the relative efficiency of the Maximum Likelihood Estimation (MLE) to Ordinary Least Squares (OLS) in a special case. Furthermore, Pantula (1988) showed a general result that the MLE is more efficient than OLS in ARCH models. For these reasons, MLE is used as the estimation technique in ARCH-type models.

ARCH models have been used in financial economics, especially in modelling foreign exchange rates, stock returns, the changing uncertainty of inflation rates and the term structure of interest rates. One of the main features of asset returns is volatility clustering. Volatility clustering is the phenomenon that large (small) values of returns are followed by large (small) values of either sign. Another feature is that the asset returns have a leptokurtic distribution, i.e. it has fat tails. Fat-tailedness brings about higher probabilities for extreme values compared to a normal distribution. The generalized ARCH (GARCH) model proposed by Bollerslev (1986) was able to successfully capture the volatility clustering and fat-tailedness of asset returns. The GARCH model is found by including lagged conditional variances in the ARCH conditional variance equation. The GARCH(p, q) model's conditional variance is

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}\alpha_i y_{t-i}^2 + \sum_{j=1}^{p}\beta_j \sigma_{t-j}^2,$$

where $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, 2, \ldots, q$, and $\beta_j \geq 0$ for $j = 1, 2, \ldots, p$. The GARCH model allows the conditional variance to depend on both its lags and on the lags of past squared returns. Hence, some kind of learning mechanism is involved in the general process. Nevertheless, there are some drawbacks of simple GARCH models. For instance, the model is restricted by nonnegativity constraints, and the conditional variance is affected only by the magnitude of the shocks, not by their signs. In order not to have a negative σ_t^2, the parameters in the conditional variance equation are set to be nonnegative. Furthermore, because we have lags of squared returns, it does not matter whether the innovations are positive or negative. However, Black (1976) suggested that while "bad news" are increasing volatility, "good news" are reducing it. Thus, the model fails to capture the asymmetric responses of σ_t^2. Therefore, GARCH model has been generalized and extended in various directions in order to overcome these problems.

1.1 Linear Generalizations of ARCH/GARCH Models

The main aim of the generalizations was to increase the flexibility of the original model and have weaker assumptions. The first generalization made was by Engle and Bollerslev in 1986. They introduced the Integrated GARCH (IGARCH) where the coefficients, except the intercept, of the conditional variance sum up to one. For example, the GARCH model defined above will be IGARCH model if we restrict $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1$. In IGARCH, the forecasts of the variance for further periods do not approach the unconditional variance. In the model, today's information is important for forecasting the conditional variance in every period. Hence, the conditional variance will be affected permanently by the current shocks. IGARCH was proven not to be covariance stationary but still strictly stationary by Nelson (1990). One of the other important and popular generalizations is exponential GARCH (EGARCH) introduced by Nelson (1991). EGARCH$(1,1)$ is defined in the following way.

$$y_t = \varepsilon_t \sigma_t \quad \text{and} \quad \log \sigma_t^2 = \omega + \gamma y_{t-1} + \alpha \left[|y_{t-1}| - E(|y_{t-1}|) \right] + \beta \log \sigma_{t-1}^2,$$

where ε_t is i.i.d. with zero mean and unit variance. EGARCH solves the two drawbacks. First, it allows asymmetries and second, both the sign and the magnitude of the shocks affect the conditional variance in EGARCH model. For "good news", i.e. where $0 < \varepsilon_t < \infty$, $\log \sigma_t^2$ is linear in y_{t-1} with a slope of $\gamma + \alpha$. However, for "bad news", where $-\infty < \varepsilon_t < 0$, the slope becomes $\gamma - \alpha$. Thus, the conditional variance responses asymmetrically to different news. Furthermore, the sign and magnitude effects are captured by the term $\alpha \left[|y_{t-1}| - E(|y_{t-1}|) \right]$.

Another important generalization is GJR-GARCH which is proposed by Glosten et al. (1993). The model allows for asymmetry as it was the case in EGARCH. GJR-GARCH$(1,1)$ is defined as

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \gamma y_{t-1}^2 I(y_{t-1} < 0) + \beta \sigma_{t-1}^2,$$

where I is the indicator function. When the data is collected from asset returns, then γ generally turns out to be positive which implies that negative news affect the conditional variance more than the positive news. A similar structure to GJR-GARCH model is introduced by Zakoïan (1994), the Threshold GARCH (TGARCH). But instead of the conditional variance, the conditional standard deviation is modelled in TGARCH. Another linear generalization is Sentana (1995)'s quadratic GARCH (QGARCH) model. It includes not only lagged returns but also cross-products of lagged returns which provide asymmetric responses. The augmented ARCH (AARCH) introduced by Bera, Higgins and Lee (1992) is a slightly more general model than QGARCH. Baillie et al. (1996) proposed fractionally integrated GARCH (FIGARCH). In the FIGARCH model, the response of the conditional variance to lagged residuals can decay slowly, contrary to GARCH and IGARCH where there is either exponentially decay or permanence in effect of shocks. Hence, the model captures slowly changing volatility where effects of shocks decay in a long time. These are some characteristics of high-frequency data captured by FIGARCH.

1.2 Nonlinear Generalizations

There are also nonlinear extensions of ARCH and GARCH models. Higgins and Bera (1992) introduced the nonlinear ARCH (NARCH) where the conditional standard deviation raised to the power δ is modelled. Another nonlinear model, that aims capturing asymmetry, is the nonlinear asymmetric GARCH (NAGARCH) of Engle and Ng (1993) where the response to shocks is centralized at a constant different from zero. A nonlinear threshold model Double Threshold ARCH (DTARCH) model is presented by Li and Li (1996). The name comes from the two threshold structure both in the autoregressive conditional mean and in the conditional variance. Then, Hagerud (1996), González-Rivera (1998) and Anderson et al. (1999) proposed nonlinear GARCH models by making smooth transitions between regimes. In this paper, we are going to analyze a similar model to that of González-Rivera (1998) defined as the following.

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \delta y_{t-1}^2 F(y_{t-1};\gamma) + \beta\sigma_{t-1}^2,$$

where $F(y_{t-1};\gamma) = (1+\exp(-\gamma y_{t-1}))^{-1}$. The choice of the transition function is important. For instance, if we put an indicator function as a transition function, then we get GJR-GARCH model. Thus, jumps will occur at the regime changing points. However, a choice of a continuous function makes the transition smoother, allowing intermediate regimes to exist. Another nonlinear model that aims to explain volatility processes with different regimes is Markov Switching GARCH (MS-GARCH) model. Authors such as Gray (1996), Klaassen (2002), and Haas et al. (2004) proposed this kind of model. For instance, Haas et al. (2004)'s model has a structure where the conditional variance equation is different for each regime.

1.3 Families of GARCH Models

In addition to all these models mentioned above, some authors defined families of GARCH processes. These families nest most of the GARCH models. Ding et al. (1993), Hentschel (1995), Duan (1997), and He and Teräsvirta (1999) are some examples who proposed families of GARCH models. For instance, the HGARCH model of Hentschel (1995) is defined as

$$\frac{\sigma_t^\lambda - 1}{\lambda} = \omega + \alpha\sigma_{t-1}^\lambda \left[|y_{t-1} - b| - c(y_{t-1} - b)\right]^v + \beta\frac{\sigma_{t-1}^\lambda - 1}{\lambda}.$$

If we put $(v, b) = (1, 0)$ and take the limit as $\lambda \to 0$ then we get EGARCH. If $(\lambda, v, b) = (2, 2, 0)$ then we end up with GJR-GARCH. TGARCH and NGARCH are among other important GARCH models that are nested in HGARCH model. Of course, the literature on the extensions of ARCH and GARCH models is vast. We tried to present the most popular and used versions. For a more general list, see Bollerslev (2008).

1.4 Theoretical Results

Because ARCH/GARCH models were capable of capturing some important features of financial data, they have been used widely in finance research and applications. Despite the wide use of these models, sampling properties of the estimators were studied scarcely in the early years after introduction of ARCH/GARCH processes. Milhøj (1985) studied conditions for the existence of moments. The first study of the asymptotic properties of the ARCH model is done by Weiss (1986). He proved that the MLE is consistent and asymptotically normal under fourth order moment condition on the ARCH process. However, this assumption is a strong assumption. Nelson (1990) provided necessary and sufficient conditions for stationarity and ergodicity of GARCH(1,1) process.

The first theory for the asymptotic properties of GARCH models is covered by Lumsdaine (1991), her doctoral dissertation. In 1992, Bollerslev and Wooldridge used strong assumptions and derived the large sample properties of Quasi-Maximum Likelihood Estimator (QML estimator). Nevertheless, they did not confirm whether these assumptions apply for GARCH models. Bougerol and Picard (1992a, b) provided necessary and sufficient conditions for the existence of a unique stationary solution of the conditional variance equation of GARCH(p, q) models. Lee and Hansen (1994) contributed to the literature by showing the consistency and asymptotic normality of QML estimator in GARCH(1,1) with weaker assumptions than those of Weiss (1986) and Lumsdaine (1991). Lee and Hansen did not make the assumption of the "i.i.d."ness of the rescaled variable - the ratio of the returns to the conditional standard deviation. However, they exclude integrated GARCH processes for consistency proof. Lumsdaine (1996) showed that even without the assumption of finite fourth moments, the QML Estimators of all the parameters in GARCH(1,1) and IGARCH(1,1) are consistent and jointly asymptotically normal. Nevertheless, both Lee and Hansen (1994) and Lumsdaine (1996)'s work were unique so that they cannot be easily generalized to GARCH (p, q) model. Ling and Li (1998) showed that MLE of unstable ARMA with GARCH errors is consistent, yet the results were found for local estimators.

Eventually, a more general result without very strong assumptions is found by Berkes, Horváth and Kokoszka (2003). They proved the consistency and asymptotic normality of QML estimator in GARCH(p, q) models in milder conditions. Moreover, they maximized the likelihood function over a general compact set, contrary to Lee and Hansen (1994) and Lumsdaine (1996) where only the local consistency is obtained. Francq and Zakoïan (2004) proved strong consistency and asymptotic normality of QML estimator under weaker conditions than those in the existing literature, and they extended their results to ARMA-GARCH models. Jensen and Rahbek (2004b) established similar asymptotic results for QML estimator in GARCH(1,1) model even when the parameters are allowed to be in the region where the process is non-stationary. Straumann and Mikosch (2006) extended the results of Berkes, Horváth and Kokoszka (2003) to a more general conditionally heteroskedastic time series model. They studied asymptotic properties of QML estimator in nonlinear pure GARCH models, including GARCH(p, q). As examples, AGARCH and EGARCH models are examined. As an extended result, Meitz and Saikkonen (2008c) established the consistency and asymptotic normality of the QML estimator in nonlinear AR-GARCH models. In their model, they allow nonlinearity in both the conditional mean and in the conditional variance. They also provide some examples where their results can be applied.

1.5 The Topic of this Thesis

The aim of this thesis is to establish strong consistency and asymptotic normality of the QML estimator in a smooth transition GARCH(1,1) model. As mentioned above, smooth transition models are presented by Hagerud (1996) and González-Rivera (1998). For the transition function, generally, either the logistic or exponential function is used. We work on the smooth transition model with logistic function because in this case both the sign and the magnitude of the innovations affect the conditional variance. But, when exponential transition is used then the model cannot capture the effects of different signs of innovations. Moreover, in a nonlinear autoregressive model, logistic function is proven to be superior to exponential function by Teräsvirta (1994).

The smooth transition model with logistic function has a similar structure to GJR-GARCH model where indicator function is used to distinguish between the regimes. Thus, there are breaks between two different regimes in GJR model. However, via smooth transition functions we allow intermediary regimes to exist as the process is passing from one regime to another. The parameters in the logistic function will determine the shape of the transition. There can be a sharp transition so that there are low and high volatility regimes or there can be a smoother transition where there are many intermediate regimes depending on the parameters of the logistic function.

This thesis can be seen as a special case of Meitz and Saikkonen (2008c) where asymptotic theory for nonlinear AR-GARCH models is analyzed. We will basically follow their arguments, therefore for more general results see Meitz and Saikkonen (2008c). Moreover, some assumptions, derivations and techniques used in this paper are similar to those of Straumann and Mikosch (2006), and Francq and Zakoïan (2004). Therefore, throughout the paper we will refer to them often.

The rest of the thesis proceeds as follows. Section 2 introduces the model to be studied. The assumptions and particular results related to strong consistency of the QML estimator are included in this section. Section 3 contains some results, an extra assumption and the strong consistency theorem. Section 4 is devoted to the asymptotic normality of QML estimator. Concluding remarks are presented in Section 5. All of the proofs and useful lemmas are left to the Appendix.

A final note on the notation used throughout this thesis. We denote $(0,\infty)$ as $\mathbb{R}_+$. The transpose of a vector or a matrix v is denoted as v'. For any scalar, vector, or matrix v, the Euclidean norm[1] is denoted by $|v|$. The L_p-norm of v is denoted by $\|v\|_p = \{E[|v|^p]\}^{1/p}$ for a random variable (scalar, vector, or matrix) v and for $p > 0$. A sequence of random elements v_t is said to converge in L_p-norm to v if $\|v_t\|_p < \infty$ for all t, $\|v\|_p < \infty$, and $\lim_{t\to\infty} \|v_t - v\|_p = 0$. The sequence of random elements $v_t(\lambda)$ is said to be L_p-dominated in Λ if there exists a sequence of positive random variables D_t such that $|v_t(\lambda)| \leq D_t$ for all $\lambda \in \Lambda$ and $\|D_t\|_p < \infty$ uniformly in t. Finally, 'a.s.' stands for 'almost surely' and 'iff' stands for 'if and only if'.

[1] Euclidean norm for a real valued vector v is defined as $|v| = (v'v)^{1/2}$. For a real valued $(m \times n)$ matrix $A = [a_{ij}]$, the norm is defined as

$$|A| = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}.$$

2 The Model

We will focus on a specific type of smooth transition GARCH, ST-GARCH$(1,1)$, model defined by, for $t = 1, 2, \ldots$,

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \textit{and } \sigma_t^2 &= g(y_{t-1}, \sigma_{t-1}^2; \lambda_0) = \omega_0 + (\alpha_{0,1} + \alpha_{0,2} G(y_{t-1}; \gamma_0)) y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{aligned} \quad (1)$$

Note that if one would set $\alpha_{0,2} = 0$, the model would reduce to the standard GARCH(1,1) model of Bollerslev(1986). In our model, the errors are assumed to be independent and identically distributed random variables with zero mean and unit variance, i.e., $\varepsilon_t \overset{iid}{\sim} (0, 1)$. Moreover, ε_t is assumed to be independent of y_s for $s < t$. In particular, ε_t is not only independent of its own past, but also independent of the past of σ_t^2. Thus, σ_t^2 determines the unobserved conditional variance of the observed process y_t, where 'conditional' refers to conditional on the past of the observed series, $\{y_{t-1}, y_{t-2}, \ldots\}$. In this paper the conditional mean is simply set to zero. We assume that the (6×1) true parameter vector $\lambda_0 = (\omega_0, \alpha_{0,1}, \alpha_{0,2}, \beta_0, \gamma_0)$, where $\gamma_0 = (\gamma_{0,1}, \gamma_{0,2})$, belongs to a compact set Λ. This permissible parameter space is a subset of $\mathbb{R}_+ \times [0, \infty)^2 \times [0, 1) \times \mathbb{R} \times \mathbb{R}_+$. Having a compact parameter space is a common assumption in nonlinear estimation problems. We can relax this assumption at a cost of some additional assumptions and a more complicated analysis.

The function G is the distribution function of a logistic distribution defined by

$$G(y_{t-1}; \gamma) = \frac{1}{1 + e^{-\gamma_2 (y_{t-1} - \gamma_1)}}.$$

Although in economic applications, usually, the nonlinear function $G(y; \gamma)$ is chosen to be the cumulative distribution function of the logistic distribution, there are also other choices for the function G. For instance, Lanne and Saikkonen (2005) used an increasing function similar to the cumulative distribution of a positive continuous random variable as the transition function where it depends on the lagged conditional variance instead of the lagged series. Another choice may be exponential distribution function.

Note that the nonlinear function G takes values in $[0, 1]$ and depends only on the first lag of y_t. The location parameter γ_1 takes values in $\mathbb{R}$, whereas γ_2 takes values in

$\mathbb{R}_+$ being a scale parameter. Consider the case $\gamma_2 = 0$. The transition function becomes a constant function and thus $\alpha_{0,1}$ and $\alpha_{0,2}$ cannot be identified. Therefore, the permissible parameter space of γ_2 is positive numbers. Essentially, γ_2 determines the smoothness of the transition. Hence, the logistic smooth transition GARCH model nests several important GARCH models, such as GJR-GARCH introduced by Glosten et al. (1993). For large values of γ_2 the transition function will become steeper and behave like an indicator function at $y = \gamma_1$. Thus, additionally, if we set $\gamma_1 = 0$ then the model will be equivalent to GJR-GARCH. These different plots of the function G are illustrated in Figure 1 for $\gamma_1 = 0$.

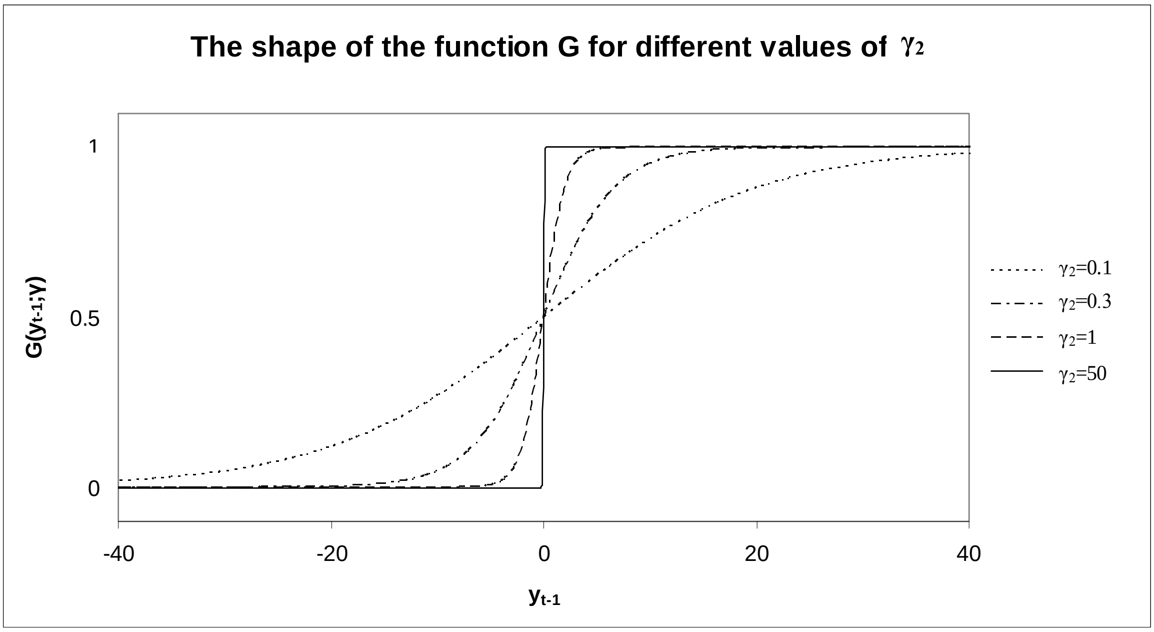

Figure 1

Other properties related to the function G, such as continuity and differentiability, will be investigated in Appendix A.

2.1 Assumptions and Results for Strong Consistency

Now, we will make some assumptions from which we obtain particular results to prove the consistency of the estimator.

Assumption 1:

a) ε_t *has (Lebesgue) density which is positive and lower-semi continuous on* $\mathbb{R}$.

b) $E\left[\left(\beta_0 + (\alpha_{0,1} + \alpha_{0,2})\,\varepsilon_t^2\right)^r\right] < 1$ *for some* $r > 0$.

c) $\alpha_{0,1} > 0$ *and* $\beta_0 > 0$.

This assumption is necessary not only for satisfying identification condition but also for using results of Meitz and Saikkonen (2008a) for ergodicity and stationarity of (y_t, σ_t^2). Assumption 1 is needed to satisfy the Proposition 1 and Theorem 1 of that paper. The number r is going to be used in further results, thus, throughout the paper we fix this r. The last part of the assumption ensures the existence of the ARCH and GARCH effects, thus the identifiability of the parameters in the conditional variance. Moreover, the positiveness of these parameters is inevitable for the asymptotic normality of the QML estimator. Under the assumption of finite 6^{th} moments of the innovations, Francq and Zakoïan (2007) shows that the asymptotic distribution is non-Gaussian if we let some parameters to be equal to zero.

Assumption 2: $\alpha_{0,2} > 0$.

This assumption is important for identifying the parameters. If $\alpha_{0,2} = 0$, then the parameters in γ cannot be identified, thus the model becomes linear GARCH(1,1) model of Bollerslev (1986).

Now let's discuss the results that are derived from assumptions and properties of the function G. They are going to be used for proving consistency of the QML estimator. The first result is obtained from Assumption 1. The proofs of this and all the subsequent results are in the Appendices.

Result 1: *Suppose Assumption 1 holds. Then the process* (y_t, σ_t^2) *defined above is ergodic and stationary with* $E[\sigma_t^{2r}] < \infty$.

Stationarity and ergodicity play a crucial role in the proofs of consistency and asymptotic normality. They enable us to use well-known results, such as Ergodic Theorem. Result 1 additionally provides us the finiteness of $E[|y_t|^{2r}]$. Similar to our Result 1, the ergodic stationarity assumption is made in many papers, for instance see Assumption

C.1 of Straumann and Mikosch (2006). Although Result 1 is sufficient in the proof of consistency of the estimator, we need stronger moment conditions for proving asymptotic normality.

Assume $y_0,\ y_1, \ldots,\ y_T$ are observed, which are generated by ergodic and stationary process defined by equation (1)(cf. Result 1). If ε_t were assumed to be Gaussian, then, conditionally on the $\sigma-$algebra generated by $\{\varepsilon_i, -\infty < i \leq t-1\}$, the ratio y_t/σ_t would also be Gaussian. However, we do not assume the normality of ε_t, but the likelihood is constructed as if ε_t would be normal. Therefore, the estimator considered in the paper is QML estimator. Hence, the conditional quasi-maximum log-likelihood function is defined as

$$\mathcal{L}(\lambda) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log \sigma_t^2 + \frac{y_t^2}{\sigma_t^2}\right). \tag{2}$$

Because we cannot observe the process of the conditional variance we do not know anything about its stationary distribution. But the log-likelihood function has the conditional variance as a variable. Therefore, we define a process to approximate the conditional variance in the following way.

$$h_t(\lambda) = \begin{cases} \varsigma_0 & t = 0 \\ g(y_{t-1}, h_{t-1}(\lambda); \lambda) & t = 1, 2..., \end{cases}$$

where λ is an (6×1) parameter vector with the true value λ_0. We assume that the initial value ς_0 is positive and independent of λ. After the specification of ς_0, we can solve $h_t(\lambda)$ recursively for any given λ. A similar approach is taken by several authors including Lee and Hansen (1994), Lumsdaine (1996), Francq and Zakoïan (2004), Straumann and Mikosch (2006). In the derivations, we will use the notation h_t for $h_t(\lambda)$ unless the explicit dependence to λ is needed. Meitz and Saikkonen (2008b) provided sufficient conditions for ergodicity and stationarity only for $h_t(\lambda_0)$. We do not know anything about the process $h_t(\lambda)$ for $\lambda \neq \lambda_0$. Thus, we additionally have to use Result 2, which is a direct consequence of properties of the function G, in order to get Result 3 where we obtain desirable properties of $h_t(\lambda)$.

Result 2: *The function* $g : \mathbb{R} \times \mathbb{R}_+ \times \Lambda \rightarrow \mathbb{R}_+$ *is continuous with respect to all of its arguments and satisfies the following.*

i) *For some* $0 < \varrho < 1$ *and* $0 < \varkappa, \varpi < \infty$, $g(y, x; \lambda) \leq \varrho x + \varkappa y^2 + \varpi$ *for all* $\lambda \in \Lambda, y \in \mathbb{R}$ *and* $x \in \mathbb{R}_+$.
ii) *For some* $0 < \kappa < 1$, $|g(y, x_1; \lambda) - g(y, x_2; \lambda)| \leq \kappa |x_1 - x_2|$ *for all* $\lambda \in \Lambda, y \in \mathbb{R}$ *and* $x_1, x_2 \in \mathbb{R}_+$.

In the property *i)*, we do not only bound the function g from above but also we specify the upper bound. This specific bound is used while achieving Result 3 where we introduce ergodic and stationary solution to the approximation of the conditional variance. The contraction property of g with respect to its second argument is given in *ii)*. Here, the constant κ is fixed throughout the paper. An assumption similar to this result is made by Straumann and Mikosch (2006), see Proposition 3.12. They also bound the function g from above and assume a contraction property.

Result 3: *Suppose Assumption 1 holds. Then, for all* $\lambda \in \Lambda$, *there exists a stationary and ergodic solution* $h_t^*(\lambda)$ *to the equation*

$$h_t(\lambda) = g(y_{t-1}, h_{t-1}(\lambda); \lambda) = \omega + (\alpha_1 + \alpha_2 G(y_{t-1}; \gamma))y_{t-1}^2 + \beta h_{t-1}^2(\lambda), \quad t = 1, 2, ... \quad (3)$$

Moreover, $h_t^*(\lambda)$ *is continuous in* λ *and is measurable with respect to the* $\sigma-$*algebra generated by* $(y_{t-1}, y_{t-2}, ...)$. *The solution to the equation is unique when it is extended to all* $t \in \mathbb{Z}$. *The solution evaluated at the true parameter is equal to the conditional variance, i.e.,* $h_t^*(\lambda_0) = \sigma_t^2$. *Furthermore,* $E\left[\sup_{\lambda \in \Lambda} h_t^{*r}(\lambda)\right] < \infty$ *and* $\left\|\sup_{\lambda \in \Lambda} |h_t^* - h_t|\right\|_r \leq C\kappa^t$, *where* $C \in \mathbb{R}$ *and* κ *is as in Result 2. Finally, for* $\lambda \in \Lambda$, *if* $h_t(\lambda)$ *is any other solution to the above equation, then for some* $\gamma > 1$,

$$\gamma^t \sup_{\lambda \in \Lambda} |h_t^*(\lambda) - h_t(\lambda)| \to 0 \quad in\ L_r - norm\ as\ t \to \infty.$$

Result 3 is essential in proving consistency of QML estimator. Since the difference between $h_t^*(\lambda)$ and $h_t(\lambda)$ is negligible, i.e., as $t \to \infty$ the difference converges uniformly to zero sufficiently fast, we can use $h_t^*(\lambda)$ in the derivations. Actually, the difference between $h_t^*(\lambda)$ and any other solution to the recurrence equation is negligible. The advantage of $h_t^*(\lambda)$ is its ergodicity and stationarity which is central in inferring the stationarity of the likelihood function and in using the Ergodic Theorem.

3 Strong Consistency of the QML estimator

As mentioned in the previous section, we will use quasi-maximum likelihood estimation in our analysis since ε_t is not assumed to be normal. We define QML estimator of λ_0 to be

$$\hat{\lambda}_T = \arg\min_{\lambda\in\Lambda} L_T(\lambda) = \arg\min_{\lambda\in\Lambda}\left\{\frac{1}{T}\sum_{t=1}^{T} l_t(\lambda)\right\}, \text{ where } l_t(\lambda) = \log(h_t(\lambda)) + \frac{y_t^2}{h_t(\lambda)}.$$

The objective function $L_T(\lambda)$ is obviously a linear transformation of the conditional log-likelihood function defined in (2), as the constant is ignored and the function is rescaled by $-2/T$. The ergodic and stationary counterpart of $L_T(\lambda)$ is defined as $L_T^*(\lambda) = T^{-1}\sum_{t=1}^{T} l_t^*(\lambda)$, where $l_t^*(\lambda) = \log(h_t^*) + y_t^2/h_t^*$ and $h_t^* = h_t^*(\lambda)$ is the stationary and ergodic solution to (3). In addition to above results, we need further results about the conditional variance function.

Result 4: *The function* $g : \mathbb{R}\times\mathbb{R}_+ \times \Lambda \to \mathbb{R}_+$ *is bounded away from* 0 *in the sense that* $\inf_{(y,x,\lambda)\in\mathbb{R}\times\mathbb{R}_+\times\Lambda} g(y,x;\lambda) = \underline{g}$ *for some* $\underline{g} > 0$.

Result 5: *Suppose Assumptions 1 and 2 hold. Then* $h_t^*(\lambda) = \sigma_t^2$ *a.s. only if* $\lambda = \lambda_0$.

Result 4 is an immediate outcome of model's definition. However, it is important for bounding the likelihood function from below uniformly. Moreover, since $h_t(\lambda)$ is constructed by the equation $h_t(\lambda) = g(y_{t-1}, h_{t-1}(\lambda); \lambda)$, we obtain that $h_t(\lambda) \geq \underline{g}$. Hence, having the definition of being the solution of this recursive relation, $h_t^*(\lambda)$ has to satisfy the properties of the relation as well, i.e., $h_t^*(\lambda) \geq \underline{g}$. These two inequalities provide us the well-definedness of y_t^2/h_t^* and y_t^2/h_t which are components of the likelihood function. A similar uniformly bounding assumption is made by Straumann and Mikosch (2006). In order to obtain Result 5, which is an identification condition, we have to use properties of the function G and other results. This condition is used in showing that the true parameter is the unique minimizer of the expected value of the likelihood function. This identifiability condition is assumed in the same way in Assumption C.4 of Straumann and Mikosch (2006). In the derivations of Francq and Zakoïan (2004), it is shown that this condition holds for the linear ARMA-GARCH models.

Assumption 3: $E\left[\sup_{\lambda\in\Lambda}\sigma_t^2/h_t^*\right] < \infty$.

Assumption 3 is an extra assumption. For proving strong consistency we actually do not need it. However, if it is assumed then the proof simplifies. The consistency proof with Assumption 3 as an extra assumption is done in another part (*Part 1*) of the proof. Similarly, Straumann and Mikosch (2006) provided 2 different proofs for strong consistency, one proof with extra assumption similar to Assumption 3 and one without it.

Now we are ready to state our strong consistency theorem whose proof is in Appendix E.

Theorem 1 *Suppose Assumptions 1 and 2 hold. Then, the QML estimator $\hat{\lambda}_T$ is strongly consistent for λ_0, i.e., $\hat{\lambda}_T \to \lambda_0$ a.s.*

In the proof, we will use the Gaussian likelihood function $L_T(\lambda)$ and the ergodic and stationary approximation $L_T^*(\lambda)$. Instead of minimizing directly the likelihood function, we will minimize its ergodic and stationary analogue $L_T^*(\lambda)$. Then, we make use of the negligible difference between $L_T(\lambda)$ and $L_T^*(\lambda)$ in large samples. Both Francq and Zakoïan (2004), and Straumann and Mikosch (2006) have a similar approach. In our proof, we basically follow the arguments of Pötscher and Prucha (1991). We introduce an extra part, named as *Part 1*, where we use a uniform Strong Law of Large Numbers (SLLN). Here, we finished the consistency proof by imposing the extra assumption, Assumption 3. Thereby, in this part it is shown how the proof simplifies when uniform SLLN can be used. In *Part 2*, we continued the arguments of Pötscher and Prucha (1991) by considering the Ergodic Theorem and the argument of Pfanzagl (1969) instead of using a uniform SLLN. The latter argument is used also in other papers to prove the strong consistency (see Jeantheau, 1998).

4 Asymptotic Normality of the QML estimator

4.1 Assumptions

In order to verify the asymptotic normality of the QML estimator of the true value λ_0, we need to analyze the limiting properties of the first and second derivatives of the objective likelihood function $L_T(\lambda)$. In particular, we will focus on the derivatives of h_t and h_t^*. In this subsection, we state the necessary assumptions for asymptotic normality. They are essential in proving the further results in the next subsection. These assumptions, along with the previous ones, ensure twice continuously differentiability of the processes h_t and h_t^*.

Assumption 4: *The true parameter value* λ_0 *is an interior point of* Λ, i.e., $\lambda_0 \in \mathring{\Lambda}$.

This assumption is necessary for differentiation of the conditional variance. Having the true parameter values as an interior point, we can use Taylor series expansion of the score vector around λ_0. Moreover, because $\hat{\lambda}_T$ is consistent by Theorem 1 we only need to consider parameter values in an open neighborhood of λ_0. Therefore, we define Λ_0, contained in the interior of Λ, to be a compact and convex set containing λ_0 as an interior point. Assumption 4 is used in several papers that prove the asymptotic normality of the QML estimator in a GARCH framework. Furthermore, Straumann and Mikosch (2006), and Francq and Zakoïan (2004) also make use of a set having similar properties like Λ_0.

Another importance of Assumption 4 is that it allows the asymptotic distribution of QML estimator to be Gaussian. For instance, if β_0 were equal to zero, then $\sqrt{T}(\hat{\beta}_T - \beta_0) = \sqrt{T}\hat{\beta}_T \geq 0$ for all T, where $\hat{\beta}_T$ is the QML estimator of β_0. However, the asymptotic distribution of an estimator which is positive for every T cannot be standard normal distribution. Thus, as pointed out in Francq and Zakoïan (2007), when the true parameter value is on the boundary then the resulting asymptotic distribution is not Gaussian anymore.

Assumption 5: $E\left[\left(\beta_0 + (\alpha_{0,1} + \alpha_{0,2})\,\varepsilon_t^2\right)^2\right] < 1.$

In other words, we assume that Assumption 1.b holds with $r = 2$. Basically, this assumption is on the distribution of ε_t, it provides us the finiteness of the fourth moments of the innovations, i.e., $E[\varepsilon_t^4] < \infty$. We need it for the finiteness of the variance of $\partial l_t^* (\lambda_0) / \partial \lambda$. Additionally, together with the other assumptions, we will be able to prove the finiteness of $E[\sigma_t^4]$ and the fourth moment of y_t. The finite fourth moments of the innovations, directly or indirectly, is assumed in many papers, including Straumann and Mikosch (2006) (assumption N.3.i.), and Francq and Zakoïan (2004)(assumption A6).

4.2 Results and the Main Theorem for Asymptotic Normality

In this section, we introduce necessary results that will be used in proving the asymptotic normality theorem. They are about the derivatives of the function g, and the differentiability properties of h_t and its ergodic and stationary counterpart h_t^*. In fact, we will show that the derivatives of h_t converge to those of h_t^*. Again, the results in this section have similarities in the corresponding results of Straumann and Mikosch (2006), and Francq and Zakoïan (2004).

First, we state a result on the derivatives of the function g. Let's denote the first and second partial derivatives of g with $g_{u_1} = \partial g(y, h; \lambda)/ \partial u_1$ and $g_{u_1 u_2} = \partial^2 g(y, h; \lambda)/ \partial u_1 \partial u_2'$, where u_1 and u_2 can be any of y, h and λ.

Result 6:

i) *For some finite C_1 and C_2, and all $(y, x, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \Lambda_0$, the quantities $|g_\lambda|$ and $|g_{\lambda\lambda}|$ evaluated at (y, x, λ) are bounded by $C_1(1 + y^2 + x)$ and $C_2 y^2$, respectively.*

ii) *For some $\kappa' < \infty$ and all $(y, x_1, x_2) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$ we have*

$$\begin{aligned} |g_u(y, x_1; \lambda) - g_u(y, x_2; \lambda)| &\leq \kappa' |x_1 - x_2|, \qquad u = h,\ \lambda, \\ |g_{u_1 u_2}(y, x_1; \lambda) - g_{u_1 u_2}(y, x_2; \lambda)| &\leq \kappa' |x_1 - x_2|, \qquad u_1,\ u_2 = h,\ \lambda. \end{aligned}$$

This result will be used to prove the existence of some certain moments involving the partial derivatives of g. Note that, κ' need not be less than 1, thus, the partial derivatives of g need not be contractions as it was the case in Result 2. The inequalities in Result 6 will trivially hold for the partial derivatives with respect to h because $g_h = \beta$. For the other inequalities, we will make use of the properties of partial derivatives of the function G.

The following two results are associated with the derivatives of $h_t(\lambda)$. The differentiation notations are kept here as well, e.g. $h_{\lambda,t} = \partial h_t(\lambda)/\partial\lambda$ and $h_{\lambda\lambda,t} = \partial^2 h_t(\lambda)/\partial\lambda\partial\lambda'$. By a straightforward differentiation of the function $h_t(\lambda) = g(y_{t-1}, h_{t-1}(\lambda); \lambda) = \omega + (\alpha_1 + \alpha_2 G(y_{t-1};\gamma))y_{t-1}^2 + \beta h_{t-1}(\lambda)$ we obtain, for $t = 1, 2, \ldots,$

$$\begin{aligned} h_{\lambda,t}(\lambda) &= g_{\lambda,t} + g_{h,t}h_{\lambda,t-1}(\lambda) \\ &= g_{\lambda,t} + \beta h_{\lambda,t-1}(\lambda), \end{aligned} \tag{4}$$

$$h_{\lambda\lambda,t}(\lambda) = g_{\lambda\lambda,t} + e_4 h'_{\lambda,t-1}(\lambda) + h_{\lambda,t-1}(\lambda)e'_4 + \beta h_{\lambda\lambda,t-1}(\lambda), \tag{5}$$

where e_4 is a (6×1) unit vector having 1 in the fourth entry and 0 otherwise, i.e., $e_4 = (0,0,0,1,0,0)'$. In the following results we will use the stationary ergodic counterparts of the derivatives of the function g. Let's define $g^*_{\lambda,t} = [g_\lambda]_{h=h^*_{t-1}(\lambda)} = \partial g(y_{t-1}, h^*_{t-1}(\lambda); \lambda)/\partial\lambda$ as the partial derivative of g evaluated at $h = h^*_{t-1}(\lambda)$, where $h^*_t(\lambda)$ is the stationary ergodic solution to the equation (3) in Result 3. We define $g^*_{\lambda\lambda,t}$ in a similar way.

Result 7: *Suppose Assumptions 1, 2, 4 and 5 hold. Then, for all $\lambda \in \Lambda_0$, there exists a stationary and ergodic solution $h^*_{\lambda,t}(\lambda)$ to the equation*

$$h_{\lambda,t}(\lambda) = g^*_{\lambda,t} + \beta h_{\lambda,t-1}(\lambda), \qquad t = 1,\ 2, \ldots. \tag{6}$$

Moreover, $h^*_{\lambda,t}(\lambda)$ is measurable with respect to the $\sigma-$algebra generated by $(y_{t-1}, y_{t-2}, \ldots)$ and is unique when the recursive equation (6) is extended to all $t \in \mathbb{Z}$. The ergodic and stationary solution $h^*_t(\lambda)$ obtained from Result 1 is continuously partially differentiable on Λ_0 for every $t \in \mathbb{Z}$ and $\partial h^*_t(\lambda)/\partial\lambda = h^*_{\lambda,t}(\lambda)$. Furthermore, $E[\sup_{\lambda\in\Lambda} |h^*_{\lambda,t}(\lambda)|^{r/2}] < \infty$ and $\|\sup_{\lambda\in\Lambda} |h^*_t - h_t|\|_{r/4} \le C' \max\{t, t^{4/r}\}\,\kappa^{t-1}$, where $C' \in \mathbb{R}$ and κ is as it was in Result 3. Finally, for $\lambda \in \Lambda_0$, if $h_t(\lambda)$ and $h_{\lambda,t}(\lambda)$ are any other solutions to the difference equations (3) and (6), respectively, then for some $\gamma > 1$,

$$\gamma^t \sup_{\lambda\in\Lambda_0} \left| h^*_{\lambda,t}(\lambda) - h_{\lambda,t}(\lambda) \right| \to 0 \ \ \textit{in } L_{r/4} - \textit{norm as } t \to \infty.$$

This result can be viewed as an analogue of Result 3 for the first derivatives. It shows that $h_t^*(\lambda)$, the stationary ergodic solution to (3), is continuously differentiable and that its derivative is also a stationary ergodic process which solves (6). This result will be used in showing that $L_T^*(\lambda)$ is continuously differentiable with an ergodic stationary derivative. Moreover, Result 7 implies that for any other solution $h_t(\lambda)$ to the equation (3), the difference between its derivative $h_{\lambda,t}(\lambda)$ and $h_{\lambda,t}^*(\lambda)$ converges to zero exponentially fast and uniformly over Λ_0. Hence, by this result, we will be able to prove that the first derivative of $L_T^*(\lambda)$ can be served as an approximation to the first derivative of $L_T(\lambda)$.

Result 8: *Suppose Assumptions 1, 2, 4 and 5 hold. Then, for all $\lambda \in \Lambda_0$, there exists a stationary and ergodic solution $h_{\lambda\lambda,t}^*(\lambda)$ to the equation*

$$h_{\lambda\lambda,t}(\lambda) = g_{\lambda\lambda,t}^* + e_4 h_{\lambda,t-1}^{*\prime}(\lambda) + h_{\lambda,t-1}^*(\lambda) e_4' + \beta h_{\lambda\lambda,t-1}(\lambda), \qquad t = 1, 2, ... \tag{7}$$

Moreover, $h_{\lambda\lambda,t}^*(\lambda)$ is measurable with respect to the $\sigma-$algebra generated by $(y_{t-1}, y_{t-2}, ...)$ and is unique when (7) is extended to all $t \in \mathbb{Z}$. The ergodic and stationary solution $h_t^*(\lambda)$ obtained from Result 1 is continuously partially differentiable on Λ_0 for every $t \in \mathbb{Z}$ and $\partial^2 h_t^*(\lambda)/\partial\lambda\partial\lambda' = h_{\lambda\lambda,t}^*(\lambda)$. Furthermore, $E[\sup_{\lambda\in\Lambda} |h_{\lambda\lambda,t}^*(\lambda)|^{r/4}] < \infty$. Finally, for $\lambda \in \Lambda_0$, if $h_t(\lambda)$, $h_{\lambda,t}(\lambda)$ and $h_{\lambda\lambda,t}(\lambda)$ are any other solutions to the difference equations (3), (6) and (7), then for some $\gamma > 1$,

$$\gamma^t \sup_{\lambda\in\Lambda_0} \left| h_{\lambda\lambda,t}^*(\lambda) - h_{\lambda\lambda,t}(\lambda) \right| \to 0 \ \text{ in } L_{r/8} - \text{norm as } t \to \infty.$$

This result, concerning the second derivatives of $h_t(\lambda)$ and $h_t^*(\lambda)$, is similar to Result 7. Regarding the moments and convergence results in the last two previous results, the values ($r/2$, $r/4$ and $r/8$) are not strict if these results hold for some positive exponents. As a matter of fact, in Result 9 we will show that r can be replaced by 2 in all of the results (Results 1, 3, 7 and 8) having r as either moment or convergence result. Result 7 and 8 are alike to Proposition 6.1 and 6.2 in Straumann and Mikosch (2006), respectively.

Result 9: *Suppose Assumptions 1, 2, 4 and 5 hold. Then, Result 1 holds with $r = 2$, specifically, $E[\sigma_t^4] < \infty$. Moreover,*

$$\left\| \sup_{\lambda\in\Lambda_0} \frac{\left| h_{\lambda,t}^*(\lambda) \right|}{h_t^*(\lambda)} \right\|_4 < \infty \quad \text{and} \quad \left\| \sup_{\lambda\in\Lambda_0} \frac{\left| h_{\lambda\lambda,t}^*(\lambda) \right|}{h_t^*(\lambda)} \right\|_2 < \infty.$$

This result is equivalent to an assumption (N.3) in Straumann and Mikosch (2006). One of the consequences of Result 9, together with the Assumption 5, is the finiteness of the fourth moments of y_t. The inequalities regarding the derivatives of h_t^* will be used in proving the finiteness of the expectation, for $\lambda \in \Lambda_0$,

$$\mathcal{J}(\lambda) \overset{def}{=} E\left[\frac{\partial^2 L_T^*(\lambda)}{\partial\lambda\partial\lambda'}\right],$$

where $\lambda \in \Lambda_0$. Thus, by the help of the other results, we will be able to use the uniform Strong Law of Large Numbers for the derivatives of $L_T^*(\lambda)$ and $h_t^*(\lambda)$. The expectation $\mathcal{J}(\lambda)$ is important because its inverse at the true parameter value λ_0 takes place in the asymptotic variance of QML estimator. The properties and some convergence results regarding $\mathcal{J}(\lambda)$ will be thoroughly analyzed in Appendix E.

Result 10:
i) *The distribution of ε_t is not concentrated at two points.*
ii) *For $x_\lambda \in \mathbb{R}^6$, $x_\lambda' \frac{\partial g\left(y_t,\sigma_t^2;\gamma_0\right)}{\partial\lambda} = 0$ a.s. only if $x_\lambda = 0$.*

The second part of our last result implies the linear independence of the elements of $\partial g\left(y_t, \sigma_t^2; \gamma_0\right)/\partial\lambda$ with probability one, hence, it is an identification result. It is parallel to Assumption N.4 of Straumann and Mikosch (2006). Together with the first part of Result 10, they guarantee the positive definiteness of the asymptotic variance matrix of the QML estimator of the true parameter λ_0. Hence, Result 10 assures that the asymptotic covariance matrix is regular.

Now, we can state the main theorem of the asymptotic normality section.

Theorem 2 *Suppose Assumptions 1, 2, 4 and 5 hold. Then,*

$$\sqrt{T}\left(\hat{\lambda}_T - \lambda_0\right) \to_d N\left(0,\ E\left[\varepsilon_t^4 - 1\right]\mathcal{J}(\lambda_0)^{-1}\right),$$

where $\mathcal{J}(\lambda_0) = E\left[\sigma_t^{-4} h_{\lambda,t}^*(\lambda_0)\, h_{\lambda,t}^{*\prime}(\lambda_0)\right]$ is positive definite.

As in the proof of Theorem 1, we will follow a similar approach like Straumann and Mikosch (2006), and Francq and Zakoïan (2004). We first establish the asymptotic normality of an infeasible QML estimator, will be denoted as $\tilde{\lambda}_T$, that minimizes the likelihood function $L_T^*(\lambda)$. A mean value expansion of the first derivative of $L_T^*(\lambda)$ will be used as a technical tool for the asymptotic normality of $\tilde{\lambda}_T$. Next, the asymptotic equivalence of the feasible and infeasible estimators will be established, so that $\sqrt{T}(\tilde{\lambda}_T - \hat{\lambda}_T) \to 0$ a.s. Finally, in order to compute consistent estimator of the asymptotic covariance matrix, we introduce consistent estimators for $E[\varepsilon_t^4 - 1]$ and $\mathcal{J}(\lambda_0)$. They are

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t^4}{\hat{h}_t^2} - 1\right) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{h}_{\lambda,t}}{\hat{h}_t}\frac{\hat{h}'_{\lambda,t}}{\hat{h}_t}\right), \tag{8}$$

respectively. Here, the hat "^" signifies that the variable is evaluated at the feasible QML estimator $\hat{\lambda}_T$. Obviously, if ε_t is distributed normally, thus $E[\varepsilon_t^4] = 3$, then the asymptotic covariance simplifies to $2\mathcal{J}(\lambda_0)^{-1}$.

5 Conclusion

In applied econometrics, conditionally heteroskedastic models are commonly used. In this paper, we have studied some asymptotic properties of QML estimator in a model where the returns are conditionally heteroskedastic. The conditional variance of the returns is specified to have a smooth transition GARCH$(1,1)$ model with a logistic function as transition mechanism. This specific nonlinear model in the conditional variance allows the asymmetric response to positive and negative shocks. It is a generalization of threshold GARCH models since our model reduces to a threshold model for specific choices of parameters.

In this thesis, we proved the strong consistency and the asymptotic normality of QML estimator. In fact, these asymptotic properties of QML estimator in our nonlinear GARCH model are proved under conditions which are as mild as in linear GARCH models (see for instance Lee and Hansen (1994), Lumsdaine (1996), Berkes, Horváth and Kokoszka (2003)).

This master thesis relies on the results of Meitz and Saikkonen (2008c). But, our approach is also similar to Straumann and Mikosch (2006), and Francq and Zakoïan (2004). References to these papers are made at several steps throughout the paper. This thesis can be generalized in various ways. An extension to nonlinear AR-GARCH models has already be done by Meitz and Saikkonen (2008c). It may be further extended to nonlinear ARMA-GARCH models. Another generalization might be letting the true parameter value λ_0 be on the boundary. This case is handled in a pure linear GARCH(p, q) model by Francq and Zakoïan (2007). However, as discussed in section 2.1, the resulting asymptotic distribution is a non-Gaussian one. For instance, if we let β_0 to be equal to 0, then for all T, $\sqrt{T}(\hat{\beta}_T - \beta_0) = \sqrt{T}\hat{\beta}_T \geq 0$ cannot have a Gaussian distribution. By making minor changes in our assumptions, a potential future work can be considered as obtaining finite moments of any order for the terms in Result 9, i.e. $|| \sup_{\lambda \in \Lambda_0} |h^*_{\lambda,t}(\lambda)| / h^*_t(\lambda) ||_v < \infty$ and $|| \sup_{\lambda \in \Lambda_0} |h^*_{\lambda\lambda,t}(\lambda)| / h^*_t(\lambda) ||_v < \infty$ for any $v > 0$.

Appendices

Appendix A: Properties of the Function G

In this part we will analyze the nonlinear function G. By taking the first derivative with respect to y, we see that the function is strictly increasing at the true parameter value γ_0.

$$\frac{\partial G(y; \gamma_0)}{\partial y} = \frac{\gamma_{0,2} e^{-\gamma_{0,2}(y - \gamma_{0,1})}}{\left[1 + e^{-\gamma_{0,2}(y - \gamma_{0,1})}\right]^2} > 0.$$

Because the exponential function always takes nonzero numbers and because the derivative of exponential function exists up to any order, we conclude that $G(\cdot; \gamma_0)$ is strictly increasing and its derivative exists up to any order and is continuous. Moreover, the function G is continuous in the parameter values, thus $G(\cdot; \cdot)$ is continuous.

We know that the function G takes values close to 1 as y_{t-1} gets larger, and values close to 0 as y_{t-1} gets smaller. However, in the conditional variance equation we have the multiplication of these two functions. Therefore, let's analyze the behavior of $G(y_{t-1}; \gamma) y_{t-1}$ as y_{t-1} takes very large and very small values. We have the following equalities for all $\gamma \in \mathbb{R} \times \mathbb{R}_+$.

$$\begin{aligned} \lim_{y \to -\infty} y^2 G(y; \gamma) &= \lim_{y \to -\infty} \frac{y^2}{1 + e^{-\gamma_2(y - \gamma_1)}} \\ &= \lim_{y \to -\infty} \frac{2}{\gamma_2^2 e^{-\gamma_2(y - \gamma_1)}} = 0 \quad \text{by L'Hôpital Rule}^2. \end{aligned} \tag{9}$$

$$\begin{aligned}
\lim_{y\to\infty} y^2\left[1-G(y;\gamma)\right] &= \lim_{y\to\infty} \frac{y^2}{\frac{1}{e^{-\gamma_2(y-\gamma_1)}}+1} = 0 \\
&= \lim_{y\to\infty} \frac{2}{\gamma_2^2 e^{\gamma_2(y-\gamma_1)}} = 0 \qquad \text{by a similar argument.} \qquad (10)
\end{aligned}$$

Hence, we see that the function G converges to its extreme values much faster than y_{t-1} does. The next analysis is done for identification. Assume that $\gamma \neq \gamma_0$. Then,

$$\begin{aligned}
y &\neq \tfrac{\gamma_2\gamma_1-\gamma_{0,2}\gamma_{0,1}}{\gamma_2-\gamma_{0,2}} && \textit{iff} \\
-\gamma_2(y-\gamma_1) &\neq -\gamma_{0,2}(y-\gamma_{0,1}) && \textit{iff} \\
\tfrac{1}{1+e^{-\gamma_2(y-\gamma_1)}} &\neq \tfrac{1}{1+e^{-\gamma_{0,2}(y-\gamma_{0,1})}} && \textit{iff} \\
G(y;\gamma) &\neq G(y;\gamma_0). && (11)
\end{aligned}$$

Hence, if $\gamma \neq \gamma_0$, then we have $G(y;\gamma) \neq G(y;\gamma_0)$ for $y \neq (\gamma_2\gamma_1-\gamma_{0,2}\gamma_{0,1})(\gamma_2-\gamma_{0,2})^{-1}$. In other words, the function G takes different values for different parameters for all y's except for a particular choice.

Next, analyze the properties of the derivatives of the function G. We denote the partial derivatives of G as it is denoted for the functions g and h (e.g. $G_\gamma = \partial G(y;\gamma)/\partial\gamma$). The first partial derivatives of G are given below.

$$\begin{aligned}
G_\gamma(y;\gamma) &= \begin{bmatrix} G_{\gamma_1}(y;\gamma) \\ G_{\gamma_2}(y;\gamma) \end{bmatrix} = \left(1+e^{-\gamma_2(y-\gamma_1)}\right)^{-2} \begin{bmatrix} -\gamma_2 e^{-\gamma_2(y-\gamma_1)} \\ (y-\gamma_1)e^{-\gamma_2(y-\gamma_1)} \end{bmatrix} \\
G_y(y;\gamma) &= \left(\gamma_2 e^{-\gamma_2(y-\gamma_1)}\right)\left[1+e^{-\gamma_2(y-\gamma_1)}\right]^{-2}
\end{aligned}$$

Now, we will prove that the partial derivative G_γ at the true values converge to zero faster than the convergence of y^2 to $\pm\infty$. We apply L'Hôpital Rule twice and three times,

[2] L'Hôpital Rule: Suppose f and g are real and differentiable in (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$, where $-\infty \leq a < b \leq \infty$. Suppose $[f'(x)/g'(x)] \to A$ as $x \to a$. If $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$, or $g(x) \to +\infty$ as $x \to a$ then

$$\frac{f(x)}{g(x)} \to A \quad \text{as} \quad x \to a.$$

The analogous statement is of course true if $x \to b$, or $g(x) \to -\infty$. Here we take $f(x)$ as y^2, and $g(x)$ as $1+\exp(-\gamma_2(y-\gamma_1))$. We apply the rule twice.

respectively, to obtain that

$$\begin{aligned}
\lim_{y\to\pm\infty} y^2 G_{\gamma_1}(y;\gamma_0) &= \lim_{y\to\pm\infty} \frac{-\gamma_{2,0} y^2 e^{-\gamma_{2,0}(y-\gamma_{1,0})}}{\left[1+e^{-\gamma_{2,0}(y-\gamma_{1,0})}\right]^2} \\
&= \lim_{y\to\pm\infty} \frac{-\gamma_{2,0} y^2}{e^{\gamma_{2,0}(y-\gamma_{1,0})}+2+e^{-\gamma_{2,0}(y-\gamma_{1,0})}} \\
&= \lim_{y\to\pm\infty} \frac{-2}{\gamma_{2,0}^2 e^{\gamma_{2,0}(y-\gamma_{1,0})}+\gamma_{2,0}^2 e^{-\gamma_{2,0}(y-\gamma_{1,0})}} \\
&= 0. \qquad (12)
\end{aligned}$$

$$\begin{aligned}
\lim_{y\to\pm\infty} y^2 G_{\gamma_2}(y;\gamma_0) &= \lim_{y\to\pm\infty} \frac{y^2 (y-\gamma_{1,0}) e^{-\gamma_{2,0}(y-\gamma_{1,0})}}{\left[1+e^{-\gamma_{2,0}(y-\gamma_{1,0})}\right]^2} \\
&= \lim_{y\to\pm\infty} \frac{y^3-\gamma_{1,0} y^2}{e^{\gamma_{2,0}(y-\gamma_{1,0})}+2+e^{-\gamma_{2,0}(y-\gamma_{1,0})}} \\
&= \lim_{y\to\pm\infty} \frac{3}{\gamma_{2,0}^3 e^{\gamma_{2,0}(y-\gamma_{1,0})}-\gamma_{2,0}^3 e^{-\gamma_{2,0}(y-\gamma_{1,0})}} \\
&= 0. \qquad (13)
\end{aligned}$$

Note that after L'Hôpital Rule is applied, in either case, i.e. $y \to +\infty$ and $y \to -\infty$, the denominators diverge to $\pm\infty$.

For the identification of the partial derivative $G_\gamma(y;\gamma_0)$ we will prove that there exists $\bar{y}$ such that $(a,b)' G_\gamma(\bar{y};\gamma_0) \neq 0$ for $(a,b) \in \mathbb{R}^2 \setminus \{(0,0)\}$.

$$(a,b)' G_\gamma(y;\gamma_0) \neq 0 \qquad \textit{iff}$$

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \frac{-\gamma_{2,0} e^{-\gamma_{2,0}(y-\gamma_{1,0})}}{\left(1+e^{-\gamma_{2,0}(y-\gamma_{1,0})}\right)^2} \\ \\ \frac{(y-\gamma_{1,0}) e^{-\gamma_{2,0}(y-\gamma_{1,0})}}{\left(1+e^{-\gamma_{2,0}(y-\gamma_{1,0})}\right)^2} \end{bmatrix} \neq 0 \qquad \textit{iff}$$

$$\frac{\left(-a\gamma_{2,0}+b(y-\gamma_{1,0})\right)e^{-\gamma_{2,0}(y-\gamma_{1,0})}}{\left[1+e^{-\gamma_{2,0}(y-\gamma_{1,0})}\right]^2} \neq 0 \qquad \textit{iff}$$

$$b(y-\gamma_{1,0}) \neq a\gamma_{2,0} \qquad \textit{iff}$$

$$\frac{a\gamma_{2,0}}{b}-\gamma_{1,0} \neq y. \tag{14}$$

Thus, if $(a,b)\neq(0,0)$ then $(a,b)'G_\gamma(y;\gamma_0)\neq 0$ for all y's except one for $y=b^{-1}a\gamma_{2,0}-\gamma_{1,0}$. In other words, the partial derivatives $G_{\gamma_1}(y;\gamma_0)$ and $G_{\gamma_2}(y;\gamma_0)$ are linearly independent unless $y=b^{-1}a\gamma_{2,0}-\gamma_{1,0}$.

Let's prove that $G(\cdot;\cdot)$ is twice continuously differentiable on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}_+$ - actually it is infinitely many times continuously differentiable, but we only need up to the second continuous derivatives of G in our analysis. Let $f_1(x)=x^{-1}$ and $f_2(y,\gamma_1,\gamma_2)=1+e^{-\gamma_2(y-\gamma_1)}$. Note that, f_1 is twice continuously differentiable for $x>0$, and f_2 is three times continuously differentiable for any $(y,\gamma_1,\gamma_2)\in\mathbb{R}\times\mathbb{R}\times\mathbb{R}_+$. The composition of two twice continuously differentiable functions is also twice continuously differentiable. Thus, $(f_1\circ f_2)(y,\gamma_1,\gamma_2)=\left[1+e^{-\gamma_2(y-\gamma_1)}\right]^{-1}$ is twice continuously differentiable because $f_2(y,\gamma_1,\gamma_2)>0$ for any y,γ_1 and γ_2. The second partial derivatives with respect to parameters are found as the following.

$$\begin{aligned} G_{\gamma\gamma}(y;\gamma) &= \begin{bmatrix} G_{\gamma_1\gamma_1}(y;\gamma) & G_{\gamma_1\gamma_2}(y;\gamma) \\ G_{\gamma_2\gamma_1}(y;\gamma) & G_{\gamma_2}(y;\gamma) \end{bmatrix} \\ &= \frac{e^{-\gamma_2(y-\gamma_1)}\left(e^{-\gamma_2(y-\gamma_1)}-1\right)}{\left[1+e^{-\gamma_2(y-\gamma_1)}\right]^3} \begin{bmatrix} \gamma_2^2 & 1-\gamma_2(y-\gamma_1)-2\left(1-e^{\gamma_2(y-\gamma_1)}\right)^{-1} \\ G_{\gamma_1\gamma_2}(y;\gamma) & (y-\gamma_1)^2 \end{bmatrix} \end{aligned}$$

Note that $G_{\gamma_2\gamma_1}(y;\gamma)=G_{\gamma_1\gamma_2}(y;\gamma)$ for any (y,γ) because G is twice continuously differentiable. Since $G(\cdot;\cdot)$ is twice continuously differentiable on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}_+$ it is obviously differentiable on any $\mathbb{R}\times N(\gamma_0)$ where $N(\gamma_0)$ denotes a neighborhood of the true parameters γ_0. Because the statement is true for any neighborhood we fix one neighborhood for the rest of the analysis throughout the paper. For convenience, let's take $N(\gamma_0)=(\gamma_0/2,\ 2\gamma_0)$. Note that, γ_2 is still strictly positive uniformly in $N(\gamma_0)$. Hence, we can prove now the boundedness of $G_{\gamma\gamma}$ in absolute value uniformly over $\mathbb{R}\times N(\gamma_0)$. Note that,

$$\frac{e^{-\gamma_2(y-\gamma_1)}\left(e^{-\gamma_2(y-\gamma_1)}-1\right)}{\left[1+e^{-\gamma_2(y-\gamma_1)}\right]^3}\leq\frac{1}{\left[e^{\frac{2}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{1}{3}\gamma_2(y-\gamma_1)}\right]^3}+\frac{1}{\left[e^{\frac{1}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{2}{3}\gamma_2(y-\gamma_1)}\right]^3}.$$

By using the inequality above we found limits of the second partials as the followings.

$$\begin{aligned}\lim_{y\to\pm\infty}\left|G_{\gamma_1\gamma_1}(y;\gamma)\right| &\leq \lim_{y\to\pm\infty}\frac{\gamma_2^2}{\left[e^{\frac{2}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{1}{3}\gamma_2(y-\gamma_1)}\right]^3}+\frac{\gamma_2^2}{\left[e^{\frac{1}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{2}{3}\gamma_2(y-\gamma_1)}\right]^3}\\ &= 0.\end{aligned}$$

$$\begin{aligned}\lim_{y\to\pm\infty}\left|G_{\gamma_2\gamma_2}(y;\gamma)\right| &\leq \lim_{y\to\pm\infty}\frac{(y-\gamma_1)^2}{\left[e^{\frac{2}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{1}{3}\gamma_2(y-\gamma_1)}\right]^3}+\frac{(y-\gamma_1)^2}{\left[e^{\frac{1}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{2}{3}\gamma_2(y-\gamma_1)}\right]^3}\\ &= 0.\end{aligned}$$

$$\begin{aligned}\lim_{y\to\pm\infty}\left|G_{\gamma_1\gamma_2}(y;\gamma)\right| &\leq \lim_{y\to\pm\infty}\frac{1+\gamma_2(y-\gamma_1)}{\left[e^{\frac{2}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{1}{3}\gamma_2(y-\gamma_1)}\right]^3}+\frac{3+\gamma_2(y-\gamma_1)}{\left[e^{\frac{1}{3}\gamma_2(y-\gamma_1)}+e^{-\frac{2}{3}\gamma_2(y-\gamma_1)}\right]^3}\\ &= 0.\end{aligned}$$

In the last two results the divergence rate of the nominator to $+\infty$ is much slower than that of the denominator. Hence, following the previous arguments the limits are zero. Also notice that, none of the three limits are affected by the parameters $\gamma\in N(\gamma_0)$. Thus, we conclude that all of the partial derivatives in $G\gamma\gamma$ are bounded in absolute value uniformly over $\mathbb{R}\times N(\gamma_0)$. We can conclude that for $\lambda\in\mathbb{R}\times N(\gamma_0)$ there exist constants M_i for $i=1,\ldots,5$ such that $|G_{\gamma_1}|\leq M_1, |G_{\gamma_2}|\leq M_2$, $|G_{\gamma_1\gamma_1}|\leq M_3, |G_{\gamma_1\gamma_2}|\leq M_4$, $|G_{\gamma_2\gamma_2}|\leq M_5$.

Appendix B: Auxiliary Lemmas

Lemma 1 *Let $x_1, x_2, \ldots, x_k$ be random variables. Then, for any $r>0$,*

$$\left\|\sum_{i=1}^k x_i\right\|_r\leq\Delta_{r,k}\sum_{i=1}^k\|x_i\|_r \ \ \textit{where}\ \Delta_{r,k}=\max\left\{1,k^{(1-r)/r}\right\}.$$

Proof. For the case $r\geq 1$ Minkowski's Inequality[3] applies, i.e.,

$$\left\|\sum_{i=1}^k x_i\right\|_r\leq\sum_{i=1}^k\|x_i\|_r\leq\Delta_{r,k}\sum_{i=1}^k\|x_i\|_r.$$

[3] Minkowski's Inequality: $\|X+Y\|_r\leq\|X\|_r+\|Y\|_r$.

A recursive generalization of Minkowski's Inequality is found by $\left\|\sum_{i=1}^k X_i\right\|_r\leq\sum_{i=1}^k\|X_i\|_r$.

For $0 < r < 1$,

$$\begin{aligned}
\left\|\sum_{i=1}^{k} x_i\right\|_r &= \left\{E\left[\left|\sum_{i=1}^{k} x_i\right|^r\right]\right\}^{1/r} \leq \left\{c_r \sum_{i=1}^{k} E\left[|x_i|^r\right]\right\}^{1/r} \text{ by Loève's } c_r\text{-Inequality}^4, \\
&= c_r^{1/r}\left(\sum_{i=1}^{k} E\left[|x_i|^r\right]\right)^{1/r} \\
&= c_r^{1/r} E\left\{\left|\sum_{i=1}^{k} E\left[|x_i|^r\right]\right|^{1/r}\right\} \\
&\leq c_r^{1/r} c_{1/r} \sum_{i=1}^{k} E\left\{\left(E\left[|x_i|^r\right]\right)^{1/r}\right\} \quad \text{by Loève's } c_r\text{-Inequality for } 1/r, \\
&= c_r^{1/r} c_{1/r} \sum_{i=1}^{k} \left(E\left[|x_i|^r\right]\right)^{1/r} \qquad \text{where } c_r^{1/r} = 1 \text{ and } c_{1/r} = k^{1/r-1}, \\
&\leq \Delta_{r,k} \sum_{i=1}^{k} \|x_i\|_r .
\end{aligned}$$

The third equality holds because $E\left[|x_i|^r\right]$ is a nonnegative scalar. ■

Lemma 2 *Suppose for some $r > 0$, $\gamma > 1$, and nonnegative process x_t, $\gamma^t x_t$ converges to zero in L_r-norm, i.e., $\|\gamma^t x_t\|_r \to 0$ as $t \to \infty$. Then $\sum_{t=1}^{\infty} x_t < \infty$ a.s. and $\|\sum_{t=1}^{\infty} x_t\|_r < \infty$ also holds.*

Proof. See Meitz and Saikkonen (2008c), p.26. ■

Lemma 3 *For all $x \geq 0$ and all $s \in (0,1)$, we have $x/(1+x) \leq x^s$.*

Proof. Let $s \in (0,1)$. We have three cases. If $x = 0$, then it is obvious. If $x \in (0,1)$, then $(1+x)^{-1} < 1 < x^{s-1}$. If $x \geq 1$, then $x/(1+x) < 1 \leq x^s$. ■

Appendix C: Derivations Of Results 1-5

Result 1: Suppose Assumption 1 holds. Our main aim is to satisfy Proposition 1 and Theorem 1 of Meitz and Saikkonen (2008a) which provides the ergodicity and stationarity of the process (y_t, σ_t^2) such that σ_t^2 have moments of order r. Proposition 1 is satisfied by the properties of (y_t, σ_t^2). Since we do not have any conditional mean in our model, all premises of Theorem 1 of Meitz and Saikkonen (2008a) associated with the conditional mean are satisfied. The rest premises are ensured by Assumption 1.a and 1.b, and by the

[4] Loève's c_r-Inequality: $E\left[\left|\sum_{i=1}^{k} X_i\right|^r\right] \leq c_r \sum_{i=1}^{k} E\left[|X_i|^r\right]$ where $c_r = 1$ when $0 < r \leq 1$, and $c_r = k^{r-1}$ when $r \geq 1$.

properties of the conditional variance function. The theorem and the proposition provide finiteness of $E\left[\sigma_t^{2r}\right]$ and the joint ergodicity and stationarity of the process (y_t, σ_t^2).

Moreover, from Result 1 we can obtain that $E[y_t^{2r}] = E[\sigma_t^{2r}]E[\varepsilon_t^{2r}] < \infty$ by the independence of ε_t's and by Assumption 1.b.

Result 2: The function $g(y, x; \lambda) = \omega + (\alpha_1 + \alpha_2 G(y;\gamma))y^2 + \beta x$ is clearly continuous. The parameter space being compact, hence bounded, there exist $\overline{\omega}$, $\varkappa$ and ρ satisfying $\infty > \overline{\omega} \geq \omega > 0$, $\infty > \varkappa \geq \alpha_1 + \alpha_2 > 0$ and $1 > \rho > \beta \geq 0$. These three properties give us

$$g(y, x; \lambda) = \omega + (\alpha_1 + \alpha_2 G(y;\gamma))y^2 + \beta x \leq \overline{\omega} + \varkappa y^2 + \rho x \text{ since } G \in [0, 1].$$

Let $\lambda \in \Lambda$, $y \in \mathbb{R}$ and $x_1, x_2 \in \mathbb{R}_+$. Then,

$$\begin{aligned} |g(y, x_1; \lambda) - g(y, x_2; \lambda)| &= \left|\left[\omega + (\alpha_1 + \alpha_2 G(y;\gamma))y^2 + \beta x_1\right] - \left[\omega + (\alpha_1 + \alpha_2 G(y;\gamma))y^2 + \beta x_2\right]\right| \\ &= \beta\,|x_1 - x_2| \\ &\leq \kappa\,|x_1 - x_2| \end{aligned}$$

where $0 \leq \beta \leq \kappa < 1$. Such a κ exists because $\beta \in [0, 1)$.

Result 3: The derivation of this result is omitted for now (work in progress).

Result 4: We have $g(y, x; \lambda) = \omega + (\alpha_1 + \alpha_2 G(y;\gamma))y^2 + \beta x \geq \omega > 0$ since $\alpha_1, \alpha_2 \in [0, \infty)$, $G \in [0, 1]$ and $\beta \in [0, 1)$. Hence, $g(y, x; \lambda)$ is bounded uniformly below in Λ.

Result 5: Suppose Assumptions 1 and 2 hold. Assume there exists $\lambda \in \Lambda$ satisfying $h_t^*(\lambda) = \sigma_t^2$. We will show that $\lambda = \lambda_0$. By Result 3, we have $h_{t+1}^*(\lambda_0) = \omega_0 + (\alpha_{0,1} + \alpha_{0,2}G(y_t;\gamma_0))y_t^2 + \beta_0\sigma_t^2 = \sigma_{t+1}^2$. If we subtract $h_{t+1}^*(\lambda_0)$ from $h_{t+1}^*(\lambda)$ and use $y_t = \sigma_t\varepsilon_t$ we obtain

$$(\omega - \omega_0) + (\alpha_1 - \alpha_{0,1})\sigma_t^2\varepsilon_t^2 + \left[\alpha_2 G(\sigma_t\varepsilon_t;\gamma) - \alpha_{0,2}G(\sigma_t\varepsilon_t;\gamma_0)\right]\sigma_t^2\varepsilon_t^2 + (\beta - \beta_0)\sigma_t^2 = 0 \quad a.s. \tag{15}$$

By definition of h_t^* and by Result 4 we have $h_t^* = g(y_{t-1}, h_{t-1}^*; \lambda) \geq \underline{g} > 0$. In particular, $h_t^*(\lambda_0) = \sigma_t^2 \geq \underline{g} > 0$. Hence, we divide both sides of (15) by σ_t^2 and arrange to obtain

$$(\alpha_1 - \alpha_{0,1})\varepsilon_t^2 = -(\beta - \beta_0) - \sigma_t^{-2}\left\{(\omega - \omega_0) + \left[\alpha_2 G(\sigma_t\varepsilon_t;\gamma) - \alpha_{0,2}G(\sigma_t\varepsilon_t;\gamma_0)\right]\sigma_t^2\varepsilon_t^2\right\} \quad a.s. \tag{16}$$

By Result 4 and Assumption 1.a, i.e. ε_t has a density that is positive everywhere, the event $\left\{\sigma_t^2 \geq \underline{g},\ \ \varepsilon_t \leq \underline{g}^{-1/2} M\right\}$ has positive probability for all $M < 0$. We have $-\varepsilon_t \geq -\underline{g}^{-1/2} M$ implying $-\varepsilon_t \sigma_t \geq \underline{g}^{1/2}(-\underline{g}^{-1/2} M) = -M$. Thus, $\sigma_t \varepsilon_t \leq M < 0$ on this event. By (9), $y_t^2 G(y_t; \gamma) = \sigma_t^2 \varepsilon_t^2 G(\sigma_t \varepsilon_t; \gamma) \rightarrow 0$ as $y_t = \sigma_t \varepsilon_t \rightarrow -\infty$. By choosing M small enough, we can make the term $[\alpha_2 G(\sigma_t \varepsilon_t; \gamma) - \alpha_{0,2} G(\sigma_t \varepsilon_t; \gamma_0)]\, \sigma_t^2 \varepsilon_t^2$ arbitrarily close to 0, thus we can make the term in the curly brackets of (16) arbitrarily close to $(\omega - \omega_0)$ on the event $\{\sigma_t \varepsilon_t \leq M\}$. We know that σ_t^{-2} is bounded since $0 < \sigma_t^{-2} \leq \underline{g}^{-1}$. If $\sigma_t \varepsilon_t = M$, then clearly the right-hand-side of (16) is bounded. Also, if we let $\sigma_t \varepsilon_t \rightarrow -\infty$, then it is bounded again by (9). Therefore, the right-hand-side of the equation is bounded on the event $\{\sigma_t \varepsilon_t \leq M\}$. But the left-hand-side may take arbitrarily large values in absolute value if $\alpha_1 \neq \alpha_{0,1}$ and M is chosen small enough. Thus, we must have $\alpha_1 = \alpha_{0,1}$ since $\sigma_t \varepsilon_t \leq M$ with positive probability for all $M < 0$. In other words, unless $\alpha_1 \neq \alpha_{0,1}$ the left-hand-side may take any value with positive probability but the right-hand-side is bounded on the event $\{\sigma_t \varepsilon_t \leq M\}$.

Now we have the restriction $\alpha_1 = \alpha_{0,1}$. Rearranging (16) gives

$$(\alpha_2 - \alpha_{0,2}) \varepsilon_t^2 = -(\beta - \beta_0) - \sigma_t^{-2} \left\{(\omega - \omega_0) + [\alpha_2 (G(\sigma_t \varepsilon_t; \gamma) - 1) - \alpha_{0,2} (G(\sigma_t \varepsilon_t; \gamma_0) - 1)]\, \sigma_t^2 \varepsilon_t^2\right\} \ a.s. \tag{17}$$

Consider the event $\{\sigma_t^2 \geq \underline{g}\ ,\ \varepsilon_t \geq \underline{g}^{-1/2} M\}$. We will follow a similar argument as above. By Result 4 and Assumption 1.a, this event has positive probability for all $M > 0$. On this event, we have $\sigma_t \varepsilon_t \geq M$. By (10), we have that $y_t = \sigma_t \varepsilon_t \rightarrow \infty$ implies that $y_t^2 (1 - G(y_t; \gamma)) = \sigma_t^2 \varepsilon_t^2 (1 - G(\sigma_t \varepsilon_t; \gamma)) \rightarrow 0$ for all γ. By choosing M large enough, we can make the term in the curly brackets of (17) arbitrarily close to $(\omega - \omega_0)$ on the event $\{\sigma_t \varepsilon_t \geq M\}$. Again, the right-hand-side becomes bounded and the left-hand-side may take any value with positive probability unless α_2 is equal to $\alpha_{0,2}$. Therefore, another restriction is $\alpha_2 = \alpha_{0,2}$.

If we arrange (17) again under restrictions, then we obtain

$$(\beta_0 - \beta) \sigma_t^2 = (\omega - \omega_0) + \alpha_{0,2} [G(\sigma_t \varepsilon_t; \gamma) - G(\sigma_t \varepsilon_t; \gamma_0)] \sigma_t^2 \varepsilon_t^2 \quad a.s. \tag{18}$$

Now consider events $\{\sigma_t^2 \in (\sigma_*, \sigma^*),\ \varepsilon_t \leq \underline{\sigma}^{-1/2} M\}$ for some $\underline{\sigma} > 0$ with $\underline{\sigma} < \sigma_* < \sigma^*$ and $M < 0$. We know that $P\{\sigma_t^2 \in (\sigma_*, \sigma^*)\} > 0$ where $0 < \underline{\sigma} < \sigma_* < \sigma^*$. By Assumption

1.a and by the independence of σ_t^2 and ε_t, these events have positive probability. Since $\sigma_t^2 > \underline{\sigma}$, on these events $\sigma_t \varepsilon_t < M$ regardless of the values σ_* and σ^*. By (9) and by similar arguments as above, the right-hand-side of (18) can be arbitrarily close to $(\omega - \omega_0)$ with positive probability if M is chosen small enough. However, the left-hand-side is varying with positive probability if we consider events with different values of σ_* and σ^*,unless $\beta = \beta_0$.

Now we have another restriction $\beta = \beta_0$. Thus, we obtain

$$(\omega_0 - \omega) = \alpha_{0,2} \left[G(\sigma_t \varepsilon_t; \gamma) - G(\sigma_t \varepsilon_t; \gamma_0)\right] \sigma_t^2 \varepsilon_t^2 \quad a.s.$$

If we take limit of both sides as $\sigma_t \varepsilon_t \to -\infty$, then the left-hand-side remains as $\omega_0 - \omega$ whereas the right-hand-side converges to 0 by (9). Thus, another restriction is $\omega = \omega_0$. Since $\alpha_{0,2} > 0$ by Assumption 2, we end up with

$$[G(\sigma_t \varepsilon_t; \gamma) - G(\sigma_t \varepsilon_t; \gamma_0)]\sigma_t^2 \varepsilon_t^2 = 0 \quad a.s. \tag{19}$$

If $\gamma \neq \gamma_0$ then by (11) there exists $\overline{y} = \overline{\sigma_t \varepsilon_t}$ such that $G(\overline{\sigma_t \varepsilon_t}; \gamma) \neq G(\overline{\sigma_t \varepsilon_t}; \gamma_0)$. Consider a neighborhood of $\overline{\sigma_t \varepsilon_t}$. Set $u_* = \overline{\sigma_t \varepsilon_t} - \xi$ and $u^* = \overline{\sigma_t \varepsilon_t} + \xi$ for some $\xi > 0$. The event $\{\sigma_t \varepsilon_t \in (u_*, u^*)\}$ has positive probability because ε_t has a density that is positive everywhere and is independent of σ_t, and additionally because σ_t^2 is nondegenerate since $P\{\sigma_t^2 \in (\sigma_*, \sigma^*)\} > 0$. Since $G(.;.)$ is continuous $[G(\sigma_t \varepsilon_t; \gamma) - G(\sigma_t \varepsilon_t; \gamma_0)]$ is bounded away from zero on the event $\{\sigma_t \varepsilon_t \in (u_*, u^*)\}$. Therefore, we have to have $\gamma = \gamma_0$ which, with other restrictions, leads us to conclude that $\lambda = \lambda_0$. Hence, Result 5 holds.

Appendix D: Derivations Of Results 6-10

Result 6: In this proof we need the first and second partial derivatives of the function $g\left(y, h; \lambda\right) = \omega + \left[\alpha_1 + \alpha_2 G\left(y; \gamma\right)\right] y^2 + \beta h$ with respect to its second and third arguments. Let's find the first and second partial derivatives of g with respect to the parameter value $\lambda = \left(\omega, \alpha_1, \alpha_2, \beta, \gamma_1, \gamma_2\right)$, where the partials are evaluated at $(y, x; \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \Lambda_0$. For short notation, G is used for $G(y;\gamma)$, and similarly for the partial derivatives of G.

$$g_\lambda = \begin{bmatrix} 1 \\ y^2 \\ y^2 G \\ x \\ \alpha_2 y^2 G_{\gamma_1} \\ \alpha_2 y^2 G_{\gamma_2} \end{bmatrix} \quad \text{and} \quad g_{\lambda\lambda} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y^2 G_{\gamma_1} & y^2 G_{\gamma_2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & y^2 G_{\gamma_1} & 0 & \alpha_2 y^2 G_{\gamma_1\gamma_1} & \alpha_2 y^2 G_{\gamma_1\gamma_2} \\ 0 & 0 & y^2 G_{\gamma_2} & 0 & \alpha_2 y^2 G_{\gamma_2\gamma_1} & \alpha_2 y^2 G_{\gamma_2\gamma_2} \end{bmatrix}$$

Because $G(\cdot;\cdot)$ is twice continuously differentiable, so is the function g. Now, we can prove the first part of Result 6, i.e., the boundedness of $|g_\lambda|$ and $|g_{\lambda\lambda}|$. As mentioned in the introduction, Euclidean Norm will be used as the vector and matrix norm.

$$\begin{aligned} |g_\lambda| &= \left[1 + y^4 + y^4 G^2 + x^2 + \alpha_2^2 y^4 \left(G_{\gamma_1}^2 + G_{\gamma_2}^2\right)\right]^{1/2} \\ &\le 1 + y^2 + y^2 G + x + \alpha_2 y^2 \left(\left|G_{\gamma_1}\right| + \left|G_{\gamma_2}\right|\right) \\ &\le 1 + \left[2 + \alpha_2 \left(\left|G_{\gamma_1}\right| + \left|G_{\gamma_2}\right|\right)\right] y^2 + x \qquad \text{since } G \in [0,1], \\ &\le 1 + \left[2 + \alpha_2 (M_1 + M_2)\right] y^2 + x \\ &\le C_1 \left(1 + y^2 + x\right). \end{aligned} \tag{20}$$

where the constant $C_1 \ge [2 + \alpha_2 (M_1 + M_2)]$. The third inequality follows by the finiteness of the first partial derivatives of G, which is discussed in Appendix A. Regarding the second partial derivative of g, we will use similar justifications.

$$\begin{aligned} |g_{\lambda\lambda}| &= \left[y^4 \left\{2\left(G_{\gamma_1}^2 + G_{\gamma_2}^2\right) + \alpha_2^2 \left(G_{\gamma_1\gamma_1}^2 + 2G_{\gamma_1\gamma_2}^2 + G_{\gamma_2\gamma_2}^2\right)\right\}\right]^{1/2} \\ &\le y^2 \left[\sqrt{2}\left(\left|G_{\gamma_1}\right| + \left|G_{\gamma 2}\right|\right) + \alpha_2 \left(\left|G_{\gamma_1\gamma_1}\right| + \sqrt{2}\left|G_{\gamma_1\gamma_2}\right| + \left|G_{\gamma_2\gamma_2}\right|\right)\right] \\ &\le y^2 \left[\sqrt{2}(M_1 + M_2) + \alpha_2 \left(M_3 + \sqrt{2} M_4 + M_5\right)\right] \\ &\le C_2 y^2. \end{aligned} \tag{21}$$

Next, we prove the second part of Result 6. Let $(y, x_1, x_2, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \Lambda_0$. Note that, because $g_h = \beta$, $g_{hh} = 0$, $g_{\lambda h} = 0$, and $g_{h\lambda} = 1$, the required inequalities in the second part of Result 6 are trivially satisfied for the partial derivatives of g with respect to h. For those with respect to λ,

$$\begin{aligned} |g_\lambda(y, x_1;\gamma) - g_\lambda(y, x_2;\gamma)| &= |x_1 - x_2| \\ |g_{\lambda\lambda}(y, x_1;\gamma) - g_{\lambda\lambda}(y, x_2;\gamma)| &= 0. \end{aligned}$$

Hence the inequalities with $\kappa' = 1$ works for the derivatives of g with respect to the second and third arguments.

Result 7 and *Result 8:* See Meitz and Saikkonen (2008c).

Result 9: First, we show that Result 1 holds for $r = 2$. As in the proof of Result 1, we refer to the Theorem 1 of Meitz and Saikkonen (2008a). Assumption 1 and 5, and the properties of σ_t^2 justifies the premises of that theorem. Then, we obtain that $E[\sigma_t^4] < \infty$.

Now, we will show the finiteness of $|| \sup_{\lambda \in \Lambda_0} |h^*_{\lambda,t}(\lambda)| / h^*_t(\lambda) ||_4$. We already assumed that $\lambda_0 \in \mathring{\Lambda}_0$. Furthermore, without loss of generality, we may assume that Λ_0 is small enough so that $\lambda = (\omega, \alpha_1, \alpha_2, \beta, \gamma) \in \Lambda_0$ satisfies $0 < \underline{\omega} \leq \omega \leq \overline{\omega} < \infty$, $0 < \underline{\alpha_1} \leq \alpha_1 \leq \overline{\alpha_1} < \infty$, $0 < \underline{\alpha_2} \leq \alpha_2 \leq \overline{\alpha_2} < \infty$, $0 < \underline{\beta} \leq \beta \leq \overline{\beta} < 1$, and $\gamma \in N(\gamma_0)$. In other words, we are interested in the parameter values that do not lie on the boundary of the general parameter set Λ and that are close to the true parameter value λ_0. We are allowed to assume these because by the help of Results 1-5 we are able to prove that the QML estimator $\hat{\lambda}_T$ is strongly consistent for λ_0.

If we consider the ergodic stationary counterpart of (4) and use recursive substitution we obtain an infinite sum.

$$\begin{aligned} h^*_{\lambda,t} &= g^*_{\lambda,t} + \beta h^*_{\lambda,t-1} \\ &= g^*_{\lambda,t} + \beta \left(g^*_{\lambda,t-1} + \beta h^*_{\lambda,t-2} \right) \\ &\vdots \\ &= \sum_{j=0}^{\infty} \beta^j g^*_{\lambda,t-j}. \end{aligned}$$

In order to prove the convergence of the sum, we need to use Lemma 2 and the upper bound for $g^*_{\lambda,t-j}$ found in (20)

$$\begin{aligned} \left| g^*_{\lambda,t-j} \right| &\leq 1 + y^2_{t-1-j} \left\{ 1 + G(y_{t-1-j}; \gamma) + \alpha_2 \left[\left| G_{\gamma_1}(y_{t-1-j}; \gamma) \right| + \left| G_{\gamma_2}(y_{t-1-j}; \gamma) \right| \right] \right\} + h^*_{t-1-j} \\ &\leq 1 + y^2_{t-1-j} \left[2 + \alpha_2 (M_1 + M_2) \right] + h^*_{t-1-j} \\ &\leq 1 + C_1 y^2_{t-1-j} + h^*_{t-1-j}. \end{aligned} \tag{22}$$

By using this inequality we can write that

$$\left|h^*_{\lambda,t}\right| = \left|\sum_{j=0}^{\infty} \beta^j g^*_{\lambda,t-j}\right| \le \sum_{j=0}^{\infty} \beta^j \left|g^*_{\lambda,t-j}\right| \le \sum_{j=0}^{\infty} \beta^j \left(1 + C y^2_{t-1-j} + h^*_{t-1-j}\right).$$

We have already found the finiteness of $E[y_t^4] = \|y_t^2\|_2^2$. Thus, $\|y_t^2\|_2 < \infty$. By letting $1 < \gamma < \beta^{-1}$ so that $\beta\gamma < 1$, we obtain that for $j \to \infty$

$$\left\|\gamma^j \beta^j \left(1 + C_1 y^2_{t-1-j} + h^*_{t-1-j}\right)\right\|_2 \le (\gamma\beta)^j + C_1 (\gamma\beta)^j \left\|y^2_{t-1-j}\right\|_2 + (\gamma\beta)^j \left\|h^*_{t-1-j}\right\|_2 \to 0 \ a.s.,$$

by stationarity of y_t^2 and h_t^*, and by $E\left[\sup_{\lambda\in\Lambda} h_t^{*2}(\lambda)\right] < \infty$ given in the Result 3 with $r = 2$. Because the term $(1 + C_1 y^2_{t-1-j} + h^*_{t-1-j})$ is non-negative, by Lemma 2 we get $\sum_{k=0}^{\infty} \beta^j (1 + C_1 y^2_{t-1-j} + h^*_{t-1-j}) < \infty$. Hence, we prove the finiteness of $|h^*_{\lambda,t}|$, which implies the convergence of $\sum_{j=0}^{\infty} \beta^j g^*_{\lambda,t-j}$. By using the inequality (22) and the fact that $\beta \in (0,1)$ we can write

$$\begin{aligned}
&\left\|\sup_{\lambda\in\Lambda_0} \frac{|h^*_{\lambda,t}|}{h_t^*}\right\|_4 \\
\le\ & \left\|\sup_{\lambda\in\Lambda_0} \frac{\sum_{j=0}^{\infty} \beta^j \left(1 + C_1 y^2_{t-1-j} + h^*_{t-1-j}\right)}{h_t^*}\right\|_4 \\
\le\ & \frac{1}{1-\beta} \left\|\sup_{\lambda\in\Lambda_0} \frac{1}{h_t^*}\right\|_4 + C_1 \left\|\sup_{\lambda\in\Lambda_0} \frac{\sum_{j=0}^{\infty} \beta^j y^2_{t-1-j}}{h_t^*}\right\|_4 + \left\|\sup_{\lambda\in\Lambda_0} \frac{\sum_{j=0}^{\infty} \beta^j h^*_{t-1-j}}{h_t^*}\right\|_4. \qquad (23)
\end{aligned}$$

In order to show the finiteness of $\|\sup_{\lambda\in\Lambda_0} |h^*_{\lambda,t}(\lambda)| / h_t^*(\lambda)\|_4$ we need to show the finiteness of the three terms in (23). The first term is clearly finite because $h_t^* \ge \underline{g}$ by Result 4. Next, let's show that the second term is finite. Rewrite h_t^* as

$$\begin{aligned}
h_t^* &= \omega + [\alpha_1 + \alpha_2 G(y_{t-1};\gamma)]\, y^2_{t-1} + \beta h^*_{t-1} \\
&= \sum_{k=0}^{\infty} \beta^k \left(\omega + [\alpha_1 + \alpha_2 G(y_{t-1-k};\gamma)]\, y^2_{t-1-k}\right) \\
&\le \sum_{k=0}^{\infty} \beta^k \left(\omega + (\alpha_1 + \alpha_2)\, y^2_{t-1-k}\right) \\
&= \omega (1-\beta)^{-1} + (\alpha_1 + \alpha_2) \sum_{k=0}^{\infty} \beta^k y^2_{t-1-k}. \qquad (24)
\end{aligned}$$

is finite by the same arguments used in proving the finiteness of $|h^*_{\lambda,t}|$. Hence, we find an upper bound for h_t^*. Although we have $\underline{g}$ as a lower bound for h_t^* we need another lower bound that has similar components as the upper bound (24). Because $G \in [0,1]$,

$\beta \in (0,1)$, $\omega \geq \underline{\omega} > 0$, and $\alpha_1 \geq \underline{\alpha_1} > 0$ we have that

$$\begin{aligned} h_t^* &= \sum_{k=0}^{\infty} \beta^k \left(\omega + \left[\alpha_1 + \alpha_2 G\left(y_{t-1-k};\gamma\right)\right] y_{t-1-k}^2\right) \\ &\geq \sum_{k=0}^{\infty} \beta^k \left(\omega + \alpha_1 y_{t-1-k}^2\right) \\ &\geq \omega\left(1-\beta\right)^{-1} + \underline{\alpha_1} \sum_{k=0}^{\infty} \beta^k y_{t-1-k}^2 \\ &\geq \underline{\omega} + \beta^j \underline{\alpha_1} y_{t-1-j}^2 \qquad \text{for any } j \geq 0. \end{aligned} \tag{25}$$

By taking any $0 < s \leq 1/2$ and by making use of Lemma 3, we obtain

$$\begin{aligned} \frac{\beta^j y_{t-1-j}^2}{h_t^*} &\leq \frac{\beta^j y_{t-1-j}^2}{\underline{\omega} + \beta^j \underline{\alpha_1} y_{t-1-j}^2} \\ &= \underline{\alpha_1}^{-1} \frac{\beta^j \underline{\alpha_1} y_{t-1-j}^2 / \underline{\omega}}{1 + \beta^j \underline{\alpha_1} y_{t-1-j}^2 / \underline{\omega}} \\ &\leq \underline{\alpha_1}^{-1} \left(\beta^j \underline{\alpha_1} y_{t-1-j}^2 / \underline{\omega}\right)^s \\ &= \underline{\alpha_1}^{s-1} \underline{\omega}^{-s} \beta^{sj} y_{t-1-j}^{2s}. \end{aligned}$$

By the fact that $s \in (0, 1/2]$, and by using the Liapunov's Inequality[5] we can write that

$$\begin{aligned} \left\|y_{t-1-j}^{2s}\right\|_4 &= \left\{E\left[y_{t-1-j}^{8s}\right]\right\}^{1/4} = \left\{\left(E\left[y_{t-1-j}^{8s}\right]\right)^{1/8s}\right\}^{2s} = \left\|y_{t-1-j}\right\|_{8s}^{2s} \\ &\leq \left\|y_{t-1-j}\right\|_4^{2s} \end{aligned}$$

Thus, we obtain

$$\begin{aligned} \left\|\sup_{\lambda \in \Lambda_0} \sum_{j=0}^{\infty} \frac{\beta^j y_{t-1-j}^2}{h_t^*}\right\|_4 &\leq \left\|\sup_{\lambda \in \Lambda_0} \sum_{j=0}^{\infty} \underline{\alpha_1}^{s-1} \underline{\omega}^{-s} \beta^{sj} y_{t-1-j}^{2s}\right\|_4 \\ &\leq \underline{\alpha_1}^{s-1} \underline{\omega}^{-s} \sum_{j=0}^{\infty} \overline{\beta}^{sj} \left\|y_{t-1-j}^{2s}\right\|_4 \qquad \text{by Minkowski's Inequality,} \\ &\leq \underline{\alpha_1}^{s-1} \underline{\omega}^{-s} \sum_{j=0}^{\infty} \overline{\beta}^{sj} \left\|y_{t-1-j}\right\|_4^{2s} \\ &= \underline{\alpha_1}^{s-1} \underline{\omega}^{-s} M'' \sum_{j=0}^{\infty} \overline{\beta}^{sj} \qquad \text{where } M'' = \left\|y_t\right\|_4^{2s} < \infty, \\ &= \underline{\alpha_1}^{s-1} \underline{\omega}^{-s} M'' \left(1 - \overline{\beta}^s\right)^{-1} \qquad \text{because } \overline{\beta}^s \in (0,1) \text{ for } s \in (0, 1/2], \\ &< \infty. \end{aligned} \tag{26}$$

[5] If $r > p > 0$, then $\|X\|_r \geq \|X\|_p$.

Hence, the finiteness of the second term in (23) is proved. Finally, let's concentrate on the third term, i.e., $||\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j h^*_{t-1-j}/h^*_t||_4$. When we analyze the summand in the norm, and use the same arguments as in proving the finiteness of the second term, we end up with a similar result. We use Lemma 3, and the lower bound of h^*_t for any $j \geq 0$, that is found in (25), and we choose $s \in (0, 1/2]$ in order to get

$$
\begin{aligned}
\frac{\beta^j h^*_{t-j-1}}{h^*_t} &\leq \beta^j \sum_{k=0}^{\infty} \frac{\beta^k \left(\omega + \left[\alpha_1 + \alpha_2 G\left(y_{t-2-j-k};\gamma\right)\right] y^2_{t-2-j-k}\right)}{\omega + \alpha_1 \beta^{j+k+1} y^2_{t-2-j-k}} \\
&\leq \beta^j \sum_{k=0}^{\infty} \beta^k \frac{\overline{\omega} + \left(\alpha_1 + \alpha_2\right) y^2_{t-2-j-k}}{\underline{\omega} + \underline{\alpha_1}\beta^{j+k+1} y^2_{t-2-j-k}} \\
&= \beta^j \sum_{k=0}^{\infty} \frac{\overline{\omega}\beta^k}{\underline{\omega} + \underline{\alpha_1}\beta^{j+k+1} y^2_{t-2-j-k}} + \left(\alpha_1 + \alpha_2\right) \sum_{k=0}^{\infty} \frac{\beta^{j+k} y^2_{t-2-j-k}}{\underline{\omega} + \underline{\alpha_1}\beta^{j+k+1} y^2_{t-2-j-k}} \\
&\leq \overline{\beta}^j \frac{\overline{\omega}}{\underline{\omega}} \sum_{k=0}^{\infty} \overline{\beta}^k + \frac{\left(\alpha_1 + \alpha_2\right)}{\beta} \sum_{k=0}^{\infty} \frac{\beta^{j+k+1} y^2_{t-2-j-k}}{\underline{\omega} + \underline{\alpha_1}\beta^{j+k+1} y^2_{t-2-j-k}} \\
&\leq \overline{\beta}^j \frac{\overline{\omega}}{\underline{\omega}} \left(1 - \overline{\beta}\right)^{-1} + \frac{\left(\alpha_1 + \alpha_2\right)}{\underline{\alpha_1}\beta} \sum_{k=0}^{\infty} \frac{\beta^{j+k+1} \underline{\alpha_1} y^2_{t-2-j-k}/\underline{\omega}}{1 + \underline{\alpha_1}\beta^{j+k+1} y^2_{t-2-j-k}/\underline{\omega}} \\
&\leq \overline{\beta}^j \frac{\overline{\omega}}{\underline{\omega}} \left(1 - \overline{\beta}\right)^{-1} + \frac{\left(\alpha_1 + \alpha_2\right) \underline{\alpha_1}^s}{\underline{\omega}^s \underline{\alpha_1} \underline{\beta}} \sum_{k=0}^{\infty} \beta^{(j+k+1)s} y^{2s}_{t-2-j-k} \\
&\leq \overline{\beta}^j \frac{\overline{\omega}}{\underline{\omega}} \left(1 - \overline{\beta}\right)^{-1} + \frac{\left(\overline{\alpha_1} + \overline{\alpha_2}\right) \underline{\alpha_1}^{s-1} \overline{\beta}^{(j+1)s}}{\underline{\omega}^s \underline{\beta}} \sum_{k=0}^{\infty} \overline{\beta}^{ks} y^{2s}_{t-2-j-k}
\end{aligned}
$$

Thus, by the help of Minkowski's Inequality and similar arguments as in (26), we obtain that $||\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j h^*_{t-1-j}/h^*_t||_4 < \infty$. As a conclusion, the first norm in Result 9 is finite, i.e.,

$$
\left\| \sup_{\lambda\in\Lambda_0} \frac{|h^*_{\lambda,t}|}{h^*_t} \right\|_4 < \infty. \tag{27}
$$

Next, we will prove the finiteness of the second norm in Result 9, which is $||\sup_{\lambda\in\Lambda_0}|h^*_{\lambda\lambda,t}|/h^*_t||_2$. As found in (7) in Result 8, the ergodic stationary second derivative of h_t evolves according to the equation

$$
\begin{aligned}
h^*_{\lambda\lambda,t} &= g^*_{\lambda\lambda,t} + e_4 h^{*\prime}_{\lambda,t-1} + h^*_{\lambda,t-1}e_4' + \beta h^*_{\lambda\lambda,t-1} \\
&= \sum_{j=0}^{\infty}\beta^j g^*_{\lambda\lambda,t-j} + \sum_{j=0}^{\infty}\beta^j e_4 h^{*\prime}_{\lambda,t-1-j} + \sum_{j=0}^{\infty}\beta^j h^*_{\lambda,t-1-j}e_4'.
\end{aligned} \tag{28}
$$

The convergence of these sums can be proven in a similar way as it is done in the case of

$h^*_{\lambda,t}$.

$$|h^*_{\lambda\lambda,t}| \leq \sum_{j=0}^{\infty} \beta^j \left|g^*_{\lambda\lambda,t-j}\right| + 2\sum_{j=0}^{\infty} \beta^j \left|h^*_{\lambda,t-j}\right|.$$

Note that $|e_4 h^{*\prime}_{\lambda,t-1-j}| = |h^*_{\lambda,t-1-j}e_4'| = |h^*_{\lambda,t-1-j}|$. This time we have to use the upper bound for $|g^*_{\lambda\lambda,t-j}|$ found in (21), i.e., $|g^*_{\lambda\lambda,t-j}| \leq C_2 y^2_{t-1-j}$. The convergence of the infinite sums follows by the finiteness of $|h^*_{\lambda,t}|$ and by Lemma 2. By using (28), the norm in Result 9 can be written as

$$\begin{aligned}\left\|\sup_{\lambda\in\Lambda_0} \frac{|h^*_{\lambda\lambda,t}|}{h^*_t}\right\|_2 &= \left\|\sup_{\lambda\in\Lambda_0} \frac{\left|\sum_{j=0}^{\infty}\beta^j g^*_{\lambda\lambda,t-j} + \sum_{j=0}^{\infty}\beta^j e_4 h^{*\prime}_{\lambda,t-1-j} + \sum_{j=0}^{\infty}\beta^j h^*_{\lambda,t-1-j}e_4'\right|}{h^*_t}\right\|_2 \\ &\leq \left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|g^*_{\lambda\lambda,t-j}\right|}{h^*_t}\right\|_2 + 2\left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|h^*_{\lambda,t-1-j}\right|}{h^*_t}\right\|_2.\end{aligned}$$

The inequality follows from Minkowski's Inequality. Therefore, $||\sup_{\lambda\in\Lambda_0}|h^*_{\lambda\lambda,t}|/h^*_t||_2$ is finite if

$$\left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|g^*_{\lambda\lambda,t-j}\right|}{h^*_t}\right\|_2 < \infty \quad \text{and} \quad \left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|h^*_{\lambda,t-1-j}\right|}{h^*_t}\right\|_2 < \infty. \tag{29}$$

Regarding the first term above,

$$\begin{aligned}\left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|g^*_{\lambda\lambda,t-j}\right|}{h^*_t}\right\|_2 &\leq \left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{C_2 y^2_{t-1-j}}{h^*_t}\right\|_2 \\ &\leq C_2\left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{y^2_{t-1-j}}{h^*_t}\right\|_4 \qquad \text{by Liapunov's Inequality,} \\ &< \infty.\end{aligned}$$

The last inequality has already been proven in (26). Regarding the other term in (29), we will follow similar steps as in previous analysis.

$$\begin{aligned}\left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|h^*_{\lambda,t-1-j}\right|}{h^*_t}\right\|_2 &\leq \left\|\sup_{\lambda\in\Lambda_0}\sum_{j=0}^{\infty}\beta^j \frac{\left|h^*_{\lambda,t-1-j}\right|}{h^*_t}\right\|_4 \\ &\leq \sum_{j=0}^{\infty}\overline{\beta}^j \left\|\sup_{\lambda\in\Lambda_0} \frac{\left|h^*_{\lambda,t-1-j}\right|}{h^*_t}\right\|_4\end{aligned}$$

by Liapunov's Inequality and Minkowski's Inequality. Choose $\gamma \in (1, \overline{\beta}^{-1})$, because the term $||\sup_{\lambda\in\Lambda_0}|h^*_{\lambda,t-1-j}|/h^*_t||_4$ is finite by (27), we have $\gamma^j\overline{\beta}^j||\sup_{\lambda\in\Lambda_0}|h^*_{\lambda,t-1-j}|/h^*_t||_4 \to 0$

a.s. Thus, Lemma 2 provides the finiteness of $\sum_{j=0}^{\infty} \overline{\beta}^j || \sup_{\lambda \in \Lambda_0} |h^*_{\lambda,t-1-j}|/h^*_t||_4$. As a conclusion, the second norm in Result 9 $|| \sup_{\lambda \in \Lambda_0} |h^*_{\lambda\lambda,t}|/h^*_t||_2$ is finite.

Result 10: Since ε_t have a density which is positive and lower semicontinuous on $\mathbb{R}$ by Assumption 1, clearly Result 10 i) holds. Now, let's verify the second condition in Result 10, which is the identification condition given as $x'_\lambda \partial g\left(y_t, \sigma_t^2; \gamma_0\right)/\partial\lambda = 0$ a.s. only if $x_\lambda = 0$. Let $x_\lambda = (x_1, x_2, \ldots, x_6) \in \mathbb{R}^6$ and suppose that $x'_\lambda \partial g\left(y_t, \sigma_t^2; \gamma_0\right)/\partial\lambda = 0$ a.s. In other words,

$$
\begin{aligned}
x'_\lambda \frac{\partial g\left(y_t, \sigma_t^2; \gamma_0\right)}{\partial \lambda} &= x'_\lambda \begin{bmatrix} 1 \\ \sigma_t^2 \varepsilon_t^2 \\ \sigma_t^2 \varepsilon_t^2 G\left(\sigma_t \varepsilon_t; \gamma_0\right) \\ \sigma_t^2 \\ \alpha_{0,2} \sigma_t^2 \varepsilon_t^2 G_\gamma\left(\sigma_t \varepsilon_t; \gamma_0\right) \end{bmatrix} \\
&= x_1 + x_2 \sigma_t^2 \varepsilon_t^2 + x_3 \sigma_t^2 \varepsilon_t^2 G\left(\sigma_t \varepsilon_t; \gamma_0\right) + x_4 \sigma_t^2 + \alpha_{0,2}\left(x_5, x_6\right)' \sigma_t^2 \varepsilon_t^2 G_\gamma\left(\sigma_t \varepsilon_t; \gamma_0\right) \\
&= 0 \qquad a.s. \qquad (30)
\end{aligned}
$$

We will follow a similar way as in the verification of Result 5. We found that for some $\underline{\sigma} > 0$ and for all $\underline{\sigma} < \sigma_* < \sigma^*$ we have

$$P\{\sigma_t^2 \in (\sigma_*, \sigma^*)\} > 0. \qquad (31)$$

Consider the events $\left\{\sigma_t^2 \in (\sigma_*, \sigma^*) \text{ and } \varepsilon_t \leq \underline{\sigma}^{-1/2} M\right\}$ with $\underline{\sigma} < \sigma_* < \sigma^*$ and $M < 0$. By the independence of σ_t and ε_t, we can separate the joint probability as

$$P\left\{\sigma_t^2 \in (\sigma_*, \sigma^*) \text{ and } \varepsilon_t \leq \underline{\sigma}^{-1/2} M\right\} = P\{\sigma_t^2 \in (\sigma_*, \sigma^*)\} P\{\varepsilon_t \leq \underline{\sigma}^{-1/2} M\}.$$

This probability is strictly positive by (31) and Assumption 1.a. Moreover, on the events $\left\{\sigma_t^2 \in (\sigma_*, \sigma^*) \text{ and } \varepsilon_t \leq \underline{\sigma}^{-1/2} M\right\}$ we have $\sigma_t \varepsilon_t \leq M$ regardless of the values of σ_* and σ^*. Thus, let σ_* and σ^* be fixed. Consider (30) as $M \to -\infty$, then, $y_t^2 = \sigma_t^2 \varepsilon_t^2 \to \infty$. By the properties of the function G, i.e. by (9), (12) and (13), the third and the fifth terms in (30) are converging to zero as $M \to -\infty$, hence they are bounded. The fourth one is also bounded since $\sigma_t^2 \in (\sigma_*, \sigma^*)$. However, the second term can take values arbitrarily large in absolute value unless $x_2 = 0$. Hence, we restrict x_2 to be zero.

By rewriting $x_3\sigma_t^2\varepsilon_t^2 G(\sigma_t\varepsilon_t;\gamma_0)$ as $x_3\sigma_t^2\varepsilon_t^2 + x_3\sigma_t^2\varepsilon_t^2[G(\sigma_t\varepsilon_t;\gamma_0)-1]$, we arrange (30) under the restriction of $x_2 = 0$ as

$$x_1 + x_3\sigma_t^2\varepsilon_t^2 + x_3\sigma_t^2\varepsilon_t^2[G(\sigma_t\varepsilon_t;\gamma_0)-1] + x_4\sigma_t^2 + \alpha_{0,2}(x_5,x_6)'\sigma_t^2\varepsilon_t^2 G_\gamma(\sigma_t\varepsilon_t;\gamma_0) = 0 \quad a.s. \tag{32}$$

Consider the events $\left\{\sigma_t^2 \in (\sigma_*,\sigma^*) \text{ and } \varepsilon_t \geq \underline{\sigma}^{-1/2}M\right\}$ with $M > 0$. As M takes arbitrarily large values, the third and the fifth terms in the above equation are bounded by the properties of the interaction between y_t^2 and the function G, i.e. (10), (12) and (13). Hence, by the same reasoning as above, x_3 should be equal to zero since $\sigma_t^2\varepsilon_t^2$ becomes unbounded as $M \to \infty$.

Under the additional restriction $x_3 = 0$, the above equation can be written as

$$x_1 + x_4\sigma_t^2 + \alpha_{0,2}(x_5,x_6)'\sigma_t^2\varepsilon_t^2 G_\gamma(\sigma_t\varepsilon_t;\gamma_0) = 0 \quad a.s.$$

Now, consider the events $\left\{\sigma_t^2 \in (\sigma_*,\sigma^*) \text{ and } \varepsilon_t \geq \underline{\sigma}^{-1/2}M\right\}$ with $M > 0$ and with different values for σ_* and σ^*. The third term in the last equation converges to zero as $M \to \infty$. However, in order to have $x_1 + x_4\sigma_t^2 = 0$ a.s., we either have to have $x_1 = x_4 = 0$ or $\sigma_t^2 = -x_1/x_4$. But, by taking σ_* and σ^* accordingly, we can exclude the possibility of σ_t^2 being equal to $-x_1/x_4$. Thus, we have $x_1 = x_4 = 0$.

Since $\alpha_{0,2} > 0$ by Assumption 2, we are left with the equation $(x_5,x_6)' G_\gamma(\sigma_t\varepsilon_t;\gamma_0)\sigma_t^2\varepsilon_t^2 = 0$ *a.s.* Note that, in (14) we found that $(x_5,x_6)' G_\gamma(\sigma_t\varepsilon_t;\gamma_0)$ is zero if and only if $\sigma_t\varepsilon_t = x_5\gamma_{2,0}/x_6 - \gamma_{1,0}$ for nonzero (x_5,x_6). Hence, again by taking σ_* and σ^* properly, we reach the conclusion that $(x_5,x_6) = (0,0)$.

As a result, x_λ should be zero if $x'_\lambda \partial g(y_t,\sigma_t^2;\gamma_0)/\partial\lambda = 0$ a.s., thus, the identification condition in Result 10 holds.

Appendix E: The Proofs of the Main Theorems

The Proof of Theorem 1

For strong consistency of $\hat{\lambda}_T$, it suffices to show that, for all $\delta > 0$

$$\liminf_{T\to\infty} \inf_{\lambda\in B(\lambda_0,\delta)^c} (L_T(\lambda) - L_T(\lambda_0)) > 0 \ \ a.s.,$$

where $L_T(\lambda) = T^{-1}\sum_{t=1}^{T} l_t(\lambda) = T^{-1}\sum_{t=1}^{T} \log h_t + y_t^2/h_t$, and $B(\lambda_0, \delta) = \{\lambda \in \Lambda \mid |\lambda_0 - \lambda| < \delta\}$ is the open ball with the center λ_0 and radius δ. Moreover, $B(\lambda_0, \delta)^c$ is the complement of this set in Λ, which will be denoted as B^c (see Pötscher and Prucha (1991, p.145))[6]. The interpretation of this inequality is as follows. Even in the best case, i.e., even if the infimum of the distance is taken, the likelihood function $L_T(\lambda)$ is greater than the likelihood function with the true parameter, $L_T(\lambda_0)$, outside all open balls centered at λ_0 with radii δ, uniformly for large T. Hence, equivalently, the inequality implies that λ_0 is the unique minimizer of $L_T(\lambda)$ in Λ.

Note that,

$$\begin{aligned}
\liminf_{T\to\infty} \inf_{\lambda \in B^c} (L_T(\lambda) - L_T(\lambda_0)) &= \liminf_{T\to\infty} \inf_{\lambda \in B^c} (L_T(\lambda) - L_T(\lambda_0) \pm L_T^*(\lambda) \pm L_T^*(\lambda_0) \pm E[l_t^*(\lambda_0)]) \\
&\geq \liminf_{T\to\infty} \inf_{\lambda \in B^c} \{(L_T(\lambda) - L_T(\lambda_0)) - (L_T^*(\lambda) - L_T^*(\lambda_0))\} \quad (33) \\
&\quad + \liminf_{T\to\infty} \inf_{\lambda \in B^c} \{E[l_t^*(\lambda_0)] - L_T^*(\lambda_0)\} \quad (34) \\
&\quad + \liminf_{T\to\infty} \inf_{\lambda \in B^c} \{L_T^*(\lambda) - E[l_t^*(\lambda_0)]\}
\end{aligned}$$

Let's analyze the first expression on the right hand side, which is denoted by (33).

$$\begin{aligned}
&\liminf_{T\to\infty} \inf_{\lambda \in B^c} \{(L_T(\lambda) - L_T(\lambda_0)) - (L_T^*(\lambda) - L_T^*(\lambda_0))\} \\
= &-\limsup_{T\to\infty} -\inf_{\lambda \in B^c} \{(L_T(\lambda) - L_T(\lambda_0)) - (L_T^*(\lambda) - L_T^*(\lambda_0))\}^{7} \\
= &-\limsup_{T\to\infty} \sup_{\lambda \in B^c} -\{(L_T(\lambda) - L_T(\lambda_0)) - (L_T^*(\lambda) - L_T^*(\lambda_0))\} \\
\geq &-\limsup_{T\to\infty} \sup_{\lambda \in B^c} |(L_T^*(\lambda) - L_T^*(\lambda_0)) - (L_T(\lambda) - L_T(\lambda_0))| \\
\geq &-\limsup_{T\to\infty} \sup_{\lambda \in \Lambda} |(L_T^*(\lambda) - L_T^*(\lambda_0)) - (L_T(\lambda) - L_T(\lambda_0))| \text{ since } B^c \subseteq \Lambda.
\end{aligned}$$

If we consider the term denoted by (34), we see that $E[l_t^*(\lambda_0)] - L_T^*(\lambda_0)$ does not depend on λ. Thus,

$$\liminf_{T\to\infty} \inf_{\lambda \in B^c} \{E[l_t^*(\lambda_0)] - L_T^*(\lambda_0)\} = \liminf_{T\to\infty} \{E[l_t^*(\lambda_0)] - L_T^*(\lambda_0)\}.$$

[6] A basic sufficient condition for $\rho_B(\hat{\beta}_n, \bar{\beta}_n) \to 0$ *a.s.* as $n \to \infty$ is that for each $\varepsilon > 0$, $\liminf_{n\to\infty} \inf_{\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon} \left[R_n(\beta) - R_n(\bar{\beta}_n)\right] > 0$ *a.s.*

To familiarize the notation: R_n is the likelihood function, $\hat{\beta}_n$ is the likelihood estimator, $\bar{\beta}_n$ is the minimizer of R_n, B is the space of parameters of interest, ρ_B is the metric defined on B, and $\beta \in B$. In the proof we simply replace $\bar{\beta}_n$ with the true parameter.

[7] Let $\{a_n\}$ be a sequence in [-∞, ∞], then $\liminf a_n = -\limsup(-a_n)$.

Hence, we obtain

$$\begin{aligned}
& \liminf_{T\to\infty} \inf_{\lambda\in B^c} (L_T(\lambda) - L_T(\lambda_0)) \\
\geq\ & -\limsup_{T\to\infty} \sup_{\lambda\in\Lambda} |(L_T^*(\lambda) - L_T^*(\lambda_0)) - (L_T(\lambda) - L_T(\lambda_0))| \qquad (35) \\
& +\liminf_{T\to\infty} \{E[l_t^*(\lambda_0)] - L_T^*(\lambda_0)\} \qquad (36) \\
& +\liminf_{T\to\infty} \inf_{\lambda\in B^c} \{L_T^*(\lambda) - E\left[l_t^*(\lambda_0)\right]\} \qquad (37)
\end{aligned}$$

Throughout the proof, we shall show that the first two terms, that is (35) and (36), are equal to zero a.s. whereas the third term (37) is strictly positive. We will proceed term by term.

First, we will show that the first term is equal to zero a.s. To deduce it, we begin by showing that $\sup_{\lambda\in\Lambda} |L_T^*(\lambda) - L_T(\lambda)| \to 0$ a.s. as $T \to \infty$. Notice that,

$$\begin{aligned}
|l_t^*(\lambda) - l_t(\lambda)| &= \left|(\log h_t^* + \frac{y_t^2}{h_t^*}) - (\log h_t + \frac{y_t^2}{h_t})\right| \\
&\leq |\log h_t^* - \log h_t| + y_t^2 \left|\frac{1}{h_t^*} - \frac{1}{h_t}\right| \\
&\leq \left|\frac{1}{\overline{h}_t}(h_t^* - h_t)\right| + y_t^2 \left|\left(-\frac{1}{\hat{h}_t^2}\right)(h_t^* - h_t)\right| \\
&= \frac{1}{\overline{h}_t} |h_t^* - h_t| + \frac{y_t^2}{\hat{h}_t^2} |h_t^* - h_t| \\
&\leq \frac{1}{\underline{g}} |h_t^* - h_t| + \frac{y_t^2}{\underline{g}^2} |h_t^* - h_t| \\
&\leq m(1 + y_t^2) |h_t^* - h_t| \qquad \text{for some } m \geq \max\{\underline{g}^{-1}, \underline{g}^{-2}\}.
\end{aligned}$$

The second inequality is satisfied by Mean Value Theorem[8] for some $\hat{h}_t$ and $\overline{h}_t$ lying between h_t and h_t^*. The third inequality is satisfied because both $h_t^* \geq \underline{g}$ and $h_t \geq \underline{g}$ by Result 4. Hence, we have

$$\begin{aligned}
\left\|\sup_{\lambda\in\Lambda} |l_t^*(\lambda) - l_t(\lambda)|\right\|_{r/2} &\leq \left\|\sup_{\lambda\in\Lambda}\{m(1 + y_t^2) |h_t^* - h_t|\}\right\|_{r/2} \\
&= \left\|m(1 + y_t^2)\sup_{\lambda\in\Lambda} |h_t^* - h_t|\right\|_{r/2}
\end{aligned}$$

[8]Mean Value Theorem: If f is a real continuous function on $[a, b]$ which is differentiable in (a, b), then there is a point $x \in (a, b)$ at which

$$f(b) - f(a) = (b - a)\frac{df(x)}{dx}.$$

$$
\begin{aligned}
&\leq \left\|m(1+y_t^2)\right\|_r \left\|\sup_{\lambda\in\Lambda}|h_t^* - h_t|\right\|_r \\
&\leq \triangle_{r,2} m\left(1+\left\|y_t^2\right\|_r\right)\left\|\sup_{\lambda\in\Lambda}|h_t^* - h_t|\right\|_r \quad \text{by Lemma 1.}
\end{aligned}
$$

The second inequality is justified by Cauchy-Schwarz Inequality[9]. The finiteness of the last term is given by the proof of Result 1, i.e., $E[y_t^{2r}] < \infty$ implies $\|y_t^2\|_r = M < \infty$. Result 3 gives that $\|\sup_{\lambda\in\Lambda}|h_t^* - h_t|\|_r \leq C\kappa^t$, where $C \in \mathbb{R}$ and $\kappa \in (0,1)$. Let $1 < \gamma < 1/\kappa$ and $\triangle_{r,2} m(1+\|y_t^2\|_r) = m'$. Then,

$$
\left\|\gamma^t \sup_{\lambda\in\Lambda}|l_t^*(\lambda) - l_t(\lambda)|\right\|_{r/2} \leq m'C\frac{\gamma^t}{\kappa^t} \to 0 \quad a.s. \quad \text{as } t \to \infty.
$$

Thus, by Lemma 2, $\sum_{t=1}^{\infty}\sup_{\lambda\in\Lambda}|l_t^*(\lambda) - l_t(\lambda)| < \infty$ a.s. We can write that

$$
\begin{aligned}
\sup_{\lambda\in\Lambda}|(L_T^*(\lambda) - L_T(\lambda)| &= \sup_{\lambda\in\Lambda}\left|\frac{1}{T}\sum_{t=1}^{T} l_t^*(\lambda) - \frac{1}{T}\sum_{t=1}^{T} l_t(\lambda)\right| \\
&\leq \frac{1}{T}\sup_{\lambda\in\Lambda}\sum_{t=1}^{T}|l_t^*(\lambda) - l_t(\lambda)| \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\lambda\in\Lambda}|l_t^*(\lambda) - l_t(\lambda)|
\end{aligned}
$$

The last term converges a.s. as $T \to \infty$, since $\sum_{t=1}^{T}\sup_{\lambda\in\Lambda}|l_t^*(\lambda) - l_t(\lambda)|$ is finite as $T \to \infty$. We showed that $\sup_{\lambda\in\Lambda}|(L_T^*(\lambda) - L_T(\lambda)| \to 0$ a.s. as $T \to \infty$. In particular, since $\lambda_0 \in \Lambda$, we have $|(L_T^*(\lambda_0) - L_T(\lambda_0)| \to 0$ a.s. as $T \to \infty$. Hence,

$$
\begin{aligned}
&\limsup_{T\to\infty}\ \sup_{\lambda\in\Lambda}|(L_T^*(\lambda) - L_T^*(\lambda_0)) - (L_T(\lambda) - L_T(\lambda_0))| \\
\leq\ &\limsup_{T\to\infty}\ \sup_{\lambda\in\Lambda}\{|L_T^*(\lambda) - L_T(\lambda)| + |L_T^*(\lambda_0) - L_T(\lambda_0)|\} \\
=\ &\limsup_{T\to\infty}\left\{|L_T^*(\lambda_0) - L_T(\lambda_0)| + \sup_{\lambda\in\Lambda}|L_T^*(\lambda) - L_T(\lambda)|\right\} \\
\leq\ &\limsup_{T\to\infty}|L_T^*(\lambda_0) - L_T(\lambda_0)| + \limsup_{T\to\infty}\ \sup_{\lambda\in\Lambda}|L_T^*(\lambda) - L_T(\lambda)| \\
=\ &0 \quad a.s.
\end{aligned}
$$

Thus, the term (35) equal to zero a.s.

To handle with the remaining two terms (36) and (37), first consider the term $L_T^*(\lambda)$, thus the term $l_t^*(\lambda)$. By Result 1 and Result 3, h_t^* and y_t^2 are ergodic and stationary

[9] Cauchy-Schwarz Inequality: For random variables X and Y, $(E[XY])^2 \leq E[X^2]E[Y^2]$ with equality attained when $Y = cX$, c a constant.

implying that $l_t^*(\lambda) = \log h_t^* + y_t^2/h_t^*$ is ergodic and stationary. Because $h_t^* \geq \underline{g}$ and y_t^2 is independent of λ, $l_t^*(\lambda)$ is bounded below uniformly in Λ. Therefore, the expectation of negative part of $l_t^*(\lambda)$ is not infinite. Hence, $E[l_t^*(\lambda)]$ is well-defined and belongs to $\mathbb{R} \cup \{\infty\}$. In particular, $E[\inf_{\lambda\in\Lambda} l_t^*(\lambda)] < \infty$. By Result 3, $E[\sup_{\lambda\in\Lambda} h_t^{*r}(\lambda)] < \infty$ for some $r > 0$. Therefore, we can write that

$$\begin{aligned}
\infty &> E\left[\sup_{\lambda\in\Lambda} h_t^{*r}(\lambda)\right] \\
&> \log E\left[\sup_{\lambda\in\Lambda} h_t^{*r}(\lambda)\right] \\
&\geq E\left[\log\left\{\sup_{\lambda\in\Lambda} h_t^{*r}(\lambda)\right\}\right] \\
&= E\left[\log\left\{\max_{\lambda\in\Lambda} h_t^{*r}(\lambda)\right\}\right] \quad \text{since } \mathit{\Lambda} \text{ is a compact set,} \\
&= E\left[\max_{\lambda\in\Lambda}\{\log h_t^{*r}(\lambda)\}\right] \\
&= rE\left[\max_{\lambda\in\Lambda}\{\log h_t^{*}(\lambda)\}\right] \\
&= rE\left[\sup_{\lambda\in\Lambda}\{\log h_t^{*}(\lambda)\}\right] \quad \text{since } \mathit{\Lambda} \text{ is a compact set.}
\end{aligned}$$

The second inequality is satisfied since $x > x-1 \geq \log x$ for all $x \in \mathbb{R}_+$. The tangent line of the curve $\log x$ passing through the point $x = 1$ is the line $y = x - 1$. Hence, the curve $\log x$ is under this tangent line for all x [10]. The third inequality is justified by Jensen 's Inequality[11] since log is a concave function. The first equality is justified because log is a strictly increasing transformation, thus, the maximizer of $h_t^{*r}(\lambda)$ and the maximizer of

[10] The function $y = \log u$ is a concave function. The tangent line at the point $u = 1$ is $y = -1 + u$. Because a concave function plots below any tangent line, the inequality $\log u \leq -1 + u$ is obtained.

[11] Jensen's Inequality: If a Borel function ϕ is convex on an interval I containing the support of an integrable random variable X, where $\phi(X)$ is also integrable, then

$$\phi(E(X)) \leq E(\phi(X)).$$

For a concave function the reverse inequality holds.

$\log h_t^{*r}(\lambda)$ is the same. Hence $E\left[\sup_{\lambda\in\Lambda}\log h_t^*(\lambda)\right]<\infty$. Notice that,

$$\begin{aligned} E\left[l_t^*(\lambda_0)\right] &= E\left[\log h_t^*(\lambda_0)+\frac{y_t^2}{h_t^*(\lambda_0)}\right] \\ &= E\left[\log\sigma_t^2+\frac{\sigma_t^2\varepsilon_t^2}{\sigma_t^2}\right] \quad \text{since } h_t^*(\lambda_0)=\sigma_t^2 \text{ by Result 3,} \\ &= E\left[\log\sigma_t^2\right]+1 \quad\quad \text{since } E[\varepsilon_t^2]=1. \end{aligned}$$

Since log is a concave function, we have, $rE[\log\sigma_t^2]=E[\log\sigma_t^{2r}]\le\log E[\sigma_t^{2r}]<E[\sigma_t^{2r}]<\infty$ by Jensen's inequality and by Result 1. Hence, we have $E\left[l_t^*(\lambda_0)\right]=E[\log\sigma_t^2]+1<\infty$. However, for $\lambda\neq\lambda_0$, we may have $E[y_t^2/h_t^*(\lambda)]=\infty$. Thus, we may not bound $E\left[l_t^*(\lambda)\right]$ from above as well.

Part 1: Note that if Assumption 3 is assumed as an extra assumption, then the proof is simplified. We, now, can bound $E\left[l_t^*(\lambda)\right]$ from above, actually uniformly, because by Assumption 3 we have $E[\sup_{\lambda\in\Lambda}\sigma_t^2/h_t^*]<\infty$. We showed above that $E\left[\sup_{\lambda\in\Lambda}\log h_t^*(\lambda)\right]<\infty$. Thus,

$$E\left[\sup_{\lambda\in\Lambda} l_t^*(\lambda)\right]=E\left[\sup_{\lambda\in\Lambda}\left\{\log h_t^*(\lambda)+\frac{\sigma_t^2}{h_t^*}\right\}\right]\le E\left[\sup_{\lambda\in\Lambda}\log h_t^*(\lambda)\right]+E\left[\sup_{\lambda\in\Lambda}\frac{\sigma_t^2}{h_t^*}\right]<\infty$$

Since $l_t^*(\lambda)$ is ergodic and stationary with $E\left[\sup_{\lambda\in\Lambda} l_t^*(\lambda)\right]<\infty$, a Uniform Strong Law of Large Numbers[12] provides $\sup_{\lambda\in\Lambda}\left|L_T^*(\lambda)-E\left[l_t^*(\lambda)\right]\right|\to 0$ a.s. as $T\to\infty$. In particular, $\left|L_T^*(\lambda_0)-E\left[l_t^*(\lambda_0)\right]\right|\to 0$ a.s. as $T\to\infty$. Therefore, the second term (36), $\liminf_{T\to\infty}\{E\left[l_t^*(\lambda_0)\right]-L_T^*(\lambda_0)\}$, is equal to zero.

Next, we will prove that $E\left[l_t^*(\lambda)\right]$ is uniquely minimized at $\lambda=\lambda_0$. Equivalently, we will show that $E\left[l_t^*(\lambda)\right]-E\left[l_t^*(\lambda_0)\right]>0$ for all $\lambda\in\Lambda\backslash\{\lambda_0\}$.

$$\begin{aligned} E\left[l_t^*(\lambda)\right]-E\left[l_t^*(\lambda_0)\right] &= E\left[\log h_t^*+\frac{y_t^2}{h_t^*}\right]-E\left[\log h_t^*\left(\lambda_0\right)+\frac{y_t^2}{h_t^*\left(\lambda_0\right)}\right] \\ &= E\left[\log h_t^*+\frac{\sigma_t^2\varepsilon_t^2}{h_t^*}\right]-E\left[\log\sigma_t^2+\frac{\sigma_t^2\varepsilon_t^2}{\sigma_t^2}\right] \\ &= E\left[\log h_t^*-\log\sigma_t^2\right]+E\left[\frac{\sigma_t^2\varepsilon_t^2}{h_t^*}\right]-1 \\ &= E\left[-\log\frac{\sigma_t^2}{h_t^*}+\frac{\sigma_t^2}{h_t^*}\right]-1 \quad \text{since } \varepsilon_t \text{ is independent of } \sigma_t^2 \text{ and } h_t^*. \end{aligned}$$

[12] Let (v_t) be a stationary ergodic sequence random elements with values in the space of continuous $\mathbb{R}^d$-valued functions equipped with the sup norm $\|v\|_K=\sup_{s\in K}|v(s)|$. Then the Uniform Strong Law of Large Numbers (SLLN) is implied by $E\|v_0\|_K<\infty$. The uniform SLLN is said to hold if $\|1/n\sum_{t=1}^n v_t-E(v_0)\|_K\to 0$ *a.s.*

By using the inequality $x - \log x \geq 1$ for all $x \in \mathbb{R}_+$ with equality if and only if $x = 1$, and by using the identification in Result 5, i.e., $\sigma_t^2/h_t^* \neq 1$ if $\lambda \neq \lambda_0$, we conclude that $E\left[l_t^*(\lambda)\right] - E\left[l_t^*(\lambda_0)\right] > 0$ for $\lambda \neq \lambda_0$. Let's consider the third term (37).

$$
\begin{aligned}
& \liminf_{T\to\infty} \inf_{\lambda\in B^c} \left\{L_T^*(\lambda) - E\left[l_t^*(\lambda_0)\right]\right\} \\
= \; & \liminf_{T\to\infty} \inf_{\lambda\in B^c} \left\{L_T^*(\lambda) - E\left[l_t^*(\lambda_0)\right] + E\left[l_t^*(\lambda)\right] - E\left[l_t^*(\lambda)\right]\right\} \\
\geq \; & \liminf_{T\to\infty} \left\{ \inf_{\lambda\in B^c} \left(L_T^*(\lambda) - E\left[l_t^*(\lambda)\right]\right) + \inf_{\lambda\in B^c} \left(E\left[l_t^*(\lambda)\right] - E\left[l_t^*(\lambda_0)\right]\right) \right\} \\
\geq \; & \left\{ \liminf_{T\to\infty} \inf_{\lambda\in B^c} \left(L_T^*(\lambda) - E\left[l_t^*(\lambda)\right]\right) \right\} + \left\{ \liminf_{T\to\infty} \inf_{\lambda\in B^c} \left(E\left[l_t^*(\lambda)\right] - E\left[l_t^*(\lambda_0)\right]\right) \right\} \\
= \; & \left\{ \liminf_{T\to\infty} \inf_{\lambda\in B^c} \left(L_T^*(\lambda) - E\left[l_t^*(\lambda)\right]\right) \right\} + \left\{ \left(\min_{\lambda\in B^c} E\left[l_t^*(\lambda)\right] \right) - E\left[l_t^*(\lambda_0)\right] \right\} \qquad (38)
\end{aligned}
$$

The last equality is satisfied because $E\left[l_t^*(\lambda_0)\right]$ does not depend on λ and T, B^c is compact being a closed set of a compact set Λ, and $E\left[l_t^*(\lambda)\right]$ does not depend on T. As a consequence, the result is positive because the first term in (38) is zero but the last term is positive by the analysis above. Thus, $\liminf_{T\to\infty} \inf_{\lambda\in B^c}\{L_T^*(\lambda) - E\left[l_t^*(\lambda_0)\right]\} > 0$ which completes the proof.

In this part, the uniform SLLN, via the help of Assumption 3, facilitated the proof by letting us to conclude that the first term in the last equality is equal to zero. But, without Assumption 3 we cannot bound $E\left[l_t^*(\lambda)\right]$ from above, thus we cannot use uniform SLLN to obtain a nonnegative number from the first term in (38). Next, we move one the second part of the proof where we use different arguments other then SLLN.

Part 2: Let's continue the proof without having Assumption 3 as an extra assumption. Since $l_t^*(\lambda_0)$ is ergodic and stationary with $E\left[l_t^*(\lambda_0)\right] < \infty$, we can use Ergodic Theorem[13] and conclude that $L_T^*(\lambda_0) = 1/T\sum_{t=1}^{T} l_t^*(\lambda_0) \to E\left[l_t^*(\lambda_0)\right]$ a.s. Thus, the second term (36), $\liminf_{T\to\infty}\{E\left[l_t^*(\lambda_0)\right] - L_T^*(\lambda_0)\}$, is equal to zero. Now, consider the third term (37), $\liminf_{T\to\infty} \; \inf_{\lambda\in B^c}\{L_T^*(\lambda) - E[l_t^*(\lambda_0)]\}$. By Lemma 3.11 of Pfanzagl (1969)[14],

[13]Ergodic Theorem: Let $\{\boldsymbol{z}_i\}$ be a stationary and ergodic process with $E(\boldsymbol{z}_i) = \boldsymbol{\mu}$. Then

$$
\bar{\boldsymbol{z}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \to \boldsymbol{\mu} \;\; a.s.
$$

[14]Let $(T, \mathcal{U})$ be a σ-compact metrizable space and $f_t : X \to [-\infty, +\infty]$, $t \in T$, a family of $\mathcal{A}$-measurable

we obtain that $E\left[l_t^*(\lambda)\right]$ is a lower semi–continuous function[15] on Λ and

$$\liminf_{T\to\infty}\inf_{\lambda\in B^c} L_T^*(\lambda) \geq \inf_{\lambda\in B^c} E\left[l_t^*(\lambda)\right] \;\; a.s. \tag{39}$$

If we can show that $E\left[l_t^*(\lambda)\right] - E\left[l_t^*(\lambda_0)\right] \geq 0$ with equality if and only if $\lambda = \lambda_0$, then the positivity of the third term (37) will be proven because

$$\begin{aligned}\liminf_{T\to\infty}\inf_{\lambda\in B^c}\left\{L_T^*(\lambda) - E\left[l_t^*(\lambda_0)\right]\right\} &= \left[\liminf_{T\to\infty}\inf_{\lambda\in B^c} L_T^*(\lambda)\right] - E\left[l_t^*(\lambda_0)\right] \\ &\geq \left\{\inf_{\lambda\in B^c} E\left[l_t^*(\lambda)\right]\right\} - E\left[l_t^*(\lambda_0)\right]\end{aligned}$$

will be strictly greater than zero by the lower semicontinuity of $E\left[l_t^*(\lambda)\right]$. The inequality is satisfied by (39). We know that $E\left[l_t^*(\lambda_0)\right] < \infty$. Thus, if $E\left[l_t^*(\lambda)\right] = \infty$ then we are done. Therefore, we assume that $E\left[l_t^*(\lambda)\right] < \infty$. The rest follows as it was shown in the part with the extra assumption, i.e. Part 1.

As a conclusion, we showed that the following inequality holds

$$\liminf_{T\to\infty}\inf_{\lambda\in B(\lambda_0,\delta)^c}\left(L_T(\lambda) - L_T(\lambda_0)\right) > 0 \;\; a.s.$$

by showing (35) and (36) are zero a.s., and (37) is strictly positive a.s. Therefore, by following Pötscher and Prucha (1991)'s argument we conclude that the QML estimator $\hat{\lambda}_T$ is strongly consistent for λ_0, that is, $\hat{\lambda}_T \to \lambda_0$ a.s.

---

function such that: (1) $t \to f_t(x)$ is l.s.c., (2) inf $f_C \in \mathcal{A}$ for any compact set $C \subset T$. Let furthermore $P|\mathcal{A}$ is a p-measure such that (3) $P[\inf\ f_T] > -\infty$. Then, (a) $t \to P[f_t]$ is l.s.c., (b) $\inf_{t\in C}\ P[f_t] \leq \underline{\lim}_{n\to\infty}\inf_{t\in C} 1/n\sum_{i=1}^n f_t(x_i)\ \ P^{\mathbb{N}} - a.e.$ for any compact set C.

l.s.c. stands for lower semicontinuous and $\underline{\lim}$ is a notation for lim inf. In our proof we take $X = \mathbb{R}$, $\mathcal{A}$ as the Borel Algebra, P as the expectation, $T = \Lambda$. Thus, $t = \lambda$, $x_i = t$, and $f_t(x_i) = l_t^*(\lambda)$. (1) is satisfied because is $l_t^*(\lambda)$ continuous, (2) and (3) are satisfied because $l_t^*(\lambda)$ is bounded below uniformly in Λ. Therefore we obtain the results (a) $E\left[l_t^*(\lambda)\right]$ is l.s.c. and (b) for $C = B^c$ (since B^c is a compact set) $\inf_{\lambda\in B^c} E\left[l_t^*(\lambda)\right] \ \leq \liminf_{T\to\infty}\inf_{\lambda\in B^c} 1/T\sum_{t=1}^T {}_t l_t^*(\lambda) = \liminf_{T\to\infty}\inf_{\lambda\in B^c} L_T^*(\lambda)$

[15] Let f be a real (or extended-real) function on a topological space. If $\{x : f(x) > \alpha\}$ is open for every real α, then, f is said to be lower semicontinuous.

The Proof of Theorem 2

We denote the first and the second derivatives of $l_t(\lambda)$ as $l_{\lambda,t}(\lambda)$ and $l_{\lambda\lambda,t}(\lambda)$. Thus, those of $L_T(\lambda)$ is given as

$$\begin{aligned} L_{\lambda,T}(\lambda) &= \frac{\partial L_T(\lambda)}{\partial \lambda} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\lambda)}{\partial \lambda} = \frac{1}{T}\sum_{t=1}^{T} l_{\lambda,t}(\lambda), \\ L_{\lambda\lambda,T}(\lambda) &= \frac{\partial^2 L_T(\lambda)}{\partial \lambda \partial \lambda'} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t(\lambda)}{\partial \lambda \partial \lambda'} = \frac{1}{T}\sum_{t=1}^{T} l_{\lambda\lambda,t}(\lambda). \end{aligned}$$

Similar notation is used for the ergodic stationary counterparts, i.e., the first and the second derivatives are denoted as $l^*_{\lambda,t}(\lambda)$, $l^*_{\lambda\lambda,t}(\lambda)$, and $L^*_{\lambda,T}(\lambda)$, $L^*_{\lambda\lambda,T}(\lambda)$. In order to prove Theorem 2, we need some intermediate steps, named as Lemma D1-D6.

Lemma D1. *Under the assumptions of Theorem 2,*

$$\sqrt{T} L^*_{\lambda,T}(\lambda_0) \to_d N(0, \mathcal{I}(\lambda_0)),$$

where $\mathcal{I}(\lambda_0) = E[l^*_{\lambda,t}(\lambda_0) l^{*\prime}_{\lambda,t}(\lambda_0)] = E[\varepsilon_t^4 - 1] E[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0) h^{*\prime}_{\lambda,t}(\lambda_0)]$ *is finite.*

Proof. Taking the first derivative of $l^*_t(\lambda) = \log h^*_t + y_t^2/h^*_t$ yields

$$l^*_{\lambda,t} = -\frac{h^*_{\lambda,t}}{h^*_t}\left(\frac{y_t^2}{h^*_t} - 1\right).$$

Evaluating it at λ_0 yields $l^*_{\lambda,t}(\lambda_0) = -\sigma_t^{-2} h^*_{\lambda,t}(\lambda_0)(\varepsilon_t^2 - 1)$. Hence,

$$\begin{aligned} E\left[l^*_{\lambda,t}(\lambda_0) l^{*\prime}_{\lambda,t}(\lambda_0)\right] &= E\left[\left(\varepsilon_t^2 - 1\right)^2 \sigma_t^{-4} h^*_{\lambda,t}(\lambda_0) h^{*\prime}_{\lambda,t}(\lambda_0)\right] \\ &= E\left[\varepsilon_t^4 - 2\varepsilon_t^2 + 1\right] E\left[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0) h^{*\prime}_{\lambda,t}(\lambda_0)\right] \quad \text{by the independence of } \varepsilon_t, \\ &= E\left[\varepsilon_t^4 - 1\right] E\left[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0) h^{*\prime}_{\lambda,t}(\lambda_0)\right] \quad \text{since } E[\varepsilon_t^2] = 1. \end{aligned}$$

An implication of Assumption 5 is the finiteness of $E[\varepsilon_t^4]$. By Result 9 and Liapunov's Inequality, we have $\|\sup_{\lambda \in \Lambda_0} |h^*_{\lambda,t}(\lambda)|/h^*_t(\lambda)\|_2 < \infty$. In particular, $E[|h^*_{\lambda,t}(\lambda_0)|^2/\sigma_t^4] < \infty$. Thus, $E[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0) h^{*\prime}_{\lambda,t}(\lambda_0)]$ is finite which implies that $\mathcal{I}(\lambda_0) = E[l^*_{\lambda,t}(\lambda_0) l^{*\prime}_{\lambda,t}(\lambda_0)]$ is finite.

The ergodic stationarity of y_t and σ_t^2, by Result 1, imply the ergodic stationarity of ε_t^2. Furthermore, $h^*_{\lambda,t}(\lambda_0)$ is ergodic stationary by Result 7. Thus, $l^*_{\lambda,t}(\lambda_0) =$

$-\sigma_t^{-2} h^*_{\lambda,t}(\lambda_0)(\varepsilon_t^2 - 1)$ is ergodic and stationary. In addition, $l^*_{\lambda,t}(\lambda_0)$ is a martingale difference sequence.

$$\begin{aligned} E\left[l^*_{\lambda,t}(\lambda_0)\,|l^*_{\lambda,t-1}(\lambda_0),\ldots,l^*_{\lambda,1}(\lambda_0)\right] &= E\left[\sigma_t^{-2} h^*_{\lambda,t}(\lambda_0)\left(1-\varepsilon_t^2\right)|l^*_{\lambda,t-1}(\lambda_0),\ldots,l^*_{\lambda,1}(\lambda_0)\right] \\ &= E\left[1-\varepsilon_t^2\right] E\left[\sigma_t^{-2} h^*_{\lambda,t}(\lambda_0)\,|l^*_{\lambda,t-1}(\lambda_0),\ldots,l^*_{\lambda,1}(\lambda_0)\right] \\ &= 0. \end{aligned}$$

The second equality follows from the independence of ε_t. The last one follows from the finiteness of σ_t^{-2} and $h^*_{\lambda,t}(\lambda_0)$, and from the fact that $E[\varepsilon_t^2] = 1$. Hence, $l^*_{\lambda,t}(\lambda_0)$ being an ergodic stationary martingale difference sequence with finite $\mathcal{I}(\lambda_0) = E[l^*_{\lambda,t}(\lambda_0) l^{*\prime}_{\lambda,t}(\lambda_0)]$, we can use Billingsley's Central Limit Theorem to conclude

$$\sqrt{T} L^*_{\lambda,T}(\lambda_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} l^*_{\lambda,t}(\lambda_0) \to_d N(0, \mathcal{I}(\lambda_0)).$$

■

Lemma D2. *Under the assumptions of Theorem 2, $l^*_{\lambda\lambda,t}(\lambda)$ is L_1-dominated in Λ_0 and $\sup_{\lambda\in\Lambda_0} |L^*_{\lambda\lambda,T}(\lambda) - \mathcal{J}(\lambda)| \to 0$ a.s. where $\mathcal{J}(\lambda) = E[l^*_{\lambda\lambda,t}(\lambda)]$ is continuous at λ_0, and $\mathcal{J}(\lambda_0) = E[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0)\, h^{*\prime}_{\lambda,t}(\lambda_0)]$.*

Proof. By taking the second derivative of $l^*_t(\lambda)$, we have

$$l^*_{\lambda\lambda,t} = -\frac{h^*_{\lambda\lambda,t}}{h^*_t}\left(\frac{y_t^2}{h^*_t} - 1\right) + \frac{h^*_{\lambda,t}}{h^*_t}\frac{h^{*\prime}_{\lambda,t}}{h^*_t}\left(2\frac{y_t^2}{h^*_t} - 1\right).$$

A sufficient condition for $l^*_{\lambda\lambda,t}(\lambda)$ being L_1-dominated in Λ_0 is the finiteness of $E[\sup_{\lambda\in\Lambda_0} |l^*_{\lambda\lambda,t}(\lambda)|]$. Via the help of Hölder's Inequality[16] we obtain

$$\begin{aligned} E\left[\sup_{\lambda\in\Lambda_0}\left|l^*_{\lambda\lambda,t}(\lambda)\right|\right] &= \left\|\sup_{\lambda\in\Lambda_0}\left|l^*_{\lambda\lambda,t}(\lambda)\right|\right\|_1 \\ &= \left\|\sup_{\lambda\in\Lambda_0}\left|-\frac{h^*_{\lambda\lambda,t}}{h^*_t}\left(\frac{y_t^2}{h^*_t} - 1\right) + \frac{h^*_{\lambda,t}}{h^*_t}\frac{h^{*\prime}_{\lambda,t}}{h^*_t}\left(2\frac{y_t^2}{h^*_t} - 1\right)\right|\right\|_1 \\ &\leq \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h^*_{\lambda\lambda,t}}{h^*_t}\left(\frac{y_t^2}{h^*_t} - 1\right)\right|\right\|_1 + \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h^*_{\lambda,t}}{h^*_t}\frac{h^{*\prime}_{\lambda,t}}{h^*_t}\left(2\frac{y_t^2}{h^*_t} - 1\right)\right|\right\|_1 \end{aligned}$$

[16] For any $p \geq 1$,

$$E[|XY|] \leq \|X\|_p \|Y\|_q,$$

where $q = p/(p-1)$ if $p > 1$, and $q = \infty$ if $p = 1$.

$$
\begin{aligned}
&\leq \left\| \sup_{\lambda\in\Lambda_0} \left| \frac{h^*_{\lambda\lambda,t}}{h^*_t} \right| \right\|_2 \left\| \sup_{\lambda\in\Lambda_0} \left| \frac{y_t^2}{h^*_t} - 1 \right| \right\|_2 + \left\| \sup_{\lambda\in\Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} \right| \right\|_4 \left\| \sup_{\lambda\in\Lambda_0} \left| \frac{h^{*\prime}_{\lambda,t}}{h^*_t} \right| \right\|_4 \left\| \sup_{\lambda\in\Lambda_0} \left| 2\frac{y_t^2}{h^*_t} - 1 \right| \right\|_2 \\
&\leq \left\| \sup_{\lambda\in\Lambda_0} \frac{\left|h^*_{\lambda\lambda,t}\right|}{h^*_t} \right\|_2 \left(\underline{g}^{-1} \left\|y_t^2\right\|_2 + 1\right) + \left\| \sup_{\lambda\in\Lambda_0} \frac{\left|h^*_{\lambda,t}\right|}{h^*_t} \right\|_4^2 \left(2\underline{g}^{-1} \left\|y_t^2\right\|_2 + 1\right) \\
&< \infty.
\end{aligned}
$$

The finiteness at the last step is justified by Result 9. Thus, $l^*_{\lambda\lambda,t}(\lambda)$ is L_1-dominated in Λ_0.

For showing $\sup_{\lambda\in\Lambda_0} |L^*_{\lambda\lambda,T}(\lambda) - \mathcal{J}(\lambda)| \to 0$ a.s. we will make use of uniform SLLN. The ergodic stationarity of the variables y_t^2, h^*_t, $h^*_{\lambda,t}$ and $h^*_{\lambda\lambda,t}$ are satisfied by Result 1, 3, 7 and 8, respectively. Moreover, because $h^*_{\lambda\lambda,t}$ is continuous by Result 8, so are h^*_t and $h^*_{\lambda,t}$ since they are differentiable. Thus, $l^*_{\lambda\lambda,t}(\lambda)$ forms an ergodic stationary sequence with finite $E[\sup_{\lambda\in\Lambda_0} |l^*_{\lambda\lambda,t}(\lambda)|]$. Hence, by the uniform SLLN

$$
L^*_{\lambda\lambda,T}(\lambda) = \frac{1}{T}\sum_{t=1}^{T} l^*_{\lambda\lambda,t}(\lambda) \to E\left[l^*_{\lambda\lambda,t}(\lambda)\right] \quad a.s. \text{ uniformly on } \Lambda_0.
$$

Hence, $\sup_{\lambda\in\Lambda_0} |L^*_{\lambda\lambda,T}(\lambda) - \mathcal{J}(\lambda)| \to 0$ a.s. where $\mathcal{J}(\lambda) = E[l^*_{\lambda\lambda,t}(\lambda)]$. The uniform almost sure convergence implies the uniform continuity of $\mathcal{J}(\lambda)$ in Λ_0, in particular, $\mathcal{J}(\lambda)$ is continuous at λ_0. At the true parameter value, $\mathcal{J}(\lambda)$ takes the following form

$$
\begin{aligned}
\mathcal{J}(\lambda_0) &= E\left[l^*_{\lambda\lambda,t}(\lambda_0)\right] \\
&= E\left[-\frac{h^*_{\lambda\lambda,t}(\lambda_0)}{\sigma_t^2}\left(\varepsilon_t^2 - 1\right) + \frac{h^*_{\lambda,t}(\lambda_0)}{\sigma_t^2}\frac{h^{*\prime}_{\lambda,t}(\lambda_0)}{\sigma_t^2}\left(2\varepsilon_t^2 - 1\right)\right] \\
&= E\left[\varepsilon_t^2 - 1\right] E\left[-\sigma_t^{-2} h^*_{\lambda\lambda,t}(\lambda_0)\right] + E\left[2\varepsilon_t^2 - 1\right] E\left[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0)\, h^{*\prime}_{\lambda,t}(\lambda_0)\right] \\
&= E\left[\sigma_t^{-4} h^*_{\lambda,t}(\lambda_0)\, h^{*\prime}_{\lambda,t}(\lambda_0)\right].
\end{aligned}
$$

In the third equality, the term $E[\varepsilon_t^2 - 1]E[-\sigma_t^{-2} h^*_{\lambda\lambda,t}(\lambda_0)]$ vanishes because $E[\varepsilon_t^2] = 1$ and $E[-\sigma_t^{-2} h^*_{\lambda\lambda,t}(\lambda_0)] \leq \underline{g}^{-2} E[h^*_{\lambda\lambda,t}(\lambda_0)]$ is finite since $h^*_{\lambda\lambda,t}(\lambda)$ is continuous, thus finite, at λ_0 by Result 8. ■

Lemma D3. *Under the assumptions of Theorem 2, $\mathcal{I}(\lambda_0)$ and $\mathcal{J}(\lambda_0)$ are positive definite.*

Proof. Let $x \in \mathbb{R}^6 \setminus \{0\}$. We want to show that $x'\mathcal{I}(\lambda_0)\, x > 0$ for all $x \in \mathbb{R}^6 \setminus \{0\}$. For a contradiction, suppose that $x'\mathcal{I}(\lambda_0)\, x = 0$. Also note that we can express $x'\mathcal{I}(\lambda_0)\, x = 0$

as

$$
\begin{aligned}
x'\mathcal{I}(\lambda_0)\, x &= x' E\left[l^*_{\lambda,t}(\lambda_0) l^{*\prime}_{\lambda,t}(\lambda_0)\right] x \\
&= E\left[x' l^*_{\lambda,t}(\lambda_0) l^{*\prime}_{\lambda,t}(\lambda_0) x\right] \\
&= E\left[\left(x' l^*_{\lambda,t}(\lambda_0)\right)^2\right] \\
&= 0.
\end{aligned}
$$

But this implies that $x' l^*_{\lambda,t}(\lambda_0) = (\varepsilon_t^2 - 1)\sigma_t^{-2} x' h^*_{\lambda,t}(\lambda_0) = 0$ a.s. By Result 10 i) ε_t is not concentrated at two points, thus $\varepsilon_t^2 \neq 1$ a.s. Moreover, σ_t^2 is bounded $\underline{g} \leq \sigma_t^2 < \infty$ by Result 4 and Result 2. Therefore, we are left with $x' h^*_{\lambda,t}(\lambda_0) = 0$ a.s. Because $h^*_{\lambda,t}$ is stationary by Result 7, it is also true that $x' h^*_{\lambda,t-1}(\lambda_0) = 0$ a.s. Hence,

$$
\begin{aligned}
x' h^*_{\lambda,t}(\lambda_0) &= x'\left[g_{\lambda,t}(\lambda_0) + \beta h^*_{\lambda,t-1}(\lambda_0)\right] \\
&= x' g_{\lambda,t}(\lambda_0) + \beta x' h^*_{\lambda,t-1}(\lambda_0) \\
&= x' \frac{\partial g\left(y_t, \sigma_t^2; \gamma_0\right)}{\partial \lambda} \\
&= 0 \quad a.s.
\end{aligned}
$$

Yet, by the identification condition in Result 10 we have to have $x = 0$. We conclude that $\mathcal{I}(\lambda_0)$ is a (6×6) positive definite matrix.

In a similar way, the positive definiteness of $\mathcal{J}(\lambda_0)$ can be proven. Actually, after the step $x'\mathcal{J}(\lambda_0)\, x = E[\sigma_t^{-4} x' h^*_{\lambda,t}(\lambda_0)] = 0$ a.s., the same steps as above should be taken.

■

Lemma D.4 *Under the assumptions of Theorem 2,* $\sqrt{T}(\widetilde{\lambda}_T - \lambda_0) \rightarrow_d N\left(0, E[\varepsilon_t^4 - 1]\mathcal{J}(\lambda_0)^{-1}\right)$ where $\widetilde{\lambda}_T = \arg\min_{\lambda \in \Lambda} L^*_T(\lambda)$.

Proof. First note that $\widetilde{\lambda}_T \to \lambda_0$ a.s., which can be seen by the first few steps of the proof of Theorem 1. The only difference is that we do not have the term in (35), which converges to zero a.s.

$$
\begin{aligned}
\liminf_{T\to\infty} \inf_{\lambda \in B^c} \left(L^*_T(\lambda) - L^*_T(\lambda_0)\right) &\geq \liminf_{T\to\infty} \{E[l^*_t(\lambda_0)] - L^*_T(\lambda_0)\} + \liminf_{T\to\infty} \inf_{\lambda \in B^c} \{L^*_T(\lambda) - E\left[l^*_t(\lambda_0)\right]\} \\
&> 0 \qquad a.s. \;\; \text{by the proof of Theorem 1.}
\end{aligned}
$$

Hence, for strong consistency of $\widetilde{\lambda}_T$, it sufficed to show the positiveness of $\liminf_{T\to\infty} \inf_{\lambda \in B^c}(L^*_T(\lambda) - L^*_T(\lambda_0))$, which is done in the proofs of (36) and (37).

The mean value expansion[17] of $L_T^*(\lambda)$ around λ_0 gives us

$$L_{\lambda,T}^*(\widetilde{\lambda}_T) = L_{\lambda,T}^*(\lambda_0) + \dot{L}_{\lambda\lambda,T}^*(\widetilde{\lambda}_T - \lambda_0)$$

where $\dot{L}_{\lambda\lambda,T}^*$ signifies the (6×6) $L_{\lambda\lambda,T}^*(\lambda)$ matrix with each row evaluated at an intermediate point $\lambda_{i,T}$, for $i = 1, 2, \ldots, 6$, lying between $\widetilde{\lambda}_T$ and λ_0. More explicitly,

$$\dot{L}_{\lambda\lambda,T}^* = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \frac{\partial^2 l_t(\lambda)}{\partial \omega^2}\Big|_{\lambda=\lambda_{1,T}} & \cdots & \frac{\partial^2 l_t(\lambda)}{\partial \omega \partial \gamma_2}\Big|_{\lambda=\lambda_{1,T}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l_t(\lambda)}{\partial \gamma_2 \partial \omega}\Big|_{\lambda=\lambda_{6,T}} & \cdots & \frac{\partial^2 l_t(\lambda)}{\partial \gamma_2^2}\Big|_{\lambda=\lambda_{6,T}} \end{bmatrix}_{(6\times 6)}.$$

Because $\lambda_{i,T}$ lies between $\widetilde{\lambda}_T$ and λ_0 for each i and because $\widetilde{\lambda}_T$ is strongly consistent for λ_0, so are each $\lambda_{i,T}$. Hence, $\lambda_{i,T} \to \lambda_0$ a.s. for $i = 1, 2, \ldots, 6$. Moreover, the convexity of Λ_0 assures that each $\lambda_{i,T}$ is contained in the interior of Λ_0 for large T.

Next, we will show that $\dot{L}_{\lambda\lambda,T}^* \to \mathcal{J}(\lambda_0)$ a.s. as $T \to \infty$. By Lemma D.2 we have $\sup_{\lambda\in\Lambda_0} |L_{\lambda\lambda,T}^*(\lambda) - \mathcal{J}(\lambda)| \to 0$ a.s. Hence, each row of $L_{\lambda\lambda,T}^*(\lambda)$ converges to the corresponding row of $\mathcal{J}(\lambda)$. Thus, i^{th} row of $L_{\lambda\lambda,T}^*(\lambda)$ evaluated at $\lambda_{i,T}$ converges almost surely to the i^{th} row of $\mathcal{J}(\lambda)$ evaluated at $\lambda_{i,T}$, for each i. As a result, we can write that $\dot{L}_{\lambda\lambda,T}^* \to \dot{\mathcal{J}}$ a.s. where $\dot{\mathcal{J}}$ signifies the matrix $\mathcal{J}(\lambda)$ with i^{th} row evaluated at $\lambda_{i,T}$ for $i = 1, 2, \ldots, 6$. Because $\mathcal{J}(\lambda)$ is continuous at λ_0 by Lemma D.2 and because $\lambda_{i,T} \to \lambda_0$ a.s. for each i, we have $\dot{\mathcal{J}} \to \mathcal{J}(\lambda_0)$ a.s.

$$\dot{L}_{\lambda\lambda,T}^* \to \dot{\mathcal{J}} \;\; a.s. \quad \text{and} \quad \dot{\mathcal{J}} \to \mathcal{J}(\lambda_0) \;\; a.s. \quad \text{imply} \quad \dot{L}_{\lambda\lambda,T}^* \to \mathcal{J}(\lambda_0) \;\; a.s.$$

Because $\mathcal{J}(\lambda_0)$ is invertible being positive definite by Lemma D.3, $\dot{L}_{\lambda\lambda,T}^*$ is also invertible for large T. Thus, we can write

$$\dot{L}_{\lambda\lambda,T}^{*-1} \to \mathcal{J}(\lambda_0)^{-1} \;\; a.s. \;\; \text{as } T \to \infty. \tag{40}$$

Let's multiply the mean value expansion with $\sqrt{T}$ and with the Moore-Penrose Inverse[18] of $\dot{L}_{\lambda\lambda,T}^*$, which exists for all T and denoted by $\dot{L}_{\lambda\lambda,T}^{*+}$. Then,

$$\sqrt{T}\dot{L}_{\lambda\lambda,T}^{*+} L_{\lambda,T}^*(\widetilde{\lambda}_T) = \sqrt{T}\dot{L}_{\lambda\lambda,T}^{*+} L_{\lambda,T}^*(\lambda_0) + \sqrt{T}\dot{L}_{\lambda\lambda,T}^{*+}\dot{L}_{\lambda\lambda,T}^*(\widetilde{\lambda}_T - \lambda_0).$$

[17] Let $\mathbf{h} : \mathbb{R}^p \to \mathbb{R}^q$ be continuously differentiable. Then, $\mathbf{h}(\mathbf{x})$ admits the mean value expansion

$$\mathbf{h}(\mathbf{x}) = \mathbf{h}(\mathbf{x}_0) + \frac{\partial \mathbf{h}(\bar{\mathbf{x}})}{\partial \mathbf{x}'}(\mathbf{x} - \mathbf{x}_0),$$

where $\bar{\mathbf{x}}$ is a mean value lying between $\mathbf{x}$ and $\mathbf{x}_0$.

[18] The $(n \times m)$ matrix A^+ is the Moore-Penrose Inverse of the $(n \times m)$ matrix A if it satisfies

By adding and subtracting $\sqrt{T}(\widetilde{\lambda}_T - \lambda_0)$ to the right-hand-side, we arrange the above equation to obtain

$$\sqrt{T}(\widetilde{\lambda}_T - \lambda_0) = (I_6 - \dot{L}^{*+}_{\lambda\lambda,T}\dot{L}^{*}_{\lambda\lambda,T})\sqrt{T}(\widetilde{\lambda}_T - \lambda_0) + \sqrt{T}\dot{L}^{*+}_{\lambda\lambda,T}L^{*}_{\lambda,T}(\widetilde{\lambda}_T) - \sqrt{T}\dot{L}^{*+}_{\lambda\lambda,T}L^{*}_{\lambda,T}(\lambda_0) \quad (41)$$

where I_6 is the (6×6) identity matrix. Now, we will show that the first two terms on the right-hand-side of (41) converge to zero a.s. Actually, these terms will be exactly equal to zero for sufficiently large T, in other words, for each event w on a set with probability one there exists T_w such that for all $T \geq T_w$ the first two terms are exactly zero. Regarding the first term, for sufficiently large T, the inverse of $\dot{L}^{*}_{\lambda\lambda,T}$ exists as shown in (40). Hence, $\dot{L}^{*+}_{\lambda\lambda,T} = \dot{L}^{*-1}_{\lambda\lambda,T}$, which gives us $\dot{L}^{*+}_{\lambda\lambda,T}\dot{L}^{*}_{\lambda\lambda,T} = I_6$. For the second term, because $\widetilde{\lambda}_T$ minimizes the function $L^{*}_{T}(\lambda)$ on Λ and because λ_0 is an interior point of Λ_0, we have $L^{*}_{\lambda,T}(\widetilde{\lambda}_T) = 0$ for all sufficiently large T. Note that, being a minimizer is not enough to have a zero derivative at that point. First of all, the derivative is taken in the vicinity of λ_0, i.e. on Λ_0. Moreover, $\widetilde{\lambda}_T$ may be on the boundary of Λ_0 for some T, which may result in a non-zero derivative. Therefore, we need $\widetilde{\lambda}_T$ to be an interior point of Λ_0, and to ensure this, given that $\widetilde{\lambda}_T$ is consistent for λ_0, T should be sufficiently large. As a result first two terms in (41) are not only converging to zero a.s. but also they are exactly equal to zero a.s. after a threshold for T. Furthermore, the Moore-Penrose inverse $\dot{L}^{*+}_{\lambda\lambda,T}$ converges to $\mathcal{J}(\lambda_0)^{-1}$ a.s. since $\dot{L}^{*+}_{\lambda\lambda,T} = \dot{L}^{*-1}_{\lambda\lambda,T}$ for sufficiently large T. Thus, (41) can be written as

$$\sqrt{T}(\widetilde{\lambda}_T - \lambda_0) = o_1(1) + \left[\mathcal{J}(\lambda_0)^{-1} + o_2(1)\right]\sqrt{T}L^{*}_{\lambda,T}(\lambda_0),$$

where $o_1(1)$ and $o_2(1)$ are vector- and matrix-valued processes, respectively, and converge to zero a.s. at rate T. Thus, as $T \to \infty$, the right-hand-side becomes $\mathcal{J}(\lambda_0)^{-1}\sqrt{T}L^{*}_{\lambda,T}(\lambda_0)$. But, by Lemma D.1, we have that $\sqrt{T}L^{*}_{\lambda,T}(\lambda_0) \to_d N(0, \mathcal{I}(\lambda_0))$ yielding the desired result $\mathcal{J}(\lambda_0)^{-1}\sqrt{T}L^{*}_{\lambda,T}(\lambda_0) \to_d N(0, \mathcal{J}(\lambda_0)^{-1}\mathcal{I}(\lambda_0)\mathcal{J}(\lambda_0)^{-1})$. Here, note that the matrix $\mathcal{J}(\lambda_0) = E[\sigma_t^{-4}h^{*}_{\lambda,t}(\lambda_0)h^{*\prime}_{\lambda,t}(\lambda_0)]$ is symmetric, so is its inverse, and moreover $\mathcal{J}(\lambda_0)^{-1}\mathcal{I}(\lambda_0) =$

i) $AA^{+}A = A$,

ii) $A^{+}AA^{+} = A^{+}$,

iii) $(AA^{+})^{H} = AA^{+}$,

iv) $(A^{+}A)^{H} = A^{+}A$,

where A^{H} is the conjugate transpose of the matrix A.

$E[\varepsilon_t^4 - 1]$. To conclude the proof,

$$\sqrt{T}(\widetilde{\lambda}_T - \lambda_0) \to_d N\left(0, E\left[\varepsilon_t^4 - 1\right]\left\{E\left[\sigma_t^{-4} h_{\lambda,t}^*(\lambda_0)\, h_{\lambda,t}^{*\prime}(\lambda_0)\right]\right\}^{-1}\right).$$

■

Lemma D.5 *Under the assumptions of Theorem 2, there exists a constant $\gamma > 1$ such that*

$$\gamma^t \sup_{\lambda \in \Lambda_0} \left| l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda) \right| \to 0 \text{ in } L_{1/3} - \text{norm as } t \to \infty.$$

Proof. In this proof and in some further analysis, we will use the inequality

$$\begin{aligned} |x^* y^* - xy| &= |x^* y^* - xy \pm xy^*| = |y^*(x^* - x) - (x^* - x)(y^* - y) + x^*(y^* - y)| \\ &\leq |y^*|\,|(x^* - x)| + |x^* - x|\,|y^* - y| + |x^*|\,|y^* - y| \end{aligned} \tag{42}$$

for any conformable vectors.

Now, let's consider the difference $h_t^{*-1} h_{\lambda,t}^* - h_t^{-1} h_{\lambda,t}$.

$$\begin{aligned} \left| \frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t} \right| &= \left| \frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t} + \frac{h_{\lambda,t}^*}{h_t} - \frac{h_{\lambda,t}^*}{h_t} \right| \\ &= \left| h_{\lambda,t}^* \left(\frac{1}{h_t^*} - \frac{1}{h_t} \right) - \frac{1}{h_t}\left(h_{\lambda,t}^* - h_{\lambda,t}\right) \right| \\ &\leq \left| h_{\lambda,t}^* \right| \left| \frac{1}{h_t^*} - \frac{1}{h_t} \right| + \left| h_t^{-1} \right| \left| h_{\lambda,t}^* - h_{\lambda,t} \right| \\ &\leq \underline{g}^{-2} \left| h_{\lambda,t}^* \right| \left| h_{\lambda,t}^* - h_{\lambda,t} \right| + \underline{g}^{-1} \left| h_{\lambda,t}^* - h_{\lambda,t} \right|, \end{aligned}$$

where the last inequality follows by the Mean Value Theorem and Result 4. Thus, we can write

$$\begin{aligned} & \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t} \right| \right\|_{1/2} \\ \leq\; & \left\| \sup_{\lambda \in \Lambda_0} \left\{ \underline{g}^{-2} \left| h_{\lambda,t}^* \right| \left| h_{\lambda,t}^* - h_{\lambda,t} \right| + \underline{g}^{-1} \left| h_{\lambda,t}^* - h_{\lambda,t} \right| \right\} \right\|_{1/2} \\ \leq\; & \triangle_{1/2,2} \left[\underline{g}^{-2} \left\| \sup_{\lambda \in \Lambda_0} \left| h_{\lambda,t}^* \right| \left| h_{\lambda,t}^* - h_{\lambda,t} \right| \right\|_{1/2} + \underline{g}^{-1} \left\| \sup_{\lambda \in \Lambda_0} \left| h_{\lambda,t}^* - h_{\lambda,t} \right| \right\|_{1/2} \right] \\ \leq\; & 2 \left[\underline{g}^{-2} \left\| \sup_{\lambda \in \Lambda_0} \left| h_{\lambda,t}^* \right| \right\|_1 \left\| \sup_{\lambda \in \Lambda_0} \left| h_{\lambda,t}^* - h_{\lambda,t} \right| \right\|_2 + \underline{g}^{-1} \left\| \sup_{\lambda \in \Lambda_0} \left| h_{\lambda,t}^* - h_{\lambda,t} \right| \right\|_{1/2} \right] \end{aligned}$$

where, in the second inequality, we make use of Lemma 1, and for the last step we use Cauchy-Schwarz Inequality and Liapunov's Inequality. In order to get a more proper upper bound for $\|\sup_{\lambda\in\Lambda_0}|h_t^{*-1}h_{\lambda,t}^* - h_t^{-1}h_{\lambda,t}|\|_{1/2}$, recall that $\|\sup_{\lambda\in\Lambda_0}|h_t^* - h_t|\|_r$ and $\|\sup_{\lambda\in\Lambda_0}|h_{\lambda,t}^* - h_{\lambda,t}|\|_{r/4}$ are bounded by $C\kappa^t$ and $C'\max\{t, t^{4/r}\}\kappa^{t-1}$ by Result 3 and 7, respectively. Moreover, by Result 7, we also have that $E[\sup_{\lambda\in\Lambda_0}|h_{\lambda,t}^*|^{r/2}] < \infty$, thus, $\|\sup_{\lambda\in\Lambda_0}|h_{\lambda,t}^*|\|_{r/2} = D < \infty$. Finally, in Result 9, we showed that we can replace r with 2 in our analysis. Therefore,

$$\begin{aligned}\left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\right\|_{1/2} &\leq 2\underline{g}^{-2}DC\kappa^t + 2\underline{g}^{-1}C'\max\left\{t, t^2\right\}\kappa^{t-1} \\ &\leq \max\left\{2\underline{g}^{-2}DC, 2\underline{g}^{-1}C'\kappa^{-1}\right\}\left(1+t^2\right)\kappa^t \\ &\leq C''t^2\kappa^t. \qquad (43)\end{aligned}$$

Next, let's consider the difference $l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)$ and use the inequality (42)

$$\begin{aligned}&\left|l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)\right| \\ =\ &\left|-\frac{h_{\lambda,t}^*}{h_t^*}\left(\frac{y_t^2}{h_t^*} - 1\right) + \frac{h_{\lambda,t}}{h_t}\left(\frac{y_t^2}{h_t} - 1\right)\right| \\ \leq\ &\left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left|\frac{y_t^2}{h_t^*} - 1\right| + \left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left|\frac{y_t^2}{h_t^*} - \frac{y_t^2}{h_t}\right| + \left|\frac{h_{\lambda,t}^*}{h_t^*}\right|\left|\frac{y_t^2}{h_t^*} - \frac{y_t^2}{h_t}\right| \\ \leq\ &\left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left(\underline{g}^{-1}y_t^2 + 1\right) + \left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left(\underline{g}^{-2}y_t^2\left|h_t^* - h_t\right|\right) + \left|h_{\lambda,t}^*\right|\left(\underline{g}^{-3}y_t^2\left|h_t^* - h_t\right|\right),\end{aligned}$$

where the last inequality follows by the Mean Value Theorem and Result 4. Hence, we now are able to find a proper upper bound of the norm $\|\sup_{\lambda\in\Lambda_0}|l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)|\|_{1/3}$.

$$\begin{aligned}&\left\|\sup_{\lambda\in\Lambda_0}\left|l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)\right|\right\|_{1/3} \\ \leq\ &\left\|\sup_{\lambda\in\Lambda_0}\left|\left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left(\underline{g}^{-1}y_t^2 + 1\right) + \left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left(\underline{g}^{-2}y_t^2\left|h_t^* - h_t\right|\right) + \left|h_{\lambda,t}^*\right|\left(\underline{g}^{-3}y_t^2\left|h_t^* - h_t\right|\right)\right|\right\|_{1/3} \\ \leq\ &\triangle_{1/3,3}\left\{\left\|\left(\underline{g}^{-1}y_t^2 + 1\right)\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\right\|_{1/3}\right. \\ &\left.+\left\|\underline{g}^{-2}y_t^2\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*} - \frac{h_{\lambda,t}}{h_t}\right|\left|h_t^* - h_t\right|\right\|_{1/3} + \left\|\underline{g}^{-3}y_t^2\sup_{\lambda\in\Lambda_0}\left|h_{\lambda,t}^*\right|\left|h_t^* - h_t\right|\right\|_{1/3}\right\}\end{aligned}$$

$$
\begin{aligned}
\leq & \ 3\Bigg\{ \left\|\underline{g}^{-1}y_t^2+1\right\|_1 \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|\right\|_{1/2} + \underline{g}^{-2}\left\|y_t^2\right\|_2 \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|\right\|_{1/2} \left\|\sup_{\lambda\in\Lambda_0}\left|h_t^*-h_t\right|\right\|_2 \\
& + \underline{g}^{-3}\left\|y_t^2\right\|_1 \left\|\sup_{\lambda\in\Lambda_0}\left|h_{\lambda,t}^*\right|\right\|_1 \left\|\sup_{\lambda\in\Lambda_0}\left|h_t^*-h_t\right|\right\|_1 \Bigg\} \\
\leq & \ 3\Bigg\{ \left(\underline{g}^{-1}\left\|y_t^2\right\|_2+1\right) \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|\right\|_{1/2} \\
& + \underline{g}^{-2}\left\|y_t^2\right\|_2 \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|\right\|_{1/2} \left\|\sup_{\lambda\in\Lambda_0}\left|h_t^*-h_t\right|\right\|_2 \\
& + \underline{g}^{-3}\left\|y_t^2\right\|_2 \left\|\sup_{\lambda\in\Lambda_0}\left|h_{\lambda,t}^*\right|\right\|_1 \left\|\sup_{\lambda\in\Lambda_0}\left|h_t^*-h_t\right|\right\|_2 \Bigg\}.
\end{aligned}
$$

where Lemma 1 is used in the second inequality, Hölder's Inequality is used in the third inequality, and Liapunov's Inequality is used in the last step. Moreover, by Result 9, we know that $E[y_t^4] < \infty$ which provides the finiteness of the norm $\|y_t^2\|_2 = M < \infty$. In addition, by using the upper bound for $|| \sup_{\lambda\in\Lambda_0} |h_t^{*-1}h_{\lambda,t}^* - h_t^{-1}h_{\lambda,t}|||_{1/2}$ found in (43), we can bound $|| \sup_{\lambda\in\Lambda_0} |l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)|||_{1/3}$ from above in the following way

$$
\begin{aligned}
\left\|\sup_{\lambda\in\Lambda_0}\left|l_{\lambda,t}^*(\lambda)-l_{\lambda,t}(\lambda)\right|\right\|_{1/3} & \leq 3\left\{\left(\underline{g}^{-1}M+1\right)C''t^2\kappa^t+\underline{g}^{-2}MC''t^2\kappa^tC\kappa^t+\underline{g}^{-3}MDC\kappa^t\right\} \\
& \leq 3\max\left\{\left(\underline{g}^{-1}M+1\right)C'',\underline{g}^{-2}MC''C,\underline{g}^{-3}MDC\right\}\kappa^tt^2\left(1+\kappa^t\right) \\
& \leq Bt^2\kappa^t\left(1+\kappa^t\right) \\
& \leq 2Bt^2\kappa^t.
\end{aligned}
$$

for some constant B. Now, let $\gamma \in (1, \kappa^{-1})$ so that $||\gamma^t \sup_{\lambda\in\Lambda_0} |l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)|||_{1/3} \leq 2Bt^2\left(\kappa\gamma\right)^t \to 0$ a.s. as $t \to \infty$. In other words, the almost sure convergence of $\gamma^t \sup_{\lambda\in\Lambda_0} |l_{\lambda,t}^*(\lambda) - l_{\lambda,t}(\lambda)|$ to zero in $L_{1/3}-$norm is proven as $t \to \infty$. ■

Lemma D.6 *Under the assumptions of Theorem 2,* $\sqrt{T}(\hat{\lambda}_T - \widetilde{\lambda}_T) \to 0$ *a.s. as* $T \to \infty$.

Proof. We know that $\lambda_0 \in \mathring{\Lambda}_0$, and both $\hat{\lambda}_T$ and $\widetilde{\lambda}_T$ are strongly consistent for λ_0 by Theorem 1 and by the proof of Lemma D.4. Thus, for sufficiently large T we can say that $\hat{\lambda}_T$ and $\widetilde{\lambda}_T$ belong to $\mathring{\Lambda}_0$ with probability one. Since $\hat{\lambda}_T$ and $\widetilde{\lambda}_T$ are minimizers of $L_T(\lambda)$ and $L_T^*(\lambda)$ on Λ, respectively, we can write $L_{\lambda,T}(\hat{\lambda}_T) = L_{\lambda,T}^*(\widetilde{\lambda}_T) = 0$ for sufficiently large T. By applying the mean value theorem we obtain

$$
L_{\lambda,T}^*(\widetilde{\lambda}_T) - L_{\lambda,T}^*(\hat{\lambda}_T) = \ddot{L}_{\lambda\lambda,T}^*(\widetilde{\lambda}_T - \hat{\lambda}_T),
$$

where $\ddot{L}^*_{\lambda\lambda,T}$ signifies the matrix $L^*_{\lambda\lambda,T}(\lambda)$ with each row evaluated at an intermediate point $\ddot{\lambda}_{i,T}$, for $i = 1, 2, \ldots, 6$, lying between $\widetilde{\lambda}_T$ and $\hat{\lambda}_T$. By using the above equation and the equality $L_{\lambda,T}(\hat{\lambda}_T) = L^*_{\lambda,T}(\widetilde{\lambda}_T) = 0$, we can write that

$$\sqrt{T}(L_{\lambda,T}(\hat{\lambda}_T) - L^*_{\lambda,T}(\hat{\lambda}_T)) = \sqrt{T}(L^*_{\lambda,T}(\widetilde{\lambda}_T) - L^*_{\lambda,T}(\hat{\lambda}_T)) = \ddot{L}^*_{\lambda\lambda,T}\sqrt{T}(\widetilde{\lambda}_T - \hat{\lambda}_T).$$

Let's concentrate on the difference in the left-hand-side.

$$\begin{aligned}
\sqrt{T}\left|L_{\lambda,T}(\hat{\lambda}_T) - L^*_{\lambda,T}(\hat{\lambda}_T)\right| &= \sqrt{T}\left|\frac{1}{T}\sum_{t=1}^{T} l_{\lambda,t}(\hat{\lambda}_T) - \frac{1}{T}\sum_{t=1}^{T} l^*_{\lambda,t}(\hat{\lambda}_T)\right| \\
&= \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\left(l_{\lambda,t}(\hat{\lambda}_T) - l^*_{\lambda,t}(\hat{\lambda}_T)\right)\right| \\
&\leq \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\sup_{\lambda\in\Lambda_0}\left(l_{\lambda,t}(\hat{\lambda}_T) - l^*_{\lambda,t}(\hat{\lambda}_T)\right)\right| \\
&\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sup_{\lambda\in\Lambda_0}\left|l_{\lambda,t}(\hat{\lambda}_T) - l^*_{\lambda,t}(\hat{\lambda}_T)\right|.
\end{aligned}$$

By Lemma D.5 and Lemma 2, $\sum_{t=1}^{\infty}\sup_{\lambda\in\Lambda_0}|l_{\lambda,t}(\hat{\lambda}_T) - l^*_{\lambda,t}(\hat{\lambda}_T)| < \infty$. Therefore, $\sqrt{T}|L_{\lambda,T}(\hat{\lambda}_T) - L^*_{\lambda,T}(\hat{\lambda}_T)| \to 0$ a.s. as $T \to \infty$. Hence,

$$\ddot{L}^*_{\lambda\lambda,T}\sqrt{T}(\widetilde{\lambda}_T - \hat{\lambda}_T) \to 0 \quad a.s. \ \ as \ T \to \infty.$$

Now, we will show that $\ddot{L}^*_{\lambda\lambda,T}$ is invertible, thus nonsingular, so that we can conclude that $\sqrt{T}(\widetilde{\lambda}_T - \hat{\lambda}_T) \to 0$ a.s. Note that, for each i, $\ddot{\lambda}_{i,T}$ is strongly consistent for λ_0 because $\ddot{\lambda}_{i,T}$ lies between $\widetilde{\lambda}_T$ and $\hat{\lambda}_T$. By Lemma D.2 we had $\sup_{\lambda\in\Lambda_0}|L^*_{\lambda\lambda,T}(\lambda) - \mathcal{J}(\lambda)| \to 0$ a.s., thus $\ddot{L}^*_{\lambda\lambda,T} \to \ddot{\mathcal{J}}$ where $\ddot{\mathcal{J}}$ denotes the matrix $\mathcal{J}(\lambda)$ with i^{th} row evaluated at $\ddot{\lambda}_{i,T}$. Since $\mathcal{J}(\lambda)$ is continuous at λ_0 and $\ddot{\lambda}_{i,T}$ is strongly consistent for λ_0, we have $\ddot{\mathcal{J}} \to \mathcal{J}(\lambda_0)$ a.s., which implies $\ddot{L}^*_{\lambda\lambda,T} \to \mathcal{J}(\lambda_0)$ a.s. Moreover, for sufficiently large T, $\ddot{L}^{*-1}_{\lambda\lambda,T}$ exists because $\mathcal{J}(\lambda_0)^{-1}$ exists by Lemma D.3. Hence, $\ddot{L}^*_{\lambda\lambda,T}$ being a nonsingular matrix, we can conclude that $\sqrt{T}(\hat{\lambda}_T - \widetilde{\lambda}_T) \to 0$ a.s. as $T \to \infty$. ■

Proof of (8) *Under the assumptions of Theorem 2,*

$$\hat{\mathcal{I}}_T = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t^4}{\hat{h}_t^2} - 1\right)\right]\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{h}_{\lambda,t}}{\hat{h}_t}\frac{\hat{h}'_{\lambda,t}}{\hat{h}_t}\right] \to \mathcal{I}(\lambda_0) \quad a.s.$$

Proof. We will prove the strong consistency of the estimators in $\hat{\mathcal{I}}_T$, i.e., we will prove

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t^4}{\hat{h}_t^2}-1\right) \quad\rightarrow\quad E\left[\varepsilon_t^4-1\right] \qquad a.s., \tag{44}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{h}_{\lambda,t}}{\hat{h}_t}\frac{\hat{h}'_{\lambda,t}}{\hat{h}_t} \quad\rightarrow\quad E\left[\frac{h_{\lambda,t}(\lambda_0)}{\sigma_t^2}\frac{h'_{\lambda,t}(\lambda_0)}{\sigma_t^2}\right] \qquad a.s. \tag{45}$$

We will use uniform SLLN and the strong consistency of the QML estimator $\hat{\lambda}_T$. First, we prove two results on uniform strong consistency.

$$\sup_{\lambda\in\Lambda_0}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_t^4}{h_t^{*2}}-E\left[\frac{y_t^4}{h_t^{*2}}\right]\right| \quad\rightarrow\quad 0 \;\; a.s. \;\; as\; T\rightarrow\infty, \tag{46}$$

$$\sup_{\lambda\in\Lambda_0}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{h_{\lambda,t}}{h_t^*}\frac{h'_{\lambda,t}}{h_t^*}-E\left[\frac{h_{\lambda,t}}{h_t^*}\frac{h'_{\lambda,t}}{h_t^*}\right]\right| \quad\rightarrow\quad 0 \;\; a.s. \;\; as\; T\rightarrow\infty. \tag{47}$$

In order to use uniform SLLN, we need to prove the finiteness of $E[\sup_{\lambda\in\Lambda_0}|h_t^{*-2}y_t^4|]$ and $E[\sup_{\lambda\in\Lambda_0}|h_t^{*-2}h_{\lambda,t}h'_{\lambda,t}|]$.

$$E\left[\sup_{\lambda\in\Lambda_0}\left|\frac{y_t^4}{h_t^{*2}}\right|\right] = E\left[y_t^4\sup_{\lambda\in\Lambda_0}\left|\frac{1}{h_t^{*2}}\right|\right] \le E\left[y_t^4\underline{g}^{-2}\right] = \underline{g}^{-2}E\left[y_t^4\right] < \infty \quad \text{by Result 4 and 9,}$$

$$E\left[\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}}{h_t^*}\frac{h'_{\lambda,t}}{h_t^*}\right|\right] = \left\|\sup_{\lambda\in\Lambda_0}\left|\frac{h_{\lambda,t}}{h_t^*}\frac{h'_{\lambda,t}}{h_t^*}\right|\right\|_1 = \left\|\sup_{\lambda\in\Lambda_0}\frac{|h_{\lambda,t}|}{h_t^*}\right\|_2^2 \le \left\|\sup_{\lambda\in\Lambda_0}\frac{|h_{\lambda,t}|}{h_t^*}\right\|_4^2 < \infty \quad \text{by Result 9.}$$

Hence, uniform SLLN applies, thus (46) and (47) hold.

Now, let's concentrate on the convergence of the first estimator, and try to show that

$$\sup_{\lambda\in\Lambda_0}\left|1/T\sum_{t=1}^{T}y_t^4\left(h_t^{*-2}-h_t^{-2}\right)\right| \rightarrow 0 \;\; a.s. \;\; as\; T\rightarrow\infty.$$

We first analyze the difference in the summand.

$$\begin{aligned}\left|\frac{y_t^4}{h_t^{*2}}-\frac{y_t^4}{h_t^2}\right| &= y_t^4\left|\frac{1}{h_t^{*2}}-\frac{1}{h_t^2}\right| \\ &\le y_t^4\left|\left(-2\dot{h}_t^{-3}\right)\left(h_t^*-h_t\right)\right| \\ &= 2y_t^4\left|\dot{h}_t^{-3}\right|\left|h_t^*-h_t\right| \\ &\le 2y_t^4\underline{g}^{-3}\left|h_t^*-h_t\right|,\end{aligned}$$

where the first inequality follows from the Mean Value Theorem for $\dot{h}_t$ lying between h_t^* and h_t. Thus, by Result 4 $\dot{h}_t \ge \underline{g}$, which justifies the last inequality. By Hölder's Inequality we get

$$\begin{aligned}\left\|\sup_{\lambda\in\Lambda_0}\left|\frac{y_t^4}{h_t^{*2}}-\frac{y_t^4}{h_t^2}\right|\right\|_{2/3} &\leq \left\|\sup_{\lambda\in\Lambda_0} 2y_t^4\underline{g}^{-3}\left|h_t^*-h_t\right|\right\|_{2/3}\\ &\leq 2\underline{g}^{-3}\left\|y_t^4\right\|_1\left\|\sup_{\lambda\in\Lambda_0}\left|h_t^*-h_t\right|\right\|_2\\ &\leq M'\kappa^t,\end{aligned}$$

where $M'=2\underline{g}^{-3}\left\|y_t^4\right\|_1 C<\infty$ by Result 3 and 9. Let $1<\gamma<\kappa^{-1}$, then $||\gamma^t\sup_{\lambda\in\Lambda_0}|y_t^4(h_t^{*-2}-h_t^{-2})|||_{2/3}\leq M'(\gamma\kappa)^t\to 0$ a.s. By Lemma 2, $\sum_{t=1}^{\infty}\left\|\sup_{\lambda\in\Lambda_0}|y_t^4\left(h_t^{*-2}-h_t^{-2}\right)|\right\|_{2/3}<\infty$. Therefore, we conclude that

$$\sup_{\lambda\in\Lambda_0}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t^4}{h_t^{*2}}-\frac{y_t^4}{h_t^2}\right)\right|\to 0 \quad a.s. \ as \ T\to\infty. \tag{48}$$

The first desired result in (44) will be derived from the uniform convergence results in (46) and (48). They, respectively, imply that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{y_t^4}{h_t^{*2}}\to E\left[\frac{y_t^4}{h_t^{*2}}\right] \ a.s. \ uniformly \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\frac{y_t^4}{h_t^{*2}}\to\frac{1}{T}\sum_{t=1}^{T}\frac{y_t^4}{h_t^2} \ a.s. \ uniformly.$$

Thus, we can infer that $1/T\sum_{t=1}^{T}h_t^{-2}y_t^4\to E[h_t^{*-2}y_t^4]$ a.s. uniformly on Λ_0. Because the convergence is uniformly, it will also hold for a particular choice of $\lambda\in\Lambda_0$, for instance for $\hat{\lambda}_T$. Therefore, we can write that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{y_t^4}{h_t^2(\hat{\lambda}_T)}\to E\left[\frac{y_t^4}{h_t^{*2}(\hat{\lambda}_T)}\right] \quad a.s.$$

Since $\hat{\lambda}_T$ is strongly consistent for λ_0 and h_t^* is continuous we have $h_t^{*2}(\hat{\lambda}_T)\to h_t^{*2}(\lambda_0)=\sigma_t^4$. Then, we can easily obtain that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t^4}{\hat{h}_t^2}-1\right)\to E\left[\frac{y_t^4}{\sigma_t^4}-1\right]=E\left[\varepsilon_t^4-1\right] \quad a.s.$$

Next, let's find similar results for the second estimator in $\hat{\mathcal{I}}_T$. We have already found the uniform convergence result in (47). What we need is to find the analogous result of (48) that contains the first derivatives of the processes h_t and h_t^*. We again use the inequality (42) with $x^*=h_t^{*-1}h_{\lambda,t}^*$ and $y^*=x^{*\prime}$,

$$\begin{aligned}\left|\frac{h_{\lambda,t}^*}{h_t^*}\frac{h_{\lambda,t}^{*\prime}}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\frac{h_{\lambda,t}'}{h_t}\right| &\leq \left|\frac{h_{\lambda,t}^{*\prime}}{h_t^*}\right|\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|+\left|\frac{h_{\lambda,t}^*}{h_t^*}\right|\left|\frac{h_{\lambda,t}^{*\prime}}{h_t^*}-\frac{h_{\lambda,t}'}{h_t}\right|+\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|\left|\frac{h_{\lambda,t}^{*\prime}}{h_t^*}-\frac{h_{\lambda,t}'}{h_t}\right|\\ &\leq 2\left|\frac{h_{\lambda,t}^*}{h_t^*}\right|\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|+\left|\frac{h_{\lambda,t}^*}{h_t^*}-\frac{h_{\lambda,t}}{h_t}\right|^2.\end{aligned}$$

Thus, we can write

$$
\begin{aligned}
& \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} \frac{h^{*\prime}_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \frac{h'_{\lambda,t}}{h_t} \right| \right\|_{1/4} \\
\leq & \left\| \sup_{\lambda \in \Lambda_0} \left\{ 2 \left| \frac{h^*_{\lambda,t}}{h^*_t} \right| \left| \frac{h^*_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \right| + \left| \frac{h^*_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \right|^2 \right\} \right\|_{1/4} \\
\leq & \triangle_{1/4,2} \left\{ 2 \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} \right| \left| \frac{h^*_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \right| \right\|_{1/4} + \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \right|^2 \right\|_{1/4} \right\} \\
\leq & 8 \left\{ 2 \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} \right| \right\|_{1/2} \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \right| \right\|_{1/2} + \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \right| \right\|^2_{1/2} \right\} \\
\leq & 8 \left\{ 2 \left\| \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} \right| \right\|_4 C'' t^2 \kappa^t + \left(C'' t^2 \kappa^t \right)^2 \right\} \\
\leq & D' t^4 \kappa^t,
\end{aligned}
$$

where the second inequality follows from Lemma 1, the third one follows from Cauchy-Schwarz Inequality, the fourth one is justified by Liapunov's Inequality and by (43), the fifth one is satisfied by Result 9. Again by using similar arguments and making use of Lemma 2,

$$
\sup_{\lambda \in \Lambda_0} \left| \frac{1}{T} \sum_{t=1}^{T} \left(\frac{h^*_{\lambda,t}}{h^*_t} \frac{h^{*\prime}_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \frac{h'_{\lambda,t}}{h_t} \right) \right| \leq \frac{1}{T} \sum_{t=1}^{T} \sup_{\lambda \in \Lambda_0} \left| \frac{h^*_{\lambda,t}}{h^*_t} \frac{h^{*\prime}_{\lambda,t}}{h^*_t} - \frac{h_{\lambda,t}}{h_t} \frac{h'_{\lambda,t}}{h_t} \right| \to 0 \quad a.s.
$$

Hence, by a very similar reasoning done for the first estimator in $\hat{\mathcal{I}}_T$, we can conclude that

$$
\frac{1}{T} \sum_{t=1}^{T} \frac{\hat{h}_{\lambda,t}}{\hat{h}_t} \frac{\hat{h}'_{\lambda,t}}{\hat{h}_t} \to E \left[\frac{h_{\lambda,t}(\lambda_0)}{\sigma^2_t} \frac{h'_{\lambda,t}(\lambda_0)}{\sigma^2_t} \right] \quad a.s.
$$

■

References

[1] ANDERSON, H. M., K. NAM, AND F. VAHID (1999): "Asymmetric nonlinear smooth transition GARCH models," In: Rothmann, P. (Ed.): *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston, Kluwer Academic Press, 191–207.

[2] BAILLIE, R. T., T. BOLLERSLEV, AND H. O. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics,* 74, 3–30.

[3] BERA, A. K., M. L. HIGGINGS, AND S. LEE (1992): "Interaction between autocorrelation and conditional heteroscedasticity: A Random-Coefficient Approach," *Journal of Business and Economic Statistics*, 10, 133–142.

[4] BERKES, I., L. HORVÁTH, AND P. S. KOKOSZKA (2003): "GARCH processes: structure and estimation," *Bernoulli,* 9, 201–227.

[5] BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

[6] BOLLERSLEV, T., AND J. WOOLDRIDGE (1992): "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric Reviews*, 11, 143–172.

[7] BOLLERSLEV, T. (2008): "Glossary to ARCH (GARCH)," *CREATES Research Paper.*

[8] BOUGEROL, P., AND N. PICARD (1992a): "Stationarity of GARCH processes and of some nonnegative time series," *Journal of Econometrics*, 52, 115–127.

[9] BOUGEROL, P., AND N. PICARD (1992b): "Strict stationarity of generalized autoregressive processes," *The Annals of Probability,* 20, 1714–1730.

[10] DAVIDSON, J. (1994): "Stochastic limit theory: an introduction for econometricians," 1st edition, *Oxford University Press*, *New York.*

[11] DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A long memory property of stockmarket returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

[12] DUAN, J.-C. (1997): "Augmented GARCH(p, q) process and its diffusion limit," *Journal of Econometrics* ,79, 97–127.

[13] ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica,* 50, 987–1008.

[14] ENGLE, R. F., AND T. BOLLERSLEV (1986): "Modelling the persistence of conditional variances," *Econometric Reviews*, 5, 1–50.

[15] ENGLE, R. F., AND V. K. NG, (1993): "Measuring and testing the impact of news on volatility,".*Journal of Finance,* 48, 1749–1777.

[16] FRANCQ, C., AND J.-M. ZAKOÏAN (2004): "Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes," *Bernoulli*, 10, 605–637.

[17] FRANCQ, C., AND J.-M. ZAKOÏAN (2007): "Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero," *Stochastic Processes and Their Applications*, 117, 1265–1284.

[18] GALLANT, R. (1997): "An introduction to econometric theory," *Princeton University Press, New Jersey.*

[19] GLOSTEN, L. W., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

[20] GONZÁLEZ-RIVERA, G. (1998): "Smooth-transition GARCH models," *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

[21] GRAY, S. F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics,* 42, 27–62.

[22] HAAS, M., S. MITTNIK, AND M. PAOLELLA (2004): "A new approach to Markov-switching GARCH models," *Journal of Financial Econometrics,* 2, 211–250.

[23] HAGERUD, G. (1997): "A new non-linear GARCH model," *EFI Economic Research Institute, Stockholm.*

[24] HAYASHI, F. (2000): "Econometrics," *Princeton University Press, New Jersey.*

[25] HE, C., AND T. TERÄSVIRTA (1999): "Properties of moments of a family of GARCH processes," *Journal of Econometrics,* 92, 173–192.

[26] HENTSCHEL, L. (1995): "All in the family: nesting symmetric and asymmetric GARCH models," *Journal of Financial Economics*, 39, 71–104.

[27] HIGGINS, M. L., AND A. K. BERA (1992): "A class of nonlinear ARCH models," *International Economic Review*, 33, 137–158.

[28] JEANTHEAU, T. (1998): "Strong consistency of estimators for multivariate ARCH models," *Econometric Theory*, 14, 70–86.

[29] KLAASSEN, F. (2002): "Improving GARCH volatility forecasts with regime-switching GARCH," *Empirical Economics*, 27, 363–394.

[30] LANNE, M., AND P. SAIKKONEN (2005): "Non-linear GARCH models for highly persistent volatility," *Econometrics Journal*, 8, 251–276.

[31] LEE, S.-W., AND B. E. HANSEN (1994): "Asymptotic theory for the GARCH (1,1) quasi- maximum likelihood estimator," *Econometric Theory*, 10, 29–52.

[32] LI, C. W., AND W. K. LI (1996): "On a double-threshold autoregressive heteroscedastic time series model," *Journal of Applied Econometrics*, 11, 253–274.

[33] LING, S. AND W. K. LI (1998): "Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors," *The Annals of Statistics*, 26, 84–125.

[34] LUMSDAINE, R. L. (1991): "Essays on time series econometrics," unpublished Ph.D. dissertation, Harvard University.

[35] LUMSDAINE, R. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64, 575–596.

[36] MEITZ, M., AND P. SAIKKONEN (2008a): "Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models," *Econometric Theory*, 24, 1291–1320.

[37] MEITZ, M., AND P. SAIKKONEN (2008b): "Stability of nonlinear AR-GARCH models," *Journal of Time Series Analysis*, 29, 453–475.

[38] MEITZ, M., AND P. SAIKKONEN (2008c): "Parameter estimation in nonlinear AR-GARCH models," *Working paper.*

[39] MILHØJ, A. (1985): "The moment structure of ARCH processes," *Scandinavian Journal of Statistics*, 12, 281–292.

[40] NELSON, D. B. (1990): "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318–334.

[41] NELSON, D. B. (1991): "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica,* 59, 347–370.

[42] NELSON, D. B., AND C. Q. CAO (1992): "Inequality Constraints in the univariate GARCH Model," J*ournal of Business and Economic Statistics*, 10, 229–235.

[43] PANTULA, S. G. (1988): "Estimation of autoregressive models with ARCH errors," *Sankhya B*, 50, 119–138.

[44] PFANZAGL, J. (1969): "On the measurability and consistency of minimum contrast estimates," *Metrika*, 14, 249–272.

[45] PÖTSCHER B. M., AND I. R. PRUCHA (1991): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: consistency and approximation concepts," *Econometric Reviews*, 10, 125–216.

[46] RUDIN, W. (1976): "Principles of mathematical analysis," 3rd Edition, *McGraw-Hill, New York.*

[47] RUDIN, W. (1987): "Real and Complex Analysis," 3rd Edition, *McGraw-Hill, New York.*

[48] SENTANA, E. (1995): "Quadratic ARCH models," *The Review of Economic Studies*, 62, 639–661.

[49] STRAUMANN, D., AND T. MIKOSCH (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach," *Annals of Statistics*, 34, 2449–2495.

[50] WEISS, A. A. (1986): "Asymptotic theory for ARCH models: estimation and testing," *Econometric Theory*, 2, 107–131.

[51] ZAKOÏAN, J.-M. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.