



Munich Personal RePEc Archive

Modelling asymmetric conditional heteroskedasticity in financial asset returns: an extension of Nelson's EGARCH model

Cassim, Lucius

Economics Department, University of Malawi

5 May 2018

Online at <https://mpra.ub.uni-muenchen.de/86615/>

MPRA Paper No. 86615, posted 11 May 2018 12:00 UTC

Modelling asymmetric conditional heteroskedasticity in financial asset returns: an extension of Nelson's EGARCH model

Lucius Cassim¹

May 5 2018

Abstract

Recently, volatility modeling has been a very active and extensive research area in empirical finance and time series econometrics for both academics and practitioners. GARCH models have been the most widely used in this regard. However, GARCH models have been found to have serious limitations empirically among which includes, but not limited to; failure to take into account leverage effect in financial asset returns. As such so many models have been proposed in trying to solve the limitations of the leverage effect in GARCH models two of which are the EGARCH and the TARCH models. The EGARCH model is the most highly used model. It however has its limitations which include, but not limited to; stability conditions in general and existence of unconditional moments in particular depend on the conditional density, failure to capture leverage effect when the parameters are of the same signs, assuming independence of the innovations, lack of asymptotic theory for its estimators et cetera. This paper therefore is geared at extending/improving on the EGARCH model by taking into account the said empirical limitations. The main objective of this paper therefore is to develop a volatility model that solves the problems faced by the exponential GARCH model. Using the Quasi-maximum likelihood estimation technique coupled with martingale techniques, while relaxing the independence assumption of the innovations; the paper has shown that the proposed asymmetric volatility model not only provides strongly consistent estimators but also provides asymptotically efficient estimators.

Key Words: GARCH, TARCH, EGARCH, Quasi Maximum Likelihood Estimation, Martingale

¹ Economics Department, Chancellor College, University of Malawi, P.O.Box 280, Zomba, Malawi : Email: luciuscassim@gmail.com or lcassim@cc.ac.mw

1. Introduction

1.1 The ARCH and GARCH models

A uni-variate ARCH model based on Engle (1982) is given by set equations given below;

$$(1) \begin{cases} \varepsilon_t = \sigma_t z_t \\ z_t \sim i.i.d \\ E(z_t) = 0, Var(z_t) = 1 \\ \sigma_t^2 = \sigma^2(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, t, x, \beta) \end{cases}$$

According to Nelson (1991) the most widely used specification for σ_t^2 are the linear ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986) respectively;

$$(2) \begin{cases} \sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j z_{t-j}^2 \sigma_{t-j}^2 \\ \sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j z_{t-j}^2 \sigma_{t-j}^2 + \sum_{i=1}^q \alpha_i \sigma_{t-i}^2 \end{cases}$$

where ω, α_i are non-negative. According to Nelson (1991) and Rossi (2004), substituting recursively for the $\alpha_i \sigma_{t-i}^2$ terms imply that the GARCH model can be written as;

$$(3) \begin{cases} \sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j z_{t-j}^2 \sigma_{t-j}^2 + \sum_{i=1}^q \alpha_i \sigma_{t-i}^2 = \omega^* + \sum_{k=i}^{+\infty} \phi_k z_{t-k}^2 \sigma_{t-k}^2 \\ \omega^* = \frac{\omega}{1 - \sum_{i=1}^q L^i \alpha_i} \end{cases}$$

According to Nelson(1991); by setting conditional variance equal to a constant plus a weighted average residuals, GARCH models elegantly capture the volatility clustering in asset returns first noted by Mandelbrot(1963);...large changes tend to be followed by large changes of either sign and small changes by small changes. This is one of the most important features of GARCH models that have made it very attractive in the empirical literature.

1.2 Limitations of the GARCH models

However there are limitations of GARCH models that have been noticed in literature, ably summarized by Nelson (1991);

- i. Stock returns are negatively correlated with changes in volatility e.g. either volatility leads to rise in response to "bad news"(excess returns lower than expected) and to fall in response to "good news"(excess returns higher than expected). GARCH models, however, assume that only the magnitude and not the positivity or negativity of unanticipated excess returns determines volatility
- ii. Non negativity constraints on parameters. These constraints imply that increasing z_t in any period increases volatility ruling out random oscillatory behavior in the volatility process.
- iii. Persistence; in GARCH models, shocks may persist in one norm and die out in another, so the conditional moments of GARCH may explode even when the process is strictly stationary and ergodic.

1.3 Models with improvements on GARCH limitations

Due to the empirical limitations faced by GARCH models outlined above, so many models have since been proposed in the literature. Here we review a few of them. As noted above there is a long tradition in finance stock return volatility are negatively correlated with stock returns. The explanation for this phenomenon is based on leverage. A drop in the value of the stock (negative return) increases financial leverage, which makes the stock riskier and increases its volatility. The news has asymmetric effects on volatility. In the asymmetric volatility models good news and bad news have different predictability for future volatility. As noted, the GARCH models do not capture the leverage effect. It should therefore be pointed out here that most of the models that have been proposed in the literature try to improve on this failure of the GARCH models apart from improving on the non-negativity restrictions of the GARCH models.

1.3.1 THE Exponential GARCH model

The Exponential GARCH introduced by Nelson (1991) is the most popular of the models that have proposed improvements on the limitations of the GARCH model outlined above. To deal with the non-negativity issue the model took natural log of the volatility. This technically means that whatever the values assumed by the parameters the implied volatility can never be negative. On top of that, in dealing with the asymmetric relation between stock returns, he made the g_t sum of z_t and $|z_t|E_t$ such that the E-GARCH model proposed was;

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^p \alpha_i \ln(\sigma_{t-i}^2) + \sum_{j=1}^q \beta_j [\phi z_{t-j} + \varphi(|z_{t-j}| - E|z_{t-j}|)]$$

This, according to Rossi (2004), can be written as;

$$\begin{aligned} \left[1 - \sum_{i=1}^p \alpha_i L^i\right] \ln(\sigma_t^2) &= \omega + \sum_{j=1}^q \beta_j [\phi z_{t-j} + \varphi(|z_{t-j}| - E|z_{t-j}|)] \\ \Rightarrow \ln(\sigma_t^2) &= \left[1 - \sum_{i=1}^p \alpha_i L^i\right]^{-1} \omega + \left[1 - \sum_{i=1}^p \alpha_i L^i\right]^{-1} \left[\sum_{j=1}^q \alpha_j L^j\right] g(z_t) \\ \therefore \begin{cases} \ln(\sigma_t^2) &= \omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \\ \omega^* &= \left[1 - \sum_{i=1}^p \alpha_i L^i\right]^{-1} \omega, g(z_{t-k}) = \phi z_{t-k} + \varphi(|z_{t-k}| - E|z_{t-k}|) \end{cases} \end{aligned}$$

The model, therefore, proposed by Nelson (1991) was;

$$\begin{cases} \ln(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) & ; \beta_1 = 1, -\infty \leq \beta \leq \infty, \{\alpha_t\}_t = -\infty, \infty \\ g(z_t) = \phi z_t + \gamma[|z_t| - E|z_t|] & \{g(z_t)\}_t = -\infty, \infty \end{cases}$$

It should be clear from here that over $0 \leq z_t \leq \infty$ $g(z_t) = \phi z_t + \gamma[|z_t| - E|z_t|]$ and $g'(z_t) = -\theta + \gamma$. This confirms that the E-GARCH model solves the asymmetric issue not solved by

the traditional GARCH model. It should be pointed out here that by introducing the log the E-

GARCH model solves the persistence issue since $\ln(\sigma_t^2)$ was proved (see Nelson (1991)) to be

strictly stationary and ergodic. In fact in EGARCH (p, q) model $\ln(\sigma_t^2)$ is a linear process, and its stationarity (covariance or strict) and ergodicity are easily checked. Given that $\phi \neq 0$ or $\varphi \neq 0$, then

$|\ln(\sigma_t^2) - \omega^*| < +\infty$ a.s when $\sum_{k=1}^{+\infty} \beta_k^2 < +\infty$ follows from the independence and finite variance of the

$g(z_t)$ and from Billingsley (1986, Theorem 22.6). From this we have that $\left|\ln\left(\frac{\sigma_t^2}{\exp(\omega^*)}\right)\right| < +\infty$ a.s

and $\left|\left(\frac{\sigma_t^2}{\exp(\omega^*)}\right)\right| < +\infty$ a.s $\{\exp(-\omega^*)\sigma_t^2\}, \{\exp(-\omega^*/2)\varepsilon_t\}$, where $\varepsilon_t = z_t\sigma_t$, z_t is i.i.d., ergodic and

strictly stationary. For all $t \in \mathbb{N}$ $E[\ln(\sigma_t^2) - \omega^*] = 0$ and $Var[\ln(\sigma_t^2) - \omega^*] = Var(g(z_t)) \sum_{k=1}^{+\infty} \beta_k^2$. Since

$Var(g(z_t))$ is finite and the distribution of $[\ln(\sigma_t^2) - \omega^*]$ is independent of t , the first two moments

of $[\ln(\sigma_t^2) - \omega^*]$ are finite and time invariant, so $[\ln(\sigma_t^2) - \omega^*]$ is covariance stationary if

$$\sum_{k=1}^{+\infty} \beta_k^2 < +\infty. \text{ If } \sum_{k=1}^{+\infty} \beta_k^2 = +\infty \text{ then } [\ln(\sigma_t^2) - \omega^*] = +\infty, \text{ almost surely.}$$

In essence, therefore, an EGARCH model does a very good job of trying to solve limitations (1)-(3) outlined above. To be more specific, there are four interesting features to notice about the EGARCH model:

- i. The equation for the conditional variance is in log-linear form. Regardless of the magnitude of $\ln(\sigma_t^2)$, the implied value of σ_t^2 can never be negative. Hence, it is permissible for the coefficients to be negative.
- ii. Instead of using the value of ε_{t-1} , the EGARCH model uses the level of standardized value of ε_{t-1} [i.e., ε_{t-1} divided by σ_t]. Nelson argues that this standardization allows for a more natural interpretation of the size and persistence of shocks. After all, the standardized value of ε_{t-1} is a unit-free measure
- iii. The E-GARCH model solves the asymmetric issue not solved by the traditional GARCH model. It should be clear from here that over $0 \leq z_t \leq \infty$; $g(z_t) = \theta + \gamma z_t$ and $g'(z_t) = \gamma$. This confirms that negative shocks have different impacts on volatility of asset returns compared to positive shocks of the same magnitude.
- iv. Persistence; it has been pointed out above that by introducing the log, the E-GARCH model solves the persistence issue since the conditional moments of $\ln(\sigma_t^2)$ are finite and do not explode.

1.3.1.1 Limitations of the Nelson's E-GARCH

Despite the nice properties of the E-GARCH model over and above the traditional GARCH models, the following limitations can be directed at it

- i. In contrast with ordinary GARCH models, as ably echoed by Sucarrat and Escibano (2010), in exponential GARCH models stability conditions in general and existence of unconditional moments in particular depend on the conditional density. Interestingly, for example, the Nelson's EGARCH model is not stable for t-distributed innovations, see Nelson (1991, p.365). As explained by Sucarrat and Escibano (2010), this is a serious shortcoming since the t-distribution is the preferred choice by practitioners among the densities that are more fat-tailed than the normal density.

ii. It does not entirely capture leverage effect. As noted above if $|z_t| \geq E|z_t|$ then $g'(z_t) = -\theta + \gamma$. This captures "leverage effect" only if $\theta < 0$ since that would mean that that $g'(z_t)$ will be greater than $g'(z_t)|_{|z_t| < E|z_t|}$ which is exactly what leverage effect is. But if $\theta \geq 0$ then obviously $g'(z_t)|_{|z_t| < E|z_t|}$ will be less than $g'(z_t)$ which is the opposite of what leverage effect states. This means that the E-GARCH model does not capture leverage effect when the parameters are positive. This means that there is need for an all-encompassing model that is able to capture leverage effect regardless of the signs of the parameters.

iii. Use of Maximum Likelihood Estimation(MLE) technique; the MLE proposed by Nelson(1991) exposes the estimators derived to high risk of inconsistency in the event that the model is not correctly specified. One may recall that the conventional ML method assumes that the postulated likelihood function is specified correct, so that specification errors are "assumed away." It is conceivable that specifying a likelihood function, while being more general and more flexible than specifying a function for conditional mean, is more likely to result in specification errors. And if that happens, the estimators derived are more likely to be inconsistent.

This suggests the need for an estimation technique of the EGARCH model that allows possible misspecification of the likelihood function.

iv. No asymptotic theory for the derived Maximum Likelihood estimates was developed. The developed MLE was neither proved for consistency nor proved for asymptotic normality. Nelson (1991) acknowledged this by stating that;

".....in the remainder of this paper we assume (as is the usual practice of researchers using GARCH models) that the maximum likelihood estimator is consistent and asymptotically normal".

This suggests that there is need for asymptotic theory of the E-GARCH estimators.

v. The assumption that z_t is *i.i.d* ; It's common knowledge currently that innovations in asset returns may not be statistically independent .It is also well documented (see Holly, 2010; Chung, 2012; Drost & Klassenn, 1996; Engle & Gonzale-Rivera, 1991) that if we assume statistical independence of innovations when in fact they are dependent we run high risk of misspecifying the likelihood function which then leads to high risk of getting inconsistent estimates. Empirical regularities of time series financial returns show that the innovations are dependent and not independent. The following characteristics are frequently observed in

financial data (see Holly, 2010; Chung, 2012; Drost & Klassenn, 1996; Engle & Gonzale-Rivera, 1991). The first is volatility clustering. This is where large changes tend to be followed by large changes and small changes tend to be followed by small changes. Second is that squared returns exhibit serial correlation whereas little or no serial dependence can be detected in the return series itself. In addition, financial returns exhibit fading memory i.e., distant innovations have little effect on financial returns compared to recent innovations. This means that the i.i.d assumption is not correct when it comes to financial time series. That being said, one may wonder as to what exactly is the problem with continuing with the statistical independence assumption when in fact the innovations are statistically dependent. The statistical independence assumption is very critical when the volatility function is allowed to be time dependent. This is because it ensures that the parameters entering the conditional mean function are time-independent (Dahl & Levine, 2010). Dahl and Levine (2010) argued that if the conditional mean function is estimated assuming time invariant parameters, when they are time variant, its estimators will be inconsistent and the effect of this misspecification will carry over into the volatility estimation. Technically, what we are saying here is that the statistical independence assumption (within the i.i.d assumption) is not appropriate in time series financial data. If we continue making it, when in fact the innovations are statistically dependent, we are bound to get inconsistent estimators. Therefore by making the *i.i.d* assumption, the E-GARCH estimates are at a high risk of being inconsistent. This means that there is a need to relax the independence assumption (within the i.i.d assumption) in E-GARCH model.

1.3.2 THE THRESHOLD GARCH (TARCH) MODEL

Glosten et al (1994) showed how to allow the effects of good and bad news to have different effects on volatility Enders (2004). They considered the threshold-GARCH (TARCH) process;

$$(4) \quad \left\{ \begin{array}{l} y_t = y_0 + \varepsilon_t, \varepsilon_t = \sigma_t z_t \\ z_t \sim i.i.d \\ E(z_t) = 0, Var(z_t) = 1 \\ \sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^q \lambda_i d_{t-i} \varepsilon_{t-i}^2 + \sum_{k=1}^w \beta_k \sigma_{t-k}^2 \\ d_{t-i} = \mathbf{I}(L^i(\varepsilon_t) < 0) \end{array} \right.$$

One clearly notices that the intuition behind the TARCH model is that positive values of ε_{t-i} are associated with a zero value of d_{t-i} . Hence, if $\varepsilon_{t-i} \geq 0$, the effect of an ε_{t-i} shock on σ_t^2 is (α_j) . When $\varepsilon_{t-i} < 0$, and the effect of an ε_{t-i} shock on σ_t^2 is $(\alpha_j + \lambda_i)$. If

$\lambda_i > 0$, negative shocks will have larger effects on volatility than positive shocks. One of the advantages of the TARCH model therefore is that it captures leverage effect. However, one notices that the following criticisms can be directed at the T-GARCH model;

- i. It still places the non-negativity constraints on the parameters, just like the GARCH
- ii. It assumes that the innovations are independent. As ably explained above, this is a serious problem.

1.3.3 The GJR-GARCH Model

The GJR-GARCH (p, q) model is another asymmetric GARCH model proposed by Glosten et al (1993). The generalized form of the GJRGARCH (p, q) model is given in the following form:

$$(5) \quad \left\{ \begin{array}{l} y_t = y_0 + \varepsilon_t, \varepsilon_t = \sigma_t z_t \\ z_t \sim i.i.d \\ E(z_t) = 0, Var(z_t) = 1 \\ \sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2 + \gamma_i d_{t-i} \varepsilon_{t-i}^2 \\ d_{t-i} = I(L^i(\varepsilon_t) < 0) \end{array} \right.$$

Obviously this model has almost the same properties as the TARCH model explained above.

It should be mentioned here that the EGARCH is the most highly used among the asymmetric volatility models in the literature due to its desirable properties. However, from the limitations of the Exponential Model, four gaps are clear;

- i. There is a need to extend the traditional E-GARCH model so that it captures leverage effect regardless of the signs of the parameters
- ii. There is need for an estimation technique of the E-GARCH model that allows possible misspecification of the likelihood function;
- iii. There is need to develop an estimation mechanism of the E-GARCH model that relaxes the serial independence assumption;
- iv. There is a need to develop an asymptotic theory for the estimators of the E-GARCH model.

This paper, therefore, aims at filling these gaps to come up with a relatively better asymmetric volatility model than the existing models in the literature. This will be partly done by combining the features of the TARCH and the EGARCH models to come up with a new model. We propose a model that captures leverage effect while taking into account the fact that the parameters can also be

positive and while relaxing the serial independence assumption. We also propose quasi-maximum likelihood estimation technique in estimating the proposed model. We then show that the derived quasi-maximum likelihood estimator is not only consistent but also efficient

.

1.5 Objectives of the study

The main objective of this paper is to develop a volatility model that solves the problems faced by the exponential GARCH model. Specifically the following objectives shall be pursued;

- i. Showing that the QMLE of the proposed E-GARCH model is consistent
- ii. Showing that the QMLE of the proposed E-GARCH model is asymptotically efficient

2. The proposed exponential GARCH model

2.1 The Model

We therefore assume the following observed model

$$(6) \quad \begin{cases} y_t = y(x_t, \delta_0) + \varepsilon_{0t}, \varepsilon_{0t} = \sigma_{0t} z_t \\ z_t \sim i.i.d \\ E(z_t) = 0, Var(z_t) = 1 \\ \ln(\sigma_{0t}^2) = \sigma^2(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, t, x, \delta_0, \beta_0) = \omega_0^* + \sum_{k=1}^{+\infty} \beta_{0k} g(z_{t-k}) \\ g(z_t) = \theta_0 |z_t| + \gamma_0 I_t [|z_t| - E|z_t|] \quad \{g(z_t)\}_t = -\infty, \infty \text{ \& } I_t = \mathbf{I}(L(z_t) < 0) \end{cases}$$

Along with Kouassi et al (2017) and Lee and Hansen (1994), one also considers the conditional variance when infinite past values of observations are available;

$$(7) \quad \begin{cases} y_t = y(x_t, \delta) + \varepsilon_t, \varepsilon_t = \sigma_t z_t \\ z_t \sim i.i.d \\ E(z_t) = 0, Var(z_t) = 1 \\ \ln(\sigma_t^2) = \sigma^2(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, t, x, \delta, \beta) = \omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \\ g(z_t) = \theta I_t |z_t| + \gamma [|z_t| - E|z_t|] \quad \{g(z_t)\}_t = -\infty, \infty \text{ \& } I_t = \mathbf{I}(L(z_t) < 0) \end{cases}$$

Which includes the unknown parameters as well as the true disturbances but over an infinitely long period; σ_{0t}^2 and σ_t^2 are called observed and unobserved conditional variance processes, respectively.

One notices here that we have taken natural log of the volatility, just like in the EGARCH model, to avoid imposing non-negativity restriction on the parameters. We have also introduced the dummy variable, just like in TARCH model, to take care of the leverage effect. So, in essence we have built this model by combining ideas from the EGARCH and the TARCH models. However, unlike both the TARCH and the EGARCH models, we are making an assumption here that the innovations are identically and dependently distributed (i.d.d) and not identically and independently distributed (i.i.d) (i.e. $z_t \sim i.d.d$).

2.2 Assumptions

This section outlines some of the assumptions made in this analysis. They include both the regularity assumptions (i.e. standard assumptions) that are necessary for identifiability, stationarity etc and some additional assumptions necessary for our analysis. To achieve the objectives, we make (following Kouassi et al, 2017; Choi, 2004; Chung, 2012; Posedel, 2005 and Hansen, 2006) the following assumptions:

Assumption one (A1): The innovations are martingale differences

This is one of very important assumptions in this analysis. Assuming that our innovations are dependent, we cannot be able to use convergence in probability since this needs the i.i.d assumption. However, if we assume that the innovations are martingale differences, then we will be able to show almost sure convergence under dependent innovations which will ultimately help us prove convergence in probability (Rao, 1973; Stout, 1974).

Assumption two (A2): Differentiability and continuity

$$\forall_{x_t, \theta} y(x_t, \beta) \text{ and the likelihood functions are both continuous and differentiable}$$

This assumption ensures the positivity of both conditional and unconditional variances.

Assumption five (A5): Existence of the moments of innovations

$$\text{For some } \delta > 0, \exists S_\delta < \infty, \text{ s.t. } E(z_t^{2+\delta}) \leq S_\delta < \infty$$

This assumption is simply saying that all moments (i.e. the mean, the variance etc.) of the innovations do exist.

Assumption six (A6): ergodicity and distribution of innovations

z_t are ergodic process that belong to a probability law that belongs to the quadratic exponential family

Assumption seven (A7): Uniform boundness of the likelihood function

$$y(x_t, \delta) = O(1)$$

This means that $y(x_t, \delta_0)$ is bounded uniformly. This ensures that our likelihood function is bounded as well. The boundness of the likelihood function is important in establishing estimator consistency.

3. Estimation Techniques

3.1 The Quasi Maximum Likelihood Function

For us to be able to derive the QMLE we need a likelihood function, the Quasi Maximum Likelihood Function (QMLF, here-in-after). However, since we have both the observed and unobserved models, we will also have observed and unobserved likelihood functions. We will firstly provide the unobserved likelihood function after which we will provide the observed likelihood.

$$\begin{aligned}
 \text{Unobserved Likelihood : } L_T(\theta) &= \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) = \sum_{t=1}^T l_t(\theta) \\
 &= \sum_{t=1}^T \left[\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right] \\
 &= \sum_{t=1}^T \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) y_t^i + \sum_{t=1}^T \chi_3(y_t)
 \end{aligned}$$

The observed log-likelihood $L_{0T}(\theta)$ is derived analogously as;

$$\begin{aligned}
 L_{0T}(\theta) &= \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \\
 &= \sum_{t=1}^T l_{0t}(\theta) = \sum_{t=1}^T \left[\sum_{i=0}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right] \\
 &= \sum_{t=1}^T \sum_{i=0}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i + \sum_{t=1}^T \chi_3(y_t)
 \end{aligned}$$

3.2 Lemmas

In order to achieve our objectives, we will provide and prove lemmas that will enable us achieve the objectives. In this section we will mainly provide these lemmas and prove them.. To achieve our objectives, we prove the following standard lemmas (see Kouassi, 2015; Choi, 2004; Chung, 2012; Hansen & Lunde, 2001; Holly & Montifort, 2010; Engle R, 1982; Bollerslev T, 1986; Rossi, 2004; Buhlman & McNeil, 2000).

3.2.1 Lemma one

$$\boxed{\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) - \frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right|^p \rightarrow 0}$$

The following lemma technically implies that the observed likelihood function which is not stationary is asymptotically approximated by the unobserved likelihood function such that we can

ignore the differences that may exist between the observed and the unobserved likelihood functions.

Proof: Check Appendix A1

3.2.2 Lemma two

The Processes $\lambda_t, l_t(\theta)$ and their derivatives are strictly stationary and ergodic.

The necessity of this lemma is therefore that it will help us apply the strict law of large numbers on expressions that are functions of λ_t , for example the likelihood functions. If we can be able to apply the strong law of large numbers then we can be able to show almost sure convergence² which ultimately implies convergence in probability (Chung, 2012; Rao, 1973). Therefore, this lemma is very important.

Proof: Check Appendix

A23.2.3 Lemma three

$$\frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \xrightarrow{p} E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)]$$

This is to say that our criterion to maximize the likelihood function converges in probability to a non-stochastic function $E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)]$. This is a necessary condition for the convergence of QMLE.

Proof: Check Appendix A3

3.2.4 Lemma four

$$E\left|\nabla l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)\right| < \infty, \& E\left|\nabla^2 l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)\right| < \infty$$

What this lemma is saying is that the absolute score function together with its derivatives is bounded. This is a sufficient condition for convergence of QMLE (Hood & Koopman, 1953).

Proof: Check Appendix A4

3.2.5 Lemma five

$$\nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) = 0 = \nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla \sum_{t=0}^T l_t(\theta^*), \text{ where } \theta^* \text{ is between } \tilde{\theta}_{SEM} \& \theta_0$$

²Check Appendix B for more information on basics of statistical convergence.

What this lemma is saying is that, we can analyse the asymptotic behavior of the likelihood function by simply analyzing the behavior of the right hand side of lemma 5 above.

Proof: Check Appendix A5

3.2.6 Lemma six

$$\boxed{\sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \left| \nabla^2 \sum_{t=0}^T l_t(\theta^*) - E \left(\nabla^2 \sum_{t=0}^T l_t(\theta^*) \right) \right| \xrightarrow{p} 0}$$

This Lemma is saying that $\nabla^2 \sum_{t=0}^T l_t(\theta^*)$ obeys the weak uniform law of large numbers

Proof: Check Appendix A6

3.2.7 Lemma seven

$$\boxed{\sqrt{T}(\tilde{\theta}_{SEM} - \theta_0) = - \left[E \left(\nabla^2 \sum_{t=0}^T l_t(\theta) \right) \right]^{-1} \sqrt{T} \nabla \sum_{t=0}^T l_t(\theta_0)}$$

Proof: Check Appendix A7

4. Theoretical Results and Discussion

4.1 Introduction

In this section we present the main results of our analysis. Given the assumptions and the lemmas above, we show in this section how the proposed E-GARCH model achieves the stated objectives.

4.2 The Leverage effect issue

Given the model;

$$\left\{ \begin{array}{l} y_t = y(x_t, \delta) + \varepsilon_t, \varepsilon_t = \sigma_t z_t \\ z_t \sim i.d.d \\ E(z_t) = 0, Var(z_t) = 1 \\ \ln(\sigma_t^2) = \sigma^2(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, t, x, \delta, \beta) = \omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \\ g(z_t) = \theta I_t |z_t| + \gamma_0 [|z_t| - E|z_t|] \quad \{g(z_t)\}_t = -\infty, \infty \text{ \& } I_t = \mathbf{I}(L(z_t) < 0) \end{array} \right.$$

One clearly notices that; for the range $-\infty < z_t \leq 0$, $g(z_t) = (\theta)z_t + \gamma [|z_t| - E|z_t|] \Rightarrow g'(z_t) = \theta + \gamma$, for the range $0 \leq z_t < +\infty$, $g(z_t) = \theta |z_t| \Rightarrow g'(z_t) = \theta$. This means that, assuming that the parameters are of the same signs; negative shock has a higher impact on volatility than a positive shock of the same magnitude which exactly what leverage effect says. This shows that the proposed extension of the E-GARCH model captures leverage effect irrespective of the sign of the parameters.

4.3 The estimator

Having shown that the proposed model is indeed capturing leverage effect, the question then is what estimation technique can one use to estimate it?. We propose use of quasi-maximum likelihood estimation technique (QMLE, hereafter) as opposed to Maximum Likelihood Estimation technique (MLE hereafter) that Nelson (1991) proposed. The QML method is essentially the same as the ML method usually seen in statistics and econometrics textbooks. A key difference between these two methods is that the former allows for possible misspecification of the likelihood function. It is conceivable that specifying a likelihood function, while being more general and more flexible than specifying a function for conditional mean, is more likely to result in specification errors. How to draw statistical inferences under potential model misspecification is thus a major concern of the QML method. By contrast, the conventional ML method assumes that the postulated likelihood function is specified correct, so that specification errors are “assumed away.” As such, the results in the ML method are just special cases of the QML method. One notices that by using QMLE we are reducing the risk of getting inconsistent estimators in the event of wrong likelihood function. It uses the Kullback-Leiber Information Criterion. From equation (2.19), it was noted that in the spirit of

Kullback-Leiber Information Criterion (KLIC) the quasi-maximum likelihood estimator, $\hat{\theta}_{QMLE}$, is given as shown below;

$$\hat{\theta}_{QMLE} = \arg \min KLIC = \arg \min \sum_{t=1}^T \ln \left(\frac{\zeta(\varepsilon)}{\zeta(\varepsilon | \theta)} \right) \zeta(\varepsilon) \partial \varepsilon = E[\ln \zeta(\varepsilon | \theta)]$$

It can be noted here that equation (4.01) implies that the quasi-maximum likelihood estimator is the one that minimizes the KLIC. But, minimizing the KLIC is the same as maximizing the unobserved function.

Therefore, using lemma 1, the semi parametric GARCH (1, 1) estimator, $\tilde{\theta}_{SEM}$ is the one that maximizes the unobserved likelihood function as given in equation (4.02). That is,

$$\begin{aligned} \tilde{\theta}_{SEM} &= \arg \max \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \\ \tilde{\theta}_{SEM} &= \arg \max \sum_{t=1}^T \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \end{aligned}$$

So the semi-parametric E-GARCH under serially dependent quadratic exponential innovations is given as;

$$\tilde{\theta}_{SEM} = \arg \max \sum_{t=1}^T \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right)$$

4.3 Theorem one

Given the above assumptions and lemmas,

$$\boxed{\tilde{\theta}_{SEM} \rightarrow \theta_0, a.s \Rightarrow \tilde{\theta}_{SEM} = \theta_0 + o_p(1)}$$

This theorem is saying that, given all the assumptions outlined above and all the lemmas proved above, the estimator is consistent. In other words, the estimator converges almost surely to the true population parameter as we increase the sample size indefinitely.

Proof: Check Appendix B1

4.4 Theorem two

$$\boxed{E[\nabla^2(L_T(\theta_0))] = -Var(\sqrt{T}(L_T(\theta_0)))}$$

This equality is called *information matrix equality*. It is a very useful result in the quasi-maximum likelihood theory. This equality shows that when the specification is correct up to certain extent, the information matrix $Var(\sqrt{T}(L_T(\theta_0)))$ is the same as the negative of the expected Hessian matrix $-E[\nabla^2(L_T(\theta_0))]$ when evaluated at $\theta = \theta_0$, i.e., $E[\nabla^2(L_T(\theta_0))] + Var(\sqrt{T}(L_T(\theta_0))) = 0$. In other words, this shows that the QMLE achieves the Cramer-Rao lower bound asymptotically (i.e. the QMLE is asymptotically efficient)

Proof: Check Appendix B2

Conclusion

This paper has proposed an asymmetric volatility model that is improving on the limitations of some of the widely used asymmetric volatility models in the literature. The paper also proposes an estimation technique one can use in estimating the proposed model. Having done that the paper has went ahead to show that if one uses the QMLE technique on the proposed model, the derived estimators are not only strongly consistent but also asymptotically efficient.

REFERENCES

- Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Havard University Press.
- Andersen, T. G. (1996). GMM estimation of stochastic volatility models:A monte Carlo Study. *Journal of Business and Economic Statistics*,48, 328-352.
- Avran, F. (1988). Weak Convergence of the variations,iterated integrals and Doleans-Dade exponentials of sequences of semi-martingales. *Anal of probability* 16, 246-250.
- Baillie, R. T., & Bollerslev, T. (1987). The message in Daily Exchange Rates: A conditional Variance Tale. *Econometrica*.
- Bollerslev. (1987). A conditional Heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*,69, 542-547.
- Bollerslev, T. (1986). Generalized Autoregressive Heteroskedasticity. *Journal of Econometrics*, 307-327.
- Bollersslev, J., & Woodridge, J. (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*,11(2), 143-172.
- Buhlman, P., & McNeil, A. J. (2000). *Non-parametric GARCH Models*. Zurich, Switzerland: Seminar Fur Statistik:CH-8092.
- Cameron, C. A., & Trivedi, P. (2005). *Microeconometrics:Methods and Applications*. New York,USA: Cambrdge University Press.
- Choi, E. J. (2004). Estimation of Stochastic Volatility Models by Simulated Maximum Likelihood Method. *University of Waterloo*.
- Chung, S. S. (2012). A class of non-parametric volatility models:Application to financial time series. *Journal of Econometrics*.

- Dahl, C. M., & Levine, M. (2010). Non-parametric estimation of volatility models under serially dependent innovations. *Econometrica*.
- Davidson, J. (2000). *Econometric Theory*. Blackwel: Oxford University Press.
- Davidson, R., & Mackinnon, J. (1993). *Estimation and inference in Econometrics*. London: Oxford University Press.
- Drost, F. C., & Klasssen, C. (1996). Efficient estimation in Semi-parametric GARCH Models. *Discussion paper;vol(1996-38),Tilburg*.
- Duan, J. (1997). Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics*,79(1), 97-127.
- Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with estimation of the variance of U.K inflation. *Econometrica*, 987-1008.
- Engle, R. F., & Gonzale-Rivera, G. (October,1991). Semiparametric ARCH Models. *Journal of Business and Economic Statistics*, 9(4), 345-359.
- Engle, R. F., & Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of finance*,48, 1747-1778.
- Fan, & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting:variable bandwidth and spatial adaptation. *Journal of Royal Statistical Society,B*,57, 371-394.
- Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*,85, 645-660.
- Gallant, A. R., & Hsieh, D. (1989). Fitting a Recalcitrant series:The Pound/Dollar Exchange Rate,1974-83. *Econometrica*.
- Geweke, J. (1986). Modelling Persistence in Conditional Variances: A Comment. *Econometric Review*,5, 56-61.

- Glosten, L. R., & Runkle, D. (1993). On the relation between the expected value and volatility of the nominal excess return of stocks. *Journal of Finance*, 48, 1779-1801.
- Gourieroux, C., & Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory, 52(1). *Econometrica*, 681-700.
- Haafner, C. M. (2003). Analytical quasi maximum likelihood inference in BEKK-GARCH models. *Econometric Institution, Erasmus University, Rotterdam*.
- Hafner, C. M., & Rombonts, J. (2002). Semiparametric multivariate GARCH models . *Discussion paper, 2002/XX, CORE*.
- Hansen, B. C. (2006). *Econometrics*. New York: Cambridge University Press.
- Hansen, P., & Heyde, C. (1980). *Martingale limit theory and its applications*. New York: Academic Press.
- Hansen, R. P., & Lunde, A. (2001). A comparison of volatility models: Does anything beat GARCH(1,1)? *Centre for analytical finance: University of AARHUS*.
- Hentshel, L. (1995). All in the family: Nesting Symetric and Asymetric GARCH Models. *Journal of Financial Economics*, 39, 71-104.
- Herwartz, H. (2004). Conditional Heteroskedasticity. In H. Lutkepoh, & M. Kratzig (Eds.) *Themes in Modern Econometrics*, pp. 197-220.
- Higging, M. L., & Bera, A. (1992). A Class of non-linear ARCH Models. *International Economic Review*.
- Holly. (2009). *Modelling Risk using fourth order Pseudo Maximum Likelihood Methods*. University of Lausanne, Institute of Healthy Economics.
- Holly, A., & Montifort, A. (2010). Fourth Order Pseudo Maximum Likelihood Methods. *Econometrica*.
- Holly, A., & Pentsak, Y. (2004). *Maximum Likelihood Estimation of the Conditional mean $E(Y|X)$ for Skewed Dependent variables in Fourth-Parameter families of Distributions*. Technical Report, University of Lausanne, Institute of Healthy Economics and Management.

- Hood, W., & Koopman, T. (1953). The estimation of simultaneous linear economic relationships. *Econometric Methods*.
- Ibragimov, R., & Philips, P. (2010). Regression asymptotics using martingale convergence. *Yale University press*.
- Kouassi, E. (2015). Consistency of Pseudomaximum likelihood estimation in ARCH(1) under dependent innovations. *Working paper*.
- Linton, O., & Mammen, E. (May, 2003). Estimating Semi-parametric ARCH(∞) models by kernel smoothing methods. *Discussion paper, No: EM/03/453*.
- McCullagh, P. (1994). Exponential mixtures and quadratic exponential families. *Biometrika*, 81(4), 721-729.
- Nadaraya, E. (1964). "On Estimating Regression". The Theory of Probability and its Applications. *Econometrica*, 141-2.
- Nielsen, B. (1978). *Information and exponential families in statistical theory*. New York: Wiley.
- Pantula, S. G. (1986). Modelling Persistence in Conditional Variances: A comment. *Econometric Review*, 5, 71-74.
- Posedel, P. (2005). Properties and estimation of GARCH(1,1) model. *Metodoloski Zvezki*, 2, 243-257.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: John Wiley & Sons.
- Rossi, E. (2004, march). A note on GARCH models. *Working paper*.
- Sentana, C. (1995). Quadratic ARCH Models. *Review of Economic Studies*, 62(4), 639-661.
- Sherphard, N. (2008). Statistical aspects of ARCH and Stochastic Volatility. In D. R. Cox, D. Hinkly, & O. Barndorff (Eds) *Time series models in Econometrics, Finance and other fields. Monographs on statistics and Applied probability*, 65, pp. 1-65.

- Sousi, P. (2013). *Advanced Probability*. New York: Cambridge University Press.
- Stout, W. F. (1974). *Almost Sure Convergence*. New York: Academic Press.
- Su, L., Ullah, A., & Mashra, S. (2011). Non-parametric and semi-parametric volatility models: specification, estimation and testing. *Econometrica*.
- Tapia, R. A., & Thompson, J. (n.d.). Nonparametric Probability Density Estimation.
- Taylor, S. (1986). *Modelling Financial Time Series*. John Wiley & Sons.
- Tsay, R. S. (2010). *Analysis of time series*. Willey.
- Tschiernig, R. (2004). Non-parametric Econometrics. In H. Lutkepoh, & M. Kratzig (Eds) *Themes in Econometrics*, pp. 243-289.
- Watson, G. (1964). Smooth Regression Analysis. *The Indian Journal of Statistics*, 359-372.
- Weiss, A. A. (1986). Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory*, 2, 107-131.
- Williams, D. (1991). *Probability with martingales*. New York: Cambridge University Press.
- Yang, L., & Song, Q. (2012). Efficient Semi-parametric GARCH Modelling of Financial Volatility. *Statistica Sinica*, 22, 249-270.
- Zakoian, M. J. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18, 931-955.

APPENDIX A: PROOF OF THE LEMMAS

Appendix A1: Proof of Lemma 1: Technically for us to prove this lemma we just have to show that

$$\lim_{T \rightarrow \infty} \text{prob} \left(\left| \frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) - \frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right| > \varepsilon \right) = 0$$

Using the notation used in this study, the above expression can also be written as;

$$\lim_{T \rightarrow \infty} \text{prob} \left(\left| \frac{1}{T} \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) - \frac{1}{T} \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right| > \varepsilon \right) = 0$$

The unobserved single observation likelihood function is given as;

$$l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) = \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t)$$

Likewise, observed single observation likelihood function is given as;

$$l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) = \sum_{i=0}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t)$$

$$\text{and, using triangle inequality; } \left| \sum_{t=1}^T X_t \right| \leq \sum_{t=1}^T |X_t|$$

$$\Rightarrow \left| \frac{1}{T} \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) - \frac{1}{T} \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right| \leq$$

$$\frac{1}{T} \sum_{\forall t} \left| \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) - \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right|$$

$$= \frac{1}{T} \sum_{\forall t} \left| \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) - \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right|$$

$$= \frac{1}{T} \sum_{\forall t} \left| \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) - \sum_{i=0}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i - \chi_3(y_t) \right|$$

$$\leq \frac{1}{T} \sum_{\forall t} \left[\left| \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i - \sum_{i=0}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i \right| \right] +$$

$$\frac{1}{T} \sum_{\forall t} [\chi_3(y_t) - \chi_3(y_t)]$$

$$\begin{aligned}
&= \frac{1}{T} \sum_{\forall t} \left[\left| \chi_0 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) \right| \right] \\
&\quad - \frac{1}{T} \sum_{\forall t} \left[\left| \chi_0 \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) \right| \right] \\
&\quad + \frac{1}{T} \sum_{\forall t} \left[\sum_{i=1}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i \right] - \\
&\quad \frac{1}{T} \sum_{\forall t} \left[\sum_{i=1}^2 \chi_i \left(y(x_t, \delta_0), \exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) y_t^i \right] \\
&\leq \frac{1}{T} \sum_{\forall t} \chi_0 |\Xi_t| + \frac{1}{T} \sum_{\forall t} \left[\chi_0 \left[\exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right] - \left[\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right] \right] \\
&\quad + \frac{1}{T} \sum_{\forall t} \left[\sum_{i=1}^2 \chi_i \left[\exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right] - \left[\exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right] \right] \\
&\Rightarrow \frac{1}{T} \sum_{\forall t} \chi_0 |\Xi_t| + \frac{1}{T} \sum_{\forall t} \left[\chi_0 \left[\exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right] - \left[\exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right] \right] \\
&\quad + \frac{1}{T} \sum_{\forall t} \left[\sum_{i=1}^2 \chi_i \left[\exp \left(\omega^* + \sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right] - \left[\exp \left(\omega_0^* + \sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right] \right] \\
&= \frac{1}{T} \sum_{\forall t} \chi_0 |\Xi_t| + \frac{1}{T} \sum_{\forall t} \chi_0 |\omega^* - \omega_0| + \frac{1}{T} \sum_{\forall t} \chi_0 \left| \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right| \\
&\quad - \frac{1}{T} \sum_{\forall t} \chi_0 \left| \exp \left(\sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right| + \frac{1}{T} \sum_{\forall t} \left[\sum_{i=1}^2 \chi_i |\exp(\omega^*) - \exp(\omega_0)| \right] \\
&\quad + \frac{1}{T} \sum_{\forall t} \sum_{i=1}^2 \chi_i \left| \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right| - \frac{1}{T} \sum_{\forall t} \sum_{i=1}^2 \chi_i \left| \exp \left(\sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right| \\
&= \frac{1}{T} \sum_{\forall t} \chi_0 |\Xi_t| + \frac{1}{T} \sum_{\forall t} \chi_0 |\Gamma| + \frac{1}{T} \sum_{\forall t} \chi_0 \left| \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right| - \frac{1}{T} \sum_{\forall t} \chi_0 \left| \exp \left(\sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right| + \frac{1}{T} \sum_{\forall t} \left[\sum_{i=1}^2 \chi_i |\Gamma| \right] \\
&\quad + \frac{1}{T} \sum_{\forall t} \sum_{i=1}^2 \chi_i \left| \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right| - \frac{1}{T} \sum_{\forall t} \sum_{i=1}^2 \chi_i \left| \exp \left(\sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right|
\end{aligned}$$

Where $\Gamma = \exp(\omega^*) - \exp(\omega_0)$

$$\begin{aligned}
& \therefore \forall \varepsilon > 0, \text{prob} \left(\left| \frac{1}{T} \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) - \frac{1}{T} \ln L_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}(0)}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right| > \varepsilon \right) \\
& = \text{Prob} \left((T)^{-1} \sum_{\forall t} \chi_0 |\Xi_t| > \varepsilon \right) + \text{Prob} \left((T)^{-1} \sum_{\forall t} \chi_{0(1)} |\Gamma| > \varepsilon \right) + \text{Prob} \left((T)^{-1} \sum_{\forall t} \chi_0 \left| \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right| > \varepsilon \right) \\
& \quad - \text{Prob} \left((T)^{-1} \sum_{\forall t} \chi_0 \left| \exp \left(\sum_{k=1}^{t-2} \beta_k g(z_{t-k}) \right) \right| > \varepsilon \right) + \text{Prob} \left((T)^{-1} \sum_{\forall t} \sum_{i=1}^2 \chi_i |\Gamma| > \varepsilon \right) + \\
& + \text{Prob} \left((T)^{-1} \sum_{\forall t} \sum_{i=1}^2 \chi_i \left| \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right| > \varepsilon \right) - \text{Prob} \left((T)^{-1} \sum_{\forall t} \sum_{i=1}^2 \chi_i \left| \exp \left(\sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right| > \varepsilon \right) \\
& \leq \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \sum_{\forall t} \chi_0 \Xi_t \right) + \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \sum_{\forall t} \chi_{0(1)} \Gamma \right) + \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \sum_{\forall t} \chi_0 \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) \\
& \quad - \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \alpha_0 \sum_{\forall t} \chi_0 \exp \left(\sum_{k=0}^{\infty} \pi_0^k L^k \varepsilon_{0t-1}^2 \right) \right) + \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \sum_{\forall t} \sum_{i=1}^2 \chi_i \Gamma \right) + \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \sum_{\forall t} \sum_{i=1}^2 \chi_i \exp \left(\sum_{k=1}^{+\infty} \beta_k g(z_{t-k}) \right) \right) - \\
& \quad \lim_{T \rightarrow \infty} E \left((T\varepsilon)^{-1} \sum_{\forall t} \sum_{i=1}^2 \chi_i \exp \left(\sum_{k=1}^{t-2} \beta_{0k} g(z_{t-k}) \right) \right) = 0 + 0 + 0 - 0 + 0 + 0 - 0 = 0 \\
& \therefore \forall \varepsilon > 0, \lim_{T \rightarrow \infty} \text{prob} \left(\left| \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) - \frac{1}{T} \sum_{t=1}^T l_{(0)}(y_t; \theta) \right| > \varepsilon \right) = 0 \\
& \Rightarrow \text{Sup}_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) - \frac{1}{T} \sum_{t=1}^T l_{(0)}(y_t; \theta) \right| \xrightarrow{p} 0
\end{aligned}$$

Appendix A2: Proof of Lemma 2: To prove this lemma we just have to show that $\hat{\vartheta}_t^*$ and its derivatives are functions of z_t or \hat{k}_t . From assumption six z_t is strictly stationary and ergodic. Since \hat{k}_t is a function of z_t , it therefore means that ε is also strictly stationary and ergodic. Technically any function of \hat{k}_t or z_t will also be strictly stationary and ergodic. That is why all we need to show is that σ_t^2 and its derivatives are functions of \hat{k}_t or z_t to prove that they are strictly stationary and ergodic.

$$\ln(\sigma_t^2) = \omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k})$$

$$\Rightarrow (\sigma_t^2) = \exp\left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k})\right)$$

$$\Rightarrow \sigma_t^2 = \exp(\omega^*) \prod_{i=0}^{+\infty} \exp(\beta_i g(z_{t-i})) = \sigma(z_{t-k})$$

This shows that σ_t^2 , by being a function of z_t , a strictly stationary and ergodic variable, is ergodic and strictly stationary. For the derivatives of σ_t^2 with respect to the respective parameters;

$$(\sigma_t^2) = \exp\left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k})\right) = \exp(\omega^*) \exp\left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k})\right)$$

$$\Rightarrow \frac{\partial(\sigma_t^2)}{\partial \omega^*} = \exp(\omega^*) \exp\left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k})\right) = \sigma(z_{t-k})$$

$$\Rightarrow \frac{\partial(\sigma_t^2)}{\partial \beta_k} = \exp(\omega^*) \exp\left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k})\right) g(z_{t-k}) = f(z_t)$$

Without loss of generality, it can be seen that even the second derivatives will be functions of \mathcal{F}_T process. Therefore, process (σ_t^2) and its derivatives are measurable functions of an ergodic process (ε_t) , and so they are also ergodic. Similarly for $l_t(\theta)$,

$$l_t(\theta) = \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) y_t^i + \chi_3(y_t)$$

This implies that the log-likelihood function is a function of an ergodic process. This means that the likelihood function itself is ergodic.

$$\begin{aligned} \Rightarrow \frac{\partial l_t(\theta)}{\partial \delta} &= \frac{\partial l_t(\theta)}{\partial y(x_t, \delta)} \frac{\partial y(x_t, \delta)}{\partial \delta} \\ &= \frac{\partial}{\partial y(x_t, \delta)} \left[\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp\left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k})\right) \right) y_t^i + \chi_3(y_t) \right] \\ &\quad \times \left(\frac{\partial y(x_t, \delta)}{\partial \delta} \right) \end{aligned}$$

$$= f(z_t)$$

Just like the log-likelihood function, its first derivative here is also a function of an ergodic process.

This means that it itself is ergodic as well.

$$\begin{aligned}
&\Rightarrow \frac{\partial^2 l_t(\theta)}{\partial \delta \partial \delta'} = \frac{\partial}{\partial \delta'} \left[\frac{\partial l_t(\theta)}{\partial \delta} \right] \\
&= \frac{\partial}{\partial \delta'} \left[\frac{\partial \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right)}{\partial y(x_t, \delta)} \times \frac{\partial y(x_t, \delta)}{\partial \delta} \right] \\
&= \frac{\partial}{\partial y(x_t, \delta)} \left[\frac{\partial \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right)}{\partial y(x_t, \delta)} \times \frac{\partial y(x_t, \delta)}{\partial \delta} \right] \times \frac{\partial y(x_t, \delta)}{\partial \delta'} \\
&= f(z_t)
\end{aligned}$$

Without loss of generality, this also implies that the second derivative of the likelihood function is also ergodic by virtue of being a function of an ergodic process.

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \beta_k} &= \frac{\partial l_t(\theta)}{\partial \delta^2_t} \frac{\partial \delta^2_t}{\partial \beta_k} \\
&\times \left(\frac{\partial \delta^2_t}{\partial \beta_k} \right) \\
&= \frac{\partial}{\partial \delta^2_t} \left[\left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right] \\
&\times \left(\exp(\omega^*) \exp \left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) g(z_{t-k}) \right) \\
&= f(z_t) \\
\frac{\partial^2 l_t(\theta)}{\partial \beta_k \partial \beta'_k} &= \frac{\partial}{\partial \beta'_k} \left[\frac{\partial l_t(\theta)}{\partial \beta_k} \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{\partial}{\partial \beta'_k} \left[\frac{\partial \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right)}{\partial \sigma_t^2} \times \left(\exp(\omega^*) \exp \left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) g(z_{t-k}) \right) \right] \\
&= f(z_t) \\
&\frac{\partial l_t(\theta)}{\partial \omega^*} = \frac{\partial l_t(\theta)}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \omega^*} \\
&= \frac{\partial}{\partial \sigma_t^2} \left[\left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right] \times \exp(\omega^*) \exp \left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \\
&\frac{\partial^2 l_t(\theta)}{\partial \omega^* \partial \omega^*} = \frac{\partial}{\partial \omega^*} \left[\frac{\partial l_t(\theta)}{\partial \omega^*} \right] = \\
&\frac{\partial}{\partial \omega^*} \left[\frac{\partial}{\partial \sigma_t^2} \left[\left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right] \times \exp(\omega^*) \exp \left(\sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right] \\
&= f(z_t)
\end{aligned}$$

This implies that σ_t^2 and its derivatives are functions of t and/or z proving that they are strictly stationary and ergodic. The above lemma implies that,

$$\begin{aligned}
&\frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \rightarrow E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] \\
&E|\nabla l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)| < \infty, \& E|\nabla^2 l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)| < \infty
\end{aligned}$$

So we will treat these as our next lemmas.

Appendix A4: Proof of Lemma 3: From lemma two above, $E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] = f(z_t)$

which is stationary, implying $E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)]$ is finite, guarantying the conditions for applying the strong law of large numbers. By the strong law of large numbers,

$$\frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \xrightarrow{a.s} E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)]$$

Since almost sure convergence imply convergence in probability (Rao, 1973), then

$$\frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \xrightarrow{p} E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)]$$

For formality sake, let's show indeed that $E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] = f(\varepsilon_t)$

$$\begin{aligned} l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) &= \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) \right) y_t^i + \chi_3(y_t) \\ E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] &= \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) E(y_t^i) + \chi_3(y_t) \right) \\ &= \chi_0 \left(y(x_t, \delta), \omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) + \chi_1 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(y_t) \\ &+ \chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(y_t^2) = \chi_0 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) + \\ &\chi_0 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y(x_t, \delta) + \chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y^2(x_t, \delta) + \\ &\chi_0 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) \varepsilon_t^2 \\ \therefore E[l_t(\theta)] &= \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y(x_t, \delta)^i + \\ &\chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(\varepsilon_t^2) \\ &= f(\varepsilon_t) \end{aligned}$$

So $E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] = f(\varepsilon_t)$, which implies that $E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] < \infty$,

since \tilde{h}_T is stationary. This technically means that we can apply the strong law of large numbers (SLL). That is;

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) &\xrightarrow{a.s.} E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] \\ \Rightarrow \frac{1}{T} \sum_{t=1}^T l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) &\xrightarrow{p} E[l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)] \end{aligned}$$

It must be mentioned here that without A1, A6, and lemma 2 above, we could not be able to show this lemma. We have derived convergence in probability of the likelihood function to its expected

value through almost sure convergence which mainly depends on the assumptions of strict stationarity, ergodicity and martingales. This lemma will be heavily used in theorem 1 below when we will be showing consistency of the estimator.

Appendix A4: Proof of Lemma 4:

$$\begin{aligned}
l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) &= \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \\
\Rightarrow E \left| \frac{\partial l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)}{\partial \delta} \right| &= E \left| \frac{\partial}{\partial \delta} \left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right| \\
&= E \left| \frac{\partial}{\partial y(x_t, \delta)} \left[\left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right] \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \\
&= E \left[\left| \frac{\partial}{\partial y(x_t, \delta)} \left[\left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \right) \right] \right| \right] \\
&\quad \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right] \\
&\leq \left| \frac{\partial}{\partial y(x_t, \delta)} \left[\left(\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(y_t^i) + \chi_3(y_t) \right) \right] \right| \\
&\quad \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right] \\
&= \left| \frac{\partial}{\partial y(x_t, \delta)} \left[\sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(y_t^i) + \frac{\partial}{\partial y(x_t, \delta)} [\chi_3(y_t)] \right] \right| \times \\
&\quad \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right] \\
&= \left| \frac{\partial}{\partial y(x_t, \delta)} \chi_0 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) \right| + \left| \frac{\partial}{\partial y(x_t, \delta)} \chi_1 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y(x_t, \delta) \right| \\
&\quad + \left| \frac{\partial}{\partial y(x_t, \delta)} \chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(\varepsilon_t^2) \right| \left| \frac{\partial}{\partial y(x_t, \delta)} \chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y^2(x_t, \delta) \right|
\end{aligned}$$

$$\begin{aligned}
& + \left| \frac{\partial}{\partial y(x_t, \delta)} [\chi_3(y_t)] \right| \times \left| \frac{\partial}{\partial y(x_t, \delta)} \chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(\varepsilon_t^2) \right| \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right] \\
& = \left| \sum_{k=0}^2 \frac{\partial}{\partial y(x_t, \delta)} \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y^i(x_t, \delta) \right| \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right] + \\
& \quad \left| \frac{\partial}{\partial y(x_t, \delta)} \chi_2 \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) E(\varepsilon_t^2) \right| \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right] + \\
& \quad \left| \frac{\partial}{\partial y(x_t, \delta)} [\chi_3(y_t)] \right| \times \left[\left| \frac{\partial y(x_t, \delta)}{\partial \delta} \right| \right]
\end{aligned}$$

By assumption seven and assumption two it implies that;

$$\begin{aligned}
& \sum_{k=0}^2 \frac{\partial}{\partial y(x_t, \delta)} \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y^i(x_t, \delta) = O(1) \\
& \frac{\partial y(x_t, \delta)}{\partial \delta} = O(1) \text{ \& } E(\varepsilon_t^2) = O(1) \Rightarrow E \left| \frac{\partial l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta)}{\partial \delta} \right| < \infty
\end{aligned}$$

Without loss of generality we can do the same for all the other parameters and indeed for the second derivatives.

$$\therefore E \left| \nabla l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right| < \infty, \text{ \& } E \left| \nabla^2 l_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_T}}(y_{t_1}, y_{t_2}, \dots, y_{t_T}; \theta) \right| < \infty$$

Lemma 3 and lemma 4 are necessary and sufficient conditions for the convergence of QMLE respectively.

Appendix A5: Proof of Lemma 5: According to the mean value theorem, Let $f : \mathbf{R}^k$ be defined on an open convex set $\Theta \subset \mathbf{R}^k$ such that f is continuously differentiable on Θ . Then there exists θ^* in the interval, $[\theta - \theta_0]$ such that $f(\theta^*) = \frac{f(\theta_0) - f(\theta)}{\theta_0 - \theta}$ (Avran, 1988). In our case, $\nabla l_t(\tilde{\theta}_{SEM})$ is differentiable in the interval $[\theta_0 - \tilde{\theta}_{SEM}]$ such that its mean value expansion about θ^* is:

$$\begin{aligned}
\nabla^2 \sum_{t=0}^T l_t(\theta^*) &= (\tilde{\theta}_{SEM} - \theta_0)^{-1} \left(\nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) - \nabla \sum_{t=0}^T l_t(\theta_0) \right) \Rightarrow (\tilde{\theta}_{SEM} - \theta_0) \nabla^2 \sum_{t=0}^T l_t(\theta^*) \\
&= \left(\nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) - \nabla \sum_{t=0}^T l_t(\theta_0) \right) \\
&\Rightarrow \nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) = (\tilde{\theta}_{SEM} - \theta_0) \nabla^2 \sum_{t=0}^T l_t(\theta^*) + \nabla \sum_{t=0}^T l_t(\theta_0), \text{ but } \nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) = 0
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) &= 0 = \nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla^2 \sum_{t=0}^T l_t(\theta^*) \Rightarrow \nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) = 0 \\
&= \nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla^2 \sum_{t=0}^T l_t(\theta^*)
\end{aligned}$$

This proves that;

$$\nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) = 0 = \nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla^2 \sum_{t=0}^T l_t(\theta^*)$$

This technically implies that the asymptotic normality of $\tilde{\theta}_{SEM}$ is determined by the RHS of lemma 5;

$$\nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla^2 \sum_{t=0}^T l_t(\theta^*)$$

Appendix A6: Proof of Lemma 6: Before we prove this; let us review some concepts in asymptotic theory that will be necessary in this section. Now, we say that $q_{T_i(Z_i; \theta)}$ obeys strong uniform law of large numbers, SULLN if;

$$\sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T |q_{T_i(Z_i; \theta)} - E(q_{T_i(Z_i; \theta)})| \rightarrow 0, a.s$$

On the other hand, $q_{T_i(Z_i; \theta)}$ is said to obey WULLN if the convergence condition above holds in probability. That is; $\sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T |q_{T_i(Z_i; \theta)} - E(q_{T_i(Z_i; \theta)})| \xrightarrow{p} 0$

Here we want to show that; $\sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \left| \nabla^2 \sum_{t=0}^T l_t(\theta^*) - E\left(\nabla^2 \sum_{t=0}^T l_t(\theta^*) \right) \right| \xrightarrow{p} 0$. There are two

conditions that a sequence, like $\nabla^2 \sum_{t=0}^T l_t(\theta^*)$, must satisfy in order to obey WULLN (Posedel, 2005);

first, $\nabla^2 \sum_{t=0}^T l_t(\theta^*)$ should follow the weak law of large numbers, WLLN. Second, $\nabla^2 \sum_{t=0}^T l_t(\theta^*)$,

should be Lipchitz continuous. Let's start with WLLN;

$$\begin{aligned}
l_t(\theta) &= \sum_{i=0}^2 \chi_i \left(y(x_t, \delta), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}) \right) \right) y_t^i + \chi_3(y_t) \\
\nabla^2 \sum_{t=0}^T l_t(\theta^*) &= \sum_{t=0}^T \nabla^2 l_t(\theta^*) = \sum_{t=0}^T \nabla^2 \chi_0 \left(y(x_t, \delta^*), \exp \left(\omega^* + \sum_{k=1}^{\infty} \beta_k^* g(z_{t-k}) \right) \right)
\end{aligned}$$

$$+ \sum_{t=0}^T \left(\sum_{i=1}^2 \nabla^2 \chi_i \left(y(x_t, \delta^*), \exp \left((\omega^*)^* + \sum_{k=1}^{\infty} \beta_k^* g(z_{t-k}) \right) \right) \right) y_t^i + \nabla^2 \chi_3(y_t) \Bigg)$$

But according to lemma two, $l_t(\theta), \nabla \sum_{t=1}^T l_t(\theta), \& \nabla^2 \sum_{t=1}^T l_t(\theta)$ are ergodic. This implies that

$l_t(\theta^*), \nabla \sum_{t=1}^T l_t(\theta^*), \& \nabla^2 \sum_{t=1}^T l_t(\theta^*)$ are also ergodic processes. This implies that

$E(l_t(\theta)), E(\nabla l_t(\theta)), \& E(\nabla^2 l_t(\theta))$ are finite. The fact that $E(\nabla^2 l_t(\theta)) < \infty$ implies that we can apply

the strong law of large numbers on $E(\nabla^2 l_t(\theta))$

$$\frac{1}{T} \sum_{t=0}^T \nabla^2 l_t(\theta^*) \rightarrow E(\nabla^2 l_t(\theta)), a.s \Rightarrow \frac{1}{T} \sum_{t=1}^T \left[\sum_{t=0}^T \nabla^2 l_t(\theta^*) - E(\nabla^2 l_t(\theta)) \right] \rightarrow 0, a.s$$

$$\text{But } \frac{1}{T} \sum_{t=1}^T \left[\sum_{t=0}^T \nabla^2 l_t(\theta^*) - E(\nabla^2 l_t(\theta)) \right] \rightarrow 0, a.s \Rightarrow \frac{1}{T} \sum_{t=1}^T \left[\sum_{t=0}^T \nabla^2 l_t(\theta^*) - E(\nabla^2 l_t(\theta)) \right]^p \rightarrow 0$$

$$\therefore \frac{1}{T} \sum_{t=1}^T \left[\sum_{t=0}^T \nabla^2 l_t(\theta^*) - E(\nabla^2 l_t(\theta)) \right]^p \rightarrow 0$$

Let's now look at (ii), Lipchitz continuity. A function f from $S \in \mathbf{R}^n$ is Lipchitz continuous at $x \in S$ if there is a constant C such that (see Posedel, 2005);

$$|f(y) - f(x)| \leq C \|y - x\|, a.s$$

For any $y \in S$ sufficiently near, C is a random variable bounded almost surely and $\|\cdot\|$ is the Euclidean norm

$$\text{Here we want to show that } \left| \sum_{t=0}^T \nabla^2 l_t(\theta) - \sum_{t=0}^T \nabla^2 l_t(\theta^*) \right| \leq C \|\theta - \theta^*\|$$

For any f , real valued function, defined and differentiable on the interval $I \in \mathbf{R}$. If f is bounded on I , then f is a Lipchitz function on I . So, any differentiable function is Lipchitz. One of our assumptions is that the likelihood function is differentiable. This means therefore that $\nabla L_T(\theta)$ is Lipchitz.

$$\therefore \left| \sum_{t=0}^T \nabla^2 l_t(\theta) - \sum_{t=0}^T \nabla^2 l_t(\theta^*) \right| \leq C \|\theta - \theta^*\|$$

$$\text{Given that } \frac{1}{T} \sum_{t=1}^T \left[\sum_{t=0}^T \nabla^2 l_t(\theta^*) - E(\nabla^2 l_t(\theta)) \right]^p \rightarrow 0, \& \left| \sum_{t=0}^T \nabla^2 l_t(\theta) - \sum_{t=0}^T \nabla^2 l_t(\theta^*) \right| \leq C \|\theta - \theta^*\|$$

$$\therefore \text{Sup}_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \left| \nabla^2 \sum_{t=0}^T l_t(\theta^*) - E \left(\nabla^2 \sum_{t=0}^T l_t(\theta^*) \right) \right| \xrightarrow{p} 0$$

Indeed $\nabla^2 \sum_{t=0}^T l_t(\theta^*)$ obeys the weak uniform law of large numbers, WULLN.

Appendix A7: Proof of Lemma 7: From lemma 6 above,

$$\text{Sup}_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \left| \nabla^2 \sum_{t=0}^T l_t(\theta^*) - E \left(\nabla^2 \sum_{t=0}^T l_t(\theta^*) \right) \right| \xrightarrow{p} 0 \Rightarrow \left| \nabla^2 \sum_{t=0}^T l_t(\theta^*) - E \left(\nabla^2 \sum_{t=0}^T l_t(\theta^*) \right) \right| \rightarrow 0$$

Using lemma 5;

$$\begin{aligned} \nabla \sum_{t=0}^T l_t(\tilde{\theta}_{SEM}) &= 0 = \nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla \sum_{t=0}^T l_t(\theta^*) \Rightarrow 0 = \nabla \sum_{t=0}^T l_t(\theta_0) + (\tilde{\theta}_{SEM} - \theta_0) \nabla \sum_{t=0}^T l_t(\theta^*) \\ (\tilde{\theta}_{SEM} - \theta_0) \nabla \sum_{t=0}^T l_t(\theta^*) &= -\nabla \sum_{t=0}^T l_t(\theta_0) \Rightarrow (\tilde{\theta}_{SEM} - \theta_0) = - \left[E \left(\nabla \sum_{t=0}^T l_t(\theta_0) \right) \right]^{-1} \nabla \sum_{t=0}^T l_t(\theta_0) \end{aligned}$$

Multiplying through by \sqrt{T} ;

$$\sqrt{T}(\tilde{\theta}_{SEM} - \theta_0) = - \left[E \left(\nabla^2 \sum_{t=0}^T l_t(\theta) \right) \right]^{-1} \sqrt{T} \nabla \sum_{t=0}^T l_t(\theta_0)$$

This technically implies that the asymptotic distribution of $\sqrt{T}(\tilde{\theta}_{SEM} - \theta_0)$ is determined by the

asymptotic distribution of the normalized score; $\nabla \sum_{t=0}^T l_t(\theta_0)$

APPENDIX B

PROOF OF THE THEOREMS

Appendix B1: Proof of Theorem 1: Before we prove this, let's look at some basic mathematical concepts necessary in this section. An adherent point (also known as closure point or point of closure or contact point) of subset A of a topological space X , is a point x in X such that every open set containing x contains at least one point of A (Hansen, 2004). Any compact set has an adherent point (Stout, 1974; Hansen, 2006). Now consider the finite series of our estimator $(\tilde{\theta}_{SEM})$ defined on Θ . Since Θ is compact by assumption, there exists an adherent point. Let this adherent point be $(\theta_{0(T)})$. There exists a sub-sequence of estimators $(\tilde{\theta}_{\psi(SEM)})$ such that $(\tilde{\theta}_{\psi(SEM)}) \rightarrow (\theta_{0(T)})$, a.s where $\psi(SEM)$ is an increasing injective function. From lemma 1 above, we can see that;

$$\begin{aligned} \tilde{\theta}_{SEM} &= \arg \max_{\theta \in \Theta} \sum_{t=1}^T l(y_t; \theta) = \arg \max_{\theta \in \Theta} \sum_{t=1}^T l(y_t; \theta) \Rightarrow \tilde{\theta}_{SEM} = \arg \max_{\theta \in \Theta} \sum_{t=1}^T l(y_t; \theta) \\ \therefore \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) &\leq \frac{1}{T} \sum_{t=1}^T l(y_t; \tilde{\theta}_{SEM}) \Rightarrow \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \leq \frac{1}{T} \sum_{t=1}^T l(y_t; \tilde{\theta}_{\psi(SEM)}) \\ &\Rightarrow \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \leq \frac{1}{T} \sum_{t=1}^T l(y_t; \tilde{\theta}_{\psi(SEM)}) \end{aligned}$$

$$\text{But } (\hat{\theta}_{\psi(T)}) \rightarrow (\theta_{0(T)}), \text{ a.s \& } \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \rightarrow E[l(y_t; \theta)] \Rightarrow \frac{1}{T} \sum_{t=1}^T l(y_t; \theta_0) \rightarrow E[l(y_t; \theta_0)], \text{ a.s,}$$

$$\& \frac{1}{T} \sum_{t=1}^T l(y_t; \hat{\theta}_{\psi(SEM)}) \rightarrow E[l(y_t; \theta_{0(T)})]; \text{ Hence, } \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \leq \frac{1}{T} \sum_{t=1}^T l(y_t; \tilde{\theta}_{\psi(SEM)}),$$

$$\text{becomes: } \forall \theta \in \Theta, E[l(y_t; \theta)] \leq E[l(y_t; \theta_{0(T)})], \text{ a.s}$$

$$\therefore \text{By definition of } \theta_0, \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \leq \frac{1}{T} \sum_{t=1}^T l(y_t; \theta_0),$$

$$\text{and; } (\tilde{\theta}_{\psi(SEM)}) \rightarrow (\theta_{0(T)}), \text{ a.s \& } \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \rightarrow E[l(y_t; \theta)] \Rightarrow \frac{1}{T} \sum_{t=1}^T l(y_t; \theta_0) \rightarrow E[l(y_t; \theta_0)], \text{ a.s,}$$

$$\& \frac{1}{T} \sum_{t=1}^T l(y_t; \tilde{\theta}_{\psi(SEM)}) \rightarrow E[l(y_t; \theta_{0(T)})] \therefore \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta) \leq \frac{1}{T} \sum_{t=1}^T l(y_t; \theta_0),$$

$$\text{can be written as; } \forall \theta \in \Theta, E[l(y_t; \theta)] \leq E[l(y_t; \theta_0)], \text{ a.s}$$

$$\theta = \theta_{0(T)} \Rightarrow \forall \theta \in \Theta, \frac{1}{T} \sum_{t=1}^T l(y_t; \theta_{0(T)}) \leq \frac{1}{T} \sum_{t=1}^T l(y_t; \theta_0) \Rightarrow \forall \theta \in \Theta, E[l(y_t; \theta_{0(T)})] \leq E[l(y_t; \theta_0)], \text{ a.s}$$

But we just showed that, $\forall \theta \in \Theta, E[l(y_t; \theta)] \leq E[l(y_t; \theta_{0(T)})], \text{ a.s}$, and now we have shown that

$$\begin{aligned} \forall \theta \in \Theta, E[l(y_t; \theta_{0(T)})] &\leq E[l(y_t; \theta_0)], a.s \Rightarrow \theta_{0(T)} = \theta_0 \\ \text{Since } (\tilde{\theta}_{\psi(SEM)}) &\rightarrow (\theta_{0(T)}), a.s \Rightarrow \tilde{\theta}_{SEM} \rightarrow \theta_0, a.s \end{aligned}$$

But almost sure convergence implies convergence in probability (see Rao, 1973; Stout, 1974). This technically implies that;

$$\tilde{\theta}_{SEM} = \theta_0 + o_p(1), i.e. \forall \varepsilon > 0, \lim_{T \rightarrow \infty} \text{Prob}(|\tilde{\theta}_{SEM} - \theta_0| > \varepsilon) = 0$$

This proves that, $\tilde{\theta}_{SEM} \rightarrow \theta, a.s \Rightarrow \tilde{\theta}_{SEM} = \theta_0 + o_p(1)$. This technically means that $\tilde{\theta}_{SEM}$ is consistent.

Appendix B2: Proof of Theorem 2: One recalls from Lemmas six and seven that it is clear that ;

$$\begin{aligned} \sqrt{T}(\tilde{\theta}_{SEM} - \theta_0) &= -E[\nabla^2(L_T(\theta))]^{-1} \left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) + o_p(1) \\ \Rightarrow \left[\text{var} \left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) \right]^{-0.5} \left[\left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) - E \left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) \right] &\xrightarrow{d} N(0, I_k) \end{aligned}$$

This, assuming there is no dynamic misspecification, implies that;

$$\begin{aligned} \left[E[\nabla^2(L_T(\theta))]^{-1} \left[\text{var} \left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) \right] E[\nabla^2(L_T(\theta))]^{-1} \right]^{-0.5} \sqrt{T}(\tilde{\theta}_{SEM} - \theta_0) &\xrightarrow{d} N(0, I_K) \\ \Rightarrow E[\nabla^2(L_T(\theta))] + \left[\text{var} \left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) \right] &= 0 \Rightarrow \left[\text{var} \left(\sqrt{T} \sum_{t=1}^T \nabla l_t(\theta) \right) \right] = -E[\nabla^2(L_T(\theta))] \end{aligned}$$

This shows that the derived estimator reaches its Cramer-Rao lower bound.