A Scalar Dynamic Conditional Correlation Model: Structure and Estimation

HUI WANG[a] and JIAZHU PAN[b]

[a] *School of Finance, Central University of Finance and Economics, China.*

[b] *Department of Mathematics and Statistics, University of Strathclyde, UK.*

Abstract. The dynamic conditional correlation (DCC) model has been popularly used for modeling conditional correlation of multivariate time series since Engle (2002). However, the stationarity conditions are established only most recently and the asymptotic theory of parameter estimation for the DCC model has not been discussed fully. In this paper, we propose an alternative model, namely the scalar dynamic conditional correlation (SDCC) model. Sufficient and easy-checking conditions for stationarity, geometric ergodicity and β-mixing with exponential decay rates are provided. We then show the strong consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE) of the model parameters under regular conditions. The asymptotic results are illustrated by Monte Carlo experiments. As a real data example, the proposed SDCC model is applied to analysing the daily returns of the FSTE 100 index and FSTE 100 futures. Our model improves the performance of the DCC model in the sense that the LiMcleod statistic of the SDCC model is much smaller and the hedging efficiency is higher.

MSC(2010): 37M10, 62F12.

Key words and phrases. Dynamic conditional correlation; stationarity; ergodicity; QMLE; consistency; asymptotic normality.

1 Introduction

The recent econometric and statistical literature has witnessed a growing interest in modeling conditional correlations of multivariate time series. Especially in the study of financial econometrics and risk management, how to efficiently measure and manage the market risk is always a core topic. This topic has become more and more important for the competitiveness and even survival of financial institutions in today's global and highly volatile markets. One of the critical inputs required by risk managers is cross-sectional (or cross-asset) correlations. For example, the estimates of the correlations between the returns of the assets are required in the financial hedge. If the correlations and volatilities are changing, then the hedge ratio should be adjusted to account for the most recent information. It is also the case for asset allocation, pricing structured products and systemic risk measurement (Brownlees and Engle (2017)). These facts motivate the construction of models that can summarize the dynamic properties of two or more asset returns. A class of models that address this topic is the multivariate GARCH models, in which volatilities and correlations at a given time are functions of lagged returns and lagged values of themselves. Seminal works in this area are the Constant Conditional Correlation (CCC) model of Bollerslev (1990), the Dynamic Conditional Correlation (DCC) model of Engle (2002) and the Varying Correlation model of Tse and Tsui (2002). Extensions of the CCC model and the DCC model have been proposed, among others, by He and Teräsvita (2004), Cappiello, Engle and Sheppard (2006), Franses and Hafner (2009), Zakoïan (2010), Francq and Zakoïan (2011), Aielli (2013), Aielli and Caporin (2014). For the reviews of the literature on multivariate GARCH models, see Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009) and Francq and Zakoïan (2010) etc.

The existence and uniqueness of stationary and ergodic solution and the existence of moments for the multivariate GARCH are important in establishing the asymptotic results of estimation. Such problems for multivariate GARCH models have been less investigated than those for univariate models, although several papers have appeared on this aspect for different specifications. Dennis et al.(2002) gave sufficient conditions for geometric ergodicity of the so-called Baba, Engle, Kraft and Kroner (BEKK) representation of the multivariate ARCH(q) model. Ling and McAleer (2003) established conditions under which the CCC model of Bollerslev (1990) has a strictly stationary solution. Kristensen (2007) provided sufficient conditions for geometric ergodicity for a variety of multivariate GARCH models, including the vector form (VEC) of the BEKK GARCH model, see also Boussama et al. (2011) which established the same results using Markovian chain theory combined with algebraic geometry theory. Hafner (2003) derived conditions for the the existence of the fourth moment of multivariate GARCH processes in vector specification which nests the BEKK model of Engle and Kroner (1995) and the factor GARCH model of Diebold and Nerlove (1989) when the innovations belong to the class of spherical distributions. He and Teräsvita (2004) gave a sufficient condition for the existence of the fourth moment and the complete fourth-moment structure for the extended CCC model considered by Jeantheau (1998). Francq and Zakoïan (2011) gave sufficient and necessary conditions for the strict stationarity of a class of multivariate asymmetric multivariate GARCH models, including the extended CCC model in Jeantheau (1998).

The asymptotic theory of estimation for multivariate GARCH is far from being coherent, compared to univariate GARCH models. Bollerslev and Wooldridge (1992) proposed the condition that the likelihood follows a uniform weak law of large numbers for consistency of the QMLE. They also assumed asymptotic normality of the score but did not verify whether any of the conditions actually holds for specific multivariate GARCH models. Jeantheau (1998) gave conditions for the strong consistency of the QMLE for multivariate GARCH models and verified the conditions for the extended CCC model. Jeantheau's work did not require conditions on the log-likelihood derivatives. Comte and Lieberman (2003) showed the consistency and asymptotic normality of the QMLE for the BEKK formulation. Asymptotic results were established by Ling and McAleer (2003) for the CCC formulation of an autoregressive moving average model with GARCH noises (ARMA-GARCH). Hafner and Preminger (2009) investigated the asymptotic theory for VEC model proposed by Bollerslev, Engle and Wooldridge (1988). McAleer et al.(2009) developed a constant conditional correlation vector ARMA-asymmetric GARCH model and established the asymptotic normality of QMLE. Francq and Zakoïan (2011) established the strong consistency and asymptotic normality of QMLE for a class of multivariate asymmetric GARCH processes. Their processes generalize the extended CCC model of Jeantheau (1998) by allowing cross leverage effects.

In contrast to the CCC model, the DCC model of Engle (2002) fit the conditional variance of each component with a univariate GARCH model and the conditional correlation with a particular function of the past standardized residuals obtained in the separate GARCH fittings for all components of returns. This model has two major advantages: capturing the dynamic structure of correlations and having a small number of parameters. Most recently, Fermanian and Malongo (2016) established the stationary conditions for the DCC model based on Tweedie's (1988) criteria. However, as Aielli (2013) pointed out, the estimator of the location parameter in the DCC model can be inconsistent, and the traditional GARCH-like interpretation of the DCC correlation parameters can lead to paradoxical conclusions. The asymptotic theory of parameter estimation for the DCC model has not been clearly established under regular conditions since Engle and Sheppard (2001) gave only general conditions which are difficult to verify. Francq and Zakoïan (2016) considered a new estimator for a wide class of multivariate volatility models including the CCC model and the BEKK model etc.. Strong consistency and asymptotic normality were established for general constant conditional correlation models. However, the asymptotic properties of their estimator for the DCC model is still an open issue. This motivates us to consider a scalar version of the DCC model of Engle (2002) in the sense that we assume the conditional variance of every single asset to be constant but the correlation of the assets is dynamic. That is, we only focus on the dynamics of correlations. To apply the proposed SDCC model to practice, one should standardize the data for every individual asset first. In the SDCC model the location parameter is treated as a free estimator and may fit the data more adequately compared with the DCC model in which the location parameter is estimated by sample estimator (see Remark 2 in section 2 and the real data example in section 4). Referring to the result of Boussama et al. (2011), we will discuss the probabilistic structure of the SDCC model and derive its stationarity under simple conditions. Furthermore, we will show that the QMLE is strongly consistent provided that the model is stationary. The asymptotic normality of the QMLE is established under the assumption that the innovation has finite $(8+\delta)$th moment for some $\delta > 0$.

The rest of this paper is organized as follows: Section 2 introduces the scalar DCC model and gives the existence of strictly stationary solution with finite second moment to the model. Section 3 establishes the consistency and asymptotic normality of the QMLE. Section 4 discusses the finite sample performance of the QMLE through Monte Carlo simulations and a real data example on empirical application of the SDCC model to financial futures hedging problem. All proofs are presented in section 5.

In the sequel, $\stackrel{\mathcal{L}}{\to}$, $\stackrel{P}{\to}$ and $\stackrel{a.s}{\to}$ denote convergence in distribution, in probability and almost surely respectively. A' denotes the transpose of a vector or a matrix A, $Tr(A)$ is the trace of a matrix A, $|A|$ denotes the determinant of matrix A, and $\|\cdot\|$ denotes the Euclidean norm for both vectors and matrices, i.e. $\|A\| = \sqrt{Tr(A'A)}$. I_m is an $m \times m$ identity matrix and K is a constant or a random variable which does not depend on sample size and may be different at different places.

2 The model and its stationarity

Consider an $m-$dimensional time series $X_t = (X_{1t}, \cdots, X_{mt})'$, e.g. a sequence of return vectors of m assets. Assuming the conditional variance of each component is unity but the conditional correlation between components is dynamic, we propose the following scalar dynamic conditional correlation (SDCC) model:

$$
\begin{cases}
X_t = R_t^{1/2}\eta_t \\
R_t = \Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1} \\
\Sigma_t = C + \sum_{i=1}^{q}\alpha_i X_{t-i}X_{t-i}' + \sum_{j=1}^{p}\beta_j\Sigma_{t-j} =: (\sigma_{ij,t}^2)_{m\times m} \\
\Sigma_{t*} = \mathrm{diag}\{\sigma_{ii,t}\}
\end{cases}
\tag{2.1}
$$

where $\alpha_i \geq 0$, $i = 1, \cdots, q$, $\beta_j \geq 0$, $j = 1, \cdots, p$, and $C = (c_{ks})_{m\times m}$ is a positive definite matrix with unit-diagonal elements. $R_t^{1/2}$ is the unique positive definite square root of R_t. Furthermore, $\{\eta_t\}$ is a sequence of independent and identically distributed (iid) random vectors with $E(\eta_t) = 0$ and $Var(\eta_t) = I_m$, and η_t is independent of $\mathcal{F}_{t-1} = \sigma\big(X_{t-k}, k \geq 1\big)$ for all t.

Remark 1. Here the condition that C is unit-diagonal is a simple normalization such that the conditional correlation process is identifiable.

Remark 2. The SDCC model (2.1) focuses on capturing the dynamics of the conditional correlation. In fact, the conditional variance of i-th component X_{it} can be assumed as σ_{it}^2. We standardize the data by dividing them by their conditional standard deviations before applying this model. So we assume the conditional variance to be unity in the model for simplicity. We have two ways to apply the proposed SDCC model to real data of financial returns. The approximate way is that we can standardize the data by dividing them using the sample standard deviations over a moving window. But the usual way is the following two steps,and see the real data example in section 4 below for illustration.

Step 1. To capture the heteroscedasticity of each asset, fit a GARCH-type model or any other type of models to the conditional variance to each asset returns, and then obtain residuals for each asset. The specification of the univariate GARCH type models is not limited to the standard GARCH model, but can include any GARCH type process such as EGARCH, TGARCH, APGARCH, depending on special features of the data.

Step 2. To capture the dynamic conditional correlation across assets, fit model (2.1) to the $m-$dimensional vector time series of residuals.

Now we establish the stationarity of X_t defined in model (2.1) under the typical basic assumptions:

A1. The distribution Γ of η_t is absolutely continuous with respect to the Lebesgue measure on R^m and the point zero is in the interior of $E := supp(\Gamma)$.

A2. $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$.

Theorem 1. *Suppose that assumptions A1-A2 hold. Then there exists a unique strictly stationary solution X_t to model (2.1), and X_t is geometrically ergodic and geometrically β-mixing. Furthermore, $E\|X_t\|^2 < \infty$ and $E\|\Sigma_t\| < \infty$.*

Remark 3. Please refer to Fan and Yao (2003) for the definition of ergodicity, geometric ergodicity and $\beta-$mixing condition.

Remark 4. The moment structure of model (2.1) is very simple. Noting that R_t is bounded in the sense that $\|R_t\| \le m$ a.s., for any $\tau > 0$ we have $E\|X_t\|^\tau < \infty$ provided that $E\|\eta_t\|^\tau < \infty$. Thus, if model (2.1) has a strictly stationary solution, it is also weakly stationary when $E\|\eta_t\|^2 < \infty$. This is different from the BEKK model. The BEKK model might have a strictly stationary solution with infinite second moment even when the innovations have finite variance. Although in principle the main problem is only to find an appropriate function for the Foster-Lyapunov drift criterion, as Boussama et al. (2011) pointed out, it seems impossible to extend the univariate result to cover the BEKK model at the moment.

Remark 5. Since one usually uses absolutely continuous innovations η_t, such as multivariate Gaussian or multivariate student-t innovations with finite second moments, assumption A2 is the only condition to be checked. Furthermore, assumption A1 can be weakened further, see Boussama et al. (2011) for details.

3 Asymptotic properties of QMLE

The parameter in the SDCC model consists of the coefficients of the lower triangular part of the intercept matrix C and α_i, β_j, $i = 1, \cdots, q$, $j = 1, \cdots, p$. The number of unknown parameters is thus $d = m(m-1)/2 + p + q$, and the parameter vector is denoted by $\theta = (\theta_1, \cdots, \theta_d)'$ with true parameter $\theta_0 = (\theta_{10}, \cdots, \theta_{d0})'$. Let $X_1, \cdots, X_n$ be observations from model (2.1). Conditionally on initial values $X_0, \cdots, X_{1-q}, \tilde{\Sigma}_0, \cdots, \tilde{\Sigma}_{1-p}$, the Gaussian quasi-likelihood function is

$$GL_n(\theta) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{m/2} |\tilde{R}_t(\theta)|^{1/2}} \exp\Big\{ -\frac{1}{2} X_t' \tilde{R}_t^{-1}(\theta) X_t \Big\},$$

where $\tilde{R}_t(\theta)$ are recursively defined, for $t \ge 1$, by

$$\begin{cases} \tilde{R}_t(\theta) = \tilde{\Sigma}_{t*}^{-1}(\theta)\tilde{\Sigma}_t(\theta)\tilde{\Sigma}_{t*}^{-1}(\theta), \\ \tilde{\Sigma}_t(\theta) = C + \sum_{i=1}^{q} \alpha_i X_{t-i}X_{t-i}' + \sum_{j=1}^{p} \beta_j \tilde{\Sigma}_{t-j}(\theta) =: (\tilde{\sigma}_{ij,t}^2(\theta))_{m\times m}, \\ \tilde{\Sigma}_{t*}(\theta) = \mathrm{diag}\{\tilde{\sigma}_{ii,t}(\theta)\}. \end{cases} \tag{3.1}$$

The model is not assumed to be necessarily Gaussian, but we work with the Gaussian quasi-likelihood. The quasi-maximum likelihood estimator (QMLE) $\hat{\theta}_n$ is defined as

$$\hat{\theta}_n = \arg\max_{\theta} GL_n(\theta) = \arg\min_{\theta} \tilde{L}_n(\theta),$$

where

$$\tilde{L}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta) \quad \text{and} \quad \tilde{l}_t(\theta) = X_t'\tilde{R}_t^{-1}(\theta)X_t + \log|\tilde{R}_t(\theta)|. \tag{3.2}$$

It will be convenient to approximate the sequence $\tilde{l}_t(\theta)$ by an ergodic stationary sequence. Therefore, we define

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \quad \text{and} \quad l_t(\theta) = X_t'R_t^{-1}(\theta)X_t + \log|R_t(\theta)|, \tag{3.3}$$

where

$$\begin{cases} R_t(\theta) = \Sigma_{t*}^{-1}(\theta)\Sigma_t(\theta)\Sigma_{t*}^{-1}(\theta) \\ \Sigma_t(\theta) = C + \sum_{i=1}^{q} \alpha_i X_{t-i}X_{t-i}' + \sum_{j=1}^{p} \beta_j \Sigma_{t-j}(\theta) =: (\sigma_{ij,t}^2(\theta))_{m\times m} \\ \Sigma_{t*}(\theta) = \mathrm{diag}\{\sigma_{ii,t}(\theta)\} \end{cases} \tag{3.4}$$

We need the following assumptions to establish the strong consistency of the QMLE $\hat{\theta}_n$.

A3. The parameter space Θ is compact, and $\theta_0 \in \Theta$.

A4. For any $\theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

A5. Any element of X_t can not be determined by the other elements of X_t and $\mathcal{F}_{t-1}$.

The following theorem gives the strong consistency of the QMLE.

Theorem 2. *Under assumptions A1-A5,* $\hat{\theta}_n \overset{a.s.}{\to} \theta_0$.

Remark 6. Assumption A5 is the identification condition for model (2.1). Jeantheau (1998) gave primitive conditions for identifiability for an extended version of the CCC model, see also Francq and Zakoïan (2011). For the discussion of the identification problem of the BEKK model and factor GARCH models, see Sherrer and Ribarits (2007), Fiorentini and Sentana (2001), and Doz and Renault (2004). However, the identifiability of the DCC model of Engle (2002) is still open.

To establish the asymptotic normality of the QMLE $\hat{\theta}_n$, we need the following additional assumptions.

A6. θ_0 is an interior point of Θ.

A7. $E|\eta_t|^{8+\delta} < \infty$ for some $\delta > 0$.

A8. $\eta_t \eta_t'$ is non-degenerate with $E\eta_t\eta_t' = I_m$.

Theorem 3. *Under assumptions A1-A8,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, J^{-1}HJ^{-1}),$$

where

$$H = E\Big[\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\Big] \quad \textit{and} \quad J = E\Big[\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\Big].$$

4 Numerical properties

4.1 Simulation

This subsection presents numerical evidence on the finite sample performance of asymptotic results of the QMLE through a simulation study. We computed the estimator for a bivariate SDCC model (2.1) of order $q = 1$ and $p = 1$ with $\eta_t \sim N(0, I_2)$ based on 1000 independent simulated trajectories with sample size $n = 500$. Table 1 lists the mean, bias, root mean square error (RMSE) of the QMLE for each parameter. The estimates are very accurate in general. To investigate the sampling distributions of the QMLE, we give the empirical distribution of the QMLE in figure 1 for each parameter. As we see, it can be well approximated by a Gaussian law.

4.2 A real data example: empirical analysis of financial hedging

4.2.1 Data description

The data used in this paper consist of the spot and futures prices of FSTE 100 index. The sample period covers from 12/11/2009 to 23/3/2012 with the sample size 597. Both the spot and futures prices are obtained from Forexpros. To avoid thin markets and expiration effects, the nearby futures contract is rolled over to the next nearest contract when it emerges as the most active contract. We split the sample into two subperiods: from 12/11/2009 to 4/11/2011 as the in-sample period, and from 7/11/2011 to 23/3/2012 as the out-sample period.

Table 2 lists the unit root and cointegration tests for the FSTE 100 futures and spot prices. The results of the Augmented Dickey-Fuller unit root test (ADF) indicate that both series of futures and spot indices are nonstationary. However, the first differences of the logarithmic stock indices (i.e. log-returns) of both futures and spot are stationary. In addition, the Engle-Granger two-step method reveals that the logarithm of spot price and the logarithm of futures price are cointegrated with a cointegrating vector near $(1, -1)$, implying that the spreads between log-prices of spot and futures can serve as the error correction term.

Table 3 presents a summary of the statistics for the in-sample data, including the mean, median, standard deviation, skewness and kurtosis. Also presented in Table 3 are the results of the Jarque and Bera normality test and the Ljung-Box test for returns and square returns. Initially, the mean returns of both spot and futures are close to 0. The sample standard deviation of the futures returns is larger than that of the spot returns, indicating that the futures market is more volatile than the spot market. For both spot and futures the skewness coefficients are negative, and the kurtosis coefficients significantly exceeds three. The Jarque-Bera test provides clear evidence to reject the null hypothesis of normality for the returns of both spot and futures. The Ljung-Box statistics of returns and squared returns indicate possible serial correlation and autoregressive conditional heteroscedasticity (ARCH) effects in both spot and futures return series.

4.2.2 Estimation results

As Lien and Yang (2008) pointed out, the theory of storage suggests that spot and futures prices move up and down together in a long run; however, the short-run deviations from the long run equilibrium could take place due to mispricing of either futures or spot price. The lagged basis helps to determine the spot and futures price movement, therefore, to facilitates adjustment of price deviation. Since the cointegration test indicates that the basis can serve as the error correction term, we impose the following mean models:

$$r_{st} = \phi_{s0}B_{t-1} + \sum_{i=1}^{p} \phi_{si} r_{s,t-i} + \sum_{i=1}^{q} \psi_{si} r_{f,t-i} + \varepsilon_{st} \tag{4.1}$$

$$r_{ft} = \phi_{s0}B_{t-1} + \sum_{i=1}^{p} \phi_{fi} r_{s,t-i} + \sum_{i=1}^{q} \psi_{fi} r_{f,t-i} + \varepsilon_{ft} \tag{4.2}$$

where $r_{st} = \log(S_t) - \log(S_{t-1})$, $r_{ft} = \log(F_t) - \log(F_{t-1})$, S_t and F_t denote spot and futures price at time t respectively, and $B_t = \log(S_t) - \log(F_t)$ is the spread at time t. Before estimating the mean equations, we use the Akaike Information Criterion to determine p and q and obtain that $p = q = 2$. Table 4 presents the estimation results for the mean equations (4.1) and (4.2). The feedback effects between the spot and futures markets are observed. That is, the lagged spot (or futures) returns help to predict current futures (or spot) returns. More specifically, the one-step lagged futures returns have positive effects on current spot returns and negative effects on current futures returns, the two-step lagged futures returns have positive effects on both current spot returns and futures returns, the one-step lagged spot returns have positive effects on current futures returns, and the two-step lagged spot returns have negative effects on both current spot returns and futures returns. Furthermore, when considering only statistically significant estimates at a conventional significance level of 1%, the futures returns tend to increase when the basis is large in order to restore the long-run equilibrium relationship. Based on the above estimation, we adopt the SDCC model with $p = q = 1$ in this paper and the DCC model of Engle (2002) to model the conditional correlation between the spot returns and the futures returns. Since the DCC model is estimated by two steps through QMLE and the SDCC is used for standardized data, we fit GARCH(1,1) models for the residuals of equation (4.1) and (4.2) first, namely,

$$\varepsilon_{it} = h_{it}^{1/2} e_{it} \quad \textit{and} \quad h_{it} = \omega_i + \theta_{i1}\varepsilon_{it}^2 + \theta_{i2}h_{it-1}, \quad i = s, f \tag{4.3}$$

where $\omega_i > 0, \theta_{i1} \geq 0, \theta_{i2} \geq 0$ and $\theta_{i1} + \theta_{i2} < 1$ for $i = s, f$. Then for the standardized residuals $e_t = (e_{st}, e_{ft})'$, we describe the dynamic conditional correlation coefficients between the spot returns and the futures returns through a DCC and a SDCC model respectively, i.e.

$$\begin{cases} e_t = R_t^{1/2}\eta_t \\ R_t = Q_{t*}^{-1} Q_t Q_{t*}^{-1} \\ Q_t = (1-\alpha-\beta)\bar{Q} + \alpha e_{t-1}e'_{t-1} + \beta Q_{t-1} = (q_{ij,t}) \\ Q_{t*} = \mathrm{diag}\{\sqrt{q_{ii,t}}\} \end{cases} \tag{4.4}$$

and

$$\begin{cases} e_t = R_t^{1/2}\eta_t \\ R_t = \Sigma_{t*}^{-1} \Sigma_t \Sigma_{t*}^{-1} \\ \Sigma_t = C + \alpha e_{t-1}e'_{t-1} + \beta \Sigma_{t-1} = (\sigma^2_{ij,t}) \\ \Sigma_{t*} = \mathrm{diag}\{\sigma_{ii,t}\} \end{cases} \tag{4.5}$$

where $\bar{Q}$ is the unconditional covariance of the standardized residuals resulting from the first stage estimation, $C = (c_{ij})$ is unit-diagonal positive definite, and $\alpha > 0, \beta > 0$. We call the above SDCC model a GARCH-SDCC model, since the volatilities are modeled by the GARCH(1,1) models and the residuals are modeled by the SDCC model. On the other hand, we estimate volatility simply by the sample standard deviations of the log-returns data in a moving window with width 10. Namely, h_{it} is estimated by the sample variance of $r_{i,t-1}, \cdots, r_{i,t-10}$, $i = s, f$. Then, we standardize the log-returns data by $\sqrt{h_{it}}$, $i = s, f$, and fit a SDCC model with $p = q = 1$ for the standardized data (this model is called a STD-SDCC model). Table 4 presents the estimation results and the LiMcLeod statistics for the conditional variance equations (4.3) and the correlation equations (4.4) and (4.5). Being consistent with Q^2 statistics (in Table 2), both ARCH (θ_{s1} and θ_{f1}) and GARCH (θ_{s2} and θ_{f2}) effects are found to be significant at 5% level in both spot and futures returns, with the GARCH effect being the dominant factor. The Q and Q^2 statistics indicate that the conditional variance models for both spot returns and futures returns are adequate. For the conditional correlation, the "GARCH" effect is the dominant factor for the DCC model. However, "ARCH" effect and "GARCH" effect are almost the same for the SDCC model. Furthermore, $\hat{\alpha} + \hat{\beta}$ is closed to 1 for both DCC and SDCC models, which implies the persistency of the past values in the conditional correlation. The LiMcleod statistics imply that the estimated DCC, GARCH-SDCC and STD-SDCC are all adequate and the SDCC model provides a better modeling for the data since the LiMcleod statistics of the SDCC models are much smaller than that of the DCC model. Intuitively, this result is natural since the SDCC model is more flexible in the sense that the constant matrix is not set as in the DCC model. Figure 2 presents the estimated conditional correlation coefficients for the in-sample period and the forecasting conditional correlation coefficients. For the in-sample period, the DCC, GARCH-SDCC and STD-SDCC give similar results at least for the trend. However, for the out-sample period, the DCC gives more violent results which contradicts the relationship between the spot and futures. This may be caused by the inconsistency of the estimator of DCC model.

4.2.3 Potential effects on dynamic hedging strategy

In this subsection, we discuss the potential impacts on dynamic hedging due to the different correlations captured by the SDCC model and the DCC model. By the minimum variance principle, the optimal hedging ratio is defined as

$$\Delta_t = \rho_{t+1}\sqrt{h_{st+1}/h_{ft+1}},$$

where ρ_t is referred to as the conditional coefficient between spot returns and futures returns at time t, and h_{st}, h_{ft} denote conditional variances of spot returns and futures returns at time t respectively. We use two measures to investigate the performance of the hedging strategy. The first measure is to compute the reduction of the variance after hedging, namely

$$H_E = \frac{Var(r_u(t)) - Var(r(\Delta_t))}{Var(r_u(t))} = 1 - \frac{Var(r(\Delta_t))}{Var(r_u(t))}$$

where $Var(r_u(t))$ is the variance of un-hedged portfolio and $Var(r(\Delta_t))$ is the variance of hedged portfolio. The second measure is the following mean-variance utility function:

$$U = E(r_t) - \gamma Var(r_t)$$

where $E(r_t)$ and $Var(r_t)$ are the expected return and variance of hedged portfolio, and γ is the degree of risk aversion, which is assumed to be 4 (see, for example, Grossman and Shiller (1981)). Table 5 presents the values of H_E and U for the DCC, GARCH-SDCC and STD-SDCC, which indicates that the SDCC model outperforms the DCC model in the hedging for both in-sample period and out-sample period.

5 Proofs

5.1 Proof of Theorem 1

The main idea of the proof of Theorem 1 is to apply the theory of Boussama et al. (2011) to the SDCC model. We introduce some notations first. Denote the $k-$dimemsional Euclidean space by R^k, the set of real $m \times d$ matrices by $M_{m\times d}(R)$, the vector space of real $m \times m$ matrices by $M_m(R)$, the subspace of symmetric matrices by S_m, the positive semi-definite cone by S_m^+ and positive definite matrices by S_m^{++}. Let $\mathcal{N}$ be the set of natural numbers and $\mathcal{N}^*$ be the set of natural numbers excluding zero (i.e. $\mathcal{N}^* \setminus \{0\}$). Put

$$Y_t = \big(\text{vech}(\Sigma_t)', \cdots, \text{vech}(\Sigma_{t-p+1})', X_t', \cdots, X_{t-q+1}'\big)', \tag{5.1}$$

where vech is a transformation mapping S_m to $R^{m(m+1)/2}$ by stacking the lower triangular portion of a matrix. By (2.1) and (5.1), Y_t has the form

$$Y_t = F(Y_{t-1}, \eta_t) \tag{5.2}$$

and F is a continuous map (i.e. C^1-map) from $U \times R^m$ into U, where U is the open set in $\big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q$ defined as

$$U = \underbrace{\text{vech}(S_m^{++}) \times \cdots \times \text{vech}(S_m^{++})}_{p} \times \underbrace{R^m \cdots \times R^m}_{q}.$$

Thus $\{Y_t\}$ is a Markov chain in U. Obviously, model (2.1) has a stationary solution if and only if (5.2) does. Suppose assumptions A1-A2 in section 2 hold. To apply the result of Boussama et al. (2011), we need to prove the following assertions, which are listed as lemmas 5.1-5.4, hold for the model under consideration.

Lemma 5.1. There exists a point $\omega \in \text{int}E$ and a point $\Psi \in U$ such that the sequence $(Y_t^z)_{t\in\mathcal{N}}$ defined by $Y_0^z = z$ and $Y_t^z = F(Y_{t-1}^z, \omega)$ for $t \geq 1$ converges to the point Ψ for all $z \in U$.

Proof. For arbitrary $y \in U$ and $t \geq 1$, define the sequence $(Y_t^z)_{t\in N}$ by $Y_0^z = z$ and $Y_t^z = F(Y_{t-1}^z, 0)$. We denote by X_t^z and $\text{vech}(\Sigma_t^z)$ in Y_{t-1}^z the associated values of X_t and $\text{vech}(\Sigma_t)$ in Y_t. Noting that $X_t = \Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1}\eta_t$, we have $X_t^z = 0$ for all $t \geq 1$ which yields that

$$\text{vech}(\Sigma_t^z) = \text{vech}(C) + \sum_{j=1}^{p} \beta_j \text{vech}(\Sigma_{t-j}^z)$$

Therefore, for all $t \geq q$ and $z \in U$, the following equality holds

$$Y_t^z = \Lambda + \tilde{B} Y_{t-1}^z, \tag{5.3}$$

where

$$\Lambda = (\text{vech}(C)', 0, \cdots, 0)' \in (R^{m(m+1)/2})^p \times (R^m)^q, \quad \tilde{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix} \in M_{pm(m+1)/2+qm},$$

and

$$B = \begin{pmatrix} \beta_1 I_{m(m+1)/2} & \beta_2 I_{m(m+1)/2} & \cdots & \beta_{p-1} I_{m(m+1)/2} & \beta_p I_{m(m+1)/2} \\ I_{m(m+1)/2} & 0 & \cdots & 0 & 0 \\ 0 & I_{m(m+1)/2} & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I_{m(m+1)/2} & 0 \end{pmatrix} \in M_{pm(m+1)/2}(R).$$

Since $\sum_{j=1}^q \beta_j < 1$, we get that the spectral radius of the matrix $\sum_{j=1}^q \beta_j I_{m(m+1)/2}$ is less than 1, which implies the same case for B by Proposition 4.5 of Boussama et al. (2011). Thus, (5.3) means Lemma 5.1 hold with $\omega = 0$ and Ψ defined by

$$\Psi = \Lambda + \tilde{B}\Psi. \tag{5.4}$$

Let

$$\bar{\Sigma} = \frac{1}{1 - \sum_{j=1}^p \beta_j} C. \tag{5.5}$$

Then, $\bar{\Sigma}$ is positive definite by assumption A2. We can easily verify that Ψ can be written as

$$\Psi = \big(\underbrace{vech(\bar{\Sigma})', \cdots, vech(\bar{\Sigma})'}_{p}, \underbrace{0, \cdots, 0}_{qm}\big)' \in U. \tag{5.6}$$

Now, we complete the proof of Lemma 5.1.

Lemma 5.2. There exist an algebraic variety $W \subseteq \left(R^{m(m+1)/2}\right)^p \times \left(R^m\right)^q$, a C^1-map f from $U \times R^m$ into R^m and a regular map ϕ from $\left(\left(R^{m(m+1)/2}\right)^p \times \left(R^m\right)^q\right) \times R^m$ into $\left(R^{m(m+1)/2}\right)^p \times \left(R^m\right)^q$ such that

(i) $F(z, y) = \phi(z, f(z, y))$ for all $(z, y) \in U \times R^m$,

(ii) $\phi(W \cap U \times R^m) \subseteq W \cap U$,

(iii) for all $z \in U$, the map $f_z(\cdot) = f(z, \cdot)$ is a C^1-diffeomorphism.

Proof. From (2.1) and (5.1), we obtain immediately a regular map (for the definition of a regular map, see Boussama et al. (2011)) as follows

$$\phi : \Big(\big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q\Big) \times R^m \to \big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q \text{ such that } Y_t = \phi(Y_{t-1}, X_t). \quad (5.7)$$

Let f be a map from $U \times R^m$ into R^m such that

$$f : U \times R^m \to R^m, \quad (Y_{t-1}, \eta_t) \to (\Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1})^{1/2}\eta_t = X_t.$$

Since $(\Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1})^{1/2}$ is invertible, we obtain that $f_z(\cdot) = f(z, \cdot)$ for $z \in U$ is a C^1-diffeomorphism, which means (iii) of Lemma 5.2 hold. Noting that $X_t = (\Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1})^{1/2}\eta_t$, for Y_t defined in (5.1), we have

$$Y_t = F(Y_{t-1}, \eta_t) := \phi\big(Y_{t-1}, f_{Y_{t-1}}(\eta_t)\big) = \phi(Y_{t-1}, X_t), \quad (5.8)$$

where F is a C^1-map from $U \times R^m$ into U. Thus, we establish (i) of Lemma 5.2. In the following, we show that *(ii)* of Lemma 5.2 hold. Define ϕ^k recursively by $\phi^1 = \phi$ and

$$\phi^{k+1} : \quad \Big(\big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q\Big) \times \big(R^m\big)^{k+1} \to \big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q,$$
$$\phi^{k+1}(y, x_1, \cdots, x_k, x_{k+1}) = \phi\big(\phi^k(y, x_1, \cdots, x_k), x_{k+1}\big).$$

By Proposition 3.7 of Boussama et al. (2011), we can get the algebraic variety of states

$$W = {}^Z\overline{\bigcup_{k \in \mathcal{N}^*} \phi^k(\Psi, (R^m)^k)},$$

where Ψ is defined in (5.4) and satisfies (5.6), ${}^Z\overline{A}$ denotes the closure of a set A in the Zariski topology. Here, Zariski topology means the topology over $\big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q$ for which the algebraic sets in $\big(R^{m(m+1)/2}\big)^p \times \big(R^m\big)^q$ are the closed sets, and the Zariski closure of a set A is defined by

$${}^Z\overline{A} = \bigcap_{\text{D Zariski closed}, D \supseteq A} D.$$

For any $k \in \mathcal{N}^*$, consider $y(k) \in \phi^k\big(\Psi, (R^m)^k\big)$ defined as

$$y(k) = \phi^k(\Psi, x_1, \cdots, x_k) =: (vech(\sigma_k)', \cdots, vech(\sigma_{k-p+1})', x_k', \cdots, x_{k-q+1}')'$$

with $x_1, \cdots, x_k \in R^m$. Denote $x(k) = \big(vech(x_k x_k')', vech(x_{k-1}x_{k-1}')', \cdots, vech(x_{k-q+1}x_{k-q+1}')'\big)'$ and $\sigma(k) = \big(vech(\sigma_k)', \cdots, vech(\sigma_{k-p+1})'\big)'$ with $\sigma(0) =: \sigma = \big(vech(\bar{\Sigma})', \cdots, vech(\bar{\Sigma})'\big)'$ and $x(0) = 0$, where $\bar{\Sigma}$ is defined in (5.5). It is easy to verify that

$$\sigma(k+1) = \Lambda_1 + Ax(k) + B\sigma(k), \quad (5.9)$$

where $\Lambda_1 = (\text{vech}(C)', 0, \cdots, 0)' \in (R^{m(m+1)/2})^p$ and

$$A = \begin{pmatrix} \alpha_1 I_{m(m+1)/2} & \alpha_2 I_{m(m+1)/2} & \cdots & \alpha_q p I_{m(m+1)/2} \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix} \in M_{pm(m+1)/2\times qm(m+1)/2}(R).$$

By (5.4), (5.6) and the definition of $\sigma(0)$, we can obtain $\sigma = \Lambda_1 + B\sigma$, which implies that

$$\sigma = \sum_{i=0}^{k-1} B^i \Lambda_1 + B^k \sigma.$$

Therefore, (5.9) yields that

$$\sigma(k) = \sum_{i=1}^{k-1} B^i \Lambda_1 + B^k \sigma + \sum_{i=1}^{k-1} B^i A x(k-i) = \sigma + \sum_{i=0}^{k-1} B^i A x(k-i).$$

Hence, we have

$$vech(\sigma_{k-j}) = vech(\bar{\Sigma}) + \sum_{i=1}^{k-1} Q_{i,j+1} vech(x_{k-i-j} x'_{k-i-j}),$$

where for all $i \in \{1, \cdots, k-1\}$ and $j \in \{0, \cdots, p-1\}$, $Q_{i,j+1}$ is defined by $Q_{i,j+1} = [B^i A]_{1,j+1}$ with the convention $B^0 = I_{pm(m+1)/2}$, $B^i = 0$ if $i < 0$ and $[Q]_{1,j+1}$ is the $m(m+1)/2 \times m(m+1)/2$ block from lines 1 to $m(m+1)/2$ and from columns $jm(m+1)/2 + 1$ to $(j+1)m(m+1)/2$ of a matrix $B^i A$. Therefore, W is the Zariski closure of

$$\begin{aligned} S_\Psi &= \bigcup_{k\in\mathcal{N}^*} \Big\{ y(k) : x_1, \cdots, x_k \in R^m \Big\} \\ &= \bigcup_{k\in\mathcal{N}^*} \Big\{ \Psi + \sum_{i=1}^{k-1} Q_{i,1} vech(x_{k-i} x'_{k-i}), \cdots, \Psi + \sum_{i=1}^{k-1} Q_{i,p} vech(x_{k-p+1-i} x'_{k-p+1-i}), x_1, \cdots, x_k \Big\}. \end{aligned}$$

Noting that $\phi(U \times R^m) \subseteq U$ by its definition and $\phi(S_\Psi \times R^m) \subseteq S_\Psi$ due to the above arguments, we have

$$\phi(W \cap U \times R^m) \subseteq \phi\big({}^Z\overline{S_\Psi} \times R^m\big) \subseteq^Z \overline{\phi({}^Z\overline{S_\Psi} \times R^m)} =^Z \overline{\phi(S_\Psi \times R^m)} \subseteq^Z \overline{S_\Psi} = W,$$

since ϕ is a regular map and thus continuous with respect to the Zariski topology. This completes the proof of Lemma 5.2.

Lemma 5.3. There exist a function V and positive constants $\tau < 1$, $b < \infty$ as well as a small set $\mathcal{K}$ in $W \cap U$ such that the Foster-Lyapunov condition is satisfied, i.e. for any $x \in W \cap U$

$$E\big[V(Y_t) | Y_{t-1} = x\big] \le \tau V(x) + b 1_{\mathcal{K}}(x).$$

Proof. Let

$$\Sigma = \frac{C}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}.$$

Since C is positive definite, assumption A2 implies that Σ is also positive definite. Furthermore, it is easy to verify that

$$\Sigma = C + \sum_{j=1}^{p} \beta_j \Sigma + \sum_{i=1}^{q} \alpha_i \Sigma. \tag{5.10}$$

Define

$$\begin{aligned} V(Y_t) &= \sum_{j=1}^{p} Tr(V_j \Sigma_{t-j+1}) + \sum_{i=1}^{q} X'_{t-i+1} V_{p+i} X_{t-i+1} + 1, \\ V_k &= \frac{p-k+1}{p+q} C + \sum_{j=k}^{p} \beta_j \Sigma, \quad k = 1, \cdots, p, \\ V_{k+p} &= \frac{q-k+1}{p+q} C + \sum_{i=k}^{q} \alpha_i \Sigma, \quad k = 1, \cdots, q. \end{aligned}$$

Thus, we have

$$\begin{aligned} E\big[V(Y_t)|Y_{t-1} = y\big] &= E\big[Tr(V_1 \Sigma_t) + X_t' V_{p+1} X_t | Y_{t-1} = y\big] \\ &\quad + \sum_{j=2}^{p} Tr(V_j \Sigma_{t-j+1}) + \sum_{i=2}^{q} X'_{t-i+1} V_{p+i} X_{t-i+1} + 1. \end{aligned} \tag{5.11}$$

By the formulation of Σ_t, we get

$$\begin{aligned} & E\big[Tr(V_1 \Sigma_t)|Y_{t-1} = y\big] \\ = \; & E\Big\{Tr\big[V_1(C + \sum_{i=1}^{q} \alpha_i X_{t-i} X'_{t-i} + \sum_{j=1}^{p} \beta_j \Sigma_{t-j})\big]\big|Y_{t-1} = y\Big\} \\ = \; & Tr(V_1 C) + \sum_{j=1}^{p} \beta_j Tr(V_1 \Sigma_{t-j}) + \sum_{i=1}^{q} \alpha_i X'_{t-i} V_1 X_{t-i}. \end{aligned} \tag{5.12}$$

Since $X_t = R_t^{1/2} \eta_t$, $R_t = R_t^{1/2} R_t^{1/2}$, $E\eta_t \eta_t' = I_m$, and every entry of Σ_{t*}^{-1} is less than 1, it follows that

$$\begin{aligned} & E\big[X_t' V_{p+1} X_t | Y_{t-1} = y\big] \\ = \; & E\big[Tr(V_{p+1} R_t^{1/2} \eta_t \eta_t' R_t^{1/2})\big|Y_{t-1} = y\big] \\ = \; & Tr(V_{p+1} \Sigma_{t*}^{-1} \Sigma_t \Sigma_{t*}^{-1}) \\ \leq \; & Tr(V_{p+1} \Sigma_t) \\ = \; & Tr(V_{p+1} C) + \sum_{j=1}^{p} \beta_j Tr(V_{p+1} \Sigma_{t-j}) + \sum_{i=1}^{q} \alpha_i X'_{t-i} V_{p+1} X_{t-i}. \end{aligned} \tag{5.13}$$

Combining (5.11)-(5.13), we obtain

$$
\begin{aligned}
& E\big[V(Y_t)|Y_{t-1}=y\big] \\
= \; & Tr\big\{[\beta_1(V_1+V_{p+1})+V_2]\Sigma_{t-1}\big\}+\cdots+Tr\big\{[\beta_{p-1}(V_1+V_{p+1})+V_p]\Sigma_{t-p+1}\big\} \\
& +Tr\big[\beta_p(V_1+V_{p+1})\Sigma_{t-p}\big]+X'_{t-1}[\alpha_1(V_1+V_{p+1})+V_{p+2}]X_{t-1}+\cdots \\
& +X'_{t-q+1}[\alpha_{q-1}(V_1+V_{p+1})+V_{p+q}]X_{t-q+1}+X'_{t-q}\alpha_q(V_1+V_{p+1})X_{t-q} \\
& +Tr[(V_1+V_{p+1})C]+1. \qquad (5.14)
\end{aligned}
$$

By the definition of V_i, we deduce

$$\beta_k(V_1+V_{p+1})+V_{k+1}=V_k-\frac{1}{p+q}C, \quad k=1,\cdots,p-1, \qquad (5.15)$$

$$\beta_p(V_1+V_{p+1})=V_p-\frac{1}{p+q}C, \qquad (5.16)$$

$$\alpha_k(V_1+V_{p+1})+V_{p+k+1}=V_{p+k}-\frac{1}{p+q}C, \quad k=1,\cdots,q-1, \qquad (5.17)$$

$$\alpha_q(V_1+V_{p+1})=V_{p+q}-\frac{1}{p+q}C. \qquad (5.18)$$

Because V_k is positive definite, and $\alpha_i>0$, $\beta_j>0$,$i=1,\cdots,q$, $j=1,\cdots,p$, we have $V_k-\frac{1}{p+q}C$ is also positive definite, $k=1,\cdots,p+q$. Consider the non-negative constants $(\tau_k)_{1\le k\le p+q}$ defined by

$$\tau_k=\max\Big\{x'(V_k-\frac{1}{p+q}C)x : x\in R^m, x'V_kx=1\Big\}.$$

Noting that $\{x : x'V_kx=1\}$ is a compact set, there exists $x_k\in R^m$ such that $x'V_kx=1$ and

$$\tau_k=x'_k(V_k-\frac{1}{p+q}C)x_k=1-\frac{x'_kCx_k}{p+q}.$$

The positive definiteness of C means that $0\le\tau_k<1$, $k=1,\cdots,p+q$. Put $\tau_0=\max\{\tau_k, \quad k=1,\cdots,p+q\}$ and we have $\tau_0\in[0,1)$ and

$$V_k-\frac{1}{p+q}C\le\tau_0V_k, \quad k=1,\cdots,p+q.$$

Therefore, for any $M\in S_d^{++}$, we get

$$Tr\big[(V_k-\frac{1}{p+q}C)M\big]\le\tau_0Tr(V_kM), \quad k=1,\cdots,p+q. \qquad (5.19)$$

From (5.14)-(5.19), we obtain that

$$E\big[V(Y_t)|Y_{t-1}=y\big]\le\tau_0V(y)+Tr(\Sigma C)+1-\tau_0.$$

If we choose $\tau = (\tau_0 + 1)/2 \in [1/2, 1)$, $b = Tr(\Sigma C) + 1 - \tau_0 \in (0, \infty)$, the (FL) condition is satisfied with the set given by

$$\mathcal{K} = \{W \cap U : V(x) \leq \frac{b}{\tau - \tau_0}\}.$$

Following the proof of Theorem 2.4 of Boussama et al. (2011), we can show that $\mathcal{K}$ is a small set. This complete the proof of Lemma 5.3.

Lemma 5.4. Any strictly stationary solution of the Markov chain $Y_{t+1} = F(Y_t, \eta_t)$ takes its values in the algebraic variety $W \cap U$.

Proof. Suppose X_n be a strictly stationary solution of model (2.1). Define

$$\Sigma(k) = \big(vech(\Sigma_k)', \cdots, vech(\Sigma_{k-p+1})'\big)', \quad X(k) = \big(vech(X_k X_k')', \cdots, vech(X_{k-q+1}X_{k-q+1}')'\big)'.$$

Noting that $\Sigma(k+1) = \Lambda_1 + AX(k) + B\Sigma(k)$, we can deduce

$$\Sigma(k+1) = \sum_{i=0}^{j-1} \Lambda_1 + B^j A \Sigma(k-j) + \sum_{i=1}^{j} B^{i-1} AX(k-i)$$

for all $k \in \mathcal{N}$. Under assumption A2, the spectral radius of B is less than 1 by Proposition 4.5 of Boussama et al. (2011). Using similar arguments to the proof of Theorem 4.8 of Boussama et al. (2011), we obtain that

$$\Sigma(k) = \sigma + \sum_{i=1}^{\infty} B^{i-1} AX(k-i) \quad a.s.$$

Thus,

$$vech(\Sigma_k) = vech\big(\bar{\Sigma}\big) + \sum_{i=1}^{\infty} Q_i vech\big(X_{k-i}X_{k-i}'\big) \quad a.s.,$$

with Q_i the the $m(m+1)/2 \times m(m+1)/2$ block from lines 1 to $m(m+1)/2$ and from columns $m(m+1)/2 + 1$ to $m(m+1)$ of a matrix $B^{i-1}A$,which implies that Y_t takes its values in the variety W and hence in $W \cap U$.

Proof of Theorem 1. Boussama et al. (2011) pointed out that W may not be of full dimension because of the non-invertibility of the coefficient matrices in the BEKK model. However, such degeneracy does not occur in the SDCC model (2.1) since none of the coefficient matrices in Σ_t is singular. So, combining assumption A1 and lemmas 5.1-5.4, we know that all conditions of Boussama et al. (2011)'s result are satisfied, and then Theorem 1 follows.

5.2 Proofs of Theorem 2 and Theorem 3

For the consistency of the QMLE, under assumptions A1-A5, we need to establish that

C1. $\lim_{n\to\infty}\sup_{\theta\in\Theta}\big|L_n(\theta)-\tilde{L}_n(\theta)\big|=0, \quad a.s.$

C2. $E|l_t(\theta_0)|<\infty$, and if $\theta\neq\theta_0$, $El_t(\theta)>El_t(\theta_0)$.

C3. For any $\theta\neq\theta_0$, there exists a neighborhood $N(\theta)$ such that

$$\liminf_{n\to\infty}\inf_{\theta^*\in N(\theta)}\tilde{L}_n(\theta^*)>El_1(\theta_0)\quad a.s.$$

For the asymptotic normality of the QMLE, under assumptions A1-A8, we need to show that

N1. $n^{-1/2}\sum_{t=1}^n\frac{\partial l_t(\theta_0)}{\partial\theta}\xrightarrow{\mathcal{L}}N(0,H)$ for a nonrandom H.

N2. $J =: E[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'] = \Big(E\big[Tr\big(\dot{R}_{t,i}R_t^{-1}\dot{R}_{t,j}R_t^{-1}\big)\big]\Big)_{i,j=1,\cdots,d} < \infty$ and J is non-singular.

N3. $E\sup_{\theta\in\Theta}\left|\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|<\infty$ for all $i,j,k=1,\cdots,d$.

N4. $\big\|n^{-1/2}\sum_{t=1}^n\big(\frac{\partial l_t(\theta_0)}{\partial\theta}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\big)\big\|\xrightarrow{P}0$ and

$$\sup_{\theta\in\Theta}\big\|n^{-1}\sum_{t=1}^n\big(\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\big)\big\|\xrightarrow{P}0.$$

(Refer to the proofs of Theorem 2.1 and 2.2 in Francq and Zakoïan (2004) for the univariate case.)

Before we prove the above assertions, some notations and elementary results on the norm of matrices and the differentiation of expressions involving matrices are introduced first. Let $f(A)$ be a real valued function of a matrix A whose elements a_{ij} are functions of some variable y. Then, for square matrices A and B we have the following equalities and inequalities (see Magnus and Neudecker (1988)):

$$\frac{\partial f(A)}{\partial y}=Tr\Big\{\frac{\partial f(A)}{\partial A'}\frac{\partial A}{\partial y}\Big\},\ \frac{\partial\log|A|}{\partial A'}=A^{-1},\ \frac{\partial A^{-1}}{\partial y}=-A^{-1}\frac{\partial A}{\partial y}A^{-1}, \tag{5.20}$$

$$|Tr(AB)|\le\|A\|\|B\|,\quad \|AB\|\le\|A\|\|B\|. \tag{5.21}$$

For any function $f_t(\theta)$, denote $\partial f_t(\theta)/\partial\theta_i$ by $\dot{f}_{t,i}(\theta)$, $\partial^2 f_t(\theta)/\partial\theta_i\partial\theta_j$ by $\ddot{f}_{t,ij}(\theta)$, $\partial^3 f_t(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$ by $\dddot{f}_{t,ijk}(\theta)$, and $f(\theta_0)$ by f. From (3.4), we get

$$
\begin{aligned}
\dot{R}_{t,i}(\theta) &= -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)R_t(\theta) + \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta) - R_t(\theta)\dot{\Sigma}_{t*,i}\Sigma_{t*}^{-1}(\theta); \qquad (5.22)\\
\ddot{R}_{t,ij}(\theta) &= \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)R_t(\theta) - \Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,ij}(\theta)R_t(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\dot{R}_{t,j}(\theta) - \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t,ij}(\theta)\Sigma_{t*}^{-1}(\theta) - \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\dot{R}_{t,j}(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta) - R_t(\theta)\ddot{\Sigma}_{t*,ij}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +R_t(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta); \qquad (5.23)\\
\dddot{R}_{t,ijk}(\theta) &= -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)R_t(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,jk}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)R_t(\theta) + \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,ik}(\theta)R_t(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)R_t(\theta) + \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\dot{R}_{t,k}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,ij}(\theta)R_t(\theta) - \Sigma_{t*}^{-1}(\theta)\dddot{\Sigma}_{t*,ijk}(\theta)R_t(\theta) - \Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,ij}(\theta)\dot{R}_{t,k}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\dot{R}_{t,j}(\theta) - \Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,ik}(\theta)\dot{R}_{t,j}(\theta) - \Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\ddot{R}_{t,jk}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,jk}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t,ik}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t,ij}(\theta)\Sigma_{t*}^{-1}(\theta) + \Sigma_{t*}^{-1}(\theta)\dddot{\Sigma}_{t,ijk}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t,ij}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t,ik}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,j}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,jk}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -\ddot{R}_{t,jk}(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta) - \dot{R}_{t,j}(\theta)\ddot{\Sigma}_{t*,ik}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\dot{R}_{t,j}(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta) - \dot{R}_{t,k}(\theta)\ddot{\Sigma}_{t*,ij}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad -R_t(\theta)\dddot{\Sigma}_{t*,ijk}(\theta)\Sigma_{t*}^{-1}(\theta) + R_t(\theta)\ddot{\Sigma}_{t*,ij}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&\quad +\dot{R}_{t,k}(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)
\end{aligned}
$$

$$
\begin{aligned}
&+R_t(\theta)\ddot{\Sigma}_{t*,ik}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&-R_t(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&+R_t(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,jk}(\theta)\Sigma_{t*}^{-1}(\theta)\\
&-R_t(\theta)\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,j}(\theta)\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,k}(\theta)\Sigma_{t*}^{-1}(\theta). \qquad (5.24)
\end{aligned}
$$

We then introduce several lemmas.

Lemma 5.5. Under assumption A5, model (2.1) is identifiable, namely For $\theta, \theta_0 \in \Theta$, $R_t(\theta) = R_t(\theta_0)$ a.s. implies $\theta = \theta_0$, where $R_t(\theta)$ is defined in (3.4).

Proof. We treat the case with $q = 1$, $p = 0$ and $m = 2$ first. Suppose $R_t(\theta) = R_t(\theta_0)$ a.s. Then we have

$$\frac{\sigma_{12}^2(\theta_0)}{\sigma_{11,t}(\theta_0)\sigma_{22,t}(\theta_0)} = \frac{\sigma_{12}^2(\theta)}{\sigma_{11,t}(\theta)\sigma_{22,t}(\theta)},$$

which means

$$\frac{(c_0+\alpha_0 X_{1,t-1}X_{2,t-1})^2}{(1+\alpha_0 X_{1,t-1}^2)(1+\alpha_0 X_{2,t-1}^2)} = \frac{(c+\alpha X_{1,t-1}X_{2,t-1})^2}{(1+\alpha X_{1,t-1}^2)(1+\alpha X_{2,t-1}^2)}.$$

After straight computation, we have

$$
\begin{aligned}
&c_0^2 - c^2 + (c_0^2\alpha - c^2\alpha_0)(X_{1,t-1}^2 + X_{2,t-1}^2) + (c_0^2\alpha^2 - c^2\alpha_0^2 + \alpha_0^2 - \alpha^2)X_{1,t-1}^2X_{2,t-1}^2 + \\
&2(c_0\alpha - c\alpha_0)X_{1,t-1}X_{2,t-1} + 2(c_0\alpha\alpha_0 - c\alpha_0\alpha)X_{1,t-1}X_{2,t-1}(X_{1,t-1}^2 + X_{2,t-1}^2) + \\
&2(c_0\alpha_0\alpha^2 - c\alpha\alpha_0^2)X_{1,t-1}^3X_{2,t-1}^3 + (\alpha\alpha_0^2 - \alpha_0\alpha^2)X_{1,t-1}^2X_{2,t-1}^2(X_{1,t-1}^2 + X_{2,t-1}^2) = 0
\end{aligned}
$$

If $\theta_0 \neq \theta$, then there exists at least one random term in above equation whose coefficient is non-zero. By implicit function theorem X_{1t} can be determined by X_{2t} which contradicts assumption A5. Thus $\theta_0 = \theta$. For general cases, $\sigma_{ij,t}(\theta)$ can be represented as the sum of constant term, the function of $X_{it}X_{jt}$ term and the function of $\mathcal{F}_{t-2}$ term. Similar method can yield the result of Lemma 5.5 under assumption A5.

Lemma 5.6. Suppose assumptions A1-A3 hold. Then it follows that, for i, j, $k = 1, \cdots, d$,

(i) for any $\Delta > 0$,

$$E\Big[\sup_{\theta\in\Theta}\|\dot{R}_{t,i}(\theta)\|\Big]^{\Delta} < \infty, \quad E\Big[\sup_{\theta\in\Theta}\|\ddot{R}_{t,ij}(\theta)\|\Big]^{\Delta} < \infty, \quad E\Big[\sup_{\theta\in\Theta}\|\dddot{R}_{t,ijk}(\theta)\|\Big]^{\Delta} < \infty;$$

(ii) if $E\|\eta_t\|^{2w} < \infty$ for some $w > 0$, we have $E\Big[\sup_{\theta\in\Theta}\|R_t^{-1}(\theta)\|\Big]^{w} < \infty$.

Proof. (i) First, similar to the univariate case (see Berkes et al. (2003) and Francq and Zakoïan (2004)), under assumptions A1-A2, we have

$$E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\big\|^{\Delta}\big]<\infty,\quad E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\big\|^{\Delta}\big]<\infty,$$

$$E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t*,ij}(\theta)\big\|^{\Delta}\big]<\infty,\quad E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^{-1}(\theta)\ddot{\Sigma}_{t,ij}(\theta)\Sigma_{t*}^{-1}(\theta)\big\|^{\Delta}\big]<\infty,$$

$$E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^{-1}(\theta)\dddot{\Sigma}_{t*,ijk}(\theta)\big\|^{\Delta}\big]<\infty,\quad E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^{-1}(\theta)\dddot{\Sigma}_{t,ijk}(\theta)\Sigma_{t*}^{-1}(\theta)\big\|^{\Delta}\big]<\infty \quad (5.25)$$

for any $\Delta>0$ and $i,j,k=1,\cdots,d$. From (5.22), we have

$$\big\|\dot{R}_{t,i}(\theta)\big\|\le\big\|\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t*,i}(\theta)\|+\|\Sigma_{t*}^{-1}(\theta)\dot{\Sigma}_{t,i}(\theta)\Sigma_{t*}^{-1}(\theta)\|+\|\dot{\Sigma}_{t*,i}(\theta)\Sigma_{t*}^{-1}(\theta)\big\|.$$

Combining with (5.25), we obtain that $E\big[\sup_{\theta\in\Theta}\big\|\dot{R}_{t,i}(\theta)\big\|\big]^{\Delta}<\infty$. Similarly, we can prove the latter two inequalities of (i) hold.

(ii) Note that each element of $\Sigma_{t*}^2(\theta)$ follows an univariate GARCH form. Since $1-\sum_{j=1}^p\beta_j>0$ under assumption A2, there exist some $0<\rho<1$ and positive constant K such that

$$\sigma_{ii,t}^2(\theta)\le 1+K\sum_{j=1}^{\infty}\rho^j X_{i,t-j}^2,$$

for $i=1,\cdots,m$. Since Θ is compact by assumption A3, we have

$$E\Big[\sup_{\theta\in\Theta}\sigma_{ii,t}^2(\theta)\Big]^w<\infty \quad (5.26)$$

if $E\|X_t\|^{2w}<\infty$, which is ensured by the definition of model (2.1) and the condition $E\|\eta_t\|^{2w}<\infty$. According to appendix of Comte and Lieberman (2003), for a positive definite matrix D and a positive semi-definite matrix G, we have

$$\|(D+G)^{-1}\|\le K\sqrt{Tr(D^{-4})}. \quad (5.27)$$

Since Θ is compact and C is positive definite, (3.1), (3.4) and (5.27) imply

$$\sup_{\theta\in\Theta}\|\Sigma_t^{-1}(\theta)\|\le K,\quad \sup_{\theta\in\Theta}\|\tilde{\Sigma}_t^{-1}(\theta)\|\le K,\quad \sup_{\theta\in\Theta}\|\Sigma_{t*}^{-1}(\theta)\|\le K,\quad \sup_{\theta\in\Theta}\|\tilde{\Sigma}_{t*}^{-1}(\theta)\|\le K. \quad (5.28)$$

Due to (5.26) and (5.28), we get

$$E\big[\sup_{\theta\in\Theta}\big\|R_t^{-1}(\theta)\big\|\big]^w\le E\big[\sup_{\theta\in\Theta}\big\|\Sigma_{t*}^2(\theta)\big\|\big]^w\big[\sup_{\theta\in\Theta}\big\|\Sigma_t^{-1}(\theta)\big\|\big]^w\le K\sum_{i=1}^d\big[\sup_{\theta\in\Theta}\big\|\sigma_{ii,t}^2(\theta)\big\|\big]^w<\infty.$$

This completes the proof of Lemma 5.6.

Lemma 5.7. Suppose assumption A8 holds. If for some constant vector $x=(x_1,\cdots,x_d)'$, $\sum_{k=1}^d x_k\dot{R}_{t,k}=0$ a.s. for any t, then we have $x=0$.

Proof. By (5.22) and $R_t = \Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1}$, we have

$$\dot{R}_{t,k} = -\Sigma_{t*}^{-1}\dot{\Sigma}_{t*,k}\Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1} + \Sigma_{t*}^{-1}\dot{\Sigma}_{t,k}\Sigma_{t*}^{-1} - \Sigma_{t*}^{-1}\Sigma_t\Sigma_{t*}^{-1}\dot{\Sigma}_{t*,k}\Sigma_{t*}^{-1}.$$

Multiplying Σ_{t*} from both the left and right sides of $\sum_{k=1}^{d} x_k\dot{R}_{t,k} = 0$ yields

$$\sum_{k=1}^{d} x_k\big(- \dot{\Sigma}_{t*,k}\Sigma_{t*}^{-1}\Sigma_t + \dot{\Sigma}_{t,k} - \Sigma_t\Sigma_{t*}^{-1}\dot{\Sigma}_{t*,k}\big) = 0. \tag{5.29}$$

Note that Σ_{t*} is diagonal. From (2.1), we have $\Sigma_{t*}^{-1}\dot{\Sigma}_{t*,k} = \dot{\Sigma}_{t*,k}\Sigma_{t*}^{-1} = \frac{1}{2}\Sigma_{t*}^{-2}\dot{\Sigma}^2{}_{t*,k} = \frac{1}{2}\dot{\Sigma}^2{}_{t*,k}\Sigma_{t*}^{-2}$, where $\dot{\Sigma}^2{}_{t*,k} = \partial\Sigma_{t*}^2/\partial\theta_k$. Multiplying Σ_{t*}^2 from both the left and right sides of (5.29) yields

$$\sum_{k=1}^{d} x_k\big(2\Sigma_{t*}^2\dot{\Sigma}_{t,k}\Sigma_{t*}^2 - \dot{\Sigma}^2{}_{t*,k}\Sigma_t - \Sigma_t\dot{\Sigma}^2{}_{t*,k}\big) = 0.$$

It is easy to verify that $\dot{\Sigma}^2{}_{t*,k} = 0$ for $k = 1, \cdots, m(m-1)/2$ for model (2.1). Thus, we have

$$\sum_{k=1}^{m(m-1)/2} 2x_k\Sigma_{t*}^2\dot{\Sigma}_{t,k}\Sigma_{t*}^2 + \sum_{k=m(m-1)/2+1}^{d} x_k\big(2\Sigma_{t*}^2\dot{\Sigma}_{t,k}\Sigma_{t*}^2 - \dot{\Sigma}^2{}_{t*,k}\Sigma_t - \Sigma_t\dot{\Sigma}^2{}_{t*,k}\big) = 0. \tag{5.30}$$

Note that the digonal elements of the first part of the left side of (5.30) are all zeros. Consider (i,i)th elements of the left side of (5.30), and we have

$$\sum_{k=m(m-1)/2+1}^{d} x_k\big(2\sigma_{ii,t}^4\frac{\partial\sigma_{ii,t}^2}{\partial\theta_k} - 2\sigma_{ii,t}^2\frac{\partial\sigma_{ii,t}^2}{\partial\theta_k}\big) = 0$$

Since the probability of $\sigma_{ii,t}^2 = 1 + \sum_{s=1}^{q}\alpha_s X_{i,t-s}^2 + \sum_{r=1}^{q}\alpha_s\sigma_{ii,t-r}^2 > 1$ is positive under assumption A8, we have

$$\sum_{k=m(m-1)/2+1}^{d} x_k\frac{\partial\sigma_{ii,t}^2}{\partial\theta_k} = 0.$$

The same argument as Francq and Zakoïan (2004) together with assumption A8 yields that $x_k = 0$ for $k = m(m-1)/2+1, \cdots, d$. Thus, from (5.30) we obtain

$$\sum_{k=1}^{m(m-1)/2} 2x_k\Sigma_{t*}^2\dot{\Sigma}_{t,k}\Sigma_{t*}^2 = 0.$$

On the other hand, the (i,j)th $(i \neq j)$ element of the left side of the above equation is $x_{m(i-1)+j}\sigma_{ii,t}^2/(1 - \sum_{s=1}^{q}\beta_s)$. Since $\sigma_{ii,t}^2/(1 - \sum_{s=1}^{q}\beta_s) > 0$, we obtain $x_k = 0$ for $k = 1, \cdots, m(m-1)/2$. Now, we complete the proof of Lemma 5.7.

Proof of C1. By the compactness of Θ and assumption A2, we get $\sup_{\theta\in\Theta}\sum_{j=1}^{p}\beta_j < 1$. Due to (3.1) and (3.4), we deduce that, almost surely

$$\sup_{\theta\in\Theta}\big\|\Sigma_t(\theta)-\tilde{\Sigma}_t(\theta)\big\| \le K\rho^t \tag{5.31}$$

for any t, where $0<\rho<1$ is a constant and K is a random variable which depends on the past values $\{X_t, t\le 0\}$. Since K does not depend on n, we can treat it as a constant. Note that

$$\begin{aligned}\sup_{\theta\in\Theta}\big|L_n(\theta)-\tilde{L}_n(\theta)\big| &\le \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\big|X_t'(R_t^{-1}(\theta)-\tilde{R}_t^{-1}(\theta))X_t\big| + \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\big|\log|R_t(\theta)|-\log|\tilde{R}_t(\theta)|\big| \\ &=: P_1+P_2. \end{aligned}\tag{5.32}$$

We deal with P_1 first. Due to (3.1) and (3.4), we have

$$\begin{aligned} R_t^{-1}(\theta)-\tilde{R}_t^{-1}(\theta) &= \Sigma_{t*}(\theta)\Sigma_t^{-1}(\theta)\Sigma_{t*}(\theta)-\tilde{\Sigma}_{t*}(\theta)\tilde{\Sigma}_t^{-1}(\theta)\tilde{\Sigma}_{t*}(\theta) \\ &= \big(\Sigma_{t*}(\theta)-\tilde{\Sigma}_{t*}(\theta)\big)\Sigma_t^{-1}(\theta)\Sigma_{t*}(\theta)+\tilde{\Sigma}_{t*}(\theta)\big(\Sigma_t^{-1}(\theta)-\tilde{\Sigma}_t^{-1}(\theta)\big)\Sigma_{t*}(\theta) \\ &\quad +\tilde{\Sigma}_{t*}(\theta)\tilde{\Sigma}_t^{-1}(\theta)\big(\Sigma_{t*}(\theta)-\tilde{\Sigma}_{t*}(\theta)\big) \\ &=: P_{11}+P_{12}+P_{13} \end{aligned}$$

Furthermore, (5.28) and (5.31) imply

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|X_t'P_{11}X_t| &= \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}Tr(P_{11}X_tX_t') \\ &\le \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\|P_{11}\|\|X_tX_t'\| \\ &\le \frac{K}{n}\sum_{t=1}^{n}\rho^t\|\Sigma_{t*}\|\|X_tX_t'\|. \end{aligned}$$

Duo to Theorem 1, there exists a $\delta>0$ such that

$$E\big(\|\Sigma_{t*}\|^{\delta}\|X_tX_t'\|^{\delta}\big)<\infty.$$

Thus, we have

$$\begin{aligned} \sum_{t=1}^{\infty}P\big(\rho^t\|\Sigma_{t*}\|\|X_tX_t'\|>\varepsilon\big) &\le \sum_{t=1}^{\infty}\frac{\rho^{t\delta}E\big(\|\Sigma_{t*}\|^{\delta}\|X_tX_t'\|^{\delta}\big)}{\varepsilon^{\delta}} \\ &= \frac{E\big(\|\Sigma_{t*}\|^{\delta}\|X_tX_t'\|^{\delta}\big)}{\varepsilon^{\delta}}\sum_{t=1}^{\infty}\rho^{t\delta} \\ &< \infty. \end{aligned}$$

By the Borel-Cantelli lemma, we have

$$\rho^t\|\Sigma_{t*}\|\|X_tX_t'\| \xrightarrow{a.s.} 0.$$

Therefore,

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|X_t'P_{11}X_t| \xrightarrow{a.s.} 0 \tag{5.33}$$

due to the Cesáro lemma. Noting that $\|\tilde{\Sigma}_{t*}\| \leq \|\Sigma_{t*}\| + K\rho^t$, and using similar arguments to the proof of (5.33), we can get

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|X_t'P_{13}X_t| \xrightarrow{a.s.} 0. \tag{5.34}$$

But, because

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|X_t'P_{12}X_t| &= \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}Tr(P_{12}X_tX_t') \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\|P_{12}\|\|X_tX_t'\| \\
&= \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left\|\tilde{\Sigma}_{t*}(\theta)\tilde{\Sigma}_t^{-1}(\theta)\big(\Sigma_t(\theta)-\tilde{\Sigma}_t(\theta)\big)\tilde{\Sigma}_t^{-1}(\theta)\Sigma_{t*}(\theta)\right\|\|X_tX_t'\| \\
&\leq \frac{K}{n}\sum_{t=1}^{n}\rho^t\|\tilde{\Sigma}_{t*}\|\|\Sigma_t\|\|X_tX_t'\|,
\end{aligned}$$

similar to (5.33), we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|X_t'P_{12}X_t| \xrightarrow{a.s.} 0. \tag{5.35}$$

Combining (5.33), (5.34) and (5.35), we see that $P_1 \overset{a.s.}{\to} 0$.

We turn to P_2 now. Notice that $\log(1+x) \leq x$ for $x \geq -1$. We have

$$\begin{aligned}
\log|R_t(\theta)| - \log|\tilde{R}_t(\theta)| &= \log\left|I_m + (R_t(\theta)-\tilde{R}_t(\theta))\tilde{R}_t^{-1}(\theta)\right| \\
&\leq m\log\left\|I_m + (R_t(\theta)-\tilde{R}_t(\theta))\tilde{R}_t^{-1}(\theta)\right\| \\
&\leq m\log\left(\sqrt{m} + \|R_t(\theta)-\tilde{R}_t(\theta))\|\|\tilde{R}_t^{-1}(\theta)\|\right) \\
&\leq K + m\log\left(1 + \|R_t(\theta)-\tilde{R}_t(\theta))\|\|\tilde{R}_t^{-1}(\theta)\|/\sqrt{m}\right) \\
&\leq K + \sqrt{m}\|R_t(\theta)-\tilde{R}_t(\theta))\|\|\tilde{R}_t^{-1}(\theta)\|.
\end{aligned}$$

Symmetrically, we get

$$\log|\tilde{R}_t(\theta)| - \log|R_t(\theta)| \leq m\|R_t(\theta)-\tilde{R}_t(\theta))\|\|R_t^{-1}(\theta)\|.$$

Using similar argument to the proof of $P_1 \overset{a.s.}{\to} 0$, we can get $P_2 \overset{a.s.}{\to} 0$. Thus, the proof of C1 is completed.

Proof of C2. We first show that $El_t(\theta)$ is well defined for all $\theta \in \Theta$. Noting that for positive semi-definite matrices G_1 and G_2, $|G_1 + G_2| \geq \max\{|G_1|, G_2|\}$, we have

$$\begin{aligned} El_t^-(\theta) &\leq E \log^- |R_t(\theta)| \\ &\leq E \max\big\{0, -\log |\Sigma_{t*}(\theta)|^{-2} - \log |\Sigma_t(\theta)|\big\} \\ &\leq E \max\big\{0, |\Sigma_{t*}(\theta)|^{\delta} - \log |C|\big\} \\ &< +\infty, \end{aligned}$$

which implies that $El_t(\theta)$ is well defined. Furthermore, we have

$$\begin{aligned} E\big|\log |R_t(\theta_0)|\big| &= E\big|\log |\Sigma_{t*}^{-2}(\theta_0)| + \log |\Sigma_t(\theta_0)|\big| \\ &\leq |\log |C|| + |\log |C_*|| + E|\Sigma_{t*}(\theta_0)| + E|\Sigma_t(\theta_0)| \\ &< +\infty, \end{aligned}$$

where $C_* = diag\{c_{ii}, i = 1, \cdots, m\}$. Therefore,

$$E|l_t(\theta_0)| \leq E|X_t' R_t(\theta_0) X_t| + E\big|\log |R_t(\theta_0)|\big| = 1 + E\big|\log |R_t(\theta_0)|\big| < \infty.$$

Let $\lambda_{i,t}$ be the eigenvalues of $R_t^{1/2}(\theta_0)R_t^{-1}(\theta)R_t^{1/2}(\theta_0)$. We have $\lambda_{i,t} > 0$ since $R_t^{1/2}(\theta_0)R_t^{-1}(\theta)R_t^{1/2}(\theta_0)$ is positive definite. Thus, we have

$$\begin{aligned} & El_t(\theta) - El_t(\theta_0) \\ = \;& E \log \frac{|R_t(\theta)|}{|R_t(\theta_0)|} + E\Big\{\eta_t'\big[R_t^{1/2}(\theta_0)R_t^{-1}(\theta)R_t^{1/2}(\theta_0) - I_m\big]\eta_t\Big\} \\ = \;& E \log \big|R_t^{1/2}(\theta_0)R_t^{-1}(\theta)R_t^{1/2}(\theta_0)\big| + Tr\Big\{E\big[R_t^{1/2}(\theta_0)R_t^{-1}(\theta)R_t^{1/2}(\theta_0) - I_m\big]E\eta_t\eta_t'\Big\} \\ = \;& E\big[\sum_{i=1}^{m}(\lambda_{i,t} - 1 - \log \lambda_{i,t})\big] \\ \geq \;& 0, \end{aligned}$$

due to $\log x \leq x - 1$ for any $x > 0$. Furthermore, $\log x = x - 1$ if and only if $x = 1$, so the inequality above is strict unless for all i, $\lambda_{i,t} = 1$ a.s., which implies $R_t(\theta) = R_t(\theta_0)$ a.s.. By Lemma 5.5, we obtain $\theta = \theta_0$ and thus assertion C2 holds.

Proof of C3. For any $\theta \in \Theta$ and any positive integer k, let $N_k(\theta)$ be the open ball with center θ and radius $1/k$. By C1, we have

$$\begin{aligned} \liminf_{n\to\infty} \inf_{\theta^* \in N_k(\theta)\cap\Theta} \tilde{L}_n(\theta^*) &\geq \liminf_{n\to\infty} \inf_{\theta^* \in N_k(\theta)\cap\Theta} L_n(\theta^*) - \limsup_{n\to\infty} \sup_{\theta^* \in \Theta} |L_n(\theta^*) - \tilde{L}_n(\theta^*)| \\ &\geq \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\theta^* \in N_k(\theta)\cap\Theta} l_t(\theta^*). \end{aligned}$$

By the ergodic theorem, we get

$$\lim_{n\to\infty}\inf \frac{1}{n}\sum_{t=1}^{n}\inf_{\theta^*\in N_k(\theta)\cap\Theta} l_t(\theta^*) = E\inf_{\theta^*\in N_k(\theta)\cap\Theta} l_1(\theta^*).$$

Due to the Beppo-Levi theorem, $E\inf_{\theta^*\in N_k(\theta)\cap\Theta} l_1(\theta^*)$ increases to $El_1(\theta)$ when k increases to ∞. By C2, we complete the proof of assertion C3.

Proof of N1. We show $H<\infty$ first. Note that, from (3.3) and (5.20),

$$\frac{\partial l_t(\theta)}{\partial\theta_i} = Tr\big(\dot{R}_{t,i}(\theta)R_t^{-1}(\theta) - X_tX_t'R_t^{-1}(\theta)\dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\big) \tag{5.36}$$

for $i=1,\cdots,d$. By Minkowski inequality and Lemma 5.6, we have

$$E\big[Tr\big(\dot{R}_{t,i}R_t^{-1}\big)\big]^2 \le E\big[\|\dot{R}_{t,i}\|^2\|R_t^{-1}\|^2\big] \le \big[E\|\dot{R}_{t,i}\|^{2+\frac{8}{\delta}}\big]^{\frac{\delta}{4+\delta}}E\big[\|R_t^{-1}\|^{2+\frac{\delta}{2}}\big]^{\frac{4}{4+\delta}} < +\infty \tag{5.37}$$

provided that $E\|\eta_t\|^{4+\delta}<\infty$ for some $\delta>0$. Similarly, we have

$$\begin{aligned} E\big[Tr\big(X_tX_t'R_t^{-1}\dot{R}_{t,i}R_t^{-1}\big)\big]^2 &= E\big[Tr\big(\eta_t\eta_t'R_t^{-1/2}\dot{R}_{t,i}R_t^{-1/2}\big)\big]^2 \\ &\le E\big[\|\eta_t\eta_t'\|^2\|R_t^{-1/2}\dot{R}_{t,i}R_t^{-1/2}\|^2\big] \\ &\le KE\big[Tr(R_t^{-1}\dot{R}_{t,i}R_t^{-1}\dot{R}_{t,i})\big] \\ &\le KE\big[\|(R_t^{-1}\dot{R}_{t,i}R_t^{-1}\dot{R}_{t,i})\|\big] < \infty. \end{aligned} \tag{5.38}$$

By (5.36), (5.37) and (5.38), we have

$$E\Big[\frac{\partial l_t(\theta_0)}{\partial\theta_i}\Big]^2 \le 2E\big[Tr\big(\dot{R}_{t,i}R_t^{-1}\big)\big]^2 + 2E\big[Tr\big(X_tX_t'R_t^{-1}\dot{R}_{t,i}R_t^{-1}\big)\big]^2 < \infty, \tag{5.39}$$

which implies

$$E\Big|\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\Big| \le \Big\{E\Big[\frac{\partial l_t(\theta_0)}{\partial\theta_i}\Big]^2E\Big[\frac{\partial l_t(\theta_0)}{\partial\theta_j}\Big]^2\Big\}^{1/2} < \infty.$$

Therefore, $H<\infty$.

On the other hand, by (5.36), we have

$$\begin{aligned} E\Big(\frac{\partial l_t(\theta_0)}{\partial\theta}\Big|\mathcal{F}_{t-1}\Big) &= E\Big[Tr\big(\dot{R}_{t,i}R_t^{-1} - R_t^{1/2}\eta_t\eta_t'R_t^{-1/2}\dot{R}_{t,i}R_t^{-1}\big)\Big|\mathcal{F}_{t-1}\Big] \\ &= Tr\big(\dot{R}_{t,i}R_t^{-1} - R_t^{1/2}I_mR_t^{-1/2}\dot{R}_{t,i}R_t^{-1}\big) \\ &= 0. \end{aligned}$$

Thus $\partial l_t(\theta_0)/\partial\theta$ are stationary martingale differences. Applying the martingale central limit theorem, we get the assertion N1.

Proof of N2. From (5.36), it follows that

$$\begin{aligned}\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j} &= Tr\Big(\ddot{R}_{t,ij}(\theta)R_t^{-1}(\theta) - \dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\\ &\quad + X_tX_t'R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\dot{R}_{t,i}(\theta)R_t^{-1}(\theta) - X_tX_t'R_t^{-1}(\theta)\ddot{R}_{t,ij}(\theta)R_t^{-1}(\theta)\\ &\quad + X_tX_t'R_t^{-1}(\theta)\dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\Big). \qquad (5.40)\end{aligned}$$

By Lemma 5.6, similarly with (5.37), we obtain that

$$\begin{aligned} & E\big|Tr\big(X_tX_t'R_t^{-1}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1}\big)\big| \\ = \ & E\big|Tr\big(\eta_t\eta_t'R_t^{-1/2}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1/2}\big)\big| \\ \le \ & E\Big[\|\eta_t\eta_t'\|\|R_t^{-1/2}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1/2}\|\Big] \\ \le \ & KE\Big[\|R_t^{-1/2}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1/2}\|\Big] \\ = \ & KE\big[Tr\big(R_t^{-1/2}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1/2}\big)'\big(R_t^{-1/2}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1/2}\big)\big]^{1/2} \\ = \ & KE\big[Tr\big(\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1}\dot{R}_{t,j}R_t^{-1}\dot{R}_{t,i}R_t^{-1}\big)\big]^{1/2} \\ \le \ & KE\big[\|\dot{R}_{t,i}\|\|\dot{R}_{t,j}\|\|R_t^{-1}\|^2\big] \\ < \ & \infty, \end{aligned}$$

provided that $E\|\eta_t\|^{4+\delta} < \infty$ for some $\delta > 0$. If $E\|\eta_t\|^{4+\eta} < \infty$, we can similarly show that the expectation of the norms for the other terms in the right side of (5.40) is finite. Thus,

$$E\Big|\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}\Big| < \infty.$$

On the other hand, from (5.40) we have ,

$$E\Big(\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}\Big|\mathcal{F}_{t-1}\Big) = Tr\big(\dot{R}_{t,i}R_t^{-1}\dot{R}_{t,j}R_t^{-1}\big), \qquad (5.41)$$

which implies that $J = \Big(E\big[Tr\big(\dot{R}_{t,i}R_t^{-1}\dot{R}_{t,j}R_t^{-1}\big)\big]\Big)_{i,j=1,\cdots,d}$. Similar proof of $E\big|\partial^2 l_t(\theta_0)/\partial\theta_i\partial\theta_j\big| < \infty$ yields $J < \infty$.

Next, we will prove J is positive definite. Let $\mathcal{A}_{t,i} = R^{-\frac{1}{2}}\dot{R}_{t,i}R_t^{-\frac{1}{2}}$, where $R^{-\frac{1}{2}}$ is the unique positive definite square root of R_t^{-1}. From (5.22), we have $\dot{R}_{t,i}$ is symmetric. Using that $Tr(AB) = vec(A')'vec(B)$ and $vec(ABC) = (C' \otimes A)vec(B)$, we have

$$\begin{aligned} Tr(\mathcal{A}_{t,i}\mathcal{A}_{t,j}) &= vec(\mathcal{A}_{t,i})'vec(\mathcal{A}_{t,j}) \\ &= vec(\dot{R}_{t,i})'(R_t^{-\frac{1}{2}} \otimes R_t^{-\frac{1}{2}})(R_t^{-\frac{1}{2}} \otimes R_t^{-\frac{1}{2}})vec(\dot{R}_{t,j}) \\ &= vec(\dot{R}_{t,i})'(R_t^{-1} \otimes R_t^{-1})vec(\dot{R}_{t,j}). \end{aligned}$$

Let $\mathcal{P}_t' = \big(vec(\dot{R}_{t,1}), \cdots, vec(\dot{R}_{t,d})\big)$ and we have

$$E\Big(\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}|\mathcal{F}_{t-1}\Big) = \mathcal{P}_t(R_t^{-1}\otimes R_t^{-1})\mathcal{P}_t'. \tag{5.42}$$

Since R_t^{-1} is positive definite, we have $R_t^{-1}\otimes R_t^{-1}$ is positive definite using the fact that the eigenvalues of $A\otimes B$ for any matrices A and B are $\lambda_i\nu_j$, where λ_i and ν_j are the eigenvalues of A and B respectively. It follows that $\mathcal{P}_t(R_t^{-1}\otimes R_t^{-1})\mathcal{P}_t'$ is at least positive semi-definite. If there exists a constant vector x such that $x'Jx = 0$, due to (5.42) we have

$$E\big[(\mathcal{P}_t x)'(R_t^{-1}\otimes R_t^{-1})\mathcal{P}_t x\big] = 0.$$

As the term under the expectation is non-negative, it is necessarily zero. Together with the positive definiteness of $R_t^{-1}\otimes R_t^{-1}$, we obtain

$$\mathcal{P}_t x = 0 \quad \text{a.s.}$$

Recall the definition of $\mathcal{P}_t$, we have

$$\sum_{k=1}^{d} x_k \dot{R}_{t,k} = 0 \quad \text{a.s.}$$

By Lemma 5.7, we have $x = 0$, which means that J is positive definite. This proves assertion N2.

Proof of N3. From (5.40), we obtain that the third order derivatives of the log-likelihood involves such terms as

$$Tr\big(\ddot{R}_{t,ij}(\theta)R_t^{-1}(\theta)\dot{R}_{t,k}(\theta)R_t^{-1}(\theta)\big),$$
$$Tr\big(\dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\dot{R}_{t,k}(\theta)R_t^{-1}(\theta)\big),$$
$$Tr\big(\dddot{R}_{t,ijk}(\theta)R_t^{-1}(\theta)\big),$$

and the traces of the same matrices pre-multiplied by $X_tX_t'R_t^{-1}(\theta)$. Therefore, for instance,

$$\begin{aligned}
& E\big[\sup_{\theta\in\Theta} Tr\big(|X_tX_t'R_t^{-1}(\theta)\dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\dot{R}_{t,k}(\theta)R_t^{-1}(\theta)\big)|\big] \\
\le\ & E\sup_{\theta\in\Theta}\Big[\big\|R_t^{1/2}\eta_t\eta_t'R_t^{1/2}\big\|\big\|R_t^{-1}(\theta)\dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\dot{R}_{t,k}(\theta)R_t^{-1}(\theta)\big)\big\|\Big] \\
\le\ & KE\sup_{\theta\in\Theta}\Big[\big\|R_t^{-1}(\theta)\dot{R}_{t,i}(\theta)R_t^{-1}(\theta)\dot{R}_{t,j}(\theta)R_t^{-1}(\theta)\dot{R}_{t,k}(\theta)R_t^{-1}(\theta)\big)\big\|\Big] \\
<\ & KE\sup_{\theta\in\Theta}\Big[\big\|R_t^{-1}(\theta)\big\|^4\big\|\dot{R}_{t,i}(\theta)\big\|\big\|\dot{R}_{t,j}(\theta)\big\|\big\|\dot{R}_{t,k}(\theta)\big\|\Big] \\
<\ & \infty
\end{aligned}$$

by Lemma 5.6 provided that $E\|\eta_t\|^{8+\delta} < \infty$ for some $\delta > 0$. Using similar method, we can show that if $E\|\eta_t\|^{8+\delta} < \infty$ for some $\delta > 0$,

$$E\big[\sup_{\theta\in\Theta} Tr\big(\big|X_t X_t' R_t^{-1}(\theta)\ddot{R}_{t,ij}(\theta)R_t^{-1}(\theta)\dot{R}_{t,k}(\theta)R_t^{-1}(\theta)\big)\big|\big] < \infty$$

and

$$E\big[\sup_{\theta\in\Theta} Tr\big(\big|X_t X_t' R_t^{-1}(\theta)\dddot{R}_{t,ijk}(\theta)R_t^{-1}(\theta)\big)\big|\big] < \infty.$$

Therefore,

$$E\sup_{\theta\in\Theta}\Big|\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\Big| < \infty$$

for all $i, j, k = 1, \cdots, d$.

Proof of N4. From (5.36), we deduce that

$$\Big|\frac{\partial l_t(\theta_0)}{\partial\theta_i} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta_i}\Big| = \big|Tr\big(\dot{R}_{t,i}R_t^{-1}\big) - Tr\big(\dot{\tilde{R}}_{t,i}\tilde{R}_t^{-1}\big) - X_t' R_t^{-1}\dot{R}_{t,i}R_t^{-1}X_t + X_t'\tilde{R}_t^{-1}\dot{\tilde{R}}_{t,i}\tilde{R}_t^{-1}X_t\big|.$$

There are thus two types of terms to study. First, due to (5.22), we have

$$\begin{aligned}
&\big|Tr\big(\dot{R}_{t,i}R_t^{-1}\big) - Tr\big(\dot{\tilde{R}}_{t,i}\tilde{R}_t^{-1}\big)\big| \\
= \; &\big|Tr\big(-2\Sigma_{t*}^{-1}\dot{\Sigma}_{t*,i} + \dot{\Sigma}_{t,i}\Sigma_t^{-1}\big) - Tr\big(-2\tilde{\Sigma}_{t*}^{-1}\dot{\tilde{\Sigma}}_{t*,i} + \dot{\tilde{\Sigma}}_{t,i}\tilde{\Sigma}_t^{-1}\big)\big| \\
\le \; &2\big|Tr\big(\Sigma_{t*}^{-1}\dot{\Sigma}_{t*,i} - \tilde{\Sigma}_{t*}^{-1}\dot{\tilde{\Sigma}}_{t*,i}\big)\big| + \big|Tr\big(\dot{\Sigma}_{t,i}\Sigma_t^{-1} - \dot{\tilde{\Sigma}}_{t,i}\tilde{\Sigma}_t^{-1}\big)\big|.
\end{aligned}$$

Similarly to the proof of Theorem 4 in Comte and Lieberman (2003), we can show that

$$\sup_{\theta\in\Theta}\big\|\dot{\Sigma}_{t*,i}(\theta) - \dot{\tilde{\Sigma}}_{t,i}(\theta)\big\| \le K t^{d_0}\rho^{t-1}, \tag{5.43}$$

where $d_0 = \max(p, q)d(d+1)/2$, and K is a random variable which depends on the past values $(X_t, t \le 0)$ and can be then treated as a constant since it does not depend on n. Using the same method as the proof of assertion C1, we can obtain that

$$\frac{1}{n}\sum_{t=1}^{n}\big|Tr\big(\dot{R}_{t,i}R_t^{-1}\big) - Tr\big(\dot{\tilde{R}}_{t,i}\tilde{R}_t^{-1}\big)\big| \xrightarrow{P} 0.$$

Second, the same method is suitable for $\big|X_t' R_t^{-1}\dot{R}_{t,i}R_t^{-1}X_t - X_t'\tilde{R}_t^{-1}\dot{\tilde{R}}_{t,i}\tilde{R}_t^{-1}X_t\big|$. Thus, the first equality in N4 holds. Noting that the result in (5.43) can be extended to $\ddot{\Sigma}_t$, namely

$$\sup_{\theta\in\Theta}\big\|\ddot{\Sigma}_{t*,i}(\theta) - \ddot{\tilde{\Sigma}}_{t,i}(\theta)\big\| \le K t^{d_0}\rho^{t-1}.$$

Then, applying the same method as above, we know that the second equality in N4 holds.

ACKNOWLEDGEMENTS

The research of Hui Wang is supported by National Natural Science Foundation of China (No. 71771224), National Social Science Foundation of China (No. 15BGJ037), the Program for National Statistics Science Research Plan (No. 2016LD02), the program for Innovation Research in Central University of Finance and Economics.

References

[1] Aielli, G.P.(2013). Dynamic conditional correlations: on properties and estimation. *Journal of Business and Economic Statistics*, **31**, 282-299.

[2] Aielli, G.P. and Caporin, M.(2014). Variance clustering improved dynamic conditional correlation MGARCH estimators. *Computational Statistics & Data Analysis*, **76**, 556-576.

[3] Bauwens, L., Laurent, S. and Rombouts, J.V.K.(2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, **21**, 79-109.

[4] Berkes I, Horváth, L. and Kokoszka P.(2003). GARCH processes: Structure and estimation. *Bernoulli*, **9**, 201-227.

[5] Bollerlsev, T. (1990). Modeling the coherence in short run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, **72**, 498-505.

[6] Bollerlsev, T., Engle, R. and Wooldridge, J. (1988). A capital asset pricing mode1 with time varying covariances. *Journal of Political Economy*, **96**, 116-131.

[7] Bollerslev, T and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**, 143-172.

[8] Boussama, F., Fuchs, F. and Stelzer, R. (2011). Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models. *Stochastic Processes and their Applications*, **121**(10), 2331-2360.

[9] Brownlees,C. and Engle, R.F.(2017). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *The Review of Financial Studies* , **30**(1), 48-79.

[10] Cappiello, L., Engle R.F. and Sheppard, K. (2006). Dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, **4**, 537-572.

[11] Comte, F. and Lieberman, O.(2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, **84**, 61-84.

[12] Diebold, F.X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4**, 1-21.

[13] Dennis, J.G., Hansen, E. and Rahbek A.(2002). ARCH innovations and their impact on cointegration rank testing. Working paper no. 22, Centre for Analytical Finance, University of Aarhus.

[14] Doz, C. and Renault, E. (2004). Conditionally heteroskedastic factor models: Identification and instrumental variables estimation. CIRANO working paper.

[15] Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, **20**, 339-350.

[16] Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variance. *Econometric Reviews*, **5**, 1-50.

[17] Engle, R. F. and Kroner, K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11**, 122-150.

[18] Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Preprint.

[19] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag: New York.

[20] Fiorentini, G. and Sentana, E. (2001). Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics*, **102**, 143-164.

[21] Francq, C. and Zakoïan, J.M. (2010). *GARCH models: Structure, statistical inference and financial applications*. John Wiley & Sons.

[22] Francq, C. and Zakoïan, J.M. (2011). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory*, **27**(4), 1-28.

[23] Francq, C. and Zakoïan, J.M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, **10**, 605-637.

[24] Francq, C. and Zakoïan, J.M. (2016). Estimating multivariate volatility models equation by equation. *Journal of Statistical Society: Series B, Statistical Methodology* , **78**, 613-635.

[25] Franses, P.H., and Hafner, C.M. (2009). A generalized dynamic conditional correlation model for many ssset returns. *Econometric Reviews*, **28**, 612-631.

[26] Fernamian, J.D., and Malongo, H. (2017).On the stationary of dynamic conditional correlation models. *Econometric Theory*, 33, 636-663.

[27] Grossman, S. J. and Shiller, R. J.(1981). The determinants of the variability of stock market prices. *American Economic Review*, **71**, 222-227.

[28] Hafner, C.M. (2003) Fourth moment structure of Multivariate GARCH models. *Journal of Financial Econometrics*, **1**(1), 26-54.

[29] Hafner, C.M. and Preminger, A. (2009) On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis*, **1**(100), 2044-2054.

[30] He, C. and Teräsvirta, T. (2004). An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory*, **20**, 904-926.

[31] Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, **14**, 70-86.

[32] Kristensen, D. (2007). Uniform ergodicity of a class of Markov chains with applications to time series models. Working paper, Economics department, Columbia University.

[33] Lien, D. and Yang, L.(2008). Asymmetric effect of basis on dynamic futures hedging: Empirical evidence from commodity markets. *Journal of Banking and Finance*, **32**, 187-198.

[34] Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, **19**, 280-310.

[35] Magnus, J.R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.

[36] McAleer, M., Hoti, S. and Chan, F. (2009). Structure and Asymptotic Theory for Multivariate Asymmetric Conditional Volity. *Econometric Reviews*, **28**, 422-440.

[37] Sherrer,W. and Ribarits, E. (2007). On the Parametrization of Multivariate GARCH Models. *Econometric Theory*, **23**, 464-484.

[38] Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH Models. *Handbook of Financial Time Series* T.G. Andersen, R.A. Davis, J-P. Kreiss and T. Mikosch, eds. New York: Springer.

[39] Tse, Y. K., and Tsui, K. C. (2002). A multivariate generalized autoregressive conditional heteroskedasticity model with time-varying correlation. *Journal of Business and Economic Statistics*, **20**, 315-362.

[40] Tweedie, R. L.(1988). Invariant measure for Markov chains with no irreductibility assumptions. *Journal of Applied Probability*, **25**, 275-285.

[41] Zakoïan, J. M. (2010). A class of DCC asymmetric GARCH models driven by exogenous variables. *Rapport de Recherche, RR-FiME-10-12*. Available at SRRN: http://www.fime-lab.org/fr/rapports-de-recherche.

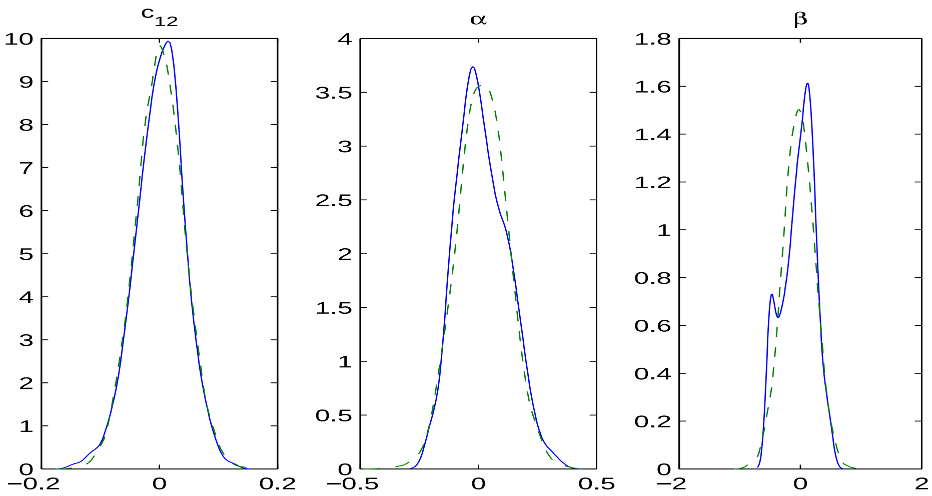

Figure 1: Kernel density estimator (full line) of the distribution of the QMLE errors for the estimation of c_{12}(the left figure), α (the middle figure) and β (the right figure) and Gaussian density (dotted line) with the same mean and variance.

TABLE 1

Estimation results of QMLE for SDCC(1, 1) model

parameter	true value	means	bias	RMSE
c_{12}	0.4	0.4007	0.0007	0.0411
α	0.2	0.2164	0.0164	0.1064
β	0.5	0.4708	-0.0292	0.2593

TABLE 2

Unit root and cointegration tests of FSTE 100 futures and spot prices

	lns	lnf	Rs	Rf	Error
ADF	-2.4039	-2.5240	-16.8242^{***}	-14.5936^{***}	-4.1916^{***}

Note: ***, ** and * refer to significance at level 1%, 5% and 10% respectively and we use the same notation in the following. The critical values at 1%, 5% and 10% for the ADF test are -3.44, -2.86 and -2.57 respectively.

TABLE 3

Summary statistics of FSTE 100 futures and spot returns

	Mean	SD	Skewness	Kurtosis	JB	Q(10)	$Q^2(10)$
Rs	0.0000	0.0123	-0.3821	5.5880	151.095^{***}	15.994^{*}	97.889^{***}
Rf	0.0001	0.0145	-0.6685	9.4576	902.375^{***}	39.108^{***}	207.488^{***}

Note: The critical value at 1% for the JB test is 9.2103. The critical values at 1%, 5% and 10% for Ljung & Box test are 23.209 18.307and 15.987 respectively.

TABLE 4

Estimation results for conditional mean, conditional variance and correlation coefficients

Estimation results for conditional mean					
	ϕ_{i0}	ϕ_{i1}	ϕ_{i2}	ψ_{i1}	ψ_{i2}
rs	-0.0050	0.0004	-0.1750	0.1581	0.0759
	(-0.2323)	(0.0065)	(-3.2663^{***})	(3.3299^{***})	(1.6323^{*})
rf	0.0975	0.1755	-0.2144	-0.1265	0.1823
	(3.9840^{***})	(3.0252^{***})	(-4.3232^{***})	(-1.9669^{***})	(3.4373^{***})
Estimation results for conditional variance					
	ω_i	θ_{i1}	θ_{i2}	Q(10)	$Q^2(10)$
hs	0.0757	0.1485	0.7998	2.9131	4.2999
	(2.0842^{**})	(2.7711^{***})	(12.6570^{***})		
hf	0.0697	0.1148	0.8500	7.4618	13.9580
	(2.3103^{**})	(2.7920^{***})	(21.1420^{***})		
Estimation results for conditional correlation coefficients of DCC, GARCH-SDCC and STD-SDCC					
	α	β	c_{12}	LiMcLeod	
DCC	0.1266	0.8706		44.6077	
	(3.9920^{***})	(24.4071^{***})			
GARCH-SDCC	0.5013	0.4985	0.7444	34.1978	
	(1.5707^{*})	(1.7983^{**})	(20.4119^{***})		
STD-SDCC	0.4443	0.5556	0.5979	35.2534	
	(1.3568^{*})	(2.1698^{**})	(4.6148^{***})		

Note: The critical value at 10% for the LiMcLeod statistic is 51.8051.

TABLE 5

The hedging performance of DCC and SDCC

	in-sample		*out-sample*	
	HE	U	HE	U
DCC	0.4835	-3.1158	0.7540	-0.9566
GARCH-SDCC	0.5140	-2.9305	0.9228	-0.8108
STD-SDCC	0.4645	-2.9980	0.7871	-0.9530

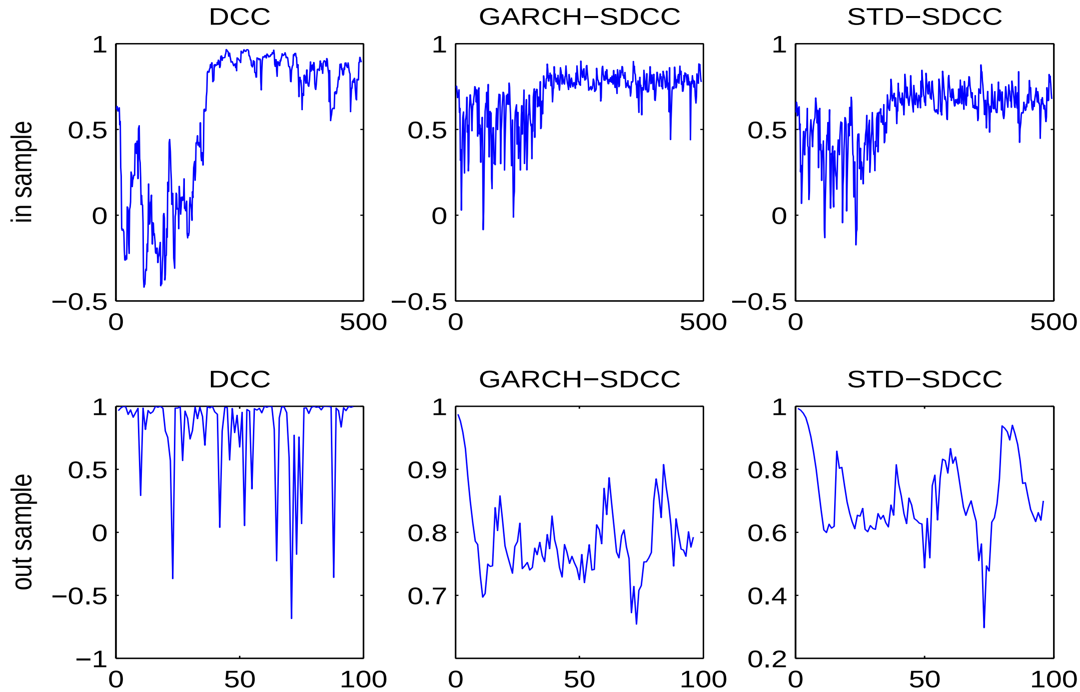

Figure 2: The estimated conditional correlation coefficients (the up panel) and the forecasting conditional correlation coefficients (the down panel) for the DCC (left), GARCH-SDCC (middle) and STD-DCC (right) .