이 학 박 사 학 위 논 문

Statistical inference in time series models with nonstandard mean and variance structure

평균 및 분산이 비표준형태인 시계열모형의 통계적 추론

2016년 8월

서울대학교 대학원

통계학과

정 환 식

Statistical inference in time series models with nonstandard mean and variance structure

평균 및 분산이 비표준형태인 시계열모형의 통계적 추론

지도교수 조 신 섭

이 논문을 이학박사 학위논문으로 제출함

2016년 **4**월

서울대학교 대학원

통계학과

정 환 식

정환식의 이학박사 학위논문을 인준함

2016년 **6**월

위 원 장	박 태 성	(인)
부위원장	조 신 섭	(인)
위 원	이 재 용	(인)
위 원	여 인 권	(인)
위 원	성 병 찬	(인)

Statistical inference in time series models with nonstandard mean and variance structure

By

Hwansik Jung

A Thesis
Submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
August, 2016

ABSTRACT

Statistical inference in time series models with nonstandard mean and variance structure

Hwansik Jung
The Department of Statistics
The Graduate School
Seoul National University

In this thesis, we consider inferences in time series model with non-standard mean and variance structure. First, an alternative GARCH model is proposed to handle the asymmetric leverage effect. The conditional variance of the proposed model consists of past conditional variances and squares of past transformed residuals. The Yeo-Johnson transformation is employed to model asymmetric leverage effect. Consistency and asymptotic normality of maximum likelihood estimator(MLE)s are derived. Real data is analyzed and the performance of the proposed model is compared with other GARCH-type models.

Second, the generalized method of moment(GMM) estimation is proposed for the cointegrated vector autoregressive(VAR) process of integrated order 1 where the process consists of endogenous variables and exogenous variables. Ahn et al. (2015) considered the MLE and the least squares estimation(LSE) of the cointegrated VAR processes assuming that the non-stationary exogenous variables are cointegrated. The same model considered by Ahn et al. (2015) was studied by the iterative GMM estimation method. The asymptotic properties of the GMM estimators are derived and the finite sample properties of the estimators are examined through a Monte Carlo simulation.

Keywords: Asymmetric GARCH, Yeo-Johnson transformation, Leverage effect, Cointegration, Generalized method of moments, Exogenuous variable
Student Number: 2009 – 20254

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A time series is a collection of random variables observed sequentially in time. Unlike many statistical models, which are developed assuming that the observations are independent, a time series analysis is concerned with describing the dependence among the elements of a sequence of random variables. Therefore, many time series models are developed based on an assumption of stationarity, meaning that the behavior does not depend on when we start to observe it. However, many economic time series data exhibit an non-stationary behavior, such as a stochastic trend or non-homogenous variances. In order to handle these nonstationary phenomena, many researchers have suggested the transformations of data, such as difference or a log transformation. In this thesis, we propose models for the inference of time series data with nonstandard mean and variance structures.

An autoregressive conditional heteroscedastic(ARCH) model was first introduced by Engle (1983) to explain heteroscedastic properties, e.g., clustering and time-varying properties of volatility. Bollerslev (1986) generalized it to propose a generalized ARCH(GARCH) model which is analogous to the extension of AR to ARMA models. Many researchers, however, have found that this model sometimes fails to capture the asymmetric property that is called the leverage effect. This led to the use of non-normal distributions, Student's t and skewed Student's t, within many non-linear extensions of the GARCH model, the exponential GARCH(EGARCH) of Nelson (1991), the GJR of Glosten et al. (1993) , and the asymmetric power ARCH(APARCH) of Ding et al. (1993), to better model the fat-tail(the excess kurtosis), the skewness, and the leverage effect. These models have been relevant in estimating and forecasting volatility as well as capturing asymmetry in volatility. However, Allen et al. (2014) studied the issue of asymmetry and leverage in conditional volatility models and showed that the leverage is not well handled in the GJR and EGARCH models. The APARCH model also has the same limitations as the Box-Cox transformation, which is the boundness and the positiveness of the transformed variables.

When we analyze vector time series data which are integrated we usually take a difference of the data. But if some of the linear combination of the series is stationary, a vector time series is said to be cointegrated. Granger (1981) and Engle and Granger (1987) proposed the concept of cointegration and developed a vector error correction model(VECM). Johansen (1988) and Ahn and Reinsel (1990) proposed a maximum likelihood estimation(MLE) based on the reduced rank approach, Phillips (1994) and Engle and Granger (1987) suggested the regression approach, and Quintos (1998), Kleibergen (1997), and Park et al. (2011) proposed GMM estimation methods. Descriptions and applications of these methods can be found in many financial empirical research. Recently, a cointegration model with exogeneity has received the attention of many authors. Since Hunter (1990) defined the cointegrating exogeneity, many researchers have developed new structural models. Pradel and Rault (2003) studied the strongly exogenous case, while Johansen (1992), Harbo et al. (1998), and Pesaran et al. (2000) studied the weakly exogenous case, Mosconi and Giannini (1992) studied the case of non-causality. But they used only the sufficient condition of exogenous assumptions and did not consider the case in which exogenous variables are cointegrated.Ahn et al. (2015) generalized the previous results and proposed MLE and LSE when the exogenous variables are cointegraed. However,according to Phillips (1994) MLE does not work properly when the data is huge and high-dimensional since the distributional assumption is easily violated and extraordinary outliers are easily encountered in the finite sample. Unlike the MLE, the GMM estimation does not need a distributional assumption and only requires the specification of moment conditions. Therefore, the GMM estimation is computationally convenient for the inference of a complex model.

In this thesis, a new asymmetric GARCH-type model is proposed to handle the leverage effect and the GMM estimation method is proposed for estimation of the cointegration model with exogenous variables.

The remainder of the thesis is organized as follows. Chapter 2 reviews the basic concepts that will aid the reader in understanding the models and methods used in this thesis. In chapter 3, an asymmetric GARCH, named YJ-GARCH, is proposed, and the asymptotic properties of the proposed model are derived. Real data is analyzed to compare the performance of the YJ-GARCH with other GARCH-type models in chapter 3. Chapter 4 contains the parameterization of the parameter sets and the iterative GMM estimation based on Ahn et al. (2015). Asymptotic properties of the iterative GMM estimators derived and the performance with MLE and LSE in finite samples are compared through a Monte Carlo simulation. Chapter 5 contains some concluding remarks.

Chapter 2

Literature Review

In this chapter, basic concepts useful to understand the models and methods used in this thesis are reviewed. Section 1 introduces the conditional heteroscedasticity and the family of transformation is summarized in section 2. In section 3, the concept of cointegration and the cointegrated model with exogenous variables are reviewed. Finally, we present the GMM estimation in section 4.

In the followings, $\xrightarrow{p}$ and $\xrightarrow{d}$ denotes convergence in probability and in distribution, respectively. $vec(\cdot)$ stacks the columns of a matrix into a column vector and $\otimes$ operator is a kronecker product.

2.1 Conditional heteroscedastic models

Consider an autoregressive process of order p, AR(p), for the observed variable X_t

$$X_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t,$$

where ϵ_t is a white noise process with

$$E(\epsilon_t) = 0$$
$$E(\epsilon_i \epsilon_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.1}$$

The process is covariance-stationary provided that the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

are outside the unit circle. The optimal linear forecast of the level of X_t for an AR(p) process is

$$E(X_t | X_{t-1}, X_{t-2} \cdots) = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p}, \tag{2.2}$$

where $E(X_t|X_{t-1}, X_{t-2}\cdots)$ denotes the linear projection of X_t on a constant and $(X_{t-1}, X_{t-2}\cdots)$. While the conditional mean of X_t changes over time according to (2.2), provided that the process is covariance-stationary, the unconditional mean of X_t is constant

$$E(X_t | X_{t-1}, X_{t-2} \cdots) = \mu / (1 - \phi_1 - \phi_2 - \cdots - \phi_p).$$

In the analysis of financial time series, investors often require higher expected returns as a compensation for holding risk assets. Since volatility means the conditional variance of the underlying asset, the modeling of the variance is as much important as the level of the series X_t. A variance that changes over time also has implications for the validity and efficiency of statistical inference about the parameters $(\mu, \phi_1, \phi_2, \cdots, \phi_p)$ that describe the dynamics of the level of X_t.

Although the unconditional variance of ϵ_t, σ^2, is constant in (2.1), the conditional variance of ϵ_t evolves over time. One approach to describe the behavior the conditional variance is denoting the square of ϵ_t as an AR(m) process

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_m \epsilon_{t-m}^2 + w_t, \tag{2.3}$$

where w_t is a new white noise process:

$$E(w_t) = 0$$

$$E(w_i w_j) = \begin{cases} \xi^2 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Since ϵ_t is an error in forecasting X_t, (2.3) implies that the linear projection of the squared error of a forecast of y_t on the previous m squared forecast errors is given by

$$E(\epsilon_t^2 | \epsilon_{t-1}^2, \epsilon_{t-2}^2 \cdots) = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_m \epsilon_{t-m}^2. \tag{2.4}$$

A white noise process ϵ_t satisfying (2.3) is described as an autoregressive conditional heteroscedastic process of order m, $ARCH(m)$, which is introduced by Engle (1983). Since ϵ_t is random and ϵ_t^2 cannot be negative, this is a sensible representation only if (2.4) is positive and (2.3) is nonnegative for all realizations of ϵ_t. This can be ensured only if w_t is bounded from below by $-\alpha_0$ with $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, 2, \cdots, m$. In order for ϵ_t^2 to be covariance-stationary, we further require that the roots of

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m = 0$$

lie outside the unit circle, which is equivalent to

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1, \tag{2.5}$$

if α_i's are nonnegative. When this condition is satisfied, the unconditional variance of ϵ_t is given by

$$\sigma^2 = E(\epsilon_t^2) = \alpha_0 / (1 - \alpha_1 - \alpha_2 - \cdots - \alpha_m).$$

Let $\hat{\epsilon}_{t+s|t}$ denote a $s-$period-ahead linear forecast

$$\hat{\epsilon}_{t+s|t} = E(\epsilon^2_{t+s}|\epsilon^2_t, \epsilon^2_{t-1}, \cdots).$$

This can be calculated iteratively. Then the $s-$period-ahead forecast $\hat{\epsilon}_{t+s|t}$ converges in probability to σ^2 as $s \to \infty$, assuming that w_t has finite variance and the condition (2.5) is satisfied.

It is often convenient to use an alternative represenation of an ARCH(m) process that imposes slightly stronger assumptions on the serial dependence of ϵ_t. Let

$$\epsilon_t = \sqrt{h_t} e_t,$$

where e_t is an i.i.d. sequence with zero mean and unit variance, then h_t evolves according to

$$h_t = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \cdots + \alpha_m \epsilon^2_{t-m}, \tag{2.6}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$. The coefficients α_i must satisfy these regularity conditions to ensure that the unconditional variance of ϵ_t is finite.

Although (2.3) and (2.6) are useful to describe the behavior of volatility, a relatively long lag is needed in the applications. To avoid this problem Bollerslev (1986) introduced a useful extension of ARCH known as a GARCH model as follows:

$$h_t = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \cdots + \alpha_m \epsilon^2_{t-m} + \beta_1 h_{t-1} + \cdots + \beta_s h_{t-s},$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i) < 1$. The last constraint on $\alpha_i + \beta_i$ ensures the finite unconditional variance of ϵ_t. The simple structure of GARCH model, however, has important drawbacks. GARCH models need an nonnegativity constraint on the parameters and does not accommodate the asymmetric relations in returns and volatility changes, for example. To accomodate the asymmetric effects between positive and negative residuals, Nelson (1991) proposed the EGARCH model

$$g(e_t) = \theta e_t + \gamma\Big(|e_t| - E\Big(|e_t|\Big)\Big) \tag{2.7}$$
$$log(h_t) = \alpha_0 + \frac{(1 + \alpha_1 B^1 + \cdots + \beta_a B^a)}{(1 - \beta_1 B^1 - \cdots - \beta_b B^b)} g(e_{t-1}),$$

where θ and γ are real constants. In (2.7), both e_t and $|e_t| - E(|e_t|)$ are i.i.d. random variable with mean zero. The EGARCH model differs from the GARCH model in several ways. First, it uses log conditional variance to relax the positiveness constraint of model coefficients. Second, the use of $g(e_t)$ enables the model to separate asymmetrically positive and negative lagged values of ϵ_t.

Another volatility model commonly used to handle the leverage effect is the threshold GARCH(TGARCH) model; see Glosten, Jagannathanm and Glosten et al. (1993) and Zakoian (1994). A TGARCH(m,s) model assumes the form

$$h_t = \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i N_{t-i})\epsilon_{t-1}^2 + \sum_{j=1}^{m} \beta_j h_{t-j},$$

where N_{t-i} is an indicator for negative ϵ_{t-i} such that

$$N_{t-i} = \begin{cases} 1 & \text{if } \epsilon_{t-i} < 0 \\ 0 & \text{if } \epsilon_{t-i} \geq 0, \end{cases}$$

α_i, β_i, and γ_i are nonnegative parameters satisfying conditions similar to those of GARCH model. Depending on whether past residual lies above or below zero, it has a different effect on the volatility model.

Even though EGARCH and TGARCH models are useful and simple for reflecting the leverage effect of positive and negative past residuals to volatility, however, when the data has a non-symmetric behavior these GARCH-type models do not work well. To avoid this weakness, transforming non-normal data to nearly normal before fitting GARCH-type model has been proposed. Asymmetric-power autoregressive conditional heteroscedastic(A-PARCH) introduced by Ding et al. (1993) is one of the ARCH family model which is useful to handle the asymmetry. The power term estimated within the A-PARCH is the means by which the data is transformed. The power term captures the volatility clustering, which is the influence of the outliers. Traditionally, data transformations involve a square term. But when data is non-normal or when it is not possible to characterize the distribution by the mean and variance, the use of a squared power transformation is not appropriate and other power transformations which use higher moments to adequately describe the distribution are needed.

2.2 Family of Power Transformations

Many statistical models are developed under the strong assumptions about the structure of data. In practice, these assumptions often fail to hold. One of the solution is to develope flexible alternatives which do not depend on the strong assumptions. Another is to transform data so that they satisfy the assumptions. The family of power transformation is reviewed in this section. Tukey (1957) introduced a family of power transformation such that the transformed values are a monotonic function of the observations over some admissible range and indexed by

$$X_t^{(\lambda)} = \begin{cases} X_t^{\lambda} \ \text{ if } \lambda \neq 0 \\ log(X_t) \ \text{ if } \lambda = 0, \end{cases} \tag{2.8}$$

for $X_t > 0$. This transformation was modified by Box and Cox (1964) to take account of the discontinuity at $\lambda = 0$ as follows:

$$X_t^{(\lambda)} = \begin{cases} (X_t^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ log(X_t) & \text{if } \lambda = 0. \end{cases} \tag{2.9}$$

It is noted that Box-Cox transformation is valid only for $X_t > 0$, hence modification is negative observations. Box and Cox (1982) proposed the shifted power transformation with the form

$$X_t^{(\lambda)} = \begin{cases} [(X_t + \lambda_2)^{\lambda_1} - 1]/\lambda_1 & \text{if } \lambda \neq 0 \\ log(X_t + \lambda_2) & \text{if } \lambda = 0, \end{cases} \tag{2.10}$$

where λ_1 is the transformation parameter and λ_2 is chosen such that $X_t > -\lambda_2$. The families (2.9) and (2.10) have the advantage that they are continuous functions of λ_1 and λ_2, respectively. Also, Box and Cox (1982) converted the selection of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ into an estimation problem. Draper and Cox (1969) attempted to derive the asymptotic properties of maximum likelihood estimator, $\hat{\lambda}$, in (2.9). Andrews et al. (1971) and (Velilla (1993),Velilla (1995)) discussed the transformation of multivariate observations. The transformation (2.10) is discussed in detail in Atkinson (1985) and Atkinson and Pericchi (1991).

Manly (1976) suggested another alternative which can be used with negative observations

$$X_t^{(\lambda)} = \begin{cases} [exp(\lambda X_t)] - 1/\lambda & \text{if } \lambda \neq 0 \\ X_t & \text{if } \lambda = 0. \end{cases}$$

He claimed that it is effective at transforming skewed unimodal distributions into nearly symmetric normal-like distributions. John and Draper (1980) introduced the so-called modulus transformation which is considered to normalize distributions already possessing some measure of approximate symmetry and carries the form

$$X_t^{(\lambda)} = \begin{cases} sign(X_t)[(|X_t|+1)^\lambda - 1/\lambda] & \text{if } \lambda \neq 0 \\ sign(X_t)log(|X_t|+1)^\lambda. \end{cases} \tag{2.11}$$

It is important to note that the range of $X_t^{(\lambda)}$ in (2.8)-(2.10) and (2.11) is restricted according to whether λ is positive or negative. This implies that the transformed values do not cover the entire range $(-\infty, \infty)$, hence their distributions are of bounded support. Consequently, only approximate normality is to be expected. But they are not appropriate for skewed data and change the prime properties of distribution like increase or sign.

Yeo and Johnson (2000) proposed an alternative family of power transformation which has properties similar to the Box-Cox transformation, in particular, convexity on the whole real line and is appropriate for reducing the skewness as follows:

$$\psi(\lambda, X_t) = \begin{cases} \dfrac{(X_t+1)^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0, X_t \geq 0 \\ log(X_t+1), & \text{if } \lambda = 0, X_t \geq 0 \\ -\dfrac{(-X_t+1)^{2-\lambda} - 1}{2-\lambda}, & \text{if } \lambda \neq 2, X_t < 0 \\ -log(-X_t+1), & \text{if } \lambda = 2, X_t < 0. \end{cases} \tag{2.12}$$

This transformation has the same form as Box-Cox transformation on the positive part, except the shift constant 1 in contained. The properties of the transformation are summarized in the following lemma.

Lemma 2.1. *(Yeo and Johnson, 2000) The transformation function $\psi(\cdot, \cdot)$ defined in (2.12) satisfies*

(1) $\psi(\lambda, X_t) \geq 0$ for $X_t \geq 0$ and $\psi(\lambda, X_t) < 0$ for $X_t < 0$.

(2) $\psi(\lambda, X_t)$ is convex in X_t for $\lambda > 1$ and concave in X_t for $\lambda < 1$.

(3) $\psi(\lambda, X_t)$ *is continuous function of* (λ, X_t).

(4) Let $\psi^{(k)} = \frac{\partial^k}{\partial \lambda^k}\psi(\lambda, x_t)$. *Then for* $k \geq 0$

$$\psi^{(k)} = \begin{cases} (X_t+1)^{\lambda} log^k(X_t+1) - k\psi^{(k-1)}/\lambda & if\ \lambda \neq 0, X_t \geq 0 \\ log^{k+1}(X_t+1)/(k+1) & if\ \lambda = 0, X_t \geq 0 \\ -(-X_t+1)^{2-\lambda}\big(-log(-X_t+1)\big)^k - k\psi^{(k-1)}/(2-\lambda) & if\ \lambda \neq 2, X_t < 0 \\ \big(-log(-X_t+1)\big)^{k+1}/(k+1) & if\ \lambda = 2, X_t < 0 \end{cases}$$

is continuous in (λ, X_t). *Here* $\psi^{(0)} \equiv \psi(\lambda, X_t)$.

(5) $\psi(\lambda, X_t)$ *is increasing in both* x *and* λ.

(6) $\psi(\lambda, X_t)$ *is convex in* λ *for* $X_t > 0$ *and concave in* λ *for* $X_t < 0$.

From Lemma 1, we can see that Yeo-Johnson transformation is monotone, continuous, and differentiable. These properties are useful to obtain the consistency and asymptotic results of the asymmetric GARCH model proposed in chapter 3.

In this thesis, a GARCH model with Yeo-Johnson transformation is proposed. In asymmetric GARCH modelling, we focus on the lagged error terms instead of independent or dependent terms. If all error terms are positive or vice versa, then some of asymmetric GARCH models just discriminate the sign of value and lead to the loss the information in error terms. For this reason, we suggest an alternative type of asymmetric GARCH model which can modify the distribution of error by Yeo-Johnson transformation and thus is useful in converting a skewed distribution to approximately symmetry. And it has quite satisfactory performance in transforming a near symmetric distribution to a normal distribution. As is well known, if a distribution of errors is normal then the skewness and the excess-kurtosis (i.e. kurtosis minus 3) are identical to zero. Thus a distribution of errors in financial time series which are far from normal would mean that at least one of the skewness or excess-kurtosis measures is significantly different from zero. Since Yeo-Johnson transformation would show good performance in skewness case, we expect that the proposed model(YJ-GARCH) specification can improve the estimates of the forecast intervals.

2.3 Cointegration model

Consider an $m-$dimensional vector process $\boldsymbol{X}_t$ generated by an autoregressive process of order p, VAR(p), given by

$$\Phi(L)\boldsymbol{X}_t = (I_m - \sum_{j=1}^{p} \Phi_j L^j)\boldsymbol{X}_t = \boldsymbol{\epsilon}_t, \tag{2.13}$$

where $\Phi_1, \cdots, \Phi_p$ are $m \times m$ coefficient matrices and L is a lag operator such that $L\boldsymbol{X}_t = \boldsymbol{X}_{t-1}$. $\boldsymbol{\epsilon}_t$ is a m-dimensional white noise vector with mean 0 and nonsingular covariance matrix Ω_ϵ.

VAR process is stable or called covariance-stationary if the characteristic polynomial equation $det(\Phi(L)) = 0$ has m characteristic roots which are outside the unit circle. When some roots of $det(\Phi(L)) = 0$ lie on or inside the unit circle, the VAR process is nonstationary. For details of the properties of VAR process, see Hamilton (1994) and Brüggemann and Lütkepohl (2005) among others. In general, any linear combination of nonstationary $\boldsymbol{X}_t$ will be also nonstationary. However, if there exists a vector B such that $B'\boldsymbol{X}_t$ is stationary, then $\boldsymbol{X}_t$ is cointegrated, following Engle and Granger (1987).

In this thesis, the cointegration models with exogenous variables are studied. Assume that $\det(\Phi(L)) = 0$ has d $(0 < d < m)$ unit roots and the remaining roots are outside the unit circle, i.e., $\text{rank}(\Phi(1)) = r(= m - d)$. This assumption implies that the $I_m - \Phi(1) = \sum_{j=1}^{p} \Phi_j$ has d liearly independent eigenvectors associated with eigenvalue 1. Under this assumption, the first difference of each component of $\boldsymbol{X}_t$ is stationary, while at least d components are integrated processes of order 1, denoted by $I(1)$. Then (2.13) can be represented as a vector error correction model(VECM) introduced by Engle and Granger (1987) as follows:

$$\Phi^*(L)(1-L)\boldsymbol{X}_t = -\Phi(1)\boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \tag{2.14}$$

where $\Phi^*(L) = I_m - \sum_{j=1}^{p-1} \Phi_j^* L^j$ with $\Phi_j^* = -\sum_{k=j+1}^{p} \Phi_k$.

If $r = m - d > 0$, then, since $\det(\Phi(L)) = 0$ has d unit roots, there are $m \times m$ matrices P and $Q = P^{-1}$ such that $Q(\sum_{j=1}^{p} \Phi_j)P = J$, where $J = diag(I_d, \Lambda_r)$ is the Jordan canonical form of $\sum_{j=1}^{p} \Phi$. We can easily see that $\Phi(1) = I_m - \sum_{j=1}^{p} \Phi_j = P(I_m - J)Q = P_2(I_r - \Lambda_r)Q_2'$ when we partition $Q' = [Q_1, Q_2]$, $Q_1' = [Q_{11}', Q_{12}']$, $Q_2' = [Q_{21}', Q_{22}']$, $P = [P_1, P_2]$, $P_1 = [P_{11}, P_{12}]'$, and $P_2 = [P_{21}, P_{22}]'$ such that Q_1 and P_1 are $m \times r$, Q_{11}' and P_{11} are $r \times d$, and Q_{12}' and P_{21} are $d \times d$, Q_{21}' and P_{12} are $r \times r$, and Q_{22}' and P_{22} are $d \times r$ matrices. The rank of the $m \times m$ matrix $\Phi(1)$ is a reduced rank r. Let $C = -\Phi(1) = AB'$, then (2.14) can be written as

$$\begin{aligned}\Delta\boldsymbol{X}_t &= -\Phi(1)\boldsymbol{X}_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* \Delta\boldsymbol{X}_{t-j} + \boldsymbol{\epsilon}_t \\ &= AB'\boldsymbol{X}_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* \Delta\boldsymbol{X}_{t-j} + \boldsymbol{\epsilon}_t, \end{aligned} \tag{2.15}$$

where Δ is the difference operator such that $\Delta\boldsymbol{X}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}$. $C\boldsymbol{X}_{t-1}$ is called the error correction term which adjusts for over-differencing. $C = AB'$ is a full rank factorization of C since $m \times r$ matrices A and B have full rank r by the definition of A and B.

Ahn and Reinsel (1990) showed that $B'\boldsymbol{X}_{t-1}$ is stationary if $A = P_2(I_r - \Lambda_r)Q_{21}'$ and $B = Q_{21}'^{-1}Q_2'$. Since $B'\boldsymbol{X}_{t-1}$ is a $r-$dimensional stationary process, $\boldsymbol{X}_t$ is cointegrated with cointegrating rank r. The matrix B' is called the cointegrating matrix and A is called the adjust matrix, or loading matrix. Note that B' is the coefficient matrix of non-stationary process $\boldsymbol{X}_{t-1}$ and A can be treated as a coefficient matrix of stationary process $B'\boldsymbol{X}_{t-1}$.

In (2.15), there is no unique identification as $C = AB'$ and $C = A^*B'^* = AMM^{-1}B'$ for some $r \times r$ matrix M. In order to avoid this problem, many researchers considered different identifying conditions. Johansen (1988) proposed a method which is motivated by the canonical correlation analysis. $B'\Omega_\epsilon B = I_m$ is used as an identifying condition based on the MLE procedure and a closed form solution of maximum likelihood equation exists. It can be done in two steps. In the first step, $\Delta\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$ are regressed on $\Delta\boldsymbol{X}_{t-j}$'s and residuals are obtained. Residuals of the regression of $\Delta\boldsymbol{X}_t$ are regressed on those of $\boldsymbol{X}_{t-1}$ in the second step.

On the other hand, Ahn and Reinsel (1990) developed a method which is motivated by the reduced rank regression analysis. This method uses the identifying condition of the form $B' = [I_r, B_0']$. Based on this method, adjusted term is expressed as $B'\boldsymbol{X}_{t-1} = \boldsymbol{X}_{1,t-1} + B_0'\boldsymbol{X}_{2,t-1}$. It can be partitioned as $\boldsymbol{X}_t = [\boldsymbol{X}_{1,t}', \boldsymbol{X}_{2,t}']$ where $\boldsymbol{X}_{1,t}$ is $r \times 1$ matrix and $\boldsymbol{X}_{2,t}$ is $d \times 1$ matrix. For this normalization, it is assumed that $\boldsymbol{X}_t$ are arranged so that $\boldsymbol{X}_{2,t}$ is purely nonstationary, that is, not cointegrated.

Unlike Johansen (1988), Ahn and Reinsel (1990) developed an estimation procedure with LSE and MLE simultaneously using the Newton-Rahpson algorithm. Ahn and Reinsel (1990)'s LSE is given by

$$\tilde{F} = \Big(\sum_{t=1}^{T} \Delta\boldsymbol{X}_t \boldsymbol{U}_{t-1}'\Big)\Big(\sum_{t=1}^{T} \boldsymbol{U}_{t-1}\boldsymbol{U}_{t-1}'\Big), \tag{2.16}$$

where $F = [C, \Phi_1^*, \cdots, \Phi_{p-1}^*]$ and $\boldsymbol{U}_{t-1} = [\boldsymbol{X}_{t-1}', \Delta\boldsymbol{X}_{t-1}', \cdots, \Delta\boldsymbol{X}_{t-p+1}']'$. Ahn and Reinsel (1990) also obtained MLE using the Newton-Raphson algorithm as follows:

$$\hat{\eta}^{(i+1)} = \hat{\eta}^{(i)} - \Big(\sum \frac{\partial \epsilon'_{t-1}}{\partial \eta}\Omega_\epsilon^{-1}\frac{\partial \epsilon_{t-1}}{\partial \eta}\Big)_{\hat{\eta}^{(i)}}\Big(\sum \frac{\partial \epsilon'_{t-1}}{\partial \eta}\Omega_\epsilon^{-1}\epsilon_t\Big)_{\hat{\eta}^{(i)}},$$

where $\eta = (B', \theta')', B = vec(B'_0)$, and $\theta = vec[A, \Phi_1^*, \cdots, \Phi_{p-1}^*]$. The gradient vector $\partial\epsilon'_t/\partial\eta$ is driven as

$$\frac{\partial \epsilon'_t}{\partial \eta} = \begin{bmatrix} \boldsymbol{X}_{2t} \otimes A' \\ \tilde{\boldsymbol{U}}_{t-1} \otimes I_m \end{bmatrix},$$

where $\tilde{\boldsymbol{U}}_{t-1} = [(B'\boldsymbol{X}_{t-1})', \Delta\boldsymbol{X}'_{t-1}, \cdots, \Delta\boldsymbol{X}'_{t-p+1}]'$. The initial estimator $\hat{\eta}^{(0)}$ can be obtained using LSE $\tilde{F}$. Then the initial estimators of A and B are obtained as

$$\tilde{A} = \tilde{C}_1$$

and

$$\tilde{B}'_0 = (\tilde{A}'\tilde{\Omega}_\epsilon^{-1}\tilde{A})^{-1}\tilde{A}'\tilde{\Omega}_\epsilon^{-1}\tilde{C}_2,$$

where $\tilde{C} = [\tilde{C}_1, \tilde{C}_2]$, $\tilde{C}_1$ and $\tilde{C}_2$ are $m \times r$ and $m \times d$ matrices, respectively and $\tilde{\Omega}_a^{-1}$ is a sample variance of the regression residuals using (2.16). Ahn and Reinsel (1990) also derived the asymptotic distributions of the above estimators. The key in the derivation of the asymptotic distribution of the estimators is the *Lemma* 1 of Ahn and Reinsel (1990) as follows:

Lemma 2.2. *(Ahn and Reinsel) Let* $\boldsymbol{Z}_t = Q\boldsymbol{X}_t = [\boldsymbol{Z}'_{1,t}\boldsymbol{Z}'_{2,t}]'$ *and* $\boldsymbol{a}_t = [\boldsymbol{a}'_{1,t}, \boldsymbol{a}'_{2,t}]' = Q\boldsymbol{\epsilon}_t$ *such that* $\boldsymbol{Z}_{1,t}$ *and* $\boldsymbol{a}_{1,t}$ *are* $d \times 1$ *and* $\boldsymbol{Z}_{2,t}$ *and* $\boldsymbol{a}_{2,t}$ *are* $r \times 1$ *with* $\Omega_a = Q\Omega_\epsilon Q'$ *and* $\Omega_{a_1} = [I_d, 0]\Omega_a[I_d, 0]'$*. In addition, define* $\Psi_{11} = [I_d, 0]\Psi[I_d, 0]'$ *with* $\Psi = \sum_{k=1}^{\infty}\Psi_k$*,* Ψ_k *are the infinity moving average coefficients in the representation* $\boldsymbol{u}_t = \sum_{k=1}^{\infty}\Psi_k\boldsymbol{a}_{t-k}$ *for the stationary process* $\boldsymbol{u}_t = Q(\Delta\boldsymbol{X}_t - C\boldsymbol{X}_{t-1}) = \boldsymbol{Z}_t - diag(I_d, \Lambda_r)\boldsymbol{Z}_{t-1}$*, and* $B_m(\boldsymbol{u})$ *and* $B_d(\boldsymbol{u}) = \Omega_{a_1}^{-1/2}[I_d, 0]\Omega_{a_1}^{1/2}B_m(\boldsymbol{u})$ *be standard Brownian motions of dimensions* m *and* d

repectively. Under the assumption of model (2.14), the following distributional results hold:

(1) $T^{-2}\sum_{t=1}^{T}\boldsymbol{Z}_{1,t-1}\boldsymbol{Z}'_{1,t-1}\xrightarrow{d}\Psi_{11}\Omega_{a_1}^{1/2}\int_0^1 B_d(\boldsymbol{u})B_d(\boldsymbol{u})'d\boldsymbol{u}\Omega_{a_1}^{1/2}\Psi'_{11}=:\mathcal{B}_{zz}$.

(2) $T^{-1}\sum_{t=1}^{T}\boldsymbol{a}_t\boldsymbol{Z}'_{1,t-1}\xrightarrow{d}\Omega_a^{1/2}\Big[\int_0^1 B_d(\boldsymbol{u})dB_m(\boldsymbol{u})'\Big]'\Omega_{a_1}^{1/2}\Psi'_{11}=:\mathcal{B}_{az}$.

(3) $T^{-3/2}\sum_{t=1}^{T}\boldsymbol{U}_{t-1}\boldsymbol{Z}'_{1,t-1}\xrightarrow{p}0$.

(4) $T^{-1}\sum_{t=1}^{T}\boldsymbol{U}_{t-1}\boldsymbol{U}_{t-1}\xrightarrow{p}\Gamma_{\boldsymbol{U}}=Cov(U)$.

(5) $T^{-1/2}\sum_{t=1}^{T}vec(\boldsymbol{a}_t\boldsymbol{U}'_{t-1})\xrightarrow{d}N(0,\Gamma_U\otimes\Omega_a)$.

The asymptotic distribution of LSE $\tilde{F}$ and MLE $\hat{B}_0$ and $\hat{\theta}$ are summarized in Theorem 1 and Theorem 2 of Ahn and Reinsel (1990).

Consider the case where the cointegrated model $\boldsymbol{X}_t$ consists of enogenous and exogenous variables. Assume that $\boldsymbol{X}_t=[\boldsymbol{Y}'_t,\boldsymbol{Z}'_t]'$, where $\boldsymbol{Y}_t$ is an m_y- dimensional vector process of endogenous variables and $\boldsymbol{Z}_t$ is an m_z- dimensional vector process of exogenous variables with $m_y+m_z=m$. '$\boldsymbol{Z}_t$ is exogenous' means that $\boldsymbol{Z}_t$ is not affected by $\boldsymbol{Y}_t$ while $\boldsymbol{Y}_t$ is affected by $\boldsymbol{Z}_t$. Exogeneity is sometimes mixed up with causality. However, 'exogeneity' and 'causality' were perfectly distinguished by mathematical definition of Engle et al. (1983). They defined that $\boldsymbol{Z}_t$ is *exogenous* if and only if the conditional probability density function(pdf) of $\boldsymbol{X}_t$ given past information $(\boldsymbol{X}_{t-1},\cdots,\boldsymbol{X}_1)$ can be represented as

$$\begin{aligned} f(\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\cdots,\boldsymbol{X}_1;\theta) &= f(\boldsymbol{Y}_t|\boldsymbol{Z}_t\boldsymbol{X}_{t-1},\cdots,\boldsymbol{X}_1;\theta_1) \\ &\quad\times f(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1},\cdots,\boldsymbol{Z}_1;\theta_2), \end{aligned} \tag{2.17}$$

where θ_1 and θ_2 are *variation free*. In the cointegration anaylsis, (2.17) should be modified, since $\boldsymbol{X}_t=[\boldsymbol{Y}'_t,\boldsymbol{Z}'_t]'$ does not have a pdf since $\boldsymbol{X}_t$ is assumed as $I(1)$ process.

$$\begin{aligned} &f(\Delta \boldsymbol{X}_t | \boldsymbol{X}_{t-1}, \Delta \boldsymbol{X}_{t-1}, \cdots, \Delta \boldsymbol{X}_1; \theta) \\ &= f(\Delta \boldsymbol{Y}_t | \Delta \boldsymbol{Z}_t, \boldsymbol{X}_{t-1}, \Delta \boldsymbol{X}_{t-1}, \cdots, \Delta \boldsymbol{X}_1; \theta_1) \\ &\times f(\Delta \boldsymbol{Z}_t | \boldsymbol{Z}_{t-1}, \Delta \boldsymbol{Z}_{t-1}, \cdots, \Delta \boldsymbol{Z}_1; \theta_2). \end{aligned}$$

Ahn et al. (2015) considered the cointegrated model with exogenous variable $\boldsymbol{Z}_t$. The coefficient matrix $C, \Phi_j^*, \boldsymbol{\epsilon}_t$, and Ω in (2.15) are separated so that they are conformable to $\boldsymbol{X}_t = (\boldsymbol{Y}_t', \boldsymbol{Z}_t')'$ as follows:

$$C = \begin{bmatrix} C_{yy} & C_{yz} \\ 0 & C_{zz} \end{bmatrix}, \Phi_j^* = \begin{bmatrix} \Phi_j^{*yy} & \Phi_j^{*yz} \\ 0 & \Phi_j^{*zz} \end{bmatrix}$$
$$\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{yt}', \boldsymbol{\epsilon}_{zt}')', \Omega = \begin{bmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{bmatrix}.$$

They also left-multiplyed

$$\begin{bmatrix} I & -\Omega_{yz}\Omega_{zz}^{-1} \\ 0 & I \end{bmatrix}$$

on both sides of (2.14) and obtain

$$\begin{aligned} \begin{bmatrix} \Delta \boldsymbol{Y}_t \\ \Delta \boldsymbol{Z}_t \end{bmatrix} &= \begin{bmatrix} C_{yy} & C_{yz} - \Omega_{yz}\Omega_{zz}^{-1}C_{zz} \\ 0 & C_{zz} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_{t-1} \\ \boldsymbol{Z}_{t-1} \end{bmatrix} + \begin{bmatrix} \Omega_{yz}\Omega_{zz}^{-1} \\ 0 \end{bmatrix} \Delta \boldsymbol{Z}_t \\ &+ \sum_{j=1}^{p-1} \begin{bmatrix} \Phi_j^{*yy} & \Phi_j^{*yz} - \Omega_{yz}\Omega_{zz}^{-1}\Phi_j^{*zz} \\ 0 & \Phi_j^{*zz} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{Y}_{t-j} \\ \Delta \boldsymbol{Z}_{t-j} \end{bmatrix} \\ &+ \begin{bmatrix} \boldsymbol{\epsilon}_{yt} - \Omega_{yz}\Omega_{zz}^{-1}\boldsymbol{\epsilon}_{zt} \\ \boldsymbol{\epsilon}_{zt} \end{bmatrix}. \end{aligned}$$

Then the conditional model of $\Delta \boldsymbol{Y}_t$ and the marginal model of $\Delta \boldsymbol{Z}_t$ are

$$\begin{aligned} \Delta \boldsymbol{Y}_t &= (C_{yy}, C_{yz})\boldsymbol{X}_{t-1} - \Omega_{yz}\Omega_{zz}^{-1}C_{zz}\boldsymbol{Z}_{t-1} + \Omega_{yz}\Omega_{zz}^{-1}\Delta \boldsymbol{Z}_t \\ &+ \sum_{j=1}^{p-1} H_j \Delta \boldsymbol{X}_{t-j} + \boldsymbol{e}_{yt} \end{aligned} \tag{2.18}$$

and

$$\Delta \boldsymbol{Z}_t = C_{zz}\boldsymbol{Z}_{t-1} + \sum_{j=1}^{p-1} \Phi_j^{*zz}\Delta \boldsymbol{Z}_{t-j} + \boldsymbol{\epsilon}_{zt}, \tag{2.19}$$

where $H_j = (\Phi_j^{*yy}, \Phi_j^{*yz} - \Omega_{yz}\Omega_{zz}^{-1}\Phi_j^{*zz})$ and $\boldsymbol{e}_{yt} = \boldsymbol{\epsilon}_{yt} - \Omega_{yz}\Omega_{zz}^{-1}\boldsymbol{\epsilon}_{zt}$ which is uncorrelated with ϵ_{zt}.

If $C_{zz} = 0$ and $\Phi_j^{*zy} \neq 0$, the model (2.18) and (2.19) are variation free and $\boldsymbol{Z}_t$ is weakly exogenous. This assumption implies that the exogenous variables do not cointegrate among themselves and the marginal model (2.19) becomes

$$\Delta \boldsymbol{Z}_t = \sum_{j=1}^{p-1} \left[\Phi_j^{*zy}, \Phi_j^{*zz}\right]\Delta \boldsymbol{X}_{t-j} + \boldsymbol{\epsilon}_{zt}. \tag{2.20}$$

Johansen (1992), Harbo et al. (1998), and Pesaran et al. (2000) studied the estimation of parameters and cointegrating rank test procedure based on the models (2.18) and (2.20). Mosconi and Giannini (1992) developed the estimating method and testing procedure for model (2.15) with only $\Phi_j^{*zy} = 0$ and $C_{zy} = 0$. If Φ_j^{*zy} and C_{zz} are zero matrices, $\boldsymbol{Z}_t$ is strongly exogenous process. Pradel and Rault (2003) considered the case of strongly exogenous $\boldsymbol{Z}_t$ where exogenous variables do not cointegrated among themselves.

Ahn et al. (2015) considered the cointegrated models where nonstationary exogenous variables are cointegrated themselves based on (2.18) and (2.19). Since the entire process $\boldsymbol{X}_t$ and exogenous process $\boldsymbol{Z}_t$ have $d = m - r$ and $d_z = m_z - r_z$ unit roots, repectively, they could have the Jordan canonical form of $\sum_{j=1}^{p} \Phi_j = I_m - \Phi(1)$ as $Q\Big(\sum_{j=1}^{p} \Phi_j\Big)P = J$, where $J = diag(I_{m_y - r_y}, \Lambda_{r_y}, I_{m_z - r_z}, \Lambda_{r_z})$, and $r_y = r - r_z$, Λ_{r_y} and Λ_{r_z} are diagonal (Jordan block) matrices whose elements are the stationary roots of the autoregressive operator. Since $C = -\Phi(1)$ is an upper block triangular matrix, there exist upper block triangular matrices P and Q such that

$$P = \begin{bmatrix} P_y & P_{yz} \\ O & P_z \end{bmatrix}$$

$$Q = P^{-1} = \begin{bmatrix} P_y & P_{yz} \\ O & P_z \end{bmatrix} = \begin{bmatrix} P_y^{-1} & -P_y^{-1}P_{yz}P_z^{-1} \\ O & P_z^{-1} \end{bmatrix} \equiv \begin{bmatrix} Q_y & Q_{yz} \\ O & Q_z \end{bmatrix} = \begin{bmatrix} Q_y & -Q_yP_{yz}Q_z \\ O & Q_z \end{bmatrix}$$

$$\Phi(1) = \begin{bmatrix} \Phi_{yy}(1) & \Phi_{yz}(1) \\ 0 & \Phi_{zz}(1) \end{bmatrix} = \begin{bmatrix} -C_{yy} & -C_{yz} \\ 0 & -C_{zz} \end{bmatrix},$$

which can be partitioned as follows:

$$Q = \begin{bmatrix} Q_y & Q_{yz} \\ O & Q_z \end{bmatrix} = \begin{bmatrix} Q'_{y1} & Q'_{yz1} \\ Q'_{y2} & Q'_{yz2} \\ 0 & Q'_{z1} \\ 0 & Q'_{z2} \end{bmatrix}, \quad P = \begin{bmatrix} P_y & P_{yz} \\ O & P_z \end{bmatrix} = \begin{bmatrix} P_{y1} & P_{y1} & P_{yz1} & P_{yz2} \\ 0 & 0 & P_{z1} & P_{z2} \end{bmatrix},$$

where $Q'_{yz} = -[Q'_zP'_{yz}Q_{y1}, Q'_zP'_{yz}Q_{y2}]$, Q_{y1} and P_{y1} are $m_y \times d_y$, Q_{y2} and P_{y2} are $m_y \times r_y$, Q_{yz1} and P_{yz1} are $m_y \times d_z$, Q_{yz2} and P_{yz2} are $m_y \times r_z$, and Q_{z1} and P_{z1} are $m_z \times d_z$, and Q_{z2} and P_{z2} are $m_z \times r_z$ matrices, respectively. Using the Jordan canonical form $\Phi(1)$ can be represented as

$$\begin{aligned}
\Phi(1) &= P(I_m - J)Q \\
&= \begin{bmatrix} P_{y2} & P_{yz2} \\ 0 & P_{z2} \end{bmatrix} \begin{bmatrix} I_{r_y} - \Lambda_{r_y} & 0 \\ 0 & I_{r_z} - \Lambda_{r_z} \end{bmatrix} \begin{bmatrix} Q'_{y2} & Q'_{yz2} \\ 0 & Q'_{z2} \end{bmatrix} \\
&= \begin{bmatrix} P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y2} & P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{yz2} + P_{yz2}I_{r_z} - \Lambda_{r_z})Q'_{z2} \\ 0 & P_{z2}(I_{r_z} - \Lambda_{r_z})Q'_{z2} \end{bmatrix} \\
&= \begin{bmatrix} P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y2} & P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y2}P_{yz}Q_z + P_{yz2}(I_{r_z} - \Lambda_{r_z})Q'_{z2} \\ 0 & P_{z2}(I_{r_z} - \Lambda_{r_z})Q'_{z2} \end{bmatrix}.
\end{aligned}$$

and (2.18) and (2.19) can be rewritten as follows:

$$
\begin{aligned}
\Delta \boldsymbol{Y}_t =& - P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y2}\boldsymbol{Y}_{t-1} \\
&+ \Big[P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y2}P_{yz}Q_z - P_{yz2}(I_{r_z} - \Lambda_{r_z})Q'_{z2}\Big]\boldsymbol{Z}_{t-1} \\
&- DP_{z2}(I_{r_z} - \Lambda_{r_z})Q'_{z2}\boldsymbol{Z}_{t-1} + D\Delta\boldsymbol{Z}_t + \sum_{j=1}^{p-1} H_j \Delta \boldsymbol{X}_{t-j} + \boldsymbol{e}_{yt} \\
=& - P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y2}[I_{m_y}, -P_{yz}Q_z] \begin{bmatrix} \boldsymbol{Y}_{t-1} \\ \boldsymbol{Z}_{t-1} \end{bmatrix} - P_{yz2}(I_{r_z} - \Lambda_{r_z})Q'_{z2}\boldsymbol{Z}_{t-1} \\
&+ D(\Delta\boldsymbol{Z}_t - P_{z2}(I_{r_z} - \Lambda_{r_z})Q'_{z2}\boldsymbol{Z}_{t-1} + \sum_{j=1}^{p-1} H_j \Delta \boldsymbol{X}_{t-j} + \boldsymbol{e}_{yt}, \qquad (2.21)
\end{aligned}
$$

$$
\Delta \boldsymbol{Z}_t = - P_{z2}(I_{r_z} - \Lambda_{r_z})Q'_{z2}\boldsymbol{Z}_{t-1} + \sum_{j=1}^{p-1} \Phi_j^{**} \Delta \boldsymbol{Z}_{t-j} + \boldsymbol{\epsilon}_{zt}, \qquad (2.22)
$$

where $D = \Omega_{yz}\Omega_{zz}^{-1}$. $Q'_{y2}(I_{m_y}, -P_{yz}Q_z)\boldsymbol{X}_{t-1}$ and $Q'_{z2}\boldsymbol{Z}_{t-1}$ are stationary by a similar argument in Ahn and Reinsel (1990).

If we define $A = -P_{y2}(I_{r_y} - \Lambda_{r_y})Q'_{y21}$,$B' = Q'^{-1}_{y21}Q'_{y2}[I_{m_y}, -P_{yz}Q_z]$, $A_{2z} = -P_{yz2}(I_{r_z} - \Lambda_{r_z})Q'_{z21}$, $A_z = -P_{z2}(I_{r_z} - \Lambda_{r_z})Q'_{z21}$ and $B_z = Q'^{-1}_{z21}Q'_{z2}$ and also conformable partition

$$
Q_y = \begin{bmatrix} Q'_{y1} \\ Q'_{y2} \end{bmatrix} = \begin{bmatrix} Q'_{y11} & Q'_{y12} \\ Q'_{y21} & Q'_{y22} \end{bmatrix}, \quad Q_z = \begin{bmatrix} Q'_{z1} \\ Q'_{z2} \end{bmatrix} = \begin{bmatrix} Q'_{z11} & Q'_{z12} \\ Q'_{z21} & Q'_{z22} \end{bmatrix},
$$

where A and B are $m_y \times r_y$ matrices, A_{2z} is $m_y \times r_z$ matrix, A_z and B_z are $m_z \times r_z$ matrices, Q'_{y21} is $r_y \times r_y$ matrix, and Q'_{z21} is $r_z \times r_z$ matrix. Then model (2.21) and (2.22) can be rewritten again as

$$
\Delta \boldsymbol{Y}_t = AB'\boldsymbol{X}_{t-1} + A_{2z}B'_z\boldsymbol{Z}_{t-1} + D(\Delta\boldsymbol{Z}_t - A_zB'_z\boldsymbol{Z}_{t-1}) + \sum_{j=1}^{p-1} \Phi_j^* \Delta X_{t-j} + \boldsymbol{e}_t \qquad (2.23)
$$

and

$$\Delta \boldsymbol{Z}_t = A_z B_z' \boldsymbol{Z}_{t-1} + \sum_{j=1}^{p-1} \Phi_j^{**} \Delta \boldsymbol{Z}_{t-j} + \boldsymbol{\epsilon}_{zt}. \tag{2.24}$$

Note that A and A_z are coefficient matrices of stationary process of $B'\boldsymbol{X}_{t-1}$ and $B_z'\boldsymbol{Z}_{t-1}$, respectively. A_{2z} is a coefficient matrix which represents the effect of long run variation of exogenous $\boldsymbol{Z}_t$ to conditional model. On the other hand, D is a coefficient matrix which represents the effect of purely short run variation of exogenous $\boldsymbol{Z}_t$ to conditional model. If $A_{2z} - DA_z = 0$, then the parameters in the conditional model and those in the marginal model are variation free as mentioned above.

Ahn et al. (2015) also reparameterized B and B_z using the same parameterization in Ahn and Reinsel (1990) which assumes that the last $d = m - r$ components do not cointegrate. To this end, they considered $\boldsymbol{X}_t^* = (\boldsymbol{Y}_{1,t}', \boldsymbol{Z}_{1,t}', \boldsymbol{Y}_{2,t}', \boldsymbol{Z}_{2,t}')$ which is a rearrangement of $\boldsymbol{X}_t$ such that $\boldsymbol{X}_{2,t} = (\boldsymbol{Y}_{2,t}', \boldsymbol{Z}_{2,t}')'$ is a $d-$dimensional purely nonstationary component of $\boldsymbol{X}_t$. The matrix of the cointegrating vector associated with $\boldsymbol{X}_t^*$ is of the form $B^* = (I_r, B_0^*)$ which can be partitioned so that it is conformable with $\boldsymbol{Y}_{2,t}$ and $\boldsymbol{Z}_{2,t}$. They rearrange $\boldsymbol{X}_t$ to form $\boldsymbol{X}_t^*$ and the column matrices corresponding to $\boldsymbol{Y}_{1,t}$ and $\boldsymbol{Y}_{2,t}$ and finally obtain the following matrix of cointegration vectors

$$\begin{bmatrix} I_{r_y} & B_{10}' & O_{r_y} & B_{20}' \\ O & O & I_{r_z} & B_{z0}', \end{bmatrix},$$

where B_{10} is $(m_y - r_y) \times r_y$, B_{20} is $(m_z - r_z) \times r_y$, and B_{z0} is $(m_z - r_z) \times r_z$ matrices of parameters.

Note that $B_z'\boldsymbol{Z}_{t-1} = \boldsymbol{Z}_{1,t-1} + B_{z0}'\boldsymbol{Z}_{2,t-1}$ is the cointegrating combinations of the exogenous variable only and $B'\boldsymbol{X}_{t-1} = \boldsymbol{Y}_{1,t-1} + B_{10}'\boldsymbol{Y}_{2,t-1} + B_{20}'\boldsymbol{Z}_{2,t-1}$ is cointegrating combinations of both endogenous variable $\boldsymbol{Y}_{t-1}$ and exogenous variable $\boldsymbol{Z}_{t-1}$. If some of the rows of B_{20}' are zero or linearly dependent, then the components corresponding to these rows are cointegrating combinations of $\boldsymbol{Y}_{t-1}$ only. For more details, see Ahn et al. (2015)

2.4 Generalized method of moment estimation

The generalized method of moments(GMM) was first introduced by Hansen (1982). Various types of GMM estimators have been proposed. Unlike the ML estimator, the GMM estimator does not need a distributional assumption and only requires the specification of moment conditions. Therefore, the GMM estimation is computationally convenient for the inference of a complicated model.

Mátyás (1999) and Hall et al. (2005) introduced a full details of estimators based on the GMM. In this section, we briefly review the GMM estimation in Mátyás (1999). The method of moment estimates the unknown parameters by matching the population moments with the corresponding sample moments. Suppose that we have an observed sample $\{x_t : t = 1, \cdots, T\}$ from which we want to estimate an unknown $p \times 1$ parameter $\boldsymbol{\theta}$ with the true value $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. In this case, the moment conditions are based on the set of q population orthogonality conditions below,

$$E\big[\boldsymbol{m}_t(x_t, \boldsymbol{\theta})\big] = \boldsymbol{0}$$

and the corresponding sample moment equation

$$\bar{\boldsymbol{m}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{m}_t(x_t, \boldsymbol{\theta}).$$

If $q = p$, i.e., the number of equations is the same as the number of unknown parameters, the system is exactly identified, and we can obtain $\hat{\boldsymbol{\theta}}_T$ by solving the system of the following moment equation

$$\bar{\boldsymbol{m}}_T(\hat{\boldsymbol{\theta}}_T) = \boldsymbol{0}.$$

If $q > p$, i.e., the number of equations are larger than the number of unknown parameters, the system is overidentified. Since there are more equations than unknown parameters, we can not find a vector $\hat{\boldsymbol{\theta}}_T$ that satisfies $\bar{\boldsymbol{m}}_T(\theta) = 0$. Instead, we will find the vector $\hat{\boldsymbol{\theta}}_T$ that makes $\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})$ as close to zero as possible using objective function and the GMM estimator is used in this situation.

Now consider the objective function of GMM estimator. For any $q \times q$ positive definite weighting matrix $W_T(\boldsymbol{\theta})$ which assigns a relative weight to each moment, define the objective function $Q_T(\boldsymbol{\theta})$ as

$$Q_T(\boldsymbol{\theta}) = \big[\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})\big]' W_T \big[\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})\big].$$

Then the GMM estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_T = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}}\, Q_T(\boldsymbol{\theta})$$

and we minimize the objective function based on the sum of square constraint violations. In some cases numerical optimization such as Gauss-Newton is needed.

Consider a linear regression model with $q > p$ valid instruments $\boldsymbol{z}_t$, $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \epsilon_t$, the moment conditions are $E(\boldsymbol{z}_t\epsilon_t) = E(\boldsymbol{z}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})) = \mathbf{0}$. The sample moments are

$$\bar{m}_T(\boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}) = T^{-1}(Z'\boldsymbol{y} - Z'X\boldsymbol{\beta}),$$

where Z, $\boldsymbol{y}$, and X are $T \times q$, $T \times 1$, and $T \times p$, respectively.
Since this is exactly the overidentified case. If we choose weighting matrix W_T,

$$W_T = \Big[\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t\boldsymbol{z}_t'\Big]^{-1} = T(Z'Z)^{-1}$$

and assume that $T^{-1}(Z'Z)$ converges in probability to a constant matrix W, i.e., the weak law of large numbers holds for $\boldsymbol{z}_t\boldsymbol{z}_t'$. The objective function is

$$Q_T(\boldsymbol{\beta}) = T^{-1}(Z'\boldsymbol{y} - Z'X\boldsymbol{\beta})'(Z'Z)^{-1}(Z'\boldsymbol{y} - Z'X\boldsymbol{\beta}).$$

By differentiating $Q_T(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and solving the first order conditions, we obtain $\hat{\boldsymbol{\beta}}_T$ as follows:

$$\hat{\boldsymbol{\beta}}_T = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'y.$$

This is the standard instruments variable(IV) estimator for the case where there are more instruments than the regressors.

For the asymptotic properties of the GMM estimator we need the following assumptions.

Assumption. 1. *(Mátyás)*

(1) $\boldsymbol{g}_t(\boldsymbol{\theta}) = E(m_t(\boldsymbol{\theta}))$ *exists and is finite forall* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and for all* t*.*

(2) There exists a $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ *such that* $vecg_t(\boldsymbol{\theta}) = 0$ *for all* t *if andonly if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*.*

(3) $\boldsymbol{\Theta}$ *is compact.*

(4) Each q *elements of* $\bar{\boldsymbol{m}}_T(\boldsymbol{\theta}) - \boldsymbol{g}_T(\boldsymbol{\theta})$ *converge in probability to 0 on* $\boldsymbol{\Theta}$ *where* $\boldsymbol{g}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\boldsymbol{g}_t(\boldsymbol{\theta})$*.*

(5) Each q *elements of* $\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})$ *are stochastically equicontinous and each* q *elements of* $\boldsymbol{g}_T(\boldsymbol{\theta})$ *are equicontinuous.*

(6) There exists a non-random sequence of positive definite matrices $\bar{W}_T$ *such that* $W_T - \bar{W}_T \xrightarrow{p} 0$*.*

Under the assumption.1 , $sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|Q_T(\boldsymbol{\theta}) - \bar{Q}_T(\boldsymbol{\theta})| \xrightarrow{p} 0$ where $\bar{Q}_T(\boldsymbol{\theta}) = \boldsymbol{g}_T(\boldsymbol{\theta})'\bar{W}_T\boldsymbol{g}_T(\boldsymbol{\theta})$. Since $\hat{\boldsymbol{\theta}}_T$ minimizes $Q_T(\boldsymbol{\theta})$ and $\boldsymbol{\theta}_0$ minimizes $\bar{Q}_T(\boldsymbol{\theta})$, $\hat{\boldsymbol{\theta}}_T$ is converge in probability to $\boldsymbol{\theta}_0$.

Assumption. 2. *(Mátyás)*

(1) $\boldsymbol{m}_t(\boldsymbol{\theta})$ *is continuously differentiable with respect to* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(2) For any sequence $\boldsymbol{\theta}_T^*$ *such that* $\boldsymbol{\theta}_T^* \xrightarrow{p} \boldsymbol{\theta}_0$, $F_T(\boldsymbol{\theta}_T^*) - \bar{F}_t \xrightarrow{p} 0$, *where* $F_T(\boldsymbol{\theta}_T^*) = \frac{\partial \bar{\boldsymbol{m}}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}_T^*}$ *and* $\bar{F}_T$ *is a sequence of* $(q \times p)$ *matrices that do not depend on* $\boldsymbol{\theta}$.

(3) $\boldsymbol{m}_t(\boldsymbol{\theta})$ *satisfies a central limit theorem, so that* $\bar{V}_T^{-1/2}\sqrt{T}\bar{\boldsymbol{m}}_T(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, I_q)$ *where* $\bar{V}_T$ *is sequence of* $(q \times q)$ *non-random positive definite matrices defined as* $\bar{V}_T = T var(\bar{\boldsymbol{m}}_T(\boldsymbol{\theta}_0))$.

Under the assumptions 1 and 2

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, V_{GMM}),$$

where $V_{GMM} = (\bar{F}_T'\bar{W}_T\bar{F}_T)^{-1}\bar{F}_T'\bar{W}_T\bar{V}_T\bar{F}_T\bar{F}_T'\bar{W}_T\bar{F}_T)^{-1}$.

Even though the GMM estimator $\hat{\boldsymbol{\theta}}_T$ has good asymptotic properties such as the consistency and the asymptotic normality, however, since arbitrary weighting matrix was used to obtain the initial estimator, those of GMM estimator do not ensure the efficiency. For any given symmetric positive definite matrix W_T asymptotic covariance matrix of the GMM estimator is satisfied with the following algebraic inequality:

$$(\bar{F}_T'\bar{W}_T\bar{F}_T)^{-1}\bar{F}_T'\bar{W}_T\bar{V}_T\bar{W}_T\bar{F}_T(\bar{F}_T'\bar{W}_T\bar{F}_T)^{-1} \geq (\bar{F}_T'\bar{V}_T^{-1}\bar{F}_T)^{-1}.$$

Therefore, choosing weighting matrix W_T such that $W_T \xrightarrow{p} \bar{V}_T^{-1}$. We can obtain the efficient GMM estimator which has the smallest asymptotic covariance matrix for the given orthogonal moment condition. By substituting the weighting matrix W_t by $\hat{V}_T^{-1}$, the objective function is obtained as follows:

$$Q_T^*(\boldsymbol{\theta}) = [\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})]'\hat{V}_T^{-1}[\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})].$$

Then the efficient GMM estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_T^* = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}}\, Q_T^*(\boldsymbol{\theta})$$

and the asymptotic distribution of the efficient GMM estimator is

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\bar{F}_T' \bar{V}_T^{-1} \bar{F}_T)^{-1}).$$

Hansen (1982) suggested two step procedure for estimating the weight matrix $\hat{V}_T$. For simplicity, consider the situation where the moment condition is serially uncorrelated. In the first step, using the weight matrix $W_T = I_q$, we obtain the consistent estimator $\hat{\boldsymbol{\theta}}_T^{(1)}$ of $\boldsymbol{\theta}$. Then, estimate $\hat{V}_T$ with $\hat{V}_T = \frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\boldsymbol{\theta}}_T^{(1)}) m_t(\hat{\boldsymbol{\theta}}_T^{(1)})'$. In the second step, we again obtain the GMM estimator $\hat{\boldsymbol{\theta}}_T^{(2)}$ using $W_T = \hat{V}_T^{-1}$. The iterative GMM estimation repeats this process until the GMM estimator converges. Hansen et al. (1996) proposed another estimator which is known as the continuous-updating estimator. The continuous-updating estimator uses the objective function $Q_T(\boldsymbol{\theta}) = [\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})]' \hat{V}_T(\theta)^{-1} [\bar{\boldsymbol{m}}_T(\boldsymbol{\theta})]$. In this case, the weight matrix V_T is also a function of the unknown parameter $\boldsymbol{\theta}$. The GMM estimators for three methods have the same asymptotic properties but the continuous-updating estimator performs better than the other methods in a finite sample.

Chapter 3

Asymmetric GARCH model via Yeo-Johnson transformation

3.1 Introduction

Financial time series often have reveal the conditional variance that evolves over time depending on the past observations. The autoregressive conditional heteroscedasticity(ARCH) model by Engle (1983) and the Generalized ARCH(GARCH) model by Bollerslev (1986) have been widely used to model financial time series having the volatility clustering property.

A number of researchers have found the leverage effect in financial time series. However, ARCH and GARCH can not describe it. Because of the square of lagged residuals, it means that positive and negative residuals have the same effect on the model, i.e., only the size, not the sign, of lagged residual effects conditional variance. For this reason, various asymmetric volatility models were recommended to overcome the weaknesses of the ARCH and

GARCH models.

Although one of the most important features of the linear ARCH and GARCH models is that it postulates a nonlinear relationship between the present and the past values of a time series, current evidence suggests that it is not nonlinear enough to model some financial time-series data. For example, Hsieh (1989) found that the GARCH model cannot fit some exchange rates satisfactorily; Scheinkman and Scheinkman and LeBaron (1989) found evidence that volatility in stock market data cannot be captured completely by the linear expression.

As these limitations, many authors mentioned alternative methods. Well-known asymmetric GARCH models which have been also found to be extremely useful in applications are nonlinear ARCH(NARCH), exponential GARCH(EGARCH), threshold ARCH(TARCH) and ARCH-in-the-mean(ARCH-M) (see Bera and Higgins (1993) for an excellent survey of GARCH models) and power ARCH(PARCH) models.

In this chapter we propose to describe the strong non-linearity and non-normal in some financial time-series data by suggesting a different approach in which instead of assuming arbitrary specifications for the conditional variance, Box-Cox family of power transformation which was reviewed in Chapter 2 and called Yeo and Johnson transformation, used in variance part at residuals. This transformation has traditionally been used to linearize otherwise non-linear models without restrictions on the variable. This has also been used for reducing heterogeneity and achieving symmetric distribution of the transformed variable. And it has many of the good properties of the Box and Cox (1964).

The remainder of this chapter is organized as follows. In section 2, suggested GARCH, named YJ-GARCH is described and devoted to inference: maximum likelihood and some of the estimation's main properties (consistency, asymptotic) are developed. we consider the forecasts of the YJ-GARCH model in section 3. Section 4 contains some real data analyze for comparisons with previous studies of GARCH-type models. Section 5 contains the concluding remarks. The proofs are given in the appendix.

3.2 Yeo and Johnson GARCH model

Let $\{X_t\}$ be a stationary sequence of time series and let $\mathcal{I}_t$ denote an increasing sequence of the sigma fields generated by $X_t, X_{t-1}, \ldots$. In this thesis, we assume that $\{X_t\}$ is generated by the GARCH(1,1)-type process as follows:

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} e_t, \quad \text{and} \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1},$$

where $\mu_t = E(X_t|\mathcal{I}_{t-1})$ is the conditional mean of X_t given $\mathcal{I}_{t-1}$ and $h_t = Var(\varepsilon_t|\mathcal{I}_{t-1})$ is the conditional variance of ε_t given $\mathcal{I}_{t-1}$. Here $\{e_t\}$ is a sequence of iid random variables with mean zero and variance unity.

In financial time series, the volatility tends to respond asymmetrically to the sign of the shocks. In other words, good and bad news have different effects on the volatility. However, the GARCH model does not describe the leverage effect since it uses the squared term of ε_{t-1} in h_t. In order to handle leverage effect, extensions of the GARCH model such as the exponential GARCH(EGARCH) model and the threshold GARCH(TGARCH) model have been proposed and reviewed in chapter 2.

We introduce an alternative GARCH model to handle the leverage effect and to make a smooth transition around zero. The conditional variance of the proposed model is written as

$$h_t = \alpha_0 + \alpha_1 \psi(\lambda, \varepsilon_{t-1})^2 + \beta_1 h_{t-1}, \tag{3.1}$$

where ψ is The Yeo-Johnson transformation in (2.12). Domain of Yeo-Johnson transformation is well defined on whole real line in λ and Yeo-

Johnson transformation is continuous in (λ, ε) and differentiable. Note that ψ reduces to the identity function for $\lambda = 1$ and so (3.1) includes the standard GARCH as a special case. We call this the Yeo-Johnson GARCH(YJ-GARCH) model.

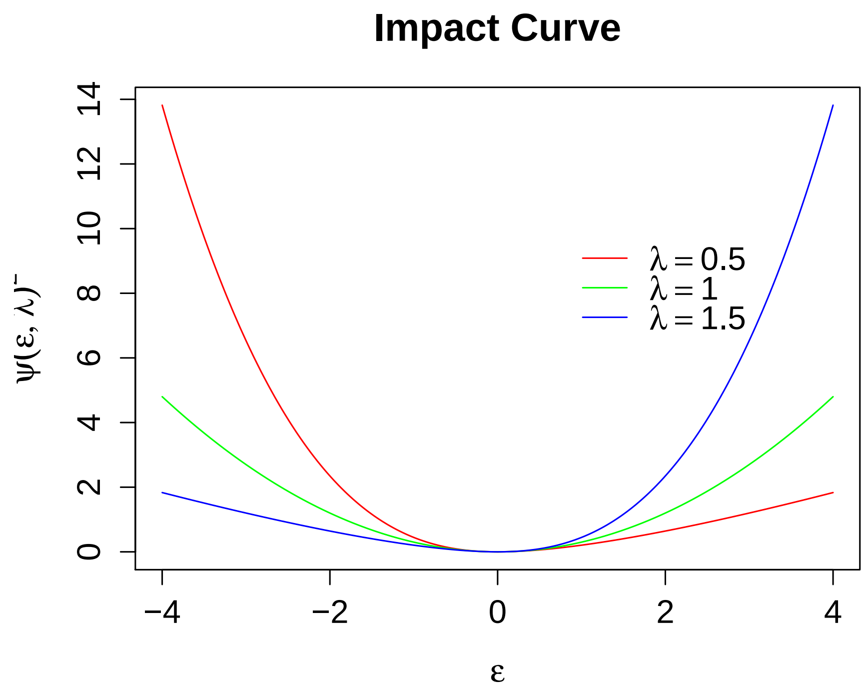

Figure 3.1: News Impact Curves for $\alpha_1 = 0.3$ and $\lambda = (0, 1, 1.5)$

Figure 3.1 shows news impact curves of the YJ-GARCH when $\alpha_1 = 0.3$ and λ =0, 1, and 1.5. The curve for $\lambda = 1$ is symmetric around zero and others are asymmetric. We see that, for $\lambda < 1$, negative shocks have larger impacts on the volatility or vice versa.

3.3 Estimation and Asymptotics

The main scope of this chapter is to investigate the impact of the alternative variance equation specifications on volatility. For high-frequency data, (Nelson (1991), Nelson (1992)) shows that the effect of misspecification of the conditional mean does not particularly affect the conditional volatility. Hence we simply assume that $\{X_t\}$ is generated by YJ-GARCH(p,q) with a constant mean such as

$$X_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} e_t, \quad \text{and} \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \psi(\lambda, \varepsilon_{t-i})^2 + \sum_{j=1}^{q} \beta_j h_{t-j}. \tag{3.2}$$

To make the conditional variance strictly positive for all realizations of ε_t, we complete the model with the positivity constraints assume as GARCH-type parameters, $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \leq 1$.

The standardized error term e_t is usually assumed to be normally distributed. In practice, we can also use a Student-t distribution to further enhance the robustness of the results. Let $f(\cdot)$ be a presumed density function of e_t and let $\boldsymbol{\theta}$ be the vector of parameters in the model (3.2). Then the likelihood function of $\boldsymbol{\theta}$ given the data $X_1, X_2, \ldots, X_n$ is written as

$$L_n(\boldsymbol{\theta}) = \prod_{t=1}^{n} p(X_t|\mathcal{F}_{t-1}) = \prod_{t=1}^{n} f((X_t - \mu)/\sqrt{h_t}) h_t^{-1/2}, \tag{3.3}$$

where $p(X_t|\mathcal{F}_{t-1})$ denotes a conditional density function of X_t given $\mathcal{F}_{t-1}$. The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is obtained by solving $S(\boldsymbol{\theta}) = \mathbf{0}$, where $S(\boldsymbol{\theta})$ is the score function given by

$$S(\boldsymbol{\theta}) = \frac{\partial \log L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial \log f((y_t - \mu)/\sqrt{h_t})}{\partial \boldsymbol{\theta}} - \frac{1}{2h_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}}$$

Let $\boldsymbol{\theta}_0$ be the vector of true parameters. If the true likelihood is different from the presumed likelihood in (3.3), $\boldsymbol{\theta}_0$ is the minimizer of the Kullback-Leibler information between the true density and the presumed density in the parameter space $\boldsymbol{\Theta}$. Let

$$\nabla l_t(\boldsymbol{\theta}_0) = \left(\frac{\partial}{\partial \theta_i} l_t(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)$$

be the gradient of the loglikelihood function of one observation for $\boldsymbol{\theta}$ and let

$$\nabla^2 l_t(\boldsymbol{\theta}_0) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_t(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)$$

be the Hessian of the loglikelihood function.

Theorem 3.1. *Suppose the parameter space $\boldsymbol{\Theta}$, the true density function $g(\cdot)$ and the log-likelihood function $l_t(\boldsymbol{\theta})$ satisfy the following conditions*

(c.1) the parameter space $\boldsymbol{\Theta}$ is a compact set

(c.2) $E_g[l_t(\boldsymbol{\theta}; X_t)]$ has a unique global maximum at $\boldsymbol{\theta}_0$,

(c.3) $\sup_i E[I_{(X_i \geq 0)} X_i^{4b}] < \infty$ and $\sup_i E[I_{(X_i < 0)} (-X_i)^{4(2-a)}] < \infty$,

Then,

(A) $\lim_{n \to \infty} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{n} \sum_{t=1}^{n} l_1(\boldsymbol{\theta}; X_t) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} E_g[l_1(\boldsymbol{\theta} | X_t)]$ with probability one.

(B) The MLE $\hat{\boldsymbol{\theta}}$ is a strongly consistent estiamtor of $\boldsymbol{\theta}_0$.

Furthermore, if

(c.5) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$,

(c.6) $\sup_i E[I_{X_i \geq 0} X_i^{6b} log^4(X_i + 1)] < \infty$ and
$\sup_i E[I_{X_i < 0} (-X_i)^{6(2-a)} log^4(-X_i + 1)] < \infty$,

(c.7) $E_g[\nabla l_1(\boldsymbol{\theta}_0; X_t)] = \mathbf{0}$,

(c.8) $E_g[\nabla^2 l_1(\boldsymbol{\theta}_0; X_t)]$ is non-singular,

(c.9) $A_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \to A(\boldsymbol{\theta})$ *and* $B_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \to B(\boldsymbol{\theta})$ *and both* $A(\boldsymbol{\theta}), B(\boldsymbol{\theta})$ *are positive definite matrix*

Then,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)),$$

where the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ *is given by*

$$\begin{aligned}
\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) &= \boldsymbol{V}(\boldsymbol{\theta}_0)\boldsymbol{W}(\boldsymbol{\theta}_0)\boldsymbol{V}(\boldsymbol{\theta}_0)^T \\
\boldsymbol{V}(\boldsymbol{\theta}_0) &= E_f[\nabla^2 l_t(\boldsymbol{\theta}_0)]^{-1} \\
\boldsymbol{W}(\boldsymbol{\theta}_0) &= E_f[\nabla l_t(\boldsymbol{\theta}_0)\{\nabla l_t(\boldsymbol{\theta}_0)\}^T].
\end{aligned}$$

The details of the proof under the normal assumption are given in Appendix.

3.4 Forecast Interval

One of the main targets in time series analysis is to forecast future observations. Let $X_t(k)$ and $h_t(k)$ denote the k-step ahead forecast of the observation and the conditional variance at time t. We assume that $\{X_t\}$ follows the YJ-GARCH(1,1) model with constant mean, that is,

$$X_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t}e_t, \quad \text{and} \quad h_t = \alpha_0 + \alpha_1\psi(\lambda, \varepsilon_{t-1})^2 + \beta_1 h_{t-1},$$

The mean square error(MSE) is a common criterion to select an optimum forecast and the conditional expectation $E[X_{n+k}|\mathcal{F}_n] = \mu + E(\sqrt{h_{n+k}}e_{n+k}|\mathcal{F}_n) = \mu$ is the minimum MSE forecast for the k-step ahead value at time n.

For 1-step ahead forecast for the conditional variance, we have

$$h_{t+1} = \alpha_0 + \alpha_1\psi(\varepsilon_t, \lambda)^2 + \beta_1 h_t,$$

where ε_t and h_t are known at the time index $1 \le t \le n$. So for $1 \le t \le n$, the 1-step ahead forecast is

$$h_t(1) = \alpha_0 + \alpha_1 \psi(\varepsilon_t, \lambda)^2 + \beta_1 h_t.$$

In order to derive general k-step ahead forecast, we consider the Taylor expansion of $\psi(\varepsilon_t, \lambda)^2$ at arbitrary e_0 as follows:

$$\begin{aligned}
\psi(\varepsilon_t, \lambda)^2 &= \psi(\sqrt{h_t} e_t, \lambda)^2 \\
&\approx \psi(\sqrt{h_t} e_0, \lambda)^2 + (e_t - e_0) 2\psi(\sqrt{h_t} e_0, \lambda)\psi'(\sqrt{h_t} e_0, \lambda) \\
&\quad + (e_t - e_0)^2 \left[\psi(\sqrt{h_t} e_0, \lambda)\psi''(\sqrt{h_t} e_0, \lambda) + \left\{ \psi'(\sqrt{h_t} e_0, \lambda) \right\}^2 \right] \\
&\equiv \hat{\psi}(\varepsilon_t, \lambda)^2,
\end{aligned}$$

where

$$\psi'(\sqrt{h_t} e_0, \lambda) = \begin{cases} \sqrt{h_t}(\sqrt{h_t} e_0 + 1)^{\lambda - 1}, & \text{if } e_0 \ge 0 \\ \sqrt{h_t}(-\sqrt{h_t} e_0 + 1)^{1 - \lambda}, & \text{if } e_0 < 0. \end{cases}$$

Then, the 2-step ahead forecast at the time n can be written as

$$h_{t+2} = \alpha_0 + \alpha_1 \psi(\varepsilon_{t+1}, \lambda)^2 + \beta_1 h_{t+1} \approx \alpha_0 + \alpha_1 \hat{\psi}(\varepsilon_{t+1}, \lambda)^2 + \beta_1 h_{t+1}$$

Here, setting $e_0 = E(e_t) = 0$, we obtain

$$E\left[\psi(\varepsilon_t, \lambda)^2\right] \approx E\left[\hat{\psi}(\varepsilon_t, \lambda)^2\right] = E(e_t^2) h_t = h_t$$

because $\psi(0, \lambda) = 0$ and

$$h_{t+2} \approx \alpha_0 + (\alpha_1 + \beta_1) h_{t+1}$$

In general, we have

$$h_n(k) \approx \alpha_0 + (\alpha_1 + \beta_1) h_n(k-1), \quad k \ge 2.$$

The forecast error in the constant mean model is $\varepsilon_n(k) = X_{n+k} - X_n(k) = \varepsilon_{n+k}$ and

$$\frac{X_{n+k} - X_n(k)}{\sqrt{h_{n+k}}} = e_{n+k} \sim f(\cdot).$$

The $100(1-\alpha)\%$ forecast interval for X_{n+k} is estimated as follows:

$$[\hat{L}(n,k,\alpha/2),\ \hat{U}(n,k,\alpha/2)] = \left[\hat{\mu} + z_{\alpha/2}\sqrt{h_n(k)},\ \hat{\mu} + z_{1-\alpha/2}\sqrt{h_n(k)}\right] \quad (3.4)$$

where z_α denotes the αth quantile of $f(\cdot)$.

3.5 Real Data Analysis

In this section, we compare the performance of the YJ-GARCH model with other GARCH models such as the GARCH, the EGARCH, and the TGARCH model by forecasting future returns of assets. Let P_t be the price of an asset at time t and let X_t be the log return, that is $X_t = 100\log(P_t/P_{t-1})$. Data sets consist of 4054 daily returns of IBM stock price from January 2000 to Febuary 2016. We compute the mean square prediction error(MEPE) and coverage probabilities of each model to compare performances. We divide the whole period into subperiods of size n by sliding window as follows:

$$X_{j+1}, X_{j+2}, \ldots, X_{j+n}, \quad j = 0, \ldots, m-1.$$

From a preliminary analysis, we conclude that the constant mean model with order $p = 1$ and $q = 1$ is a proper choice as the analyzing model in most periods. For each $j = 0, \ldots, m-1$, we obtain the k-step ahead forecast $\hat{X}_{j+n}(k)$ and prediction intervals $[\hat{L}(j+n,k,\alpha/2), \hat{U}(j+n,k,\alpha/2)]$. In risk management, the one-sided prediction interval is more important and so we also compute $[\hat{L}(j+n,k,\alpha), \infty)$ and $(-\infty, \hat{U}(j+n,k,\alpha)]$. Here we use $k = (1, 5, 10)$, $\alpha = (0.01, 0.05)$, and $n = 500$. Student t-distribution is employed as distribution of $f(\cdot)$ and so z in (3.4) depends on the estimated degree of freedom and is varying for each j.

The MSPE for k-step ahead forecast is defined as

$$\text{MSPE}(k) = \frac{1}{m}\sum_{j=0}^{m-1}\left(X_{j+n+k} - \hat{X}_{j+n}(k)\right)^2,$$

Table 3.1: MSPE of IBM stock price

Model	MSPE(1)	MSPE(5)	MSPE(10)
GARCH	2.209	2.211	2.215
EGARCH	2.207	2.210	2.215
TGARCH	2.208	2.211	2.215
YJ-GARCH	2.208	2.211	2.215

Table 3.2: 95% Coverage Probability for forecast intervals of IBM stock price

CP	GARCH	EGARCH	TGARCH	YJ-GARCH
CP(1,1)	0.944	0.941	0.944	0.945
CP(1,2)	0.935	0.936	0.937	0.946
CP(1,3)	0.957	0.953	0.957	0.951
CP(5,1)	0.947	0.930	0.945	0.945
CP(5,2)	0.937	0.932	0.938	0.945
CP(5,3)	0.958	0.943	0.953	0.948
CP(10,1)	0.945	0.920	0.943	0.941
CP(10,2)	0.936	0.930	0.942	0.942
CP(10,3)	0.955	0.935	0.951	0.943

and the coverage probability for k-step ahead forecast interval is

$$\text{CP}(k,l) = \frac{1}{m}\sum_{j=0}^{m-1} I(X_{j+n+k} \in V_l(j+n,k)),$$

where

$$\begin{aligned} V_1(n,k) &= [\hat{L}(n,k,\alpha/2), \hat{U}(n,k,\alpha/2)] \\ V_2(n,k) &= [\hat{L}(n,k,\alpha), \infty) \\ V_3(n,k) &= (-\infty, \hat{U}(n,k,\alpha)]. \end{aligned}$$

Table 3.3: 99% Coverage Probability for forecast intervals of IBM stock price

cp	GARCH	EGARCH	TGARCH	YJ-GARCH
cp(1,1)	0.988	0.986	0.988	0.989
cp(1,2)	0.988	0.987	0.987	0.988
cp(1,3)	0.991	0.991	0.990	0.990
cp(5,1)	0.986	0.983	0.986	0.985
cp(5,2)	0.987	0.985	0.986	0.985
cp(5,3)	0.989	0.989	0.989	0.989
cp(10,1)	0.986	0.983	0.986	0.986
cp(10,2)	0.986	0.984	0.985	0.986
cp(10,3)	0.990	0.988	0.989	0.989

Table 3.1 shows that MSPE of each model. As Nelson(Nelson (1991), Nelson (1992)) considered, conditional mean shows very similar results for all models. Tables 3.2-3.3 summarize the coverage probability of each model for each forecast interval at specific levels which is 95% and 99%, respectively. TGARCH and EGARCH fail to converge 5 times for all periods, while YJ-GARCH and GARCH are stable. We discard these cases from the summary tables. For Table 3.2, we can see that YJ-GARCH stably keeps the level in coverage probability of one-sided forecast intervals, while others does not. Although performances of multi-step-ahead forecast is similiar to other GARCH-type models, YJ-GARCH with 1-step-ahead forecast confirms the superior performance of the YJ-GARCH model over other GARCH-type models. It is worth to mention that YJ-GARCH model can handle the non-constant structure of asymmetry when the relative impacts of negative and positive shocks on the current volatility are not the same.

Chapter 4

Generalized method of moments estimation of cointegration model with exogenous variables

4.1 Introduction

Since the GMM was first introduced by Hansen (1982), class of GMM estimators have been broadly applied to financial and economic data analysis. Unlike the MLE, GMM does not need a distributional assumption, but only requires specification of moment conditions. Therefore, the GMM estimation is a computationally convenient method, especially for the inference of a complicated model.

Kitamura and Phillips (1997) introduced GMM estimation of a nonstationary regression model and Quintos (1998) and Kleibergen (1997) extended it to cointegration model. Park et al. (2011) suggested two iterative GMM estimation with identification condition for the parameters in cointegration model.

Since Granger (1981) introduced the idea of cointegration, several estimation methods for cointegration models have been proposed. These include ordinary least squares(OLS) by Engle and Granger (1987), nonlinear least squares(NLS) by Stock (1987), MLE by Johansen (1988) and reduced rank approach by Ahn and Reinsel (1990). Johansen (1992), Harbo et al. (1998), and Pesaran et al. (2000) considered inference of the processes which was cointegrated vector autoregressive process of integrated order 1, where the process consists of endogenous and exogenous variables, assuming that the nonstationary exogenous variables are not cointegrated. Ahn et al. (2015) considered the more general condition that nonstationary exogenous variables are cointegrated themselves.

However, due to strong parametric assumptions of MLE, it should be checked before real data analysis. Since collected data is huge and high-dimensional, Those of MLE are easily violated. Therefore, many researchers are recently been searching for other estimation methods which are free from assumptions. As mentioned above, an alternative method satisfying this requisite is GMM.

In this thesis, we consider GMM estimate on the basis of the process by Ahn et al. (2015). Following part is organized as follows, section 2 contains the parameterization of the parameter sets. In section 3, parameters are estimated by the iterative GMM based on Park et al. (2011) method. Asymptotic properties of the iterative GMM estimators are presented in section 3. Section 4 examines the finite sample properties of the estimators through a Monte Carlo simulation and a relatively simple numerical example to illustrate the methods in section 2. Conclusions are in Section 5.

4.2 VECM with exogenous variables

We consider $m-$dimensional vector autoregressvie(VAR) process of order p with cointegrating(CI) rank r. The vector error correction model(VECM) form is expressed as

$$\Delta \boldsymbol{X}_t = C\boldsymbol{X}_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* \Delta \boldsymbol{X}_{t-j} + \boldsymbol{a}_t. \tag{4.1}$$

Let the process be decomposed as $\boldsymbol{X}_t = (\boldsymbol{Y}_t', \boldsymbol{Z}_t')'$, where $\boldsymbol{Y}_t$ is an m_y-dimensional vector process of endogenous variables with and $\boldsymbol{Z}_t$ is an m_z-dimensional vector process of exogenous variables with $m_y + m_z = m$. The error term $\boldsymbol{a}_t$ is also decomposed as $\boldsymbol{a}_t = (\boldsymbol{\epsilon}_{yt}', \boldsymbol{\epsilon}_{zt}')$ where $\boldsymbol{\epsilon}_{yt}$, $\boldsymbol{\epsilon}_{zt}$ has the same dimension as $\boldsymbol{Y}_t$ and $\boldsymbol{Z}_t$ respectively. Then, the covariance matrix $\Omega = Cov(\boldsymbol{a}_t)$ are denoted by

$$\Omega = \begin{bmatrix} Cov(\boldsymbol{\epsilon}_{yt}) & Cov(\boldsymbol{\epsilon}_{yt}, \boldsymbol{\epsilon}_{yt}) \\ Cov(\boldsymbol{\epsilon}_{zt}, \boldsymbol{\epsilon}_{yt}) & Cov(\boldsymbol{\epsilon}_{zt}) \end{bmatrix} = \begin{bmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{bmatrix}.$$

When $\boldsymbol{Z}_t$ is subcointegrated with cointegration rank $r_z < r$, the VECM in (4.1) can be represented as the following conditional model of $\Delta\boldsymbol{Y}_t$ and marginal model of $\Delta\boldsymbol{Z}_t$ as in Ahn et al. (2015)

For simplicity of exposition, we consider the case with $p = 1$ as follows:

$$\Delta\boldsymbol{Y}_t = AB'\boldsymbol{X}_{t-1} + A_{2z}B_z'\boldsymbol{Z}_{t-1} + D(\Delta\boldsymbol{Z}_t - A_zB_z'\boldsymbol{Z}_{t-1}) + \boldsymbol{e}_{yt} \tag{4.2}$$

$$\Delta\boldsymbol{Z}_t = A_zB_z'\boldsymbol{Z}_{t-1} + \boldsymbol{\epsilon}_{zt}, \tag{4.3}$$

where $D = \Omega_{yz}\Omega_{zz}^{-1}$, $\boldsymbol{e}_{yt} = \boldsymbol{\epsilon}_{yt} - D\boldsymbol{\epsilon}_{zt}$. Note that $B' = [I_{r_y}, B_{10}', O_{r_y \times r_y}, B_{20}']$ and $B_z' = [I_{r_z}, B_{z0}']$ is expressed as normalization by Ahn and Reinsel (1990) for the identifying condition where $r_y = r - r_z$. This condition is motivated by the regression analysis since, $B'\boldsymbol{X}_{t-1} = \boldsymbol{Y}_{1,t-1} + B_{10}'\boldsymbol{Y}_{2,t-1} + B_{20}'\boldsymbol{Z}_{2,t-1}$ with partitioning $\boldsymbol{X}_t = [\boldsymbol{Y}_{1,t}', \boldsymbol{Y}_{2,t}', \boldsymbol{Z}_{1,t}', \boldsymbol{Z}_{2,t}']$ where $\boldsymbol{Y}_{1,t}$ is $r_y \times 1$ and $\boldsymbol{Y}_{2,t}$

is $(m_y - r_y) \times 1$ and $\boldsymbol{Z}_{1,t}$ is $r_z \times 1$ and $\boldsymbol{Y}_{2,t}$ is $(m_z - r_z) \times 1$.

The models in (4.2) and (4.3) can be rewritten in compact form as follows:

$$\Delta \boldsymbol{Y}_t = \Pi_1 \boldsymbol{\xi}_{t-1} + \boldsymbol{e}_{yt} \tag{4.4}$$

$$\Delta \boldsymbol{Z}_t = \Pi_2 \boldsymbol{u}_{t-1} + \boldsymbol{\epsilon}_{zt}, \tag{4.5}$$

where

$$\boldsymbol{\xi}_{t-1} = [\boldsymbol{X}'_{t-1}, (B'_z \boldsymbol{Z}_{t-1})', (\Delta \boldsymbol{Z}_t - A_z B'_z \boldsymbol{Z}_{t-1})']', \, \boldsymbol{u}_{t-1} = [\boldsymbol{Z}'_{t-1}]',$$

$$\Pi_1 = [\Pi'_{11}, \Pi'_{12}]', \Pi_{11} = vec(B'_0), \Pi_{12} = vec(A, A_{2z}, D)$$
$$\Pi_2 = [\Pi'_{21}, \Pi'_{22}]' \Pi_{21} = vec(B'_{z0}), \Pi_{22} = vec(A_z)$$

$B'_0 = [B'_{10}, B'_{20}]'$ and A is $m_y \times r_y$, A_{2z} is $m_y \times r_z$, A_z is $m_z \times r_z$, B_{10} is $(m_y - r_y) \times r_y$, B_{20} is $(m_z - r_z) \times r_y$, B_{z0} is $(m_z - r_z) \times r_z$.

4.3 GMM estimation of VECM

In order to estimate the parameters of A_z and B_Z both in (4.4) and (4.5), we consider the iterative GMM. In the first stage, we estimate the parameters in (4.5). In the second stage, we estimate the parameters in (4.4) with estimated parameters in (4.5). One set constitutes the parameters of the marginal model (4.5), Π'_{21}, Π'_{22}. The other set constitutes the parameters of conditional model (4.4), Π'_{21}, Π'_{22}.

We can obtain the simple orthogonal condition that the regressor of the processes $\boldsymbol{\xi}_{t-1}$, $\boldsymbol{u}_{t-1}$ are orthogonal to the error terms $\boldsymbol{e}_{yt}$, $\boldsymbol{\epsilon}_{zt}$ respectively. Thus, in this setting, moment conditions become

$$E[m(\Pi_1)] = E[vec(\boldsymbol{e}_{yt} \boldsymbol{\xi}'_{t-1})]$$
$$E[m(\Pi_2)] = E[vec(\boldsymbol{\epsilon}_{zt} \boldsymbol{u}'_{t-1})].$$

Since we consider the cointegrated model of the reduced rank structure following Ahn and Reinsel (1990), the number of unknown parameters is smaller than the number of equations. In this case, no parameters satisfy the moment condition. But, we can find the parameters which make the moment conditions close to zero using the quadratic form of the objective function below:

$$Q(\Pi_1) = Tvec\Big(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{e}_{yt}\boldsymbol{\xi}'_{t-1}\Big)'\hat{V}_T^{-1}vec\Big(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{e}_{yt}\boldsymbol{\xi}'_{t-1}\Big),$$
$$Q(\Pi_2) = Tvec\Big(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{zt}\boldsymbol{u}'_{t-1}\Big)'\hat{V}_T^{*-1}vec\Big(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{zt}\boldsymbol{u}'_{t-1}\Big).$$

To secure the efficiency of the GMM estimator we use the following weighting matrix $\hat{V}_T = \big(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}'_{t-1}\big)\otimes\hat{\Omega}_e$ and $\hat{V}_T^* = \big(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_{t-1}\boldsymbol{u}'_{t-1}\big)\otimes \hat{\Omega}_\epsilon$ with a consistent estimators $\hat{\Omega}_e$ and $\hat{\Omega}_\epsilon$ for Ω_e and Ω_ϵ respectively. Then we obtain GMM estimator of parameters from the minimization of objective functions.

The iterative GMM estimation method also divides unknown parameters into two parameter sets as cointegrated parameters and adjustment parameters in each stage. In total, it consists of four steps. At each step, we consider GMM estimator of the parameters and their asymptotic distributions.

In the first step, we assume that the initial estimators of $\Pi_{22}(vec(A_z))$ and Ω_ϵ are given by $\tilde{\Pi}_{22}$ and $\tilde{\Omega}_\epsilon$. Then the model in (4.5) can be represented as

$$\Delta\dot{\boldsymbol{Z}}_t = \tilde{A}_z B_z' \boldsymbol{Z}_{2,t-1} + \boldsymbol{\epsilon}_{zt} \quad \text{for} \quad \Delta\dot{\boldsymbol{Z}}_t = \Delta\boldsymbol{Z}_t - \tilde{A}_z\boldsymbol{Z}_{1,t-1}. \tag{4.6}$$

The objective function is

$$Q_T^*(\Pi_{21}) = T\Big(\frac{1}{T}\sum_{t=1}^{T}vec(\Delta\dot{\boldsymbol{Z}}_t\boldsymbol{u}'_{t-1}) - \frac{1}{T}K_1\Pi_{21}\Big)'\hat{V}_T^{-1}$$
$$\times\Big(\frac{1}{T}\sum_{t=1}^{T}vec(\Delta\dot{\boldsymbol{Z}}_t\boldsymbol{u}'_{t-1}) - \frac{1}{T}K_1\Pi_{21}\Big), \tag{4.7}$$

where $\hat{V}_T = (\boldsymbol{u}\boldsymbol{u}'/T \otimes \tilde{\Omega}_\epsilon)$, $K_1 = \sum_{t=1}^{T}(\boldsymbol{u}_{t-1}\boldsymbol{Z}'_{2,t-1} \otimes \tilde{A}_z)$, and $\boldsymbol{u} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{T-1}]$. Then the efficient GMM estimator for Π_{21} is given by

$$\tilde{\Pi}_{21} = \big((J_2'\boldsymbol{Z}\boldsymbol{Z}'J_2 \otimes \tilde{A}_z'\tilde{\Omega}_\epsilon^{-1}\tilde{A}_z)\big)^{-1} vec\big(A_z'\tilde{\Omega}_\epsilon^{-1}\Delta\dot{\boldsymbol{Z}}\boldsymbol{Z}'J_2\big), \tag{4.8}$$

where $J_2' = [O_{(m_z-r_z)\times r_z}, I_{(m_z-r_z)}]$, $\Delta\dot{\boldsymbol{Z}}$, and $\boldsymbol{Z}$ are a matrix form representation as $\boldsymbol{u}$.

In the second step, we use the initial estimator of Π_{21} and the updataed $\tilde{\Omega}_\epsilon$ are obtained in the first step. Then the efficient GMM estimator for Π_{22} is obtained by the same process of the first step as follows:

$$\tilde{\Pi}_{22} = \big((\tilde{\boldsymbol{\omega}}\tilde{\boldsymbol{\omega}}' \otimes \tilde{\Omega}_\epsilon^{-1})\big)^{-1} vec\big(\tilde{\Omega}_\epsilon^{-1}\Delta\boldsymbol{Z}\tilde{\boldsymbol{\omega}}'\big), \tag{4.9}$$

where $\boldsymbol{\omega} = [\boldsymbol{\omega}_0, \cdots, \boldsymbol{\omega}_{T-1}]$ with $\boldsymbol{w}_{t-1} = [(B_z'\boldsymbol{Z}_{t-1})']'$.

In the third step, we use the estimator of Π_2 obtained in the first and second step and assume that the initial estimators of $\Pi_{12}(vec(A, A_{2z}, D))$ and Ω_e are given by $\tilde{\Pi}_{12}$ and $\tilde{\Omega}_e$. Then VECM (4.4) can be represented as follows:

$$\begin{aligned} \Delta\dot{\boldsymbol{Y}}_t &= \tilde{A}B'\boldsymbol{X}^*_{2,t-1} + \boldsymbol{e}_{yt} \quad \text{for} \\ \Delta\dot{\boldsymbol{Y}}_t &= \Delta\boldsymbol{Y}_t - \tilde{A}\boldsymbol{X}^*_{1,t-1} - \tilde{A}_{2z}\tilde{B}_z'\boldsymbol{Z}_{t-1} - D(\Delta\boldsymbol{Z}_t - \tilde{A}_z\tilde{B}_z'\boldsymbol{Z}_{t-1}). \end{aligned} \tag{4.10}$$

where $\boldsymbol{X}^*_{2,t-1} = [\boldsymbol{Y}_{2,t-1}, \boldsymbol{Z}_{2,t-1}]$. Similarly, in the first and the second step, the efficient GMM estimator for Π_{11} and Π_{12} is

$$\tilde{\Pi}_{11} = \big((J_1'\boldsymbol{X}^*\boldsymbol{X}^{*'}J_1 \otimes \tilde{A}'\tilde{\Omega}_e^{-1}\tilde{A})\big)^{-1} vec\big(\tilde{A}'\tilde{\Omega}_e^{-1}\Delta\dot{\boldsymbol{Y}}\boldsymbol{X}^{*'}J_1\big), \tag{4.11}$$

where $J_1' = [O_{(m-r)\times r}, I_{(m-r)}]$, $\Delta\dot{\boldsymbol{Y}}$ and $\boldsymbol{X}^*$ are matrix represenation and

$$\tilde{\Pi}_{12} = \big((\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}' \otimes \tilde{\Omega}_e^{-1})\big)^{-1} vec\big(\tilde{\Omega}_e^{-1}\Delta\boldsymbol{Y}\tilde{\boldsymbol{\eta}}'\big), \tag{4.12}$$

where $\tilde{\boldsymbol{\eta}} = [\tilde{\boldsymbol{\eta}}_0, \cdots, \tilde{\boldsymbol{\eta}}_{T-1}]$ with $\tilde{\boldsymbol{\eta}}_{t-1} = [(\tilde{B}'\boldsymbol{X}_{t-1})', (\tilde{B}_z'\boldsymbol{Z}_{t-1})', (\Delta\boldsymbol{Z}_t - \tilde{A}_z\tilde{B}_z'\boldsymbol{Z}_{t-1})']'$.

These four steps are repeated until the objective functions or estimates converge. In each step, estimators of iterative GMM are closed form solution. For this reason, iterative GMM has simple and easy computation and more useful when m and p are large.

We also obtain the asymptotic results in the following theorem using by application of lemma 1 of Ahn and Reinsel (1990) and functional of stochastic integrals of vector Brownian motions.

Theorem 4.1. *Let* $\tilde{\Pi}$ *denote the iterative GMM estimator for* $\Pi, \Omega_e, \Omega_\epsilon$ *using the equation as above. When an initial consistenct estimators of* Π_{12} *and* Π_{22} *is given, the asymptotic distribution of the* $\tilde{\Pi}$ *is given by*

(1)

$$T\big(\tilde{\Pi}_{21} - \Pi_{21}\big) \xrightarrow{d} (\mathcal{R}_{zz} \otimes A_z'\Omega_\epsilon^{-1}A_z)^{-1}vec(A_z'\Omega_\epsilon^{-1}\mathcal{B}_{z\epsilon}')$$

(2)

$$T^{1/2}\big(\tilde{\Pi}_{22} - \Pi_{22}\big) \xrightarrow{d} \big(\Omega_\omega \otimes \Omega_\epsilon^{-1}\big)^{-1}N(0, \Omega_\omega \otimes \Omega_\epsilon^{-1})$$

(3)

$$T\big(\tilde{\Pi}_{11} - \Pi_{11}\big) \xrightarrow{d} (\mathcal{R}_{xx} \otimes A'\Omega_e^{-1}A)^{-1}vec(A'\Omega_e^{-1}\mathcal{B}_{ye}')$$

(4)

$$T^{1/2}\big(\tilde{\Pi}_{12} - \Pi_{12}\big) \xrightarrow{d} \big(\Omega_\eta \otimes \Omega_e^{-1}\big)^{-1}N(0, \Omega_\eta \otimes \Omega_e^{-1})$$

where

(a) $\frac{1}{T^2}\sum_{t=1}^{T}\boldsymbol{X}_{2,t-1}\boldsymbol{X}_{2,t-1}' \xrightarrow{d} \Psi_{22}\Omega_{a_2}^{1/2}\int_0^1 B_d(\boldsymbol{u})B_d(\boldsymbol{u})'d\boldsymbol{u}\Omega_{a_2}^{1/2}\Psi_{22}' =: \mathcal{R}_{xx}$

(b) $\frac{1}{T^2}\sum_{t=1}^{T}\boldsymbol{Z}_{2,t-1}\boldsymbol{Z}_{2,t-1}' \xrightarrow{d} \Psi_{22}^{z}\Omega_{\epsilon_2}^{1/2}\int_0^1 B_{d_z}(\boldsymbol{u})B_{d_z}(\boldsymbol{u})'d\boldsymbol{u}\Omega_{\epsilon_2}^{1/2}\Psi_{22}^{z'} =: \mathcal{R}_{zz}$

(c) $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{2,t-1}\boldsymbol{e}_{y,t}' \xrightarrow{d} \Psi_{22}\Omega_{a_2}^{1/2}\int_0^1 B_d(\boldsymbol{u})dB_{m_y}'(\boldsymbol{u})\Omega_e^{1/2} =: \mathcal{B}_{ye}$

(d) $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_{2,t-1}\boldsymbol{e}'_{y,t} \xrightarrow{d} \Psi^z_{22}\Omega^{1/2}_{\epsilon_2}\int_0^1 B_{d_z}(\boldsymbol{u})dB'_{m_y}(\boldsymbol{u})\Omega^{1/2}_e =: \mathcal{B}_{ze}$

(e) $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{w}_t\boldsymbol{w}'_t \xrightarrow{p} \Omega_w = cov(\boldsymbol{w}_t)$

(f) $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\eta}_t\boldsymbol{\eta}'_t \xrightarrow{p} \Omega_\eta = cov(\boldsymbol{\eta}_t)$

(g) $\frac{1}{T^{1/2}}\sum_{t=1}^{T} vec(\boldsymbol{\epsilon}_{t-1}\boldsymbol{w}'_t) \xrightarrow{d} N(0, \Omega_w \otimes \Omega_\epsilon)$

(h) $\frac{1}{T^{1/2}}\sum_{t=1}^{T} vec(\boldsymbol{e}_{t-1}\boldsymbol{\eta}'_t) \xrightarrow{d} N(0, \Omega_\eta \otimes \Omega_e)$,

where $\boldsymbol{a}'_{2,t} = (\boldsymbol{e}'_{2,yt}, \boldsymbol{\epsilon}'_{2,zt}) = (\boldsymbol{e}'_2, \boldsymbol{\epsilon}'_2)$, *and* $\Psi_{22} = [0, I_{m-r}]\Psi[0, I_{m-r}]'$. $\Psi = \sum_{k=1}^{\infty}\Psi_k$ *is sum of the infinite moving average coefficients of model (4.2),* $\Omega_{e_2} = [O, I_{m_y-r+r_z}]\Omega_e[O, I_{m_y-r+r_z}]'$, *and* $\Psi^z_{22} = [0, I_{m_z-r_z}]\Psi^z[0, I_{m_z-r_z}]'$ *with* $\Psi = \sum_{k=1}^{\infty}\Psi^z_k$ *is sum of the infinite moving average coefficients of model (4.3),* $\Omega_{\boldsymbol{\epsilon}_2} = [O, I_{m_z-r_z}]\Omega_\epsilon[O, I_{m_z-r_z}]'$. *Let* $B_{m_y}(u)$ $B_{m_z}(u)$*and* $B_d(u) = \Omega^{1/2}_{a1}[I_d, 0]\Omega^{1/2}_{a1}B_m(u)$, $B_{d_z}(u) = \Omega^{1/2}_{a1}[I_d, 0]\Omega^{1/2}_{a1}B_m(u)$ *be standard Browian motions of dimensions.*

The proof of Theorem 4.1 is given in the appendix. Ahn and Reinsel (1990) showed that if initial estimates of adjustment parameters Π_{12} and Π_{22} are consistent, then cointergated parameters Π_{11} and Π_{21} can be obtained by two-step estimation. Therefore, we use the initial estimator of adjustment parameter obtained by LSE in Thorem. 4.1

4.4 Monte-carlo simulation and numerical example

In this section, we perform a simulation study to examine the properties of ML, LS and GMM estimator for the simple 4-dimensional vector cointegration model. The objectives of the simulation are to verify the main theorem empirically and to examine the performance of the GMM estimator in small samples.

For the simulation we use the data generating processes suggested by Ahn and Reinsel (1990) with $m_z = m_y = 2$ and $r = 2$, $r_z = r_y = 1$. The true parameters are

$$\Delta \boldsymbol{Y}_t = AB'\boldsymbol{X}_{t-1} + A_{2z}B_z'\boldsymbol{Z}_{t-1} + D(\Delta \boldsymbol{Z}_t - A_zB_z'\boldsymbol{Z}_{t-1}) + \boldsymbol{e}_t$$
$$\Delta \boldsymbol{Z}_t = A_zB_z'\boldsymbol{Z}_{t-1} + \boldsymbol{\epsilon}_{zt},$$

where $A = [-0.33, 0.66]', B = [1, -0.4, 0, -3.2]', A_{2z} = [0.21 - 0.24]', A_z = [-0.29, 0.08]', B_z = [1, -2.5]$ and

$$\boldsymbol{a}_t \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 25 & 6.1 & 4 & -1.2 \\ 6.1 & 9 & 1.8 & 3.9 \\ 4 & 1.8 & 25 & 5.4 \\ -1.2 & 3.9 & 5.4 & 9 \end{bmatrix}\right)$$
$$D = \begin{bmatrix} 4 & -1.2 \\ 1.8 & 3.9 \end{bmatrix} \begin{bmatrix} 25 & 5.4 \\ 5.4 & 9 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2169 & -0.2635 \\ 0.0248 & 0.4482 \end{bmatrix}.$$

In this simulation, the length of series $T = 50, 100$, and 200 are used, and 10000 replications of the sample series are generated for each value of T. For each series, estimates of the parameters are computed by ML, LS and iterative GMM. The empirical results from the simulation for the three estimation procedures are summarized in Table 4.1 and 4.2 where, as the conventional measures of accuracy, the means and mean squares errors(MSE's) of the each estimators are given. Brüggemann and Lütkepohl (2005) show through a Monte Carlo simulation study that the LSE performed better than the MLE in terms of smaller MSE for cointegrating vectors. However, Ahn et al. (2015) show that the MLE performed better in terms of smaller MSE and bias for stationary parameter. Our results are little different from their result. The bias and MSE of GMM are quite smaller than other estimation methods when sample size is small. In addition, GMM estimations are more robust from distribution assumptions. Especially, MSE of GMM is smaller in the estimation of marginal model.

Table 4.1: Means and mean squared errors(MSE)×100 of the GMM and MLE, LSE for various sample size in normal distribution

			GMM	MLE	LSE
$b_{10} = -0.4$	$T = 50$	mean	-0.4006	-0.4010	-3.9924
		MSE	0.0632	0.0645	0.0633
	$T = 100$	mean	-0.4000	-0.4001	-3.9969
		MSE	0.0137	0.0137	0.0137
	$T = 200$	mean	-0.3999	-0.3999	-3.9998
		MSE	0.0034	0.0034	0.0034
$b_{20} = -3.2$	$T = 50$	mean	-3.2017	-3.2081	-3.1882
		MSE	3.0199	3.0257	2.832
	$T = 100$	mean	-3.1996	-3.2005	-3.1948
		MSE	0.6319	0.6262	0.6672
	$T = 200$	mean	-3.1999	-3.1999	-3.1984
		MSE	0.1879	0.1638	0.1887
$b_{z0} = -2.5$	$T = 50$	mean	-2.5060	-2.5096	-2.4896
		MSE	4.4457	4.5620	3.9873
	$T = 100$	mean	-2.5005	-2.5014	-2.4952
		MSE	0.9549	0.9045	0.9687
	$T = 200$	mean	-2.5001	-2.5001	-2.4283
		MSE	0.3170	0.2706	0.3322
$a_1 = -0.33$	$T = 50$	mean	-0.3451	-0.3544	-0.4026
		MSE	0.9856	1.0684	1.7390

Continued on next page

Table 4.1 – *Continued from previous page*

			GMM	MLE	LSE
	$T = 100$	mean	-0.3388	-0.3432	-0.3672
		MSE	0.4422	0.4605	0.6369
	$T = 200$	mean	-0.3359	-0.3381	-0.3505
		MSE	0.1980	0.2018	0.2484
$a_2 = -0.66$	$T = 50$	mean	0.6644	0.6665	0.6480
		MSE	0.3299	0.3359	0.4180
	$T = 100$	mean	0.6640	0.6663	0.6570
		MSE	0.1355	0.1374	0.1590
	$T = 200$	mean	0.6637	0.6669	0.6621
		MSE	0.0638	0.0641	0.0697
$a_{2z,1} = 0.21$	$T = 50$	mean	0.2178	0.2382	0.2433
		MSE	2.6207	2.9503	2.9396
	$T = 100$	mean	0.2151	0.2250	0.2276
		MSE	1.2098	1.2892	1.2992
	$T = 200$	mean	0.2143	0.2191	0.2210
		MSE	0.5483	0.5655	0.5657
$a_{2z,2} = -0.24$	$T = 50$	mean	-0.2371	-0.2311	-0.2298
		MSE	1.0084	1.0366	1.0637
	$T = 100$	mean	-0.2400	-0.2370	-0.2372
		MSE	0.4623	0.4703	0.4800
	$T = 200$	mean	-0.2423	-0.2408	-0.2411
		MSE	0.2122	0.2139	0.2157
$a_{z1} = -0.29$	$T = 50$	mean	-0.2996	-0.2964	-0.3046
		MSE	0.8481	0.8456	0.8535
	$T = 100$	mean	-0.2940	-0.2926	-0.2972

Continued on next page

Table 4.1 – *Continued from previous page*

			GMM	MLE	LSE
		MSE	0.3814	0.3870	0.3950
	$T = 200$	mean	-0.2906	-0.2900	-0.2928
		MSE	0.1715	0.1796	0.1844
$a_{z2} = 0.09$	$T = 50$	mean	0.0965	0.0954	0.0935
		MSE	0.3013	0.3028	0.3100
	$T = 100$	mean	0.0909	0.0905	0.0885
		MSE	0.1357	0.1371	0.1406
	$T = 200$	mean	0.0909	0.0905	0.0885
		MSE	0.1387	0.1371	0.1406

Table 4.2: Means and mean squared errors(MSE)×100 of the GMM and MLE, LSE for various sample size in t-distribution with 3 degree of freedom

			GMM	MLE	LSE
$b_{10} = -0.4$	$T = 50$	mean	-0.4001	-0.4004	-0.3984
		MSE	0.0653	0.0666	0.0653
	$T = 100$	mean	-0.3998	-0.3998	-0.3994
		MSE	0.0142	0.0142	0.0143
	$T = 200$	mean	-0.4000	-0.4000	-0.3999
		MSE	0.0034	0.0034	0.0034
$b_{20} = -3.2$	$T = 50$	mean	-3.2052	-3.2077	-3.1902
		MSE	2.4716	2.7160	2.4899
	$T = 100$	mean	-3.2008	-3.2017	-3.1957
		MSE	0.6202	0.6244	0.6841

Continued on next page

Table 4.2 – *Continued from previous page*

			GMM	MLE	LSE
	$T = 200$	mean	-3.1995	-3.1996	-3.1980
		MSE	0.1721	0.1725	0.1972
$b_{z0} = -2.5$	$T = 50$	mean	-2.5039	-2.5058	-2.4911
		MSE	3.2128	3.7077	3.2530
	$T = 100$	mean	-2.5038	-2.5048	-2.4982
		MSE	1.0431	0.9200	1.0444
	$T = 200$	mean	-2.4989	-2.4990	-2.4973
		MSE	0.2583	0.2572	0.3158
$a_1 = -0.33$	$T = 50$	mean	-0.3444	-0.3541	-0.4032
		MSE	0.9672	1.0522	1.7392
	$T = 100$	mean	-0.3380	-0.3426	-0.3670
		MSE	0.4383	0.4574	0.6288
	$T = 200$	mean	-0.3348	-0.3369	-0.3491
		MSE	0.2008	0.2048	0.2479
$a_2 = -0.66$	$T = 50$	mean	0.6697	0.6656	0.6465
		MSE	0.3235	0.3305	0.4213
	$T = 100$	mean	0.6692	0.6674	0.6580
		MSE	0.1377	0.1395	0.1592
	$T = 200$	mean	0.6679	0.6670	0.6623
		MSE	0.0645	0.0650	0.0701
$a_{2z,1} = 0.21$	$T = 50$	mean	0.2173	0.2388	0.2476
		MSE	2.8748	3.2331	3.3443
	$T = 100$	mean	0.2138	0.2243	0.2282
		MSE	1.3172	1.3999	1.4178
	$T = 200$	mean	0.2129	0.2179	0.2194

Continued on next page

Table 4.2 – *Continued from previous page*

			GMM	MLE	LSE
		MSE	0.5738	0.5921	0.5991
$a_{2z,2} = -0.24$	$T = 50$	mean	-0.2369	-0.2304	-0.2270
		MSE	1.0615	1.0979	1.1459
	$T = 100$	mean	-0.2415	-0.2383	-0.2381
		MSE	0.4671	0.4737	0.4850
	$T = 200$	mean	-0.2419	-0.2404	-0.2409
		MSE	0.2196	0.2214	0.2249
$a_{z1} = -0.29$	$T = 50$	mean	-0.3008	-0.2979	-0.3049
		MSE	0.9156	0.9162	0.9188
	$T = 100$	mean	-0.2954	-0.2940	-0.2985
		MSE	0.4107	0.4113	0.4176
	$T = 200$	mean	-0.2903	-0.2897	-0.2925
		MSE	0.1951	0.1937	0.1980
$a_{z2} = 0.09$	$T = 50$	mean	0.0956	0.0945	0.0934
		MSE	0.2962	0.2970	0.3023
	$T = 100$	mean	0.0902	0.0898	0.0878
		MSE	0.1313	0.1328	0.1373
	$T = 200$	mean	0.0884	0.0882	0.0867
		MSE	0.0627	0.0622	0.0636

We also analyze the grain-meat data which was considered in Ahn et al.Ahn et al. (2015) and compare the performance of estimators, LSE, MLE and GMM, using the MSPE. The grain-meat data consists of six series where three grains are corn, soybean, and wheat and three meats are beef, pork, and chicken. Data set consists of 131 monthly price from January 1980 to November 2008. We divide the whole period into subperiod of size 60 by

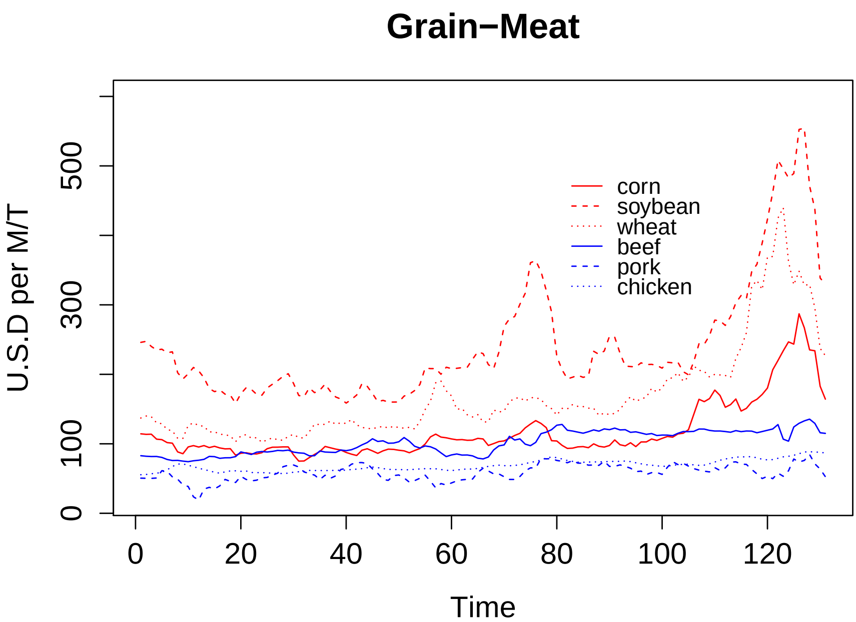

Figure 4.1: Grain-Meat prices

sliding window as follows:

$$\boldsymbol{X}_{j+1}, \boldsymbol{X}_{j+2}, \cdots, \boldsymbol{X}_{j+60}, \quad j = 0, \cdots, 70$$

From a preliminary analysis, we conclude that the data is specified by VAR(2) model and grain prices are exogenous by Akaike information criterion and log-likelihood test statistics, respectively. We also identify the cointegrating rank as 3 in grain-meat prices and 1 in grain prices. We estimate the parameters using 60 observations, and calculate the MSPE for 1-step ahead forecast.

Let y_{1t}, y_{2t}, and y_{3t} be the prices of beef, pork, and chicken and z_{1t}, z_{2t}, and z_{3t} be the prices of corn, soybean, and wheat respectively.

Table 4.3: 1-month ahead mean squared prediction errors(MSPE) of the MLE , LSE and GMM of grain-meat data

	beef	pork	chicken	corn	soybean	wheat
MLE	40.88	55.72	1.02	100.44	465.81	375.73
LSE	41.01	59.31	1.03	107.13	508.59	393.23
GMM	40.40	53.08	1.01	103.73	480.17	373.95

MLE and LSE estimators are obtained from in Ahn et al. (2015). We compare the accuracy of prediction of estimation methods and summarize it in table 3. As Brüggemann and Lütkepohl (2005) mentioned, we observed in some, but rare, cases that the estimators fail to converge by the MLE method in grain-meat data analysis. Table 4.3 constructed using the convergent cases out of 9 subperiods based on MLE. Table 4.3 supports that GMM performs a little bit better in terms of the accuracy of prediction when sample size is relatively small.

Chapter 5

Conclusion

In this thesis, we propose a new asymmetric GARCH-type model which can handle the leverage effect using Yeo-Johnson transformation. We also propose a GMM estimation method to estimate the cointegration model with exogenous variables.

In chapter 3, the YJ-GARCH model is proposed as an effective tool for handling the leverage effect. It is already mentioned in chapter 2 that classical ARCH and GARCH models suffer from many imposed limitations. To avoid those limitations, many researchers have suggested alternative models such as EGARCH, GJR, and TGARCH to name a few. However, some models are difficult to apply to interpreting the meaning of parameters, and while others have arbitrary specification assumptions. The proposed YJ-GARCH model is simple to use and can easily reflect the asymmetric relations using the transformation parameter λ. Its generality with respect to the treatment of asymmetry is illustrated in the empirical examples. The empirical results confirm the recent evidence that the volatility of stock returns reacts differently to the increase and decrease. Although the performance of the multi-step-ahead forecast is simillar to that of other GARCH-type models, YJ-GARCH with

1-step-ahead forecast confirms the superior performance of the YJ-GARCH model compared to other GARCH-type models. In some applications, it is more appropriate to assume that e_t follows a heavy-tailed and skewed distribution, such as the skewed t-distribution Fernández and Steel (1998) used in e_t. We also compare YJ-GARCH with other GARCH-type models considering these conditions in e_t. But EGARCH and TGARCH do not satisfy stationary conditions of the parameter in most of the subperiod windows, and the performance of the GARCH model with skewed t-distribution is similar to those with normal distribution. It is worth to mention that YJ-GARCH model can handle the nonconstant structure of asymmetry when the relative impacts of negative and positive shocks on the current volatility are not the same. Our data typically illustrates this possibility, which leads to moderately skewed shocks concerning Yeo and Johnson transformation. In future research, It would be worth considering various power parameters on variance equation instead of 2 as Ding et al. (1993) suggested.

In chapter 4, we propose an iterative GMM estimation method for the estimation of the cointegrated model with exogenous variables considered by Ahn et al. (2015). Park et al. (2011) suggested the iterative GMM estimation, but only considered the simple cointegrated model without exogenous variables. In this thesis, the GMM estimation is extended to the cointegrated model with exogenous variables. Although the GMM estimation requires the moment conditions the calculation is simple compared to that of the ML in the research of Ahn et al. (2015). Phillips (1994) notes that MLE on cointegrated model occasionally does not work or results in cointegration parameter estimates which are far different from the true parameters in small samples. It was observed through the Monte-Carlo simulations and numerical examples that GMM does not show such phenomenon. It was also observed, however, that the efficiency of the iterative GMM estimator depends on the choice of the weighting matrix in the objective function. For this reason, it is important to choose the proper weighting matrix in the objective function. Hansen et al. (1996) suggest the continuous-updating method and Imbens et al. (1995) propose an alternative method that uses the information theoretic approaches. Further investigations in terms of new methods of weighting matrix choice remain in order to understand exactly when and which of these methods should be used. With this consideration in mind, therefore, this thesis suggests the utilization of GMM estimation in a complex cointegrated model.

Appendix A

Lemmas

Lemma A.1. *For all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\left|\frac{\partial h_t}{\partial \boldsymbol{\theta}} h_t^{-1}\right| < \boldsymbol{M}^*$$

for $\boldsymbol{\theta}' = (\alpha_0, \alpha_1, \beta_1, \mu, \lambda)$*, where* $\boldsymbol{M}^*$ *does not depend on* $\boldsymbol{\theta}$*. Therefore, each of these partial derivatives has bounded expectation and all higher moments are bounded.*

Proof of Lemma A.1. First consider the the conditional variance of $h_t = \alpha_0 + \alpha_1 \psi(\lambda, \epsilon_{t-1})^2 + \beta_1 h_{t-1}$ and assume the parameter space as below

$$\boldsymbol{\Theta} = \Big\{ \boldsymbol{\theta}' = (\alpha_0, \alpha_1, \beta_1, \mu, \lambda) | c \leq \alpha_0 \leq d, a \leq \lambda \leq b, \delta \leq \alpha_1, \beta_1 \leq 1 - \delta, |\mu| \leq r$$
$$\textit{with} - \infty < a < 0 < \delta, c, d < \infty, \textit{ and } 2 < b < \infty \Big\}$$

we can easily show that $h_t \geq \alpha_0 \geq c$, so that $h_t^{-1} \leq c^{-1}$. Differentiating the h_t with respect to each of the parameters in $\boldsymbol{\theta}$,

1. $\dfrac{\partial h_t}{\partial \alpha_0} = 1 + \beta_1 \dfrac{\partial h_{t-1}}{\partial \alpha_0} = \displaystyle\sum_{i=0}^{t-1} \beta_1^i + \beta_1^t \dfrac{\partial h_0}{\partial \alpha_0} \leq \dfrac{1}{1 - \beta_1}$

2. $\dfrac{\partial h_t}{\partial \alpha_1} = \psi(\lambda, \epsilon_{t-1})^2 + \beta_1 \dfrac{\partial h_{t-1}}{\partial \alpha_0} = \sum_{i=0}^{t-1} \beta_1^i \psi(\lambda, \epsilon_{t-1-i})^2 + \beta_1^t \dfrac{\partial h_0}{\partial \alpha_1}$

3. $\dfrac{\partial h_t}{\partial \beta_1} = h_{t-1} + \beta_1 \dfrac{\partial h_{t-1}}{\partial \beta_1} = \sum_{i=0}^{t-1} \beta_1^i h_{t-1-i} + \beta_1^t \dfrac{\partial h_0}{\partial \beta_1}$

4. $\Big|\dfrac{\partial h_t}{\partial \mu}\Big| = \Big| - 2\alpha_1 \psi(\lambda, \epsilon_{t-1}) \dfrac{\partial \psi(\lambda, \epsilon_{t-1})}{\partial \mu} + \beta_1 \dfrac{\partial h_{t-1}}{\partial \mu}\Big|$
$\leq 2\alpha_1 \sum_{i=0}^{t-1} \beta_1^i \Big|\psi(\lambda, \epsilon_{t-1-i}) \dfrac{\partial \psi(\lambda, \epsilon_{t-1-i})}{\partial \mu}\Big| + \beta_1^t \dfrac{\partial h_0}{\partial \mu}$

5. $\Big|\dfrac{\partial h_t}{\partial \lambda}\Big| = \Big|2\alpha_1 \psi(\lambda, \epsilon_{t-1}) \dfrac{\partial \psi(\lambda, \epsilon_{t-1})}{\partial \lambda} + \beta_1 \dfrac{\partial h_{t-1}}{\partial \lambda}\Big|$
$\leq 2\alpha_1 \sum_{i=0}^{t-1} \beta_1^i \Big|\psi(\lambda, \epsilon_{t-1-i}) \dfrac{\partial \psi(\lambda, \epsilon_{t-1-i})}{\partial \lambda}\Big| + \beta_1^t \dfrac{\partial h_0}{\partial \lambda}$

conditioning on h_0, note that $\partial h_0 / \partial \boldsymbol{\theta} = \mathbf{0}$. the expressions for $\partial h_t / \partial \alpha_0$ and $\partial h_t / \partial \alpha_1$, $\partial h_t / \partial \beta_1$ appear in the expression for h_t, we have with probability one

$$\Big|\frac{\partial h_t}{\partial \alpha_0} h_t^{-1}\Big| \leq \Big|\frac{(1-\beta_1^t)/(1-\beta_1)}{\alpha_0 (1-\beta_1^t)/(1-\beta_1)}\Big| = \frac{1}{\alpha_0}$$

Similarly,

$$\Big|\frac{\partial h_t}{\partial \alpha_1} h_t^{-1}\Big| \leq \frac{1}{\alpha_1} \quad \text{and} \quad \Big|\frac{\partial h_t}{\partial \beta_1} h_t^{-1}\Big| \leq \frac{t}{\beta_1}$$

using the inequality $|x| \leq x^2 + 1$

$$\Big|\frac{\partial h_t}{\partial \mu} h_t^{-1}\Big| \leq \frac{2\Big\{2\big(\psi_\mu^{(1)}(b, y_t - r)^2 + \psi_\mu^{(1)}(a, y_t + r)^2\big)\alpha_0 + \alpha_1\Big\}}{\alpha_0}$$
$$\Big|\frac{\partial h_t}{\partial \lambda} h_t^{-1}\Big| \leq \frac{2\Big\{2\big(\psi^{(1)}(a, y_t - r)^2 + \psi^{(1)}(b, y_t + r)^2\big)\alpha_0 + \alpha_1\Big\}}{\alpha_0}$$

Since $\boldsymbol{\Theta}$ is a compact parameter space and bounded function $\boldsymbol{M}^*$s are continuous function of $\boldsymbol{\theta}$

$$\Big|\frac{\partial h_t}{\partial \boldsymbol{\theta}} h_t^{-1}\Big|$$

are uniformly bounded, and it follows immediately that all moments and croos moments exist and are finite. □

Lemma A.2. *Let* $\boldsymbol{\Theta}$ *be tha parameter space and define that is as above (9). Then, all the second order partial derivatives of* $l_t(\boldsymbol{\theta})$ *are continuous in* $(\boldsymbol{\theta}, \boldsymbol{y})$ *and, for all* $\boldsymbol{\theta}' = (\alpha_0, \alpha_1, \beta_1, \mu, \lambda)$,

$$\Big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\Big| \le \boldsymbol{D}(y_t)$$

where the $\boldsymbol{D}$*'s are given in the proof*

Proof of Lemma A.2. We list the derivatives and provide their corresponding dominating function. Since $\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_i}$ for $\boldsymbol{\theta}' = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\alpha_0, \alpha_1, \beta_1, \mu, \lambda)$, we only present D_{ij} for $i \le j$

Redifine the conditional variance h_t

$$h_t = \alpha_0 + \alpha_1 \psi(\lambda, \epsilon_{t-1})^2 + \beta_1 h_{t-1} = \alpha_0 \sum_{i=0}^{t-1} \beta_1^i + \alpha_1 \sum_{i=0}^{t-1} \psi(\lambda, \epsilon_{t-1-i})^2 + \beta_1^t h_0$$

1. $$\Big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \alpha_0^2}\Big| = \Big|\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_0}\frac{\partial h_t}{\partial \alpha_0}\Big(\frac{1}{2} - \frac{\epsilon_t^2}{h_t}\Big)\Big| \le \frac{1}{c^2}\Big(\frac{1}{2} + \frac{2(y_t^2 + r^2)}{c}\Big) = D_{11}(y_t) < \infty$$

2. $$\Big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \alpha_0 \partial \alpha_1}\Big| = \Big|\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_0}\frac{\partial h_t}{\partial \alpha_1}\Big(\frac{1}{2} - \frac{\epsilon_t^2}{h_t}\Big)\Big| \le \frac{1}{c\delta}\Big(\frac{1}{2} + \frac{2(y_t^2 + r^2)}{c}\Big) = D_{12}(y_t) < \infty$$

3. $$\Big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \alpha_0 \partial \beta_1}\Big| = \Big|\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_0}\frac{\partial h_t}{\partial \beta_1}\Big(\frac{1}{2} - \frac{\epsilon_t^2}{h_t}\Big) + \frac{1}{2h_t}\frac{\partial^2 h_t}{\partial \alpha_0 \partial \beta_1}\Big(\frac{\epsilon_t^2}{h_t} - 1\Big)\Big|$$
$$\le \frac{t}{c\delta}\Big(\frac{1}{2} + \frac{2(y_t^2 + r^2)}{c}\Big) + \frac{1}{2c\delta}\Big(\frac{2(y_t^2 + \mu^2)}{c} + 1\Big) = D_{13}(y_t) < \infty$$

4. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\mu}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha_0}\frac{\partial h_t}{\partial\mu}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big) - \frac{\epsilon_t}{h_t^2}\frac{\partial h_t}{\partial\alpha_0}\right|$$
$$\leq \frac{2\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}}{c^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)$$
$$+\frac{y_t+r}{c^2}$$
$$= D_{14}(y_t) < \infty$$

5. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\alpha_0\partial\lambda}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha_0}\frac{\partial h_t}{\partial\lambda}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big)\right|$$
$$\leq \frac{2\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)$$
$$= D_{15}(y_t) < \infty$$

6. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\alpha_1^2}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha_1}\frac{\partial h_t}{\partial\alpha_1}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big)\right| \leq \frac{1}{\delta^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big) = D_{22}(y_t) < \infty$$

Since proof (2) of Lemma A.1

7. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\alpha_1\partial\beta_1}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha_1}\frac{\partial h_t}{\partial\beta_1}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big) + \frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\alpha_1\partial\beta_1}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)\right|$$
$$\leq \frac{t}{\delta^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big) + \frac{1}{2\delta^2}\Big(\frac{2(y_t^2+r^2)}{c}+1\Big) = D_{23}(y_t) < \infty$$

Since $\left|\frac{\partial^2 h_t}{\partial\alpha_1\partial\beta_1}h_t^{-1}\right| \leq \frac{1}{\alpha_1(1-\beta_1)} \leq \frac{1}{\delta^2}$ and proof (2) (3) of Lemma A.1

8. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\alpha_1\partial\mu}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha_1}\frac{\partial h_t}{\partial\mu}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big) + \frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\alpha_1\partial\mu}\Big(\frac{\epsilon_t^2}{h_t}-1\Big) - \frac{\epsilon_t}{h_t^2}\frac{\partial h_t}{\partial\alpha_1}\right|$$
$$\leq \frac{2\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)$$
$$+\frac{2\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{2(y_t^2+r^2)}{c}+1\Big)$$
$$+\frac{y_t+r}{d\delta} = D_{24}(y_t) < \infty$$

Since $\frac{\partial^2 h_t}{\partial\alpha_1\partial\mu}h_t^{-1} = \frac{1}{\alpha_1}\frac{\partial h_t}{\partial\mu}$ and proof (2) (4) of Lemma A.1

9. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\alpha_1\partial\lambda}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha_1}\frac{\partial h_t}{\partial\lambda}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big) + \frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\alpha_1\partial\lambda}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)\right|$$
$$\le \frac{2\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$
$$+\frac{\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{2(y_t^2+\mu^2)}{c}+1\Big)$$
$$= D_{25}(y_t) < \infty$$

Since $\frac{\partial^2 h_t}{\partial\alpha_1\partial\lambda}h_t^{-1} = \frac{1}{\alpha_1}\frac{\partial h_t}{\partial\lambda}$ and proof (2) (5) of Lemma A.1

10. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\beta_1^2}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\beta_1}\frac{\partial h_t}{\partial\beta_1}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big) + \frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\beta_1\partial\beta_1}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)\right|$$
$$\le \frac{t^2}{\delta^2}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)+\frac{1}{2}\frac{t}{\delta^2}\Big(\frac{2(y_t^2+\mu^2)}{c}+1\Big) = D_{33}(y_t) < \infty$$

By Proof (3) of Lemma A.1

11. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\beta_1\partial\mu}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\beta_1}\frac{\partial h_t}{\partial\mu}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big) + \frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\beta_1\partial\mu}\Big(\frac{\epsilon_t^2}{h_t}-1\Big) - \frac{\epsilon_t}{h_t^2}\frac{\partial h_t}{\partial\beta_1}\right|$$
$$\le \frac{2t\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2}+\frac{2(y_t^2-\mu^2)}{h_t}\Big)$$
$$+\frac{\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}}{2c\delta}\Big(\frac{2(y_t^2-\mu^2)}{c}+1\Big)$$
$$+\frac{(y_t+r)t}{c\delta} = D_{34}(y_t) < \infty$$

By Proof (3) (4) of Lemma A.1

$$12.\quad \left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\beta_1\partial\lambda}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\beta_1}\frac{\partial h_t}{\partial\lambda}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big)+\frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\beta_1\partial\lambda}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)\right|$$

$$\leq \frac{2t\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$

$$+\frac{\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{2(y_t^2+\mu^2)}{c}+1\Big)$$

$$= D_{35}(y_t) < \infty$$

By Proof (3) (5) of Lemma A.1

$$13.\quad \left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\mu^2}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\mu}\frac{\partial h_t}{\partial\mu}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big)\right.$$

$$\left.+\frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\mu\partial\mu}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)-\frac{2\epsilon_t}{h_t^2}\frac{\partial h_t}{\partial\mu}-\frac{1}{h_t}\right|$$

$$\leq \frac{4\Big\{4\big(\psi_\mu^{(1)}(b,y_t-r)^4+\psi_\mu^{(1)}(a,y_t+r)^4\big)d^2+(1-\delta)^2\Big\}}{c^2}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$

$$+\frac{(1-\delta)}{c}\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)$$

$$+\big(|\psi(a,y_t+r)|+|\psi(b,y_t+r)|\big)\big(\psi_\mu^{(2)}(b,y_t+r)+\psi_\mu^{(2)}(a,y_t-r)\big)\Big\}$$

$$+\frac{4\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}(y_t+r)}{c^2}+\frac{1}{c}$$

$$= D_{44}(y_t) < \infty$$

By Proof (4) of Lemma A.1

14. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\mu\partial\lambda}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\mu}\frac{\partial h_t}{\partial\lambda}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big)+\frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\mu\partial\lambda}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)\right|$$
$$\leq \frac{4}{c^2}\Big\{2\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}$$
$$\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}\Big\}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$
$$+\frac{(1-\delta)}{c}\Big\{\big(\psi_\mu^{(1)}(b,y_t-r)+\psi_\mu^{(1)}(a,y_t+r)\big)\big(\psi^{(1)}(a,y_t-r)+\psi^{(1)}(b,y_t+r)\big)$$
$$+\big(|\psi(b,y_t+r)|+|\psi(b,y_t+r)|\big)\big(|\psi_{\lambda\mu}^{(2)}(b,y_t-r)|+|\psi_{\lambda\mu}^{(2)}(a,y_t+r)|\big)\Big\}$$
$$\Big(\frac{2(y_t^2-\mu^2)}{c}-1\Big)=D_{45}(y_t)<\infty$$

15. $$\left|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\lambda^2}\right| = \left|\frac{1}{h_t^2}\frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda}\Big(\frac{1}{2}-\frac{\epsilon_t^2}{h_t}\Big)+\frac{1}{2h_t}\frac{\partial^2 h_t}{\partial\lambda\partial\lambda}\Big(\frac{\epsilon_t^2}{h_t}-1\Big)\right|$$
$$\leq \frac{4\Big\{4\big(\psi^{(1)}(a,y_t-r)^4+\psi^{(1)}(b,y_t+r)^4\big)d^2+(1-\delta)^2\Big\}}{c^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)$$
$$+\frac{(1-\delta)}{c}\Big\{2\big(\psi^{(1)}(b,y_t-r)^2+\psi^{(1)}(a,y_t+r)^2\big)$$
$$+\big(|\psi(b,y_t+r)|+|\psi(b,y_t+r)|\big)\big(|\psi^{(2)}(a,y_t-r)|+|\psi^{(2)}(b,y_t+r)|\big)\Big\}$$
$$\Big(\frac{2(y_t^2-\mu^2)}{c}-1\Big)=D_{55}(y_t)<\infty$$

Since $\psi,\psi^{(1)}$ and $\psi^{(1)}$ are continuous functions of (λ, y) according to (4) of Lemma A. 2 in Yeo and JohnsonYeo and Johnson (2000), all the second order partial derivative of $l_t(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$ and y. □

Lemma A.3. *Let* $\boldsymbol{\Theta}$ *and* D_{ij} *be defined as in Lemma A.1, A.2. and let* y_t *be a random variable and assume that*

$$E[I_{(y_t<0)}(-y_t)^{4(2-a)}log^4(-y_t+1)]<\infty \quad and \quad E[I_{(y_t\geq 0)}y_t^{4b}log^4(y_t+1)]<\infty \tag{A.1}$$

Then, $E[D_{ij}^2(y_t)]$ *is finite.*

Proof of Lemma A.3. Fist, we recall two ineqalities (Hardy, Littlewood and Polya Hardy et al. (1952), theorem 150 and equation (5.2.4)), $log(x) \leq x - 1$ for $x > 0$ and

$$\frac{z^a - 1}{a} < \frac{z^b - 1}{b} \quad \text{for } a < b \text{ and } z > 0,$$

where the second follows directly from (5.2.4) and the fact that for $a < 0, log(z^a) \leq z^a - 1$ or $log(z) \geq (z^a - 1)/a$. Letting $z = y + 1, x = (y + 1)^{-b}$ for $b > 2$ and $x = (y + 1)^{-a}$ for $a < 0$, we obtain that

$$\begin{aligned} 0 \leq (y+1)^a log(y+1) &\leq \frac{(y+1)^a - 1}{a} \leq log(y+1) \\ &\leq \frac{(y+1)^b - 1}{b} \leq (y+1)^b log(y+1), \qquad \text{for } y \geq 0 \end{aligned} \tag{A.2}$$

Similarly, for $y < 0$, we let $z = -y + 1$ and first take $x = (-y + 1)^{-2+b}$ and then $x = (-y + 1)^{-2+a}$.

$$\begin{aligned} 0 \leq (-y+1)^{2-b} log(-y+1) &\leq \frac{(-y+1)^{2-b} - 1}{2-b} \leq log(-y+1) \\ &\leq \frac{(-y+1)^{2-a} - 1}{2-a} \leq (-y+1)^{2-a} log(-y+1), \qquad \text{for } y \geq 0 \end{aligned}$$

By (2.12), $0 \leq \psi(a, y) \leq \psi(b, y)$ for $y \geq 0$ and $0 < |\psi(b, y)| < |\psi(a, y)|$ for $y < 0$. Consequently,

$$max\Big\{|\psi(a,y)|, \psi(b,y)\Big\} \leq I_{(y \geq 0)}(y+1)^b/b + I_{(y<0)}(-y+1)^{2-a}/(2-a) \tag{A.3}$$

According to the expression (4) in Lemma 2 in Yeo and JohnsonYeo and Johnson (2000),

$$\psi^{(1)}(\lambda, y) = \begin{cases} \Big[(y+1)^\lambda log(y+1) - \psi(\lambda, y)\Big]/\lambda & \text{for } y \geq 0, \\ \Big[(-y+1)^{2-\lambda} log(-y+1) - \psi(\lambda, y)/(2-\lambda)\Big] & \text{for } y < 0, \end{cases}$$

Thus, for $y \geq 0, \psi^{(1)}(b, y) \geq (y+1)^b log(y+1)/b$ and $\psi^{(1)}(a, y) \leq \frac{(-y+1)^a - 1}{-a^2} \leq (y+1)^b log(y+1)/(-a)$, where the last inequality follow by (A.2). Similarly for $y < 0, \psi^{(1)}(a,y) \leq (-y+1)^{2-a} log(-y+1)/(2-a)$. Also $\psi^{(1)}(b,y) \geq \dfrac{(-y+1)^{2-b}-1}{-(2-b)^2} \leq (-y+1)^{2-a} log(-y+1)/(b-2)$ so we conclude that

$$\begin{aligned} max\Big\{\psi^{(1)}(a,y), \psi^{(1)}(b,y)\Big\} \leq\ & I_{(y\geq 0)}\frac{a-b}{ab}(y+1)^b log(y+1) \\ & + I_{(y<0)}\frac{b-a}{(2-a)(b-2)}(-y+1)^{2-a} log(-y+1) \end{aligned}$$

Applying Lemma A.1 in Yeo and JohnsonYeo and Johnson (2000) to $\phi(\lambda, y)$ is convex in λ for $y > 0$ and concave in λ for $y < 0$, we obtain that $\psi^{(2)} \geq 0$ if $y \geq 0$ and < 0 otherwise. According to the expression (4) in Lemma 2 in Yeo and JohnsonYeo and Johnson (2000),

$$\psi^{(2)}(\lambda, y) = \begin{cases} \Big[(y+1)^{\lambda} log(y+1) - \psi^{(1)}(\lambda, y)\Big]/\lambda & \text{for } y \geq 0, \\ \Big[(-y+1)^{2-\lambda} log(-y+1) - \psi^{(1)}(\lambda, y)/(2-\lambda)\Big] & \text{for } y < 0, \end{cases}$$

Hence, for $y \geq 0, \psi^{(2)}(a,y) \leq \psi^{(1)}(a,y)/(-a)$ and $\psi^{(2)}(b,y) \leq (y+1)^b log^2(y+1)/b$.

Similarly, since $\psi^{(2)} < 0$ for $y < 0$, $\Big|\psi^{(2)}(a,y)\Big| \leq (-y+1)^{2-a} log^2(-y+1)/(2-a)$ and $\Big|\psi^{(2)}(b,y)\Big| \leq 2\psi^{(1)}(b,y)/(b-2)$. We conclude that

$$\Big|\psi^{(2)}(a,y)\Big| \leq I_{(y\geq 0)} 2\psi(1)(a,y)/(-a) + I_{(y<0)}(-y+1)^{2-a} log^2(-y+1)/(2-a)$$

and

$$\Big|\psi^{(2)}(b,y)\Big| \leq I_{(y\geq 0)}(y+1)^b log^2(y+1)/b + I_{(y<0)} 2\psi^{(1)}(b,y)/(b-2)$$

(1) Since (A.1) implies that $E[y_t^4]$ are finte and $D_{11}(y_t) = \dfrac{1}{c^2}\Big(\dfrac{1}{2} + \dfrac{2(y_t^2+r^2)}{c}\Big)$,

$$E[D_{11}^2(y_t)] = \frac{2}{c^4} E\Big(\frac{1}{4} + \frac{4(y_t^4+r^4)}{c^2}\Big) < \infty$$

(2) Since (A.1) implies that $E[y_t^4]$ are finte and $D_{12}(y_t) = \dfrac{1}{c\delta}\Big(\dfrac{1}{2} + \dfrac{2(y_t^2+r^2)}{c}\Big)$,

$$E[D_{12}^2(y_t)] = \frac{2}{c^2\delta^2} E\Big(\frac{1}{4} + \frac{4(y_t^4+r^4)}{c^2}\Big) < \infty$$

(3) Since (A.1) implies that $E[y_t^4]$ are finte and $D_{13}(y_t) = \dfrac{t}{c\delta}\Big(\dfrac{1}{2}+\dfrac{2(y_t^2+r^2)}{c}\Big)+\dfrac{1}{2c\delta}\Big(\dfrac{2(y_t^2+r^2)}{c}+1\Big)$,

$$E[D_{13}^2(y_t)] = E\left[\left(\frac{t}{c\delta}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)+\frac{1}{2c\delta}\Big(\frac{2(y_t^2+r^2)}{c}+1\Big)\right)^2\right]$$
$$\leq \frac{2t^2}{c^2\delta^2}E\left[\frac{1}{4}+\frac{4(y_t^4+r^4)}{c^2}\right]+\frac{1}{c^2\delta^2}E\left[\frac{(y_t^4+r^4)}{c^2}+1\right] < \infty$$

(4) Since (A.1) implies that $E[y_t^4]$ are finte, so

$$E[D_{14}(y_t)] = E\left[\left(\frac{2\Big\{2\big(\psi_\mu^{(1)}(b, y_t-r)^2+\psi_\mu^{(1)}(a, y_t+r)^2\big)d+(1-\delta)\Big\}}{c^2}\right.\right.$$
$$\left.\left.\times\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)+\frac{y_t+r}{c^2}\right)^2\right] < \infty$$

(5) Since $D_{15}(y_t)$,

$$E[D_{15}^2(y_t)] = E\left[\left(\frac{2\Big\{2\big(\psi^{(1)}(a, y_t-r)^2+\psi^{(1)}(b, y_t+r)^2\big)d+(1-\delta)\Big\}}{c^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)\right)^2\right]$$
$$< \infty$$

(6) Since $D_{22}(y_t)$,

$$E[D_{22}^2(y_t)] = E\left[\left(\frac{1}{\delta^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)\right)^2\right] < \infty$$

(7) Since $D_{23}(y_t)$,

$$E[D_{23}^2(y_t)] = E\left[\left(\frac{t}{\delta^2}\Big(\frac{1}{2}+\frac{2(y_t^2+r^2)}{c}\Big)+\frac{1}{2\delta^2}\Big(\frac{2(y_t^2+r^2)}{c}+1\Big)\right)^2\right] < \infty$$

(8) Since $D_{24}(y_t)$,

$$E[D_{24}^2(y_t)] = E\Bigg[\Bigg(\frac{2\Big\{2\big(\psi_\mu^{(1)}(b, y_t - r)^2 + \psi_\mu^{(1)}(a, y_t + r)^2\big)d + (1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2} + \frac{2(y_t^2 + r^2)}{c}\Big)$$
$$+ \frac{2\Big\{2\big(\psi_\mu^{(1)}(b, y_t - r)^2 + \psi_\mu^{(1)}(a, y_t + r)^2\big)d + (1-\delta)\Big\}}{c\delta}\Big(\frac{2(y_t^2 + r^2)}{c} + 1\Big)$$
$$+ \frac{y_t + r}{d\delta}\Bigg)^2\Bigg] < \infty$$

(9) Since $D_{25}(y_t)$,

$$E[D_{25}^2(y_t)] = E\Bigg[\Bigg(\frac{2\Big\{2\big(\psi^{(1)}(a, y_t - r)^2 + \psi^{(1)}(b, y_t + r)^2\big)d + (1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2} + \frac{2(y_t^2 + \mu^2)}{c}\Big)$$
$$+ \frac{\Big\{2\big(\psi^{(1)}(a, y_t - r)^2 + \psi^{(1)}(b, y_t + r)^2\big)d + (1-\delta)\Big\}}{c\delta}\Big(\frac{2(y_t^2 + \mu^2)}{c} + 1\Big)\Bigg)^2\Bigg]$$
$$< \infty$$

(10) Since $D_{33}(y_t)$,

$$E[D_{33}^2(y_t)] = E\Bigg[\Bigg(\frac{t^2}{\delta^2}\Big(\frac{1}{2} + \frac{2(y_t^2 + \mu^2)}{c}\Big) + \frac{1}{2}\frac{t}{\delta^2}\Big(\frac{2(y_t^2 + \mu^2)}{c} + 1\Big)\Bigg)^2\Bigg] < \infty$$

(11) Since $D_{34}(y_t)$,

$$E[D_{34}^2(y_t)] = E\Bigg[\Bigg(\frac{2t\Big\{2\big(\psi_\mu^{(1)}(b, y_t - r)^2 + \psi_\mu^{(1)}(a, y_t + r)^2\big)d + (1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2} + \frac{2(y_t^2 - \mu^2)}{h_t}\Big)$$
$$+ \frac{\Big\{2\big(\psi_\mu^{(1)}(b, y_t - r)^2 + \psi_\mu^{(1)}(a, y_t + r)^2\big)d + (1-\delta)\Big\}}{2c\delta}\Big(\frac{2(y_t^2 - \mu^2)}{c} + 1\Big)$$
$$+ \frac{(y_t + r)t}{c\delta}\Bigg)^2\Bigg] < \infty$$

(12) Since $D_{35}(y_t)$,

$$E[D_{35}^2(y_t)] = E\Bigg[\Bigg(\frac{2t\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$

$$+\frac{\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}}{c\delta}\Big(\frac{2(y_t^2+\mu^2)}{c}+1\Big)\Bigg)^2\Bigg] < \infty$$

(13) Since $D_{44}(y_t)$,

$$E[D_{44}^2(y_t)] = E\Bigg[\Bigg(\frac{4\Big\{4\big(\psi_\mu^{(1)}(b,y_t-r)^4+\psi_\mu^{(1)}(a,y_t+r)^4\big)d^2+(1-\delta)^2\Big\}}{c^2}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$

$$+\frac{(1-\delta)}{c}\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)$$

$$+\big(|\psi(a,y_t+r)|+|\psi(b,y_t+r)|\big)\big(\psi_\mu^{(2)}(b,y_t+r)+\psi_\mu^{(2)}(a,y_t-r)\big)\Big\}$$

$$+\frac{4\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}(y_t+r)}{c^2}+\frac{1}{c}\Bigg)^2\Bigg] < \infty$$

(14) Since $D_{45}(y_t)$,

$$E[D_{45}^2(y_t)] = E\Bigg[\Bigg(\frac{4}{c^2}\Big\{2\Big\{2\big(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\big)d+(1-\delta)\Big\}$$

$$\Big\{2\big(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\big)d+(1-\delta)\Big\}\Big\}\Big(\frac{1}{2}+\frac{2(y_t^2+\mu^2)}{c}\Big)$$

$$+\frac{(1-\delta)}{c}\Big\{\big(\psi_\mu^{(1)}(b,y_t-r)+\psi_\mu^{(1)}(a,y_t+r)\big)\big(\psi^{(1)}(a,y_t-r)+\psi^{(1)}(b,y_t+r)\big)$$

$$+\big(|\psi(b,y_t+r)|+|\psi(b,y_t+r)|\big)\big(|\psi_{\lambda\mu}^{(2)}(b,y_t-r)|+|\psi_{\lambda\mu}^{(2)}(a,y_t+r)|\big)\Big\}$$

$$\Big(\frac{2(y_t^2-\mu^2)}{c}-1\Big)\Bigg)^2\Bigg] < \infty$$

(15) Since $D_{55}(y_t)$,

$$E[D_{55}^2(y_t)] = E\Bigg[\Bigg(\frac{4\Big\{4\big(\psi^{(1)}(a, y_t - r)^4 + \psi^{(1)}(b, y_t + r)^4\big)d^2 + (1-\delta)^2\Big\}}{c^2}\Big(\frac{1}{2} + \frac{2(y_t^2 + r^2)}{c}\Big)$$

$$+ \frac{(1-\delta)}{c}\Big\{2\big(\psi^{(1)}(b, y_t - r)^2 + \psi^{(1)}(a, y_t + r)^2\big)$$

$$+ \big(|\psi(b, y_t + r)| + |\psi(b, y_t + r)|\big)\big(|\psi^{(2)}(a, y_t - r)| + |\psi^{(2)}(b, y_t + r)|\big)\Big\}$$

$$\Big(\frac{2(y_t^2 - \mu^2)}{c} - 1\Big)\Bigg)^2\Bigg] < \infty$$

Therefore, $E[D_{ij}^2(y_t)] < \infty$, for $i = 1, \cdots, 5$ and $j = 1, \cdots, 5$ □

Appendix B

Proofs of Theorem

Proof of Theorem 3.1.

(A) We employ Lemma A. 1 in Yeo and JohnsonYeo and Johnson (2000) to establish the result (A). Let $l_t : \boldsymbol{\Theta} \times \boldsymbol{R} \rightarrow \boldsymbol{R}$ be determined a normal likelihood after transformation as

$$l_t(\boldsymbol{\theta}) = -\frac{1}{2}log(2\pi) - \frac{1}{2}log(h_t) - \frac{\epsilon_t^2}{2h_t} \tag{B.1}$$

where $h_t = \alpha_0 + \alpha_1\psi(\lambda, \epsilon_{t-1})^2 + \beta_1 h_{t-1}$. Then, for all $\theta \in \Theta$,

$$\begin{aligned} |l_t(\theta|y_i)| \leq g(y_i) &= log(2\pi) \\ &\quad + log\left(\frac{1}{\delta}\left\{d + 2(1-\delta)\left[\psi(a, y_{t-1} - r)^2 + \psi(b, y_{t-1} + r)^2\right]\right\}\right) \\ &\quad + \frac{2(y_t^2 + r^2)}{c} \end{aligned} \tag{B.2}$$

Next, recall the C_r inequality (see Loeve(1978)), $|x + y|^r \leq C_r\big(|x|^r + |y|^r\big)$ for any x, y and $r \geq 0$, where $C_r = 2^{r-1}$ if $r \geq 1$ and $C_r = 1$ if

$r < 1$. we obtain from (A.3)

$$\begin{aligned} max|\psi(\lambda, y_t-\mu)| \leq & I_{y_t-\mu\geq 0}(y_t-\mu+1)^b/b + I_{y_t-\mu<0}(-y_t+\mu+1)^{(2-a)}/(2-a) \\ \leq & I_{y_t-\mu\geq 0}C_b(y_t-\mu+1)/b \\ & + C_a I_{y_t-\mu<0}((-y_t)^{(2-a)}+\mu^{(2-a)}+1)/(2-a) \\ \leq & I_{y_t-\mu\geq 0}C_b(y_t^b+r^b+1)/b \\ & + C_a I_{y_t-\mu<0}((-y_t)^{(2-a)}+r^{(2-a)}+1)/(2-a), \end{aligned}$$

for $a \leq \lambda \leq b$, where C_a and C_b are given above. According to this bound, assumption (c.3) guarantees from (B.2) that $\sup_i E[g^2(Y_i)]$ is finite so that the first condition of Lemma 4.1 in Yeo and JohnsonYeo and Johnson (2000) holds with $\varphi(\cdot) = g(\cdot)$. Let $S_M = [-M, M]$. Because $|Y_i| \leq |Y_i^{(a)}| + |Y_i^{(b)}|$, Markov's inequality allows us to conclude that $\sup_i P(|Y_i| > M) \leq \sup_i E|Y_i|/M \to 0$ as $M \to \infty$. Thus the second condition is verified. Since $l_t(\boldsymbol{\theta})$ is continuous in $(\boldsymbol{\theta}', \boldsymbol{Y})$ over the compact set $\boldsymbol{\Theta} \times S_M$, $l_t(\boldsymbol{\theta})$ is equicontinuous in $\boldsymbol{\theta}$ for $Y_i \in S_M$ (see KosmalaKosmala (1995)). Thus all of the conditions of Lemma 4.1 in Yeo and JohnsonYeo and Johnson (2000) are satified. We conclude that

$$\frac{1}{n}L_n(\boldsymbol{\theta}) - \Pi_n(\boldsymbol{\theta}) \xrightarrow{a.s} 0$$

uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\Pi_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} E[l_t(\boldsymbol{\theta})]$. Equivalently,

$$\lim_{n\to\infty}\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}L_n(\boldsymbol{\theta}) - \Pi_n(\boldsymbol{\theta})\right\} = 0 \qquad \text{(B.3)}$$

with probability one. The result (A) follows directly since

$$\left|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} = \left|\frac{1}{n}L_n(\boldsymbol{\theta})\right| - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\Pi_n(\boldsymbol{\theta})\right|\right| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}L_n(\boldsymbol{\theta}) - \Pi_n(\boldsymbol{\theta})\right| \to 0$$

with probability one.

(B) To establish the strong consistency of $\hat{\boldsymbol{\theta}}$, we introduce the notation ω for a generic outcome and A for the set where the uniform almost sure convergence (A) holds. To obtain a contradiction, we assume that $\hat{\boldsymbol{\theta}}$ does not converge to $\boldsymbol{\theta}_0$ almost surely, so there exists a set of outcome B where $\hat{\boldsymbol{\theta}}$ does not converge to $\boldsymbol{\theta}_0$ almost surely and $P(B) > 0$. Without loss of generality, we can restrict our attention to the set $C = A \cap B$ with $P(C) > 0$.

Since $\boldsymbol{\Theta}$ is compact, for each $\omega \in C$, there exists a subsequence $\{m\} \subset \{n\}$ and a limit point $\boldsymbol{\theta}_*(\omega)$ with $\hat{\boldsymbol{\theta}}_m(\omega) \to \boldsymbol{\theta}_*(\omega) \neq \boldsymbol{\theta}_0$, where $\hat{\boldsymbol{\theta}}_m$ is a maximum likelihood estimator based on m observations. However, according to the definition of $\hat{\boldsymbol{\theta}}_m$,

$$\frac{1}{m}L_m(\hat{\theta}_m) \geq \frac{1}{m}L_m(\theta_0) \quad \text{for each } \omega \in C \tag{B.4}$$

Also

$$\begin{aligned}
|\frac{1}{m}L_m(\hat{\boldsymbol{\theta}}_m) - \Pi(\boldsymbol{\theta}_*)| \leq & |\frac{1}{m}L_m(\hat{\boldsymbol{\theta}}_m) - \Pi_m(\hat{\boldsymbol{\theta}}_m)| \\
& + |\Pi_m(\hat{\boldsymbol{\theta}}_m) - \Pi_m(\boldsymbol{\theta}_*)| + |\Pi_m(\boldsymbol{\theta}_*) - \Pi(\boldsymbol{\theta}_*)| \\
\leq & \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\frac{1}{m}L_m(\boldsymbol{\theta}) - \Pi_m(\hat{\boldsymbol{\theta}}_m)| \\
& + |\Pi_m(\hat{\boldsymbol{\theta}}_m) - \Pi_m(\boldsymbol{\theta}_*)| + |\Pi_m(\boldsymbol{\theta}_*) - \Pi(\boldsymbol{\theta}_*)|
\end{aligned} \tag{B.5}$$

For $\omega \in C$, we take the limit as $m \to \infty$ in (B.5). The first term on the right hand side goes to zero by (B.3), the second term also goes to zero by lemma 2 and the fact $\hat{\theta}_m(\omega) \xrightarrow{a.s} \theta_*(\omega)$, and the last term also goes to zero by the assumption (c.4). Intersecting this set of convergence with C and taking limit in (B.4) as $m \to \infty$, $\frac{1}{m}L_m(\boldsymbol{\theta}_0) \xrightarrow{a.s} \Pi(\boldsymbol{\theta}_0)$ by the strong law of large numbers so we obtain $\Pi(\boldsymbol{\theta}_0) \leq \Pi(\boldsymbol{\theta}_*)$ on a set of positive probability. This contradicts assumption (c.4) which states that $\boldsymbol{\theta}_0 = arg\ max\ \Pi(\boldsymbol{\theta})$, is unique. Thus

$$\hat{\boldsymbol{\theta}} \xrightarrow{a.s} \boldsymbol{\theta}_0$$

(C) To establish the asymptotic normality of $\hat{\boldsymbol{\theta}}$, we now expand the product $1/\sqrt{n}$ times the gradient of the log-likelihood function

$$\frac{1}{\sqrt{n}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1}{\sqrt{n}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad \text{(B.6)}$$

where $\boldsymbol{\theta}_* = c_n \hat{\boldsymbol{\theta}} + (1-c_n)\boldsymbol{\theta}_0, c_n \in (0,1)$ for $n \geq 1$. Since $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$ and $\hat{\boldsymbol{\theta}}$ is a strongly consistenct estimator of $\boldsymbol{\theta}_0$, the left hand side of (B.6) goes to zero in probability because $\frac{1}{\sqrt{n}} \frac{\partial_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0}$ at the maximum. Consequently,

$$\frac{1}{\sqrt{n}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{0}$$

To establish the asymptotic normality of $\frac{1}{\sqrt{n}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, we now check conditions of Lemma 4.3 in Yeo and JohnsonYeo and Johnson (2000). From (B.1) and Lemma A.1, for

$$\begin{aligned} \left|\frac{\partial l_t(\boldsymbol{\theta})}{\partial \alpha_0}\right| &= \left|\tfrac{1}{2h_t} \tfrac{\partial h_t}{\partial \alpha_0} \left(\tfrac{(y_t-\mu)^2}{h_t} - 1\right)\right| \\ &\leq \tfrac{1}{c}\left(\tfrac{2(y_t^2+r^2)}{c} + 1\right). \end{aligned}$$

$$\begin{aligned} \left|\frac{\partial l_t(\boldsymbol{\theta})}{\partial \alpha_1}\right| &= \left|\tfrac{1}{2h_t} \tfrac{\partial h_t}{\partial \alpha_1} \left(\tfrac{(y_t-\mu)^2}{h_t} - 1\right)\right| \\ &\leq \tfrac{1}{\delta}\left(\tfrac{2(y_t^2+r^2)}{c} + 1\right) \end{aligned}$$

$$\begin{aligned} \left|\frac{\partial l_t(\boldsymbol{\theta})}{\partial \beta_1}\right| &= \left|\tfrac{1}{2h_t} \tfrac{\partial h_t}{\partial \beta_1} \left(\tfrac{(y_t-\mu)^2}{h_t} - 1\right)\right| \\ &\leq \tfrac{t}{\delta}\left(\tfrac{2(y_t^2+r^2)}{c} + 1\right) \end{aligned}$$

$$\left|\frac{\partial l_t(\boldsymbol{\theta})}{\partial \mu}\right| = \left|\frac{1}{2h_t}\frac{\partial h_t}{\partial \mu}\left(\frac{(y_t-\mu)^2}{h_t}-1\right)+\frac{y_t-\mu}{h_t}\right|$$
$$\leq \frac{2\left\{2\left(\psi_\mu^{(1)}(b,y_t-r)^2+\psi_\mu^{(1)}(a,y_t+r)^2\right)d+(1-\delta)\right\}}{c}$$
$$\times\left(\frac{2(y_t^2+r^2)}{c}+1\right)+\frac{y_t+r}{c}$$

$$\left|\frac{\partial l_t(\boldsymbol{\theta})}{\partial \lambda}\right| = \left|\frac{1}{2h_t}\frac{\partial h_t}{\partial \lambda}\left(\frac{(y_t-\mu)^2}{h_t}-1\right)+\frac{y_t-\mu}{h_t}\right|$$
$$\leq \frac{2\left\{2\left(\psi^{(1)}(a,y_t-r)^2+\psi^{(1)}(b,y_t+r)^2\right)d+(1-\delta)\right\}}{c}$$
$$\times\left(\frac{2(y_t^2+r^2)}{c}+1\right)$$

Next (c.6) implies that $\sup_i E\left|\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|^3 \infty$ so that for any $c \in R^5$

$$\sum_{i=1}^{n} E\left|c'\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|^3 / n^{3/2} \to 0 \ \text{ as } \ n \to \infty$$

This conclusion and (c.6)-(c.9) allow us to apply Hoadley's Lemma 3 to $\frac{1}{\sqrt{n}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. We obtain

$$\frac{1}{\sqrt{n}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{d} N_5(\mathbf{0}, B(\boldsymbol{\theta}_0))$$

According to Lemma A.2, $\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$ is continuous in $(\boldsymbol{\theta}', \boldsymbol{Y})$ and is equicontinuous in $\boldsymbol{\theta}$ for $y_i \in S_M$ with $S_M = [-M, M]$. Using Lemma A.3, it is easy to show that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \leq D(\boldsymbol{Y})$ where D are defined in Lemma A.2 and $\sup_i E[D^2(\boldsymbol{Y})]$ is finite according to assumption (c.6). Thus, applying Lemma 4.1 in Yeo and JohnsonYeo and Johnson (2000), we conclude that with probability one,

$$\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\| -\frac{1}{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} - A_n(\theta)\right\| = 0 \qquad \text{(B.7)}$$

Also, the difference

$$\begin{aligned}
\left\| -\frac{1}{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} - A_n(\boldsymbol{\theta}_0)\right\| \leq & \left\| -\frac{1}{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} - A_n(\boldsymbol{\theta}_*)\right\| \\
& + \left\| A_n(\boldsymbol{\theta}_*) - A_n(\boldsymbol{\theta}_0)\right\| + \left\| A_n(\boldsymbol{\theta}_0) - A(\boldsymbol{\theta}_0)\right\| \\
\leq & \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| -\frac{1}{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - A_n(\boldsymbol{\theta})\right\| \\
& + \left\| A_n(\boldsymbol{\theta}_*) - A_n(\boldsymbol{\theta}_0)\right\| + \left\| A_n(\boldsymbol{\theta}_0) - A(\boldsymbol{\theta}_0)\right\|
\end{aligned} \tag{B.8}$$

For each ω in the almost sure set where (B.7) holds and $\hat{\boldsymbol{\theta}}_n(\omega) \xrightarrow{a.s} \boldsymbol{\theta}_n$, we take the limit as $n \to \infty$ in (B.8). The first term on the right hand side goes to zero by (B.7), the second term goes to zero by Lemma 4.2 in Yeo and JohnsonYeo and Johnson (2000) and the fact that $\hat{\boldsymbol{\theta}}(\omega) \xrightarrow{a.s} \boldsymbol{\theta}_0$, and the last term also goes to zero by assumption (c.9). This is, $-\frac{1}{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} \xrightarrow{a.s} A(\boldsymbol{\theta}_0)$. By Slutsky's theorem, we conclude that

$$\sqrt{n}\Big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\Big) \xrightarrow{d} N_5\Big(\mathbf{0}, A(\boldsymbol{\theta}_0)^{-1}B(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)^{-1}\Big)$$

□

Proof of Theorem 4.1. Applying the continuous mapping theorem to theorem 1, we obtain following result.

(1) using (b) and (e)

$$
\begin{aligned}
T\big(\tilde{\Pi}_{21} - \Pi_{21}\big) &= \Big(\frac{1}{T}(J_2'\boldsymbol{Z}\boldsymbol{Z}'J_2 \otimes \tilde{A}_z'\hat{\Omega}_\epsilon^{-1}\tilde{A}_z)\frac{1}{T}\Big)^{-1}\frac{1}{T}vec\big(\tilde{A}_z'\hat{\Omega}_\epsilon^{-1}\boldsymbol{\epsilon}\boldsymbol{Z}'J_2\big) \\
&\xrightarrow{d} \Big((\Psi_{22}^z\Omega_{\epsilon_2}^{1/2}\int_0^1 B_{d_z}(\boldsymbol{u})B_{d_z}'(\boldsymbol{u})d\boldsymbol{u}\Psi_{22}^{z'} \otimes A_z'\Omega_\epsilon^{-1}A_z\Big)^{-1} \\
&\times vec\Big(A_z'\Omega_\epsilon^{-1}\Omega_\epsilon^{1/2}\big(\int_0^1 B_{d_z}(\boldsymbol{u})dB_{m_z}'(\boldsymbol{u})\big)'\Omega_{\epsilon_2}^{1/2}\Psi_{22}^{z'}\Big) \\
&= (\mathcal{R}_{zz} \otimes A_z'\Omega_\epsilon^{-1}A_z)^{-1}vec(A_z'\Omega_\epsilon^{-1}\mathcal{B}_{z\epsilon}')
\end{aligned}
$$

(2) using (g) and (i)

$$
\begin{aligned}
T^{1/2}\big(\tilde{\Pi}_{22} - \Pi_{22}\big) &= \Big(\frac{1}{T}(\hat{\boldsymbol{\omega}}\hat{\boldsymbol{\omega}}' \otimes \hat{\Omega}_\epsilon^{-1})\Big)^{-1}\frac{1}{\sqrt{T}}vec\big(\hat{\Omega}_\epsilon^{-1}\boldsymbol{\epsilon}\hat{\boldsymbol{\omega}}'\big) \\
&\xrightarrow{d} \big(\Omega_\omega \otimes \Omega_\epsilon^{-1})^{-1}N(0, \Omega_\omega \otimes \Omega_\epsilon^{-1})
\end{aligned}
$$

(3) using (a) and (d)

$$
\begin{aligned}
T\big(\tilde{\Pi}_{11} - \Pi_{11}\big) &= \Big(\frac{1}{T}(J_1'\boldsymbol{X}\boldsymbol{X}'J_1 \otimes \tilde{A}'\hat{\Omega}_e^{-1}\tilde{A})\frac{1}{T}\Big)^{-1}\frac{1}{T}vec\big(\tilde{A}'\hat{\Omega}_e^{-1}\boldsymbol{e}\boldsymbol{X}'J_1\big) \\
&\rightarrow^{d} \Big(\Psi_{22}\Omega_{a_2}^{1/2}\int_0^1 B_dB_d'd\boldsymbol{u}\Omega_{a_2}^{1/2}\Psi_{22}' \otimes A'\Omega_e^{-1}A\Big)^{-1} \\
&\times vec\Big(A'\Omega_e^{-1}\Omega_e^{1/2}\big(\int_0^1 B_d((u))dB_{m_y}'(\boldsymbol{u})\big)'\Omega_{a_2}^{1/2}\Psi_{22}'\Big) \\
&= (\mathcal{R}_{xx} \otimes A'\Omega_e^{-1}A)^{-1}vec(A'\Omega_e^{-1}\mathcal{B}_{ye}')
\end{aligned}
$$

(4) using (h) and (j)

$$
\begin{aligned}
T^{1/2}\big(\tilde{\Pi}_{12} - \Pi_{12}\big) &= \big(\frac{1}{T}(\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}' \otimes \hat{\Omega}_e^{-1})\big)^{-1}\frac{1}{\sqrt{T}}vec\big(\hat{\Omega}_e^{-1}\boldsymbol{e}\hat{\boldsymbol{\eta}}'\big) \\
&\xrightarrow{d} \big(\Omega_\eta \otimes \Omega_e^{-1}\big)^{-1}N(0, \Omega_\eta \otimes \Omega_e^{-1})
\end{aligned}
$$

□

Bibliography

Sung K Ahn and Gregory C Reinsel. Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85(411):813–823, 1990.

Sung K Ahn, Hanwoom Hong, and Sinsup Cho. Estimation of cointegrated models with exogenous variables. *Journal of Statistical Computation and Simulation*, 85(10):1945–1962, 2015.

David E Allen, Michael McAleer, and Marcel Scharth. Asymmetric realized volatility risk. *Journal of Risk and Financial Management*, 7(2):80–109, 2014.

DF Andrews, R Gnanadesikan, and JL Warner. Transformations of multivariate data. *Biometrics*, pages 825–840, 1971.

Anthony C Atkinson and Luis R Pericchi. Grouped likelihood for the shifted power transformation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 473–482, 1991.

Anthony Curtis Atkinson. *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis.* Number 519.536 A875. Clarendon Press, 1985.

Anil K Bera and Matthew L Higgins. Arch models: properties, estimation and testing. *Journal of economic surveys*, 7(4):305–366, 1993.

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

GEP Box and DR Cox. An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, 77(377):209–210, 1982.

GP Box and DRJ Cox. Box-cox commutation. royal statistical. *Society Series B*, 21(3):211–243, 1964.

Ralf Brüggemann and Helmut Lütkepohl. Practical problems with reduced-rank ml estimators for cointegration parameters and a simple alternative*. *Oxford Bulletin of Economics and Statistics*, 67(5):673–690, 2005.

Zhuanxin Ding, Clive WJ Granger, and Robert F Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

Norman R Draper and David R Cox. On distributions and their transformation to normality. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 472–476, 1969.

Robert F Engle. Estimates of the variance of us inflation based upon the arch model. *Journal of Money, Credit and Banking*, 15(3):286–301, 1983.

Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

Robert F Engle, David F Hendry, and Jean-Francois Richard. Exogeneity. *Econometrica: Journal of the Econometric Society*, pages 277–304, 1983.

Carmen Fernández and Mark FJ Steel. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441): 359–371, 1998.

Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.

Clive WJ Granger. Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1):121–130, 1981.

Alastair R Hall et al. *Generalized method of moments*. Oxford University Press Oxford, 2005.

James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

Ingrid Harbo, Søren Johansen, Bent Nielsen, and Anders Rahbek. Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics*, 16(4):388–399, 1998.

Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.

David A Hsieh. Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics*, 7(3):307–317, 1989.

John Hunter. Cointegrating exogeneity. *Economics Letters*, 34(1):33–35, 1990.

Guido Imbens, Phillip Johnson, and Richard H Spady. Information theoretic approaches to inference in moment condition models, 1995.

Søren Johansen. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2):231–254, 1988.

Søren Johansen. Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of econometrics*, 52(3):389–402, 1992.

JA John and NR Draper. An alternative family of transformations. *Applied Statistics*, pages 190–197, 1980.

Yuichi Kitamura and Peter CB Phillips. Fully modified iv, give and gmm estimation with possibly non-stationary regressors and instruments. *Journal of Econometrics*, 80(1):85–123, 1997.

FR Kleibergen. Reduced rank regression using generalized method of moments estimators with extensions to structural breaks in cointegration models. Technical report, Econometric Institute Research Papers, 1997.

Witold AJ Kosmala. *Introductory Mathematical Analysis*. Wm. C. Brown, 1995.

BFJ Manly. Exponential data transformations. *The Statistician*, pages 37–42, 1976.

László Mátyás. *Generalized method of moments estimation*, volume 5. Cambridge University Press, 1999.

Rocco Mosconi and Carlo Giannini. Non-causality in cointegrated systems: representation estimation and testing. *Oxford Bulletin of Economics and Statistics*, 54(3):399–417, 1992.

Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

Daniel B Nelson. Filtering and forecasting with misspecified arch models i: Getting the right variance with the wrong model. *Journal of Econometrics*, 52(1):61–90, 1992.

Suk K Park, Sung K Ahn, and Sinsup Cho. Generalized method of moments estimation for cointegrated vector autoregressive models. *Computational Statistics & Data Analysis*, 55(9):2605–2618, 2011.

M Hashem Pesaran, Yongcheol Shin, and Richard J Smith. Structural analysis of vector error correction models with exogenous i (1) variables. *Journal of Econometrics*, 97(2):293–343, 2000.

Peter CB Phillips. Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. *Econometrica: Journal of the Econometric Society*, pages 73–93, 1994.

Jacqueline Pradel and Christophe Rault. Exogeneity in vector error correction models with purely exogenous long-run paths. *Oxford Bulletin of Economics and Statistics*, 65(5):629–653, 2003.

Carmela E Quintos. Analysis of cointegration vectors using the gmm approach. *Journal of econometrics*, 85(1):155–188, 1998.

Jose A Scheinkman and Blake LeBaron. Nonlinear dynamics and stock returns. *Journal of Business*, pages 311–337, 1989.

James H Stock. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica: Journal of the Econometric Society*, pages 1035–1056, 1987.

John W Tukey. On the comparative anatomy of transformations. *The Annals of Mathematical Statistics*, pages 602–632, 1957.

Santiago Velilla. A note on the multivariate box—cox transformation to normality. *Statistics & probability letters*, 17(4):259–263, 1993.

Santiago Velilla. Diagnostics and robust estimation in multivariate data transformations. *Journal of the American Statistical Association*, 90(431): 945–951, 1995.

In-Kwon Yeo and Richard A Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 2000.

Jean-Michel Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955, 1994.

국문초록

본 논문에서는 평균 및 분산이 비표준형태인 시계열모형의 통계적 추론에 대하여 고려하였다. 첫째로 비대칭적인 지렛대 효과를 표현하기 위해서 과거 조건분산과 여-존슨 변환을 통해 변환된 잔차들로 이뤄진 새로운 형태의 조건부 이분산 모형을 연구하였다. 모형의 최대가능도 추정량이 일치성과 점근적 정규성을 만족하는 것을 증명하였으며, 실제 자료 분석을 통해 기존의 조건부 이분산 모형들과의 성능을 비교하였다.

둘째로 외생성 변수가 포함된 공적분 차수가 1인 벡터 공적분 모형을 일반화 적률 추정방법을 이용하여 모수를 추정하는 연구를 하였다. 외생성 변수가 공적분 관계가 있을때 최대가능도추정과 최소제곱추정을 고려한 기존 연구에서 가정한 모형을 배경으로 반복적인 일반화 적률 추정을 고려하였다. 일반화 적률 추정의 점근적특성을 유도하였으며, 몬테카를로 시뮬레이션을 통해 추정량의 유한 표본에서의 특징을 표현하였다.

주요어 : 비대칭 조건부 이분산모형, 여-존슨 변환, 지렛대 효과, 공적분, 일반화 적률추정방법, 외생변수

학　번 : 2009 – 20254