GENERALIZED MULTIPLICATIVE ERROR MODELS:
ASYMPTOTIC INFERENCE AND EMPIRICAL ANALYSIS

A DISSERTATION IN
Mathematics
and
Economics

Presented to the Faculty of the University
of Missouri-Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
QIAN LI

B.S., Sichuan University, 2009
M.S., University of Missouri-Kansas City, 2012

Kansas City, Missouri
2015

GENERALIZED MULTIPLICATIVE ERROR MODELS WITH ASYMPTOTIC INFERENCE AND EMPIRICAL ANALYSIS

Qian Li, Candidate for the Doctor of Philosophy Degree

University of Missouri-Kansas City, 2015

ABSTRACT

This dissertation consists of two parts. The first part focuses on extended Multiplicative Error Models (MEM) that include two extreme cases for nonnegative series. These extreme cases are common phenomena in high-frequency financial time series. The Location MEM(p,q) model incorporates a location parameter so that the series are required to have positive lower bounds. The estimator for the location parameter turns out to be the minimum of all the observations and is shown to be consistent. The second case captures the nontrivial fraction of zero outcomes feature in a series and combines a so-called Zero-Augmented general F distribution with linear MEM(p,q). Under certain strict stationary and moment conditions, we establish a consistency and asymptotic normality of the semiparametric estimation for these two new models.

The second part of this dissertation examines the differences and similarities between trades in the home market and trades in the foreign market of cross-listed stocks. We exploit the multiplicative framework to model trading duration, volume per trade and price volatility for Canadian shares that are cross-listed in the New York Stock Exchange (NYSE) and the Toronto Stock Exchange (TSX). We explore the clustering effect, interaction between trading variables, and the time needed for price equilibrium after a perturbation for each market. The clustering effect is studied through the use of univariate MEM(1,1) on each variable, while the interactions among duration, volume and price volatility are captured by a multivariate system of MEM(p,q). After estimating these models by a standard QMLE procedure, we exploit the Impulse Response function to compute the calendar time for a perturbation in these variables to be absorbed into price variance, and use common statistical tests to identify the difference between the two markets in each aspect. These differences are of considerable interest to traders, stock exchanges and policy makers.

The undersigned, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Generalized Multiplicative Error Models: Asymptotic Inference and Empirical Analysis" presented by Qian Li, candidate for the Doctor of Philosophy degree, and certify that in their opinion it is worthy of acceptance.

<u>Supervisory Committee</u>

Yong Zeng, Ph.D., Committee Chair
Department of Mathematics and Statistics

Kamel Rekab, Ph.D.
Department of Mathematics and Statistics

David Spade, Ph.D.
Department of Mathematics and Statistics

Peter Eaton, Ph.D.
Department of Economics

Maria Pacurar, Ph.D.
Department of Economics

Erik Olsen, Ph.D.
Department of Economics

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGMENTS

I am grateful to Dr. Zeng for his sincere encouragement, support and valuable guidance for this thesis. I would like to thank Dr. Pacurar for providing support and sharing expertise for the second part. I would like to show gratitude to my husband and our parents for their support in the face of stress and challenge.

PART 1

CHAPTER 1

INTRODUCTION

A common model employed in determining financial time-varying volatility is the AutoRegressive Conditional Heteroscedasticiy (ARCH) model, as proposed by Engle (1982), for which he was awarded the 2003 Nobel Memorial Prize in Economic Sciences. It was later generalized as the GARCH (G=Generalized) model in Bollerslev (1986). Following the publication of ARCH, a variety of structure specifications were applied to GARCH, such as Multiplicative GARCH by Geweke (1986), Exponential GARCH by Nelson (1991), Nonlinear Asymmetric GARCH by Engle and Ng (1993) and Threshold GARCH (TGARCH) by Zakoian (1994).

There have been several extensions of GARCH-type models for analysis in financial data. One of the most notable examples is the Autoregressive Conditional Duration model (ACD) of Engle and Russell (1998). It models irregularly-spaced transaction data, via parameterization of the conditional distribution by focusing on time intervals (or durations) between events. In fact, the ACD model is isomorphic to the GARCH model through the use of the square root of duration as dependent variable. Engle (2002b) generalizes ACD as the Multiplicative Error Model (MEM), which focuses on non-negativity and allows the entry of predetermined variables into the model. The MEM can be applied to a wide range of non-negative variables in finance, such as volume of shares traded, the daily high-low range of price, and the ask-bid spread.

Motivation

In many empirical analyses concerning financial data, diurnal adjustment is a technique used in data preparation for modeling financial variables including duration and price volatility (see Engle and Russell (1998), Hautsch (2002) and Manganelli (2005)). With the assistance of cubic spline or piecewise regression smoothing, diurnal adjustment removes the intradaily seasonality in the data and provides a general daily pattern for these variables. Although it is reasonable to apply this technique to many variables, signs of a significant daily pattern have not been found in variables like volume per trade or cumulative trading volume. If typical phenomena remain present in these variables and this variation cannot be captured by a seasonality factor, it becomes necessary to extend the financial time-series model.

The first feature of a non-negative series is the existence of a non-zero lower bound. Such a time series can be used to model volume per trade, for example. However, the multiplicative error structure lacks such lower bounds, and this leads to a poor empirical fit. This inspired us to add a parameter to the existing model so that a 'shift' in the lower bound from zero to any positive number is allowed. Since this added parameter plays a role analogous to a 'location' parameter in a density function, we call it the location parameter and the extended model the Location MEM. Following existing analysis in GARCH, we are able to conduct asymptotic analysis on Location MEM.

Another common feature of a non-negative process is that its minimum may be exactly zero and the proportion of zero outcomes is nontrivial (significantly larger than zero). As specified in Tsay (2010), the density of the innovation term in a standard MEM (or ACD) is usually assumed to be Exponential, Weibull or Generalized Gamma (GG), and since the corresponding log likelihood functions exclude zero outcomes, a series is strictly positive when it follows a Weibull-MEM or GG-MEM. Although the presence of zero outcomes is allowed in Exponential-MEM, the density function implies that the proportion of zeroes must be trivial. In a word, the occurrence of a high proportion of zeroes cannot be captured by any of these models. Distributional misspecification will produce inconsistency, inefficiency and lack of asymptotic distribution in the estimates. To address these concerns, Hautsch et al. (2014) introduce a so-called Zero-Augmented (ZA) distribution, which is a discrete-continuous mixture distribution with a clustering of zeroes. This ZA distribution decomposes into a point-mass at zero and a continuous distribution for the positive values. Furthermore, they take the Generalized F (GF) as the continuous part of the density, so that the Exponential, Weibull, GG and Log-Logistic distributions all apply.

The other part of our research is the application of a multivariate MEM in microstructure analysis. In the study of market microstructure for high-frequency trading, duration, volume, spread and return are typical variables of interest. Each of these random variables is associated with the arrival time of a trade and is usually called a mark. Considerable research has been conducted in the econometrics modeling of a mark. One of the most notable models is the vector MEM (vMEM) that handles irregular duration between two trades and other variables related to the arrival time. In fact, before the general vMEM was officially introduced by Engle

(2002b), Simone Manganelli applied this multivariate structure to model market microstructure of transactions, as published later in Manganelli (2005). He presented the results by comparing two groups of stocks categorized by trading intensity, i.e. frequently and non-frequently traded. In our work, we apply a similar framework to investigate the contrasts and similarities between two markets based on dynamic behavior of the same stock.

Contribution

Estimation for the aforementioned augmented models is of great interest to market participants, since the models play an important role in the analysis of market microstructure and successfully capture characteristics of trading volume with irregular time intervals. Existing analysis in financial econometrics indicates that consistency and asymptotic normality are two of the most crucial properties for estimates. The first part of this thesis follows the framework of Theorem 4.1.1 and 4.1.3 in Amemiya (1985) and uses up-to-date results about stationarity, ergodicity and existence of finite moments to develop asymptotic properties of estimates.

For the first augmented model, the key in estimation is to deal with the location parameter, which must be less than all the observations. Inspired by the exponential family density function, we have an intuition that the minimum of a series might be a consistent estimator for the location parameter. However, even if this conjecture can be verified, some other questions arise. If the location parameter is estimated by the minimum, how do we develop estimators for other parameters? Since the location parameter can be estimated without a likelihood function, shall we separate it from the others in further asymptotic analysis? Fortunately, we found answers to these questions.

The second augmented model incorporates the ZAF distribution, as proposed by Hautsch et al. (2014). They empirically illustrate consistency and efficiency of the maximum likelihood estimation for mean equation parameters. They compare these results to exponential QML estimates in terms of these two properties. Their work provides evidence that MLE by ZAF density is superior to the standard QMLE in consistency and efficiency in the presence of a non-trivial proportion of zero outcomes. However, they only consider the model at order (1,1) and estimates of the conditional mean without investigating density parameters and the asymptotic distribution of the exact MLE. Hence, our task is to establish a complete asymptotic inference on such a ZAF model.

In the second part of this thesis, inspired by the sharp contrast between two groups of stocks (frequently and less-frequently traded) as discussed by Manganelli (2005), we investigate the role of trading location in determining a stock's dynamic behavior. Manganelli (2005) estimated duration, volume and price volatility with a trivariate MEM and examined the time needed to converge to price equilibrium after a perturbation in the market. The multiplicative error model used in his work is composed of a system of autoregressive equations allowing interaction between variables. Our work applies a similar framework to U.S.-listed Canadian stocks in the home market and foreign market and draws conclusions on the difference or similarity between these markets.

Overview

This thesis is constructed as follows. Chapter 2 gives a thorough review of existing research on MEM and related applications in financial markets. In Chapter 3 we specify Location MEM(1,1), expand two classical theorems related to extremum estimation, establish asymptotic inference on estimates, including the estimate of the location parameter, and apply this model to IBM volume per trade data to identify improvement in goodness-of-fit. Chapter 4 follows the approach in Chapter 3, generalizing the results for Location MEM(p,q). Specification for the (p,q) case is presented in two optional matrix equations, which are also employed in assumptions of stationarity, ergodicity and higher-order moments. The zero-augmented model is discussed in Chapter 5, with emphasis on error distribution of the GF and asymptotic properties of the corresponding exact MLE. Chapter 6 presents the contrast and similarities between trading of cross-listed Canadian stocks in New York and Toronto, by analyzing three aspects related to market microstructure.

CHAPTER 2

LITERATURE REVIEW

As mentioned in the previous chapter, MEM (or ACD) can be viewed as a GARCH model by taking the square root of the dependent variable in the multiplicative equation. A linear GARCH(p,q) model for a time series $\{y_t\}_{t=1}^{\infty}$ is defined as

$$\begin{aligned} y_t &= \sigma_t \xi_t \\ \sigma_t^2 &= \omega_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{2.1}$$

where ξ_t's are i.i.d., $E(\xi_t) = 0$ and $E(\xi_t^2) = 1$. Squaring both sides of the first equation in (2.1), the model turns out to be a linear MEM (p,q), that is

$$\begin{aligned} x_t &= \psi_t \eta_t \\ \psi_t &= \omega_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \eta_{t-j} \end{aligned} \tag{2.2}$$

Hence, most of the analysis on GARCH can be carried over to MEM. A variety of GARCH models have been successfully applied to financial data and investigated regarding mixing or moment properties, model extension and diagnostic tests in the past 30 years. These applications are summarized by Pacurar (2008) and Hautsch (2012). The following sections give an up-to-date review of existing research on GARCH and MEM.

Model Extension

There are a many number of extensions of the GARCH model. A Fractionally Integrated (FI) model, i.e. FIGARCH in Baillie et al. (1996) or FIACD in Jasiak (1998) is suggested for significant autocorrelations up to long lags in a process. A special case of FIACD is IACD, which is shown to be strictly stationary and ergodic under the same stationarity and ergodicity conditions that Bougerol and Picard (1992a) derived for the IGARCH model. Accounting for additional variables in GARCH requires the more flexible logarithmic expression proposed by Geweke (1986). This expression does not restrict nonzero parameters to be positive in the autoregressive equation defined as:

$$\ln \sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \alpha_i \ln \xi_{t-i}^2 + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 \tag{2.3}$$

Similarly, Bauwens and Giot (2000) propose a logarithmic structure for ACD model in the following two parameterizations, called Log-ACD$_1$ and Log-ACD$_2$, respectively:

$$\ln \psi_t = \omega_0 + \sum_{i=1}^{p} \alpha_i \ln \eta_{t-i} + \sum_{j=1}^{q} \beta_j \ln \psi_{t-j} \tag{2.4}$$

$$\ln \psi_t = \omega_0 + \sum_{i=1}^{p} \alpha_i \eta_{t-i} + \sum_{j=1}^{q} \beta_j \ln \psi_{t-j} \tag{2.5}$$

The Log-ACD$_2$ model is usually preferred, since it has the better-fitting autocorrelation function (ACF) in empirical analysis. Dufour and Engle (2000) later point out the drawback of this logarithmic structure in application to duration: overadjustment of the conditional mean equation when the duration is very short. They propose the Exponential ACD (EACD), which is similar to EGARCH (see Nelson (1991)), to incorporate an asymmetric news impact function kinked at $\eta_t = 1$ while specifying the conditional mean:

$$\ln \psi_t = \omega + \sum_{i=1}^{p} \alpha_i (\eta_{t-i} + \delta_i |\eta_{t-i} - 1|) + \sum_{j=1}^{q} \beta_j \ln \psi_{t-j} \tag{2.6}$$

In this specification, the impact of observed values on the conditional mean varies depending on whether the observed value is larger, i.e. $\eta_t > 1$.

Another class of asymmetric GARCH model (AGARCH) is also developed by Hentschel (1995) and then generalized for ACD in Fernandes and Grammig (2006), called Augmented ACD (AACD). These types of models allow asymmetric responses to small and large shocks, by applying a Box-Cox transformation and a kinked news impact curve. The specifications of these models at order (1,1) are

AGARCH:

$$\sigma_t^{\nu_1} = \omega + \alpha((\xi_{t-1} - b) + \delta|\xi_{t-1} - b|])^{\nu_2} + \beta \sigma_{t-1}^{\nu_1} \tag{2.7}$$

AACD:

$$\psi_t^{\nu_1} = \omega + \alpha \psi_{t-1}((\eta_{t-1} - b) + \delta|\eta_{t-1} - b|])^{\nu_2} + \beta \psi_{t-1}^{\nu_1} \tag{2.8}$$

The difference between the above two specifications is that σ^{ν_1} in AGARCH acts additively with a function of η_t, while the AACD is based on a multiplicative stochastic component. A more general asymmetric ACD model has been introduced by Hautsch (2006), which encompasses both additive and multiplicative impact of past shocks on the conditional mean.

An alternative approach to capture the nonlinearity illustrated by Engle and Russell (1998) are Regime-switching models, which incorporate different conditional means and error distributions corresponding to different regimes of observed data (see Zhang et al. (2001)). In particular, the Threshold ACD (TACD) model introduced by Zhang et al. (2001) can be seen as a generalization of Threshold GARCH (TGARCH) in Rabemananjara and Zakoian (1993), given by

$$\begin{aligned} x_t &= \psi_t \eta_t^{(k)}, \quad x_t \in R_k \\ \eta_t &= \omega^{(k)} + \sum_{i=1}^{p} \alpha_i^{(k)} x_{t-i} + \sum_{j=1}^{q} \beta_j^{(k)} \eta_{t-j}^{(k)} \end{aligned} \tag{2.9}$$

where $R_k = [r_{k-1}, r_k], k = 1, 2, \cdots, M,$ with $0 = r_0 < r_1 < \cdots < r_M = \infty$ as threshold values. Hereafter, different regimes in this model give more flexibility compared to the linear ACD model.

Recent research sheds light on other nonlinear specifications as well. Brownlees and Gallo (2011) incorporate seasonalities and trends into the conditional expected mean of MEM with a flexible deterministic component. A shrinkage type of estimation is used to jointly estimate the dynamic (parametric) and flexible components. Different methodologies are utilized to choose the amount of smoothing in such estimation. Saart et al. (2015) propose a semiparametric regression approach to specify the conditional mean equation. It allows for generalization of dynamics of both the conditional mean and the shape of the hazard function, which are the two most essential components. An iterative estimation algorithm is used in this model and asymptotic properties of the estimator are investigated.

Mixing, Moments and Asymptotic Properties

There has been a vast amount of research concentrating on stationarity, ergodicity, existence of moments and β-mixing for GARCH; for instance, Nelson (1990), Bougerol and Picard (1992b), Carrasco and Chen (2002) and Meitz and Saikkonen (2008). These properties are of great importance, because the existence of moments and stationary β-mixing verify how well theoretical models describe stylized facts and also support large sample statistical properties, such as consistency and asymptotic normality for nonparametric/semiparametric estimators. Carrasco and Chen (2002) summarize sufficient mixing and moment conditions for the generalized Random Coefficient Autoregressive (RCA) framework below, which includes both linear and nonlinear GARCH.

$$X_{t+1} = A(e_{t+1})X_t + B(e_{t+1}) \tag{2.10}$$

where $\{X_t\}$ is a R^m-valued process and $\{e_t\}$ is a R^p-valued i.i.d. sequence with absolutely continuous marginal probability distribution and independent of the sigma-field generated by $(X_0, \cdots, X_t)$. Building on results by Mokkadem (1990), they propose the keys to establish properties regarding stationarity, ergodicity, higher-order moments and β-mixing for this class of models. According to their conclusions, theoretical properties of X_t are determined by assumptions on e_t's moment and constraints on the polynomial function $A(x)$. In fact, specific forms of the conditions in Carrasco and Chen (2002) were already employed to develop asymptotic properties of estimators for various GARCH models (see e.g. Lee and Hansen (1994), Lumsdaine (1996), Ling and McAleer (2003), Jensen and Rahbek (2004a), Jensen and Rahbek (2004b) and Berkes et al. (2003)). Generalization of results in Carrasco and Chen (2002) for the ACD model are given in Meitz and Saikkonen (2008).

Assumption and Testing on the Innovation Term

The standardized innovation term in MEM is often assumed to be independently and identically distributed. As for the innovation term distribution, any density with zero mean or a positive support can be used for GARCH or MEM, respectively. The most convenient choice is the standard normal for GARCH, or equivalently the standard exponential for MEM, which ignores the true distribution and leads to a QMLE with ideal properties (see Lee and Hansen (1994), Engle and Russell (1998)). However, this i.i.d. assumption seems to be too restrictive and inappropriate for some financial data. Drost and Werker (2004) extend semiparametric techniques to an adequate model for durations allowing arbitrary dependencies between innovation terms. The nonparametric specification of these dependencies provides their model with flexibility and an efficiency gain compared to the exponential QML.

Due to the cost of efficiency in pseudo likelihood procedures, exact MLE on the basis of a true density assumption is still preferred in certain situations. The student-t and mixture-of-normal distributions are two plausible options for the innovation term in GARCH as discussed in Xu and Wirjanto (2010), while Weibull and Generalized Gamma are commonly used in MEM. Xu et al. (2011) apply the mixture-of-normal distribution approach to a stochastic conditional duration model, which not only captures various density shapes of the durations but also accommodates a richer dependence among innovation terms. Hautsch (2003) also provides a more general distribution, i.e. Generalized F based on three parameters a, m and η to nest Weibull for $m = 1$ and $\eta \to \infty$, Generalized Gamma for $\eta \to \infty$ and Log-Logistic for $m = \eta = 1$. This generalized F distribution is also adopted in the recent work of

Hautsch et al. (2014). There are more flexible distributions for MEM, for instance, mixture distributions with time-varying weights and log-normal distribution for the Log-ACD model (see Luca and Gallo (2008) and Allen et al. (2008)). Vuorenmaa (2009) generalizes the ACD model using a q-Weibull distribution for the innovation term, which outperforms the standard specification.

Given numerous alternative distributions, it is of substantial importance to examine the error term. The model is considered to be adequate if there is no evidence of dependence in the residuals. The most classical approach adopted to check remaining serial dependence is using Ljung-Box Q statistics, which is however less ideal for MEM because the test statistics do not have an asymptotic χ^2 distribution for some densities, as stressed by Pacurar (2008). Li and Mak (1994) and Li and Mak (2003) propose corrected test statistics for goodness-of-fit in the ACD model, which should be used when one needs exact inference about the dynamic properties of residuals. Additional residual examinations can be found in the work of Bauwens and Giot (2000), Ghysels et al. (2004) and Bauwens and Veredas (2004).

The tests discussed above only detect autocorrelations, but they do not examine moment conditions implied by the assumed distribution. In practice, moment conditions should be investigated to evaluate the goodness-of-fit. Engle and Russell (1998) propose a statistic to test excess dispersion of the residual based on an Exponential or Weibull distribution. However, this test was found to have a drawback compared to the nonparametric test proposed by Fernandes and Grammig (2005). This general nonparametric testing is performed on the distance between the estimated parametric density function and its non-parametric estimation.

Another cause for misspecification of a model is the density assumed for the error term. Although the QML approach produces consistent estimates, poor performance is found in finite samples. Thus, it is necessary to inspect density assumptions. Fernandes and Grammig (2005) introduce two tests for distribution of the error by the contrast between parametric and nonparametric estimates of the density and hazard rate function. These tests not only detect moment conditions as discussed before, but also inspect the distribution of the residuals.

A different evaluation method was developed by Diebold et al. (1997) based on the probability integral transform. They demonstrate that distribution of the sequence of probability integral transforms are i.i.d. Uniform(0,1) under the null hypothesis of correct model specification. Performance of the density forecasts are evaluated by testing the sequence against uniform distribution. However, both tests are based on the assumption of a right conditional mean parameterization. They do not provide identification whether a rejection is due to a false distributional assumption or a violation of the conditional mean specification. The validity of the conditional mean function can be examined using the generalized moment test proposed by Chen and Hsieh (2010), which also tests for independence and distributional misspecification. Saart and Gao (2012) develop a procedure to test the distribution of error terms in various ACD class models, including the one with semiparametric regression dynamics discussed in Saart et al. (2015).

Application to High-Frequency Data

Use of various MEMs can be widely seen in the empirical literature. Hautsch (2012) summarizes seven univariate MEM (or ACD) specifications applied to different types of financial duration, i.e. trade durations and \$ 0.05 mid-quote change durations for stocks traded in the NYSE. Other nonnegative financial series, for instance, realized volatility, number of trades and cumulative volumes in a fixed interval of time, can be modeled by a multiplicative error structure as well (see Jasiak (1998), Engle (2002b) and Hautsch (2003)). It is of substantial interest to model some of the above nonnegative variables simultaneously, especially allowing for interaction between them. A vMEM specified by Manganelli (2005) and Engle and Gallo (2006) provides a framework to do so when the multivariate series are synchronized in time. Hautsch and Jeleskovic (2008) apply such a joint model with a logarithmic specification to 1-minute squared returns, average trade sizes, number of trades and average trading costs. Correlations between all variables are revealed in their work.

As employed in most of the empirical illustrations of MEM, Schwarz Bayesian Information Criteria (BIC) based on QML estimation is the typical method to select lag orders. Nevertheless, validity of the conditional mean function is essential for the QML estimation, and tests for the remaining ACD effect in the residual should be conducted before lag order selection. The Ljung-Box Q-statistic is often used to detect any remaining autocorrelation in residuals with lag of 15 or 20. But it is proven to be less powerful than the following tests. Meitz and Teräsvirta (2006) suggest an approach to examine the adequacy of the ACD model, which is similar to the Lagrange Multiplier test for GARCH and shown to be equivalent to the Li and Mak (2003) test. Duchesne and Pacurar (2008) also construct an adequacy test with kernel estimator of the normalized spectral density of residuals. Hong and Lee (2011) later propose a generalized spectral derivative test for misspecification of the conditional expected mean, without assuming the standardized innovation term to be i.i.d. This test requires a $\sqrt{n}$-consistent estimator under the null hypothesis, which can be QMLE.

PART 2

CHAPTER 3

LOCATION MULTIPLICATIVE ERROR MODEL(1,1): ASYMPTOTIC PROPERTIES OF A MODIFIED QMLE

In this chapter, we introduce an extension to the linear MEM(1,1) and develop estimation for the conditional mean equation. There has been substantial research on GARCH type models that directly apply to MEM, such as local estimation for all parameters in stationary GARCH(1,1) in Lee and Hansen (1994), analysis on vector ARMA-GARCH with unit root restriction in Ling and McAleer (2003) and asymptotic properties of the estimator for (G)ARCH parameters in nonstationary ARCH(1) and GARCH(1,1) in Jensen and Rahbek (2004a) and Jensen and Rahbek (2004b). Apparently, this research is all based on Quasi Maximum Likelihood Estimation (QMLE), as well as requires certain assumptions about the stationarity of the process. Ling and McAleer (2003) extend asymptotic inference for all parameters in a more general model, vector ARMA-GARCH(p,q), but it requires assumptions of unit root and higher order moment of the unconditional error. Although Jensen and Rahbek (2004a) and Jensen and Rahbek (2004b) investigate asymptotic theories without requiring moment conditions and stationarity, the results in their work only apply to (G)ARCH parameters. Our work in this paper will follow the assumptions and methods adopted in Lee and Hansen (1994), which insures stationary distribution of the process, develops estimators for all parameters and allows the model to to be integrated or mildly-explosive. We do not consider any exogenous variable in the current case, because exogenous variables, especially those negatively correlated with the dependent variable, may cause more restrictions on parameters.

Because our goal is to estimate parameters in the extended MEM, which does not depend on the true density of disturbance, it is ideal to employ a quasi likelihood function without bringing in other parameters and assumptions on the correct density. For MEM, Engle (2002b) has clearly shown that a log likelihood function based on unit exponential disturbance can be interpreted as a quasi likelihood function. In fact, this function coincides with the quasi likelihood for GARCH models in literature such as Lumsdaine (1996) and Jensen and Rahbek (2004a). Hence, we will follow the idea of deriving estimators by a likelihood function of exponential error. In addition, the minimum of the observed data will be demonstrated to be a consistent estimator for the location parameter, under an assumption that we set and a theory in Nelson (1990). This estimator will be substituted into the quasi likelihood function and the QMLE for other parameters is the maximizer of this modified quasi likelihood function. Theorems in Amemiya (1985) pp. 106-111 will be extended to cover the modified QMLE, although the estimator for the location parameter is not asymptotically normal. Eventually, consistency and asymptotic normality will both be developed for the univariate Location MEM(1,1).

The rest of this chapter is organized as follows. Section 2 shows specification and assumptions for the extended model, as well as the connection between MEM(1,1) and the extended model. In Section 3, two theorems about the likelihood-based estimator in Amemiya (1985), which were used in the proofs of Theorem 1 and 3 in

Lee and Hansen (1994), are extended for a modified estimator. A consistent estimator for the location parameter will be verified in this section. Some lemmas in Lee and Hansen (1994) will be carried over to the current case in a subsection to demonstrate consistency and asymptotic normality. Section 4 provides simulation examples and empirical improvement using IBM trade volume. Outlines of proofs for consistency, asymptotic normality of the modified QMLE and some important lemmas are given in Appendix A.

The Model

Suppose that we observe a process $\{r_t\},\ t = 1, ..., n$ and $\theta_0 = (\mu_0, \omega_0, \alpha_0, \beta_0)'$ are the true values of parameters describing this process. A MEM (1,1) process without exogenous variables in the mean equation is usually written as:

$$r_t = h_{0t} z_t, z_t | \mathscr{F}_{t-1} \sim D(1, \phi^2), z_t \geq 0 \tag{3.1}$$

$$h_{0t} = \omega_0 + \alpha_0 r_{t-1} + \beta_0 h_{0t-1} \tag{3.2}$$

where $D(1, \phi^2)$ represents the distribution with mean 1 and variance ϕ^2. $\sqrt{h_{0t}}$ is also called the conditional scaling parameter, since the above process can be viewed as a GARCH(1,1) by taking the square root of both sides of equation (3.1).

Location MEM(1,1)

Now we consider an improvement on the above model by adding a constant to the right hand side of equation (3.1) and changing the ARCH term in equation (3.2). Then we have

$$r_t = \mu_0 + \epsilon_t,\ \epsilon_t = h_{0t} z_t, z_t | \mathscr{F}_{t-1} \sim D(1, \phi^2), z_t \geq 0 \tag{3.3}$$

$$h_{0t} = \omega_0(1-\beta_0) + \alpha_0(r_{t-1} - \mu_0) + \beta_0 h_{0t-1} \tag{3.4}$$

where $r_t - \mu_0 \geq 0$ for $t = 1, ..., n$. In other words, the upper bound for μ_0 is the minimum of r_t, denoted by $r_{n(1)}$, that is $\mu_0 \leq r_{n(1)}$. Since $\sqrt{h_{0t}}$ can be viewed as 'scale' , correspondingly, we name μ_0 as the location parameter. The model for the unknown parameters $\theta = (\mu, \omega, \alpha, \beta)'$ is

$$r_t = \mu + \ e_t, \mu \leq r_{n(1)} \tag{3.5}$$

$$h_t^* = \omega(1-\beta) + \alpha \ e_{t-1} + \beta \ h_{t-1}^*, \ \ h_1^* = \omega \tag{3.6}$$

For the observed sequence, we have

$$h_t^* = \omega + \alpha \sum_{k=0}^{t-2} \beta^k \ e_{t-1-k} \tag{3.7}$$

Analogous to the quasi-likelihood estimation of MEM in Engle (2002b) and other literature, the observed log likelihood function takes the form:

$$L_n^*(\theta) = \frac{1}{n} \sum_{t=1}^{n} l_t^*(\theta) \qquad l_t^*(\theta) = -\left(\ln h_t^*(\theta) + \frac{e_t}{h_t^*(\theta)} \right) \tag{3.8}$$

Here we ignore the distribution of z_t, and use the above log likelihood function to derive estimation, because any assumption about the density function will bring in additional parameters. Usually, QMLE is the maximizer of $L_n^*(\theta)$. However, as mentioned the upper bound for μ depends on the observed data and its sample size, hence, the parameter space for the extended model varies with observations. It would be problematic to locate a maximizer of the score function in such a space. Since Location MEM(1,1) can be reduced to MEM(1,1) at $\mu = \mu_0$, it occurred to us that a plausible solution is to separate μ from the other parameters in estimation. The first question that arises is whether there exists a consistent estimator for μ_0 without involving the other parameters. If we can find such an estimator, how do we derive an estimator for the remaining parameters and verify its econometrics properties using the above quasi-likelihood function? These questions will be answered in the following section.

Analogous to the corollary about the exponential ACD model in Engle and Russell (1998), certain assumptions about true innovation terms are necessary.

ASSUMPTION 3.1. *Suppose the following conditions are met.*

(1). z_t *is stationary and ergodic.*

(2). z_t *is nondegenerate.*

(3). $E(z_t^2|\mathscr{F}_{t-1}) < \infty$ *a.s.*

(4). $\sup_t E(\ln(\beta_0 + \alpha_0 z_t)|\mathscr{F}_{t-1}) < 0$*, a.s.*

(5). η_0 *is in the interior of* Θ^**.*

Note that condition (3) is actually stronger than necessary for consistency only. According to Lee and Hansen (1994), it is sufficient to establish local consistency for MEM(1,1), if there is some δ such that $E(z_t^{1+\delta}|\mathscr{F}_{t-1}) < \infty$. Existence of the second moment of z_t is just a prerequisite for asymptotic normality. For simplicity, we choose the stronger version of the moment condition in this paper. Condition (4), which is also required in proposition 3.1 below, not only assures the consistency of $r_{n(1)}$, but also serves as a sufficient condition for stationarity and ergodicity of h_{0t}. Although the analysis in this paper only focuses on stationary processes, we plan to investigate similar properties for a nonstationary case in the future, that is when $\sup_t E(\ln(\beta_0 + \alpha_0 z_t)|\mathscr{F}_{t-1}) \geq 0$, a.s. as in Jensen and Rahbek (2004a) and Jensen and Rahbek (2004b). When $\alpha_0 + \beta_0 \leq 1$, condition (4) is automatically satisfied due to Jensen's inequality. But it is not necessary to require $\alpha_0 + \beta_0 \leq 1$ in current case. Asymptotic properties of the local estimator can be established under condition (4), which allows integrated and mildly explosive cases.

Connection

An obvious relation between Location MEM and MEM is that Location MEM reduces to MEM when μ is zero. In fact, there is another transition from Location MEM(1,1) to MEM(1,1) with certain assumptions. Suppose r_t is a process described by equation (3.3). Define $x_t = \dfrac{\mu_0 + h_{0t} z_t}{\mu_0 + h_{0t}}$. It can be shown that $E(x_t|\mathscr{F}_{t-1}) = 1$ and x_t is stationary and ergodic. Let $\psi_{0t} = \mu_0 + h_{0t}$, then

$$\begin{aligned} r_t &= \psi_{0t} x_t \\ \psi_{0t} &= \omega_0' + \alpha_0 r_{t-1} + \beta_0 \psi_{0t-1} \end{aligned} \tag{3.9}$$

where $\omega_0' = \omega_0 + (1 - \alpha_0 - \beta_0)\mu_0$. Hence, Location MEM(1,1) can be transformed to MEM(1,1) with ARCH and GARCH parameters unchanged. In order to ensure positivity of ω_0' and distinguish it from ω, here we only consider $\alpha_0 + \beta_0 < 1$. Thus, Assumption 3.1 (1),(3),(4) and (5) are easily satisfied by x_t. However, the range of x_t is slightly different from that of z_t. When $\mu_0 \neq 0$, x_t is strictly positive and the exponential distribution is excluded from the alternative true densities. Lee and Hansen (1994) have demonstrated that no matter what density function the innovation term has, the log likelihood function in the form of unity exponential always produces a consistent and asymptotically normal estimation under certain conditions. Therefore, if all parts of Assumption 3.1 are satisfied by x_t, the model represented in (3.9) can still be estimated by QMLE regardless of the range of x_t. The only concern is whether condition (2) can be trivially met if μ_0 is too large. When $h_{0t}/\mu_0 \to 0$, $x_t \to 1$ a.s., that is x_t is degenerate. In that case, it is impossible to ensure consistency and asymptotic normality for the MEM(1,1). Therefore, Location MEM(1,1) can be correspondingly viewed as MEM(1,1) only if μ_0 is not too large compared to h_{0t}, which is difficult to evaluate in real data.

Asymptotic Properties for Modified QMLE

In order to estimate Location MEM(1,1) with QML as mentioned in subsection 3, the first task is to locate a consistent estimator of μ_0. The second question that arises is how to derive a consistent and asymptotically normal estimator for the remaining parameters. If consistency in estimation for μ_0 is established, is it valid to substitute the estimator of μ_0 for μ in likelihood function (3.8) and then maximize it over $\omega_0, \alpha_0, \beta_0$? Before we explore the asymptotic properties of the estimator for $\omega_0, \alpha_0, \beta_0$, the primary issue is to validate convergence results based on a score function composed from observed data, unknown parameters and a partial estimator.

A Consistent Estimator of the Location Parameter

If the observations are independently and identically distributed as Exponential, Weibull or Generalized Gamma, it is not hard to show that the minimum statistic of the sample converges to the location parameter in the density function. Inspired by this property, for a process described by equations (3.3) and (3.4), $r_{n(1)}$ might be a consistent estimator for μ_0. However, as mentioned we ignore the true distribution of z_t. The distribution of r_t cannot be specified due to the lack of an assumption on

z_t's density. Therefore, it is challenging to draw the following conclusion.

PROPOSITION 3.1. *Under Assumption 3.1,* $r_{n(1)} \rightarrow_p \mu_0$.

Estimating μ_0 so far has not involved any of the other parameters. For Location MEM(p,q), the convergence of $r_{n(1)}$ can also be validated under certain conditions of stationarity. Note that normality is not applicable to $r_{n(1)}$, because it has a lower bound of μ_0. Now that the consistency of the estimator for μ is demonstrated, we will move on to the modified QMLE for ω, α and β.

Extended Theorems of Extremum Estimators

Theorem 4.1.1 and 4.1.3 in Amemiya (1985) are examples of the most frequently-used estimation theorems, employed in both Lee and Hansen (1994) and Lumsdaine (1996). There are some other convergence theorems adopted in asymptotic likelihood-based inference, e.g., Lemma 1 of Jensen and Rahbek (2004a) and similar results stated in Lehmann (1999). As mentioned in the introduction, before discussing the extended model, we need to develop theorems about likelihood-based estimators by separating one parameter from the others and rewriting Theorem 4.1.1 and Theorem 4.1.3 in Amemiya (1985).

Suppose $Q_T(\underline{y}, \theta)$ is a measurable function for $\underline{y}, \theta$, where $\underline{y}$ is the observed data set with size T, θ is a K-vector of parameters, and Θ is the whole parameter space. The traditional extreme likelihood estimator is usually defined as $\hat{\theta}_n = \arg\max_{\theta \in \Theta} Q_T(\underline{y}, \theta)$. In Amemiya 1985, Theorem 4.1.1 and Theorem 4.1.3 prove consistency and asymptotic normality for $\hat{\theta}_n$. Now we want to expand these two theorems to a different case. Suppose $\Theta = \Theta_\mu \times \Theta_\eta$ is the whole space for a K-vector parameter

$\theta' = (\mu, \eta')$, where Θ_η is the (K-1)-dim subspace for η. Let $\hat{\mu}_T$ be an estimator for the true value μ_0.

THEOREM 3.2. *Suppose that that the following three conditions are satisfied:*
(A). The parameter subspace Θ_η is a compact subset of the Euclidean (K-1)-space.
(B). $Q_T(\underline{y}, (\mu, \eta))$, which is a function of the parameters and a T-vector $\underline{y}$, is continuous in $\theta \in \Theta$ for all $\underline{y}$ and is a measurable function of $\underline{y}$ for all $\theta \in \Theta$.
(C). $T^{-1}Q_T(\hat{\mu}_T, \eta)$ converges to $Q(\mu_0, \eta)$ in probability uniformly in $\eta \in \Theta_\eta$ as $T \to \infty$. $Q(\mu_0, \eta)$ attains a unique global maximum at η_0.
Define $\hat{\eta}_T = \arg\max_{\eta \in \Theta_\eta} Q_T(\hat{\mu}_T, \eta)$. Under the above assumptions, $\hat{\eta}_T \to_p \eta_0$.

THEOREM 3.3. *Suppose $\hat{\mu}_T \to_p \mu_0$ and that the following conditions hold:*
(A) For fixed μ, $\partial^2 Q_T(\mu, \eta)/\partial\eta\partial\eta'$ exists and continuous in an open convex neighborhood of η_0. $|\partial Q_T(\hat{\mu}_T, \eta_0)/\partial\eta - \partial Q_T(\mu_0, \eta_0)/\partial\eta| / \sqrt{T}$ converges to 0 in probability.
(B) $T^{-1}\partial^2 Q_T(\mu, \eta)/\partial\eta\partial\eta'|_{(\hat{\mu}_T, \eta_T^*)}$ converges to a finite nonsingular matrix $A(\theta_0) = \lim ET^{-1}\partial^2 Q_T/\partial\eta\partial\eta'|_{\theta_0}$ in probability for any sequence η_T^* s.t $\eta_T^* \to_p \eta_0$.
(C) $T^{-1/2}(\partial Q_T(\mu, \eta)/\partial\eta)|_{\theta_0} \to N[0, B(\theta_0)]$, where

$$B(\theta_0) = \lim ET^{-1}\left[\frac{\partial Q_T(\mu, \eta)}{\partial\eta}\bigg|_{\theta_0} \times \frac{\partial Q_T(\mu, \eta)}{\partial\eta'}\bigg|_{\theta_0}\right].$$

Then $\sqrt{T}(\hat{\eta}_T(\hat{\mu}_T) - \eta_0) \to N[0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}]$.

The key conditions for consistency are convergence of the observed likelihood function and its limit being uniquely maximized at the true values of parameters. For asymptotic normality, in addition to Theorem 4.1.3 in Amemiya (1985), we add one more condition to restrict the difference in the first gradient of likelihood, as presented in Theorem 3.3 (A). Analogous to Lee and Hansen (1994), these theorems are the foundation for the following analysis on the quasi likelihood estimator of the extended model.

Asymptotic Properties of The Modified QMLE

In order to utilize the above two theorems, we have to first specify and draw some conclusions on the limit of $L_n^*(r_{n(1)}, \eta)$ at first, which is an unobserved likelihood, defined as

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \qquad l_t(\theta) = -\left(\ln h_t(\theta) + \frac{e_t}{h_t(\theta)}\right)$$

where

$$h_t = \omega + \alpha \sum_{k=0}^{\infty} \beta^k e_{t-1-k} \tag{3.10}$$

Equation (3.10) indicates that $h_t(\theta_0) = h_{0t}$, because when $\theta = \theta_0$, $e_t = r_t - \mu_0 = \epsilon_t$. Conclusions on $L_n(\theta)$ and $L_n^*(r_{n(1)}, \eta)$ are all presented as lemmas in Appendix A.

In order to establish asymptotic properties, the parameter space is restricted to be: $\Theta = \{\eta : 0 \le \omega_l \le \omega \le \omega_u, 0 \le \alpha_l \le \alpha \le \alpha_u, 0 \le \beta_l \le \beta \le \beta_u < 1\}$, where $\eta = (\omega, \alpha, \beta)$. There is no need to discuss the estimator for μ, as the consistency of $r_{n(1)}$ is demonstrated in proposition 3.1 and asymptotic normality is not applicable to it. We will directly apply results in Lee and Hansen (1994) to the current model with μ fixed at μ_0. Similar to the analysis about GARCH (1,1) in that paper, we need to split the parameter space in the same way to bound h_t.

Let $\mathscr{R}_l = \mathscr{R}(K_l^{-1}\alpha_l) < 1$ and pick positive constants η_l and η_u, which satisfy $\eta_l < \beta_0(1 - \mathscr{R}_l^{1/6})$ and $\eta_u < \beta_0(1 - \mathscr{R}_0^{1/6})$. Define for $1 \le r \le 6$ the constants

$\beta_{rl} = \beta_0 \mathscr{R}_l^{1/r} + \eta_l < \beta_0,\ \beta_{ru} = \dfrac{\beta_0 - \eta_u}{\mathscr{R}_0^{1/r}} > \beta_0$, and the subspaces
$\Theta_r^l = \{\eta \in \Theta : \beta_{rl} \leq \beta \leq \beta_0\},\ \Theta_r^u = \{\eta \in \Theta : \beta_0 \leq \beta \leq \beta_{ru}\},$
$\Theta_r = \Theta_r^l \bigcup \Theta_r^u$. When $r = 2,\ \Theta_2 = \Theta_2^l \bigcup \Theta_2^u$.

Now we can define the modified QMLE and prove its local asymptotic properties.

DEFINITION 1. *Local modified QMLE:* $\hat{\eta}_n(r_{n(1)}) = \arg\max_{\eta \in \Theta_2} L_n^*(r_{n(1)}, \eta)$.

Let ∇_η denote the first order gradient w.r.t $\eta' = (\omega, \alpha, \beta)$. Set $\hat{G}_n(\theta) = -\frac{1}{n}\sum_{t=1}^{n} \nabla_\eta^2 l_t^*(\theta)$ and $\hat{C}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \nabla_\eta l_t^*(\theta) \nabla_\eta l_t^*(\theta)'$.

ASSUMPTION 3.2. $E|r_{n(1)} - \mu_0|^s = o(\frac{1}{\sqrt{n}})$ *for some* $s < 1$.

THEOREM 3.4. *Under Assumptions 3.1 and 3.2,*

(1) $\hat{\theta}_n \to_p \theta_0$.

(2) $\sqrt{n}(\hat{\eta}_n(r_{n(1)}) - \eta_0) \to_D N(0, V_0)$, *where* $V_0 = G_0^{-1} C_0 G_0^{-1}$, $C_0 = E(\nabla_\eta l_t(\theta_0) \nabla_\eta l_t(\theta_0)')$ *and* $G_0 = -E\nabla_\eta^2 l_t(\theta_0)$.

(3) $\hat{V}_n = \hat{G}_n^{-1}(\hat{\theta}_n) \hat{C}_n(\hat{\theta}_n) \hat{G}_n^{-1} \to_p V_0 = G_0^{-1} C_0 G_0^{-1}$.

Now let's restrict attention to nonintegrated processes. We can establish consistency of the global modified QMLE with the same assumptions.

DEFINITION 2. *Global modified QMLE:* $\tilde{\theta}_n = (r_{(1)}, \arg\max_{\eta \in \Theta} L_n^*(r_{(1)}, \eta))$.

THEOREM 3.5. *Under Assumption 3.1 and* $\alpha_0 + \beta_0 < 1, \tilde{\theta}_n \to_p \theta_0$.

PROPOSITION 3.6. *If* $\alpha_0 + \beta_0 < 1$, $L_n^*(\theta)$ *is maximized at* $\mu = r_{n(1)}$ *for any* *given* η.

Therefore, if we derive the global estimator for a nonintegrated Location MEM(1,1) by maximizing $L_n^*(\theta)$ over all parameters, the optimal solution coincides with the global modified QMLE. That is, $\tilde{\theta}_n$ is equivalent to $\arg\max_{\theta\in\Theta} L_n^*(\theta)$ when $\alpha_0 + \beta_0 < 1$.

Simulation and Empirical Results

In this section , we will illustrate asymptotic results in Theorem 3.4 as well as the relation between Location MEM(1,1) and MEM(1,1), by both simulation and real data. GARCH software with QMLE can be used to fit a positive series into either Location MEM or MEM, by taking the square root of the input variable and setting the mean to be zero. For Location MEM(1,1), the input is the observed data from which is subtracted the minimum of the sample.

Simulation

In the first step, we generate 500 Location MEM(1,1) data sets at sample size n=10000 and parameter values at $(\mu_0, \omega_0, \alpha_0, \beta_0) = (0.5, 0.03, 0.1, 0.85)$ for alternative innovation term distributions: Exponential(1), Weibull(2, $1/\Gamma(3/2)$) and Gamma(4,0.25). Asymptotic properties of modified QMLE for Location MEM(1,1) are confirmed by the estimates in table 4 and corresponding QQ plots of standardized estimates displayed in figure 4. The standardized estimate in each QQ plot is $\sqrt{n}(\hat{\eta}(r_{n(1)}) - \eta_0)$ standardized by $\hat{V}_n$ (defined in Theorem 3.4 (3)).

Second, in order to see model improvement from MEM(1,1) to Location MEM(1,1) and the transition between them, we generate 500 data sets at sample size 5000 and

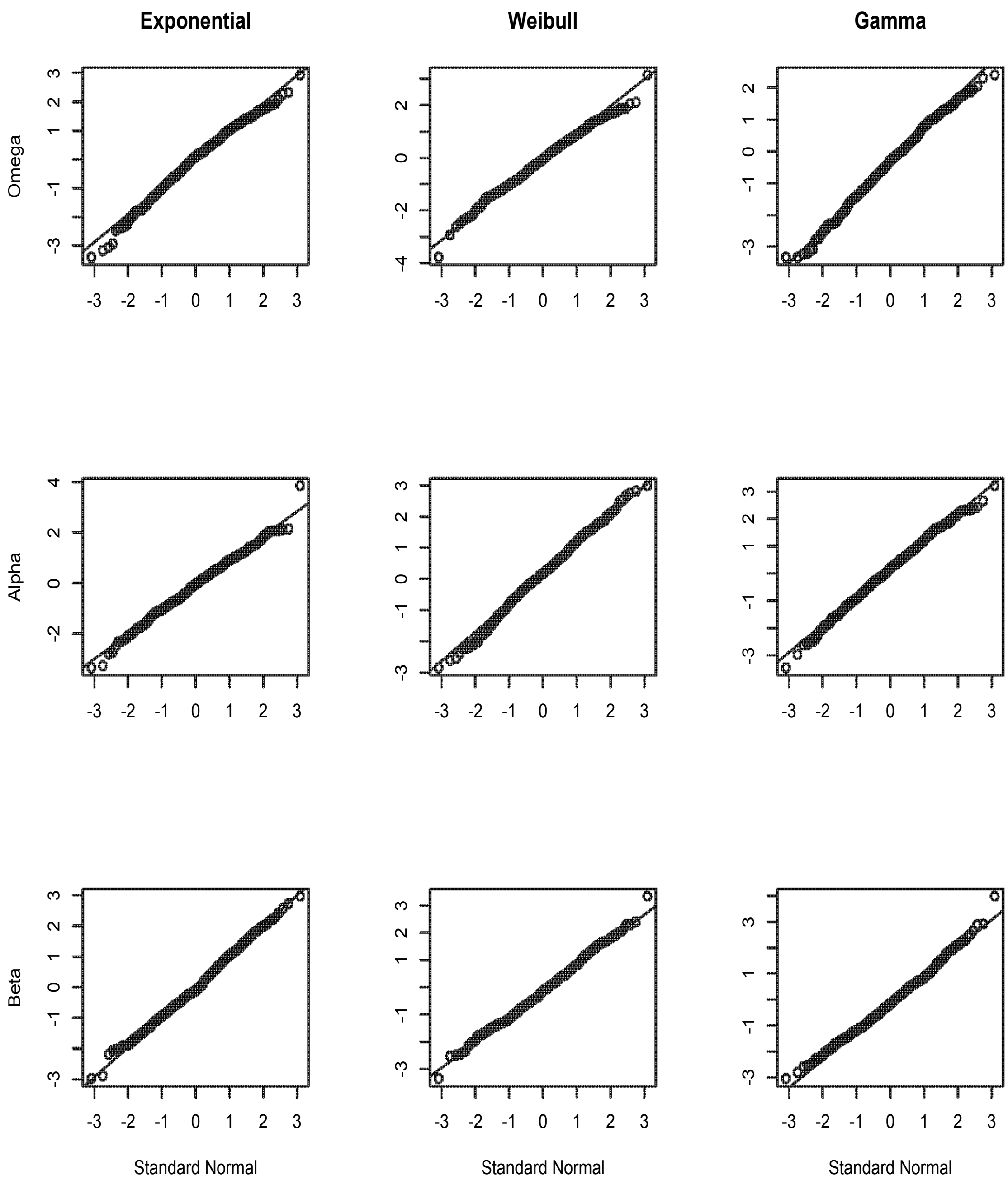

Figure 1. QQ Plot of standardized estimates for different distributions

Table 1. Modified QMLE of Location MEM(1,1) for Different Distributions.

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
True Value	0.03	0.1	0.85
Exponential			
Mean	0.030890	0.100333	0.84802
S.D.	0.005573	0.009727	0.016241
Weibull			
Mean	0.030661	0.101692	0.846497
S.D.	0.005223	0.008585	0.014517
Gamma			
Mean	0.029305	0.102111	0.845881
S.D.	0.0052	0.008652	0.015456

10000 respectively for the location parameter μ_0 at values 0.5 and 2. Since the true distribution has no impact on consistency and asymptotic normality as verified in the first step, we select Gamma(4,0.25) as the distribution of the rescaled innovations. The true values of the other parameters are still (0.03,0.1, 0.85). We use both Location MEM(1,1) and MEM(1,1) to fit each data set and compare the estimates. Although the data sets are all generated from Location MEM(1,1), they can be rewritten as MEM(1,1) according to section 3.

Tables 2 and 3 display the sample mean and standard deviation of estimates from 500 data sets at each sample size for different values of μ_0. QQ plots of the standardized estimates at $\mu_0 = 0.5$ are all similar to figure 4, no matter which model we employ. However, when $\mu_0 = 2$, it can be seen from figure 3 that asymptotic normality is slightly violated in MEM(1,1) for α, β.

Although parameter estimates are still consistent as shown in table 3, the covariance matrix V_0 is estimated with bias, which leads to a slight deviation from the standard normal in figure 3. When $\mu_0 = 8$, which is large compared to h_{0t}, the estimate by MEM is neither consistent nor asymptoticaly normal. Therefore, simulated data indicate that Location MEM(1,1) guarantees consistent and asymptotically normal estimates regardless of the value of μ_0. On the other hand, when μ_0 is not large, which can be evaluated by the minimum of observed data, MEM(1,1) can be a robust alternative for a Location MEM(1,1) process.

Model Improvement on IBM Trading Volume

In existing articles, the first application of MEM is to model trading duration in a high-frequency market, proposed as ACD by Engle and Russell (1998). Later, it was applied to other positive processes, such as volume, bid-ask spread and price return volatility. Hence, at first, we tried to find the difference between Location MEM(1,1) and MEM(1,1) by fitting IBM transaction duration data to each model. However, we saw only trivial improvement in log likelihood and the Ljung-Box test statistics brought by the location parameter. The trivial improvement is caused by the fact that the minimum of the observations is too small, since the ultimate transaction duration-one millisecond-is frequently reached and the average level of transaction

Table 2. $\mu = 0.5$

	Location MEM			MEM		
	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\omega}'$	$\hat{\alpha}$	$\hat{\beta}$
True Value	0.03	0.1	0.85	0.055	0.1	0.85
n=5000						
Mean	0.029073	0.101659	0.846894	0.056651	0.101152	0.847335
S.D.	0.005111	0.008916	0.015270	0.010198	0.009020	0.015535
n=10000						
Mean	0.029012	0.101074	0.848073	0.056028	0.100857	0.848192
S.D.	0.003647	0.006465	0.011098	0.007145	0.006501	0.011267

duration is much larger than 0.001 second for IBM, as mentioned in section 3.

Fortunately, we found that trading volumes have a relatively large minimum compared to duration. Furthermore, according to Manganelli (2005), high volume may increase price volatility in the next trade. Hence, volume at each trade is a key economic element to be molded and forecasted. In this section, we use IBM trading volume in the NYSE with the time span April 8th-12th, 2013, taken from Trade and Quotes(TAQ) database, to demonstrate the contrast between Location MEM(1,1) and MEM(1,1).

Applying these two models to IBM trading volume data, we found significant improvement in Location MEM(1,1) as shown in the following two tables. In the first step, intraday trend of trading volume is not removed yet. Contrast between two models is shown in table 5.

Next, we removed the daily pattern by smoothing with a piecewise regression function, as used by Engle and Russell (1998). We select knots with interval of 30 minutes from 10 AM to 4 PM to compute diurnal factors. The adjusted series is the original data divided by diurnal factors. Estimates for the deseasonalized data by two models are shown in Table 6. The minima of raw and adjusted data are 100 and 0.3524, which can be viewed as the estimates for location parameters. Ljung-Box statistics for Location MEM(1,1) in both tables indicate that z_t is i.i.d, which satisfies Assumption 3.1 (1) to (3). Condition (4) is trivially met, because $\alpha_0 + \beta_0$ is assumed to be less than 1. According to the theorems in previous sections, estimate by Location MEM(1,1) for either data set is asymptotically unbiased. If we round all estimates in this model to the second decimal and take them as true values, then $(\mu_0, \omega_0, \alpha_0, \beta_0)$ is (100, 1.41, 0.02, 0.97) for the raw series and (0.35, 0.03, 0.02, 0.92) for the adjusted series. Hence, by equation 3.9, true values for $(\omega_0', \alpha_0, \beta_0)$ in the corresponding MEM(1,1) should be (2.41, 0.02, 0.97) for raw data and (0.051, 0.02, 0.92) for deseasonalized data. However, results in the above two tables illustrate that the estimates by MEM(1,1) for both data sets are biased. Therefore, combined with improvement in log likelihood and Q-statistics, Location MEM is superior to MEM for order (1,1).

Table 3. $\mu = 2$

	Location MEM			MEM		
	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\omega}'$	$\hat{\alpha}$	$\hat{\beta}$
True Value	0.03	0.1	0.85	0.13	0.1	0.85
n=5000						
Mean	0.029073	0.101659	0.846894	0.134295	0.100922	0.847419
S.D.	0.005111	0.008916	0.015270	0.024812	0.009246	0.015920
n=10000						
Mean	0.029012	0.101074	0.848073	0.132734	0.100770	0.848170
S.D.	0.003647	0.006465	0.011098	0.017427	0.006619	0.011507

Table 4. $\mu = 8$

	Location MEM			MEM		
	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\omega}'$	$\hat{\alpha}$	$\hat{\beta}$
True Value	0.03	0.1	0.85	0.43	0.1	0.85
n=5000						
Mean	0.029073	0.101659	0.846894	0.000936	0.095014	0.904886
S.D.	0.005111	0.008916	0.015270	0.013216	0.010481	0.010588
n=10000						
Mean	0.029012	0.101074	0.848073	0.008649	0.095002	0.904
S.D.	0.003647	0.006465	0.011098	0.048859	0.007256	0.009527

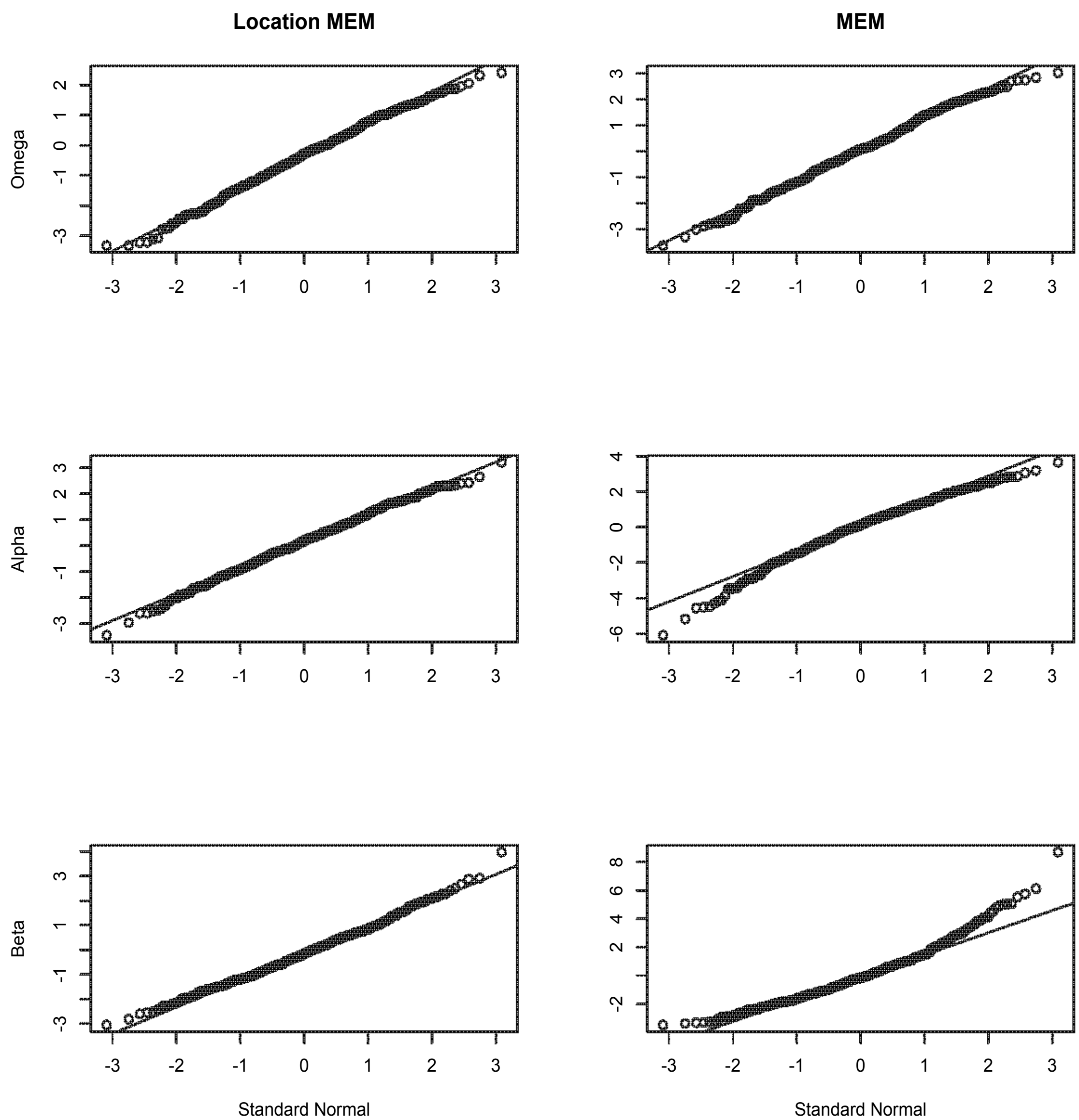

Figure 2. QQ Plot of standardized estimates for $\mu_0 = 2$ at order (1,1)

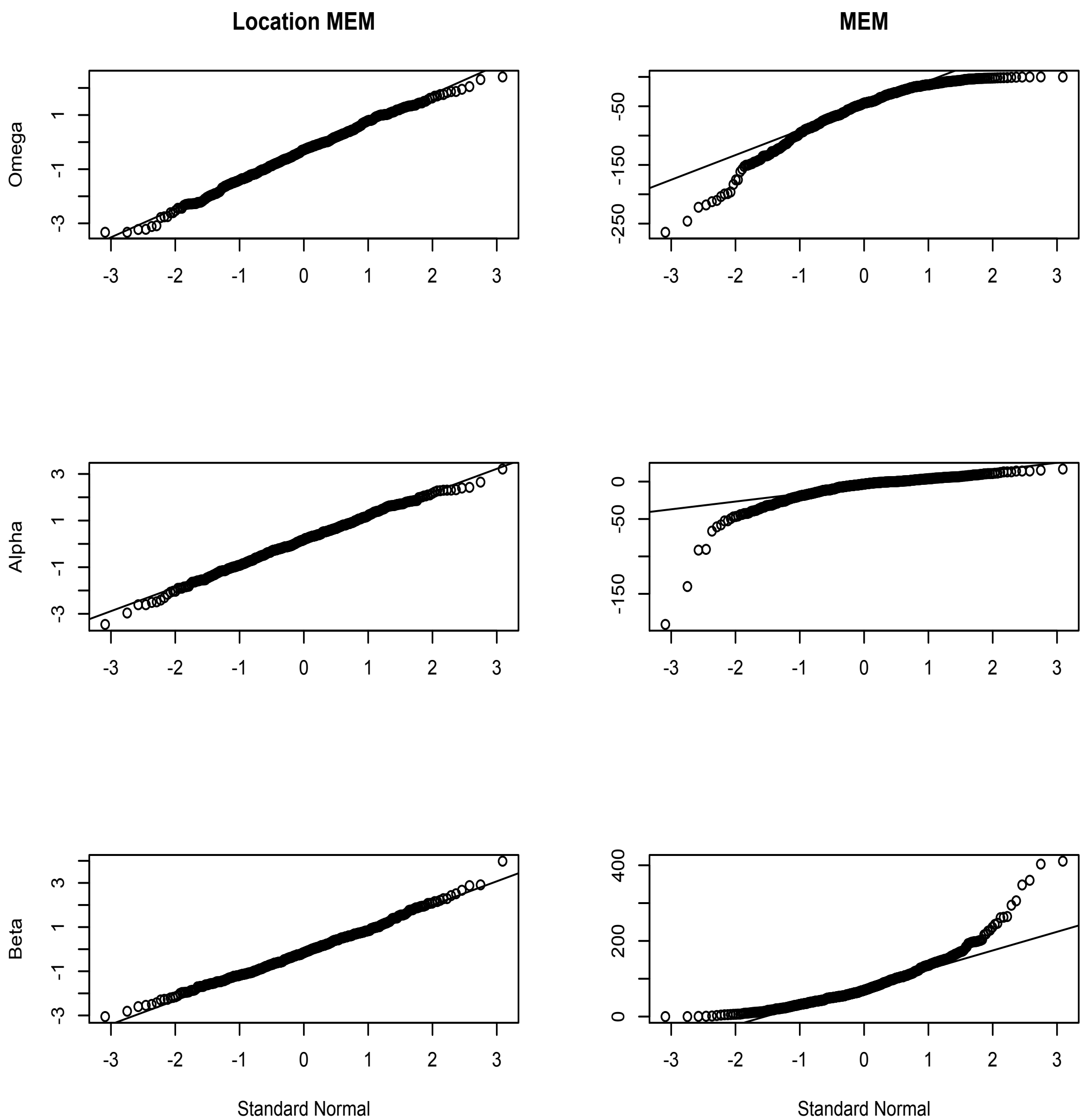

Figure 3. QQ Plot of standardized estimates for $\mu_0 = 8$ at order (1,1)

Table 5. Model Improvement on Raw IBM Trading Data

	Location MEM			MEM		
	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\omega}'$	$\hat{\alpha}$	$\hat{\beta}$
Estimate	1.4057	0.019	0.966	3.331	0.022	0.961
Std. Error	0.4043	0.0029	0.0063	0.9216	0.0029	0.0063
Log Likelihood		-56778.3			-62600.27	
Ljung-Box Q(30)		38.611			49.0837	
p Value of Q(30)		0.04027			0.00276	

Table 6. Model Improvement on Adjusted IBM Trading Data

	Location MEM			MEM		
	ω	α	β	ω'	α	β
Estimate	0.034	0.024	0.922	0.077	0.028	0.893
Std. Error	0.0189	0.0069	0.0361	0.0482	0.0096	0.0578
Log Likelihood		-18210.63			-21718.21	
Ljung-Box Q(30)		37.6396			39.4853	
p Value of Q(30)		0.05014			0.03291	

Summary

This chapter illustrates that, under weak conditions, the estimator for the location parameter and the modified QMLE for the other parameters in Location

MEM(1,1) are both consistent. With one more assumption on the deviation of the location parameter's estimate, the modified QMLE has an asymptotically normal distribution. Asymptotic normality and the connection between MEM(1,1) and the extended model as described by equation (3.9), are illustrated by both simulated data sets and real data of IBM trading.

The model and assumptions discussed in this chapter focus on order (1,1). There exist open questions in other versions of Location MEM. The first and the most interesting one must be estimation of Location MEM(p,q) under conditions of stationarity, ergodicity and moments for GARCH(p,q) as generalized in Bougerol and Picard (1992b). Berkes et al. (2003) discuss the structure of GARCH(p,q) and establish its asymptotic properties under this condition. Berkes and Horvath (2004) propose a class of estimators under the same condition by introducing various density-based likelihood functions and investigates their efficiency. It would be ideal to extend current results to Location MEM(p,q) following the methodology of Berkes et al. (2003).

CHAPTER 4

ASYMPTOTIC ANALYSIS FOR LOCATION MEM(p,q)

As discussed in the previous chapter, asymptotic analysis on GARCH(1,1) under certain conditions can be carried over to Location MEM(1,1). Motivated by existing research about GARCH(p,q), we want to extend the conclusions about Location MEM(1,1) to the (p,q) order and explore alternative conditions on which to base the results.

The main purpose of this chapter is to demonstrate consistency and asymptotic normality of the QMLE for Location MEM(p,q). Analogous to the (1,1) case, the QML in our work still takes the form of exponential density with an estimator of the location parameter included. Consistency of the estimator for the location parameter can be verified by adopting an equivalent representation of the mean equation, which is proposed by Berkes et al. (2003). This representation, which serves as the key to asymptotic properties of the QMLE, is yielded by a weak assumption on the mean equation that cannot be guaranteed by stationarity, ergodicity and finite moments. In other words, once strict stationarity and existence of finite moments are established, asymptotic properties of the standard QMLE can be obtained by imposing an additional weak condition on the mean equation. As generalized in Carrasco and Chen (2002), a process described by GARCH-type models is strictly stationary and has finite moment, if certain conditions about the innovation term are met. Hence, the QMLE employed in the current case is ideal for Location MEM(p,q) under various assumptions.

The contribution of this chapter not only lies in the asymptotic inference about Location MEM at a higher order, but also provides generalizations of various GARCH-type models. Properties of these models can be developed using matrix forms presented by Berkes et al. (2003) and Carrasco and Chen (2002). Therefore, combined with the alternative representation for the mean equation in Berkes et al. (2003), asymptotic analysis on Location MEM(p,q) with linear and power specifications is conducted under various conditions for stationarity and finite moments.

The remaining sections are structured as follows. Section 2 provides two types of specifications for Location MEM(p,q), and summarizes strict stationary and finite moment conditions for these models. Section 3 presents an alternative form of the mean equation and asymptotic inference results for each model. Section 4 gives a simulation study for linear Location MEM(2,2), illustrating the theorems in the previous section.

LOCATION MEM(p,q)

Suppose r_t is an observed process. A Linear Location Multiplicative Error Model (MEM) with order (p,q) is defined as

$$r_t = \mu_0 + \epsilon_t,\ \epsilon_t = h_{0t} z_t, z_t \geq 0 \tag{4.1}$$

$$h_{0t} = \omega + \sum_{i=1}^{p} \alpha_i (r_{t-i} - \mu_0) + \sum_{j=1}^{q} \beta_j h_{0t-j} \tag{4.2}$$

where the z_t are i.i.d. with $E(z_t) = 1$ and $\theta_0 = (\mu_0, \omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)'$ is the vector of true parameters, $\omega > 0, \alpha_i \geq 0, 1 \leq i \leq p, \beta_j \geq 0, 1 \leq j \leq q$. Here $D(1, \phi^2)$ represents distribution with unity expectation and variance ϕ^2. μ_0 is the location parameter with upper bound $r_{n(1)}$, which is the minimum of $\{r_t\}_{t=1}^n$. Similarly, a Location Power MEM(p,q) and Location Logarithmic MEM(p,q) can be specified as follows. r_t is still the process described by equation (4.1).

Location Power MEM(p,q):

$$h_{0t}^{\delta} = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^{\delta} + \sum_{j=1}^{q} \beta_j h_{0t-j}^{\delta} \tag{4.3}$$

Obviously, Power MEM(p,q) reduces to linear LMEM(p,q) when the power is 1.

Location Logarithmic MEM(p,q):

$$\ln h_{0t} = \omega + \sum_{i=1}^{p} \alpha_i \ln \epsilon_{t-i} + \sum_{j=1}^{q} \beta_j \ln h_{0t-j} \tag{4.4}$$

No matter which specification that h_{0t} follows, analogous to the (1,1) order, there exists a connection between MEM(p,q) and Location MEM(p,q). Take the linear case for example. Set $x_t = \dfrac{\mu_0 + h_{0t} z_t}{\mu_0 + h_{0t}}$. It can be illustrated that $E(x_t|\mathscr{F}_{t-1}) = 1$ and x_t is stationary and ergodic.

Let $\psi_{0t} = \mu_0 + h_{0t}$, then

$$\begin{aligned} r_t &= \psi_{0t} x_t \\ \psi_{0t} &= \omega' + \sum_{i=1}^{p} \alpha_i r_{t-i} + \sum_{j=1}^{q} \beta_j \psi_{0t-j} \end{aligned} \tag{4.5}$$

where $\omega' = \omega + \left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{q} \beta_j\right) \mu_0$. Similarly, when μ_0 is relatively large compared to h_{0t}, x_t becomes degenerate and the QML estimation for process (4.5) is not consistent.

Model Specification

The above location multiplicative error models can be expressed by two matrix forms. Define $e_t = f(z_t)$. The first form uses the following notation as employed by Berkes et al. (2003). Let

$$\begin{aligned} \boldsymbol{\tau}_t &= (\beta_1 + \alpha_1 e_t, \beta_2, ..., \beta_{q-1}) \in \mathbb{R}^{q-1}, \\ \boldsymbol{\xi}_t &= (e_t, 0, ..., 0) \in \mathbb{R}^{q-1}, \\ \boldsymbol{\alpha} &= (\alpha_2, ..., \alpha_{p-1}) \in \mathbb{R}^{p-2}. \end{aligned}$$

Define matrix C_t in block form as

$$\boldsymbol{C}_t = \begin{bmatrix} \boldsymbol{\tau}_t & \beta_q & \boldsymbol{\alpha} & \alpha_p \\ I_{q-1} & 0 & 0 & 0 \\ \boldsymbol{\xi}_t & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{bmatrix},$$

where I_n is the identity matrix of size n.

Let

$$\begin{aligned} \boldsymbol{X}_t &= (g(h_{0t}), \ldots, g(h_{0t-q-1}), \Lambda(\epsilon_{t-1}), \ldots, \Lambda(\epsilon_{t-p-1}))' \in \mathbb{R}^{p+q-1}, \\ \boldsymbol{D} &= (\omega, 0, \ldots, 0)' \in \mathbb{R}^{p+q-1} \end{aligned}$$

Equation (4.2) gives

$$\boldsymbol{X}_{t+1} = \boldsymbol{C}_t \boldsymbol{X}_t + \boldsymbol{D} \tag{4.6}$$

For Location Power MEM, $g(h_{0t}) = h_{0t}^{\delta}$, $e_t = z_t^{\delta}$, and $\Lambda(\epsilon_t) = \epsilon_t^{\delta}$.

For Location Logarithmic MEM, $g(h_{0t}) = \ln h_{0t}$, $e_t = \ln z_t$, $\Lambda(\epsilon_t) = \ln \epsilon_t$, $\boldsymbol{\tau}_t = (\beta_1, \beta_2, ..., \beta_{q-1})$, $\boldsymbol{D} = (\omega + \alpha_1 e_t, \ldots, e_t, \ldots, 0)'$ and $\boldsymbol{\xi}_t = (1, 0, ..., 0)$.

The system in (4.2) is not only applied to power GARCH and type I logarithmic MEM, but also to nonlinear asymmetric GARCH, GJR-GARCH and TGARCH.

The second form adopts the structure of power GARCH(p,q) proposed by Carrasco and Chen (2002). Let

$$\boldsymbol{Y}_{t+1} = (\Lambda(\epsilon_{t+1}), \ldots, \Lambda(\epsilon_{t+2-p}), g(h_{t+1}), \ldots, g(h_{t+2-q}))' \in \mathbb{R}^{p+q},$$

$$\boldsymbol{B}(e_{t+1}) = (e_{t+1}, \ldots, \omega, \ldots, 0)' \in \mathbb{R}^{p+q}$$

$$\boldsymbol{A}(e_{t+1}) = \begin{bmatrix} \alpha_1 e_{t+1} & \ldots & \alpha_{p-1} e_{t+1} & \alpha_p e_{t+1} & \beta_1 e_{t+1} & \ldots & \beta_{q-1} e_{t+1} & \beta_q e_{t+1} \\ 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \ldots, & 1 & 0 & 0 & \ldots & 0 & 0 \\ \alpha_1 & \ldots & \alpha_{p-1} & \alpha_p & \beta_1 & \ldots & \beta_{q-1} & \beta_q \\ 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

Equation (4.2) yields

$$\boldsymbol{Y}_{t+1} = \boldsymbol{A}(e_{t+1})\boldsymbol{Y}_t + \boldsymbol{B}(\omega e_{t+1}) \tag{4.7}$$

where $g(h_{0t}) = h_{0t}^{\delta}$, $e_t = z_t^{\delta}$, and $\Lambda(\epsilon_t) = \epsilon_t^{\delta}$.

If the model adopts a type I logarithmic specification, i.e. h_{0t} is determined by equation (4.4), then the matrix form is

$$\boldsymbol{Y}_{t+1} = \boldsymbol{A}(1)\boldsymbol{Y}_t + \boldsymbol{B}(\omega + e_{t+1}) \tag{4.8}$$

where $g(h_{0t}) = \ln h_{0t}$, $e_t = \ln z_t$ and $\Lambda(\epsilon_t) = \ln \epsilon_t$.

The above expression is also applicable to the models covered in the first matrix form. For the type II logarithmic MEM(p,q), as introduced by Bauwens and Giot (2000), the term $\ln r_{t-i}$ in (4.4) is replaced by r_{t-i}. This type of model cannot be represented by either (4.2) or (4.8). There are some other GARCH-type models that cannot be covered by this general matrix form, such as EGARCH and VGARCH.

Furthermore, if the mean equation includes functions defined by ϵ_t, e.g. dummy variables, specified as

$$g(h_{0t}) = \omega + \sum_{i=1}^{p} \alpha_i \Lambda_1(\epsilon_{t-i}) + \sum_{i=1}^{p} \gamma_i \Lambda_2(\epsilon_{t-i}) + \sum_{j=1}^{q} \beta_j g(h_{0t-j}) \tag{4.9}$$

the model can be described by the matrices below.

$$\boldsymbol{W}_{t+1} = (\Lambda_1(\epsilon_{t+1}), \ldots, \Lambda_1(\epsilon_{t+2-p}), \Lambda_2(\epsilon_{t+1}), \ldots, \Lambda_2(\epsilon_{t+2-p}), g(h_{t+1}), \ldots, g(h_{t+2-q}))' \in \mathbb{R}^{2p+q},$$

$$\boldsymbol{G}(e_{t+1}) = (e_{t+1}, \ldots, \omega, \ldots, 0)' \in \mathbb{R}^{2p+q}$$

$$\boldsymbol{H}(e_{t+1}) = \begin{bmatrix} \boldsymbol{\alpha} e_{t+1} & \alpha_p e_{t+1} & \boldsymbol{\gamma} e_{t+1} & \gamma_p e_{t+1} & \boldsymbol{\beta} e_{t+1} & \beta_q e_{t+1} \\ I_{p-1} & \mathbf{0}_{p-1} & \mathbf{0} & \mathbf{0}_{p-1} & \mathbf{0} & \mathbf{0}_{p-1} \\ \mathbf{0} & \mathbf{0}_p & I_p & \mathbf{0}_p & \mathbf{0} & \mathbf{0}_p \\ \boldsymbol{\alpha} & \alpha_p & \boldsymbol{\gamma} & \gamma_p & \boldsymbol{\beta} & \beta_q \\ \mathbf{0} & \mathbf{0}_{q-1} & \mathbf{0} & \mathbf{0}_{q-1} & I_{q-1} & \mathbf{0}_{q-1} \end{bmatrix}$$

$$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{p-1}), \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_{q-1}), \quad \mathbf{0}_n = (0, \ldots, 0)' \in \mathbb{R}^n.$$

Models (4.1) and (4.9) can be represented as

$$\boldsymbol{W}_{t+1} = \boldsymbol{H}(e_{t+1}) \boldsymbol{W}_t + \boldsymbol{G}(\omega e_{t+1}) \tag{4.10}$$

when h_{0t} is expressed by (4.3) or

$$\boldsymbol{W}_{t+1} = \boldsymbol{H}(1)\boldsymbol{W}_t + \boldsymbol{G}(\omega + e_{t+1}) \tag{4.11}$$

when h_{0t} is expressed by (4.4).

Assumptions

Previous work on Location MEM(1,1) is all based on the assumption of strict stationarity and finite moment for a process. For linear GARCH(1,1), strict stationarity can be enforced by $E\ln(\beta_{01} + \alpha_{01}z_t) < 0$, according to theorem 2 in Nelson (1990), while finite moment is obtained by imposing the finite moment condition on the innovation term. Bougerol and Picard (1992b) establishes necessary and sufficient condition for a unique stationary solution to GARCH(p,q), which is widely employed in later inference about GARCH models. For example, Berkes et al. (2003) combines it with finite moment to establish asymptotic properties for linear GARCH(p,q). In fact, there exist other conditions for stationarity and finite moments of GARCH type models, as investigated by Carrasco and Chen (2002). All of these conditions can be carried over to our research about Location MEM(p,q). They are summarized into three propositions as follows. Throughout this papaer, we assume that $z_t's$ are independent, identically distributed and nondegenerate.

The first proposition is a summary of some results in Berkes et al. (2003). Define the norm of any $d \times d$ matrix M by

$$\|M\| = \sup\{\|M\boldsymbol{x}\|_d / \|\boldsymbol{x}\|_d : \boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} \neq \boldsymbol{0}\},$$

where $\|.\|_d$ is the Euclidean norm in $\mathbb{R}^d$. The top Lyapunov exponent γ_L is

$$\gamma_L = \inf_{0 \le n \le \infty} \frac{1}{n+1} E \ln \|C_0 C_1 \cdots C_n\|.$$

ASSUMPTION 4.1. *Suppose the following conditions are satisfied.*

(a) $E(\ln \|C_0\|) < \infty$.

(b) $\gamma_L < 0$.

(c) $E|e_t|^s < \infty$ *for some* $s \ge 1$.

Assumption 4.1 (a) and (b) enforce strict stationarity of the process, while the finite moments of h_{0t} and r_t are established by condition (c).

PROPOSITION 4.1. *Under Assumption 4.1,*

(1) $\boldsymbol{X}_{t+1} = \boldsymbol{D} + \sum_{0 \le k < \infty} C_t \ldots C_{t-k} \boldsymbol{D}$

(2) r_t *and* h_{0t} *are strictly stationary, and* $E|\lambda(\epsilon_t)|^s < \infty$, $E|g(h_{0t})|^s < \infty$ *for the* s *in Assumption 4.1.*

For the models described by 4.7 and 4.8, the following conclusions follow from propositions 12 and 13 in Carrasco and Chen (2002).

ASSUMPTION 4.2. *Suppose the following conditions are satisfied.*

(a) $E|e_t|^s < \infty$ *for some* $s \ge 1$.

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

PROPOSITION 4.2. *Under Assumption 4.2,* r_t *and* h_t *are strictly stationary, and* $E|\Lambda(\epsilon_t)|^s < \infty$, $E|g(h_{0t})|^s < \infty$ *for the* s *in Assumption 4.2.*

ASSUMPTION 4.3. *Suppose the following conditions are satisfied.*

(a) $E|e_t|^s < \infty$ *for some even integer* $s \geq 2$.

(b) $E|\rho|^s < 1$ *for the* s *in (a), where* ρ *is the largest root of* $1 = \sum_{i=1}^{\max(p,q)} \lambda^{-i}(\beta_i + \alpha_i e_t)$.

PROPOSITION 4.3. *Under Assumption 4.3,* r_t *and* h_t *are strictly stationary and* $E|\Lambda(\epsilon_t)|^s < \infty$, $E|g(h_{0t})|^s < \infty$ *for the integer* s *in Assumption 4.3.*

Observing the difference between the above three propositions, it's not hard to see that the stronger the assumptions imposed on e_t, the higher the order of finite moments for $\Lambda(\epsilon_t)$. Propositions 4.1, 4.2 and 4.3 are the foundation for asymptotic inference in the next section. Each of the assumptions is sufficient for consistency and asymptotic normality of the models in power specification. Moreover, for the models in logarithmic specification, the same asymptotic properties can be established as long as the number s in each assumption satisfies $s \geq 4$. Therefore, for simplicity, the analysis in the next section is based on Assumption 4.1 and concentrates on the linear model. Similar to the (1,1) model, the moment condition below is also required in building consistency and asymptotic normality.

ASSUMPTION 4.4. $E|r_{n(1)} - \mu_0|^p = o(\frac{1}{\sqrt{n}})$ *for some* $p < 1$.

Asymptotic Analysis

Infinite Representation

In Berkes et al. (2003), the conditional variance of a strictly stationary GARCH(p,q) process can be represented by the observed data and a series of exponentially decaying coefficients. We will adopt this infinite representation in our work as well .

Let

$$\mathcal{A}(x) = \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_p x^p$$
$$\mathcal{C}(x) = 1 - \beta_1 x - \beta_2 x^2 - \cdots - \beta_q x^q$$

Then equation (4.2) can be rewritten as

$$\mathcal{C}(B) h_{0t} = \omega + \mathcal{A}(B)\epsilon_t \tag{4.12}$$

where B is the backward shift operator. Bougerol and Picard (1992a) showed that Assumption 4.1 (b) implies $\beta_1 + \beta_2 + \cdots + \beta_q < 1$, which ensures that $\mathcal{C}(x) = 0$ has all roots outside of the unit circle. Hence, we have

$$\sum_{j=0}^{\infty} d_j x^j = \frac{1}{\mathcal{C}(x)}, \; |x| \le 1$$

and $d_0, d_1, d_2, \cdots$ decay exponentially fast. Then h_{0t} can be represented as

$$h_{0t} = \frac{\omega}{\mathcal{C}(1)} + \frac{\mathcal{A}(B)}{\mathcal{C}(B)}\epsilon_t \tag{4.13}$$

ASSUMPTION 4.5. *(a)$E \ln h_{0t} < \infty$. (b)z_t are nondegenerate*

Theorems 2.1 and 2.3 in Berkes et al. (2003) verify the following conclusion.

PROPOSITION 4.4. *Under Assumptions 4.1 and 4.5, h_{0t} can be expressed as*

$$h_{0t} = c_0 + \sum_{1 \le i \le \infty} c_i \epsilon_{t-i} \tag{4.14}$$

where

$$c_0 = \omega \sum_{m=0}^{\infty} d_m$$
$$c_j = \alpha_1 d_{j-1} + \alpha_2 d_{j-2} + \cdots + \alpha_p d_{j-p}, \; 1 \le j < \infty$$

and the c_j *'s are unique.*

Note that the uniqueness of this expression is guaranteed by Assumption 4.5 (b), which serves as an important factor for precision of estimates.

Let $\mathcal{D}(x) = \sum_{1 \le i \le \infty} c_i x^i$, then $\mathcal{A}(x) = \mathcal{C}(x)\mathcal{D}(x)$, which yields

$$\alpha_1 = c_1, \;\; \alpha_2 = c_2 - \beta_1 c_1, \;\; \alpha_3 = c_3 - \beta_1 c_2 - \beta_2 c_1, \cdots$$

Therefore, $c_1, c_2, \cdots$ are functions of the true parameters $\eta_0 = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)'$. Denote the corresponding unknown parameters for η_0 by $\boldsymbol{u} = (x, s_1, \cdots, s_p, t_1, \cdots, t_q)$. Berkes et al. (2003) give an expression of coefficients $c_i, \; 1 \le i < \infty$ in terms of $\boldsymbol{u}$.

If $q \ge p$, this expression is given by

$$\begin{aligned} c_0(\boldsymbol{u}) &= \tfrac{x}{1-(t_1+\cdots+t_q)}, \\ c_1(\boldsymbol{u}) &= s_1, \\ c_2(\boldsymbol{u}) &= s_2 + t_1 c(\boldsymbol{u}), \\ &\vdots \\ c_p(\boldsymbol{u}) &= s_p + t_1 c_{p-1}(\boldsymbol{u}) + t_{p-1} c_1(\boldsymbol{u}), \\ c_{p+1}(\boldsymbol{u}) &= t_1 c_p(\boldsymbol{u}) + \cdots + t_p c_1(\boldsymbol{u}), \\ &\vdots \\ c_q(\boldsymbol{u}) &= t_1 c_{q-1}(\boldsymbol{u}) + \cdots + t_{q-1} c_1(\boldsymbol{u}) \end{aligned}$$

If $q < p$, then

$$\begin{aligned} c_0(\boldsymbol{u}) &= \tfrac{x}{1-(t_1+\cdots+t_q)}, \\ c_1(\boldsymbol{u}) &= s_1, \end{aligned}$$

$$c_2(\boldsymbol{u}) = s_2 + t_1 c(\boldsymbol{u}),$$
$$\vdots$$
$$c_{q+1}(\boldsymbol{u}) = s_{q+1} + t_1 c_q(\boldsymbol{u}) + \cdots + t_q c_1(\boldsymbol{u}),$$
$$\vdots$$
$$c_p(\boldsymbol{u}) = s_p + t_1 c_{p-1}(\boldsymbol{u}) + \cdots + t_q c_{p-q}(\boldsymbol{u})$$

If $i > \max(p, q)$,then $c_i(\boldsymbol{u}) = t_1 c_{i-1}(\boldsymbol{u}) + t_2 c_{i-2}(\boldsymbol{u}) + \cdots + t_q c_{i-q}(\boldsymbol{u})$. Let $0 < \underline{u} < \bar{u},\ 1 < \rho_0 < 1,\ q\underline{u} < \rho_0$ and define the parameter space

$$U = \left\{\boldsymbol{u} : \sum_{j=1}^{q} t_j \leq \rho_0\right\} \cup \{\boldsymbol{u} : \underline{u} \leq \min(x, s_1, s_2, \cdots, s_p, t_1, t_2, \cdots, t_q) \\ \leq \max(x, s_1, s_2, \cdots, s_p, t_1, t_2, \cdots, t_q) \leq \bar{u}\}.$$

Lemmas 3.1 and 3.2 in Berkes et al. (2003) are summarized and restated as

PROPOSITION 4.5. *Under Assumption 4.5,*

(1)$C_1 \underline{u}^i \leq c_i(\boldsymbol{u}) \leq C_2 \rho_0^{i/q},\ 0 \leq i < \infty$

(2) $|c_0'(\boldsymbol{u})| \leq \min\{C_0, \frac{\bar{u}}{(1-\rho_0)^2}\},\ |c_i'(\boldsymbol{u})| \leq C_3 c_i(\boldsymbol{u}) i,\ 1 \leq i < \infty$

(3)$|c_0''(\boldsymbol{u})| \leq C_4,\ |c_i''(\boldsymbol{u})| \leq C_5 c_i(\boldsymbol{u}) i^2,\ 1 \leq i < \infty$

(4)$|c_0^{(3)}(\boldsymbol{u})| \leq C_6,\ |c_i^{(3)}(\boldsymbol{u})| \leq C_7 c_i(\boldsymbol{u}) i^3,\ 1 \leq i < \infty$

where $C_0, ..., C_7$ *are constants and* $\boldsymbol{u} \in U$.

Results

In order to discuss asymptotic properties of the estimator for Location MEM(p,q), we follow the same path as used in Location MEM(1,1). First, we find a consistent estimator for the location parameter μ_0. Second, a modified quasi-maximum likelihood estimator for the remaining parameters is defined and its asymptotic properties are investigated. Extended theorems about extremum estimators found in the analysis of Location MEM(1,1) are still the basis for current work in this section.

The quasi likelihood function for the current case takes the form

$$L_n^*(\mu, \boldsymbol{u}) = \frac{1}{n}\sum_{t=1}^{n} l_t^*(\mu, \boldsymbol{u}) \qquad l_t^*(\theta) = -\left(\ln h_t^*(\mu, \boldsymbol{u}) + \frac{r_t - \mu}{h_t^*(\mu, \boldsymbol{u})}\right) \tag{4.15}$$

where

$$h_t^*(\mu, \boldsymbol{u}) = c_0(\boldsymbol{u}) + \sum_{i=1}^{t-1} c_i(\boldsymbol{u})(r_{t-i} - \mu) \tag{4.16}$$

Before investigating asymptotic behavior of the observed likelihood above, we need to define the following:

$$L(\mu, \boldsymbol{u}) = -E(l_t(\mu, \boldsymbol{u})),\ l_t(\mu, \boldsymbol{u}) = \ln h_t(\mu, \boldsymbol{u}) + \frac{r_t - \mu}{h_t(\mu, \boldsymbol{u})} \tag{4.17}$$

where

$$h_t(\mu, \boldsymbol{u}) = c_0(u) + \sum_{i=1}^{\infty} c_i(\boldsymbol{u})(r_{t-i} - \mu) \tag{4.18}$$

DEFINITION 3. *Modified QMLE* $\hat{\eta}_n$ *is defined as* $\hat{\eta}_n = \arg\max_{\boldsymbol{u}\in U} L_n^*(r_{n(1)}, \boldsymbol{u})$.

Set $\hat{G}_n(\mu, \boldsymbol{u}) = -\frac{1}{n}\sum_{t=1}^{n} \nabla_{\boldsymbol{u}}^2 l_t^*(\mu, \boldsymbol{u})$ and $\hat{C}_n(\mu, \boldsymbol{u}) = \frac{1}{n}\sum_{t=1}^{n} \nabla_{\boldsymbol{u}} l_t^*(\mu, \boldsymbol{u}) \nabla_{\boldsymbol{u}} l_t^*(\mu, \boldsymbol{u})'$.

THEOREM 4.6. *Under Assumptions 4.1, 4.4 and 4.5,*

(1) $\hat{\theta}_n \to_p \theta_0$.

(2) $\sqrt{n}(\hat{\eta}_n(r_{n(1)}) - \eta_0) \to_D N(0, V_0)$ *where* $V_0 = G_0^{-1} C_0 G_0^{-1}$, $C_0 = -E(\nabla_{\boldsymbol{u}} l_t(\theta_0) \nabla_u l_t(\theta_0)')$ *and* $G_0 = -E\nabla_{\boldsymbol{u}}^2 l_t(\theta_0)$.

(3)$\hat{V}_n = \hat{G}_n^{-1}(\hat{\theta}_n)\hat{C}_n(\hat{\theta}_n)\hat{G}_n^{-1} \to_p V_0 = G_0^{-1} C_0 G_0^{-1}$.

As demonstrated in Location MEM(1,1), when it is non-integrated i.e., $\alpha_0 + \beta_0 < 1$, the modified QMLE coincides with the maximizer of the quasi-likelihood function. However, for the (p,q) case, even if the sum of parameters η_0 is restricted to be less than one, we are unable to validate such a connection in estimators based on the general representation of h_{0t} in (5.10).

Simulation and Empirical Results

In this section, we illustrate asymptotic results in Theorem 4.6 along with the contrast between MEM(2,2) and Location MEM(2,2) by simulated replications and IBM trading data from the previous chapter.

We generate 500 replicates with sample size n=30000 by a DGP (Data Generating Process) of Location Gamma-MEM(2,2) with parameter values $(\omega_1, \alpha_1, \alpha_2, \beta_1, \beta_2) = (0.03, 0.05, 0.1, 0.5, 0.3)$. Consequently, parameters for the equivalent MEM(2,2) process described by (4.5) are $(\omega_2, \alpha_1, \alpha_2, \beta_1, \beta_2) = (0.08, 0.05, 0.1, 0.5, 0.3)$.

The estimation result in table 7 not only confirms consistency and asymptotic normality of the modified QMLE for a Location MEM(2,2) process, but also presents the bias in the estimation of MEM(2,2) when μ_0 is relatively large. Violation of asymptotic normality in MEM(2,2) is shown by the QQ-plot in figure 5 as well.

Applying both models with lag (2,1) to the IBM trade data in chapter 3, we find significant improvement on goodness-of-fit brought by the location parameter. For each model, (2,1) is the optimal order of lags based on BIC and significance of parameters. Tables 8 and 9 present estimates, standard errors, log likelihood and Ljung-Box test statistics for data before and after diurnal adjustment. There is no doubt that location MEM(2,2) greatly outperforms the traditional model on raw data.

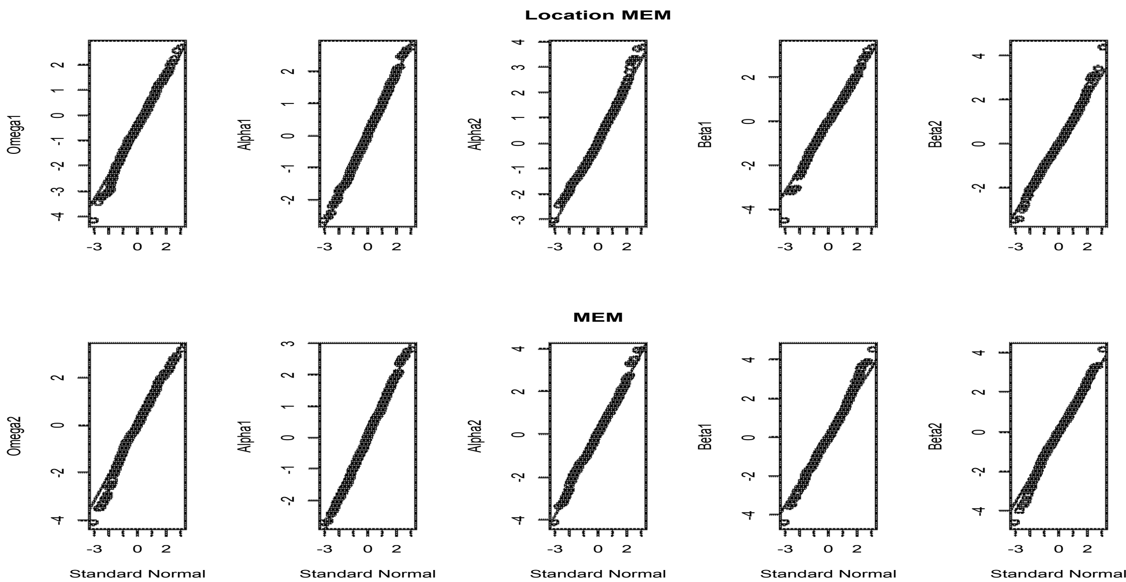

Figure 4. QQ Plot of standardized estimates for $\mu_0 = 1$ at order (2,2)

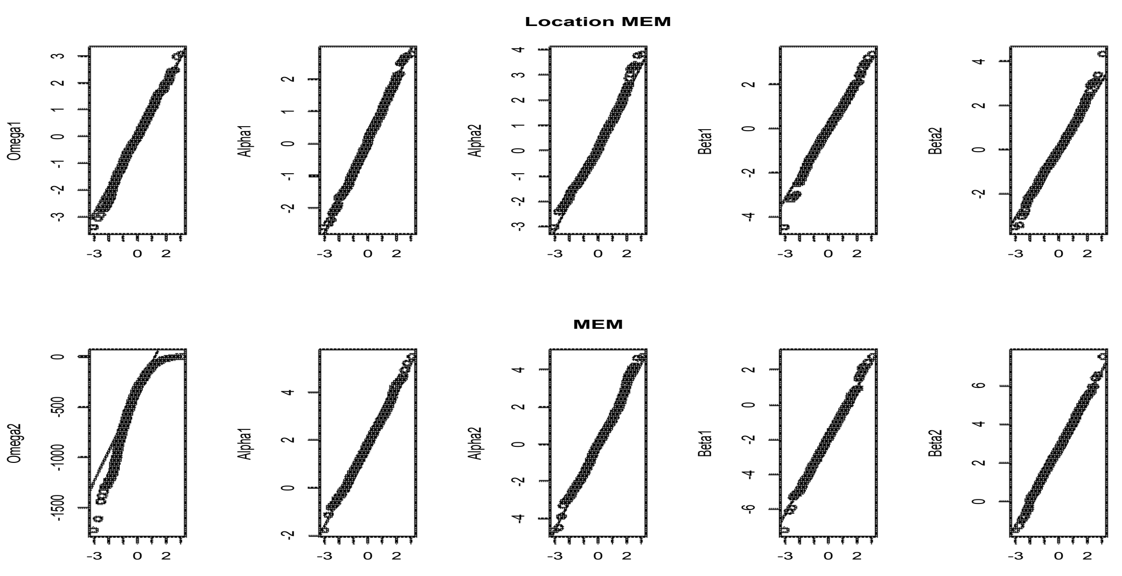

Figure 5. QQ Plot of standardized estimates for $\mu_0 = 8$ at order (2,2)

Table 7. Estimates by Two Models

	Location MEM					MEM				
	$\hat{\omega}_1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\omega}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
True	0.03	0.05	0.1	0.5	0.3	0.08	0.05	0.1	0.5	0.3
					$\mu_0 = 1$					
Mean	0.029	0.05	0.1	0.503	0.297	0.081	0.05	0.1	0.5	0.297
S.D.	0.002	0.006	0.009	0.085	0.074	0.007	0.006	0.009	0.087	0.076
					$\mu_0 = 8$					
Mean	0.03	0.05	0.1	0.503	0.296	0.0055	0.06	0.1	0.408	0.433
S.D.	0.003	0.006	0.009	0.084	0.074	0.0001	0.006	0.008	0.072	0.066

This contrast is more pronounced at lag (2,1), compared to the result in the (1,1) case.

Table 8. Improvement on Raw IBM Trading Data at Order of (1,2)

	Location MEM				MEM			
	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\omega}'$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$
Estimate	2.0548	0.0285	0.4398	0.5097	4.8176	0.0328	0.4137	0.5286
Std. Error	0.6803	0.0047	0.1248	0.1251	1.6745	0.0052	0.1509	0.1531
Log Likelihood	-56772.35				-62598.45			
L-B Q(15)	23.6975				32.2416			
p Value of Q(15)	0.07041				0.005968			

Table 9. Improvement on Adjusted IBM Trading Data at Order of (1,2)

	Location MEM				MEM			
	ω	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\omega}'$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$
Estimate	0.0429	0.032	0.4802	0.4187	0.0892	0.0353	0.4755	0.3975
Std. Error	0.0218	0.0081	0.1549	0.1601	0.0453	0.0097	0.1655	0.1741
Log Likeilhood	-18208.07				–21717.05			
L-B Q(12)	20.3357				21.1042			
p Value of Q(12)	0.061				0.04888			

CHAPTER 5

ASYMPTOTIC ANALYSIS FOR ZERO-AUGMENTED GENERALIZED F DISTRIBUTION MEM(p,q)

In addition to the analysis proposed by Hautsch et al. (2014), we investigate not only consistency and efficiency, but also asymptotic normality for estimates of parameters in the mean equation of ZAF MEM(p,q). We establish the exact MLE for the model, as well as modify the standard (exponential) QML to improve the estimates. Asymptotic properties will be theoretically and empirically validated in our work. Following the analysis in that article, our simulation study compares ZAF MLE with exponential QMLE, in terms of efficiency and normality.

Similar to the estimation of Location MEM, the parameter π is separately estimated by an empirical frequency, and the estimator for the remaining parameters is obtained by employing ZA-GF log likelihood with this empirical estimate for π. Asymptotic properties of such an estimator can be established with the support of extended theorems about extremum estimators. In other words, we have to ensure finite expectation on different order gradients of the ZAF log likelihood function, which can be guaranteed by assuming stationarity of the process and a finite moment of disturbance, as shown by Berkes et al. (2003). Because the density function is specified in the current model, a restriction on the density parameters is sufficient for the finite moment of disturbance.

Moreover, it is promising to develop the same analysis on various multiplicative error structures incorporating ZAF. Although this chapter focuses on a basic MEM instead of any other MEM structures, the key to such an analysis on any type of MEM are the conditions of stationarity, exponential decay and finite moment, as long as ZAF density is employed. Those conditions for various MEM (or GARCH) structures are thoroughly discussed and established in the existing literature, such as Carrasco and Chen (2002) and Meitz and Saikkonen (2008). Therefore, conclusions in Hautsch et al. (2014) can be demonstrated as well with the conditions specified.

The rest of this chapter is composed as follows. In Section 2, ZA-GF distribution is specified and applied to MEM(p,q). Section 3 presents sufficient conditions for a MEM(p,q) process being strictly stationary, exponentially decaying and having finite moment, which is the preparation for asymptotic inference. In Section 4, we develop the theoretical and observed log likelihood based on ZA-GF distribution, as well as draw asymptotic conclusions on the MLE. In Section 5, we use simulation to illustrate consistency and asymptotic normality of MLE for all parameters except π. We also apply exponential QML to the simulated ZA-GF MEM(p,q) data set, and the histogram of estimates shows an obvious evidence of misspecification.

ZAF in MEM(p,q)

As proposed in Hautsch et al. (2014), a zero-augmented generalized F (ZAF) distribution is applied to a nonnegative random variable X with a high proportion of zero outcomes. Because of the non-trivial proportion of zero outcomes, a discrete probability mass is assigned at the exact zero value as follows:

$$\pi = P(X > 0),\ 1 - \pi = P(X = 0) \tag{5.1}$$

Suppose that when $X > 0$, X follows a Generalized F distribution with density function

$$g(x) = g(x|\boldsymbol{\nu}, \lambda) = \frac{ax^{am-1}[\eta + (x/\lambda)^a]^{(-\eta-m)}\eta^\eta}{\lambda^{am}\mathcal{B}(m,\eta)} \tag{5.2}$$

where $\boldsymbol{\nu} = (a, m, \eta)$, $a > 0, m > 0, \eta > 0$ and $\lambda > 0$. $\mathcal{B}(\cdot)$ is the full Beta function $\mathcal{B}(m,\eta) = \frac{\Gamma(m)\Gamma(\eta)}{\Gamma(m+\eta)}$. The unconditional distribution of X, i.e. the ZAF distribution, is semicontinuous with the density function

$$f(x) = (1-\pi)I_{(x=0)} + \pi\frac{ax^{am-1}[\eta + (x/\lambda)^a]^{(-\eta-m)}\eta^\eta}{\lambda^{am}\mathcal{B}(m,\eta)}I_{(x>0)} \tag{5.3}$$

$$E[z_t^s] = \pi\lambda^s\eta^{s/a}\frac{\Gamma(m+s/a)\Gamma(\eta-s/a)}{\Gamma(m)\Gamma(\eta)};\ a\eta > s > 0 \tag{5.4}$$

and

$$E[z_t^{-s}] = \pi\lambda^{-s}\eta^{-s/a}\frac{\Gamma(m-s/a)\Gamma(\eta+s/a)}{\Gamma(m)\Gamma(\eta)};\ am > s > 0 \tag{5.5}$$

Consequently, the log likelihood function of the ZA-GF distribution based on sample data $\{x_t\}_{t=1}^n$ is

$$\begin{aligned} L(\boldsymbol{\gamma}) \quad = \quad & n_z \ln(1-\pi) + n_{nz}\ln\pi + \textstyle\sum_{t\in\mathcal{J}_{nz}}\{\ln a + (am-1)\ln x_t + \eta\ln\eta \\ & -(\eta+m)\ln\{\eta + [x_t\lambda^{-1}]^a\} - \ln\mathcal{B}(m,\eta) - am\ln\lambda\} \end{aligned} \tag{5.6}$$

where $\boldsymbol{\gamma} = (a, m, \eta, \lambda)'$. $\mathcal{J}_{nz}$ is the set of all subscripts t associated with nonzero observations of x_t, while n_z and n_{nz} denote the number of zero and nonzero observations respectively.

Now we are specifying a linear multiplicative error model incorporating the ZAF distribution. Let $\{r_t\}_{t=1}^n$ be a series of nonnegative observations,

$$r_t = h_{0t}z_t \tag{5.7}$$

$$h_{0t}^{\delta} = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^{\delta} + \sum_{j=1}^{q} \beta_j h_{0t-j}^{\delta} \tag{5.8}$$

Note that δ is a positive number, and the z_t's are independent and identically distributed as ZAF with true parameter values satisfying $\lambda_0 \xi_0 = \pi_0^{-1}$, where $\xi_0 = \eta_0^{1/a_0}[\Gamma(m_0 + 1/a_0)\Gamma(\eta_0 - 1/a_0)][\Gamma(m_0)\Gamma(\eta_0)]^{-1}$. Suppose the unknown parameter for ξ_0 is ξ. The parameter space for $\boldsymbol{\nu}$ is denoted as V. The constraint on the density parameters ensures the unity mean of z_t. There is no need to include a location parameter in the current case, due to the non-trivial portion of zero outcomes. When δ=1, the model reduces to a linear MEM(p,q), which is actually a GARCH(p,q). If asymptotic properties of a linear MEM(p,q) can be successfully established, the same properties of a power MEM(p,q) are trivially obtained by replacing h_{0t} with h_{0t}^{δ}. Therefore, the following analysis focuses on a linear ZAF MEM(p,q), i.e. the case $\delta = 1$.

Assumptions

Existing research has shown that the keys to establishing asymptotic properties for a power MEM(p,q) are stationarity and the finite moment of r_t, which can be guaranteed by three alternative sufficient conditions, as summarized in the previous chapter. Throughout this chapter, assumptions 4.3(b) and 4.5 always hold, which implies that we exploit the second matrix form in the previous chapter.

Since the z_t's are i.i.d. ZAF, they have finite moments given by (5.4). By Proposition 4.3, (5.4) and Assumption 4.3(b), r_t is strictly stationary and for $s \geq 1$.

$$E(h_{0t})^s < \infty \tag{5.9}$$

Meanwhile, according to Theorem 2.1 in Berkes et al. (2003), Assumption 4.5 gives an alternative representation of h_{0t} as

$$h_{0t} = c_0 + \sum_{1 \le i \le \infty} c_i r_{t-i} \tag{5.10}$$

where the c_i's are exponentially decaying coefficients, and this representation is unique.

c_i can be computed recursively as

$$\alpha_1 = c_1,\ \alpha_2 = c_2 - \beta_1 c_1,\ \alpha_3 = c_3 - \beta_1 c_2 - \beta_2 c_1, \cdots$$

Therefore, $c_1, c_2, \cdots$ are functions of the true parameters $\boldsymbol{u}_0 = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)'$. Denote the corresponding unknown parameters by $\boldsymbol{u} = (x, s_1, \cdots, s_p, t_1, \cdots, t_q)$. Coefficients $c_i,\ i \le i < \infty$ in terms of $\boldsymbol{u}$ are denoted by $c_i(\boldsymbol{u})$. Let $0 < \underline{u} < \bar{u},\ 1 < \rho_0 < 1,\ q\underline{u} < \rho_0$ and define the parameter space

$$\begin{aligned} U = \left\{ \boldsymbol{u} : \sum_{j=1}^{q} t_j \le \rho_0 \right\} \cup \{ \boldsymbol{u} : \underline{u} \le \min(x, s_1, s_2, \cdots, s_p, t_1, t_2, \cdots, t_q) \\ \le \max(x, s_1, s_2, \cdots, s_p, t_1, t_2, \cdots, t_q) \le \bar{u} \}. \end{aligned}$$

In contrast to our work, the model specification in Hautsch et al. (2014) incorporates logarithmic MEM, which brings in dummy variables and additional parameters in the recursive mean equation. In that case, an observed process, such as cumulative volumes, y_t, are given by

$$y_t = e^{\mu_{0t}} \epsilon_t \tag{5.11}$$

$$\mu_{0t} = \omega + \sum_{i=1}^{p} \alpha_i \ln r_{t-i} I_{(y_{t-i} > 0)} + \sum_{i=1}^{p} \alpha_i^0 I_{(y_{t-i} = 0)} + \sum_{j=1}^{q} \beta_j \mu_{0t-j} \tag{5.12}$$

Although Log-MEM with ZAF is not discussed in the current case, its asymptotic inference can be deduced similarly from stationarity, exponential decaying and moment conditions, which are already established by Propositions 4.2 and 4.3.

Asymptotic Inference

Exact MLE

The theoretical log-likelihood function that we use for the current model is derived by the density function specified in (5.3), with sample data $x_t = \frac{r_t}{h_t}$ and parameter restriction $\lambda = (\pi\xi)^{-1}$. Let ∇_θ, ∇_θ^2 denote the first and second gradients w.r.t θ and let θ_0 be the true value of θ and set

$$l_{1t}(\pi,\theta) = \ln a + (am-1)\ln\frac{r_t}{h_t} + \eta\ln\eta - (\eta+m)\ln\left\{\eta + \left[\frac{\pi\xi r_t}{h_t}\right]^a\right\} - \ln h_t - \ln\mathcal{B}(m,\eta) + am\ln\pi\xi \tag{5.13}$$

where $\theta = (\boldsymbol{\nu}, \boldsymbol{u})$ and $h_t = c_0(\boldsymbol{u}) + \sum_{1\le i\le t-1} c_i(\boldsymbol{u}) r_{t-i}$.

The observed log likelihood function based on the ZAF density is

$$L_{1n}(\pi,\theta) \quad = \quad n_z\ln(1-\pi) + n_{nz}\ln\pi + \sum_{t\in\mathcal{J}_{nz}} l_{1t} \tag{5.14}$$

Hautsch et al. (2014) use empirical frequency $\hat{\pi} = \frac{n_{nz}}{n}$ as a substitute for π_0 in the log likelihood function, to derive robust estimates for the other parameters. In fact, it can be trivially proved by the law of large numbers that $\hat{\pi}$ is a consistent estimate of π_0, as shown in lemma C.1 in Appendix C. Hence, according to the extended theorem about extremum estimators, we adopt the same substitute for π_0 to derive MLE for θ in our case. Although simulation results in that article illustrate consistency for the parameters in the conditional mean equation, it is more ideal to illustrate both consistency and asymptotic normality for all the parameters except π.

DEFINITION 4. $L_1(\pi, \theta) = (1-\pi)\ln(1-\pi) + \pi\ln\pi + \pi E l_{1t}(\pi, \theta)$

The assumption of true density trivially yields that $L_1(\pi_0, \theta)$ is maximized at θ_0, but the uniqueness of this maximizer within $V \times U$ cannot be guaranteed due to the complex expression of $l_{1t}(\pi, \theta)$. In order to establish consistency, and consequently asymptotic normality, we start with a local estimator under the following assumption. Let $D_1 = (\nabla_{\boldsymbol{\nu}'} g(x|\boldsymbol{\nu}, (\xi\hat{\pi})^{-1}), \nabla_b g(bx|\boldsymbol{\nu}, (\xi\hat{\pi})^{-1}))'$ and $D_2 = \nabla_{\boldsymbol{\nu}} g(x|\boldsymbol{\nu}, (\xi\hat{\pi})^{-1})$.

ASSUMPTION 5.1. *$D_1 D_1'$ is positive definite.*

The local maximum likelihood estimator is defined as

DEFINITION 5. $\hat{\theta} = \arg\max_{\theta \in V \times U} L_{1n}(\hat{\pi}, \theta)$.

As mentioned in the previous section, the lower bound of $a\eta$ determines the order of the finite moments of r_t and h_{0t}, therefore, we need to impose a restriction on $a\eta$, that is

ASSUMPTION 5.2. $\eta > \max\{\frac{2}{a}, 2\}$

Note that Assumption 5.2 is stronger than needed for developing the consistency of $\hat{\theta}$. Indeed, the first part of the following theorem can be demonstrated under Assumptions 5.1 and $\eta > \max\{\frac{1}{a}, 1\}$.

THEOREM 5.1. *Under Assumptions 5.1 and 5.2,*

(1) $\hat{\theta} \to_p \theta_0$.

(2) $\sqrt{n}(\hat{\theta} - \theta_0) \to_D N(0, V_1)$ *where* $V_1 = G_1^{-1} B_1 G_1^{-1}$ *and*

$$B_1 = E(\nabla_\theta l_t(\pi_0, \theta_0))\nabla_\theta l_t(\pi_0, \theta_0)'), \quad G_1 = E\nabla_\theta^2 l_t(\pi_0, \theta_0)).$$

A concern for the above results is the validity of assumption 5.1. The nonlinear constraint $\lambda_0 \xi_0 = {\pi_0}^{-1}$ imposed on density function $g(x)$ may lead to a violation of assumption 5.1. Since the primary goal in MEM is to estimate the conditional mean h_{0t}, inaccuracy in the density parameters can be ignored as long as the estimate for $\boldsymbol{u}_0$ is consistent. Fortunately, our goal can be achieved by the partial estimator and a weak assumption below.

Suppose $\hat{\theta} = (\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{u}})$, where $\hat{\boldsymbol{u}}$ is the estimate for $\boldsymbol{u}_0$.

ASSUMPTION 5.3. *There exists a vector* $\boldsymbol{v}$ *such that* $\boldsymbol{v}' D_2 = 0$.

THEOREM 5.2. *Under Assumptions 5.2 and 5.3,* $\hat{\boldsymbol{u}} \rightarrow_p \boldsymbol{u}_0$.

Note that consistency of $\hat{\nu}$ cannot be established and the asymptotic normality of $\hat{\boldsymbol{u}}$ may be invalid as well. In fact, $\hat{\boldsymbol{u}}$ can be viewed as a semiparametric estimator although it requires the assumption of a ZAF distribution, since it doesn't include estimation for density parameters. Furthermore, as discussed in Hautsch (2002), density functions of other positive random variables, e.g. Weibull and Gamma, can be approximated by the density of a GF distribution (when η in GF approaches infinity). Hence, the estimator $\hat{\boldsymbol{u}}$ can serve as a QMLE when the distribution of the error term is given exactly or approximated by a GF distribution and conditions for stationarity and finite moments are satisfied, no matter whether the fraction of zeroes in the series is trivial or not.

Exponential QMLE

Hautsch et al. (2014) state that the exact MLE of the ZAF model is superior to the standard (exponential) QMLE, according to their empirical illustration of consistency and efficiency. The exponential log likelihood used in their work is defined as

$$L_{2n} = L_{2n}(\boldsymbol{u}) = -\sum_{t=1}^{n}\left(\ln h_t + \frac{r_t}{h_t^*}\right) \tag{5.15}$$

In fact, this log likelihood function still gives robust estimation for the current model when n is large enough. The reason that this QML is not ideal compared to the exact likelihood is not inconsistency and the cost of asymptotic normality, but the need for a large number of observations. Specifically, when L_{2n} is applied to a ZAF-MEM series, L_{2n} is reduced to

$$L_{2n} = -\sum_{t=1}^{n}(\ln h_t^*) - \sum_{t\in\mathcal{J}_{nz}}\frac{r_t}{h_t} \tag{5.16}$$

and

$$\frac{1}{n}L_2(\boldsymbol{u}) = -\frac{1}{n}\sum_{t=1}^{n}(\ln h_t) - \frac{\hat{\pi}}{n_{nz}}\sum_{t\in\mathcal{J}_{nz}}\frac{r_t}{h_t} \tag{5.17}$$

Let

$$L_2 = L_2(\boldsymbol{u}) = -E(\ln h_t) - E\frac{h_{0t}}{h_t} \tag{5.18}$$

Since $E(z_t|t\in\mathcal{J}_{nz}) = \frac{1}{\pi_0}$, $E\left(\frac{r_t}{h_t}|t\in\mathcal{J}_{nz}\right) = E\left(\frac{h_{0t}}{h_t}\right)E(z_t|t\in\mathcal{J}_{nz}) = \frac{1}{\pi_0}E\frac{h_{0t}}{h_t}$. Hence, $\frac{1}{n}L_{2n} \to_p L_2$ uniformly if $\sup\left|(\hat{\pi}-\pi_0)\frac{r_t}{h_t}\right| \to_p 0$, $E|\ln h_t| < \infty$ and $E|\frac{h_{0t}}{h_t}| < \infty$. The assumed stationarity and finite moments ensure that $E|\ln h_t| < \infty$ and $E|\frac{h_{0t}}{h_t}| < \infty$. As demonstrated in Lemma C.1, $\hat{\pi} \to_p \pi_0$. Consequently, the condition that $\sup\left|(\hat{\pi}-\pi_0)\frac{r_t}{h_t}\right| \to_p 0$ can be easily met by a weak assumption in Theorem 5.3 below.

Due to the involvement of the estimator $\hat{\pi}$ in the likelihood function, the

QMLE in the current case converges to the true value slower than it does in the case without zero outcomes. In a fashion similar to the Location MEM(p,q) model, extended theorems of the extremum estimator in chapter 2 yield that consistency and asymptotic normality of the exponential QMLE still hold under sufficient conditions in ZAF-MEM(p,q).

DEFINITION 6. $\tilde{\boldsymbol{u}} = \arg\max_{\boldsymbol{u} \in U} L_{2n}(\boldsymbol{u})$

THEOREM 5.3. *If $Ez_t^s < \infty$ for some $s \geq 1$ and Assumption* **??** *is satisfied,*

(1) $\tilde{\boldsymbol{u}} \to_p \boldsymbol{u}_0$.

(2) $\sqrt{n}(\tilde{\boldsymbol{u}} - \boldsymbol{u}_0) \to_D N(0, V_3)$ *where* $V_3 = G_3^{-1} B_3 G_3^{-1}$ *and*

$$B_3 = E(\nabla_{\boldsymbol{u}} l_{2t}(\boldsymbol{u}_0))\nabla_{\boldsymbol{u}} l_{2t}(\boldsymbol{u}_0)'), \quad G_3 = E\nabla^2_{\boldsymbol{u}} l_{2t}(\boldsymbol{u}_0)).$$

Compared to the GF QMLE, this standard estimator is more ideal in certain empirical analyses. It not only provides a consistent estimate for the conditional mean of a variable, but also guarantees the asymptotic distribution of the estimates, and consequently, asymptotic distribution of standard residuals can be further discussed. Typical diagnostic tests based on the exponential QMLE for standardized residuals is more reliable as long as the sample size is large, which can be easily met in high-frequency data.

Simulation

The simulation study contains two steps, illustrating consistency of each estimator and comparing consistency and efficiency between them. We generate 500 replications following ZAF-MEM(2,2) in each step. The parameters for the conditional

Table 10. MLE for $\pi_0 = 0.5$

	$\hat{\omega}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}_1$	$\hat{\beta}_2$
True Value	0.03	0.05	0.05	0.4	0.4
Mean	0.030249	0.049234	0.049749	0.388496	0.408057
Median	0.030123	0.049103	0.049834	0.393548	0.406582
S.D.	0.004117	0.007942	0.010195	0.040136	0.034488

Table 11. MLE and QMLE for $\pi_0 = 0.9$

	$\hat{\omega}$	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\beta}_1$	$\hat{\beta}_2$
True Value	0.03	0.05	0.05	0.4	0.4
			Exact MLE		
Mean	0.029684	0.049613	0.049119	0.403335	0.396847
Median	0.029827	0.049799	0.04925	0.402438	0.396777
S.D.	0.002214	0.004695	0.005058	0.021476	0.013815
			QMLE		
Mean	0.029379	0.049265	0.048698	0.410924	0.39009
Median	0.029595	0.049233	0.04912	0.397885	0.401564
S.D.	0.004115	0.005265	0.011746	0.18579	0.163278

mean equation are set at $\boldsymbol{u}_0 = (0.03, 0.05, 0.05, 0.4, 0.4)$ and the density parameters are $\nu_0 = (1, 5, 3.3)$. The first step confirms the consistency of global estimator $\hat{\boldsymbol{u}}$ with replication sample size at 8000. The theoretical proportion of zero outcomes is $\pi_0 = 0.5$.

The second step not only presents the consistency of QMLE $\tilde{\boldsymbol{u}}$, but also exhibits its inefficiency compared to exact MLE $\hat{\boldsymbol{u}}$. Due to the density misspecification in $L_{2n}(\boldsymbol{u})$, a large sample size is required for $\tilde{\boldsymbol{u}}$ to converge to the true value. Thus, each generated replication contains 40000 data and $\pi_0 = 0.9$. The results displayed in Tables 10 and 11 illustrate that global MLE provides a consistent estimate for the conditional mean equation and the exponential QMLE is an appropriate semiparametric method when the size of observed data is large.

PART 3

CHAPTER 6

CROSS-LISTED SHARES BEHAVIOR IN DOMESTIC AND FOREIGN MARKETS: AN APPLICATION OF VECTOR MEM

In this part, we apply another extended version of MEM, i.e. vector MEM (vMEM) generalized in Engle and Gallo (2006) to investigate structures of different markets based on the same stocks.

The analysis in Manganelli (2005) draws three valuable conclusions about market microstructure in high-frequency trading by comparing two groups of stocks classified according to trading frequency. First, there is a significant trading volume clustering or autocorrelation effect in all stocks, and the effect is higher for frequently traded stocks. Second, the higher the trading intensity and the greater the volume per trade, the more likely new information and informed trading exist in the market. In other words, both duration and volume have lagged impact on the price variance. In addition, shorter duration follows larger price movements and greater volumes. These correlations are significant only for frequently-traded stocks. Third, it takes less time for new information to be incorporated into price when the stocks are more frequently traded.

In Manganelli (2005), trading intensity over a selected period is considered as the primary factor that affects characteristics of trading activity. Our interest arises from the fact that the location of trading may have an impact on the stocks' dynamic behavior as well. Specifically, we want to investigate the differences and similarities between the home market and the foreign market for cross-listed stocks. Many existing studies such as Gagnon and Karolyi (2010) and Chen and Choi (2012), have documented cross-border price differential and information imbalance concerning Canadian shares traded on both sides of Niagara Falls. A large set of cross-listed firms were also analyzed by Halling et al. (2013). They use a volume-based measure of multimarket trading to provide empirical evidence of how investors actively exploit multimarket environments. Our goal is to analyze multi-market trading in the sense of the three aspects discussed in Manganelli (2005). The following questions provide a framework for our research: is the clustering effect more significant in one market than the other? How long does it take before prices converge to full information values on each market? These questions are of considerable interest to any stock exchange and policy makers.

To address these issues, our research using tick-by-tick data over a 4-week sample period (April 2013) is conducted on 32 cross-listed Canadian shares traded in New York and Toronto. Clustering effect, interaction between indicators and the speed of convergence to long run equilibrium for both the NYSE and TSX are to be evaluated and compared. We adopt the econometric framework employed by Manganelli (2005) in the empirical analysis. The clustering effect can be investigated by a univariate MEM without any exogenous variables, while the interaction between duration, volume and price volatility is obtained by a multivariate system of MEM. After estimating models by a standard QMLE, we employ the Impulse Response function to compute the lagged effect that a shock to these variables has on price variance. Typical statistical tests are used to identify the difference between the two markets in each aspect.

The rest of this chapter is organized as follows. Section 2 gives the specification of a multivariate MEM and its covariance stationarity condition. Section 3 displays a recursive way to compute Impulse Response based on the estimates of a vMEM. The empirical analysis pertaining to cross-listed Canadian shares is presented in section 4.

Multivariate Model for Duration, Volume and Return

A variety of specifications have been proposed in existing research on multivariate GARCH type models (see, for example, Bauwens et al. (2006)). The Dynamic Conditional Correlation (DCC) model introduced by Engle (2002a) combines the multivariate linear system of the BEKK model in Engle and Kroner (1995) and the Constant Conditional Correlation (CCC) model in Bollerslev (1990). Engle and Gallo (2006) jointly model financial asset volatility indicators: absolute daily returns, daily high-low range and daily realized volatility by a dynamic system of MEM, which is similar to the model utilized in Manganelli (2005).

The linear system in Manganelli (2005) is not exactly a vMEM, as one of the variables is described by an ARMA-GARCH process. Since our analysis focuses on three aspects of market information: duration between two trades, volume per trade and price variance, a vMEM for non-negative processes is more appropriate. Suppose that we have observations for trade duration, volume and price return, denoted by d_t, v_t and y_t, respectively. The framework that we employ is specified as follows:

$$\begin{aligned} d_t &= \psi_t \epsilon_t \\ v_t &= \phi_t \eta_t \\ y_t &= \mu_t + \xi_t,\ \xi_t = h_t \delta_t \end{aligned} \tag{6.1}$$

where ϵ_t ~i.i.d.(1, σ_ϵ^2), η_t ~i.i.d.(1, σ_η^2) and δ_t ~i.i.d.(0,1). μ_t is described by an ARMA process and ξ_t is the volatility of price return. Equivalently, we have $E(d_t|\Omega_t) = \psi_t$, $E(v_t|\Omega_t) = \phi_t$ and $E(\xi_t|\Omega_t) = h_t$ where Ω_t is the information set by time t. Note that random errors ϵ_t, η_t and δ_t are independent of each other. The conditional expectations of d_t, v_t and ξ_t are presented as:

$$\begin{aligned} \psi_t &= a_0 + \sum_{i=1}^{p}(a_{1i}d_{t-i} + a_{2i}v_{t-i} + a_{3i}\xi_{t-i}^2) + \sum_{j=1}^{q}(a_{4j}\psi_{t-j} + a_{5i}\phi_{t-j} + a_{6i}h_{t-j}^2) \\ \phi_t &= b_0 + \sum_{i=1}^{p}(b_{1i}d_{t-i} + b_{2i}v_{t-i} + b_{3i}\xi_{t-i}^2) + b_4 d_t + \sum_{j=1}^{q}(b_{5j}\psi_{t-j} + b_{6i}\phi_{t-j} \\ &\quad + b_{7i}h_{t-j}^2) \\ h_t^2 &= c_0 + \sum_{i=1}^{p}(c_{1i}d_{t-i} + c_{2i}v_{t-i} + c_{3i}\xi_{t-i}^2) + c_4 d_t + c_5 v_t + \sum_{j=1}^{q}(c_{6j}\psi_{t-j} + c_{7i}\phi_{t-j} \\ &\quad + c_{8i}h_{t-j}^2) \end{aligned} \tag{6.2}$$

The system in (6.2) can also be represented by a vector autoregressive moving average framework in the following matrix form:

$$X_t = \gamma + \sum_{i=1}^{q} A_i X_{t-i} + \sum_{i=0}^{p} B_i \tau_{t-i} \tag{6.3}$$

where $X_t' = (\psi_t, \phi_t, \sigma_t)$ and $\tau_t' = (d_t, v_t, \xi_t)$. Note that matrix B_0 is a lower triangular matrix. As demonstrated by Manganelli (2005), this vector autoregressive (VAR) system is covariance stationary if all roots of $|I_n\lambda^p - H_1\lambda^{p-1} - H_2\lambda^{p-2} - \cdots - H_p| = 0$ are less than 1 in absolute value, where $H_i = (B_0 - I)^{-1} \times (B_i + A_i)$, for $i = 1, \cdots, p$.

Impulse Response

The VAR model specified in (6.3) allows us to compute the effect that an unexpected disturbance in the market has on future duration, volume and volatility. Specifically, we want to derive the rate at which each variable is expected to change after a shock occurs in the market.

This measure of shock effects on the behavior of a series is called the Impulse Response (IR) function. Many papers have adopted IR on macroeconomic time series. Campbell and Mankiw (1986) use IR on univariate linear models to investigate persistence of shocks in a business cycle. A generalized IR function is introduced by Koop et al. (1996) and applied to both linear and nonlinear multivariate models on U.S. output and unemployment rate. Phillips (1998) improves impulse response and forecast error variance decomposition for unrestricted VAR by reduced rank regressions. Building on Koop et al. (1996), Pesaran and Shin (1998) discuss properties of the generalized IR for unrestricted VAR and cointegrate VAR models and illustrate the results using U.S. quarterly data of investment, consumption and output. Baillie and Kapetanios (2013) concentrate on the estimation and confidence intervals for the impulse response on strongly persistent time series. They conclude that semi-parametric AR approximations are a good strategy for analyzing IR.

Manganelli (2005) explicitly derives IR for the trivariate model with order (1,1) at time $t > 0$ denoted by $\frac{\partial E[X_t|\Omega_0]}{\partial \tau_0}$. That is

$$\begin{aligned} X_t &= \gamma + AX_{t-1} + B\tau_0 + C\tau_{t-1} \\ \frac{\partial E[X_t|\Omega_0]}{\partial \tau_0} &= D^{t-1}(I-B)^{-1}(AB+C) \end{aligned} \tag{6.4}$$

where $D = (I-B)^{-1}(A+C)$.

We provide the following recursive equations computing IR for a higher order case. For simplicity, we impose $m = p = q$ and $B_0 = 0$ on the VAR system in (6.3). Then the IR can be recursively computed as

$$\begin{aligned} \frac{\partial E[X_1|\Omega_0]}{\partial \tau_0} &= B_1 \\ \frac{\partial E[X_t|\Omega_0]}{\partial \tau_0} &= \sum_{i=1}^{t-1}(A_i + B_i)\frac{\partial E[X_{t-i}\Omega_0]}{\partial \tau_0}, \ for\ 1 < t \leq m \\ \frac{\partial E[X_t|\Omega_0]}{\partial \tau_0} &= \sum_{i=1}^{m}(A_i + B_i)\frac{\partial E[X_{t-i}\Omega_0]}{\partial \tau_0}, \ for\ t > m \end{aligned} \tag{6.5}$$

We set $B_0 = 0$ because the emphasis of this research is on the lagged effect of a disturbance in the market. The correlation between variables at the same moment is not considered in the current case.

Empirical Results

Data

We apply the multivariate framework discussed in the previous section to 50 Canadian Stocks that are cross-listed in both the NYSE and TSX. The data for the NYSE are drawn from the Trades and Quotes (TAQ) database. Stocks are selected by the criteria that the number of transactions is no less than 5000 in both markets.

There are three steps in preparation of the data. First, we delete the transactions that occur before 9:30 AM and the overnight durations. Second, we adopt the standard procedure described in Engle and Russell (1998) to obtain the price for each transaction as the average of the bid and ask quotes. The returns y_t are the difference of log of the prices. Third, since there is a typical intra-daily pattern in durations and volatilities in the course of a trading day, we remove this seasonality in the series of duration, volume and price by smoothing with a piecewise regression spline at knots 10:00, 10:30,...,16:00. The adjusted series are obtained by dividing original series by the spline predictions. As expected, the durations are the shortest in the morning and prior to the close with a peak around noon, while price returns volatility is high at the opening and flat for the rest of the day. Fourth, there are three variables to be modeled by vMEM. The first two are duration and volume, but the third one is not the deseasonalized log return y_t. Because autoregressiveness in the expectation of y_t may be compounded into its volatility, we estimate μ_t, the mean of y_t, by an ARMA model and then take the residual as the third variable in vMEM. BIC (Schwarz Bayesian Information Criteria) and Ljung-Box statistics are the criteria in model selection for μ_t. If the Ljung-Box statistics for the fitted residual are significantly smaller than that for the raw data, the ARMA process successfully captures the autoregresiveness in μ_t and the model with the smallest Ljung-Box statistics is an optimal choice.

Clustering Effect

To compare the clustering effect in two markets, the autoregressive coefficient of the univariate MEM for duration, volume and volatility will be estimated. These models are independent, hence, A_i and B_i for $i \neq 0$ are assumed to be diagonal. They are presented as

$$\begin{aligned} d_t &= \psi_t \epsilon_t \\ v_t &= \phi_t \eta_t \\ y_t &= \mu_t + \xi_t,\ \xi_t = h_t \delta_t \end{aligned} \tag{6.6}$$

Table 12. β Coefficients for Each Variable

Percentile	95%	75%	*Median*	25%	5%
			Duration		
NYSE	0.9153	0.897	0.8717	0.8446	0.7873
TSX	0.9740	0.9139	0.9014	0.8716	0.8251
			Volume		
NYSE	0.9366	0.8986	0.8515	0.7526	0.4979
TSX	0.9475	0.8837	0.7748	0.6255	0.3439
			Volatility		
NYSE	0.9553	0.9328	0.9189	0.8978	0.8453
TSX	0.9543	0.9438	0.9267	0.9006	0.8455

$$\begin{aligned} \psi_t &= a_0 + a_1 d_{t-1} + a_2 \psi_{t-1} \\ \phi_t &= b_0 + b_1 v_{t-1} + b_2 \phi_{t-1} \\ h_t &= c_0 + c_1 \xi_{t-1} + c_2 h_{t-1} \end{aligned} \tag{6.7}$$

We compare the β coefficients, i.e. (a_2, b_2, c_2) of each stock in two markets by percentiles at first. As shown in Table 12, the β coefficient for volume is higher in general in the NYSE due to the higher frequency of trading, which confirms the answer to the first question in Manganelli (2005). We also employ Wilcoxon and K-S tests to verify the difference between two markets. The test hypothesis and results are shown in table 13.

Table 13. Statistical Tests on β Coefficients in Two Markets

	Duration	Volume	Volatility
Wilcoxon Test	$H_0 : M_N > M_T$	$H_0 : M_N < M_T$	$H_0 : M_N = M_T$
	p-value=8.344×10^{-05}	p-value=0.02452	p-value=0.3328
K-S Test	$H_0 : F_N$ below F_T	$H_0 : F_N$ above F_T	$H_0 : F_N = F_T$
	p-value=5.98×10^{-03}	p-value=5.98×10^{-03}	p-value=0.3959

Note: M_N and M_T in the Wilcoxon test represent the medians of β_N and β_T, while F_N and F_T in the K-S test represent the cumulative distribution functions of β_N and β_T, where N and T represent NYSE and TSX, respectively. If F_N is below F_T, then $F_N^{-1}(x) \geq F_T^{-1}(x)$ for $0 < x < 1$. That is, if $0 < P(\beta_N < x_1) = P(\beta_T < x_2) < 1$, then $x_1 > x_2$, which confirms the summarized results in Table 12.

Table 14. Coefficient for Volume in Duration Equation

	NYSE			TSX	
	Insignificant	Significant		Insignificant	Significant
Negative	8	32	Negative	7	1
Positive	6	4	Positive	14	28

Table 15. Coefficient for Volatility in Duration Equation

NYSE			TSX		
	Insignificant	Significant		Insignificant	Significant
Negative	14	21	Negative	22	3
Positive	7	8	Positive	12	13

Interaction Between Duration, Volume and Price Variance

The model described by (6.2) or (6.3) needs to be estimated prior to further analysis. Since we only consider lagged interaction at order 1, matrices B_i for $i \geq 2$ in (6.3) are set as diagonal. The standard quasi maximum likelihood in exponential form is used in the current case. The sum of all parameters in each equation of (6.2) is less than one. Model selection, i.e., the order of the equation for each variable is based on Ljung-Box statistics and BIC. We conduct inference on the coefficients in the equations of price variance and duration.Table 14 -Table 17 summarize coefficients corresponding to volume, variance in the duration equation and coefficients for duration, volume in the variance equation in both markets.

The results presented in these tables are the number of stocks with coefficients falling into each category. For example, in the NYSE, there are 32 stocks with negative coefficient for volume in the equation of duration and these coefficients are statistically significant at the 5% level. In other words, there is a significant negative lagged effect of volume disturbance on duration for 32 shares traded in the NYSE.

Table 16. Coefficient for Duration in Volatility Equation

NYSE			TSX		
	Insignificant	Significant		Insignificant	Significant
Negative	10	15	Negative	8	32
Positive	10	15	Positive	5	5

Comparing results for each coefficient in two markets, we find the following differences and similarities. First, the NYSE presents a large fraction of negative correlation between current volume and subsequent duration, while TSX exhibits a very different and abnormal picture in that this coefficient appears to be significantly positive with a big proportion among the stocks. This unexpected difference in the duration equation cannot be confirmed by the conclusion in Manganelli (2005) and implies an open research question about the market policy and microstructure of TSX. Second, the impact of volatility change on expected duration in the next trade is prone to be negative in the NYSE and could be either positive or negative in the TSX, but neither of the two markets shows a predominant proportion of such impact.

Third, although greater activities (or volumes) coincide with higher subsequent price volatility in both markets, positive serial correlation between movement of trading intensity and change in price volatility is only found in TSX. Furthermore, the interaction between current volume and future price change is more significant in TSX, compared to the NYSE. This striking contrast can be induced by the difference in frequency, i.e. number of observations of each stock between two markets,

Table 17. Coefficient for Volume in Volatility Equation

NYSE			TSX		
	Insignificant	Significant		Insignificant	Significant
Negative	1	3	Negative	2	0
Positive	7	39	Positive	0	48

according to the statement in Manganelli (2005) that correlation between times of bigger activities and higher probability of an informed trader is significant only for frequently traded stocks.

Impulse Reponse and Time to Converge

For each stock, the time for a perturbation to be absorbed is computed by impulse response and the average duration of transactions within the selected period. First, we calculate the empirical impulse response at each future transaction i.e. when $t > 0$ after a shock to $\tau_t = (d_t, v_t, \xi_t)$ at $t = 0$, using the recursive formula in (6.5).The third row of the IR matrix is the impulse response or the speed at which price variance changes after shocks to duration, volume and volatility occur. When the value of these elements is approaching zero, i.e. less than the threshold value 10^{-14}, the disturbance is considered to be completely absorbed. Second, we use the unconditional mean of duration by now $(t = 0)$ as an estimate for the duration between two future trades. Hence, the time for a shock to be incorporated into price is the product of average duration by now and the number of transactions that it takes for the response to fall

Table 18. Hours for Disturbance To Be Absorbed into Price

Percentile	95%	75%	*Median*	25%	5%
Duration					
NYSE	3.11	6.44	8.71	17.49	29.09
TSX	5.15	10.11	16.95	22.26	33.19
Volume					
NYSE	2.81	5.8	8.23	16.25	29.57
TSX	4.43	8.77	14.5	19.66	30.53
Volatility					
NYSE	2.74	5.56	7.84	16.1	29.35
TSX	4.43	8.45	13.96	19.79	31.39

below 10^{-14}.

As shown in Tables 18 and 19, new information seems to be absorbed faster on the NYSE than on TSX. One should note that the choice of threshold at 10^{-14} might be inappropriate for some stocks. Contrast between response time in two markets for the same stock may differ when we apply a different threshold value. For instance, Figure 6 and Table 20 illustrate that the response for stock BMO diminishes faster in TSX within the first few minutes, while it takes more time for the response in TSX to eventually converge to zero. This variation by threshold opens a door to the criteria for news being considered as completely incorporated into price.

Table 19. Statistical Tests on IR of Two Markets

	Duration	Volume	Volatility
Wilcoxson Test	$H_0 : M_N > M_T$	$H_0 : M_N < M_T$	$H_0 : M_N < M_T$
	p-value=0.0132	p-value=0.03368	p-value=0.04651
K-S Test	$H_0 : F_N$ above F_T	$H_0 : F_N$ below F_T	$H_0 : F_N$ below F_T
	p-value=0.00309	p-value=0.00598	p-value=0.00598

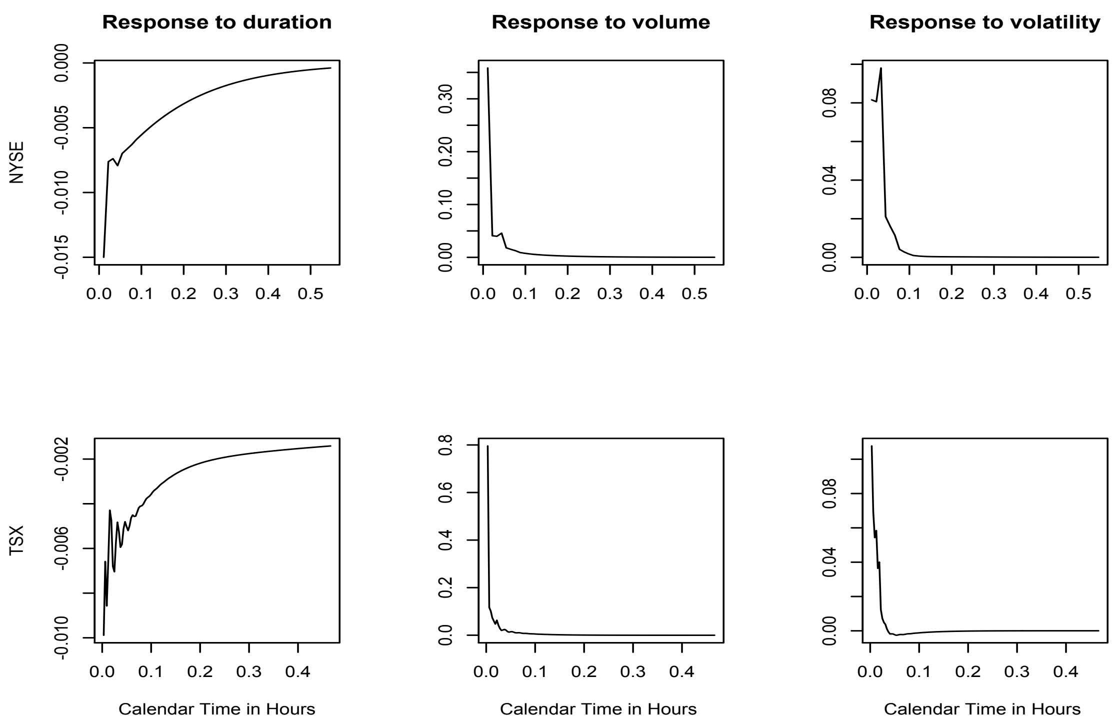

Figure 6. Response Time in Hours for Stock BMO in Two Markets

Table 20. Hours and Number of Transactions Needed for IR Convergence

	Duration	Volume	Volatility
	Time in Hours		
NYSE	4.52	4.36	4.17
TSX	22.52	20.66	19.12
	Number of Transactions		
NYSE	413	398	381
TSX	7243	6645	6148

Note: threshold of IR is 10^{-14}.

Since U.S. markets have attracted a large number of Canadian companies and increased their market share in the total traded value of Canadian cross-listed stocks, comparison between New York and Toronto in terms of impulse response lead to the question of whether the home market of Canadian shares still plays a dominant role under globalization of financial markets.

CHAPTER 7
CONCLUSION

This thesis first extends theorems of an extremum estimator, and then introduces a modified QMLE for Location MEM under various assumptions on stationarity, ergodicity and existence of moments. The location parameter in this model is separately estimated by the minimum statistic $r_{n(1)}$, which is proved to be consistent but may not be $\sqrt{n}$-consistent. As the results are built on the assumption of $r_{n(1)}$ being $\sqrt{n}$-consistent, it is crucial to investigate the validity of this condition in future work. In addition, since the asymptotic analysis for order (1,1) is conducted under mild conditions for stationarity and ergodicity, it will be of great interest to consider a nonstationary process in the future by the approach used in Jensen and Rahbek (2004a) and Jensen and Rahbek (2004b), which utilizes other convergence theorems and imposes different assumptions. Another potential direction of asymptotic inference on this model is allowing for explanatory variables (or covariates) in the dynamic component, as in GARCH-X in Han and Kristensen (2014). Similar to the investigation about GARCH-X in Han (2013), asymptotic results concerning sample autocorrelation, variance and kurtosis can be built for Location MEM as well.

For the (p,q) model, in contrast to the recursive form of the conditional mean equation, an infinite representation is employed in chapter 4 to develop asymptotic limits of the estimator. The second matrix form in this chapter, which generalizes various GARCH-type models implies that considerable attention should be paid in the future to nonlinear structures of Location MEM in the sense of Log-ACD in Bauwens and Giot (2000), TGARCH in Zakoian (1994) and AACD in Fernandes and Grammig (2006). We can also adopt an alternative methodology to capture nonlinearity in our model-the regime-switching approach or TACD-proposed by Zhang et al. (2001). Obviously, the location parameter in a regime-switching model varies along with threshold values. One should be careful while selecting the thresholds for Location MEM, because there is a possible impact of threshold values on the estimate of the location parameter, and precision in this estimate largely depends on the number of observations within each regime. Although existing results provide evidence that nonlinear structures are usually more adequate to model duration, a linear system, on the other hand, is more ideal while taking into account the interaction between variables. It would be practical to develop an estimator for a linear multivariate location MEM, in which different variables of a mark are jointly modeled. Asymptotic properties of the system of autoregressive equations can be established under weak conditions, because entry of predetermined variables in each equation excludes integrated and explosive processes.

In chapter 5, we demonstrated asymptotic properties of MLE for ZAF-MEM(p,q) with an extended extremum estimator theorem, as well as verify that exponential QMLE is still valid in this model. Further investigation on zero-augmented MEM can be summarized as follows. The ZAF MLE defined in chapter 5 actually serves as a quasi likelihood estimator, since it doesn't provide valid estimates for density parameters. An interesting extension of current work would be to adopt Weibull and Generalized Gamma as alternatives for the continuous part of the innovation term's density, and then apply the ZAF MLE to each case. We are interested in comparing exponential (Q)MLE and ZAF MLE in consistency, efficiency and asymptotic normality for each distribution. Since ZAF MLE is consistent if and only if the density function can be nested by generalized F, distributional diagnostic testing plays a crucial role in subsequent analysis of this estimator. Recent work such as Duchesne and Pacurar (2008), Chen and Hsieh (2010), Hong and Lee (2011) and Saart and Gao (2012) should be taken into account for better performance in specification tests. Moreover, as proved by current work, restrictions on the density parameters must be specified in order to establish asymptotic distributions of the ZAF MLE. Finally, although the model investigated in this thesis is restricted to be linear, analysis can be extended to nonlinear models with explanatory variables under different specifications about innovation terms, for example, a flexible component for seasonalities introduced by Brownlees and Gallo (2011) and the semiparametric regression approach in Saart et al. (2015).

In the second part of this thesis, we modeled duration, volume and price variance of Canadian cross-listed shares by a vector multiplicative framework with parameter restrictions. Clustering effect and interaction between the variables were investigated based on the estimate of specific coefficients in the models. We also employed the estimated multivariate system to compute time in hours for price to converge to a long-run equilibrium. This methodology is applied to trading data from the NYSE and TSX individually for the same stocks. We found significant evidence that higher trading frequency and better access to new information cannot guarantee the characteristics of stocks' dynamic behavior or the dominant status of the home market. Therefore, it is necessary to conduct further analysis including, but not limited to Probability of Informed Trading (PIN) in each market as proposed by Tay et al. (2009), impact of type of the industry on market microstructure, and how contrast in trading intensity or size is related to the difference in time for convergence to equilibrium between two markets.

APPENDIX A

PROOFS FOR CHAPTER 3

Proof of Theorem 3.2: This conclusion can be shown by the method in the proof of Amemiya 1985 Theorem 4.1.1. Let N be an open neighborhood around η_0, then $\bar{N}$ is a compact space. There exists a maximum of $Q(\underline{y}, (\mu_0, \eta))$ within $\bar{N}$. Let $\epsilon = Q(\mu_0, \eta_0) - \max_{\eta \in \bar{N} \bigcap \Theta_\eta} Q(\mu_0, \eta)$, and A_T be the event: $\{|T^{-1}Q_T(\hat{\mu}_T, \eta) - Q(\mu_0, \eta)| < \epsilon/2 \ for\ all\ \eta\}$. Using the method in Amemiya (1985) Theorem 4.1.1. by definition of ϵ, it is not hard to see that $A_T \Rightarrow \hat{\eta}_T \in N$. Therefore, $P\left(A_T\right) \leq P\left(\hat{\eta}_T \in N\right) \to 1$ as $T \to \infty$. $\square$

Proof of Theorem 3.3: By Taylor expansion,

$$\frac{\partial Q_T}{\partial \eta}\bigg|_{(\hat{\mu}_T, \hat{\eta}_T(\hat{\mu}_T))} = \frac{\partial Q_T}{\partial \eta}\bigg|_{(\hat{\mu}_T, \eta_0)} + \frac{\partial^2 Q_T}{\partial \eta \partial \eta'}\,|_{(\hat{\mu}_T, \eta^*)}(\hat{\eta}_T(\hat{\mu}) - \eta_0) \tag{A.1}$$

where η^* lies on the line between $\hat{\eta}_T(\hat{\mu}_T)$ and η_0. Thus, $\eta^* \to_p \eta_0$. By definition of $\hat{\eta}_T(\hat{\mu}_T)$, the left hand side of equation (A.1) equals 0,

$$\sqrt{T}(\hat{\eta}_T(\hat{\mu}_T) - \eta_0) = -\left[\frac{1}{T}\frac{\partial^2 Q_T}{\partial \eta \partial \eta'}\,|_{(\hat{\mu}_T, \eta^*)}\right]^+ \frac{1}{\sqrt{T}}\frac{\partial Q_T}{\partial \eta}\bigg|_{(\hat{\mu}_T, \eta_0)} \tag{A.2}$$

where + denotes the Moore-Penrose generalized inverse. By assumption (B),

$$\frac{1}{T}\frac{\partial^2 Q_T(\mu, \eta)}{\partial \eta \partial \eta'}\bigg|_{(\hat{\mu}_T, \eta^*)} \to_p A(\theta_0) \tag{A.3}$$

$\frac{1}{\sqrt{T}}\frac{\partial Q_T}{\partial \eta}\bigg|_{(\hat{\mu}_T, \eta_0)} - \frac{1}{\sqrt{T}}\frac{\partial Q_T}{\partial \eta}\bigg|_{(\mu_0, \eta_0)} \to_p 0$ by the second condition in assumption (A). By assumption (C),

$$\frac{1}{\sqrt{T}}\frac{\partial Q_T(\mu, \eta)}{\partial \eta}\bigg|_{\theta_0} \to N[0, B(\theta_0)] \tag{A.4}$$

Therefore, by Slutsky's Theorem, $\frac{1}{\sqrt{T}}\frac{\partial Q_T}{\partial \eta}\bigg|_{(\hat{\mu}_T, \eta_0)} \to N[0, B(\theta_0)]$. By repeated application of Slutsky's Theorem, we have $\sqrt{T}(\hat{\eta}_T(\hat{\mu}_T) - \eta_0) \to N[0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}]$

$\square$

LEMMA A.1. *Under Assumption 3.1 :*

(1) h_{0t} *is strictly stationary and ergodic for all t.*

(2) $l_t(\mu_0, \eta)$ *and its first and second derivatives are strictly stationary and ergodic for all* $\eta \in \Theta$.

(3) For some $p_2 > 0$, *and all* $\eta \in \Theta$, $E|h_t(\mu_0, \eta)|^{p_2} \leq H < \infty$

(4) $\sup_{\eta \in \Theta} |L_n(\mu_0, \eta) - L_n^*(\mu_0, \eta)| \to_p 0$.

(5)$L_n(\mu_0, \eta) \to_p L(\mu_0, \eta) = E|l_t(\mu_0, \eta)|$ *for all* $\eta \in \Theta_2$.

(6) $\sup_{\eta \in \Theta_2} E|\nabla l_t(\mu_0, \eta)| < \infty$.

Proofs for Lemma A.1 and Lemma A.4 are omitted below, since Location MEM(1,1) is reduced to MEM(1,1) at $\mu = \mu_0$ and results in Lee and Hansen (1994) can be directly applied.

This lemma, especially parts (2) and (3) are of great importance for proofs of consistency and normality. They demonstrate an important condition for WLLN of stationary and ergodic processes, as well as explain why a nonstationary process is excluded by Assumption 3.1 (4). By Theorem 3.2, in order to establish consistency, we need to show $L_n^*(r_{n(1)}, \eta) \to_p L(\mu_0, \eta)$ uniformly in $\eta \in \Theta_2$ and that $L(\mu_0, \eta)$ is uniquely maximized at η_0. Before adopting the method in the proof of Lee and Hansen (1994) Theorem 1, we have to prove one more proposition to demonstrate local consistency.

Proof of Proposition 3.1: For any $x > 0$,

$$P(r_{n(1)} - \mu_0 > x) = P(h_{0t} z_t > x \ for \ 1 \leq t \leq n) = P\left(z_t > \frac{x}{h_{0t}} \ for \ 1 \leq t \leq n\right) \tag{A.5}$$

For a fixed positive integer N, if $z_t \geq C$ for $N < t \leq n$,

$$\begin{aligned} h_{0n+1} &\geq \omega_0(1-\beta_0) + (\alpha_0 C + \beta_0) h_{0n} \\ &\geq (\alpha_0 C + \beta_0)^{n+1-N} h_{0N} \\ &\geq \omega_0 (\alpha_0 C + \beta_0)^{n+1-N} \end{aligned} \tag{A.6}$$

Let A_n be the event: $\{$there exists N s.t. $z_t \geq C \; for \; N < t \leq n\}$ and $B = \alpha_0 C + \beta_0$. Then $P(A_n) \leq P\left(h_{0n+1} \geq \omega_0 B^{n+1-N}\right) \leq P\left(h_{0n+1}^{\delta} \geq \omega_0^{\delta} B^{(n+1-N)\delta}\right) \leq \frac{E h_{0n+1}^{\delta}}{\omega_0^{\delta} B^{(n+1-N)\delta}}$.

By lemma A.1 (3), when $\delta = p_2$ and $\alpha_0 C + \beta_0 > 1$, we have $\lim_{n\to\infty} P(A_n) = 0$.

Let $C = \min\{\frac{x}{h_{0t}}\}_{t=1}^{n}$, then $\alpha_0 C + \beta_0 > 1$ when x is small enough.

$$P\left(z_t > \frac{x}{h_{0t}} \; for \; 1 \leq t \leq n\right) \leq P(A_n) \tag{A.7}$$

Hence, $\lim_{n\to\infty} P\left(z_t > \frac{x}{h_{0t}} \; for \; 1 \leq t \leq n\right) = 0$. That is $\lim_{n\to\infty} P(r_{n(1)} - \mu_0 > x) = 0$. □

LEMMA A.2. *Under Assumption 3.1,* $\sup_{\eta\in\Theta_2} |L_n^*(r_{n(1)}, \eta) - L_n^*(\mu_0, \eta)| \to_p 0$.

Proof of Lemma A.2: Because

$$\mid L_n^*(r_{n(1)}, \eta) - L_n^*(\mu_0, \eta) \mid \leq \tfrac{1}{n} \sum_{t=1}^{n} \left(\left| \ln \frac{h_t^*(r_{n(1)}, \eta)}{h_t^*(\mu_0, \eta)} \right| + \left| \frac{e_t}{h_t^*(r_{n(1)}, \eta)} - \frac{\epsilon_t}{h_t^*(\mu_0, \eta)} \right| \right)$$

and

$$\begin{aligned} \left| \ln \frac{h_t^*(r_{n(1)}, \eta)}{h_t^*(\mu_0, \eta)} \right| &= \left| \ln \left(1 + \frac{h_t^*(r_{n(1)}, \eta) - h_t^*(\mu_0, \eta)}{h^*(\mu_0, \eta)} \right) \right| \\ &\leq \left| \frac{h_t^*(r_{n(1)}, \eta) - h_t^*(\mu_0, \eta)}{h^*(\mu_0, \eta)} \right| \\ &= \left| \frac{\alpha \sum_{k=0}^{n} \beta^k (\mu_0 - r_{n(1)})}{h_t^*(\mu_0, \eta)} \right| \\ &\leq \frac{\alpha_u (r_{(1)} - \mu_0)}{\omega_l (1 - \beta_u)} \end{aligned} \tag{A.8}$$

$$\begin{aligned}
\left|\frac{e_t}{h_t^*(r_{n(1)},\eta)}-\frac{\epsilon_t}{h_t^*(\mu_0,\eta)}\right| &= \left|\frac{\epsilon_t+\mu_0-r_{n(1)}}{h_t^*(r_{n(1)},\eta)}-\frac{\epsilon_t}{h_t^*(\mu_0,\eta)}\right| \\
&\leq \epsilon_t\frac{|h_t^*(\mu_0,\eta)-h_t^*(r_{n(1)},\eta)|}{h_t^*(r_{n(1)},\eta)h_t^*(\mu_0,\eta)}+\frac{|r_{n(1)}-\mu_0|}{h_t^*(r_{n(1)},\eta)} \\
&\leq \epsilon_t\alpha\sum_{k=0}^{\infty}\beta^k\frac{(r_{n(1)}-\mu_0)}{\omega^2}+\frac{r_{n(1)}-\mu_0}{\omega} \\
&\leq \frac{\alpha\epsilon_t(r_{n(1)}-\mu_0)}{\omega^2(1-\beta)}+\frac{r_{n(1)}-\mu_0}{\omega}
\end{aligned} \tag{A.9}$$

then we have

$$\left|L_n^*(r_{n(1)},\eta)-L_n^*(\mu_0,\eta)\right| \leq \frac{1}{n}\sum_{t=1}^{n}\frac{h_{t+1}(\mu_0,\eta)(r_{n(1)}-\mu_0)}{\omega_l^2(1-\beta_u)}+K(r_{n(1)}-\mu_0) \tag{A.10}$$

where $K=\dfrac{\alpha_u}{\omega_l^2(1-\beta_u)}+\dfrac{1}{\omega_l}$. By Lemma A.1(3), there exists some $p_2>0$ and some $H>0$ such that, for all $\eta\in\Theta$, $||h_t(\mu_0,\eta)||\leq H^{1/p_2}$. Then

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{t=1}^{n}h_{t+1}(\mu_0,\eta)(r_{(1)}-\mu_0)\right\|_{p_2/2} &\leq \left\|\frac{1}{n}\sum_{t=1}^{n}h_{t+1}(\mu_0,\eta)\right\|_{p_2}||r_{n(1)}-\mu_0||_{p_2} \\
&\leq \frac{1}{n}\sum_{t=1}^{n}||h_{t+1}(\mu_0,\eta)||_{p_2}||r_{n(1)}-\mu_0||_{p_2} \\
&\leq H^{1/p_2}||r_{n(1)}-\mu_0||_{p_2}
\end{aligned} \tag{A.11}$$

By Assumption 3.2 , $||r_{n(1)}-\mu_0||_{p_2}\to 0$. □

LEMMA A.3. *Under Assumption 3.1 and 3.2, the following uniform convergence is obtained* $\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{n}\left(\nabla_\eta l_t^*(r_{n(1)},\eta_0)-\nabla_\eta l_t^*(\mu_0,\eta_0)\right)\right| \to_p 0$

Proof of Lemma A.3: Let $h_{\eta t}^*(\mu,\eta)=\dfrac{\nabla_\eta h_t^*(\mu,\eta)}{h_t^*(\mu,\eta)}$ and $g_0=\mu_0-r_{n(1)}$.

$$\nabla_\eta l_t^*(r_{n(1)},\eta_0)-\nabla_\eta l_t^*(\mu_0,\eta_0) \leq |h_{\eta t}^*(r_{n(1)},\eta_0)-h_{\eta t}^*(\mu_0,\eta_0)|+\left|\frac{e_t h_{\eta t}^*(r_{n(1)},\eta_0)}{h_t^*(r_{n(1)},\eta_0)}-\frac{\epsilon_t h_{\eta t}^*(\mu_0,\eta_0)}{h_t^*(\mu_0,\eta_0)}\right| \tag{A.12}$$

Let $X_1 = |h^*_{\eta t}(r_{n(1)},\eta) - h^*_{\eta t}(\mu_0,\eta)|$ and $X_2 = \left|\frac{e_t h^*_{\eta t}(r_{n(1)},\eta)}{h^*_t(r_{n(1)},\eta)} - \frac{\epsilon_t h^*_{\eta t}(\mu_0,\eta)}{h^*_t(\mu_0,\eta)}\right|$. Since $h^*_t(\theta) = \omega + \alpha\sum_{k=0}^{t-2}\beta^k\ e_{t-1-k}$, then $\nabla_\eta h^*_t(\theta) = \begin{pmatrix} 1 \\ (h^*_t - \omega)/\alpha \\ \alpha\sum_{k=1}^{t-2} k\beta^{k-1}e_{t-1-k}\end{pmatrix}$. It is easy to show that $X_1 \leq C_0|r_{n(1)} - \mu_0|$ for $\eta = (\omega,\alpha)$, where $C_0 = \frac{1}{\alpha\omega(1-\beta)}\max\{1, 1/\omega\}$. For $\eta = \beta$,

$$\begin{aligned} X_1 &\leq \alpha\sum_{k=1}^{t-2} k\beta^{k-1}\left|\frac{e_{t-1-k}}{h^*_t(r_{n(1)})} - \frac{\epsilon_{t-1-k}}{h^*_t(\mu_0)}\right| \\ &\leq \alpha\sum_{k=1}^{t-2} k\beta^{k-1}\left[e_{t-1-k}\left|\frac{1}{h^*_t(r_{n(1)})} - \frac{1}{h^*_t(\mu_0)}\right| + \frac{|r_{n(1)} - \mu_0|}{h^*_t(\mu_0)}\right] \\ &\leq \alpha\sum_{k=1}^{\infty} k\beta^{k-1}\left[\left(\frac{(\epsilon_{t-1-k} + \mu_0)}{\alpha\omega^2(1-\beta)} + \frac{1}{\omega}\right)|r_{n(1)} - \mu_0|\right] \end{aligned} \tag{A.13}$$

By Lyapunov's Inequality, for $0 < p < \min\{1, p_2\}, E(z^p_t|\mathscr{F}_{t-1}) \leq (E(z_t|\mathscr{F}_{t-1}))^p = 1$. $\|\epsilon_t\|^p_p = E[h^p_{0t}E[z^p_t|\mathscr{F}_{t-1}]] \leq E|h^p_{0t}| \leq H$. Thus, it is not hard to show that $\|e_t\|_p < \infty$ and $\|h^*_{\eta t}(\theta)\|_p < \infty$. Let $S = \alpha\sum_{k=1}^{\infty} k\beta^{k-1}$ and $p = \min\{1, p_1, p_2\}$, then

$$\|X_1\|_{p/2} \leq S\left[\left(\frac{\|\epsilon_t\|_p + \mu_0)}{\alpha\omega^2(1-\beta)} + \frac{1}{\omega}\right)\|r_{n(1)} - \mu_0\|_p\right] \leq S\left(\frac{H_p^{1/p} + \mu_0}{\alpha\omega^2(1-\beta)} + \frac{1}{\omega}\right)\|r_{n(1)} - \mu_0\|_p \tag{A.14}$$

Because

$$\begin{aligned} X_2 &\leq e_t\left|\frac{h^*_{\eta t}(r_{n(1)})}{h^*_t(r_{n(1)})} - \frac{h^*_{\eta t}(\mu_0)}{h^*_t(\mu_0)}\right| + \frac{h^*_{\eta t}(\mu_0)}{h^*_t(\mu_0)}|r_{n(1)} - \mu_0| \\ &\leq e_t\left[\left|\frac{h^*_{\eta t}(r_{n(1)})}{h^*_t(r_{n(1)})} - \frac{h^*_{\eta t}(r_{n(1)})}{h^*_t(\mu_0)}\right| + \left|\frac{h^*_{\eta t}(r_{n(1)}) - h^*_{\eta t}(\mu_0)}{h^*_t(\mu_0)}\right|\right] + \frac{h^*_{\eta t}(\mu_0)}{h^*_t(\mu_0)}|r_{n(1)} - \mu_0| \end{aligned} \tag{A.15}$$

$$\begin{aligned} \|X_2\|_{p/4} &\leq \|e_t\|_{p/2}\left(\tfrac{\alpha}{\omega^2(1-\beta)}\|h^*_{\eta t}(r_{n(1)})\|_p\|r_{n(1)} - \mu_0\|_p + \tfrac{1}{\omega}\|X_1\|_{p/2}\right) \\ &+ \tfrac{\|h^*_{\eta t}(\mu_0)\|_{p/2}}{\omega}\|r_{n(1)} - \mu_0\|_{p/2} \end{aligned} \tag{A.16}$$

By Assumption 3.2, $\sqrt{n}\|r_{n(1)} - \mu_0\|_p \to_0$. This lemma is proved. $\square$

LEMMA A.4. *Under Assumptions 3.1 and 3.2:*

(1) $\frac{1}{\sqrt{n}} C_0^{-1/2} \sum_{t=1}^n \nabla_\eta l_t^*(\theta_0) \Rightarrow N(0, C_0)$, *where* $C_0 = E(\nabla_\eta l_t(\theta_0)\nabla_\eta l_t(\theta_0)')$.

(2) $\sup_{\eta\in\Theta_4} |\hat{G}_n(\mu_0,\eta) - G(\mu_0,\eta)| \to_p 0$ *and* $G(\mu_0,\eta)$ *is continuous in* Θ_4.

(3) $\sup_{\eta\in\Theta_6} \left|\hat{C}_n(\mu_0,\eta) - C(\mu_0,\eta)\right| \to_p 0$.

Proof of Theorem 3.4: (1) First, Θ_2 is compact. Second, Lemma A.1 (5) yields $L_n(\mu_0,\eta) \to_p L(\mu_0,\eta)$ pointwise. Third, Lemma A.1 (6) implies that $L_n(\mu_0,\eta)$ satisfies the weak Lipschitz condition in Andrews (1992). By Theorem 4.3 in that paper, $L_n(\mu_0,\eta) \to_p L(\mu_0,\eta)$ uniformly in Θ_2 and $L(\mu_0,\eta)$ is continuous in Θ_2.

Lemma A.1 (4) and the fact

$$\sup_{\eta\in\Theta_2} |L_n^*(\mu_0,\eta) - L(\mu_0,\eta)| \leq \sup_{\eta\in\Theta_2} |L_n^*(\mu_0,\eta) - L_n(\mu_0,\eta)| + \sup_{\eta\in\Theta_2} |L_n(\mu_0,\eta) - L(\mu_0,\eta)|,$$

give that $L_n^*(\mu_0,\eta) \to_p L(\mu_0,\eta)$ uniformly. Combined with Lemma A.2, it is easy to see that $L_n^*(r_{n(1)},\eta) \to_p L(\mu_0,\eta)$.

Lumsdaine (1996), Theorem 1, showed that the limiting likelihood $L(\theta)$ for GARCH(1,1) is uniquely maximized at θ_0. In the present case, this statement still holds when we fix μ to be μ_0 and the proof carries over as follows.

The maximization of $L(\mu_0,\eta)$ over Θ_2 is equivalent to $\max_{\theta\in\Theta_2} (L(\mu_0,\eta) - L(\theta_0))$.

$$E(l_t(\theta)) - E(l_t(\theta_0)) = -E\left(\ln\frac{h_t}{h_{0t}}\right) - E\left(\frac{e_t}{h_t} - \frac{\epsilon_t}{h_{0t}}\right).$$

Because $e_t = \epsilon_t + \mu_0 - \mu$,

$$E\left(\frac{e_t}{h_t} - \frac{\epsilon_t}{h_{0t}}\right) = E\left(\frac{h_{0t}z_t + \mu_0 - \mu}{h_t}\right) - E(z_t) = E\left(\frac{h_{0t}}{h_t}\right) + E\left(\frac{\mu_0 - \mu}{h_t}\right) - 1,$$

$$E(l_t(\mu_0,\eta)) - E(l_t(\theta_0)) = E\left(\ln\frac{h_{0t}}{h_t(\mu_0,\eta)} - \frac{h_{0t}}{h_t(\mu_0,\eta)}\right) + 1.$$

Let $x = \dfrac{h_{0t}}{h_t(\mu_0, \eta)}$. Then $E\left(l_t(\mu_0, \eta)\right) - E\left(l_t(\theta_0)\right) = E\left(\ln x - x\right) + 1$.
It is easy to show that $g(x) = \ln x - x$ is maximized at $x = 1$.That is $h_t(\mu_0, \eta) = h_{0t}$. Similar to the proof of Lumsdaine 1996 Theorem 1, using the Mean Value Theorem and Lemma 5 in that paper, η_0 is the unique global maximizer of $(L(\mu_0, \eta) - L(\theta_0))$ in Θ_2, which satisfies the second part of Theorem 3.2(C). By Theorem 3.2, it can be directly shown that $\hat{\eta}_n(r_{n(1)}) \to_p \eta_0$.

(2) Theorem 3.3 has established the standard conditions for asymptotic normality in nonlinear estimation. $\hat{\eta}_n(r_{n(1)})$ can be proved to be consistent by Theorem 3.2. First, Lemma A.3 established the condition in Theorem 3.3 (A). Second, $\dfrac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla_\eta l_t^*(\theta_0) \to_D N(0, C_0)$ by Lemma A.4 (1). Third, because $\nabla_\eta^2 l_t^*(\mu, \eta)$ is continuous in μ, and $r_{n(1)} \to_p \mu_0$, $\hat{G}_n(r_{n(1)}, \eta) - \hat{G}_n(\mu_0, \eta) \to_p 0$ for $\eta \in \Theta_4$. Fourth, for any η_n^* satisfying $\eta_n^* \to \eta_0$, $\hat{G}_n(\mu_0, \eta_n^*) - G(\mu_0, \eta_n^*) \to_p 0$, by Lemma A.4(2). Fifth, since $G(\mu_0, \eta)$ is continuous in Θ_4, $G(\mu_0, \eta_n^*) \to_p G(\theta_0)$. Then $\hat{G}_n(r_{n(1)}, \eta_n^*) \to G_0 = G(\theta_0)$.

(3) By the consistency of $\hat{\theta}_n = (r_{n(1)}, \hat{\eta}_n(r_{n(1)}))$ and continuity of $\hat{C}_n(\theta)$ for any θ, $\hat{C}_n(\hat{\theta}_n) - \hat{C}_n(\theta_0) \to_p 0$. Combined with part (1), it is obvious that $\hat{C}_n(\hat{\theta}_n) \to_p C(\theta_0)$. In the proof for the second statement of this theorem, it's already shown that $\hat{G}_n \to G_0$. Therefore, $\hat{V}_n \to_p V_0$. $\square$

Proof of Proposition 3.6: The conditional distribution of z_t given the past information only depends on ϕ^2.The first partial derivatives of L_n^* w.r.t. μ, ω are

$$\frac{\partial n L_n^*}{\partial \mu} = -\sum_{t=1}^{n} \frac{1}{h_t^*} \frac{\partial h_t}{\partial \mu} \left(1 - \frac{r_t - \mu}{h_t^*}\right) + \sum_{t=1}^{n} \frac{1}{h_t^*},$$

$$\frac{\partial n L_n^*}{\partial \omega} = -\sum_{t=1}^{n} \frac{1}{h_t^*} \frac{\partial h_t^*}{\partial \omega} \left(1 - \frac{r_t - \mu}{h_t^*}\right).$$

Since $\dfrac{\partial h_t^*}{\partial \omega} = 1,\ \dfrac{\partial h_t^*}{\partial \mu} = -\alpha \sum_{k=0}^{t-2} \beta^k = -\dfrac{\alpha(1-\beta^{t-1})}{1-\beta}$,

$$\frac{\partial n L_n^*}{\partial \mu} = \sum_{t=1}^{n} \frac{\alpha(1-\beta^{t-1})}{(1-\beta)h_t^*}\left(1 - \frac{r_t - \mu}{h_t^*}\right) + \sum_{t=1}^{n} \frac{1}{h_t^*},$$

$$\frac{\partial n L_n^*}{\partial \omega} = -\sum_{t=1}^{n} \frac{1}{h_t^*}\left(1 - \frac{r_t - \mu}{h_t^*}\right).$$

For any μ, if $\dfrac{\partial n L_n^*}{\partial \omega} = 0$, then $\dfrac{\partial n L_n^*}{\partial \mu} = \sum_{t=1}^{n} \dfrac{1}{h_t^*}\left(1 - \dfrac{\alpha\beta^{t-1}}{1-\beta}\right) + \sum_{t=1}^{n} \dfrac{\alpha\beta^{t-1}}{1-\beta}\dfrac{r_t-\mu}{h_t^*}$.

Because $1 - \dfrac{\alpha\beta^{t-1}}{1-\beta} > 1 - \dfrac{\alpha}{1-\beta}$, when $\alpha + \beta < 1,\ 1 - \dfrac{\alpha}{1-\beta} > 0$,

then $\dfrac{\partial n L_n^*}{\partial \mu} > \sum_{t=1}^{n} \dfrac{\alpha\beta^{t-1}}{1-\beta}\dfrac{r_t-\mu}{h_t^*} > 0$. $\square$

APPENDIX B

PROOFS FOR CHAPTER 4

LEMMA B.1. *Under Assumption 4.1,* $r_{n(1)} \to_p \mu_0$.

Proof of Lemma B.1: For any $x > 0$,

$$P(r_{n(1)} - \mu_0 > x) = P(h_{0t}z_t > x \ for \ 1 \le t \le n) = P\left(z_t > \frac{x}{h_{0t}} \ for \ 1 \le t \le n\right) \tag{B.1}$$

For a fixed integer N, if $z_t \ge C$ for $N < t \le n$,

$$\begin{aligned} h_{0n+1} &= c_0(u) + \sum_{i=1}^{\infty} c_i(u) h_{0n+1-i} z_t \\ &\ge c_1(u) h_{0n} z_n \\ &\ge (c_1(u)C)^{n+1-N} h_{0N} \\ &\ge (c_1(u)C)^{n+1-N} c_0(u) \end{aligned} \tag{B.2}$$

Let A_n be the event: $\{$there exist N s.t. $z_t \ge C \ for \ N < t \le n\}$. Then

$$\begin{aligned} P(A_n) &\le P\left(h_{0n+1} \ge (c_1(u)C)^{n+1-N} c_0(u)\right) \\ &= P\left(h^s_{0n+1} \ge ((c_1(u)C)^{(n+1-N)s} c_0(u)^s\right) \\ &\le \frac{Eh^s_{0n+1}}{((CC_1\underline{u})^{(n+1-N)s} C_1^s} \end{aligned} \tag{B.3}$$

Since $Eh^s_{0n+1} < \infty$, $\lim_{n\to\infty} P(A_n) = 0$. The rest of the proof is identical to that for Location MEM(1,1). □

LEMMA B.2. *Under Assumptions 4.1, 4.4 and 4.5, the following uniform convergence is yielded* $\sup_{\eta \in \Theta} |L_n^*(r_{n(1)}, \boldsymbol{u}) - L_n^*(\mu_0, \boldsymbol{u})| \to_p 0$.

Proof of Lemma B.2:

$$\left|L_n^*(r_{n(1)}, \boldsymbol{u}) - L_n^*(\mu_0, \boldsymbol{u})\right| \le \frac{1}{n} \sum_{t=1}^{n} \left(\left|\ln \frac{h_t^*(r_{n(1)}, \boldsymbol{u})}{h_t^*(\mu_0, \boldsymbol{u})}\right| + \left|\frac{e_t}{h_t^*(r_{n(1)}, \boldsymbol{u})} - \frac{\epsilon_t}{h_t^*(\mu_0, \boldsymbol{u})}\right|\right) \tag{B.4}$$

$$\begin{aligned}
\left|\ln \frac{h_t^*(r_{n(1)},\boldsymbol{u})}{h_t^*(\mu_0,\boldsymbol{u})}\right| &= \left|\ln\left(1+\frac{h_t^*(r_{n(1)},\boldsymbol{u})-h_t^*(\mu_0,u)}{h^*(\mu_0,\boldsymbol{u})}\right)\right| \\
&\leq \left|\frac{h_t^*(r_{n(1)},\boldsymbol{u})-h_t^*(\mu_0,\boldsymbol{u})}{h^*(\mu_0,\boldsymbol{u})}\right| \\
&= \left|\frac{\sum_{i=0}^{t-1} c_i(\boldsymbol{u})(r_{n(1)}-\mu_0)}{c_0(\boldsymbol{u})}\right|
\end{aligned} \tag{B.5}$$

By Proposition 4.5 (1), $\displaystyle\sum_{i=1}^{t-1}\frac{c_i(\boldsymbol{u})}{c_0(\boldsymbol{u})} \leq \sum_{i=1}^{t-1}\frac{C_2\rho_0^{i/q}}{C_1} \leq \frac{C_0}{C_1}$, where $C_0 = \displaystyle\sum_{i=1}^{t-1} C_2\rho_0^{i/q}$.

Hence, $\frac{1}{n}\displaystyle\sum_{t=1}^{n}\left|\ln\frac{h_t^*(r_{n(1)},\boldsymbol{u})}{h_t^*(\mu_0,\boldsymbol{u})}\right| \to_p 0$.

$$\begin{aligned}
\left|\frac{e_t}{h_t^*(r_{n(1)},\boldsymbol{u})}-\frac{\epsilon_t}{h_t^*(\mu_0,\boldsymbol{u})}\right| &= \left|\frac{\epsilon_t+\mu_0-r_{n(1)}}{h_t^*(r_{n(1)},\boldsymbol{u})}-\frac{\epsilon_t}{h_t^*(\mu_0,\boldsymbol{u})}\right| \\
&\leq \epsilon_t\frac{|h_t^*(\mu_0,\boldsymbol{u})-h_t^*(r_{n(1)},\boldsymbol{u})|}{h_t^*(r_{n(1)},\boldsymbol{u})h_t^*(\mu_0,\boldsymbol{u})}+\frac{|r_{n(1)}-\mu_0|}{h_t^*(r_{n(1)},\boldsymbol{u})} \\
&\leq \epsilon_t\sum_{i=1}^{t-1} c_i(\boldsymbol{u})\frac{|r_{n(1)}-\mu_0|}{C_1^2}+\frac{|r_{n(1)}-\mu_0|}{C_1} \\
&\leq \frac{C_0\epsilon_t|r_{n(1)}-\mu_0|}{C_1^2}+\frac{|r_{n(1)}-\mu_0|}{C_1}
\end{aligned} \tag{B.6}$$

It suffices to show $\epsilon_t|r_{n(1)}-\mu_0| \to_p 0$. By Assumption 4.1(c), $E\epsilon_t^s < \infty$, so

$$||\epsilon_t(r_{n(1)}-\mu_0)||_{p_1/2} \leq ||\epsilon_t||_{p_1/2}||r_{n(1)}-\mu_0||_{p_1/2} \to 0 \tag{B.7}$$

$\square$

LEMMA B.3. *Under Assumptions 4.1, 4.4 and 4.5,* $\sqrt{n}\left|\nabla_u L_n^*(r_{n(1)},\boldsymbol{u})-\nabla_\eta L_n^*(\mu_0,\boldsymbol{u})\right|$ $\to_p 0$, *that is* $\frac{1}{\sqrt{n}}\left|\displaystyle\sum_{t=1}^{n}\left(\nabla_u l_t^*(r_{n(1)},\boldsymbol{u})-\nabla_\eta l_t^*(\mu_0,\boldsymbol{u})\right)\right| \to_p 0$

Proof of Lemma B.3: Let

$$\begin{aligned}
X_1 &= |h_{ut}^*(r_{n(1)},\boldsymbol{u})-h_{ut}^*(\mu_0,\boldsymbol{u})| \\
X_2 &= \left|\frac{e_t h_{ut}^*(r_{n(1)},\boldsymbol{u})}{h_t^*(r_{n(1)},\boldsymbol{u})}-\frac{\epsilon_t h_{ut}^*(\mu_0,\boldsymbol{u})}{h_t^*(\mu_0,\boldsymbol{u})}\right|
\end{aligned} \tag{B.8}$$

where $h^*_{ut}(\mu, \boldsymbol{u}) = \dfrac{\nabla_u h^*_t(\mu, \boldsymbol{u})}{h^*_t(\mu, \boldsymbol{u})}$.

According to the proof of this result for Location MEM(1,1), it suffices to show that there exists $0 < p < 1$ s.t. the p-th moment of X_1 and X_2 approach zero faster than $1/\sqrt{n}$.

$$\begin{aligned} X_1 &\leq \left|\frac{\nabla_u h^*_t(r_{n(1)}, \boldsymbol{u})}{h^*_t(r_{n(1)}, \boldsymbol{u})} - \frac{\nabla_u h^*_t(\mu_0, \boldsymbol{u})}{h^*_t(r_{n(1)}, \boldsymbol{u})}\right| + \left|\frac{\nabla_u h^*_t(\mu_0, \boldsymbol{u})}{h^*_t(r_{n(1)}, u)} - \frac{\nabla_u h^*_t(\mu_0, \boldsymbol{u})}{h^*_t(\mu_0, \boldsymbol{u})}\right| \\ &\leq \left|\sum_{i=1}^{t-1} \frac{c'_i(\boldsymbol{u})}{C_1}\right| (r_{n(1)} - \mu_0) + |\nabla_u h^*_t(\mu_0, \boldsymbol{u})| \sum_{i=1}^{t-1} \frac{c_i(\boldsymbol{u})}{C_1^2} (r_{n(1)} - \mu_0) \end{aligned} \tag{B.9}$$

$\sum_{i=1}^{t-1} \frac{c'_i(u)}{C_1}$ and $\sum_{i=1}^{t-1} \frac{c_i(u)}{C_1^2}$ are finite by Proposition 4.5(1) and (2). In order to bound the pth moment of X_1, let $p = s_0/2$. Then it suffices to show

$$||\nabla_u h^*_t(\mu_0, \boldsymbol{u})||_{s_0} < \infty \tag{B.10}$$

Since $\nabla_u h^*_t(\mu_0, \boldsymbol{u}) = c'_0(u) + \sum_{i=1}^{t-1} c'_i(u)\epsilon_t$, equation (B.10) is easily proved by Proposition 4.5 (2).

$$\begin{aligned} X_2 &\leq \epsilon_t \left|\frac{h^*_{ut}(r_{n(1)}, \boldsymbol{u})}{h^*_t(r_{n(1)})} - \frac{h^*_{ut}(\mu_0, \boldsymbol{u})}{h^*_t(\mu_0, \boldsymbol{u})}\right| + \frac{h^*_{ut}(r_{n(1)}, \boldsymbol{u})}{h^*_t(r_{n(1)}, \boldsymbol{u})}(r_{n(1)} - \mu_0) \\ &\leq \epsilon_t \left[\left|\frac{h^*_{ut}(r_{n(1)}, \boldsymbol{u})}{h^*_t(r_{n(1)}, \boldsymbol{u})} - \frac{h^*_{ut}(r_{n(1)}, \boldsymbol{u})}{h^*_t(\mu_0, \boldsymbol{u})}\right| + \frac{X_1}{h^*_t(\mu_0, \boldsymbol{u})}\right] + \frac{h^*_{ut}(\mu_0, \boldsymbol{u})}{h^*_t(\mu_0, \boldsymbol{u})}(r_{n(1)} - \mu_0) \\ &\leq \epsilon_t \nabla_u h^*_t(r_{n(1)}, \boldsymbol{u}) \left(\frac{|h^*_t(r_{n(1)}, \boldsymbol{u}) - h^*_t(\mu_0, \boldsymbol{u})|}{C_1^3} + \frac{X_1}{C_1}\right) + \frac{\nabla_u h^*_t(\mu_0, \boldsymbol{u})}{C_1^2}(r_{n(1)} - \mu_0) \end{aligned} \tag{B.11}$$

Let $p = s_0/4$. Obviously, the $p - th$ moment of X_2 is bounded if

$$||\nabla_u h^*_t(r_{n(1)}, \boldsymbol{u})||_{s_0} < \infty \tag{B.12}$$

Since $\nabla_u h^*_t(r_{n(1)}, \boldsymbol{u}) \leq \nabla_u h^*_t(\mu_0, \boldsymbol{u}) + \sum_{i=1}^{t-1} c'_i(\boldsymbol{u})(r_{n(1)} - \mu_0)$, (B.12) is trivially proved by (B.10). □

Proof of Theorem 4.6: (1)According to Berkes et al. (2003) Lemma 5.4, 5.8 and 5.9, $L_n^*(\mu_0, u)$ uniformly converges to $L(\mu_0, u)$. Lemma 5.5 in that paper shows that $L(\mu_0, u)$ has a unique maximum at η_0. Hence, combined with Lemma A.2, the consistency of $\hat{\eta}_n(r_{n(1)})$ is easily confirmed.

(2)Berkes et al. (2003) Lemma 5.6 gives that $\nabla_u^2 L_n^*(\mu_0, u)$ uniformly converges to $\nabla_u^2 L(\mu_0, u)$. Since $\nabla_u^2 l_t^*(\mu, u)$ is continuous and $r_{n(1)} \rightarrow_p \mu_0$ by Lemma B.1, then for any $u^* \rightarrow_p \eta_0$, $\nabla_u^2 L_n^*(r_{n(1)}, u^*) \rightarrow_p G_0$. The proof of Theorem 4.2 in that paper verifies that G_0 is non-singular. Lemma 5.7 in that paper also confirms the non-singularity of C_0, which indicates $\sqrt{n}\nabla_u L_n^*(\theta_0) \rightarrow N(0, C_0)$ by the CLT. Combined with Lemma B.3, asymptotic normality is proved. □

APPENDIX C

PROOFS FOR CHAPTER 5

LEMMA C.1. $\hat{\pi} \to_p \pi_0$.

Proof of Lemma C.1: Let $Y_t = I_{(r_t > 0)}$. Then Y_t is a simple random variable with expectation π_0 and $\hat{\pi} = \frac{1}{n}\sum_{t=1}^{n} Y_t$. By Weak Law of Large Numbers, convergence of $\hat{\pi}$ is trivially proved. □

LEMMA C.2. $E|l_t(\pi, \theta)| < \infty, \quad E|\nabla_\theta l_t(\hat{\pi}, \theta)| < \infty$

Proof of Lemma C.2: It suffices to show inequalities (C.1)-(C.4):

$$E\left|\ln \frac{r_t}{h_t}\right| < \infty \tag{C.1}$$

$$E\left|\ln\left\{\eta + \left[\frac{\pi\xi r_t}{h_t}\right]^a\right\}\right| < \infty \tag{C.2}$$

$$E|\nabla_{\boldsymbol{u}} \ln h_t| < \infty \tag{C.3}$$

$$E\left|\nabla_{\boldsymbol{u}} \ln\left\{\eta + \left[\frac{\pi\xi r_t}{h_t}\right]^a\right\}\right| < \infty \tag{C.4}$$

$$E\left|\ln \frac{r_t}{h_t}\right| \le E(|\ln r_t| + |\ln h_t|) \le E|\ln z_t| + E|\ln h_{0t}| + E|\ln h_t| \tag{C.5}$$

Because $h_t = c_0(\boldsymbol{u})\left(1 + \sum_{1 \le i \le t-1} \frac{c_i(\boldsymbol{u})}{c_0(\boldsymbol{u})} r_{t-i}\right)$ and $c_i(\boldsymbol{u})$, $0 \le i < \infty$ are all positive, by Jensen's inequality, for $0 < p < 1$,

$$\begin{aligned} E|\ln h_t| &\leq |\ln c_0(\boldsymbol{u})| + \tfrac{1}{p} E \ln \left(1 + \sum_{1\leq i\leq t-1} \frac{c_i(\boldsymbol{u})}{c_0(\boldsymbol{u})} r_{t-i}\right)^p \\ &\leq |\ln c_0(\boldsymbol{u})| + \tfrac{1}{p} \ln E \left(1 + \sum_{1\leq i\leq t-1} \frac{c_i(\boldsymbol{u})}{c_0(\boldsymbol{u})} r_{t-i}\right)^p \\ &\leq |\ln c_0(\boldsymbol{u})| + \tfrac{1}{p} \ln \left(1 + \sum_{1\leq i\leq t-1} \frac{c_i(\boldsymbol{u})}{c_0(\boldsymbol{u})} E r_{t-i}^p\right) \end{aligned} \tag{C.6}$$

$E|\ln h_t| < \infty$ is trivially shown by applying Proposition 4.5(1) and (2) to the above, which also yields $E|\ln h_{0t}| < \infty$. Since z_t is a GF random error, which can be seen as the ratio of two independent chi square random variables, $E|\ln z_t|$ is finite. Inequality (C.1) is proved.

By lemma 5.1 in Berkes et al. (2003) and (5.4),

$$E\frac{h_{0t}}{h_t} < \infty,\ E\left(\frac{h_{0t}}{h_t}\right)^a < \infty \tag{C.7}$$

Let $b = Ez_t^a = E(z_t^a|\mathscr{F}_{t-1})$, by Jensen's inequality and the fact that $E(\frac{r_t}{h_t})^a = E[(\frac{h_{0t}}{h_t})^a E(z_t^a|\mathscr{F}_{t-1})]$,

$$\begin{aligned} E|\ln[\eta + (\tfrac{\pi\xi r_t}{h_t})^a]| &\leq |\ln\eta| + \ln E[1 + \tfrac{1}{\eta}(\tfrac{\pi\xi r_t}{h_t})^a] \\ &\leq |\ln\eta| + \ln[1 + \tfrac{1}{\eta}E(\tfrac{\pi\xi r_t}{h_t})^a] \\ &= |\ln\eta| + \ln[1 + \tfrac{b(\pi\xi)^a}{\eta}E(\tfrac{h_{0t}}{h_t})^a] \end{aligned} \tag{C.8}$$

$$E|\nabla_{\boldsymbol{u}} \ln h_t| = E\left|\frac{\nabla_{\boldsymbol{u}} h_t}{h_t}\right| \leq |c_0'(\boldsymbol{u})| + \sum_{i=1}^{\infty} |c_i'(\boldsymbol{u})| E\frac{r_{t-i}}{h_t} \tag{C.9}$$

$$E\left|\nabla_{\boldsymbol{u}} \ln[\eta + (\frac{\hat{\pi}\xi r_t}{h_t})^a]\right| = E\left|\frac{a\nabla_{\boldsymbol{u}} h_t}{h_t[1+\eta(\frac{h_t}{\hat{\pi}\xi r_t})^a]}\right| \leq E\left|\frac{a\nabla_{\boldsymbol{u}} h_t}{h_t}\right| \tag{C.10}$$

Applying (C.7) to the above, inequalities (C.3) and (C.4) are proved.

□

LEMMA C.3. $\sup_{\theta \in V \times U} \left| \frac{1}{n} L_n(\hat{\pi}, \theta) - \frac{1}{n} L_n(\pi, \theta) \right| \rightarrow_p 0\,.$

Proof of Lemma C.3:

$$\begin{aligned} |\tfrac{1}{n} L_n(\hat{\pi}, \theta) - \tfrac{1}{n} L_n(\pi, \theta)| \quad &\leq \quad |(1-\hat{\pi})\ln(1-\hat{\pi}) - (1-\pi)\ln(1-\pi)| + |\hat{\pi}\ln\hat{\pi} - \pi\ln\pi| \\ &\quad + \tfrac{1}{n} \textstyle\sum_{t=1}^{n} |l_t(\hat{\pi}, \theta) - l_t(\pi, \theta)| \end{aligned} \tag{C.11}$$

Suppose $\hat{\pi} \geq \pi$. Since $\ln x \leq x - 1$

$$\begin{aligned} |l_t(\hat{\pi}, \theta) - l_t(\pi, \theta)| \quad &\leq \quad \ln \tfrac{\eta h_t^a + (\hat{\pi}\xi r_t)^a}{\eta h_t^a + (\pi\xi r_t)^a} + am \ln \tfrac{\hat{\pi}}{\pi} \\ &\leq \quad \tfrac{((\hat{\pi}-\pi)\xi r_t)^a}{\eta h_t^a + (\pi\xi r_t)^a} + am \tfrac{\hat{\pi}-\pi}{\pi} \\ &\leq \quad (\tfrac{\hat{\pi}-\pi}{\pi})^a + \tfrac{am(\hat{\pi}-\pi)}{\pi} \end{aligned} \tag{C.12}$$

$$\begin{aligned} |\tfrac{1}{n} L_n(\hat{\pi}, \theta) - \tfrac{1}{n} L_n(\pi, \theta)| \quad &\leq \quad |(1-\hat{\pi})\ln(1-\hat{\pi}) - (1-\pi)\ln(1-\pi)| + |\hat{\pi}\ln\hat{\pi} - \pi\ln\pi| \\ &\quad + (\tfrac{\hat{\pi}-\pi}{\pi})^a + \tfrac{am(\hat{\pi}-\pi)}{\pi} \end{aligned} \tag{C.13}$$

which obviously converges to zero in probability, by the consistency of $\hat{\pi}$. □

LEMMA C.4. $\frac{1}{\sqrt{n}} \sum_{t \in \mathcal{J}_{nz}} |\nabla_\theta l_t(\hat{\pi}, \theta) - \nabla_\theta l_t(\pi, \theta)| \rightarrow_p 0$

Proof of Lemma C.4: It suffices to show that the following two terms are approaching zero in probability.

(a) Let $A = \frac{1}{\sqrt{n}} \sum_{t \in \mathcal{J}_{nz}} \left| \ln \frac{\eta + (\frac{\hat{\pi}\xi r_t}{h_t})^a}{\eta + (\frac{\pi\xi r_t}{h_t})^a} \right|$. For any $\epsilon > 0$, $P(A > \epsilon) \leq P(A > \epsilon | \hat{\pi} > \pi) + P(A > \epsilon | \hat{\pi} \leq \pi)$. When $\hat{\pi} > \pi$,

$$\begin{aligned}
A &\le \tfrac{1}{n}\sum_{t\in\mathcal{J}_{nz}}\left|\frac{\eta+(\frac{\hat\pi\xi r_t}{h_t})^a}{\eta+(\frac{\pi\xi r_t}{h_t})^a}-1\right| \\
&\le \tfrac{1}{\sqrt{n}}\sum_{t\in\mathcal{J}_{nz}}\left|\frac{(\hat\pi^a-\pi^a)(\frac{\xi r_t}{h_t})^a}{\eta+(\frac{\pi\xi r_t}{h_t})^a}\right| \\
&\le \tfrac{1}{\sqrt{n}}\sum_{t\in\mathcal{J}_{nz}}\frac{(\hat\pi^a-\pi^a)}{\pi^a} \\
&= \sqrt{n}\tfrac{\hat\pi}{\pi^a}(\hat\pi^a-\pi^a)
\end{aligned} \tag{C.14}$$

$$P(A>\epsilon|\hat\pi>\pi)<P\left(\frac{\hat\pi}{\pi^a}(\hat\pi^a-\pi^a)>\epsilon/\sqrt{n}|\hat\pi>\pi\right) \tag{C.15}$$

Because $\hat\pi\to_p\pi$, the right hand side of the above inequality approaches zero. So does $P(A>\epsilon|\hat\pi>\pi)$. Similarly, we can show $P(A>\epsilon|\hat\pi<\pi)$. Hence, $A\to_p 0$.

(b)

$$\begin{aligned}
\tfrac{1}{\sqrt{n}}\sum_{t\in\mathcal{J}_{nz}}\left|\frac{1}{\eta+(\frac{\hat\pi\xi r_t}{h_t})^a}-\frac{1}{\eta+(\frac{\pi\xi r_t}{h_t})^a}\right| &= \tfrac{1}{\sqrt{n}}\sum_{t\in\mathcal{J}_{nz}}\frac{(\hat\pi^a-\pi^a)(\frac{\xi r_t}{h_t})^a}{[\eta+(\frac{\hat\pi\xi r_t}{h_t})^a][\eta+(\frac{\pi\xi r_t}{h_t})^a]} \\
&\le \tfrac{1}{\sqrt{n}}\sum_{t\in\mathcal{J}_{nz}}\frac{\hat\pi^a-\pi^a}{\eta\pi^a} \\
&= \tfrac{\sqrt{n}\hat\pi}{\eta\pi^a}(\hat\pi^a-\pi^a)\to_p 0
\end{aligned} \tag{C.16}$$

$\square$

LEMMA C.5. $E|\nabla_\theta l_t(\hat\pi,\theta)\times\nabla_\theta l_t(\hat\pi,\theta)'|<\infty$

Proof of Lemma C.5: (a). Since $\ln^2(\eta+(\frac{\pi\xi r_t}{h_t})^a)=[\ln\eta+\ln(1+\frac{1}{\eta}(\frac{\pi\xi r_t}{h_t})^a)]^2$, in order to bound the first moment, it suffices to show the finiteness of $E[\ln^2(1+\frac{1}{\eta}(\frac{\pi\xi r_t}{h_t})^a)]$. As $\ln^2(1+\frac{1}{\eta}(\frac{\pi\xi r_t}{h_t})^a)<[1+\frac{1}{\eta}(\frac{\pi\xi r_t}{h_t})^a]^2$, we just need to bound $E(\frac{r_t}{h_t})^{2a}$, which is equivalent to $E(\frac{h_{0t}}{h_t})^{2a}E(z_t^{2a})$. Because $\eta>2$ and $a\eta>2a$, by equation (5.4), there exist $\xi_3\in(2a,a\eta)$ s.t. $E(z_t^{\eta_3})<\infty$. By lemma 5.1 in Berkes et al. (2003), $E(\frac{r_t}{h_t})^{2a}<\infty$.

(b).

$$
\begin{aligned}
E(\ln \tfrac{r_t}{h_t})^2 &= E[(\ln \tfrac{r_t}{h_t})^2)|r_t \geq h_t] + E[(\ln \tfrac{r_t}{h_t})^2|r_t < h_t] \\
&\leq E[(\tfrac{r_t}{h_t})^2|r_t \geq h_t] + E[(\tfrac{r_t}{h_t})^2|r_t < h_t] \\
&= E[z_t^2(\tfrac{h_{0t}}{h_t})^2|r_t \geq h_t] + E[\tfrac{1}{z_t^2}(\tfrac{h_t}{h_{0t}})^2|r_t < h_t] \\
&\leq E[z_t^2(\tfrac{h_{0t}}{h_t})^2] + E[\tfrac{1}{z_t^2}(\tfrac{h_t}{h_{0t}})^2] \\
&= E(z_t^2)E[(\tfrac{h_{0t}}{h_t})^2] + E(\tfrac{1}{z_t^2})E[(\tfrac{h_t}{h_{0t}})^2]
\end{aligned}
\tag{C.17}
$$

By the assumption $a\eta > 2$, $E(z_t^2)$ and $E(\frac{1}{z_t^2})$ are bounded above by K_1 and K_2. Furthermore, lemma 5.1 in Berkes et al. (2003) gives $E[(\frac{h_{0t}}{h_t})^2] < \infty$. Proposition 4.5(2) can be used to bound $\frac{h_t}{h_{0t}}$. Therefore, $E(\ln \frac{r_t}{h_t})^2 < \infty$.

(c) By equation (5.14) in Berkes et al. (2003) ,

$$
\left|\frac{\nabla_{u_k} h_t \nabla_{u_j} h_t}{h_t^2}\right| \leq \left(\sup_{u \in U} \frac{c_0(u) + \sum_{i=1}^{\infty} i^3 c_i(u) r_{t-i}}{c_0(u) + \sum_{i=1}^{n} c_i(u) r_{t-i}}\right)^2, \tag{C.18}
$$

Lemma 5.2 in Berkes et al. (2003) and the fact $E(r_t^p) < \infty$ yield that for any $\nu > 0$

$$
E\left(\sup_{u \in U} \frac{c_0(u) + \sum_{i=1}^{\infty} i^3 c_i(u) r_{t-i}}{c_0(u) + \sum_{i=1}^{n} c_i(u) r_{t-i}}\right)^{\nu} < \infty, \tag{C.19}
$$

Hence, $E\left|\frac{\nabla_{u_k} h_t \nabla_{u_j} h_t}{h_t}\right| < \infty$. Moreover, we can obtain $E|h_{uut}|^{\nu} < \infty$, $E|h_{ut} h_{uut}| < \infty$ by applying the same method and proposition 4.5. □

LEMMA C.6. $E|\nabla_\theta^2 l_t(\hat{\pi}, \theta)| < \infty$, $E|\nabla_\theta^3 l_t(\hat{\pi}, \theta)| < \infty$

Proof of Lemma C.6: For the first conclusion, we need to bound $E[\nabla_u^2 \ln(\eta + (\frac{\hat{\pi}\xi r_t}{h_t})^a)]$.

Since $\nabla_u^2 \ln(\eta + (\frac{\hat{\pi}\xi r_t}{h_t})^a) = \frac{h_{uut} - h_{ut}^2}{1+\eta(\frac{h_t}{\hat{\pi}\xi r_t})^a} - \frac{a\eta h_{ut} h_t^{a-1}/(\hat{\pi}\xi r_t)^a}{[1+\eta(\frac{h_t}{\hat{\pi}\xi r_t})^a]^2}$, it suffices to show $E|h_{ut}^2| < \infty$.

The second conclusion can be verified by showing the finiteness of $E[\nabla_u^3 \ln(\eta + (\frac{\hat{\pi}\xi r_t}{h_t})^a)]$, which only requires $E|h_{ut}^3| < \infty$ and $E|h_{ut} h_{uut}| < \infty$. □

Proof of Theorem 5.1: (1) The first conclusion in Lemma B.1 yields pointwise convergence of $\frac{1}{n}L_{1n}(\pi,\theta)$ to $L_1(\pi,\theta)$, with application of the strong law of large numbers for stationary and ergodic sequences, for example in Stout (1974). The second conclusion in lemma C.3 confirms the weak Lipschitz condition for uniform convergence in Andrews (1992). Hence, $\frac{1}{n}L_{1n}(\pi_0,\theta)$ uniformly converges to $L_1(\pi_0,\theta)$ in probability. Combined with lemma C.4, $\frac{1}{n}L_{1n}(\hat{\pi},\theta) \rightarrow_p L_1(\pi_0,\theta)$ uniformly. Next, we need to show that $El_t(\pi_0,\theta)$ is maximized at $\boldsymbol{u}_0$ regardless of ν as long as z_t is distributed as generalized F for positive values.

Maximizing $El_t(\pi_0,\theta) - El_t(\pi_0,\theta_0)$ is equivalent to maximizing $E\ln\frac{1}{h_t}g(\frac{r_t}{h_t},\nu) - E\ln\frac{1}{h_{0t}}g(\frac{r_t}{h_{0t}},\nu_0)$.

$$
\begin{aligned}
E\ln\frac{1}{h_t}g(\tfrac{r_t}{h_t},\nu) - E\ln\frac{1}{h_{0t}}g(\tfrac{r_t}{h_{0t}},\nu_0) &= E\ln\frac{1}{h_t}g(\tfrac{z_t h_{0t}}{h_t}|\nu) - E\ln\frac{1}{h_{0t}}g(z_t|\nu_0) \\
&= E\ln\frac{h_{0t}}{h_t}\frac{g(\frac{z_t h_{0t}}{h_t}|\nu)}{g(z_t,\nu_0)} \\
&\leq E\left(\frac{h_{0t}}{h_t}\frac{g(\frac{z_t h_{0t}}{h_t}|\nu)}{g(z_t|\nu_0)} - 1\right)
\end{aligned}
\tag{C.20}
$$

Let $c = \frac{h_{0t}}{h_t}$, then

$$
\begin{aligned}
E\ln\frac{1}{h_t}g(\tfrac{r_t}{h_t}|\nu) - E\ln\frac{1}{h_{0t}}g(\tfrac{r_t}{h_{0t}}|\nu_0) &\leq E\left(c\frac{g(cz_t|\nu)}{g(z_t,\nu_0)} - 1\right) \\
&= \int_0^\infty \left(\frac{cg(cz|\nu)}{g(z|\nu_0)} - 1\right) g(z|\nu_0)dz \\
&= \int_0^\infty cg(cz|\nu)dz - \int_0^\infty g(z|\nu_0)dz \\
&= 1 - 1 \\
&= 0
\end{aligned}
\tag{C.21}
$$

Hence, $El_{1t}(\pi_0,\theta)$ has a global maximum of $El_{1t}(\pi_0,\theta_0)$ at $cg(cz_t,\nu) \equiv g(z_t,\nu_0)$. Since $f(c,\nu) = cg(cz_t,\nu)$, by the Mean Value Theorem,

$$(f(c,\nu) - f(1,\nu_0))^2 = ((\nu-\nu_0)', c-1)D_1 D_1' \begin{pmatrix} \nu - \nu_0 \\ c-1 \end{pmatrix} \tag{C.22}$$

By assumption 5.1, $\nu = \nu_0$ and $c = 1$. Since z_t is nondegenerate, theorem 2.3 in Berkes et al. (2003) yields that $c = 1$ implies that $\hat{\boldsymbol{u}} = \boldsymbol{u}_0$.

REMARK 1. *When assumption 5.1 is not satisfied, one cannot assure the uniqueness of the optimal point for* $El_t(\pi_0, \nu, \boldsymbol{u}_0)$.

$$f(c,\nu) - f(1,\nu_0) = (\nu-\nu_0)' D_2 + (c-1)(g(c^* z_t) + c^{*2} g(c^* z_t)) \tag{C.23}$$

Since $g(c^* z_t) + c^{*2} g(c^* z_t) \neq 0$ *for any* $c > 0, \nu > 0$, *then assumption 5.3 gives that* $\nu = \nu_0 + \boldsymbol{v}$ *and* $c = 1$, *i.e.* $h_t = h_{0t}$. *Hence, theorem 5.2 is proved.*

(2) Lemma C.4 implies that $\frac{1}{\sqrt{n}} |\nabla_\theta l_{1t}(\hat{\pi}, \theta) - \nabla_\theta l_{1t}(\pi, \theta)| \rightarrow_p 0$. Lemma C.5 along with the second conclusion in lemma C.2 indicate that $\frac{1}{\sqrt{n}} \nabla_\theta l_{1t}(\pi, \theta)|_{(\pi_0, \theta_0)} \rightarrow N(0, B_0)$. With finite expectation $\nabla^2_\theta l_{1t}(\hat{\pi}, \theta)$, the weak law of large numbers yields

$$\frac{1}{n} \sum_{t=1}^{n} \nabla^2_\theta l_{1t}(\hat{\pi}, \theta) \rightarrow_p E \nabla^2_\theta l_{1t}(\hat{\pi}, \theta) \tag{C.24}$$

By the extended theorem about the extremum estimator, we obtain asymptotic normality for $\hat{\theta}$. $\square$

REFERENCES

Allen, David et al. (2008). "Finite Sample Properties of the QMLE for the Log-ACD Model: Application to Australian Stocks". *Journal of Econometrics* 147.1, pp. 163–185.

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

Andrews, D.W.K. (1992). "Generic Uniform Convergene". *Econometric Theory* 8, pp. 241–257.

Baillie, Richard T and Kapetanios, George (2013). "Estimation and Inference for Impulse Response Functions from Univariate Strongly Persistent Processes". *The Econometrics Journal* 16.3, pp. 373–399.

Baillie, R.T., Bollerslev, T, and Mikkelsen, H.O. (1996). "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics* 52, pp. 91–113.

Bauwens, L and Veredas, D (2004). "The Stochastic Conditional Duration Model: a Latent Factor Model for the Analysis of Financial Durations." *Journal of Econometrics* 119, pp. 381–412.

Bauwens, Luc and Giot, Pierre (2000). "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks". *Annales d'conomie et de Statistique* 60, pp. 117–149.

Bauwens, Luc, Laurent, Sébastien, and Rombouts, Jeroen VK (2006). "Multivariate GARCH Models: a Survey". *Journal of applied econometrics* 21.1, pp. 79–109.

Berkes, Istvan and Horvath, Lajos (2004). "The Efficiency of the Estimators of the Parameters in GARCH Processes". *The Annals of Statistics* (32), pp. 633–655.

Berkes, Istvan, Horvath, Lajos, and Kokoszka, Piotr (2003). "GARCH Processes: Structure and Estimation". *Bernoulli* (9), pp. 201–227.

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

Bollerslev, Tim (1990). "Modelling the Coherence in Short-Run Nominal Exchange Rates: a Multivariate Generalized ARCH Model". *The Review of Economics and Statistics*, pp. 498–505.

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

— (1992b). "Strict Stationarity of Generalized Autoregressive Processes". *The Annals of Probability* 20, pp. 1714–1730.

Brownlees, Christian T and Gallo, Giampiero M (2011). "Shrinkage Estimation of Semiparametric Multiplicative Error Models". *International Journal of Forecasting* 27.2, pp. 365–378.

Campbell, John Y and Mankiw, N Gregory (1986). *Are Output Fluctuations Transitory?*

Carrasco, Marine and Chen, Xiaohong (2002). "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models". *Econometric Theory* 8, pp. 17–39.

Chen, Haiqiang and Choi, Paul Moon Sub (2012). "Does Information Vault Niagara falls? Cross-listed Trading in New York and Toronto". *Journal of Empirical Finance* 19.2, pp. 175–199.

Chen, Yi-Ting and Hsieh, Chih-Sheng (2010). "Generalized Moment Tests for Autoregressive Conditional Duration Models". *Journal of Financial Econometrics.*

Diebold, Francis X, Gunther, Todd A, and Tay, Anthony S (1997). *Evaluating Density Forecasts.*

Drost, Feike C and Werker, Bas J M (2004). "Semiparametric Duration Models". *Journal of Business & Economic Statistics* 22.1, pp. 40–50.

Duchesne, Pierre and Pacurar, Maria (2008). "Evaluating Financial Time Series Models for Irregularly Spaced Data: a Spectral Density Approach". *Computers & Operations Research* 35.1, pp. 130–155.

Dufour, A and Engle, R.F. (2000). "The ACD Model: Predictability of the Time between Consecutive Trades". *Discussion Paper, ISMA Centre, University of Reading.*

Engle, Robert (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom inflation". *Econometrica* 50, pp. 987–1007.

— (2002a). "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models". *Journal of Business & Economic Statistics* 20.3, pp. 339–350.

— (2002b). "New Frontiers For ARCH Models". *Journal of Applied Econometrics* 17.

Engle, Robert and Gallo, Giampiero M. (2006). “A Multiple Indicators Model for Volatility Using Intra-daily Data”. *Journal of Econometrics* 131, pp. 3–27.

Engle, Robert and Ng, Victor K. (1993). “Measuring and Testing the Impact of News on Volatility”. *The Journal of Finance* 48, pp. 1749–1778.

Engle, Robert and Russell, Jefferey (1998). “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data”. *Econometrica* 66, pp. 1127–1162.

Engle, Robert F and Kroner, Kenneth F (1995). “Multivariate Simultaneous Generalized ARCH”. *Econometric theory* 11.01, pp. 122–150.

Fernandes, M and Grammig, J (2006). “A Family of Autoregressive Conditional Duration Models.” *Journal of Econometrics* 130, pp. 1–23.

Fernandes, Marcelo and Grammig, Joachim (2005). “Nonparametric Specification Tests for Conditional Duration Models”. *Journal of Econometrics* 127.1, pp. 35–68.

Gagnon, Louis and Karolyi, G Andrew (2010). “Multi-market Trading and Arbitrage”. *Journal of Financial Economics* 97.1, pp. 53–80.

Geweke, John (1986). “Modeling the Persistence of Conditional Variance: A Comment”. *Econometric Reviews* 5, pp. 57–61.

Ghysels, E, Gourieroux, C, and Jasiak, J (2004). “Stochastic volatility durations”. *Journal of Econometrics* 119, pp. 413–435.

Halling, Michael, Moulton, Pamela C, and Panayides, Marios (2013). “Volume Dynamics and Multimarket Trading”. *Journal of Financial and Quantitative Analysis* 48.02, pp. 489–518.

Han, Heejoon (2013). “Asymptotic Properties of GARCH-X Processes”. *Journal of Financial Econometrics*, nbt023.

Han, Heejoon and Kristensen, Dennis (2014). “Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Nonstationary Covariates”. *Journal of Business & Economic Statistics* 32.3, pp. 416–429.

Hautsch, Nikolaus (2002). “Modelling Intraday Trading Activity Using Box-Cox ACD Models”. *Discussion Paper, CoFE, University of Konstanz.*

— (2003). “Assessing the Risk of Liquidity Suppliers on the Basis of Excess Demand Intensities”. *Journal of Financial Econometrics* 1.2, pp. 189–215.

— (2006). “Testing the Conditional Mean Function of Autoregressive Conditional Duration Models”. *University of Copenhagen Discussion Paper.*

— (2012). *Econometrics of Financial High-Frerquency Data*. Springer -Verlag Berlin Heidelberg.

Hautsch, Nikolaus and Jeleskovic, Vahidin (2008). “High-frequency Volatility and Liquidity”. *Applied Quantitative Finance*. Springer, pp. 379–397.

Hautsch, Nikolaus, Malec, Peter, and Schienle, Melanie (2014). “Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes”. *Journal of Financial Econometrics* 12, pp. 89–121.

Hentschel, L. (1995). “All in the Family: Nesting Symmetric and Asymmetric GARCH Models”. *Journal of Financial Economics* 39, pp. 71–104.

Hong, Yongmiao and Lee, Yoon-Jin (2011). “Detecting Misspecifications in Autoregressive Conditional Duration Models and Non-Negative Time-Series Processes”. *Journal of Time Series Analysis* 32.1, pp. 1–32.

Jasiak, J (1998). “Persistence in Intergrade Durations”. *Finance* 19, pp. 166–195.

Jensen, S.T. and Rahbek, Anders (2004a). “Asymptotic Inference for Nonstationary GARCH”. *Econometric Theory* 20.

— (2004b). “Asymptotic Normality of the QMLE Estimator for ARCH in the Non-Stationary Case”. *Econometrica* 72, pp. 641–646.

Koop, Gary, Pesaran, M Hashem, and Potter, Simon M (1996). “Impulse Response Analysis in Nonlinear Multivariate Models”. *Journal of Econometrics* 74.1, pp. 119–147.

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

Lehmann, E.L. (1999). *Elements of Large-Sample Theory*. Springer-Verlag.

Li, W.K. and Mak, T.K. (1994). “On the Squared Residual Autocorrelations in Non-Linear Time Series with Conditional Heteroskedasticity”. *Journal of Time Series Analysis* 15, pp. 627–636.

— (2003). “On the Residual Autocorrelation of the Autoregressive Conditional Duration Model”. *Economics Letters* 79, pp. 169–175.

Ling, Shiqing and McAleer, Michael (2003). “Asymptotic Theory for a Vector ARMA-GARCH Model”. *Econometric Theory* 19, pp. 280–310.

Luca, Giovanni De and Gallo, Giampiero M (2008). “Time-Varying Mixing Weights in Mixture Autoregressive Conditional Duration Models”. *Econometric Reviews* 28.1-3, pp. 102–120.

Lumsdaine, R.L. (1996). "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models". *Econometrica* 64, pp. 575–596.

Manganelli, Simone (2005). "Duration, Volume and Volatility Impact of Trades". *Journal of Financial Markets* 8, pp. 377–399.

Meitz, Mika and Saikkonen, Pentti (2008). "Ergodicity, Mixing, and Existence of Moments of A Class of Markov Models with Applications to GARCH and ACD Models". *Econometric Theory* 24, pp. 1291–1320.

Meitz, Mika and Teräsvirta, Timo (2006). "Evaluating models of autoregressive conditional duration". *Journal of Business & Economic Statistics* 24.1.

Mokkadem, A (1990). "Propriétés De Mélange Des Modèles Autorégressifs Polynomiaux". *Ann. Inst. Henri.*

Nelson, D.B. (1990). "Stationarity and Persistence in the GARCH(1,1) Model". *Econometric Theory* 6, pp. 318–334.

— (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". *Econometrica* 59, pp. 347–370.

Pacurar, Maria (2008). "Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical literature". *Journal of Economic Surveys* 22, pp. 711–751.

Pesaran, H Hashem and Shin, Yongcheol (1998). "Generalized Impulse Response Analysis in Linear Multivariate Models". *Economics letters* 58.1, pp. 17–29.

Phillips, Peter CB (1998). "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs". *Journal of econometrics* 83.1, pp. 21–56.

Rabemananjara, Roger and Zakoian, Jean-Michel (1993). "Threshold ARCH Models and Asymmetries in Volatility". *Journal of Applied Econometrics* 8.1, pp. 31–49.

Saart, Patrick W and Gao, Jiti (2012). "An Alternative Nonparametric Specification Test in Autoregressive Conditional Duration Models". *Available at SSRN 2130454.*

Saart, Patrick W, Gao, Jiti, and Allen, David E (2015). "Semiparametric Autoregressive Conditional Duration Model: Theory and Practice". *Econometric Reviews* 34.6-10, pp. 848–880.

Stout, W.F. (1974). *Almost Sure Convergence.* New York: Academic Press.

Tay, Anthony et al. (2009). "Using High-frequency Transaction Data to Estimate the Probability of Informed Trading". *Journal of Financial Econometrics*, nbp005.

Tsay, Ruey (2010). *Analysis of Financial Time Series.* John Willey and Sons.

Vuorenmaa, Tommi A (2009). "A q-Weibull Autoregressive Conditional Duration Model with an Application to NYSE and HSE data". *Available at SSRN 1952550.*

Xu, Dinghai and Wirjanto, Tony S (2010). "An Empirical Characteristic Function Approach to VaR Under a Mixture-of-Normal Distribution with Time-Varying Volatility". *The Journal of Derivatives* 18.1, pp. 39–58.

Xu, Dinghai, Knight, John, and Wirjanto, Tony S (2011). "Asymmetric Stochastic Conditional Duration Model-A Mixture-of-Normal Approach". *Journal of Financial Econometrics* 9.3, pp. 469–488.

Zakoian, Jean-Michel (1994). "Threshold Hereroskedastic Models". *Journal of Economic Dynanmic and Control* 18, pp. 931–955.

Zhang, Michael Yuanjie, Russell, Jeffrey R, and Tsay, Ruey S (2001). "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data". *Journal of Econometrics* 104.1, pp. 179–207.

VITA

Qian Li was born on September 7th, 1987, in Liuzhou, China. In 2009 she graduated from Sichuan University in China with a B.S. degree in Mathematics. She received a M.S. degree in Statistics from the University of Missouri-Kansas City in 2012.