Abstract

SHEN, JI. Three Essays on Realized Volatility Models for High-Frequency Data. (Under the direction of Denis Pelletier.)

The dissertation centers on modeling volatility of stock returns with high frequency data. It contains three chapters.

In Chapter 1, we propose a multivariate realized GARCH model where the conditional variance of daily returns depends on lagged realized measures of volatility computed with high-frequency data. In turn, a measurement equation contemporaneously links the realized measures and the conditional variance. To greatly reduce the number of parameters in the measurement equation, we use adaptive Lasso and the least-angle regression algorithm. For the GARCH dynamic of the conditional variance, we favor a diagonal BEKK structure applied to rotated variables. Our empirical results suggest that the use of variable selection with adaptive Lasso for the measurement equation instead of OLS leads to improved out-of-sample forecasting performances.

In Chapter 2, we introduced a realized Markov regime-switching BEKK model which generalizes the BEKK model by using two regimes featuring different variance dynamic specifications. Within each regime, we use a BEKK model to govern the variance. The persistence of both regimes yields an extra source of volatility persistence compared with standard, single-regime BEKK, thereby enhancing the flexibility in describing the volatility persistence of shocks. To reduce the parameters in estimation, we implement covariance targeting with rotated variables. The empirical evidence shows that the regime-switching BEKK model has a significantly better performance in both in-sample fit and out-of-sample forecasting.

In Chapter 3, we generalize the realized Markov regime-switching GARCH model in Chapter 2 by allowing the regime transition probability to vary over time using pre-determined observable variables which is the realized kernel in this case. This model has more flexibility regarding volatility persistence compared with the one in Chapter 2. To avoid the computational difficulty due to an increasing number of parameters, we utilize this model with one stock only. Our empirical research shows that the the use

of time-varying transition probabilities leads to better in-sample fit and improved forecasting ability in the near future (about 10 days).

Three Essays on Realized Volatility Models for High-Frequency Data

by
Ji Shen

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2017

APPROVED BY:

Barry Goodwin	Walter Thurman
Peter Bloomfield Minor Member	Denis Pelletier Chair of Advisory Committee

Dedication

To my parents and the memories of my grandparents.

Biography

Ji Shen was born in a nice small town Qintong, Jiangsu in 1984. In 2006, he graduated from Xiamen University where he studied advertising and began a career in marketing and advertising in Guangzhou and later Beijing. In 2010, Ji began the pursuit of a Master of Economics degree in Capital University of Economics and Business in Beijing. In August of 2012, Ji began work on his Ph.D. in Economics at North Carolina University and finally focuses on Econometrics.

Acknowledgements

I would like to thank my advisor Dr. Denis Pelletier for his direction and guidance to my research. Through patient discussion he has provided me the knowledge and motivation to overcome one hurdle after another.

I would also like to thank my committee members: Dr. Peter Bloomfield, Dr. Barry Goodwin and Dr. Walter Thurman. Their valuable comments and suggestions greatly enriched this dissertation. I am also deeply grateful for their support and encouragement.

I also want to express my appreciation to my friends providing help along the way of changing my career path. Last but not least, I would like to thank my family for their endless support and love. Without them, none of this could have happened.

Table of Contents

List of Tables

List of Figures

Chapter 1

Realized Rotated BEKK Model

1.1 Introduction

With the advent of cheaper data storage and improved computational resources, high-frequency financial data have seen increased availability. This growing availability has led researchers to focus on methods for utilizing high-frequency data to study the statistical properties of financial assets. A number of realized measures of volatility, including realized variance, bipower variation, the realized kernel, and many related quantities have been introduced in the literature (Andersen and Bollerslev, 1998; Andersen, Bollerslev, Diebold, and Ebens, 2001; Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen, 2002; Barndorff-Nielsen and Shephard, 2004; Andersen and Benzoni, 2008; Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; Hansen and Horel, 2009). Any of the aforementioned measures is far more informative about the current level of volatility than the squared return. As a result, realized measures are more useful for modeling and forecasting future volatility. Engle (2002) estimated a GARCH model that included a realized volatility measure in the GARCH equation. This model, which specifically included lagged realized variance to drive the dynamics of conditional variance of returns, has since come to be known as the GARCH-X (Forsberg and Bollerslev, 2002). Engle and Gallo (2006) model a 3-variable system comprising the squared return, the high-minus-low price range and the realized variance in an multiplicative error model (MEM) setup suggested by Engle (2002). They apply

the MEM to squared returns and realized volatility as separate models. The univariate and multivariate HEAVY models introduced by Shephard and Sheppard (2010) and Noureldin, Shephard, and Sheppard (2012) consider a system including a measurement equation for multi-step forecasting of the conditional variance of daily returns. The mathematical structure of the HEAVY model is a generalization of the MEM framework.

While the HEAVY model postulates GARCH-type dynamics for the realized measure by modelling its conditional expectation, the univariate realized GARCH model introduced by Hansen, Huang, and Shek (2012) relate the realized measure itself to conditional variance and a term that captures leverage effects. The basic idea is that the connection between the realized measure and the conditional variance can be used to update the realized measure process, thus providing a recursive prediction of the conditional variance of the asset. Almost universally, the incorporation of realized measures into volatility models has led to large economic and statistical gains (Christoffersen, Jacobs, and Mimouni, 2010; Dobrev and Szerszen, 2010). Forecast gains tend to be more pronounced at short time horizons, typically the first several days. All of the models mentioned above focus on a small number of assets. For example, multivariate HEAVY models separate ten assets into five groups, and deal with two assets in each group. In general, most of these models suffer from a curse-of-dimensionality problem that constrains their practical application.

Automatic model-building methods are familiar in the linear model literature. They are used to automatically produce better models in terms of prediction accuracy and parsimony. One promising model-building method is the Lasso; its computationally simpler variation called Least Angle Regression (LARS) will be applied here.

In this paper, we generalize the realized GARCH to the multivariate case. The number of coefficients in the measurement equation is reduced using the Lasso method, which makes this modeling approach tractable for relatively large portfolios of assets.

1.2 Literature Review

1.2.1 Univariate Volatility Models

In financial econometrics, the autoregressive conditional heteroskedasticity (ARCH) model is a benchmark tool for analyzing the time-variation of conditional variance. The ARCH model is essentially an ARMA model for the squared return and many tools of standard, linear time-series analysis can be directly applied. The family of ARCH models is easy to estimate; these parsimonious models are capable of providing good descriptions of the dynamics of asset volatility. The complete ARCH(p) model relates the current level of volatility to past p square shocks (Engle, 1982). Consider r_t, which denotes the log return of a single asset, and is characterized as a real-valued discrete-time stochastic process. We model r_t as

$$r_t = \mu_t + \epsilon_t$$

where μ_t can be any general model for the conditional mean (e.g. autoregressive (AR) model, moving average (MA) model, autoregressive integrated moving average (ARIMA) model, etc). In practice, the model for the conditional mean should be flexible enough to capture the dynamics present in the data. For many financial time series, particularly those measured over short intervals such as one day or one week, assuming $\mu_t = 0$ is appropriate. As for the error term,

$$\begin{aligned} var(r_t|\mathcal{F}_{t-1}) &= E(\epsilon_t^2|\mathcal{F}_{t-1}) = h_t, \\ \epsilon_t &= h_t^{1/2} e_t, \end{aligned}$$

where $h_t^{1/2}$ is a positive real value and $\mathcal{F}_{t-1}$ is the set of all information up to time $t-1$. Also, assume the random variable e_t meets two moment conditions:

$$\begin{aligned} E(e_t|\mathcal{F}_{t-1}) &= 0, \\ var(e_t|\mathcal{F}_{t-1}) &= 1. \end{aligned}$$

For an ARCH(1), the variance process is given by

$$h_t = \omega + \alpha\epsilon_{t-1}^2.$$

They key feature of this model is that the conditional variance of the shock, ϵ_t, is time varying and depends on past values through its square. The variance h_t is the time $t-1$ conditional variance and is measurable in $\mathcal{F}_{t-1}$. The generalized ARCH (GARCH) process, introduced by Bollerslev (1986), improves the original specification by adding lagged conditional variances. These lagged terms essentially act as "smoothing" terms. GARCH models typically fit as well as higher ordered ARCH models while remaining relatively parsimonious. The conditional variance in a GARCH(1,1) process evolves according to

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}.$$

In this specification, the conditional variance is an average of the lagged shock, ϵ_{t-1}^2, and the lagged variance, h_{t-1}, plus a constant. The effect of the lagged variance is to produce a model that is actually an ARCH(∞). This result is readily obtained by continued backward substitution of the conditional variance. To ensure the conditional variances are always positive, restrictions must be imposed on the parameters of the GARCH model. In a GARCH(1,1), these restrictions are $\omega > 0, \alpha \geq 0$, and $\beta \geq 0$. The unconditional variance can be derived as

$$\bar{h} \equiv E[h_t] = \frac{\omega}{1-\alpha-\beta}.$$

The GARCH model with $1-\alpha-\beta > 0$ is covariance stationary as long as the model for the conditional mean corresponds to a stationary process.

1.2.2 The BEKK Model

The BEKK model (Baba, Engle, Kraft and Kroner) is a multivariate GARCH model designed to ensure that the conditional covariance is positive definite (Engle and Kroner, 1995). The primary insight of

the BEKK is that quadratic forms are positive semi-definite, and the sum of a positive semi-definite matrix and a positive definite matrix is positive definite. The model specification is an extension of the univariate model. Specifically, consider a vector of returns r_t of dimension $k \times 1$. Similar to the setup in the univariate case,

$$r_t = \mu_t + \epsilon_t, \tag{1.1}$$

where μ_t is a general time series model for the conditional mean vector. The conditional variance is

$$cov(\epsilon_t|\mathcal{F}_{t-1}) = H_t$$

where H_t is a $k \times k$ positive definite matrix and

$$\epsilon_t = H_t^{1/2} e_t$$

where $H_t^{1/2}$ can be obtained by the Cholesky factorization of H_t. The moment conditions

$$E(e_t|\mathcal{F}_{t-1}) = 0$$
$$cov(e_t|\mathcal{F}_{t-1}) = I_k$$

must also hold, with I_k denoting the k dimensional identity matrix. The covariance in a BEKK(1,1) model evolves according to

$$H_t = CC' + B\, H_{t-1}\, B' + A\, \epsilon_{t-1}\epsilon_{t-1}'\, A'$$

where C is a $k \times k$ lower triangular matrix and A and B are $k \times k$ parameter matrices. The elements of H_t depend on both squared returns and the cross-product of returns.

1.2.3 Realized Volatility

Over the past decade, extensive research has shown that the rich information in high-frequency data can be effectively used to construct realized measures of volatility. For univariate processes, the corre-

sponding measure is the realized variance. Consider a log-price process, p_t, driven by a standard Wiener process with time-varying drift and stochastic volatility,

$$\mathrm{d}\, p_t = \mu_t \,\mathrm{d}\, t + \sigma_t \,\mathrm{d}\, W_t. \tag{1.2}$$

where μ_t is of finite variation and σ_t is strictly positive and square integrable. Realized variance is constructed by frequently sampling p_t throughout the trading day. Suppose that prices on day t were sampled on a regular grid of $n+1$ points $(0, 1, ..., n)$ and let $p_{i,t}$ denote the i^{th} observation of the log price on day t. The n-sample realized variance is defined as

$$RV_t^n = \sum_{i=1}^{n} (p_{i,t} - p_{i-1,t})^2 = \sum_{i=1}^{n} r_{i,t}^2. \tag{1.3}$$

RV_t^n is a consistent estimator of the integrated variance under the assumption of no market microstructure noise, which includes bid-ask bounce, discreteness of price changes, etc:

$$\lim_{n\to\infty} RV_t^n = \int_t^{t+1} \sigma_s^2 ds,$$

where the bounds t and $t+1$ represent the (arbitrary) interval over which the realized covariance is computed and n is typically somewhere between 13 (30-minute returns) and 390 (1-minute returns).

The realized variance measure suffers from a number of problems. The most pronounced is that observed prices are contaminated by noise since they are only observed at the bid and ask. The observed return is the sum of the true return and an MA(1) shock. As a result, returns are autocorrelated and computing the RV using (1.3) will result in upwardly biased estimates. This bias increases with the number of sampling points. Several solutions to this problem have been proposed. The simplest solution is to avoid the issue by sampling prices infrequently, which usually means limiting samples to somewhere between 1 to 30 minutes (Bandi and Russell, 2008). Another problem is the continuity of return process in equation (1.2) may be interrupted by sudden discrete movements (jumps) when unexpected news hits the market, which adds an upward bias to the realized variance. Barndorff-Nielsen, Hansen, Lunde, and

Shephard (2008) propose a realized kernel estimator which is consistent under the general conditions, e.g., higher-order dependence or endogeneity in the noise process, and robust to jumps. The idea is to capture (potentially noise induced) serial correlations in trade-to-trade returns by a kernel. The realized kernel estimator (RK) is defined as:

$$RK = \gamma_0 + \sum_{q=1}^{Q} k\left(\frac{q-1}{Q}\right)(\gamma_q + \gamma_{-q}), \tag{1.4}$$

with

$$\gamma_q = \sum_{i=q+1}^{n} r_i r_{i-q}, \quad for\, q \geq 0,$$

where $k(\cdot)$ is a kernel weighting function and $k(0) = 1, k(1) = 0$; Q is a bandwidth parameter which is chosen to minimize mean square error (MSE).

The multivariate case retains many of the features of the univariate case. For a multivariate process, the measure of interest is the realized variance/covariance matrix. Assume that the $k \times 1$ log-price vector, $P(t)$, is governed by the following multivariate diffusion process

$$\mathrm{d}\, P_t = U_t dt + \Omega_t\, \mathrm{d}\, W_t$$

where U_t is the $k \times 1$ drift vector and $\Sigma_t = \Omega_t \Omega_t'$ is instantaneous covariance. $\mathrm{d}\, W_t$ denotes k-variate mutually independent Brownian motion. Realized covariance estimates the integrated covariance matrix

$$\int_t^{t+1} \Sigma_s\, \mathrm{d}\, s.$$

Realized covariance is computed using the outer-product of the high-frequency returns,

$$RC_t = \sum_{i=1}^{n} r_{i,t} r_{i,t}' = \sum_{i=1}^{n} (p_{i,t} - p_{i-1,t})(p_{i,t} - p_{i-1,t})'$$

In principle, prices should be sampled as frequently as possible to maximize the precision of the realized

covariance estimator. However, frequent sampling leads to practical difficulties since transaction prices are contaminated by noise (e.g. bid-ask bounce) and not perfectly synchronized (Epps, 1979).

Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) proposed the multivariate realized kernel which uses a "refresh time" sampling procedure to adapt the sampling scheme to the trading intensity of the least active asset at any given time within a day. The kernel also includes a well-chosen weight function for the lead and lag returns used in the computation of the covariance matrix. This ensures consistency of the estimator, while also guaranteeing positive semi-definiteness of the estimated covariance matrix. Formally, the multivariate realized kernel (RK) is defined as

$$RK = \Gamma_0 + \sum_{q=1}^{Q} K\left(\frac{q-1}{Q}\right)(\Gamma_q + \Gamma_q') \tag{1.5}$$

with

$$\Gamma_q = \sum_{i=q+1}^{\tilde{n}} \tilde{r}_i \tilde{r}_{i-q}', \quad \text{for } q \geq 0,$$

where $\tilde{r}$ are refresh time returns, $\tilde{n}$ is the number of refresh time prices, $K(\cdot)$ is a kernel weighting function, and Q is a bandwidth parameter which is chosen to minimize mean square error (MSE). Refresh time returns are computed by sampling all prices and using last-price interpolation only when all assets have been traded. The recommended kernel is the non-flat Parzens kernel:

$$K(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 > x \geq 1/2, \\ 2(1-x)^3 & 1/2 > x \geq 1, \\ 0 & x > 1. \end{cases}$$

Jittering is another common technique applied in the calculation of the realized kernel. Jittering averages $m \leq \tilde{n}$ prices at the very beginning and end of the day to eliminate end-effects and ensure that the realized kernel is consistent. Larger values of m imply smaller end-effects. However, increasing m reduces the amount of available data for computing the realized kernel. The optimal amount of jittering

will minimize the MSE caused by end-effects and the increased asymptotic variance due to reduction of the effective sample size. Setting $m = 2$ is a common rule of thumb.

1.3 Realized BEKK Model

1.3.1 Definitions and Notation

In this section we introduce a new model, the realized BEKK (RBEKK) model. Similarly to the BEKK model in section 2.2, we start by assuming $E(r_t|\mathcal{F}_{t-1}) = 0$. This is a legitimate assumption based on several empirical studies. Otherwise, this condition could be met by reinterpreting r_t as the return minus its conditional mean. As a result,

$$\begin{aligned} r_t &= H_t^{1/2} e_t, \\ e_t &\overset{i.i.d}{\sim} (0, I_k), \end{aligned} \tag{1.6}$$

where I_k is the $k \times k$ identity matrix. The key quantity of interest in many financial applications is the conditional variance/covariance, $H_t = cov(r_t|\mathcal{F}_{t-1})$. For the conditioned variance/covariance, we assume it follows the following dynamic:

$$H_t = CC' + B\, H_{t-1}\, B' + A\, X_{t-1}\, A'. \tag{1.7}$$

In the standard GARCH(1, 1) model H_t is a function of H_{t-1} and $r_{t-1}r'_{t-1}$. In the present framework, a realized measure of variance/covariance X_{t-1} will replace $r_{t-1}r'_{t-1}$. In practice, X_{t-1} could be the matrix of realized variance/covariance or a realized kernel. The $k \times k$ matrices A and B each have k^2 free parameters, while C is a $k \times k$ lower triangular matrix with $k^* = k(k+1)/2$ free parameters. Under the assumption that the starting value H_0 is positive semidefinite, the parameterization in (1.7) guarantees that H_t is positive semidefinite for all t. Moreover, if C is a full rank matrix, then H_t is positive definite for all t. The coefficient matrix B determines the "smoothness" of the RBEKK model.

We will refer to (1.6) as the return equation and (1.7) as the RBEKK equation. We define the mea-

surement equation in (1.8) to close the system:

$$x_t = d + \Gamma\, h_t + u_t \tag{1.8}$$
$$u_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_u)$$

where x_t, d, h_t, and u_t are all $k^* \times 1$ vectors, d is the intercept vector, and u_t is the innovation vector. Moreover, Γ is a $(k^* \times k^*)$ constant coefficient matrix. This matrix characterizes the dependence between the realized measure and all of the elements of the conditional covariance matrix. A major feature of the RBEKK model is that the realized measure X_t can be linked with the contemporaneous logarithmic H_t embedded in the measurement equation. Define $x_t = vech(\operatorname{logm} X_t)$ and similarly $h_t = vech(\operatorname{logm} H_t)$ where the vech operator stacks the lower triangular portion, including the main diagonal, of a $(k \times k)$ symmetric matrix into a $(k^* \times 1)$ vector. The logm operator gives the matrix logarithm which is a generalization of the element-wise logarithm (Higham, 2008). In our case, since the realized measure of variance/covariance is a real symmetric matrix, it can be decomposed into the product of a diagonal matrix of eigenvalues and a matrix of eigenvectors. Specifically, assuming X is the realized measure of variance/covariance, we can write $X = V\tilde{X}V^{-1}$, where V is the matrix of eigenvectors of X (each column of V is an eigenvector of X), and $\tilde{X}$ is a diagonal matrix whose diagonal elements are eigenvalues of X. Replace each diagonal element of $\tilde{X}$ by its natural logarithm to obtain $\log \tilde{X}$. Then we have $\operatorname{logm} X = V(\log \tilde{X})V^{-1}$. Since the realized measure of variance/covariance is positive definite, the eigenvalues of X are all positive. As a result, $\operatorname{logm} X$ is always a real, symmetric, and unique matrix.

The BEKK-type parameterization in (1.7) has $O(k^2)$ parameters. If the elements of A and B are unrestricted, which is a full parameterization, the evolution of every element in H_t will be influenced by its own lag as well as every cross-asset effect in the RBEKK model. A parsimonious specification is diagonal BEKK where the coefficient matrix A and B are assumed be diagonal. In this case, the equations for each element in the matrix H_t characterize a group of univariate realized GARCH models in which the conditional variances or covariances are driven by their own lags and the corresponding

realized measure.

For demonstration, consider the 2 by 2 case of (1.7), with diagonal coefficient matrices.

$$\begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} c_{11}^2 \\ c_{11}c_{21} \\ c_{21}^2 + c_{22}^2 \end{pmatrix} + \begin{pmatrix} b_{11}^2 h_{11,t} \\ b_{11}b_{22}h_{12,t} \\ b_{22}^2 h_{22,t} \end{pmatrix} + \begin{pmatrix} a_{11}^2 x_{11,t} \\ a_{11}b_{22}x_{12,t} \\ a_{22}^2 x_{22,t} \end{pmatrix}$$

Although equation (1.8) has $O(k^4)$ parameters, the vectorized form is easy to estimate since its structure is amenable to seemingly unrelated regression (SUR). If a model has too many parameters, it begins to model the random noise in the data; this is commonly known as overfitting. Overfitting produces misleadingly high R-squared values and a reduced ability to make meaningful predictions. We use variable selection methods in estimation to reduce the number of elements in the coefficient matrix Γ.

For demonstration, extend the previous example by writing out the 2 by 2 case of (1.8) in vech form:

$$\begin{pmatrix} x_{11,t} \\ x_{21,t} \\ x_{22,t} \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{21} & \gamma_{22} & \gamma_{33} \end{pmatrix} \begin{pmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{pmatrix} + \begin{pmatrix} u_{11,t} \\ u_{21,t} \\ u_{22,t} \end{pmatrix}$$

It is clear that each element in the realized covariance is a function of every component of the conditional covariance matrix. The RBEKK model fully specifies the dynamic system of daily returns and realized measure.

1.3.2 Covariance Targeting Under Rotation

The RBEKK equation is a nonlinear model that is hard to estimate when the number of parameters is high. Covariance targeting suggested by Mezrich and Engle (1996) allows a reparameterization of the model and a first-step estimation of some parameters by empirical moments. The remaining parameters are estimated by QML in a second step. This can substantially reduce the number of parameters to be estimated through non-linear optimization. For multivariate models, we can apply the rotation tech-

nique to keep the flexibility of dynamics in the full specification model while using the restricted model to reduce the parameters. The rotated GARCH models were introduced by Noureldin, Shephard, and Sheppard (2014). They compared scalar and diagonal rotated BEKK, DCC, and OGARCH models in the 10- and 30-dimensional cases. We apply this idea to the high-frequency volatility model. Assuming the process is stationary, the unconditional covariance of the daily return and expected realized measure are, respectively, $Var[r_t] = E[H_t] = \bar{H}$ and $E[X_t] = \bar{X}$, the covariance targeting for H_t is

$$H_t = (\bar{H} - B\bar{H}B' - A\bar{X}A') + B\,H_{t-1}\,B + A\,X_{t-1}\,A, \tag{1.9}$$

assuming that $(\bar{H} - B\bar{H}B' - A\bar{X}A')$ is positive semidefinite. It is difficult to specify sufficient parameter restrictions to ensure that $(\bar{H} - B\bar{H}B' - A\bar{X}A')$ is positive semidefinite. This feature of the model prevents the estimation of more flexible dynamics.

We fit the model with rotated variables to extend the idea of variance targeting to covariance targeting in multivariate models of any dimension. With the rotated variables we can estimate a restricted version of BEKK model which has fewer parameters to estimate while keep the flexibility of full specified model. Define the rotated return as

$$e_t = (\bar{H}^{1/2})^{-1} r_t = \bar{H}^{-1/2} r_t, \tag{1.10}$$

we have

$$Var(e_t) = I_k.$$

The conditional covariance of the rotated returns is $Var[e_t|\mathcal{F}_{t-1}] \equiv \tilde{H}_t$, noting that $H_t = \bar{H}^{1/2}\tilde{H}_t\bar{H}^{1/2}$. Similarly, for the realized measure of covariance, we have $X_t = \bar{X}^{1/2}\tilde{X}_t\bar{X}^{1/2}$, where $\tilde{X}_t$ is the rotated realized measure of covariance. This leads to $\tilde{H}_t = \bar{H}^{-1/2}H_t\bar{H}^{-1/2}$ and $\tilde{X}_t = \bar{X}^{-1/2}X_t\bar{X}^{-1/2}$. The basic idea is that the variables H_t and X_t are normalized by their unconditional expectations so that the expectation of rotated variables are equal to identity matrices. The covariance targeting RBEKK model

is

$$\tilde{H}_t = (I_k - \tilde{B}\tilde{B}' - \tilde{A}\tilde{A}') + \tilde{B}\,\tilde{H}_{t-1}\,\tilde{B} + \tilde{A}\,\tilde{X}_{t-1}\,\tilde{A} \tag{1.11}$$

with

$$E[\tilde{H}_t] = E[\tilde{X}_t] = I_k.$$

This model has $2k$ dynamic parameters since we assume $\tilde{B}$ and $\tilde{A}$ are diagonal. It is easy to impose positive definiteness on $(I_k - \tilde{B}\tilde{B}' - \tilde{A}\tilde{A}')$ under this specification. For the conditional covariance matrix of unrotated returns H_t, the BEKK model in (1.10) and (1.11) implies the original model in (1.7)

$$H_t = \bar{H}^{1/2}\tilde{H}_t\bar{H}^{1/2} = CC' + B\,H_{t-1}\,B' + A\,X_{t-1}\,A' \tag{1.12}$$

where

$$B = \bar{H}^{1/2}\tilde{B}\bar{H}^{-1/2}, \quad A = \bar{H}^{1/2}\tilde{A}\bar{X}^{-1/2}, \quad CC' = \bar{H}^{1/2}(I_k - \tilde{B}\tilde{B}' - \tilde{A}\tilde{A}')\bar{H}^{1/2} \tag{1.13}$$

The rotated BEKK leads to a BEKK with full specification for the original variables, although the specification is constrained, since A and B depend on the Cholesky decomposition of $\bar{H}$ and $\bar{X}$.

Bauwens, Laurent, and Rombouts (2006) discuss the invariance of multivariate models to linear transformations of returns. A transformed model is invariant if it keeps the same dynamic specification (in scalar, diagonal, and full parameterizations). They also note that BEKK models are invariant to linear transformations except in the case of diagonal dynamics. Since $\tilde{A}$ is diagonal, directly derived from (1.13), A is a full asymmetric matrix. This results in a fully parameterized BEKK model for the unrotated variables. In this sense, fitting a diagonal rotated RBEKK model allows for richer dynamics for the unrotated returns in practice because we avoid the curse-of-dimensionality.

In the diagonal specification, $\tilde{A}$ and $\tilde{B}$ are assumed to be diagonal with elements $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$ for

$i = 1, ..., k$. This specification implies

$$\tilde{h}_{ij,t} = (1 - \kappa_{ij})1_{[i=j]} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}\tilde{h}_{ij,t-1} + \alpha_{ii}^{1/2}\alpha_{jj}^{1/2}x_{ij,t-1} \tag{1.14}$$

$$\kappa_{ij} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}. \tag{1.15}$$

for $i, j = 1, ..., k$. It is assumed that $\alpha_{ii}^{1/2} > 0$ and $\beta_{ii}^{1/2} > 0$. The cross-equation parameter restrictions between the variance and covariance equations are a feature of BEKK-type parameterizations. Covariance stationary is determined by the eigenvalues of the diagonal matrix $\tilde{A} \otimes \tilde{A} + \tilde{B} \otimes \tilde{B}$. For the rotated model given in (1.10) and (1.11), we ensure covariance stationarity and also the positive definiteness of $(I_k - \tilde{B}\tilde{B}' - \tilde{A}\tilde{A}')$, where for any $k \times k$ matrix $\tilde{A}$ with eigenvalues $\kappa_1, \ldots, \kappa_k$ and all $k \in \mathbb{Z}^+$, $\rho(A) := \max_i |\kappa_i|$.

Assumption 1 *In the rotated RBEKK model given by (1.10) and (1.11),* $\rho(\tilde{A} \otimes \tilde{A} + \tilde{B} \otimes \tilde{B}) \leq 1$, $\alpha_{ii}^{1/2} > 0$, *and* $\beta_{ii}^{1/2} > 0$.

In practice, $\alpha_{ii} + \beta_{ii} < 1$ is a necessary and sufficient condition for this assumption to hold and is easily imposed in estimation (see Engle and Kroner, 1995).

1.3.3 Common Persistence Specification

In the rotated model, $(\tilde{A} \otimes \tilde{A} + \tilde{B} \otimes \tilde{B})$ is a diagonal matrix with diagonal elements $\kappa_{ij} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}$. We adopt the Noureldin, Shephard, and Sheppard's (2014) common persistence (CP) specification for the diagonal structures. The CP specification imposes

$$\kappa_{ii} = \kappa, \quad i = 1, ..., k$$

implying common persistence for the diagonal elements of $\tilde{H}_t$. This results in the dynamic equation

$$\tilde{H}_t = (1 - \kappa)I_k + \tilde{B}\,\tilde{H}_{t-1}\,\tilde{B} + \tilde{A}\,\tilde{X}_{t-1}\,\tilde{A} \tag{1.16}$$

where $\tilde{B}$ is a diagonal matrix with elements $\beta_{ii}^{1/2} > 0$, $\tilde{A}$ is a diagonal matrix with elements $(\kappa - \beta_{ii})^{1/2}$, and $0 < \kappa < 1$ is a scalar parameter satisfying $\kappa \geq \max \beta_{ii}$. This implies

$$\tilde{h}_{ij,t} = (1-\kappa)1_{[i=j]} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}\tilde{h}_{ij,t-1} + (\kappa - \beta_{ii})^{1/2}(\kappa - \beta_{jj})^{1/2}x_{ij,t-1} \quad i,j = 1,...,k \qquad (1.17)$$

The CP specification has $k+1$ dynamic parameters compared to 2k dynamic parameters in the diagonal model. It imposes common persistence on the diagonal elements of $\tilde{H}_t$ through a common persistence parameter, κ, for the conditional variances. The condition for covariance stationarity is simply $\kappa < 1$ and $\beta_{ii} > 0$, which also guarantees that $(1-\kappa)I_k$ is positive definite. This specification allows the elements of $\tilde{H}_t$ to load freely on the lagged variances/covariances. Corresponding shocks will have different levels of ''smoothness", but the variances are restricted to have common persistence through κ.

Assumption 2 *In the rotated RBEKK model given by (1.16)* , $0 < \kappa < 1$ *and* $\kappa \geq \max \beta_{ii}$, *and* $\beta_{ii} > 0$.

1.4 Adaptive Lasso method

1.4.1 Definition and Notation

Since the measurement equation (1.8) is a fully parameterized linear model with $k^* = k(k+1)/2$ parameters in the coefficient matrix Γ that must be estimated. Equation (1.8) can be seen as a group of seemingly unrelated equations, each with one dependent variable corresponding to an element of the matrix logarithm of the realized matrix X_t. There are thus k^* univariate linear equations. Each equation has k^* parameters to link to the matrix logarithm of the realized covariance with all of the conditional covariances across different assets. From the literature on univariate realized GARCH models, the coefficient of the measurement equation is usually less than one if the measurement equation is specified in level as $x_t = \beta_0 + \beta_1 h_t + u_t$, and around one if is is specified in natural logarithm form $\log x_t = \beta_0 + \beta_1 \log h_t + u_t$ (Hansen, Huang, and Shek, 2012). Ideally, the coefficient matrix Γ should be a sparse matrix with elements on the diagonal between 0 and 1 and elements off the diagonal close to

zero. We achieve model parsimony by using a variable shrinkage technique to select the most significant variables and reduce all others to zero.

The Lasso method is a regularization technique for simultaneous estimation and variable selection in linear regression (Tibshirani, 1996). Parameters are estimated by least squares subject to an ℓ_1 norm slackness parameter penalty. Zou (2006) showed that the Lasso estimator is possibly inconsistent, so we adopt the Adaptive Lasso (ALasso) method. The ALasso method avoids inconsistency by utilizing an adaptively weighted ℓ_1 penalty. Define the ALasso estimator $\hat{\beta}^{ALasso}$ as

$$\hat{\beta}^{ALasso} = arg\, min_{\beta} \|y - X\beta\|^2 + \lambda \hat{w} |\beta|, \tag{1.18}$$

where $(n \times 1)$ vector y is the observed response variable, $(n \times q)$ matrix X is the matrix of prediction variables, $(q \times 1)$ vector β is the set of model coefficients determining the load on each predictor variable, and $\hat{w} = 1/|\hat{\beta}_{OLS}|$ is a vector of weights. The unpenalized OLS estimator is given by $\hat{\beta}_{OLS}$. The first term of (1.18) is the ℓ_2 Euclidean vector norm and the second term is the so-called "ℓ_1 penalty" where λ is a nonnegative regularization parameter. Equation (1.18) constitutes a convex optimization problem that does not suffer from the possibility of multiple local minima. The global minimizer can be found by applying the Least Angle Regression (LARS) algorithm of Efron, Hastie, Johnstone, and Tibshirani (2004).

1.4.2 Least Angle Regression

There are several algorithms that may be used to compute the solution of the shrinkage linear regression problem. These include the quadratic programming approach in Tibshirani (1996), the LARS in Efron, Hastie, Johnstone, and Tibshirani (2004), and so on. Different algorithms should produce similar results, and with minor modifications, they will produce identical solutions. Zou (2006) used the LARS algorithm to compute the solution for the ALasso estimator. Denote by $\mathcal{A}$ the active set including all non-zero variables selected by the LARS algorithm. The algorithm starts with the empty set of active variables, which is a constant model, and moves towards the full OLS regression. The correlation be-

tween each variable x and the response y is measured, and the variable with the highest correlation becomes the first variable to be included in the active set. The first direction of the move is towards the least squares solution using this single active variable. The angles between the variables and the residual vector shrink along this direction until another variable obtains the same correlation with the residual vector as the active variable. Then the new variable moves to the active set. The resulting direction is towards the least squares solution of the two active variables. After q steps, the full OLS solution is reached. The LARS algorithm is efficient since there is a closed form solution for the step length at each stage. Denoting the model estimate of y at iteration k by $\hat{y}^{(k)}$,$1 \leq k \leq q$, and the least squares solution including the newly added active variable by $\hat{y}_{OLS}^{(k+1)}$, the walk from $\hat{y}^{(k)}$ towards $\hat{y}_{OLS}^{(k+1)}$ can be formulated as $(1-\gamma^{(k)})\hat{y}^{(k)} + \gamma^{(k)}\hat{y}_{OLS}^{(k+1)}$ where $0 \leq \gamma^{(k)} \leq 1$. Estimating $\hat{y}_{OLS}^{(k+1)}$, the position where the next active variable is to be added, then amounts to estimating $\hat{y}_{LARS}$. We seek the smallest positive γ where correlations become equal, that is

$$x'_{i\in\mathcal{I}}(y - (1-\gamma)\hat{y}^{(k)} - \gamma\hat{y}_{OLS}^{(k+1)}) = x'_{j\in\mathcal{A}}(y - (1-\gamma)\hat{y}^{(k)} - \gamma\hat{y}_{OLS}^{(k+1)})$$

where $\mathcal{I}$ and $\mathcal{A}$ represent the respective inactive and active sets. Solving this expression for γ,

$$\gamma^{(k)} = \frac{(x_i - x_j)'(y - \hat{y}^{(k)})}{(x_i - x_j)'(\hat{y}_{OLS}^{(k+1)} - \hat{y}^{(k)})} = \frac{(x_i - x_j)'\epsilon^{(k)}}{(x_i - x_j)'\mathbf{d}}$$

where $\epsilon^{(k)}$ is the residual, $\mathbf{d} = \hat{y}_{OLS}^{(k+1)} - \hat{y}^{(k)}$ is the direction of the step, and $j \in \mathcal{A}$. As $\mathbf{d}$ is the orthogonal projection of $\epsilon^{(k)}$ onto the plane spanned by the variables in $\mathcal{A}$, we obtain $x'_j\epsilon^{(k)} = x'_j\mathbf{d} \equiv c$, representing the angle at the current breakpoint $\hat{y}^{(k)}$. Furthermore, the sign of the correlation between variables is irrelevant. Therefore,

$$\gamma^{(k)} = \min\left(\frac{x'_i\epsilon^{(k)} - c}{x'_i\mathbf{d} - c}, \quad \frac{x'_i\epsilon^{(k)} + c}{x'_i\mathbf{d} + c}\right), \quad 0 \leq \gamma^{(k)} \leq 1$$

where the two terms are for correlations of equal and opposite signs. The coefficients at the next step are given by

$$\beta^{(k+1)} = (1 - \gamma^{(k)})\beta^{(k)} + \gamma^{(k)}\beta_{OLS}^{(k+1)}$$

The full LARS algorithm is described in Algorithm 1.

Assuming the explanatory variables are standardized to have zero means and unit variance and that our response variable also has zero mean
Initialize the coefficient vector $\beta^{(0)} = 0$ and the fitted vector $\hat{y}^{(0)} = 0$
Initialize the active set $\mathcal{A} = \emptyset$ and the inactive set $\mathcal{I} = 1...q$, where q is the number of predicted variables
for $k = 0$ *to* $q - 2$ **do**
    Update the residual $\epsilon^{(k)} = y - \hat{y}^{(k)}$
    Find the maximal correlation $c = max_{i \in \mathcal{I}} |x_i' \epsilon^{(k)}|$
    Move the predictor $X^{(k)}$ corresponding to c from $\mathcal{I}$ to $\mathcal{A}$
    Calculate the least squares solution $\beta_{OLS}^{(k+1)} = (X_{\mathcal{A}}' X_{\mathcal{A}})^{(-1)} X_{\mathcal{A}}' y$, where $X_{\mathcal{A}}$ si the predictors in the active set
    Calculate the current direction $\mathbf{d} = X_{\mathcal{A}} \beta_{OLS}^{(k+1)} - \hat{y}^{(k)}$
    Calculate the step length $\gamma = \min_{i \in \mathbf{I}}^{+} \left(\frac{x_i' \epsilon - c}{x_i' \mathbf{d} - c}, \frac{x_i' \epsilon + c}{x_i' \mathbf{d} + c} \right),\ 0 \leq \gamma \leq 1$
    Update regression coefficients $\beta^{(k+1)} = (1 - \gamma)\beta^{(k)} + \gamma \beta_{OLS}^{(k+1)}$
    Update the fitted vector $\hat{y}^{(k+1)} = \hat{y}^{(k)} + \gamma \mathbf{d}$
Let $\beta^{(q)}$ be the full least squares solution $\beta^{(q)} = (X'X)^{(-1)} X'y$
Output the full set of estimated coefficient vectors $[\beta^{(0)}...\beta^{(q)}]$ with different vectors corresponding to models with different sets of regressors

Algorithm 1: LARS algorithm

1.4.3 Model selection

Each step of Algorithm 1 adds a covariate to the model until the full least squares solution is reached. It is natural to parameterize this process by the size of the coefficient vector at each step of the algorithm. Define K as the number of non-zero elements of the parameter vector β. The model with larger K will achieve a better fit. The question of how to choose a suitable model involves the trade-off between bias and variance, along with a desire for parsimony.

We choose the optimal model in terms of K by minimizing an information criterion. Two of the most popular criteria are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for the ALasso method proposed by Zou (2006)

$$\begin{aligned} AIC^{(K)} &= \left\| y - X\beta^{(K)} \right\|^2 + 2\sigma_\epsilon^2 K \\ BIC^{(K)} &= \left\| y - X\beta^{(K)} \right\|^2 + \log(n)\sigma_\epsilon^2 K \end{aligned} \tag{1.19}$$

The measure σ_ϵ^2 represents the residual variance of the OLS model, $\sigma_\epsilon^2 = \frac{1}{n-K}||Y - X'\beta_{ols}||^2$. AIC and BIC have different asymptotic properties. AIC tends to select the model with optimal predictive performance, while BIC identifies the true sparse model if the true model is in the candidate list (Zou, Hastie, Tibshirani, et al., 2007). Our intention in using the measurement equation (1.8) is to update the realized volatility measure beyond one step and make multiple-period predictions. Accuracy of these predictions is our main concern. From a forecasting perspective, AIC is the better information criterion for model selection. With the entire ALasso solution path computed by the LARS algorithm, we select the optimal model by identify the smallest value of AIC and the corresponding K.

As a selection criterion, AIC is a better choice than the Residual Sum of Squares (RSS). The RSS carries no penalty for additional parameters, and if used to select the optimal model, it will select the model that gives the smallest residuals. This invariably corresponds to the model with the most parameters. Fully parameterized OLS minimizes the sum of squared residuals so the RSS of a linear regression fit by OLS is minimal. Such models have a tendency to overfit the data and are unlikely to reliably predict future observations.

For comparison of "shrunk" models with fully-specified models, we define φ as the relative parameter size

$$\varphi(\beta) = \sum_{j=1}^{q} |\hat{\beta}_j| / \sum_{j=1}^{q} |\hat{\beta}_{j_{ols}}|, \varphi \in [0, 1] \tag{1.20}$$

Picking a suitable model for a particular analysis corresponding to selecting a suitable value of φ. The quantity φ is used in the empirical analysis to show the relative size of the coefficient vector.

Given the ALasso problem defined in (1.18) for $\hat{\beta}$, computational details and model selection pro-

cedures are shown in Algorithm 2. Applying the LARS to the estimation of the measurement equation

1 Define new covariates ,$x_j^* = x_j/\hat{w}_j, j = 1, 2, ..., q$. where $\hat{w}$ is the adaptive weight [1]
2 Solve the ALasso problem defined in (1.18) with the LARS algorithm (1) ,

$$\beta^{ALasso} = arg\, min_\beta \|y - X^*\beta\|^2 + \lambda|\beta|$$

which generates a sequence of solutions $[\beta^{(0)}, ..., \beta^{(q)}]$
3 Substitute the ALasso estimators from step 2 into the information criterion equation (1.19). This provides us a selection criteria value for each step.
4 Pick the $\tilde{\beta}$ which minimizes AIC (BIC)
5 Output $\hat{\beta}_j = \tilde{\beta}_j/\hat{w}_j, j = 1, 2, ..., q$ to reverse to original measure. This is the ALasso estimate.

Algorithm 2: The ALasso method with the LARS Algorithm

(1.8) requires each prediction variable in H to be normalized by subtracting the mean and scaling so that $\sum_{i=1}^{T} \|h_{1i,t}^*\| = 1$. The measurement equation can then be estimated in the normalized form

$$\begin{bmatrix} x_{11,1}^* \\ x_{11,2}^* \\ \vdots \\ x_{11,T}^* \end{bmatrix} = \begin{bmatrix} h_{11,1}^* & h_{12,1}^* & h_{13,1}^* \\ h_{11,2}^* & h_{12,2}^* & h_{13,2}^* \\ \vdots & \vdots & \vdots \\ h_{11,T}^* & h_{12,T}^* & h_{13,T}^* \end{bmatrix} \begin{bmatrix} \gamma_{11}^* \\ \gamma_{12}^* \\ \gamma_{13}^* \end{bmatrix} + \begin{bmatrix} u_{11,1} \\ u_{11,2} \\ \vdots \\ u_{11,T} \end{bmatrix} \tag{1.21}$$

where $x_{11,t}^*$ is centered and $h_{1i,t}^*$ is centered and normalized such that each variable has zero mean and unit Euclidean length. $\gamma_{1i}^* = e \times \gamma_{1i}$, such that $h_{1i,t}^* \times z = h_{1i,t}$, where $z = \sum_{i=1}^{T} \|h_{1i,t}\|$. The intercept can then be estimated as $\hat{d}_i = (1/T)\sum_{t=1}^{T}(x_{i,t} - \hat{\Gamma}_i h_t)$ where $\hat{\Gamma}_i$ is the estimate of the ist row of Γ in measurement equation (1.8).

[1] Since in ALasso coefficients are weighted penalized, we need to rescale the predictors to acheive the ALasso method. As the sample size grows, the weights for zero-coefficient predictors get inflated (to infinity), whereas the weights for nonzero coefficient predictors converge to a finite constant. It is numerically stable.

1.5 Estimation and Inference

1.5.1 Estimation for RBEKK Equation

The asymptotic properties of the quasi-maximum likelihood estimator within the RBEKK model given by (1.6) and (1.7) are important for making reliable inference. We estimate the parameters in a rotated model (1.16) and then reverse the linear transformation to recover the unrotated parameters. The BEKK equation (1.7) and rotated model (1.16) are parameterized with a finite-dimensional $(\delta \times 1)$ parameter vector $\theta \in \Theta \subset \mathbb{R}^{\delta}$. Decompose $\theta = (\theta_s', \theta_d')'$, where $\theta_s = (\bar{h}, \bar{x})$ denotes the model's static parameters that can be estimated by unconditional moments. Let θ_d denote the vector of dynamic parameters indexing $\tilde{A}$ and $\tilde{B}$ in the rotated model. The covariance targeting RBEKK (1.11) can be estimated by QMLE.

Specifically, the static parameters are estimated by the following moment estimator

$$\hat{\bar{h}}' = vech(\operatorname{logm}(\hat{\bar{H}})), \qquad \hat{\bar{x}}' = vech(\operatorname{logm}(\hat{\bar{X}}))$$

where

$$\hat{\bar{H}} = T^{-1} \sum_{t=1}^{T} r_t r_t', \qquad \hat{\bar{X}} = T^{-1} \sum_{t=1}^{T} X_t$$

These estimates are then decomposed into $\hat{\bar{H}}^{1/2}$ and $\hat{\bar{X}}^{1/2}$ and used to construct the time series of rotated returns and the realized kernel $\hat{e_t} = \hat{\bar{H}}^{-1/2} r_t$ and $\hat{\bar{X}}_t = \hat{\bar{X}}^{-1/2} X_t \hat{\bar{X}}^{-1/2}$ for $t = 1, 2, ..., T$. In the second step, θ_d is estimated by Gaussian QMLE. The joint density is the product of all conditional densities, so the log-likelihood function of the joint distribution is the sum of the logarithms of the multivariate-normal distribution in this case. Letting $L_t(\theta_d; \theta_s)$ be the likelihood function at time t, and

$l(\theta_d; \theta_s)$ be the log of the joint likelihood function,

$$l(\theta_d; \theta_s) = \sum_{t=1}^{T} \log L_t(\theta_d; \theta_s) \tag{1.22}$$

$$logL_t(\theta_d; \theta_s) = constant - \frac{1}{2}(\log|\tilde{H}_t| + r_t'\tilde{H}_t^{-1}r_t) \tag{1.23}$$

where θ_s is estimated in the previous step. We take the unconditional variance of the return of assets as the initial value of $\tilde{H}_t$ in the first period. Then the joint likelihood can be calculate in a recursive way. The QML estimator $\hat{\theta}_d$ is

$$\hat{\theta}_d = arg\, max_{\theta_d \in \Theta}\, l(\theta_d; \hat{\theta}_s)$$

Asymptotic theory for QML estimation of the BEKK model is discussed in several literature, for example, Comte and Lieberman (2003) and Hafner and Preminger (2009). For the closely related covariance-targeting BEKK model, Pedersen and Rahbek (2014) prove consistency and asymptotic normality. The rotated BEKK model applies the same parameterizations but to rotated returns.

Using two-step maximum likelihood estimation, if the usual assumptions for the validity of QMLE are satisfied, the estimates are consistent and their asymptotic distribution is

$$\sqrt{T}\left(\begin{bmatrix}\hat{\theta}_s \\ \hat{\theta}_d\end{bmatrix} - \begin{bmatrix}\theta_s \\ \theta_d\end{bmatrix}\right) \to \mathbb{N}(0, V)$$

with

$$V = \begin{bmatrix} G_{\theta_s}^{-1} & -G_{\theta_s}^{-1}G_{\theta_d}M^{-1} \\ 0 & M^{-1} \end{bmatrix} E\left[\frac{\partial l}{\partial \theta}\frac{\partial l}{\partial \theta'}\right] \begin{bmatrix} G_{\theta_s}^{-1} & -G_{\theta_s}^{-1}G_{\theta_d}M^{-1} \\ 0 & M^{-1} \end{bmatrix}$$

where

$$G_{\theta_s} = E\left[\frac{\partial g(\theta_s, \theta_d)}{\partial \theta_s'}\right]; G_{\theta_d} = E\left[\frac{\partial g(\theta_s, \theta_d)}{\partial \theta_d'}\right]; M = E\left[\frac{\partial m(\theta_d)}{\partial \theta_d'}\right]$$
$$g(\theta_s, \theta_d) = \left[\frac{\partial l(\theta_s, \theta_d)}{\partial \theta_s}\right]; m(\theta_d) = \left[\frac{\partial l(\theta_s, \theta_d)}{\partial \theta_d}\right]$$

The proof can be found in Newey and McFadden (1994).

1.5.2 Estimation of the Measurement Equation

Since the measurement equation is over-parameterized, the ALasso method is used to simultaneously estimate and shrink its parameters. As mentioned, the measurement equation is actually a system of equations that can be estimated using SUR methods. Using the LARS algorithm, we can estimate the measurement equation efficiently. Although the formula for the variance of estimator of the ALasso method is given in Zou (2006),

$$var(\hat{\beta}_{\mathcal{A}}) = \sigma^2(X'_{\mathcal{A}}X_{\mathcal{A}} + \lambda \sum(\hat{\beta}_{\mathcal{A}}))^{-1}X'_{\mathcal{A}}X_{\mathcal{A}}(X'_{\mathcal{A}}X_{\mathcal{A}} + \lambda \sum(\hat{\beta}_{\mathcal{A}}))^{-1}$$

where the subscript $\mathcal{A}$ denoting the sub-matrices consisting of the columns (variables) of X corresponding to the indices in the active set selected by the ALasso method. and $\sum(\hat{\beta}_{\mathcal{A}}) = diag(\frac{\hat{w}_1}{|\beta_1|}, ..., \frac{\hat{w}_k}{|\beta_k|})$. σ^2 is the variance of the error term in the model, including all the regressor residuals, which is estimated with the sample variance of the OLS residuals. Difficulties arise in extracting λ from the LARS algorithm. Instead of that, we typically avoid this issue by resorting to the bootstrap. $var(\hat{\beta}_{\mathcal{A}})$ can be calculated using the bootstrapping method proposed by Efron (1979). Bootstrapping is a nonparametric method which allows statistics to be computed or distributions to be estimated by drawing randomly with replacement from a given dataset. The Bootstrap Algorithm is described as following:

1 Draw N pairs of independent bootstrap samples$(y_1^*, x_1^*), (y_2^*, x_2^*), ..., (y_n^*, x_n^*)$ with replacement from the original data
2 Estimate the ALasso parameters $\tilde{\beta}$ for these bootstrap samples
3 Repeat step 1,2 for M times, so we get M of estimates: $\tilde{\beta}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_M$
4 Estimate the variance $var(\hat{\beta}$ using empirical moments $\widehat{var(\tilde{\beta})} = \left[\frac{1}{M-1}\sum_{m=1}^{M}(\tilde{\beta}_m - \bar{\tilde{\beta}})^2\right]$ where $\bar{\tilde{\beta}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}_m$

Algorithm 3: The Bootstrap Algorithm

1.6 Empirical Illustration

To examine the empirical effects of model parsimony, we estimate the RBEKK model with and without a sparse coefficient matrix in the measurement equation. Results are then compared in terms of their forecast abilities.

1.6.1 Data Description

The model is applied to high-frequency prices for ten liquid stocks from the S&P 100 [2]. The sample period is Jan 03, 2006 to April 30, 2012 and the source of data is the TAQ database. We focus on the noise-robust realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) as our choice for X_t. To control for overnight effects, the opening and closing 15 minutes of trading are excluded from the analysis. Returns are calculated as the first difference of the log-price and open-to-close returns are used at the daily frequency.

Plots of the daily return and annualized realized volatility of the ten stocks can be found in the appendix. The sharp increase in realized volatility from 2008 to 2009 is associated with major upheavals in financial markets during the subprime mortgage crisis. The increase in volatility is especially pronounced after the collapse of Lehman Brothers in the middle of September 2008. Visual inspection of the plots of annualized volatility suggests that the financial crisis lasted approximately from August 2008 to April 2009.

1.6.2 Estimation Results for RBEKK model

The following results were obtained when the model is applied to rotated data. Estimates of $\tilde{A}$ and $\tilde{B}$ along with their standard errors are shown in Tables 1.1 and 1.2. The estimates of the unrotated parameters A and B are recovered according to (1.13). From the literature on GARCH models using realized measures, the variance/covariance process in this model is more responsive to innovations. As a result, the elements (in the diagonal positions) in $\tilde{A}$ in (1.11) will be larger than those in traditional

[2]The ticker symbols are AXP(American Express), BAC(Bank of America), DD(DuPont), GE(General Electric), IBM, INTC(Intel), JPM(JPMorgan Chase), KO(Coca-Cola), MSFT(Microsoft), and XOM(Exxon Mobil).

GARCH models. The means of the diagonal elements in the coefficient matrices are 0.722 for $\tilde{B}$ and 0.209 for $\tilde{A}$, which is in line with estimates found in other studies. The estimate for $\tilde{B}$ implies that the dynamic of H_t will be smooth, though less smooth than the traditional GARCH model, which typically has a smoothing parameter around 0.9. The estimate of $\tilde{A}$ indicates that the realized measure plays an important role in updating H_t.

Table 1.1: Parameter Estimates of the matrix $\tilde{A}$ in the rotated RBEKK model

	AXP	BAC	DD	GE	IBM	INTC	JPM	KO	MSFT	XOM
AXP	0.5139 (0.0257)									
BAC		0.8070 (0.0614)								
DD			0.6297 (0.0652)							
GE				0.5578 (0.034)						
IBM					0.3817 (0.0097)					
INTC						0.2727 (0.0008)				
JPM							0.5759 (0.0272)			
KO								0.4016 (0.0143)		
MSFT									0.3961 (0.0155)	
XOM										0.3938 (0.0072)

Table 1.2: Parameter Estimates of the matrix $\tilde{B}$ in the rotated RBEKK model

	AXP	BAC	DD	GE	IBM	INTC	JPM	KO	MSFT	XOM
AXP	0.8466 (0.0414)									
BAC		0.5741 (0.0434)								
DD			0.7644 (0.0791)							
GE				0.8183 (0.0492)						
IBM					0.9138 (0.0225)					
INTC						0.9521 (0.0036)				
JPM							0.8057 (0.0375)			
KO								0.9053 (0.0317)		
MSFT									0.9077 (0.035)	
XOM										0.9087 (0.0157)

For the BEKK paramerization, using (1.13), the estimates of A and B are

$\hat{A} = \hat{\tilde{H}}^{1/2} \tilde{A} \hat{\tilde{X}}^{-1/2} =$

$$\begin{pmatrix}
0.4620 & 0.1319 & 0.0567 & 0.0060 & -0.0349 & 0.0286 & -0.0222 & -0.0029 & -0.0068 & 0.0243 \\
-0.0129 & 0.8201 & 0.0298 & 0.0048 & -0.0646 & 0.0699 & -0.0289 & 0.0261 & 0.0041 & 0.0030 \\
0.0091 & 0.0535 & 0.5698 & 0.0103 & -0.0369 & 0.0181 & 0.0039 & 0.0111 & 0.0154 & 0.0155 \\
0.0008 & 0.0919 & 0.0480 & 0.4424 & -0.0264 & 0.0679 & -0.0168 & 0.0892 & 0.0484 & 0.0427 \\
0.0107 & 0.0172 & 0.0358 & 0.0143 & 0.2984 & 0.0817 & -0.0383 & 0.0583 & 0.0391 & 0.0439 \\
0.0297 & 0.0417 & 0.0484 & 0.0223 & 0.0137 & 0.1994 & -0.0643 & 0.0076 & 0.0351 & 0.0523 \\
-0.0060 & 0.1748 & 0.0438 & 0.0159 & -0.0207 & 0.0485 & 0.4628 & 0.0301 & 0.0128 & 0.0057 \\
0.0126 & 0.0041 & 0.0188 & 0.0073 & -0.0158 & 0.0800 & -0.0471 & 0.3495 & 0.0641 & 0.0553 \\
0.0118 & 0.0289 & 0.0336 & 0.0004 & -0.0061 & 0.0917 & -0.0453 & 0.0464 & 0.3390 & 0.0465 \\
0.0173 & 0.0347 & 0.0403 & 0.0133 & -0.0179 & 0.1451 & -0.0586 & 0.0596 & 0.0810 & 0.3228
\end{pmatrix}$$

$\hat{B} = \hat{\tilde{H}}^{1/2} \tilde{B} \hat{\tilde{H}}^{-1/2} =$

$$\begin{pmatrix}
0.8693 & -0.0003 & 0.0132 & 0.0082 & 0.0020 & -0.0210 & 0.0052 & -0.0010 & -0.0026 & -0.0003 \\
0.0008 & 0.4887 & 0.0231 & 0.0258 & 0.0065 & -0.0543 & 0.0255 & -0.0103 & -0.0100 & 0.0070 \\
-0.0074 & -0.0085 & 0.7479 & 0.0087 & -0.0033 & -0.0208 & 0.0034 & -0.0051 & -0.0112 & -0.0020 \\
-0.0005 & -0.0041 & 0.0190 & 0.8257 & 0.0076 & -0.0672 & 0.0191 & -0.0306 & -0.0278 & 0.0006 \\
0.0025 & 0.0034 & 0.0270 & 0.0378 & 0.9185 & -0.0710 & 0.0173 & -0.0136 & -0.0235 & 0.0056 \\
0.0216 & 0.0386 & 0.0167 & 0.0641 & 0.0489 & 0.9680 & 0.0438 & 0.0101 & 0.0143 & 0.0859 \\
-0.0020 & -0.0133 & 0.0147 & 0.0151 & 0.0032 & -0.0449 & 0.8449 & -0.0135 & -0.0117 & 0.0035 \\
0.0003 & 0.0069 & 0.0210 & 0.0722 & 0.0237 & -0.0799 & 0.0288 & 0.9065 & -0.0245 & 0.0148 \\
0.0024 & 0.0070 & 0.0238 & 0.0497 & 0.0248 & -0.0710 & 0.0194 & -0.0019 & 0.9109 & 0.0276 \\
0.0085 & 0.0089 & 0.0256 & 0.0546 & 0.0266 & -0.1012 & 0.0226 & -0.0061 & -0.0243 & 0.9163
\end{pmatrix}$$

Asymmetry in A and B is typical with rotated models. The standard error of the estimates of A and B can be obtained from the estimation of the rotated model using the delta method.

1.6.3 The ALasso method Under Different Selection Criteria

The performance of the three different estimators can be ascertained by comparing the Residual Sum-of-Squares (RSS) and MSE, as shown in Table 1.3. This comparison is made for the measurement equation in (1.8). The OLS estimator has the smallest RSS and the BIC-based shrinkage estimator has the largest.

Since the OLS estimator is fully specified, this ranking should come as no surprise. OLS will generally obtain the best in-sample fit, while BIC-based shrinkage selects the most parsimonious model. The AIC-based shrinkage estimator has the smallest MSE since it balances the competing concerns of model fit and parsimony.

Table 1.3: Fit for Measurement Equation Estimators (1.8)

Minimizing based on	OLS	AIC	BIC
RRS	348.9	351.91	372.06
MSE	0.227	0.225	0.236

Note the $MSE = RSS/(N - K)$, where N is the number of observations and K is the number of variables.

The two information criteria, AIC and BIC, can be compared in terms of the relationship between relative model size $\varphi(\beta)$ defined in (1.20) and the model fit MSE. Figure 1.1 shows the relationship between the ALasso estimates and the $\varphi(\beta)$. The ALasso method continuously shrinks the coefficients toward zero as $\varphi(\beta)$ decreases; some coefficients shrink exactly to zero if the penalty is sufficiently large. In Figure 1.2, we can see the number of non-zero estimates decreasing as the relative bound decreases (penalty constraint increases). Both figures show the differences in the models that result from variable selection according to different criteria. The broken line gives the AIC break point while the solid line gives the BIC break point. The selected model under AIC has AIC=365.54 and φ= 0.4377, while the selected model under BIC has BIC=390.48 and φ= 0.1196. These values of φ validate the sparsity of the true model. Roughly, this corresponds to keeping just 40% of the predictors when using AIC as the criterion for selection. In Figure 1.3, the mean square error of the ALasso estimate decreases as the relative bound φ increases (penalty constraints decreases). The curve has a pronounced kink, which indicates that a portion of the parameters contribute only minimally to reducing the MSE. As a result, optimal selection of φ will occur near the kink. It seems AIC and BIC has similar effect in reducing MSE in Figure 1.3.

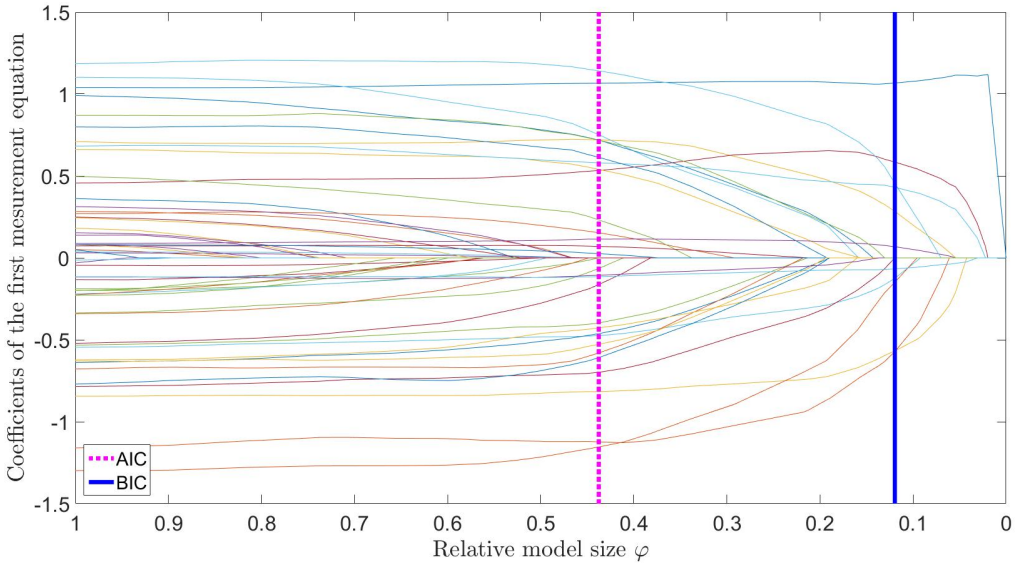

Figure 1.1: ALasso shrinkage of the coefficients of the first measurement equation in (1.8).

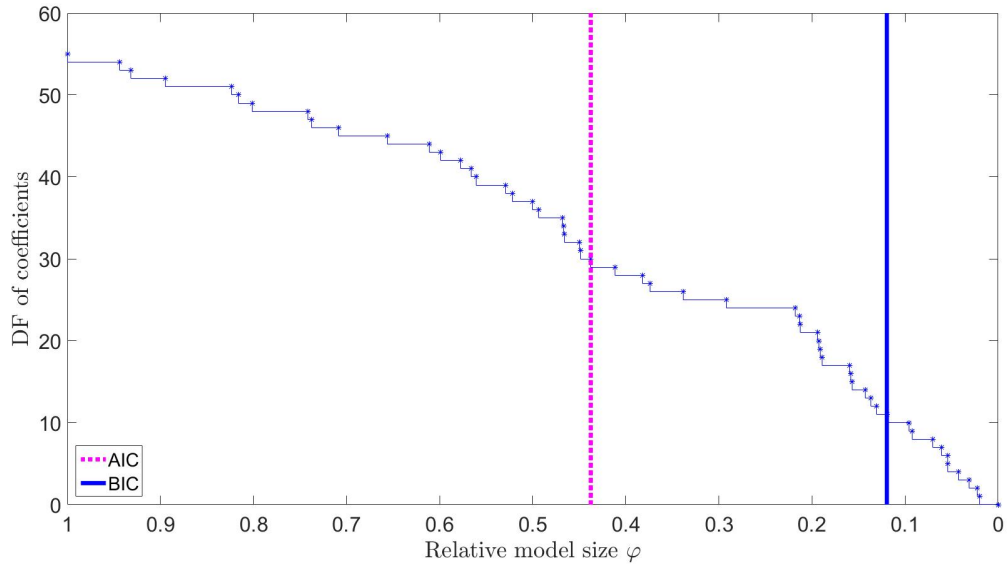

Figure 1.2: Shrinkage of Degrees of Freedom. DF refers to the number of non-zero parameters.

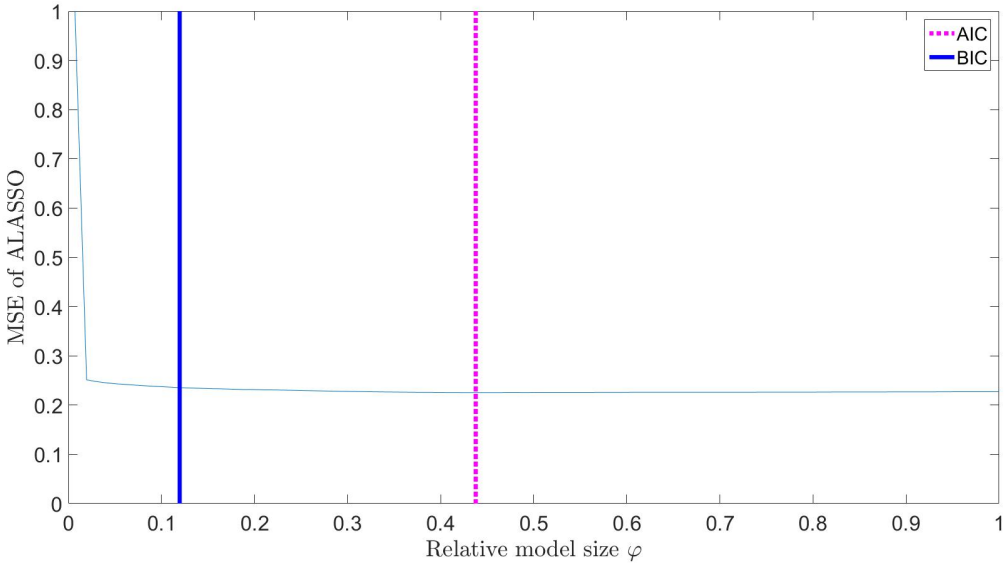

Figure 1.3: Mean Squared Error ALasso as a function of relative model size

1.6.4 Estimation Results for the Measurement Equation

For the measurement equations, each realized variance (covariance) is a function of the corresponding conditional variance (covariance) and the conditional covariances of the pairs of all other assets (after taking the matrix logarithm). Assuming that there are three assets – A, B, and C – the realized variance of asset A is updated by the conditional variance of A, and conditional covariance of A and B, the conditional covariance of A and C, and the conditional covariance of B and C. Each conditional variance/covariance is the main driver in updating its counterpart in the realized measure. As a result, the coefficient matrix should have larger elements in the diagonal positions. Most elements shrunk by the ALasso method are in the off-diagonal positions. Some diagonal elements will also shrink due to the nonlinear transformation by matrix logarithm reallocate the linkage between the corresponding elements in conditional variance/covariance matrix and realized kernel.

In the case of ten assets, the size of the coefficient matrix Γ is 55*55. On the diagonal, the elements in rows 1, 11, 20, 28, 35, 41, 46, 50, 53, and 55 belong to the matrix logarithm of the realized variance of each asset. These are much larger than the elements corresponding to covariances. In Table 1.4,

estimates of the matrix corresponding to the variance of each asset are compared under OLS and the ALasso method (AIC and BIC) estimation methods. In the last two rows, the percentage of non-zero coefficients in the matrix Γ, and the MSE, are given. As the OLS estimator does not shrink any variables, the percentage of non-zero coefficients is 100%. The number of non-zero coefficients under the AIC-based shrinkage estimator is 40.63%, much larger than the BIC-based shrinkage estimator's 12.79%. Again, we find that BIC emphasizes the parsimony of the model. However, AIC is more accurate in terms of MSE. In the second to last row of the table, we show the number of non-zero elements on the diagonal of the coefficient matrix. The AIC-based shrinkage estimator results in 74.55% of the elements being non-zero, while the BIC-based estimator results in 67.27% of the elements being non-zero. The coefficient matrix estimated by the ALasso method is essentially a diagonal matrix with a high percentage of shrinkage occuring in off-diagonal positions. The fourth to last row of the table reports the mean of the coefficient for each variance. The value for BIC is 0.889, the largest of all the models, since this estimation method implies the most shrinkage.

To check the fit of the measurement equation, we focus on the test of autocorrelation of the residuals. If there are autocorrelation of residuals, we may need to add lags of dependent variables. The Lagrange multiplier (LM) test for autocorrelation developed by Breusch (1978) and Godfrey (1978) has become a standard tool in applied econometrics. It is often applied to the residuals of a multivariate regression such VAR or SUR of time series model. A vectorized version of LM test is proposed Doornik (1996) by testing the significance of lagged residuals $\hat{u}_{t-h}$ (obtained under the null hypothesis) in the auxiliary regression:

$$u_t = d + \Gamma\, h_t + b_1 \hat{u}_{t-1} + ... + b_h \hat{u}_{t-h} + v_t \tag{1.24}$$
$$v_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_v)$$

We wish to test the significance of lagged residuals, so the null hypothesis is

$$H0 : b_1 = ... = b_h = 0. \tag{1.25}$$

Table 1.4: Estimates of coefficient matrix Γ

Index of diagonal position	OLS	AIC	BIC
1	1.04(0.05)	1.07(0.03)	1.01(0.03)
11	0.72(0.05)	0.75(0.03)	0.85(0.02)
20	0.73(0.07)	0.75(0.05)	0.80(0.03)
28	0.78(0.09)	0.77(0.05)	0.81(0.06)
35	1.00(0.07)	1.04(0.03)	1.12(0.03)
41	0.80(0.04)	0.78(0.02)	0.87(0.03)
46	0.72(0.07)	0.76(0.05)	0.89(0.06)
50	0.62(0.08)	0.67(0.04)	0.95(0.03)
53	0.50(0.06)	0.51(0.04)	0.56(0.02)
55	0.86(0.09)	0.88(0.03)	0.93(0.03)
Mean	0.78	0.80	0.88
Nonzero Pct.	100%	74.55%	62.27%
Overall Nonzero Pct.	100%	40.63%	12.79%
MSE of the measurement equation	0.201	0.199	0.203

Estimates of the coefficient matrix Γ of the measurement equation under OLS and ALasso using AIC and BIC criteria. The numbers in parentheses are standard errors. The Nonzero Pct means the percentage of nonzero estimates in the diagonal position of coefficient matrix Γ. The Overall Nonzero Pct means the percentage of nonzero estimates of all position in coefficient matrix Γ.

which implies no autocorrelation of residuals for lag one through lag h. The test statistic is constructed as:

$$LM = TnR_m^2 \tag{1.26}$$

with an asymptotic $\chi^2_{hn^2}$ under null . n is the dimension of the system, h is the number added lagged regressors in the each auxiliary regression and T is the sample size. R_m^2 is a R^2-type measures of goodness of fit:

$$R_m^2 = 1 - (1/n)tr(\hat{v}\hat{v}')(\hat{u}\hat{u}') \tag{1.27}$$

where v and u are p^* by T matrix of residuals from the auxiliary regressions and measurement equations. Since the dimension is large relative to the sample size, we only test the lag for 2. LM value for the measurement equation with OLS and LASSO are 5717 and 2421, and the p-value are 0.999 and 1, which can't reject the null hypothesis.

1.7 Model predictive Ability Evaluation

We compare the out-of-sample performance of the ALasso method and OLS, when estimating the measurement equation. Let the loss function $L_{t,s}(H^0_{t+s}, \hat{H}^i_{t+s|t})$ denote the loss at time t resulting from the s-step ahead forecast using model i. Under this notation, H^0_{t+s} denotes the true (latent) covariance matrix and $\hat{H}^i_{t+s|t}$ is the s-step ahead forecast of model i conditional on the information set at time t. Since the latent covariance H^0_{t+s} is unobservable, we use the realized kernel as a proxy such that $\hat{H}^0_{t+s} = X_{t+s}$. We use the quasi-likelihood (QLIK) loss function:

$$L_{t,s}(\hat{H}^0_{t+s}, H^i_{t+s|t}) = \log|H^i_{t+s|t}| + tr((H^i_{t+s|t})^{-1}\hat{H}^0_{t+s}). \tag{1.28}$$

The loss function in (1.28) is the negative of the log-likelihood of a Wishart density function, excluding the constant terms. The QLIK loss function can be used to consistently rank volatility models since it

is robust to noise in the proxy realized measure X_{t+s} (Patton and Sheppard, 2009; Laurent, Rombouts, and Violante, 2013).

Denote the total sample size by is T, dividing it into T_{in} of in-sample and T_{out} of out-of-sample portions. Then $T_{out} - s + 1$ is the number of s-step ahead predictions, the out-of-sample observations span $T_{in} + s$ through $T_{in} + T_{out}$. The estimation window (in-sample) portion runs from January 4, 2006 to April 29, 2011 with a sample size of 1340. The out-of-sample portion runs from May 2, 2011 to April 30, 2012 with a sample size of 252. Given the loss function, the loss difference is

$$D_{t,s} = L_{t,s}(\hat{H}^0_{t+s|t}, \hat{H}^{Lasso}_{t+s|t}) - L_{t,s}(\hat{H}^0_{t+s|t}, \hat{H}^{OLS}_{t+s|t}),$$
$$t = T_{in}, T_{in} + 1, ..., T - s.$$

The average loss is then

$$\bar{D}_s = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} D_{t,s}. \tag{1.29}$$

Average loss was used in tests of predictive ability by Diebold and Mariano (1995) and West (1996). The null and alternative hypotheses are

$$H0 : E[D_{t,s}] = 0,$$
$$H1 : E[D_{t,s}] > 0,$$
$$or \quad H2 : E[D_{t,s}] < 0.$$

The test statistic is constructed as

$$T_{loss} = (T_{out} - s + 1)^{1/2} \frac{\bar{D}_s}{\hat{S}_d^{1/2}}, \tag{1.30}$$

where

$$\hat{S}_d = \sum_{j=-s}^{s} \hat{\tau}(j),$$

$$\hat{\tau}(j) = \hat{\tau}(-j) = (T_{out} - s + 1)^{-1} \sum_{t=T_{in}}^{T-s+1} (D_{t-j,s} - \bar{D}_s)(D_{t,s} - \bar{D}_s),$$

$$j = 1, 2, ..., s.$$

Significantly negative values of the test statistic indicate that the measurement equation estimated with the ALasso method has superior forecast performance compared to the OLS estimate. The test of forecast ability is conducted on the rotated model using two different estimators for the measurement equations. We estimate the model with an expanding window (recursively), starting from observation number 1342, and then use the parameter estimates to obtain forecasts of H_t at horizons $s = 1, 2, 3, ..., 22$ days. The parameters are updated every 5 days, so the expanding window grows by 5. The estimation and prediction are repeated 50 and 250 times respectively.

Results for the test are given in Table 1.5. The results for 10 assets show that the difference in loss function is statistically significant and positive at the 10% level at horizons $s = 8, 9, 10$. The difference is statistically significant and negative at the 10% level at horizon $s = 20$ and at the 5% level at horizons $s = 21, 22$. The results with 15 stocks[3] show significantly negative values at the 10% significance level at horizons $s = 3, 4$ and at the 5% significance level at horizons $s = 5, 6, ..., 22$. With 20 stocks,[4] significantly negative values occur at the 10% significance level at horizons $s = 4, 20, 21$ and at the 5% significance level at horizons $s = 5, 6, ..., 19$. Broadly, the measurement equation estimated with the ALasso method outperforms the measurement equation estimated with OLS. The difference in performance between the two estimators increases as the number of assets under consideration increases. The statistics $\bar{D}_s$ show improvement in the 15 and 20 stock cases, except at the far end of the forecast horizon. The advantageous predictive power of the ALasso method in the 20 stock case is diminished at long forecasting horizons. Overall, the model using the ALasso method performs significantly better

[3]The ticker symbols are the original 10 plus C, CSCO, DELL, F, and WFC

[4]The ticker symbols are original 15 plus FCX,GS,MS,ORCL, and PFE.

than OLS methods, and this effect is more pronounced when additional assets are considered.

Table 1.5: Statistical Test of Predictive Ability

Prediction Step	Number of Stocks		
	10	15	20
2	0.116	-1.285	0.980
3	0.263	-1.645*	-0.248
4	0.503	-1.679*	-1.539*
5	0.864	-1.686**	-1.984**
6	1.184	-1.681**	-2.077**
7	1.271	-1.692**	-2.117**
8	1.389*	-1.761**	-2.143**
9	1.392*	-1.763**	-2.08**
10	1.370*	-1.754**	-2.066**
11	1.255	-1.847**	-2.053**
12	1.078	-1.787**	-2.04**
13	1.005	-1.924**	-2.026**
14	0.73	-2.003**	-2.032**
15	0.486	-2.054**	-2.015**
16	0.187	-2.095**	-2.018**
17	-0.149	-2.216**	-1.996**
18	-0.543	-2.263**	-1.92**
19	-1.009	-2.273**	-1.78**
20	-1.462*	-2.246**	-1.59*
21	-1.894**	-2.213**	-1.32*
22	-2.216**	-2.248**	-1.015

Test of predictive ability: Loss(ALasso)-Loss(OLS). A negative test value indicates model using the ALasso method has better predictive ability. The critical values are: 2.4 (1%), 1.68 (5%), 1.3 (10%). * denotes significance at the 10% level, ** significance at the 5% level, and *** significance at the 1% level

There are several different loss functions other than the likelihood value (see Patton, 2011). We choose the loss function constructed by mean squared error (MSE) to check how much gain using ALasso to forecast the covariance matrices compared with the model using OLS. The sign of the difference of predictive covariance matrices is negative if the Alasso is better. The loss function is constructed

as:

$$L_{t,s}(\hat{H}^0_{t+s}, H^i_{t+s|t}) = 1/N^2 \sum_{m,n=1}^{N} (h(mn)^i_{t+s|t} - \hat{h}(mn)^0_{t+s})^2. \tag{1.31}$$

where $h(mn)^i_{t+s|t}$ is the element in mth row and nth column in $H^i_{t+s|t}$, similar definition goes with $\hat{h}^{\ 0}_{mnt+s}$. Comparing (1.28) with (1.31), the former loss function represents which method are more possibly predict the true covariance the later indicate which method is closer to the true covariance with respect of the estimated value. The result of portfolio with 10, 15 and 20 assets for each forecast step are reported in Table 1.6. From the table, we can see that the mode with Alasso performs better for the portfolio with more number of assets and for the further steps in forecasting. Further, we can check the economical difference by inserting these predictive covariances into asset pricing formula or calculating the Value at risk.

1.8 Conclusion

In this chapter we present the realized rotated BEKK model to model vectors of daily returns and their corresponding realized measures. This allows for a more accurate depiction of the dynamics of the conditional covariance matrix of daily asset returns. To deal with the curse-of-dimensionality, the ALasso method is applied to estimate a sparse coefficient matrix in the measurement equation, and a parsimonious parameterized common persistence specification is used in conjunction with diagonal structures in the BEKK equation. With a rotation technique we can keep a richer dynamic while fitting a diagonal BEKK on rotated returns. In an empirical application, we find that the ALasso method, imposing a sparse coefficient matrix in the measurement equation, leads to superior predictive performance, especially for relatively large portfolios. Estimating the model for portfolios of 10, 15, and 20 stocks, our model leads to statistically significant gains in the loss function evaluation. AIC-based shrinkage method strikes a balance between in-sample fit and model parsimony, which leads to gains in out-of-sample performance at short and moderate horizons. Shrinkage methods hold promise as larger and more complete models of asset dynamics continue to be developed.

Table 1.6: Gain of predictive covariance using ALasso

Prediction Step	Number of Stocks		
	10	15	20
2	-1.037	0.004	-1.835**
3	-0.334	-0.145	-0.917
4	0.628	0.763	0.048
5	1.392*	1.116	0.025
6	1.223	0.883	-0.049
7	1.036	0.678	-0.264
8	0.658	0.259	-0.791
9	0.590	-0.199	-1.132
10	0.073	-0.699	-1.396*
11	0.041	-1.248	-1.658*
12	-0.218	-1.935**	-1.978**
13	-0.716	-2.444***	-2.282**
14	-0.766	-2.876***	-2.440***
15	-1.045	-3.381***	-2.475***
16	-0.884	-3.668***	-2.548***
17	-1.354*	-4.075***	-2.672***
18	-1.525*	-4.369***	-2.790***
19	-2.008**	-4.686***	-3.037***
20	-1.713*	-4.752***	-3.228***
21	-1.371*	-4.765***	-3.430***
22	-1.191	-4.995***	-3.605***

Test of predictive ability: Loss(ALasso)-Loss(OLS). A negative test value indicates model using the ALasso method has better predictive ability. The critical values are: 2.4 (1%), 1.68 (5%), 1.3 (10%). * denotes significance at the 10% level, ** significance at the 5% level, and *** significance at the 1% level

Chapter 2

Realized Markov Regime-Switching BEKK Model

2.1 Introduction

In the last few decades, GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models, which were introduced by Engle (1982) and Bollerslev (1986) have been the most popular and successful models in characterizing financial volatility. However, GARCH models often give good in-sample fit, but poor out-of-sample forecasting performance. According to Andersen and Bollerslev (1998), one reason for this is we typically compare the model forecasts with a weak measure of ex-post volatility, such as squared returns. The solution is to use the "realized volatility" calculated with intra-daily data. The other common problem of GARCH models in forecasting is that the volatility forecasts are usually too smooth and too high across periods due to structure breaks. Hamilton and Susmel (1994), for example, found that, for their stock return data, a shock on a given week would produce non-negligible effects on the variance more than one year later. Lamoureux and Lastrapes (1990), among others, show that this persistence may originate from structural changes in the variance process.

Empirical work has often shown that a single standard GARCH model is not valid over an extended period. Ignoring these breaks by assuming constant parameters in econometric models typically leads

to forecasts that are far from realizations (see Stock and Watson, 1996). The accumulated evidence from empirical research suggests that the volatility of financial markets displays a type of persistence that cannot be appropriately explained by classical GARCH models. In particular, these models usually show high persistence in the conditional volatility which is not compatible with the poor out-of-sample forecasting results of these models. These findings indicate a potential source of misspecification in that the structural form of the conditional mean and variance is relatively inflexible and fixed throughout the entire sampling period. Therefore, models in which the parameters are allowed to change over time may be more appropriate for modeling volatility.

Since the seminal paper by Hamilton (1989), which characterized the U.S. business cycle by periodic shifts from recessions to expansions and vice versa, the use of Markov-switching models has become increasingly popular in dynamic econometrics. The dynamics of the observed process is driven by a standard GARCH model as long as the persistence mechanism remains in the same regime and is powered by a different standard GARCH model whenever the working mechanism switches to another regime. Introducing change in the regime increases the model flexibility substantially and allows for interesting economic interpretations. See Kaufmann and Scheicher (1996) for a survey on Markov-switching models.

In the present work, we incorporate a GARCH model into a Markov regime-switching framework to allow for different volatility regimes. Within each regime, we use a GARCH model to govern the variance. The persistence of both regimes yields an extra source of volatility persistence compared with standard, single-regime GARCH, thereby enhancing the flexibility in describing the volatility persistence of shocks.

When the GARCH process is subject to regime switching, however, the basic form of the model is intractable due to a path-dependency problem (see Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996). Path-dependency occurs when the conditional variance at time t depends on the entire information set and path of the regime switching process up to time t due to the recursive nature of the GARCH process and state-dependent GARCH coefficients. Gray (1996) solves the path-dependency problem by collapsing the conditional variances in each regime into a single variance at each point in time. Klaassen

(2002) also integrates out the unobserved regimes, using all available information up to time t, whereas Gray (1996) only integrates conditional on information up to $t-1$. By doing so, the model becomes path-independent and the likelihood is tractable even with a large sample size.

Our realized Markov regime-switching GARCH model is as an extension of Klaassen's (2002) univariate generalized Markov regime-switching model to the bivariate case (except the forecasting mechanism) and potentially multiple-variate case. To improve not only the out-of-sample forecasting performance but also the accuracy of forecast evaluation, we use the realized measure as the ex-post volatility. In our model, a latent state variable governing the regime shifts follows a first-order, two-state Markov process, the parameters of which are estimated via quasi-maximum likelihood along with other unknown parameters, including transition probability and prediction probability of regimes.

2.2 The Markov Regime-Switching GARCH Model

The basic set up of the univariate Markov Regime-Switching GARCH Model is shown in this section. Let $\mathbf{p}_t$ be the price of a financial asset at the end of day t, then define $r_t = log(\mathbf{p}_t) - log(\mathbf{p}_{t-1})$ as the daily return. Let s_t be the unobserved variance regime at time t, which represents different state of volatility dynamics. Note that the time subscript t of the expectation operator E_t and probability p_t means the arguments are conditional on some information sets, e.g. $p_t(x) = p(x|\mathcal{F}_t)$, in which $\mathcal{F}_t$ represents the observable information set till period t, and we use the notation $p_{t-1}(s_t|\tilde{s}_{t-1}) = p(s_t|\mathcal{F}_{t-1}, \tilde{s}_{t-1})$ to denote the ex-ante probability of some series going to regime s_t at a specific time t conditional on the information set of the data-generating process, which consists of two parts. The first part, $\mathcal{F}_{t-1}$, denotes the information that can be observed, which is $(r_{t-1}, r_{t-2}, \cdots, r_1)$ in our case. The second part, $\tilde{s}_{t-1}$, is the regime path $(s_{t-1}, s_{t-2}, \cdots, s_1)$, which can not be observed. The dynamic of the conditional variance of return, $h_{\tilde{s}_t,t} = var_{t-1}(r_t|\tilde{s}_t)$, is our research interest, where the subscription t, s_t of h means the conditional variance predicted for time t based on the information set including the regime path until $t-1$. Without loss of generality, we assume that s_t follows a two-regime first-order

Markov process with constant transition probabilities independent with time stamp t defined as:

$$p_{ij} = p(s_t = j | s_{t-1} = i) \tag{2.1}$$

with $p_{t,i1} + p_{t,i2} = 1$ for $i, j = 1, 2$. Equation (2.1) indicates the probability of switching from state i at time $t-1$ to state j at t. The transition probability matrix is a 2 by 2 matrix composed with these probabilities

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix} \tag{2.2}$$

The unconditional probability of staying in state $s_t = 1$ is equal to $\pi_1 = (1 - p_{11})/(2 - p_{11} - p_{22})$. We choose to model the mean of the daily return as zero since the focus of this paper is the volatility rather than the mean. If the mean of daily return is not zero, it would also typically be modeled as a Markov regime-switching quantity, but not bring much change to our structure. Therefore, under the full regime path-dependent specification, the return equation is modeled as

$$\begin{aligned} r_t &= h_{\tilde{s}_t,t}^{1/2} z_t \\ z_t &\overset{i.i.d}{\sim} (0, 1) \end{aligned} \tag{2.3}$$

where $h_{t,\tilde{s}_t} = var_t(r_t | I_{t-1}, \tilde{s}_t)$ denotes the time-varying state-dependent variance given the whole regime path. The Markov regime-switching conditional covariance $h_{\tilde{s}_t,t}$ follows a GARCH(1,1) process:

$$h_{\tilde{s}_t,t} = \omega_{s_t} + b_{s_t} h_{t-1,\tilde{s}_{t-1}} + a_{s_t} r_{t-1}^2.$$

2.3 The Model Development

We generalize the previous univariate specification to the multivariate case in this section. By applying the realized BEKK model in Chapter 1 for such a Markov regime-switching context, we would obtain a

model similar to the one described above except all the variables would become bivariate. Specifically, let the $K \times 1$ vector $\mathbf{p}_t$ be the price of two assets at the end of day t, then define $r_t = log(\mathbf{p}_t) - log(\mathbf{p}_{t-1})$ as the daily return. The time-varying state-dependent covariance matrix $H_{t,\tilde{s}_t}$ with subscript $\tilde{s}$ implying it depends on the whole regime path. Therefore, $H_{t,\tilde{s}_t} = cov_{t-1}(r_t | r_1, r_2, ..., r_{t-1}, s_1, s_2, ..., s_t)$. Since H_t is positive semidefinite, we define $H_t^{1/2}$ as the symmetric square root of H_t using the Cholesky decomposition. The Markov regime-switching conditional covariance $H_{t,\tilde{s}_t}$ follows a BEKK(1,1) process:

$$H_{t,\tilde{s}_t} = CC' + B_{s_t} H_{t-1,\tilde{s}_{t-1}} B'_{s_t} + A_{s_t} X_{t-1} A'_{s_t} \tag{2.4}$$

where X_{t-1} denotes the realized measure of volatility (see Chapter 1). As discussed in the introduction, the Markov regime-switching BEKK model (MRS-BEKK) with a state-dependent past conditional variance is practically intractable. The conditional covariance matrix $H_{t,\tilde{s}}$ would in fact depend not only on the observable realized measure X_{t-1} and the current regime s_t, but also on all past states $\tilde{s}_{t-1}$. For a sample of length T, the evaluation of the likelihood function of the model with N regimes would require the integration over all the N^T possible paths, which renders the model impossible to estimate unless T is uselessly small. The intercept is set to be regime independent since we use the covariance target in the estimation.

Cai (1994) and Hamilton and Susmel (1994) were the first to solve this difficulty by dropping the lagged variance that causes the path dependence using only ARCH terms in (2.4) and several lags. Since $H_{t,\tilde{s}}$ in such a dynamic system construction only depends on the current regime s_t, there is no path-dependence problem anymore. Another way to avoid the path-dependence problem in the BEKK model is proposed by Gray (1996), who suggests keeping the lagged variance and integrating out the regime path $\tilde{s}_{t-1}$ in the BEKK term in (2.4) by using a weighted sum of the past variances over the probability of possible regimes period by period. As a result, the lagged conditional covariance $H_{t-1,s_{t-1}}$ in (2.6)

depends only on regime s_{t-1} instead of the path $\tilde{s}_{t-1}$. Following his method, we redefine the daily return

$$\begin{aligned} r_t &= H_{t,s_t}^{1/2} z_t \\ z_t &\overset{i.i.d}{\sim} (0, I_K) \end{aligned} \tag{2.5}$$

with subscript s_t to indicate the variable is solely conditional on current regime. The conditional covariance matrix H_{t,s_t} follows

$$\begin{aligned} H_{t,s_t} &= CC' + B_{s_t}\, E_{t-2}(H_{t-1,s_{t-1}})\, B_{s_t}' + A_{s_t}\, X_{t-1}\, A_{s_t}', \\ E_{t-2}(H_{t-1,s_{t-1}}) &= \sum_{s_{t-1}=1}^{2} p_{t-2}(s_{t-1}) \cdot \left(H_{t-1,s_{t-1}}\right), \end{aligned} \tag{2.6}$$

where

$$p_{t-2}(s_{t-1}) = \sum_{s_{t-2}=1}^{2} p_{t-2}(s_{t-2}) \cdot p_{t-2}(s_{t-1}|s_{t-2}) \tag{2.7}$$

is the ex-ante probability of the regime in period $t-1$ based on information in period $t-2$ and $p_{t-2}(s_{t-2})$ is the so-called "filtered probability" of volatility, which represents the probability of the volatility process being in the different regime conditional on information set at the same period. The filtered probability is a key term, from which we could derive all other probabilities in this paper. The filtered probability is computed as follows:

$$\begin{aligned} p_t(s_t) &= p_{t-1}(s_t|r_t) \\ &= \frac{f_{t-1}(r_t|s_t) \cdot p_{t-1}(s_t)}{f_t(r_t)} \\ &= \frac{f_{t-1}(r_t|s_t) \cdot p_{t-1}(s_t)}{\sum_{s_t=1}^{2} f_t(r_t|s_t) \cdot p_{t-1}(s_t)} \end{aligned} \tag{2.8}$$

Note the ex-ante probability is the product of filtered probabilities $p_{t-2}(s_{t-2})$ and the constant transition probability p_{ij}. The term $f_t(r_t|s_t)$ is the density function of the return in regime s in period t, which will

be defined below. Dueker (1997) uses a collapsing procedure in the spirit of Kim's (1994) algorithm to overcome the path-dependence problem, which is essentially as the same with Gray (1996).

Klaassen (2002) suggest an improved RS-GARCH model by calculating the expectation of lagged conditional variance with a broader information set. Klaassen (2002) has two differences compared with Gray (1996): first, the lagged term in (2.6) is calculated as conditional on $\mathcal{F}_{t-1}$ instead of $\mathcal{F}_{t-2}$, which allows us to use more observable information about previous regimes; second, Klaassen includes s_t in calculating the expected conditional covariance in period $t-1$. As a result, we replace $E_{t-2}(H_{t-1,s_{t-1}})$ with $E_{t-1}(H_{t-1,s_{t-1}}|s_t)$ in (2.6); this modification will include a lot of information, particularly if the regimes are highly persistent. For our MRS-BEKK(1,1) model, the specification using broader information would be:

$$H_{t,s_t} = CC' + B_{s_t}\, E_{t-1}(H_{t-1,s_{t-1}}|s_t)\, B'_{s_t} + A_{s_t}\, X_{t-1}\, A'_{s_t}, \tag{2.9}$$

where the expectation of conditional covariance in period $t-1$ is integrated across the regime s_{t-1} conditional on $\mathcal{F}_{t-1}$ and the current regime s_t:

$$E_{t-1}(H_{t-1,s_{t-1}}|s_t) = \sum_{s_{t-1}=1}^{2} p_{t-1}(s_{t-1}|s_t)(H_{t-1,s_{t-1}}), \tag{2.10}$$

where the probabilities are calculated in a Bayesian fashion:

$$p_{t-1}(s_{t-1} = j|s_t = i) = \frac{p_{ij}p_{t-1}(s_{t-1} = j)}{p_{t-1}(s_t = i)}. \tag{2.11}$$

Therefore, there is no problem of path dependence in the model.

2.3.1 Measurement Equation and Prediction of Conditional Covariance

The difference between our model and Klaassen (2002) is that we use the realized measure X_{t-1} in the model and forecasting. In the realized BEKK framework, we can easily make one-period-ahead predictions with the observed realized measure X_t. However, when we want to predict the conditional covariance beyond one period, such as H_{t+2}, we need to forecast the realized measure of covariance

X_{t+1}, which we cannot observe. In our model, the realized measure X_{t+1} would be updated by the contemporaneous conditional covariance H_{t+1}. However, we only have $H_{t+1,s_{t+1}}$ with different regimes. Therefore, we need to collapse $H_{t+1,s_{t+1}}$ in period $t+1$ to obtain the expected conditional covariance H_{t+1}, which will be used to update X_{t+1} in a logarithm manner with a linear function. Similarly with Chapter 1, we define the measurement equation in a vectorized form as following,

$$\begin{aligned} x_t &= d + \Gamma\, h_t + u_t \\ u_t &\overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_u) \end{aligned} \tag{2.12}$$

where

$$\begin{aligned} x_t &= vech(\operatorname{logm} X_t) \\ h_t &= vech(\operatorname{logm} H_t). \end{aligned} \tag{2.13}$$

The vech operator stacks the lower triangular part including the main diagonal of a $K \times K$ symmetric matrix into a $K(K+1)/2 \times 1$ vector. Hence, x_t, d, h_t and u_t are all $K(K+1)/2 \times 1$ vectors, with d being the intercept and u_t the innovation vector. Moreover, Γ is a $K(K+1)/2 \times K(K+1)/2$ constant coefficient matrix. This matrix characterizes the contemporaneous link between the realized measure and all the elements of the conditional covariance matrix. This vectorized form is easy to estimate since its structure is similar to the seemingly unrelated regression. The H_t in (2.13) is taken as

$$H_t = E_{t-1}(H_{t,s_t}) = \sum_{s_t=1}^{2} p_{t-1}(s_t) \cdot H_{t,s_t}. \tag{2.14}$$

The logm is defined in Chapter 1 as the matrix logarithm. The $p_{t-1}(s_t)$ is the predicted probability of regime s_t, which is calculated in (2.7). We refer to (2.9) as the realized BEKK equation. Since the realized measure X_t could be interpreted as a measurement of conditional covariance H_t, equation (2.12) is referred to as the measurement equation, tying the realized measure X_t to the expected conditional covariance H_t. Thus, the realized BEKK model fully specifies the dynamic system of daily returns and

the realized measure.

2.3.2 Covariance Targeting Under Rotation

The covariance targeting parameterization was introduced by Mezrich and Engle (1996) for the univariate GARCH model. Covariance targeting in large models is useful as it allows for two-step estimation, where in the first step, we estimate parameters representing unconditional moments with empirical moments. It substantially reduces the number of parameters to be estimated through nonlinear optimization. Rotated ARCH models were introduced by Noureldin, Shephard, and Sheppard (2014). They compared scalar and diagonal rotated BEKK and DCC models for 10 and 30 dimension cases. We cannot directly use this method in equation (2.9) due to the complexity of Markov regime-switching model in computing the unconditional expectation. However, we can still build a model imitating the idea of scaling to reduce the parameters. Let us denote the unconditional variance and covariance of daily returns as $var[r_t] = E[H_{t,s_t}] = \bar{H}$ and $E(E(H_{t-1,s_{t-1}}|s_t)) = \bar{H}$, the conditional covariance H_{t,s_t} can be expressed by covariance-targeting as following,

$$H_{t,s_t} = (\bar{H} - B_{s_t}\bar{H}B_{s_t} - A_{s_t}\bar{X}A_{s_t}) + B_{s_t}\, E_{t-1}(H_{t-1,s_{t-1}}|s_t)\, B_{s_t} + A_{s_t}\, X_{t-1}\, A_{s_t}, \tag{2.15}$$

where the unconditional mean of the realized measure is denoted as $E[X_t] = \bar{X}$. We assume that $(\bar{H} - B_{s_t}\bar{H}B_{s_t} - A_{s_t}\bar{X}A_{s_t})$ is positive semidefinite.

From (2.15), it is difficult to specify sufficient parameter restrictions to ensure that $(\bar{H} - B_{s_t}\bar{H}B_{s_t} - A_{s_t}\bar{X}A_{s_t})$ is positive semidefinite. This feature has restrained us from estimating more flexible dynamics of this model rather than just assuming that A and B are scalars. However, we can fit the model to rotated variables to extend the idea of covariance targeting in multivariate models of any dimension.

Using the unconditional variance of daily returns, $var(r_t) \equiv \bar{H}$, we define the rotated return e_t as

$$e_t = \bar{H}^{-1/2} r_t. \tag{2.16}$$

The conditional variance of the rotated returns is $Var[e_t|I_{t-1}, s_t] \equiv \tilde{H}_{t,s_t}$, noting that $H_{t,s_t} = \bar{H}^{1/2}\tilde{H}_{t,s_t}\bar{H}^{1/2}$

and $E(H_{t-1,s_{t-1}}|s_t, I_{t-1}) = \bar{H}^{1/2} E(\tilde{H}_{t-1,s_{t-1}}|s_t, I_{t-1}) \bar{H}^{1/2}$. Similarly, for the realized measure of covariance, we have $X_t = \bar{X}^{1/2} \tilde{X}_t \bar{X}^{1/2}$, where $\tilde{X}_t$ is the rotated realized measure of variance of the returns. Therefore, we obtain $E(\tilde{H}_{t-1,s_{t-1}}|s_t, I_{t-1}) = \bar{H}^{-1/2} E(H_{t-1,s_{t-1}}|s_t, I_{t-1}) \bar{H}^{-1/2}$, $\tilde{X}_{t-1} = \bar{X}^{-1/2} X_{t-1} \bar{X}^{-1/2}$. Basically, we rotate the variables with their unconditional variance to make their unconditional expectation equal to an identity matrix. A variance-targeting realized MRS-BEKK model is:

$$\tilde{H}_{t,s_t} = (I_K - \tilde{B}_{s_t} \tilde{B}'_{s_t} - \tilde{A}_{s_t} \tilde{A}'_{s_t}) + \tilde{B}_{s_t} \, E_{t-1}(\tilde{H}_{t-1,s_{t-1}}|s_t) \, \tilde{B}_{s_t} + \tilde{A}_{s_t} \, \tilde{X}_{t-1} \, \tilde{A}_{s_t} \tag{2.17}$$

$$E[\tilde{H}_{t,s_t}] = E[E_{t-1}(\tilde{H}_{t-1,s_{t-1}}|s_t)] = E[\tilde{X}_t] = I_K$$

This model has 2K dynamic parameters when $\tilde{B}_{s_t}$ and $\tilde{A}_{s_t}$ are diagonal. It is easy to impose the positive semidefinite of $(I_K - \tilde{B}_{s_t} \tilde{B}'_{s_t} - \tilde{A}_{s_t} \tilde{A}'_{s_t})$ for this specification.

For unrotated returns, the diagonal BEKK model defined in equation (2.17) for rotated returns implies a non-diagonal BEKK model:

$$H_{t,s_t} = \bar{H}^{1/2} \tilde{H}_{t,s_t} \bar{H}^{1/2} = CC' + B_{s_t} \, E_{t-1}(H_{t-1,s_{t-1}}|s_t) \, B'_{s_t} + A_{s_t} \, X_{t-1} \, A'_{s_t}$$

where

$$B_{s_t} = \bar{H}^{1/2} \tilde{B}_{s_t} \bar{H}^{-1/2}, \quad A_{s_t} = \bar{H}^{1/2} \tilde{A}_{s_t} \bar{X}^{-1/2}, \quad CC' = \bar{H}^{1/2} (I_2 - \tilde{B}_{s_t} \tilde{B'}_{s_t} - \tilde{A}_{s_t} \tilde{A'}_{s_t}) \bar{H}^{1/2}, \tag{2.18}$$

It is clear the rotated BEKK implies a BEKK specification for the original variables in a constrained version since $\tilde{A}_{s_t}$ and $\tilde{B}_{s_t}$ depend on the spectral decomposition of $\bar{H}$ and $\bar{X}$.

Bauwens, Laurent, and Rombouts (2006) discuss the invariance of multivariate models to the linear transformation of returns. In their definition, invariance means the transformed model is in the same model class and maintain the same dynamic specification (e.g. scalar, diagonal, or full parameterization). In the case of diagonal specification, our model is clearly not invariant to linear transformation. Supposing $\tilde{A}_{s_t}$ is diagonal, A_{s_t} derived from (2.18) is a full asymmetric matrix, which means we obtain

a full parameterized BEKK model for the unrotated variables. For the rotated model given in (2.16) and (2.17), similar to Chapter 1, we could impose the Assumption 3 to ensure the long-run covariance stationarity and the positive definiteness of $(I_K - B_{s_t}B'_{s_t} - A_{s_t}A'_{s_t})$. For any $K \times K$ matrix M with eigenvalues $\lambda_1, \lambda_2 ... \lambda_K$, define $\rho(M) := \max(|\lambda_1|, |\lambda_2|, ..., |\lambda_K|)$.

Assumption 3 *In the realized rotated MRS-BEKK model given by (2.16) and (2.17),* $\rho(\tilde{A}_{s_t} \otimes \tilde{A}_{s_t} + \tilde{B}_{s_t} \otimes \tilde{B}_{s_t}) \leq 1$.

This assumption is easy to impose in the estimation.

2.4 Estimation and Inference

2.4.1 Estimation for the BEKK Equation

In this section we discuss the asymptotic properties of the quasi-maximum likelihood estimator within the MRS-BEKK model. We could estimate the parameters in the rotated model, and then reverse the linear transformation to retract the unrotated measure.

The realized BEKK equation (2.9) and (2.17) can be parameterized with a finite-dimensional $(\delta \times 1)$ parameter vector $\theta \in \Theta \subset \mathbb{R}^\delta$. Decompose $\theta = (\bar{h}', \bar{x}', \theta'_d)'$, where $\bar{h}$ and $\bar{x}$ denote the model's static parameters, the unconditional moment, we can estimate by sample counterparts. Let θ_d denote the vector of the dynamic parameters indexing $\tilde{A}$ and $\tilde{B}$ in (2.17) and the transition probabilities p, q in (2.2), which can be estimated by maximum likelihood method conditional on estimate $\hat{\bar{x}}$ and $\hat{\bar{h}}$. We denote the true parameter vector with θ_0. Specifically, the static parameters are to be estimated by a moment estimator:

$$\hat{\bar{h}}' = vech(\hat{\bar{H}}), \qquad \hat{\bar{x}}' = vech(\bar{X}),$$

where

$$\hat{\bar{H}} = T^{-1} \sum_{t=1}^{T} r_t r'_t, \qquad \hat{\bar{X}} = T^{-1} \sum_{t=1}^{T} X_t.$$

These estimates are then decomposed into $\hat{\bar{H}}^{1/2}$ and $\hat{\bar{X}}^{1/2}$. Subsequently, we construct the time series of rotated returns and realized measure $\hat{e_t} = \hat{\bar{H}}^{-1/2} r_t, \hat{\bar{X}}_t = \hat{\bar{X}}^{-1/2} X_t \hat{\bar{X}}^{-1/2} t = 1, 2, ..., T$. Then θ_d will be estimated by QML estimation in the second step,conditional on $\hat{\bar{x}}$ and $\hat{\bar{h}}$.

Let $L_t(\theta_d; \theta_s | s_t)$ be the contribution of observation r_t to the likelihood function $L_T(\theta_d; \theta_s)$. For $r_t \overset{i.i.d}{\sim} \mathcal{N}(0, H_{t,s_t})$, the log-likelihood function constructed from observation r_t is defined as:

$$l_T(\theta_d; \theta_s | s_t) = \sum_{t=1}^{T} \log L_t(\theta_d; \theta_s | s_t) = \sum_{t=1}^{T} (constant - \frac{1}{2}(\log |H_{t,s_t}| + r_t' H_{t,s_t}^{-1} r_t)) \tag{2.19}$$

According to (2.5), the multivariate distribution of r_t is not assumed normal. (2.19) is the quasi-likelihood function and the estimation is performed by QML, which provides consistent estimates regardless of the underlying distribution of observation r_t (see Bollerslev and Wooldridge, 1992). We assume that the initial value of $\tilde{H}_{t,s_t}$ in the first period is known and positive semidefinite. For our path-independent model, we maximize the sum of log-likelihood function with expectation in each period with respect to the distribution of the regimes. Let

$$l_T(\theta_d; \hat{\theta}_s) = \sum_{t=1}^{T} \log E_{t-1}(L_t(\theta_d; \hat{\theta}_s | s_t)) = \sum_{t=1}^{T} \log \sum_{s_t=1}^{2} L_t(\theta_d; \hat{\theta}_s | s_t) * p_{t-1}(s_t) \tag{2.20}$$

where the ex-ante probability $p_{t-1}(s_t)$ is given in equation (2.7)). The QML estimator is $\hat{\theta}_d$, where $\hat{\theta}_d = arg\, max_{\theta_d \in \Theta} l_T(\theta_d; \hat{\theta}_s)$. The asymptotic theory for QML estimation of the GARCH model is discussed by Comte and Lieberman (2003) and Hafner and Preminger (2009). For the closely related covariance-targeting GARCH model, Pedersen and Rahbek (2014) prove its consistency and asymptotic normality. We note that the Rotated BEKK model applies the same parameterizations but to rotated returns.

Since we used two-step maximum likelihood estimation, if the usual assumptions for the validity of QMLE are satisfied, the estimator is consistent and the asymptotic distribution:

$$\sqrt{T} \left(\begin{bmatrix} \hat{\theta_s} \\ \hat{\theta_d} \end{bmatrix} - \begin{bmatrix} \theta_s \\ \theta_d \end{bmatrix} \right) \to \mathbb{N}(0, V),$$

with

$$V = \begin{bmatrix} G_{\theta_s}^{-1} & -G_{\theta_s}^{-1} G_{\theta_d} M^{-1} \\ 0 & M^{-1} \end{bmatrix} E\left[\frac{\partial l_T}{\partial \theta}\frac{\partial l_T}{\partial \theta'}\right] \begin{bmatrix} G_{\theta_s}^{-1} & -G_{\theta_s}^{-1} G_{\theta_d} M^{-1} \\ 0 & M^{-1} \end{bmatrix},$$

where

$$G_{\theta_s} = E\left[\frac{\partial g(\theta_s, \theta_d)}{\partial \theta_s'}\right]; G_{\theta_d} = E\left[\frac{\partial g(\theta_s, \theta_d)}{\partial \theta_d'}\right]; M = E\left[\frac{\partial m(\theta_d)}{\partial \theta_d'}\right]$$
$$g(\theta_s, \theta_d) = \left[\frac{\partial l_T(\theta_s, \theta_d)}{\partial \theta_s}\right]; m(\theta_d) = \left[\frac{\partial l_T(\theta_s, \theta_d)}{\partial \theta_d}\right].$$

The proof is in Newey and McFadden (1994).

2.4.2 Estimation of the Measurement Equation

Since the measurement equation in our model is just a linear model, we will use linear regression to estimate the parameters in the measurement equation. As mentioned before, the measurement equation is actually a system of equations estimated in a similar manner as using SUR method. To be concrete, we express (2.12) as the system of equations. For example, fo the first element of x_t we have:

$$\begin{bmatrix} x_{11,1} \\ x_{11,2} \\ \vdots \\ x_{11,T} \end{bmatrix} = \begin{bmatrix} d_{11} \\ d_{11} \\ \vdots \\ d_{11} \end{bmatrix} + \begin{bmatrix} h_{11,1} & h_{12,1} & h_{13,1} \\ h_{11,2} & h_{12,2} & h_{13,2} \\ \vdots & \vdots & \vdots \\ h_{11,T} & h_{12,T} & h_{13,T} \end{bmatrix} \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix} + \begin{bmatrix} u_{11,1} \\ u_{11,2} \\ \vdots \\ u_{11,T} \end{bmatrix} \tag{2.21}$$

which is just the stacked first rows of (2.12) along the time dimension. We can estimate the first row of coefficient matrix H in (2.12) with linear regression of the equation (2.21). Other rows of H can be estimated in similar way.

2.4.3 Smoothed Regime Inference

Although the regimes are not observed, one can estimate the probability that the process is in a particular regime in each period. This is useful if one wants to check how much time the series spends in each state.

Following Kim's (1994) smoothing algorithm, we compare the smoothed regime probabilities $p_T(s_t) = p(s_t|I_T)$, which are the probabilities of a series being in one specific regime conditional on the complete observations up till time T. Hence, it gives the most informative answer to the question of which regime the process was in at time t. The smoothed probabilities can be calculated recursively. According to the assumption of Markov properties, the hidden regime s_t depends on past observation r_{t-1} only through the value of s_{t-1}. Similarly, when we estimate the past regimes from present observations, the s_t depends on observations r_{t+1} only through the value of s_{t+1} :

$$p_T(s_t|s_{t+1}) = p_t(s_t|s_{t+1}) \tag{2.22}$$

Note that

$$\begin{aligned} p_t(s_t = i|s_{t+1} = j) &= \frac{p_t(s_t = i, s_{t+1} = j)}{p_t(s_{t+1} = j)} \\ &= \frac{p_t(s_t = i) * p_t(s_{t+1} = j|s_t = i)}{p_t(s_{t+1} = j)} \\ &= \frac{p_{ij} * p_t(s_t = i)}{p_t(s_{t+1} = j)} \\ i, j &= 2 \end{aligned} \tag{2.23}$$

where p_{ij} is just the constant transition probability in (2.1). It is therefore the case that

$$\begin{aligned} p_T(s_t = i, s_{t+1} = j) &= p_T(s_{t+1} = j) * p_T(s_t = i|s_{t+1} = j) \\ &= p_T(s_{t+1} = j) * p_t(s_t = i|s_{t+1} = j) \\ &= p_T(s_{t+1} = j) * \frac{p_{ij} * p_t(s_t = i)}{p_t(s_{t+1} = j)} \\ i, j &= 2 \end{aligned} \tag{2.24}$$

The second equality follows from (2.22) and the third equation follows from (2.23). The smoothed inference for date t is the sum of (2.24) over the value of s_{t+1}:

$$
\begin{aligned}
p_T(s_t = i) &= \sum_{j=1}^{2} p_T(s_t = i, s_{t+1} = j) \\
&= \sum_{j}^{2} p_T(s_{t+1} = j) * \frac{p_{ij} * p_t(s_t = i)}{p_t(s_{t+1} = j)} \\
&= p_t(s_t = i) \sum_{j=1}^{2} \frac{p_{ij} * p_T(s_{t+1} = j)}{p_t(s_{t+1} = j)} \\
&= p_t(s_t = i) * [p_{i1}, p_{i2}] * \begin{pmatrix} p_T(s_{t+1} = 1)/p_t(s_{t+1} = 1) \\ p_T(s_{t+1} = 2)/p_t(s_{t+1} = 2) \end{pmatrix} \\
&= p_t(s_t = i) * P_i * (P_T(s_{t+1} = j)(\div)P_t(s_{t+1} = j)), \\
& i, j = 2
\end{aligned}
\tag{2.25}
$$

where 1-by-2 vector P_i denotes the ith row of the transition matrix P and the sign $(\div)$ indicates the element-by-element division. Collecting the equations (2.25) for $i = 1, 2$ in a 2-by-1 vector, we can express the smoothed probability in as

$$
\begin{aligned}
&P_T(s_t) = P_t(s_t) \odot \{P_i * (P_T(s_{t+1})(\div)P_t(s_{t+1}))\}, \\
&i, j = 2
\end{aligned}
\tag{2.26}
$$

We can iterate the equation (2.26) backward from $t = T - 1$ to $t = 1$. This iteration is started with $P_T(s_T)$, which is just the filtered probability obtained from (2.8). For more details of Kim's algorithm, please refer to Kim (1993) and Hamilton and Susmel (1994).

2.5 Empirical illustration

So far, we have generalized the single-regime BEKK to MRS-BEKK to obtain more flexibility in modeling the volatility persistence. In this section we estimate both models and examine whether this gen-

eralization pays off empirically in terms of improved volatility forecasts, the central issue of the paper.

2.5.1 Data Description

The model is applied to high-frequency prices for two liquid stocks from the technology sector, IBM and INTC. The sampling period is from January 3, 2006, to April 30, 2012, and the source of data is the TAQ database. We focus on the noise-robust realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) as our choice for X_t. We compute daily returns and the realized kernel from EST 9:30-16:00, which are the weekday hours of the stock market. To eliminate the overnight effect, the prices recorded during the 15 minutes at the beginning and end of opening hours are excluded.

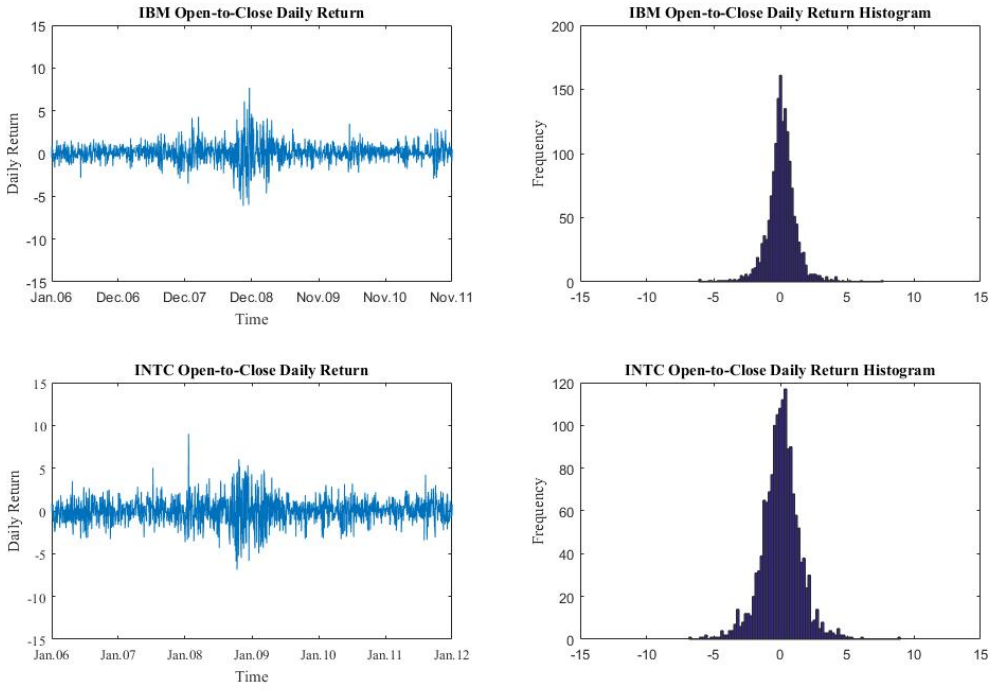

Figure 2.1: The open-to-close log daily return of IBM and INTC.

The summary statistics for the daily log open-to-close returns are presented in the left part of Table 2.1. As expected, the average daily returns are around 0%. These two stocks from the technology sector

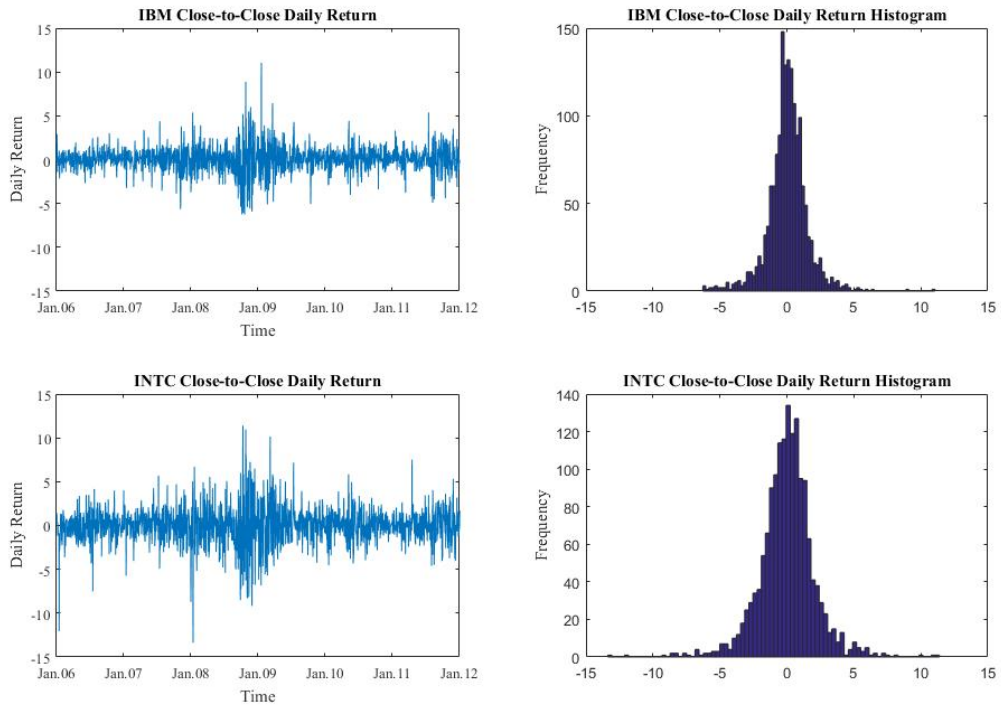

Figure 2.2: The close-to-close log daily return of IBM and INTC.

appears less volatile than some we have seen from the financial sector. The ranges of both returns are within 20%. Additionally, the summary statistics for the daily log close-to-close returns are presented in the second part of Table 2.1. Due to the overnight effect, the close-to-close log daily return is more volatile than the open-to-close log daily return. For example, one of the biggest drops of daily return of INTC in April 2006 is not showing in the open-to-close return graph. The ranges of both return also increases from under 20% to the interval between 20% and 25%. All the histograms of open-to-close (Figure 2.1) and close-to-close (Figure 2.2) daily returns appear slightly right-skewed.

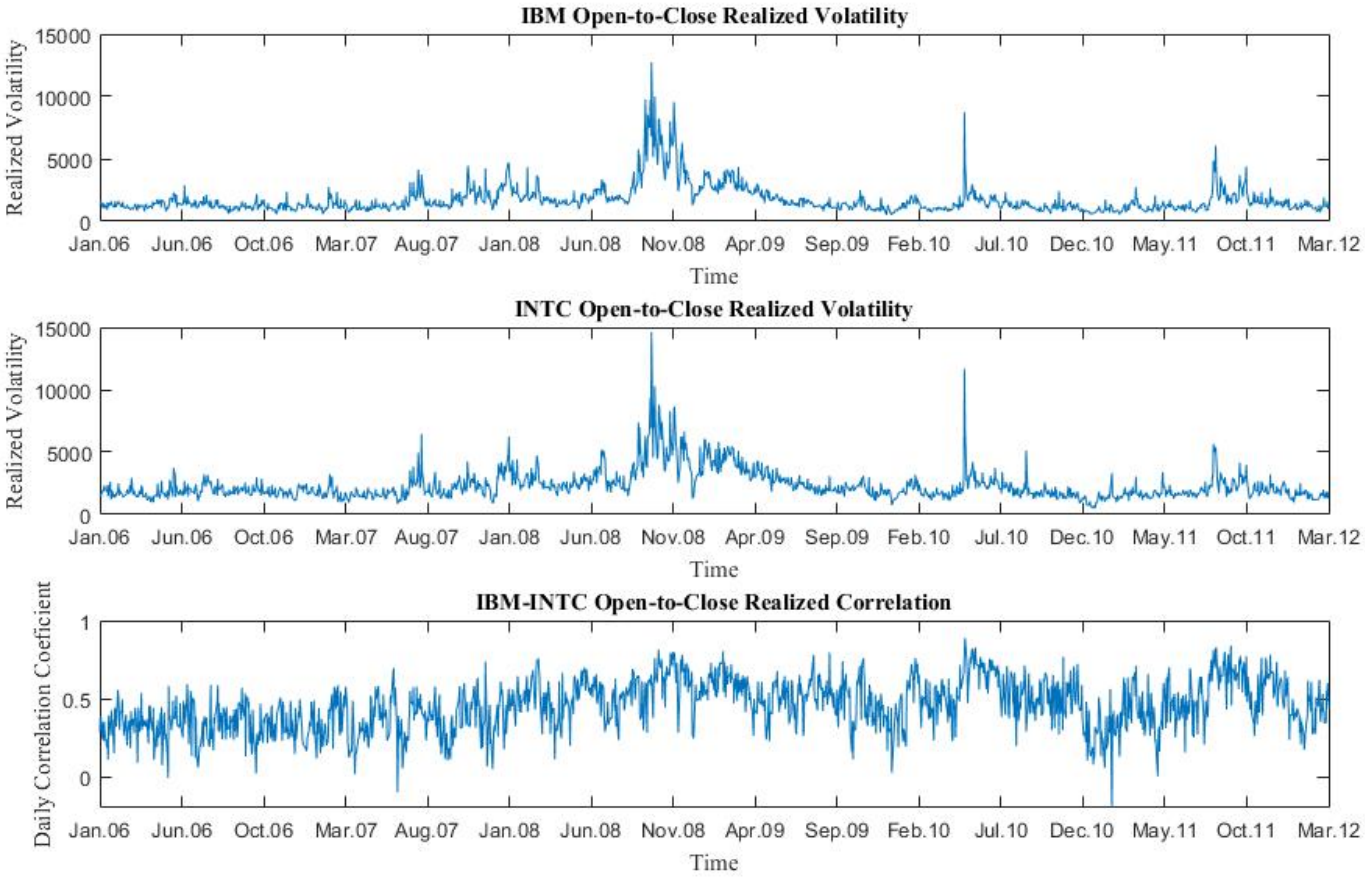

Figure 2.3: The annualized realized volatility and co-volatility of IBM and INTC.

We plot the annualized realized volatility and co-volatility of these two stocks in Figure 2.3. The sharp increase in realized volatility from 2008 to 2009 is associated with the major upheavals in the financial markets during the subprime mortgage crisis. The increase in volatility is much more pronounced especially after the collapse of Lehman Brothers in mid-September 2008. According to the visual inspection of the plot of annualized volatility, the financial crisis period approximately spanned from August 2008 to April 2009. We also record the summary statistics of annualized volatility in Table 2.2.

Table 2.1: Summary Statistics of Daily Return

Statistics	Open-to-Close Log Returns		Close-to-Close Log Returns	
	IBM	INTC	IBM	INTC
Mean	0.098	0.034	0.056	0.006
St.dev	1.116	1.454	1.498	2.088
Max	7.708	9.000	11.051	11.393
Min	-6.119	-6.857	-6.274	-13.366

The log return are in percentage point.

Table 2.2: Summary Statistics of Realized Volatility

Stocks	Open-to-Close Realized Volatility and correlation			
	Mean	St.dev	Max	Min
IBM	17.192	11.572	127.455	5.035
INTC	23.182	12.150	146.214	4.822
IBM-INTC	0.455	0.161	0.888	-0.208

The annualized realized volatility is the square root of 252 times the realized kernel, the last row is the realized correlation of IBM and INTC.

2.5.2 Estimation Results

This subsection presents the estimation results for the MRS-BEKK model. For comparison, we also estimate the single-regime BEKK model. Table 2.3 presents the QML estimation results. The smooth effect and the innovation effect correspond to B and A in equation (2.9). As expected, the MRS-BEKK model provides a better fit than the single-regime BEKK model. The gain in the log-likelihood value of the MRS-BEKK model over the regular BEKK model is 14. The standard LR test is not valid for nested model, we will test the significance of the second regime in a later section with a bootstrap method. In the single regime BEKK model and regime 1 of MRS-BEKK model, the smooth coefficients B have larger values than the innovation coefficients A. This implies that the conditional variance is smoother in regime 1. Interestingly, the result for regime 2 is characterized by more sensitivity to recent shocks and less persistence than in regime 1. The effect of individual shocks die out quickly during the periods of regime 2, but has a longer lasting effect during regime 1. As a result, regime 1 features a smooth regime while regime 2 characterize a responsive regime. An unexpected shock in the stock market induces an observable jump of volatility which is not persistent unless the volatility process stay in regime 1 in the following days. This result is consistent with the stylized fact called pressure relieving in the stock market; that is, some shocks are not persistent at all, but are followed by a tranquil period. Another finding is that the persistence of single-regime BEKK is larger than that of the second-regime of MRS-BEKK, which indicates the shock will be less persistent in a responsive regime.

The estimate of the constant Markov transition probability matrix is presented in Table 2.4. The result shows the value of staying probabilities p_{11} and p_{22} (values in the diagonal position) is larger than the switching probabilities p_{12} and p_{21} (values in the off-diagonal), which means that the regimes are persistent. Another finding is that $p_{11} > p_{22}$. Since the unconditional probability of being in regime, 1 $p_t(s_1)$ is $(1 - p_{22})/(2 - p_{11} - p_{22}) = 0.619$, we can conclude that regime 1 is more likely than regime 2. We define two sources of volatility persistence in regime-switching model. One source is the within-regime persistence which is indicated by the parameters in BEKK equation. The other source is the persistence of regimes. The estimates of the BEKK coefficient reported in Table 2.4 indicate that the variance is persistent in each regime.

Table 2.3: Estimation of coefficient

Coefficients in different model	Regime 1	Regime 2
	MRS-BEKK	
a_{11}	0.294(.066)	0.815(.002)
a_{12}	-0.028(0.013)	0.031(0.005)
a_{21}	0.07(0.021)	0.082(0.006)
a_{22}	0.139(0.093)	0.772(.056)
b_{11}	0.908(.043)	0.109(.007)
b_{12}	0.018(0.006)	0.08(0.022)
b_{21}	-0.024(0.012)	-0.108(0.039)
b_{22}	0.974(.028)	0.407(.232)
η_1	0.911	0.676
η_2	0.9968	0.761
Loglike	-4027	
	BEKK	
a_{11}	0.488(.031)	-
a_{12}	-0.018(0.005)	-
a_{21}	0.088(0.009)	-
a_{22}	0.33(.022)	-
b_{11}	0.802(.03)	-
b_{12}	0.034(0.005)	-
b_{21}	-0.046(0.009)	-
b_{22}	0.929(.022)	-
η_1	0.882	-
η_2	0.971	-
Loglike	-4041	

Standard errors in parentheses. "logLik" denotes the log-likelihood value. η_1, η_2 denotes the "persistence" of the model. The estimated models belong to the model described by equations (2.9), a_{mn},b_{mn} are element in mth row and nth column in A and B.

Table 2.4: Markov Transition Probability

Regime	$s_{t+1} = 1$	$s_{t+1} = 2$
$s_t = 1$	0.95(0.024)	0.05
$s_t = 2$	0.081	0.919(0.031)

2.6 Likelihood Ratio (LR) Test

In this section, we focus on the likelihood ratio test of the null of a single-regime model against a Markov regime-switching model. There are two problems that make testing the stability of coefficients particularly challenging. The first is known as the problem of nuisance parameters (the transition probabilities p_{ij}) that are unidentified under the null hypothesis, so that the asymptotic distribution of the likelihood ratio is not the usual χ^2-distribution (see Hansen et al., 1996). The second problem is that the information matrix is singular under the null hypothesis. This is due to the fact that underlying regimes are not observable.

A few papers propose tests for the Markov regime-switching models. Garcia (1998) studied the asymptotic distribution of a sup-type likelihood ratio test. Hansen(1992) treated the likelihood as an empirical process indexed by all the parameters (those identified and those unidentified under the null). His test relied on taking the supremum of LR over the nuisance parameters. Carrasco, Hu, and Ploberger (2014) proposed a likelihood test using parametric bootstrap which derives the null distribution of the test and a bootstrap method to compute its p-values. We use a similar strategy to derive the null distribution of LR test to approximate the α critical values c_α and the p-value of the estimated LR value. The procedure is described in algorithm (4).

Suppose our model has log-likelihood $l_T(\theta, \gamma) = \sum_{t=1}^{T} l_t(\theta, \gamma)$ with parameter vectors θ and γ where θ is fully identified, but γ is not identified under H_0. The hypothesis takes the form

H_0: The true model of the volatility process is the single-regime BEKK model

vs. H_1: The true model of the volatility process is the two-regime BEKK model

Define the log-likelihood ratio (LR) function:

$$LR_T(\theta) = 2 * (l_T(\theta_1) - l_T(\theta_0)) \tag{2.27}$$

where θ_0 is the value of θ under H_0 while θ_1 is the value of θ under H_1. Letting the number of simulated samples $S \to \infty$ will force the empirical critical values to converge to the critical values for the

1 Using the observed data, estimate θ by $\hat{\theta}_0$ and $\hat{\theta}_1$, the maximum likelihood estimator under H_0 and H_1. Compute the test statistic LR.
2 Given $\hat{\theta}_0$, generate S independent samples $y_1^s, ..., y_{T\,s=1,2,...,S}^s$ under H_0.
3 Using each simulated sample $y_1^s, ..., y_T^s$, estimate θ by $\hat{\theta}_0^s$ and $\hat{\theta}_1^s$ the MLE under both hypothesis and compute LR^s, $s = 1, 2, ..., S$.
4 Define $c_{\alpha,s}$ as the empirical critical values using LR^s, $s = 1, 2, ..., S$, which is the $(1-\alpha)$ quantile of LR^s. Moreover, the bootstrap p-value is given by $\frac{1}{S}\sum_{s=1}^{S} \mathbb{1}(LR^s > LR)$, which $\mathbb{1}(\cdot)$ is the indicator function.

Algorithm 4: The Parametric Bootstrap procedure for the LR test

distribution of the test statistic under H_0. By performing the algorithm (4) with $S = 100$, we obtained LR $= 28.967$ and the pvalue $= 0.02$, which means the null hypothesis is rejected at significance level above 2%.

2.6.1 Result of Regime Inference

Figure 2.4 shows the estimates of the probability of the variance process being in regime 2. We can see that the smooth regime (regime 1) dominates the process, which implies the volatility of the asset price is smooth most of the time. The second finding is that the process does not stay too long in the responsive regime but quickly switches back to the smooth regime. The regimes appear to switch frequently and randomly. By comparing this figure with the graph of realized volatility and daily return, we could determine that the volatility process is likely to stay in the responsive regime in early 2008 before the financial crisis hits the market, which means the stock market was very sensitive in that period. Then, a few major shocks hit the market, and the regime appears to switch back to a smooth regime for a short time. The persistently high level of volatility at that time may be the reason for this. However, the smooth regime does not last for long before switching back to the responsive regime, which indicates that the market was still not stable due to the aftershock. We also check the mean of the covariance matrix of each regime and find that the elements of the covariance matrix are larger in the responsive regime. This fact implies that the stock market is more volatile in the responsive regime, or the market experiences great shocks when in a responsive regime and then switches back to a smooth regime after the shock. Therefore, the scenario is that the market starts in a responsive regime and a shock hits the market,

inducing a volatile period; the market switches to the smooth regime, keeping a weakened volatility for a period before it switches back to the responsive regime again until the next shock.

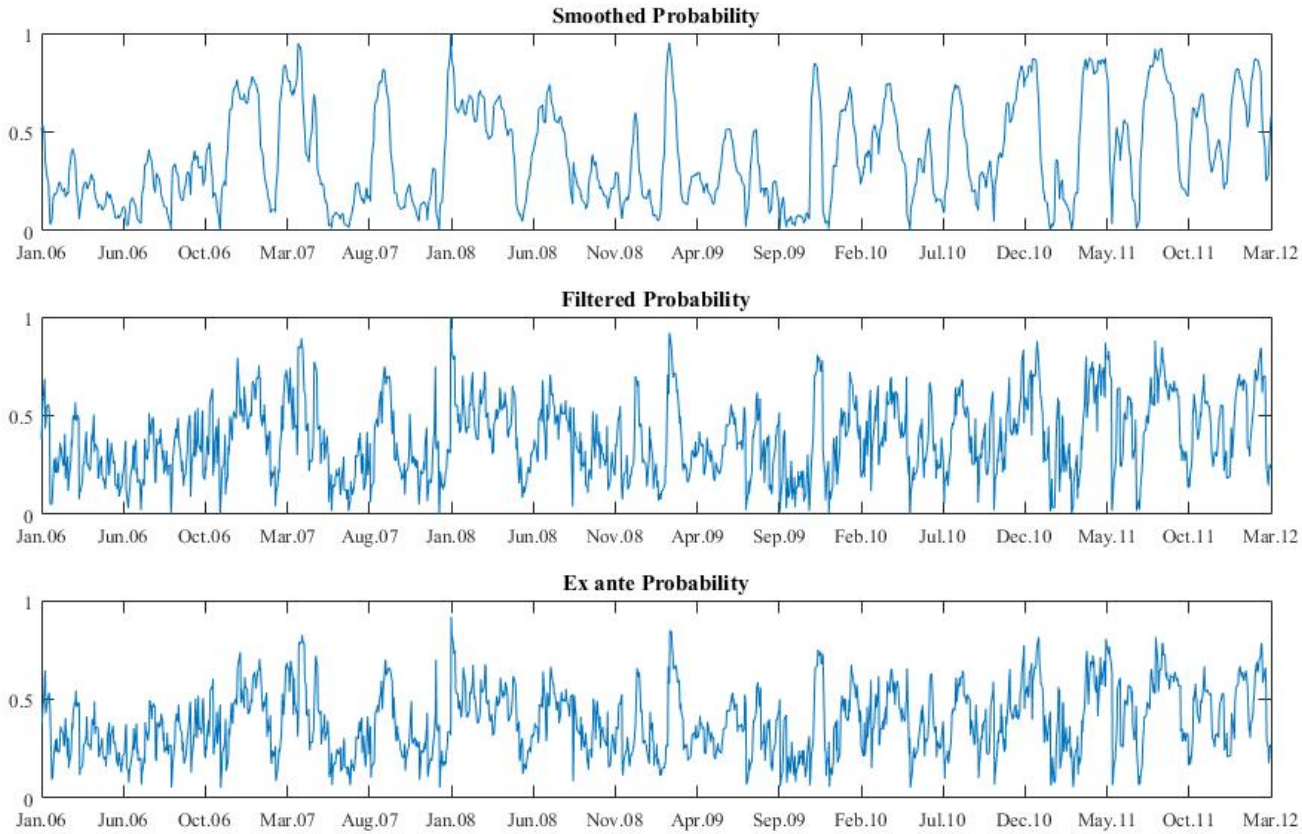

Figure 2.4: Probability of staying in responsive regime (regime 2)

2.6.2 Estimation Result of Measurement Equation

Following the structure of the measurement equations, each realized variance (covariance) is the function of the corresponding conditional variance (covariance) and the conditional covariance of the pairs of all the other assets. We wish to emphasize one empirical observation that concerns the difference between open-to-close returns and close-to-close returns. With open-to-close returns our estimates of the coefficient matrix of Γ in (2.12) for the MRS-BEKK models are reported in (2.28)

$$\hat{\Gamma} = \begin{pmatrix} 1.054(0.04) & -0.103(0.103) & 0.096(0.05) \\ -0.0081(0.014) & 1.082(0.033) & 0.048(0.017) \\ 0.1(0.04) & -0.053(0.102) & 1.057(0.05) \end{pmatrix} \tag{2.28}$$

where the numbers in parentheses are standard errors. Note that the empirical estimates of elements in the diagonal position are consistent with the belief that the realized kernel is roughly an unbiased measurement of (open-to-close) H_t. With close-to-close returns, we obtain the estimates of the coefficient matrix of Γ (2.12) for the MRS-BEKK(1,1) models in (2.29):

$$\hat{\Gamma} = \begin{pmatrix} 0.975(0.035) & 0.0737(0.087) & 0.206(0.04) \\ 0.007(0.012) & 0.837(0.03) & 0.05(0.014) \\ 0.237(0.034) & 0.138(0.084) & 0.84(0.039) \end{pmatrix} \tag{2.29}$$

It is not surprising that the estimate of elements in the diagonal position is less than one in this case since realized measure only covers partial period (open-to-close) of the close-to-close daily return. The point estimate suggests that the volatility during the open period amounts to about 84% of the daily volatility for Microsoft. However, this effect is not as obvious for IBM; one possible reason is that the overnight effect is not significant for IBM.

Similar to Chapter 1, to check the fit of the measurement equation we focus on the test of autocorrelation of the residuals. If there are autocorrelated residuals, we may need to add lags of dependent variables. The Lagrange multiplier (LM) test for autocorrelation developed by Breusch (1978) and Godfrey (1978) has become a standard tool in applied econometrics. It is often applied to the residuals of a

multivariate regression such VAR or SUR of time series model. A vectorized version of the LM test is proposed Doornik (1996) by testing the significance of lagged residuals $\hat{u}_{t-h}$ (obtained under the null hypothesis) in the auxiliary regression:

$$u_t = d + \Gamma\, h_t + b_1\hat{u}_{t-1} + ... + b_h\hat{u}_{t-h} + v_t \tag{2.30}$$
$$v_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_v)$$

We wish to test the significance of lagged residuals, so the null hypothesis is

$$H_0 : b_1 = ... = b_h = 0. \tag{2.31}$$

which implies no autocorrelation of residuals for lag one through lag h. The test statistic is constructed as:

$$LM = TnR_m^2 \tag{2.32}$$

with an asymptotic $\chi^2_{hn^2}$ under null. n is the dimension of the system, h is the number added lagged regressors in the each auxiliary regression and T is the sample size. R_m^2 is a R^2-type measures of goodness of fit:

$$R_m^2 = 1 - (1/n)tr(\hat{v}\hat{v}')(\hat{u}\hat{u}') \tag{2.33}$$

where v and u are $K(K+1)/2$-by-T matrix of residuals from the auxiliary regressions and measurement equations. We only test the lag up to 10, which is enough for the stock data. The LM value for the measurement equation with 1 regime and 2 regime switching models are 0.754 and 0.993, and the p-value are 0.999 and 0.521, from which we can't reject the null hypothesis.

2.7 Volatility Impulse Response Functions

We use Volatility Impulse Response Functions (VIRF) to check the persistence of different regimes of the MRS-BEKK model. Since they were introduced by Sims (1980), Impulse Response Functions have been used in VAR models to analyze the response of a variable to a shock. Every BEKK model possesses a unique equivalent vec specification (see Engle and Kroner, 1995); we investigate the VIRF for our MRS-BEKK model by transforming it to vec form. Still considering the BEKK(1,1) case,

$$vech(H_t) = c + B^* vech(H_{t-1}) + A^* vech(H_{t-1}^{1/2} \epsilon_{t-1} \epsilon_{t-1}' H_{t-1}^{1/2}) \tag{2.34}$$

with B^* and A^* as function of A and B. Denote the unconditional covariance matrix H, and

$$vech(H) = (I - A^* - B^*)^{-1} c. \tag{2.35}$$

At time t=0 some independent news is reflected by ϵ_0. Assuming the independence of the components of ϵ_t, one may assume a shock in one component and set all the other components to zero. The volatility state at time $t = 0$ is assumed to be the steady state $H_0 = H$. The conditional covariance matrix H_t is a function of the innovations $\epsilon_1, ..., \epsilon_{t-1}$, the initial shock ϵ_0, and H_0. VIRF is defined as the expectation of volatility conditional on an initial shock and volatility, i.e.

$$V_t(\epsilon_0) = vech(H_t | \epsilon_0, H_0 = H). \tag{2.36}$$

In (2.36),$V_t(\epsilon_0)$ is a $K(K+2)/2$ dimensional vector, where K is the number of assets. For our case, $K = 2$, then the first and third elements of $V_t(\epsilon_0)$ represent the VIRF of the conditional variances of the first and second assets, respectively, and the second element of $V_t(\epsilon_0)$ is the VIRF of the conditional covariance. Starting at $t = 1$, it can be seen that

$$V_1(\epsilon_0) = c + B^* vech(H_0) + A^* vech(H_0^{1/2} \epsilon_{t-1} \epsilon_0' H_0^{1/2}), \tag{2.37}$$

and for $t \geq 2$,

$$V_t(\epsilon_0) = c + (A^* + B^*)V_{t-1}(\epsilon_0). \tag{2.38}$$

This recursive formula can be efficiently used in algorithms. Assuming a shock in ϵ_0 and reducing the volatility profile relative to a baseline where ϵ_0 is set to its expectation zero.

$$V_t'(\epsilon_0) = E[vech(H_t)|\epsilon_0 = (\delta_1, \delta_2)] - E[vech(H_t)|\epsilon_0 = (0, 0)] \tag{2.39}$$

We will choose one of element in (δ_1, δ_2) to be zero and the other to be positive and see how the component of H_t is affected by them. Note that $\lim_{t\to\infty} V_t(\epsilon_0) = 0$, and we can use a graph to detect how persistent our model is within different regimes.

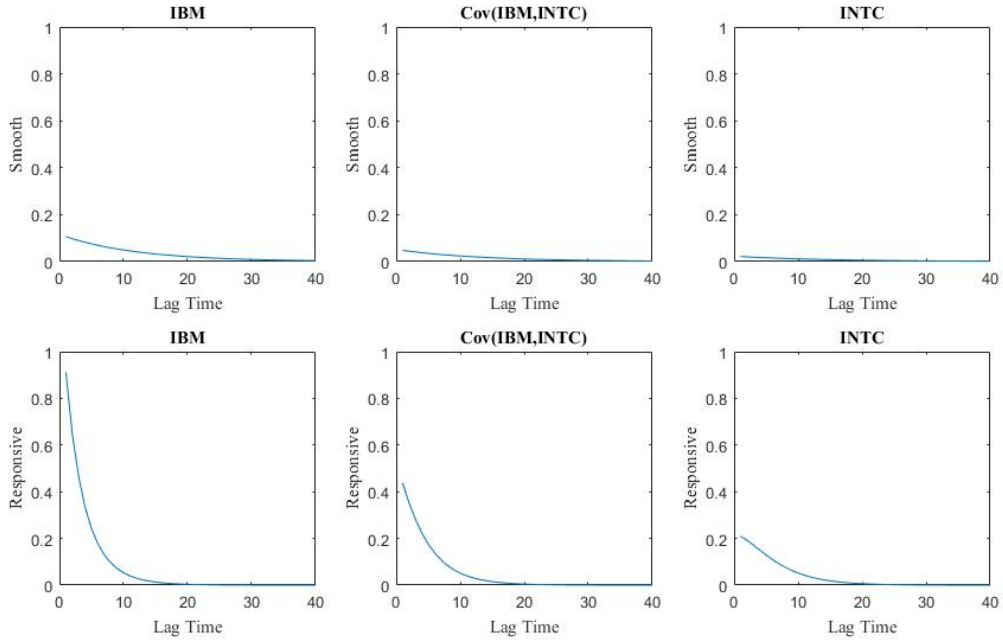

Figure 2.5: The impact of one standard deviation innovation of IBM.

In Figure 2.5, we generate one standard deviation innovation of IBM and see how it affects the IBM,

the INTC, and their covariance. The upper row shows the VIRF in the smooth regime, and the lower row shows the VIRF in the responsive regime. It is obvious that, in the smooth regime, the volatilities are persistent to the shock, which lasts for nearly one month for IBM itself and hardly affects INTC. However, in the responsive regime, the volatilities have a much larger response to the shock, but lower persistence. These results are consistent with the analysis of the dynamic parameters of the MRS-BEKK model.

2.8 Model Predictive Ability Evaluation

McCracken (2007) shows that the pseudo-out-of-sample procedure produces costly power reduction relative to the full-sample statistic while simultaneously producing no offsetting benefits. Nevertheless, We find the prediction has importance in the financial world. The goal of this section is to evaluate the out-of-sample performance of MRS-BEKK compared with the single-regime BEKK. We consider the t statistic or equal predictive ability developed by Diebold and Mariano (1995) and the F statistic proposed by McCracken (2007). Let the loss function $L_{t,s}(H^0_{t+s}, H^i_{t+s|t})$ denote the loss at time t resulting from the s-step forecast using model i, where H^0_{t+s} denotes the true (latent) covariance matrix and $H^0_{t+s|t}$ is the model i's s-step forecast conditional on the information set at time t. Since the latent covariance H^i_{t+s} is unobservable, we will use the realized kernel as the proxy such that $\hat{H}^0_{t+s} = X_{t+s}$. We use the Mean Squared error (MSE) loss function in our analysis:

$$L_{t,s}(\hat{H}^0_{t+s}, H^i_{t+s|t}) = 1/N^2 \sum_{mn=1}^{N} (\hat{h}(m,n)^0_{t+s} - h(mn)^i_{t+s|t})^2 \tag{2.40}$$

where $\hat{h}(mn)^i_{t+s|t}$ is the element in mth row and nth column in prediction covariance matrix $H^i_{t+s|t}$, similar definition goes with $\hat{h}(m,n)^0_{t+s}$. The index $i \in (1,2)$ represents the models in the comparison. The MSE-type loss function consistently ranks the volatility models as it is robust to noise in the proxy covariance X_{t+s} (see Patton and Sheppard, 2009; Laurent, Rombouts, and Violante, 2013). The total sample size $T = T_{in} + T_{out}$ is divided into in-sample and out-of-sample portion. The in-sample observations span from 1 to T_{in}. The out-of-sample observations span from $T_{in} + 1$ through $T - s$, so the

out-of-sample size is $T_{out} - s$. In our case, the estimation window (in-sample) part runs from January 4th, 2006, to April 29th, 2011, with $T_{in} = 1340$, and the out-of-sample spans from May 2, 2011, to April 7, 2012 with $T_{out} = 230$. Given the loss function, we calculate the loss difference as:

$$D_{t,s} = L_{t,s}(\hat{H}^0_{t+s}, H^{BEKK}_{t+s|t}) - L_{t,s}(\hat{H}^0_{t+s}, H^{MRS-BEKK}_{t+s|t})$$
$$t = T_{in}, T_{in}+1, ..., T-s$$

We construct the following test hypothesis:

$$H_0 : E(D_s) = 0 \tag{2.41}$$

$$vs \quad H_1 : E(D_s) > 0 \tag{2.42}$$

Diebold and Mariano (1995) suggest a test of the form using t-statistic:

$$T_{loss} = (T_{out} - s + 1)^{1/2} \frac{\bar{D}_s}{\hat{S}_d^{1/2}}, \tag{2.43}$$

with mean loss differential

$$\bar{D}_s = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} D_{t,s} \tag{2.44}$$

where $\hat{S}_d$, has been defined in Chapter 1, denotes a consistent estimate of the asymptotic variance of the mean loss differential $\bar{D}_s$. In the nested context, the null hypothesis means the extra regime does not have predictive power, which shows that the single-regime BEKK model is the true model. The significant positive value of $\bar{D}_s$ means the MRS-BEKK model has better predictive ability. As a result, we consider the one sided alternative. Note that Diebold and Mariano's (1995) statistics are probably not valid in the nested model framework. While West (1996) proves directly that the t statistic in the test we just described can be asymptotically standard normal when applied to nonnested forecasts, this is not the case when the models are nested. In particular, the results of West (1996) require that under the null, the population-level long-run variance of $D_{t,s}$ is positive. This requirement is violated with nested models.

Intuitively, with nested models, the null hypothesis that the restrictions imposed in the benchmark model are true implies that the population errors of the competing forecasting models are exactly the same. As a result, in population, the loss difference is $D_{t,s} = 0$ for all t, which makes the corresponding variance also equal to 0. Because $D_{t,s}$ and its variance converge to zero at the same rate, the test statistics have nondegenerate null distributions, but asymptotic normal distribution does not hold. Motivated by (i) the degeneracy of the long-run variance of $\bar{D}_s$ and (ii) the functional form of the standard in-sample F-test, McCracken (2007) developed an out-of-sample F-type test of equal predictive power, expressed by

$$F_{loss} = (T_{out} - s + 1)\frac{\bar{L}_{BEKK} - \bar{L}_{MRS-BEKK}}{\bar{L}_{MRS-BEKK}} = (T_{out} - s + 1)\frac{\bar{D}_s}{\bar{L}_{MRS-BEKK}} \tag{2.45}$$

where

$$\bar{L}_{BEKK} = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} L_{t,s}(\hat{H}^0_{t+s}, H^{BEKK}_{t+s|t})$$

$$\bar{L}_{MRS-BEKK} = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} L_{t,s}(\hat{H}^0_{t+s}, H^{MRS-BEKK}_{t+s|t})$$

Clark and McCracken (2001) and McCracken (2007) show that, for one-step-ahead forecasts from well-specified nested models, the F-statistics converge in the distribution to functions of stochastic integrals of quadratics of Brownian motion with limiting distributions that depend on the sample split parameter $\pi = T_{out}/T_{in}$ and the number of exclusion restrictions k, but not on any unknown nuisance parameters. The critical values of t-test for $\pi = 0.2, k = 6$ are 1.639 for 1% significance level, 0.998 for 5% significance level and 0.642 for 10% significance level; The critical values of F-test for $\pi = 0.2, k = 6$ are 3.846 for 1% significance level, 2.124 for 5% significance level and 1.313 for 10% significance level.

Another fact is that the F-test provide more significant values than t-test, which is in line with Clark and McCracken's (2005) claim that the F-type test is often more powerful than the t-type counterparts test in such a nested scenario. He also suggests using the F-test instead of the t-test because the latter is often conservative. We decide to conduct both t-test and F-test for an equal predictive ability test. The asymptotic null distributions of these statistics are non-standard. We use the critical values provided by

McCracken (2007) with the Monte Carlo method. Note that these critical values are not robust to the nuisance parameter problem, so we decide to leave them for reference, since, to our knowledge, there is no published research that solves this problem in out-of-sample. We estimate the model using an expanding window starting from 1342 observations and then use the parameter estimates to obtain forecasts of H_t at horizons $s = 1, 2, 3, ..., 22$ days. Since the estimation of MRS-BEKK takes quite a long time, the parameters are updated recursively every 5 days instead of every day, so the expanding window will grow by 5. The size of the out-of-sample is 230, in each day we predict the conditional covariance matrix for the step of 1 to 22. We report test statistics to test the equality of predictive ability in Table 2.5. The results of the F-statistic show that the difference of the value of loss function are statistically significantly positive at the 10% significance level at horizons $t = 12, 21$, statistically significantly positive at the 5% significance level at horizons $s = 4, 22$ and positive at 1% significance level at horizons $s = 5, ..., 11$. The t-statistics show that the difference of the value of loss function are statistically significantly positive at the 10% significance level at horizons $s = 4, 21$, statistically significantly positive at the 5% significance level at horizons $s = 5, 6, 7, 8, 11, 22$ and statistically significantly positive at the 1% significance level at horizons $s = 9, 10$. Therefore, we conclude that the MRS-BEKK model outperforms the single-regime BEKK model with just one regime. The effect of the performance keeps from horizons 4 to 11 but goes weaker as the prediction horizon grows. This result means the Markov regime-switching mechanism plays a very important role in improving forecasting volatility. The reason for the good forecasting performance of MRS-BEKK compared with the single-regime BEKK is that the regular realized BEKK, which only assumes one regime drifts in a certain direction when there is a random change in a real-world situation as time passes. This also explains why the MRS-BEKK model needs several days to outperform the model with one regime; the later model needs several days to drift away from the true state. In other words, the regime switching grasps the essence of randomness of reality by simulating a parallel world with different volatility dynamics and forms an expectation from them in constructing the predicted conditional covariance.

Table 2.5: Statistical Test of Predictive Ability

Prediction steps	F statistics	T statistics
1	1.248	0.582
2	1.12	0.38
3	0.505	0.177
4	2.940**	0.894*
5	4.573***	1.223**
6	6.051***	1.417**
7	6.841***	1.52**
8	5.254***	1.446**
9	5.341***	1.901***
10	4.670***	1.707***
11	4.149***	1.43**
12	2.022*	0.65*
13	1.209	0.427
14	-0.331	-0.114
15	-0.25	-0.084
16	-0.41	0.142
17	-0.129	-0.046
18	-0.49	-0.174
19	-0.322	-0.118
20	0.112	0.043
21	2.048*	0.860*
22	3.121**	1.464**

A positive value means the model using regmie switching has better predictive ability. ** means significant at 95%, *** means significant at 99%.

2.9 Conclusion

This chapter introduced the realized MRS-BEKK model for vectors of daily returns and corresponding realized measures. This model has more flexibility regarding volatility persistence; that is, not all shocks have to be highly persistent. The model generalizes the BEKK model by using two regimes with different dynamic specifications to describe the variance within the regimes. Its specification combines the attractive feature of Hansen, Huang, and Shek's (2012) realized GARCH and Klaassen's (2002) Markov regime-switching GARCH, such as the recursiveness of the likelihood function, resolved path dependence problem, and the more accurate realized measure of volatility. We also apply Noureldin, Shephard, and Sheppard's (2014) rotated techniques to ease the burden of estimation in multivariate models due to the large number of parameters. Kim's (1994) smoothed algorithm is used for the inference of the probability of regimes. Our model is easy to use for estimation and multi-period-ahead volatility forecasting for multivariate cases.

In the empirical part of the paper we estimate the model using approximately seven years of daily data on two U.S. stock returns (IBM and INTC). A comparison of our regime switching model with the single-regime BEKK model shows that our model has a better performance in and out-of-sample. The interchange of the probability of regimes seems in line with the pattern of volatility disturbance.

One drawback of our model is we cannot model the different levels of unconditional volatility in different regimes due to the rotation trick used in the model. We leave the solution of this problem for future research. Another possible extension in future work includes building Markov regime-switching models with time-varying Markov transition probabilities. Identification of the number of regimes in the Markov chain is also a promising research direction.

Chapter 3

Realized Markov Regime-Switching GARCH with time-varying transition probability

3.1 Introduction

Markov regime-switching models have been widely used in economics and finance. Good examples are changes in fiscal or monetary policies and exchange rate regime. In terms of econometrics, analyzing regime switching is difficult because econometricians rarely (directly) observe such a process, but must infer them from data. Since the seminal paper by Hamilton (1989), Markov regime-switching models have been adopted in numerous applications and extensions, see Hamilton (2010) for a survey. The key attractive feature of Markov regime-switching models is that the conditional distribution of a time series depends on an underlying latent regime, which can take only a finite number of values. The usual assumption is that the regime evolves over time as a discrete Markov chain, in which a regime at time t only depends on the regime at time $t-1$. We can summarize the statistical properties with a transition probability matrix.

Although Hamilton's (1989) fixed transition probability Markov regime-switching model (we call

it FTP model for brevity) has yielded great success, an extension of the FTP model has led to many interesting studies because of its intuitive appeal. The extension allows transition probabilities to vary over time as a function of exogenous or pre-determined variables (e.g., economic indicators). We call it the time-varying transition probability Markov regime-switching model (TVTP model). Examples of this extension appear in many fields. Researchers have used TVTP model to study business cycle fluctuations (Filardo, 1994), interest rate dynamics (Gray, 1996), bubbles and asset pricing (Schaller and Norden, 1997) and exchange rates (Diebold, Lee, and Weinbach, 1994; Engel and Hakkio, 1994). The Hamilton filter can easily be generalized to handle such cases of the time-varying transition probability.

3.2 The Framework

In this chapter, we generalize the model in Chapter 2 to allow the transition probability to vary over time using observable variables. Compared to the FTP model there are more parameters involved in the TVTP model. Therefore we limit our focus to the univariate case.

Following the realized Markov regime-switching GARCH model in Chapter 2, let $\mathbf{p}_t$ be the price of a financial asset at the end of day t, then define $r_t = log(\mathbf{p}_t) - log(\mathbf{p}_{t-1})$ as the daily return. We are using the univariate realized kernel x_t as the measure of volatility, which was proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). This realized measure has the form introduced in (1.4). Let $s_t \in Z^+$ be the unobserved regime at time t, which represents different regime of volatility dynamics. As in the last chapter, the time subscription t of expectation operators E_t and probability p_t means the arguments are conditional on some information sets, e.g., $E_t(x) \equiv E(x|\mathcal{F}_t), p_t(x) \equiv p(x|\mathcal{F}_t)$, where $\mathcal{F}_t$ represents the observable information set till period t. Therefore, we use $p_{t-1}(s_t|\tilde{s}_{t-1}) = p(s_t|\mathcal{F}_{t-1}, \tilde{s}_{t-1})$ to denote the ex-ante probability of a series going to regime s_t at a specific time t conditional on the information set of the data-generating process, which consists of two parts. The first part, $\mathcal{F}_{t-1}$, denotes the information that can be observed, which is $(r_{t-1}, r_{t-2}, \cdots)$ in our case. The second part, $\tilde{s}_{t-1}$, is the regime path $(s_{t-1}, s_{t-2}, \cdots)$, which is not observed. The dynamic of the conditional variance of the return, $h_{\tilde{s}_t,t} = var_{t-1}(r_t|\tilde{s}_t)$, is our research interest, where the subscript t of h means the conditional variance predicted for time t based on the information set at $t-1$ and the regime path. Without

loss of generality, we assume that s_t follows a two-regime first-order Markov process with transition probabilities defined as:

$$p_{t,ij} = p_t(s_t = j | s_{t-1} = i) \tag{3.1}$$
$$p_{t,i1} + p_{t,i2} = 1$$
$$\text{for} \quad i, j = 1, 2.$$

Equation (3.1) represents the probability of switching from regime i at time $t-1$ to regime j at t. The model can be easily generated to more than two regimes.

3.2.1 Markov Transition Probability

Table 3.1: Markov Transition Probability

Regime	$s_t = 1$	$s_t = 2$
$s_{t-1} = 1$	$p_{t,11}$	$1 - p_{t,11}$
$s_{t-1} = 2$	$1 - p_{t,22}$	$p_{t,22}$

We illustrate the transition matrix with these probabilities in Table 3.1 with $N = 2$ different regimes, it is a 2 by 2 matrix. Compared with the setting in Chapter 2, we drop the constant restriction upon the transition probabilities to let them be time-varying. To guarantee the probability $p_t \in [0, 1]$, we model it with the logistic function of $z_{t-1}\prime\zeta_i, i = 1, 2$, where the conditioning vector z_{t-1} contains economic variables that affect the regime transition probabilities and ζ_i is a vector of parameters. Intuitively, the volatility of the stock return is the main reason that drives the market to switch between different regimes. Hence, we choose the realized kernel x_t as the variable to affect the transition probabilities. In this way, the dynamics of the regime can be determined endogenously. We also notice that the negative returns have a larger impact on the conditional volatility than positive returns (see Engle and Ng, 1993; Ding, Granger, and Engle, 1993). The GJR GARCH model (Glosten, Jagannathan, and Runkle, 1993)

introduces a dummy variable to carry the asymmetric effect of return in the dynamic process of volatility. They allow the parameter before the innovation of return in GARCH model to change to a different level whenever the daily return is negative. In our Markov regime-switching setting, we introduce the asymmetric effect of return into the transition probabilities by adding a dummy variable into the logistic regression. As a result, the leverage effect play a role in the Markov regime-switching process. Specifically, the transition probabilities are functions of the realized volatility of stock return as follows:

$$
\begin{aligned}
p_{t,11} &= \frac{\exp(\omega_1 + \beta_1 x_{t-1} + \alpha_1 x_{t-1} D_{t-1})}{1 + \exp(\omega_1 + \beta_1 x_{t-1} + \alpha_1 x_{t-1} D_{t-1}))} \\
p_{t,22} &= \frac{\exp(\omega_2 + \beta_2 x_{t-1} + \alpha_2 x_{t-1} D_{t-1})}{1 + \exp(\omega_2 + \beta_2 x_{t-1} + \alpha_2 x_{t-1} D_{t-1})} \\
D_t &= \begin{cases} 1, \text{if } r_t < 1 \\ 0, \text{otherwise} \end{cases}
\end{aligned}
\tag{3.2}
$$

It is worth noting that when the α_1 and α_2 are set to zero, then the model is TVTP model. If β_1, α_1 and β_2, α_2 are set to zero, the transition probability functions are time-invariant so the TVTP model degenerates into the FTP model with $p_{t,11}$ and $p_{t,22}$ simply constants. If no restrictions are applied to the parameters, we call it time-varying transition probability Markov regime-switching model with negative return effect (TVTP-N model). We will estimate the three nested models in the empirical section later.

3.2.2 Realized Markov Regime-switching GARCH Model

In this section, we introduce our Realized Markov regime-switching GARCH model (MRS-GARCH model). As discussed in the introduction of Chapter 2, the MRS-GARCH model is intractable due to the regime-dependent structure and unobserved discrete regimes unless the sample size T is uselessly small. Several approaches have been applied to tackle this problem. The one we are using is Gray's (1996) collapsing method. Under the full regime path-dependent specification, the conditional variance of return r_t is defined as $h_{t,\tilde{s}_t} = var_{t-1}(r_t|\tilde{s}_t)$ which follows a GARCH(1,1) process (the following

specification can be generalized to GARCH(P,Q) straightforwardly),

$$h_{t,\tilde{s}_t} = c + b_{s_{t-1}}\, h_{t-1,\tilde{s}_{t-1}} + a_{s_{t-1}}\, x_{t-1}, \tag{3.3}$$

where c is the intercept independent of regime, b_{s_t-1} is the smoothing parameter, which represents how smooth the variance process is, and a_{s_t-1} is the innovation parameter which represents the responsiveness of volatility dynamic to shocks. The subscript $\tilde{s}_t$ means the variable is conditional on all the regime along the historical path. Since $h_{\tilde{s}_t-1,t-1}$ depends on the whole regime path, this model becomes intractable due to the number of paths grows at an exponential rate. Therefore, we specify a different dynamic for the conditional variance that does not have this path dependence. As a result, we switch subscript $\tilde{s}_t$ to s_t to indicate the variable is solely conditional on current regime. Therefore, the return r_t is assumed to follow

$$\begin{aligned} r_t &= h_{t,s_t}^{1/2} z_t, \\ z_t &\overset{i.i.d}{\sim} (0,1), \end{aligned} \tag{3.4}$$

where h_{t,s_t} follows a MRS-GARCH(1,1) process:

$$h_{t,s_t} = c + b_{s_{t-1}}\, E_{t-1}[h_{t-1,s_{t-1}}|s_t] + a_{s_{t-1}}\, x_{t-1}. \tag{3.5}$$

The expected value is calculated as:

$$E_{t-1}[h_{t-1,s_{t-1}}|s_t] = \sum_{s_{t-1}=1}^{2} p_{t-1}(s_{t-1}|s_t) h_{t-1,s_{t-1}} \tag{3.6}$$

where the probability is calculated as

$$p_{t-1}(s_{t-1}|s_t) = \frac{p_{t-1}(s_{t-1}, s_t)}{p_{t-1}(s_t)}. \tag{3.7}$$

The term in the nominator is a product of the transition probabilities and filtered probabilities, and the denominator is the ex-ante probability. We have calculated the filtered probability and ex-ante probability recursively in Chapter 2, and the transition probability can be calculated by the logistic function in our model, i.e. equation (3.2).

3.3 Estimation and Inference

3.3.1 Estimation for GARCH Equation

In this section we discuss the asymptotic properties of the quasi-maximum likelihood estimator within the MRS-GARCH(1,1) model given by (3.4) and (3.5). The estimation is similar to that in Chapter 2 for the Markov regime-switching model with constant transition probabilities. In addition, we include the parameters ω_i, α_i, and β_i $(i = 1, 2)$ in modeling the time-varying transition probabilities beside the regime-dependent GARCH parameters. We follow the recursive estimation process introduced in Chapter 2. The Realized MRS-GARCH model defined by (3.5) and (3.2) is parameterized with a finite-dimensional $(\delta \times 1)$ parameter vector $\theta \in \Theta \subset \mathbb{R}^{\delta}$. Therefore, we write parameters as $\theta = (\theta_d', \theta_s')'$ with dynamic parameters $\theta_d = [c, b_1, b_2, a_1, a_2, \omega_1, \beta_1, \alpha_1, \omega_2, \beta_2, \alpha_2]$ denote the coefficients in GARCH model and logistic regression for transition probabilities. The static parameters $\theta_s = (\bar{h}, \bar{x})'$ denote the unconditional expectation of the variance and the realized kernel, which can be estimated by sample averages. Specifically, the static parameters are estimated by a moment estimator:

$$\hat{\bar{h}} = T^{-1} \sum_{t=1}^{T} r_t^2, \qquad \hat{\bar{x}} = T^{-1} \sum_{t=1}^{T} x_t. \tag{3.8}$$

These moment estimators are used in the variance targeting introduced later in section 3.3.3. Letting $L_t(\theta)$ be the likelihood of observation t and $l_T(\theta)$ be the joint log-likelihood function, gives

$$l_{s_t,T}(\theta_d; \hat{\theta}_s) = \sum_{t=1}^{T} \log L_t(\theta_d; \hat{\theta}_s | s_t) = \sum_{t=1}^{T} \left[constant - \frac{1}{2} (\log h_{t,s_t}^2 + \frac{r_t^2}{h_{t,s_t}}) \right] \tag{3.9}$$

According to (3.4), the distribution of r_t is not assumed normal. Equation (3.9) is the quasi-likelihood function and the estimation is performed by QML, which provides consistent estimates regardless of the true distribution of z_t (see Bollerslev and Wooldridge, 1992). We assume the initial value of h_{t,s_t} in first period is known and nonnegative. For our path-independent model, we maximize the sum of log-likelihood function with expectation in each period with respect to the distribution of the regime. Let

$$l_T(\theta_d; \hat{\theta}_s) = \sum_{t=1}^{T} \log E(L_t(\theta_d; \hat{\theta}_s | s_t)) = \sum_{t=1}^{T} \log \sum_{s_t=1}^{N} L_t(\theta_d; \hat{\theta}_s | s_t) p_{t-1}(s_t) \tag{3.10}$$

where the ex-ante probability $p_{t-1}(s_t)$ is defined as

$$p_{t-1}(s_t) = \sum_{s_{t-2}=1}^{2} p_{t-1}(s_{t-1}) \cdot p_{t-1}(s_t | s_{t-1}) \tag{3.11}$$

where the latter term is the transition probability defined in (3.2). The QML estimator is $\hat{\theta}_d$, where $\hat{\theta}_d = arg\, max_{\theta_d \in \Theta} l_T(\theta_d; \hat{\theta}_s)$. The asymptotic theory for QML estimation of the GARCH model is discussed by Comte and Lieberman (2003) and Hafner and Preminger (2009). For the closely related variance-targeting GARCH model, Pedersen and Rahbek (2014) prove its consistency and asymptotic normality.

Since we use two-step QML estimation, if the usual assumptions for the validity of QML estimator are satisfied, the estimates are consistent and their asymptotic distribution is:

$$\sqrt{T} \left(\begin{bmatrix} \hat{\theta}_s \\ \hat{\theta}_d \end{bmatrix} - \begin{bmatrix} \theta_s \\ \theta_d \end{bmatrix} \right) \rightarrow \mathbb{N}(0, V)$$

with

$$V = \begin{bmatrix} G_{\theta_s}^{-1} & -G_{\theta_s}^{-1} G_{\theta_d} M^{-1} \\ 0 & M^{-1} \end{bmatrix} E \left[\frac{\partial l_T}{\partial \theta} \frac{\partial l_T}{\partial \theta'} \right] \begin{bmatrix} G_{\theta_s}^{-1} & -G_{\theta_s}^{-1} G_{\theta_d} M^{-1} \\ 0 & M^{-1} \end{bmatrix}$$

where

$$G_{\theta_s} = E\left[\frac{\partial g(\theta_s, \theta_d)}{\partial \theta_s'}\right]; G_{\theta_d} = E\left[\frac{\partial g(\theta_s, \theta_d)}{\partial \theta_d'}\right]; M = E\left[\frac{\partial m(\theta_d)}{\partial \theta_d'}\right]$$
$$g(\theta_s, \theta_d) = \left[\frac{\partial l_T(\theta_s, \theta_d)}{\partial \theta_s}\right]; m(\theta_d) = \left[\frac{\partial l_T(\theta_s, \theta_d)}{\partial \theta_d}\right]$$

The proof is in Newey and McFadden (1994).

3.3.2 Measurement Equation in Forecasting

As in the previous chapters, we use a logarithm form for the measurement equation to update the realized kernel x_t when doing multi-step ahead forecasting. The univariate measurement equation is defined as:

$$\log x_t = d + \gamma \log(h_t) + u_t \tag{3.12}$$
$$u_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_u)$$

where

$$h_t = E_{t-1}(h_{t,s_t}) = \sum_{s_t=1}^{N} p_{t-1}(s_t) h_{t,s_t} \tag{3.13}$$

where h_{t,s_t} are computed in the MRS-GARCH equation and the forecasting step runs in a recursive manner.

3.3.3 Variance Targeting

We use variance targeting to estimate the static coefficients in the MRS-GARCH equation as described in the previous chapters. Variance targeting allows for two-step estimation with a first step consisting in estimating parameters representing unconditional moments with empirical moments and a second step of QML estimation which requires less computing resources. Two-step estimation is useful as it reduces the number of parameters needed to estimate in the second step. Following the notations in (3.8) the

variance targeting is demonstrated in the following equation:

$$h_{t,s_t} = (\bar{h} - b_{s_t}\bar{h} - a_{s_t}\bar{x}) + b_{s_t}\, E_{t-1}(h_{t-1,s_{t-1}}|s_t) + a_{s_t}\, x_{t-1} \tag{3.14}$$

assuming $(\bar{h} - b_{s_t}\bar{h} - a_{s_t}\bar{x})$ is nonnegative.

3.4 Empirical Illustration

So far, we have generalized FTP model to TVTP and TVTP-N models to obtain more flexibility in modeling volatility dynamics. In this section we estimate all the models and examine whether this generalization pays off empirically in terms of improving model fitting and volatility forecasts.

3.4.1 Data Description

The model is applied to the high-frequency price records for stock with sticker name IBM, one of the most liquid stocks from the technology sector. The sampling period is from January 3rd, 2006, to April 30th, 2012, and the source of data is the TAQ database. We compute daily returns and the noise-robust realized kernel x_t from EST 9:30-16:00, which are the weekday hours of the stock market. To eliminate the overnight effect, the prices recorded during the 15 minutes at the beginning and end of opening hours are excluded.

3.4.2 Estimation Result

This subsection presents the estimation results for three models: the FTP model, the TVTP model and the TVTP-N model. Table 3.2 presents the QML estimation results. As expected, the TVTP model provides a better fit than the FTP model. The gain in the log-likelihood value is 6.

From the estimates of FTP, it is obvious that the process is split into two distinct regimes. In the first regime, b_1 is much larger than a_1; in the other regime, b_2 is smaller than a_2. The parameters a_1 and b_2 show are not statistically significant at the usual significance levels. We call regime 1 the smooth regime, while regime 2 is the responsive regime. The transition probability estimates (0.843, 0.874) in

the FTP model are fairly large which shows that the regime does not switch frequently.

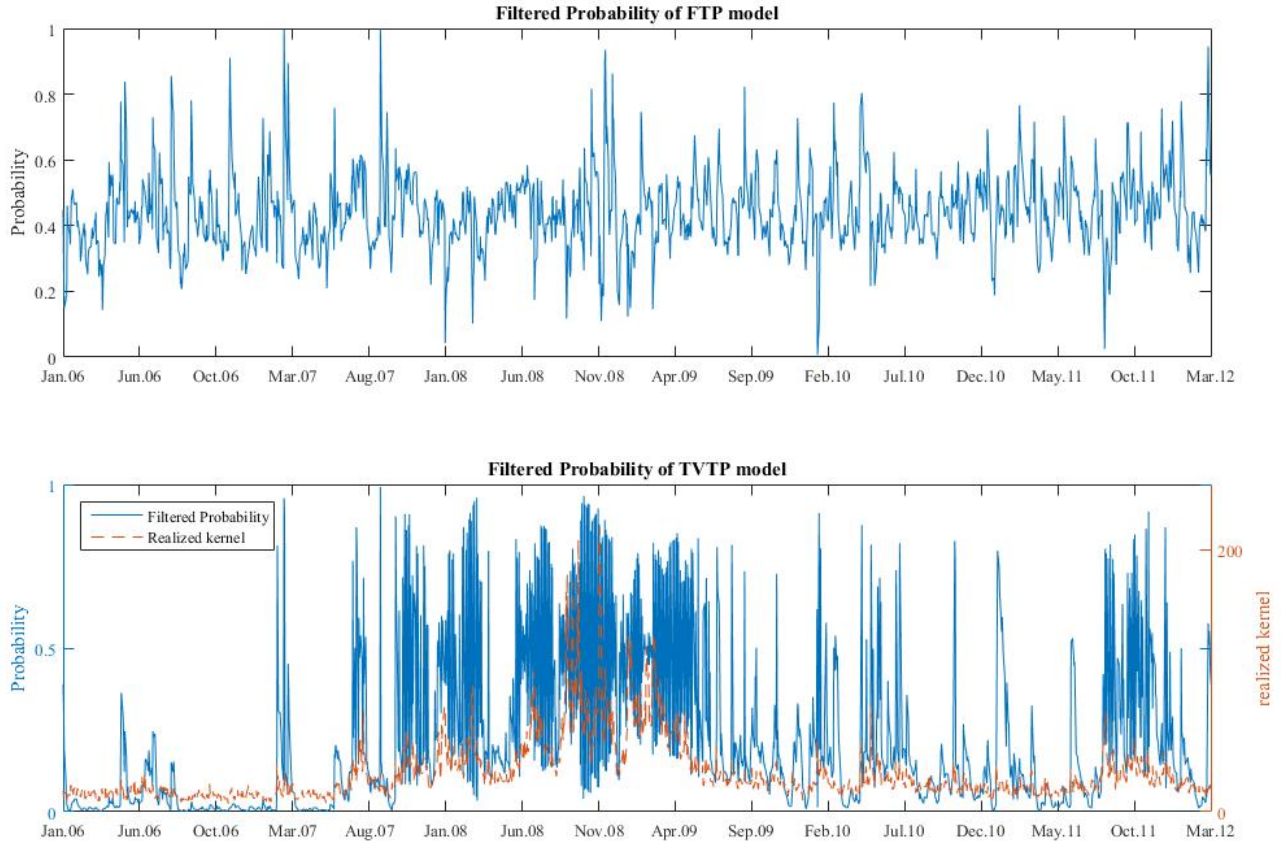

Figure 3.1: The filtered probabilities of FTP (upper panel) and TVTP (lower panel) in the responsive regime (regime 2).

In the TVTP model, we obtain a similar classification of regime with respect to the estimate of dynamic parameters b and a. However, the parameters in the responsive regime (b_2 and a_2) are not as extreme as the counterpart in the FTP model, and they are both significant at the 1% significance level. We compare the filtered probability of staying in the responsive regime (regime 2) in Figure 3.1. From the graph, we can tell that the occurring of transition of regime in FTP model is more evenly distributed than the one in the TVTP model. For the TVTP model, during the financial crisis in 08-09, the transition between regimes of volatility process is much more frequent than other periods due to the time varying transition probabilities affected by the surging volatility.

The estimates of the parameters in the logistic regression $(\omega_i, \beta_i, \alpha_i)$ for the TVTP and TVTP-N model reveal how the volatility affects Markov transition probabilities. Both slopes β_1 and β_2 are negative, which indicates that the larger volatility leads to larger probability regime switching regardless of the regime as illustrated in Figure 3.1. Furthermore, α_1 and α_2 in the TVTP-N model are both negative, which indicates that negative returns make the regimes even less persistent. The difference of log-likelihood between FTP and TVTP is 5.7. The value for the LR test is defined as twice the difference $LR_T = 2*(L_{TVTP} - L_F) = 11.4$, and the p-value is 0.002 based on a chi-square distribution with 2 degree of freedom. Therefore, the null hypothesis is rejected at the 1% significance level. However, there is no significant difference ($p = 0.511$) of log-likelihood between TVTP and TVTP-N model to conclude that negative return effect play any important rule in modeling transition probabilities. Since the TVTP-N model has no improvement in likelihood compared with the TVTP model and the estimates of α_1 is not significant, we decide to keep the analysis of TVTP-N as only an illustration and will not consider it in the forecast evaluation below.

The measurement equation (3.12) is a static time series model. To see how the model fits, we want to check if there is any autocorrelation of the residuals of the OLS estimation. The Lagrange multiplier (LM) test for autocorrelation developed by Breusch (1978) and Godfrey (1978) is a standard tool in applied econometrics due to its simplicity and flexibility. The test is performed through an auxiliary

Table 3.2: Estimation of GARCH Coefficient

Parameters in GARCH model	FTP	TVTP	TVTP-N
b_1	0.990*(0.345)	0.943*(0.08)	0.952*(0.028)
a_1	0.010(0.017)	0.010(0.057)	0.010(0.04)
b_2	0.010(0.041)	0.376*(0.112)	0.378*(0.114)
a_2	0.980*(0.189)	0.621*(0.109)	0.619*(0.111)
ω_1	0.841*(0.071)	8.101*(1.97)	9.887*(6.876)
β_1	0.73*(0.069)	-2.787*(0.721)	-3.208*(1.896)
α_1	-	-	-0.828(1.813)
ω_2	-	7.373*(2.224)	8.151*(4.75)
β_2	-	-1.647*(0.785)	-1.663*(1.158)
α_2	-	-	-0.444(0.586)
Log-likelihood	-2933.1	-2928.4	-2927.7

Parameter estimates of three different GARCH models: FTP (fixed transition probabilities); TVTP (time varying probabilities driven by realized kernel); TVTP-N (time varying probabilities driven by realized kernel with a leverage effect). Values in parentheses are the standard errors.
$*$-estimates are statistically significant at the 5% level.

regression of the residuals $\hat{u}_t$ on their l lags $\hat{u}_{t-l}$ and original regressors:

$$\hat{u}_t = \rho + \rho_0 \log(h_t) + \rho_1 \hat{u}_{t-1} + ... + \rho_l \hat{u}_{t-l} + e_t, \tag{3.15}$$

where ρ is the intercept and the error term e_t is a zero mean normal process. Then the significance of all regressors is tested. We can drop the original independent variable $\log(h_t)$ if we have orthogonal assumption in (3.12), which is valid according to the our empirical result. The test statistic is computed as TR^2, where T is the sample size, and R^2 is the usual r-square of auxiliary regression (3.15). This statistic has an asymptotic χ^2_l distribution under the null of no serial correlation when l lagged residuals are used. This is usually called the Breusch-Godfrey test for AR(l) serial correlation. There is strong evidence of AR(2) in the error u_t for the FTP model. The significant estimates, standard errors, between parameters, in (3.15) are $\rho_1 = 0.052(0.025)$, $\rho_2 = 0.078(0.025)$ with sample size $T - 2 = 1590$ and $R^2 = 0.01$. As a result the LM statistic for testing (3.15) is 14.6 which re-

ject the null at 1% significance level. For the TVTP model, there is only one significant estimate in (3.15), $\rho_1 = 0.186(0.025)$ with sample size 1591 and $R^2 = 0.03$. The LM statistic for testing is 52.9 which reject the null at 1% significance level. We use the Cochrane-Orcutt (CO) estimation developed by Cochrane and Orcutt (1949) to correct the autocorrelation with the quasi-differenced data. For the FTP model the modeled data are differenced as $\log x_t^* = \log x_t - 0.052 \log x_{t-1} - 0.078 \log x_{t-2}$ and $\log h_t^* = logh_t - 0.052 \log h_{t-1} - 0.078 \log h_{t-2}$; for the TVTP model, the data are differenced as $\log x_t^* = \log x_t - 0.186 \log x_{t-1}$ and $logh_t^* = logh_t - 0.186 logh_{t-1}$. After the correction, we re-run the LM test using $\log x_t^*$ and $\log h_t^*$ only find that there is no significant evidence to reject the null of no autocorrelation in u_t up to lag 10. Table 3.3 presents the estimates of the measurement equation corrected for serial correlations. The estimates of γ in (3.12) for both FTP and TVTP model are significant and close to 1 (not rejecting equality of 1 with t test at 1% significance level) indicating that the realized kernel is a good measure of the conditional variance.

Table 3.3: Estimation of coefficient measurement equation

Parameters in GARCH model	FTP	TVTP
d	-0.104(0.017)	-0.06(0.019)
γ	0.987(0.012)	0.971(0.013)
σ_u^2	0.215	0.226

Estimates of the coefficients in the measurement equation for the FTP and TVTP models. The values in parentheses are the standard error. All the parameters are statistically significant at the 1% level.

In multi-step ahead forecasting, the measurement equation is used to project the realized kernel. We utilize the detected serial correlation of error u_t in (3.12) in computing the forecasts. Specifically, for the FTP model when predict the conditional variance h_{t+2}, we want to compute

$$E_t(\log x_{t+1}) = d + \gamma\, E_t(\log(h_{t+1})) + E_t(u_{t+1}) \tag{3.16}$$

where

$$E_t(u_{t+1}) = d + \hat{\rho}_1 \hat{u}_t + \hat{\rho}_2 \hat{u}_{t-1}. \tag{3.17}$$

We can recursively update the error term $E_t(u_{t+s})$ for $s+1$-step ahead forecasting. To predict $E(x_{t+s})$, we use the property of log-normal distribution.

$$E_t(x_{t+s}) = \exp(E_t(\log x_{t+s}) + \hat{\sigma}_u^2/2)$$

For the TVTP model, we follow the same procedure except adapting the (3.17) to a AR(1) structure. The coefficients for (3.17) are re-estimated at the same time with other parameters in the whole model when forecasting.

3.5 Model Predictive Ability Evaluation

In this section, our goal is to evaluate the out-of-sample performance of the two competing models. Similarly to Chapter 2, we consider the t statistic developed by Diebold and Mariano (1995) and the F statistic proposed by McCracken (2007) for the equal predictive ability test. Let the loss function $L_{t,s}(h^0_{t+s}, h^i_{t+s|t})$ denote the loss at time t resulting from the s-step forecast using model i, where h^0_{t+s} denotes the true (latent) variance and $h^i_{t+s|t}$ is model i's s-step forecast conditional on the information set at time t. Since the latent variance h^0_{t+s} is unobservable, we will use the realized kernel as the proxy such that $\hat{h}^0_{t+s} = x_{t+s}$. The mean squared error (MSE) type loss function is used in our analysis:

$$L_{t,s}(\hat{h}^0_{t+s}, h^i_{t+s|t}) = (\hat{h}^0_{t+s} - (h^i_{t+s}))^2 \tag{3.18}$$

The total sample size $T = T_{in} + T_{out}$ is divided into in-sample and out-of-sample portion. The in-sample observations span from 1 to T_{in}. The out-of-sample observations span from $T_{in}+1$ through $T-s$, so the out-of-sample size is $T_{out} - s$. In our case, the estimation window (in-sample) part runs from January 4th, 2006, to April 29th, 2011, with $T_{in} = 1340$, and the out-of-sample spans from May 2, 2011, to

April 7, 2012 with $T_{out} = 230$.

Given the loss function, we could calculate the loss difference as:

$$D_{t,s} = L_{t,s}(\hat{h}^0_{t+s}, h^{FTP}_{t+s|t}) - L_{t,s}(\hat{h}^0_{t+s}, h^{TVTP}_{t+s|t}) \qquad (3.19)$$
$$t = T_{in}, T_{in}+1, ..., T-s$$

We construct the following test hypothesis:

$$\begin{aligned} & H_0 : E[D_{t,s}] = 0 \\ vs \quad & H_1 : E[D_{t,s}] > 0 \end{aligned} \qquad (3.20)$$

The null implies the informative variables included in the transition probabilities are not providing more forecasting ability. The alternative, positive value of $E[D_{t,s}]$, means the TVTP model has better predictive ability. As a result, the test is one-sided to the right. The out-of-sample t-type test:

$$T_{loss} = (T_{out} - s + 1)^{1/2} \hat{S}_d^{-1/2} \bar{D}_s \qquad (3.21)$$

with mean loss differential

$$\bar{D}_s = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} D_{t,s} \qquad (3.22)$$

where $\hat{S}_d$, has been defined in Chapter 1, denotes a consistent estimate of the asymptotic variance of the mean loss differential $\bar{D}_s$. The out-of-sample F-type statistic of equal predictive power is as follows:

$$F_{loss} = (T_{out} - s + 1)\frac{\bar{L}_{FTP} - \bar{L}_{TVTP}}{\bar{L}_{TVTP}} = (T_{out} - s + 1)\frac{\bar{D}_s}{\bar{L}_{TVTP}} \qquad (3.23)$$

where

$$\bar{L}_{FTP} = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} L_{t,s}(\hat{h}^0_{t+s}, h^{FTP}_{t+s|t})$$

$$\bar{L}_{TVTP} = \frac{1}{(T_{out} - s + 1)} \sum_{t=T_{in}}^{T-s+1} L_{t,s}(\hat{h}^0_{t+s}, h^{TVTP}_{t+s|t})$$

Clark and McCracken (2001) and McCracken (2007) show that, for one-step-ahead forecasts from well-specified nested models, both the t-statistic and the F-statistics converge in the distribution to the functions of stochastic integrals of quadratics of Brownian motion, with limiting distributions that depend on the sample split parameter $\pi = T_{out}/T_{in}$ and the number of exclusion restrictions k, but not on any unknown nuisance parameters. Clark and McCracken (2005) claims that F-type test is often more powerful than the t-type counterparts test in such nested scenario. They also suggest using the F-test instead of the t-test since the latter is often conservative. The asymptotic null distributions of the F-test are nonstandard. Each can be written as functions of the stochastic integrals of Brownian motion. We use the critical value simulated by McCracken (2007) with the Monte Carlo method. In our test, The critical values of t-test for $\pi = 0.2, k = 4$ are 1.731 for 1% significance level, 1.101 for 5% significance level and 0.742 for 10% significance level; The critical values of F-test for $\pi = 0.2, k = 4$ are 3.44 for 1% significance level, 1.964 for 5% significance level and 1.225 for 10% significance level.

We estimate the model using an expanding window starting from 1342 observations and then use the parameter estimates to obtain forecasts of h_t at horizons $s = 1, 2, 3, ..., 22$ days. To save the computation time, the parameters are updated recursively every 5 days instead of every day, so the expanding window grows by 5. The size of the out-of-sample is 230, in each day we predict the conditional variance for the step for 1 to 22 step ahead. We report both F-statistic and t-statistic to test the equality of predictive ability in Table 3.4. The result of the F-test shows that the difference of the value of loss function is statistically significantly positive at the 5% significance level at horizons $s = 3, 7, 11$ and statistically significantly positive at the 1% significance level at horizon $s = 1, 2, 8, 9, 10$; statistically

significantly negative at the 10% at horizon $s = 13$, statistically significantly negative at the 5% significance level at horizons $s = 14, ..., 19, 21$ and statistically significantly negative at the 1% significance level at horizons $s = 20, 22$. The result of the t-test shows that the difference of the value of loss function is statistically significantly positive at the 10% significance level at horizons $s = 2, 7$ and statistically significantly positive at the 5% significance level at horizons $s = 1, 8, 9, 10$; statistically significantly negative at the 10% at other horizons $s = 7, 14, 15, 16$, statistically significantly negative at the 5% horizons $s = 18, 19, 21$ and statistically significantly negative at the 1% horizons $s = 20, 22$. In summary, the result of both test show that these two models can't beat the other uniformly. In near future the TVTP model statistically significantly outperforms the FTP model in terms of volatility forecasting. The FTP model becomes better in forecasting performance as the prediction horizon grows. The possible reason is that the predicted realized kernel included in the logistic function for the transition probabilities has a predictive error. Such error brought into the transition matrix grows along with the forecast horizons.

3.6 Conclusion

The final chapter introduces the univariate realized Markov regime-switching model featuring time-varying transition probabilities. This model has more flexibility regarding volatility persistence than does the Markov regime-switching model in Chapter 2; that is, the Markov transition probability reflects market information through a logistic regression model. The realized measure of volatility is added as an independent variable in the logistic model. Its specification combines the attractive features of Hansen, Huang, and Shek's (2012) realized GARCH, Klaassen's (2002) Markov regime-switching GARCH and Diebold, Lee, and Weinbach's (1994) Markov regime-switching model, such as recursiveness of the likelihood function, resolution of the path dependence problem, and a more accurate realized measure of volatility. Our model is easy to use for estimation and multi-period-ahead forecasting for financial market volatility.

In the empirical part of the paper we examine the model using about seven years of daily data on one of the most liquid U.S. stock (IBM). A comparison of two Markov regime-switching models with constant and time-varying transition probabilities shows that the TVTP model has better in-sample per-

Table 3.4: Statistical Test of Predictive Ability

Prediction steps	F statistics	t statistics
1	9.5530***	1.7249**
2	6.0054***	1.0130*
3	2.1648**	0.4587
4	0.3178	0.0817
5	-0.5712	-0.1586
6	0.1134	0.0341
7	2.3535**	0.7517*
8	4.8890***	1.4623**
9	4.8438***	1.4328**
10	4.8826***	1.2505**
11	1.7261**	0.5287
12	-0.0144	-0.0049
13	-1.3843*	-0.4867
14	-2.3857**	-0.8998*
15	-2.2866**	-0.9386*
16	-2.2708**	-0.8839*
17	-1.4457**	-0.5281
18	-2.6572**	-1.0931**
19	-2.9010**	-1.4237**
20	-3.6368***	-1.7587***
21	-3.1044**	-1.3073**
22	-3.7268***	-1.5665**

Test of predictive ability: Loss(FTP)-Loss(TVTP). A positive value means the model using TVTP has better predictive ability.* means statistically significant at 10%, ** means statistically significant at 5%, *** means significant at 1%.

formance and better out-of-sample performance in the near future than the FTP model. The realized measure of volatility seems to provide useful information for updating the Markov transition probabilities.

One drawback of our empirical application is that we cannot obtain stronger evidence that the time-varying transition probability offer a better forecasting result uniformly across forecasting horizons. However, this might due to the incomplete list of informative variables in the logistic regression. A possible future work is identifying and including more economic indicators in modeling the Markov transition probability to enable the regime transition process to reflect the ongoing economic regime.

Bibliography

Andersen, T. G., and Benzoni, L. (2008). 'Realized Volatility', *SSRN eLibrary*, 71(2): 579–625.

Andersen, T. G., and Bollerslev, T. (1998). 'Towards a unified framework for high and low frequency return volatility modeling', *Statistica Neerlandica*, 52(3): 273–302.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). 'The distribution of realized stock return volatility', *Journal of financial economics*, 61(1): 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). 'Modeling and Forecasting Realized Volatility', *Econometrica*, 71(2): 579–625.

Bandi, F. M., and Russell, J. R. (2008). 'Microstructure noise, realized variance, and optimal sampling', *The Review of Economic Studies*, 75(2): 339–369.

Barndorff-Nielsen, O. E. (2002). 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2): 253–280.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica*, pp. 1481–1536.

——— (2011). 'Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics*, 162(2): 149–169.

Barndorff-Nielsen, O. E., and Shephard, N. (2004). 'Power and bipower variation with stochastic volatility and jumps', *Journal of financial econometrics*, 2(1): 1–37.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). 'Multivariate GARCH models: a survey', *Journal of applied econometrics*, 21(1): 79–109.

Bollerslev, T. (1986). 'Generalized Autorgressive Conditional Heteroskedasticity', *Journal of Econometrics*, 31(3): 307–327.

Bollerslev, T., and Wooldridge, J. M. (1992). 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric reviews*, 11(2): 143–172.

Breusch, T. S. (1978). 'Testing for autocorrelation in dynamic linear models', *Australian Economic Papers*, 17(31): 334–355.

Cai, J. (1994). 'A Markov model of switching-regime ARCH', *Journal of Business & Economic Statistics*, 12(3): 309–316.

Carrasco, M., Hu, L., and Ploberger, W. (2014). 'Optimal test for markov switching parameters', *Econometrica*, 82(2): 765–784.

Christoffersen, P., Jacobs, K., and Mimouni, K. (2010). 'Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices', *Review of Financial Studies*, 23(8): 3141–3189.

Clark, T. E., and McCracken, M. W. (2001). 'Tests of equal forecast accuracy and encompassing for nested models', *Journal of econometrics*, 105(1): 85–110.

——— (2005). 'Evaluating direct multistep forecasts', *Econometric Reviews*, 24(4): 369–404.

Cochrane, D., and Orcutt, G. H. (1949). 'Application of least squares regression to relationships containing auto-correlated error terms', *Journal of the American statistical association*, 44(245): 32–61.

Comte, F., and Lieberman, O. (2003). 'Asymptotic theory for multivariate GARCH processes', *Journal of Multivariate Analysis*, 84(1): 61–84.

Diebold, F. X., Lee, J.-H., and Weinbach, G. C. (1994). 'Regime switching with time-varying transition probabilities', *Business Cycles: Durations, Dynamics, and Forecasting*, pp. 144–165.

Diebold, F. X., and Mariano, R. S. (1995). 'Comparing Predictive Accuracy', 20(1): 134–144.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). 'A long memory property of stock market returns and a new model', *Journal of empirical finance*, 1(1): 83–106.

Dobrev, D., and Szerszen, P. (2010). 'The information content of high-frequency data for estimating equity return models and forecasting risk', FEDS Working Paper.

Doornik, J. A. (1996). 'Testing vector error autocorrelation and heteroscedasticity', .

Dueker, M. J. (1997). 'Markov switching in GARCH processes and mean-reverting stock-market volatility', *Journal of Business & Economic Statistics*, 15(1): 26–34.

Efron, B. (1979). 'Bootstrap methods: another look at the jackknife', *The annals of Statistics*, 7(1): 1–26.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). 'Least angle regression', *The Annals of statistics*, 32(2): 407–499.

Engel, C., and Hakkio, C. S. (1994). 'The distribution of exchange rates in the EMS', Discussion paper, National Bureau of Economic Research.

Engle, R. (2002). 'New frontiers for ARCH models', *Journal of Applied Econometrics*, 17(5): 425–446.

Engle, R. F. (1982). 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, pp. 987–1007.

Engle, R. F., and Gallo, G. M. (2006). 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics*, 131(1): 3–27.

Engle, R. F., and Kroner, K. F. (1995). 'Multivariate simultaneous generalized ARCH', *Econometric theory*, 11(01): 122–150.

Engle, R. F., and Ng, V. K. (1993). 'Measuring and testing the impact of news on volatility', *The journal of finance*, 48(5): 1749–1778.

Epps, T. W. (1979). 'in Stock Prices in the Very Short Run Comovements to Three Days', *Journal of the American Statistical Association*, 74(366): 291– 298.

Filardo, A. J. (1994). 'Business-cycle phases and their transitional dynamics', *Journal of Business & Economic Statistics*, 12(3): 299–308.

Forsberg, L., and Bollerslev, T. (2002). 'Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): The GARCH-NIG model', *Journal of Applied Econometrics*, 17(5): 535–548.

Garcia, R. (1998). 'Asymptotic null distribution of the likelihood ratio test in Markov switching models', *International Economic Review*, pp. 763–788.

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

Godfrey, L. G. (1978). 'Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables', *Econometrica: Journal of the Econometric Society*, pp. 1293–1301.

Gray, S. F. (1996). 'Modeling the conditional distribution of interest rates as a regime-switching process', *Journal of Financial Economics*, 42(1): 27–62.

Hafner, C. M., and Preminger, A. (2009). 'On asymptotic theory for multivariate GARCH models', *Journal of Multivariate Analysis*, 100(9): 2044–2054.

Hamilton, J. D. (1989). 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica: Journal of the Econometric Society*, pp. 357–384.

——— (2010). 'Regime switching models', in *Macroeconometrics and Time Series Analysis*, pp. 202–209. Springer.

Hamilton, J. D., and Susmel, R. (1994). 'Autoregressive conditional heteroskedasticity and changes in regime', *Journal of Econometrics*, 64(1): 307–333.

Hansen, B. E., et al. (1996). 'Erratum: the likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP', *Journal of Applied econometrics*, 11(2): 195–198.

Hansen, P. R., and Horel, G. (2009). 'Quadratic variation by Markov chains', .

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). 'Realized garch: a joint model for returns and realized measures of volatility', *Journal of Applied Econometrics*, 27(6): 877–906.

Higham, N. J. (2008). *Functions of matrices: theory and computation*. Siam.

Kaufmann, S., and Scheicher, M. (1996). 'Markov-regime switching in economic variables: Part I. Modelling, estimating and testing.-Part II. A selective survey', Discussion paper, Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS).

Kim, C.-J. (1993). 'Unobserved-component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty', *Journal of Business & Economic Statistics*, 11(3): 341–349.

——— (1994). 'Dynamic linear models with Markov-switching', *Journal of Econometrics*, 60(1): 1–22.

Klaassen, F. (2002). 'Improving GARCH volatility forecasts with regime-switching GARCH', in *Advances in Markov-Switching Models*, pp. 223–254. Springer.

Lamoureux, C. G., and Lastrapes, W. D. (1990). 'Persistence in variance, structural change, and the GARCH model', *Journal of Business & Economic Statistics*, 8(2): 225–234.

Laurent, S., Rombouts, J. V., and Violante, F. (2013). 'On loss functions and ranking forecasting performances of multivariate volatility models', *Journal of Econometrics*, 173(1): 1–10.

McCracken, M. W. (2007). 'Asymptotics for out of sample tests of Granger causality', *Journal of Econometrics*, 140(2): 719–752.

Mezrich, J., and Engle, R. (1996). 'GARCH for groups', *Risk*, 9(8): 36–40.

Newey, W. K., and McFadden, D. (1994). 'Large sample estimation and hypothesis testing', *Handbook of econometrics*, 4: 2111–2245.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). 'Multivariate high-frequency-based volatility (HEAVY) models', *Journal of Applied Econometrics*, 27(6): 907–933.

——— (2014). 'Multivariate rotated ARCH models', *Journal of Econometrics*, 179(1): 16–30.

Patton, A. J. (2011). 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics*, 160(1): 246–256.

Patton, A. J., and Sheppard, K. (2009). 'Evaluating volatility and correlation forecasts', in *Handbook of financial time series*, pp. 801–838. Springer.

Pedersen, R. S., and Rahbek, A. (2014). 'Multivariate variance targeting in the BEKK–GARCH model', *The Econometrics Journal*, 17(1): 24–55.

Schaller, H., and Norden, S. V. (1997). 'Regime switching in stock market returns', *Applied Financial Economics*, 7(2): 177–191.

Shephard, N., and Sheppard, K. (2010). 'Realising the future: forecasting with high-frequency-based volatility (HEAVY) models', *Journal of Applied Econometrics*, 25(2): 197–231.

Sims, C. A. (1980). 'Macroeconomics and reality', *Econometrica: Journal of the Econometric Society*, pp. 1–48.

Stock, J. H., and Watson, M. W. (1996). 'Evidence on structural instability in macroeconomic time series relations', *Journal of Business & Economic Statistics*, 14(1): 11–30.

Tibshirani, R. (1996). 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

West, K. D. (1996). 'Asymptotic inference about predictive ability', *Econometrica: Journal of the Econometric Society*, pp. 1067–1084.

Zou, H. (2006). 'The adaptive lasso and its oracle properties', *Journal of the American statistical association*, 101(476): 1418–1429.

Zou, H., Hastie, T., Tibshirani, R., et al. (2007). 'On the "degrees of freedom" of the lasso', *The Annals of Statistics*, 35(5): 2173–2192.

APPENDIX

Appendix A

Graphs of Realized Measure of Stocks

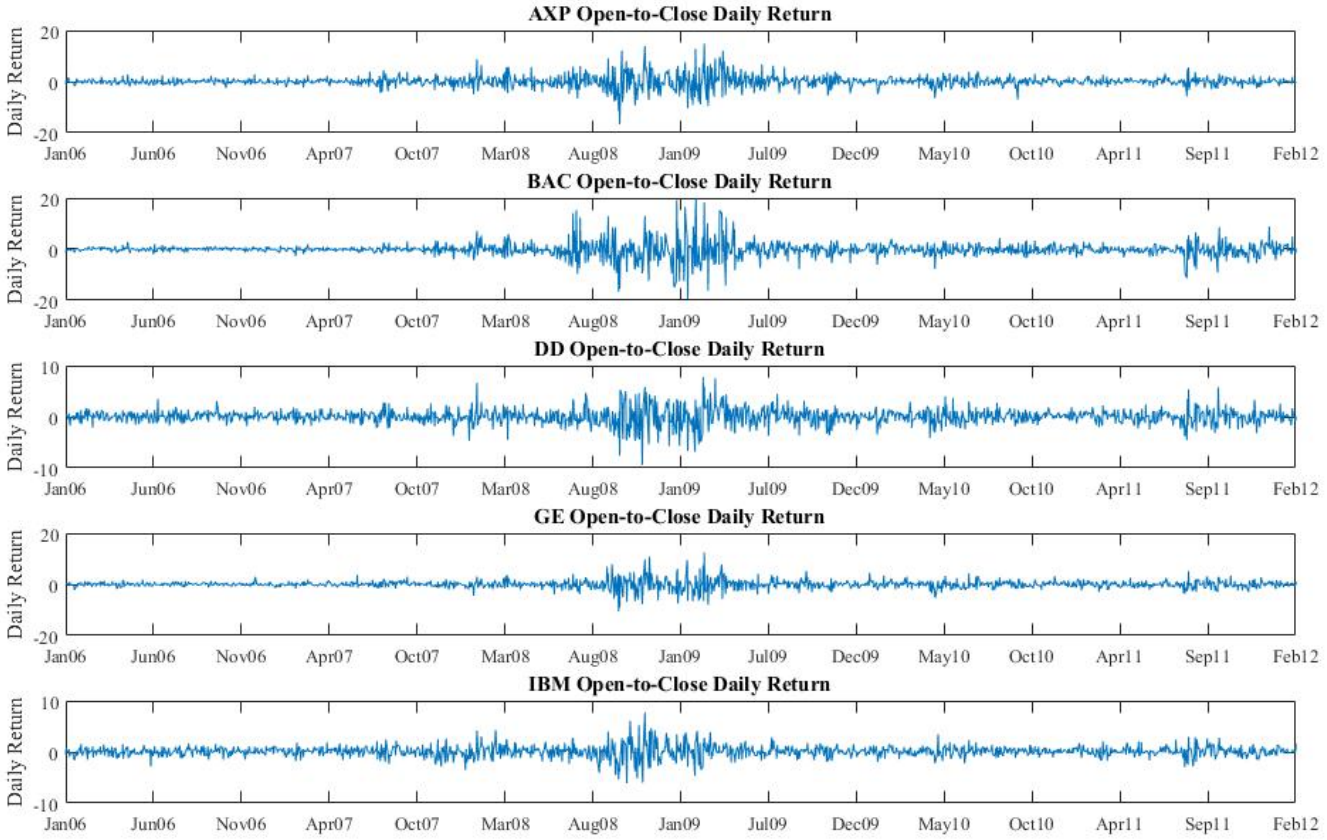

Figure A.1: The open-to-close log daily return of AXP,BAC,DD,GE,IBM.

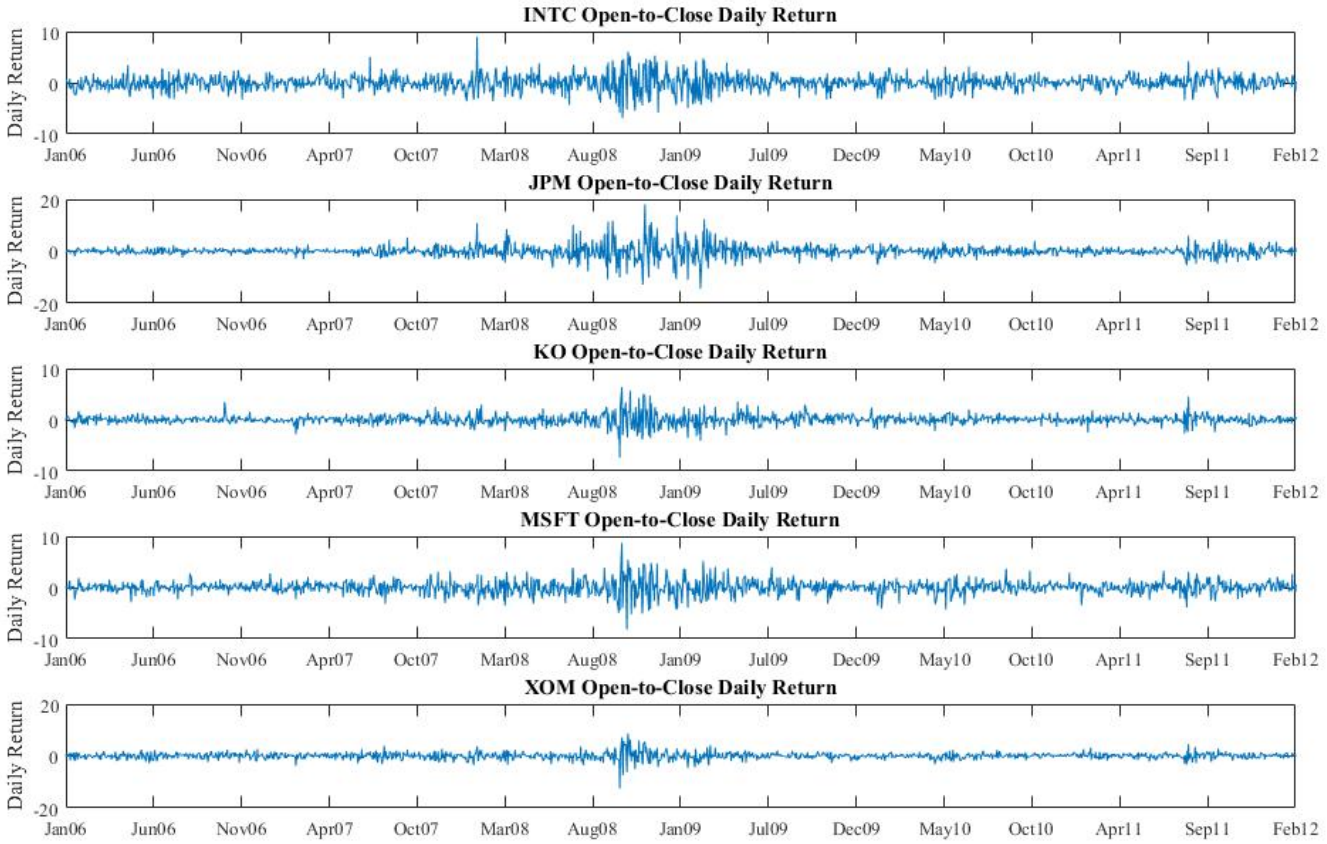

Figure A.2: The open-to-close log daily return of INTC,JPM,KO,MSFT,XOM.

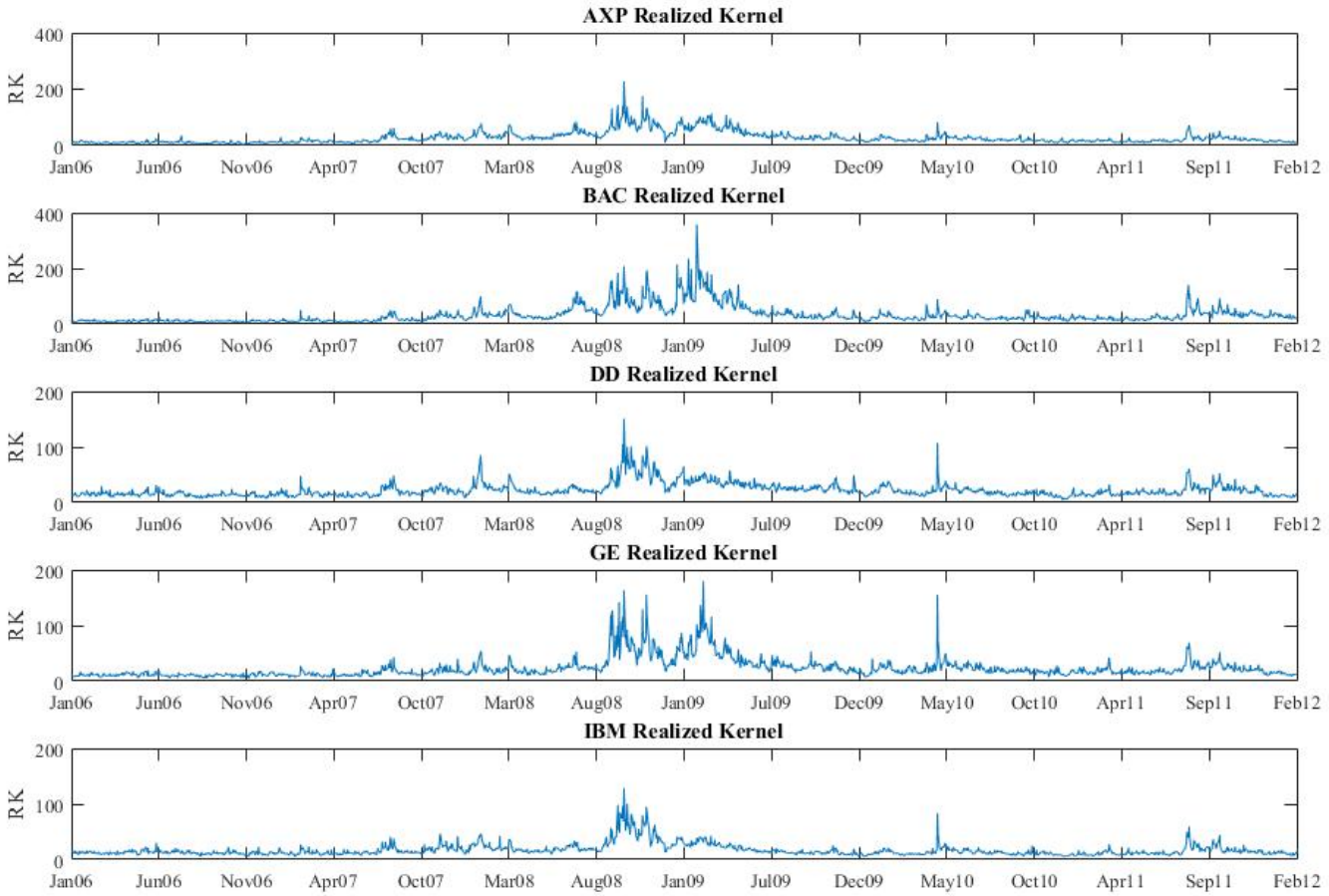

Figure A.3: The annualized realized volatility of AXP,BAC,DD,GE,IBM.

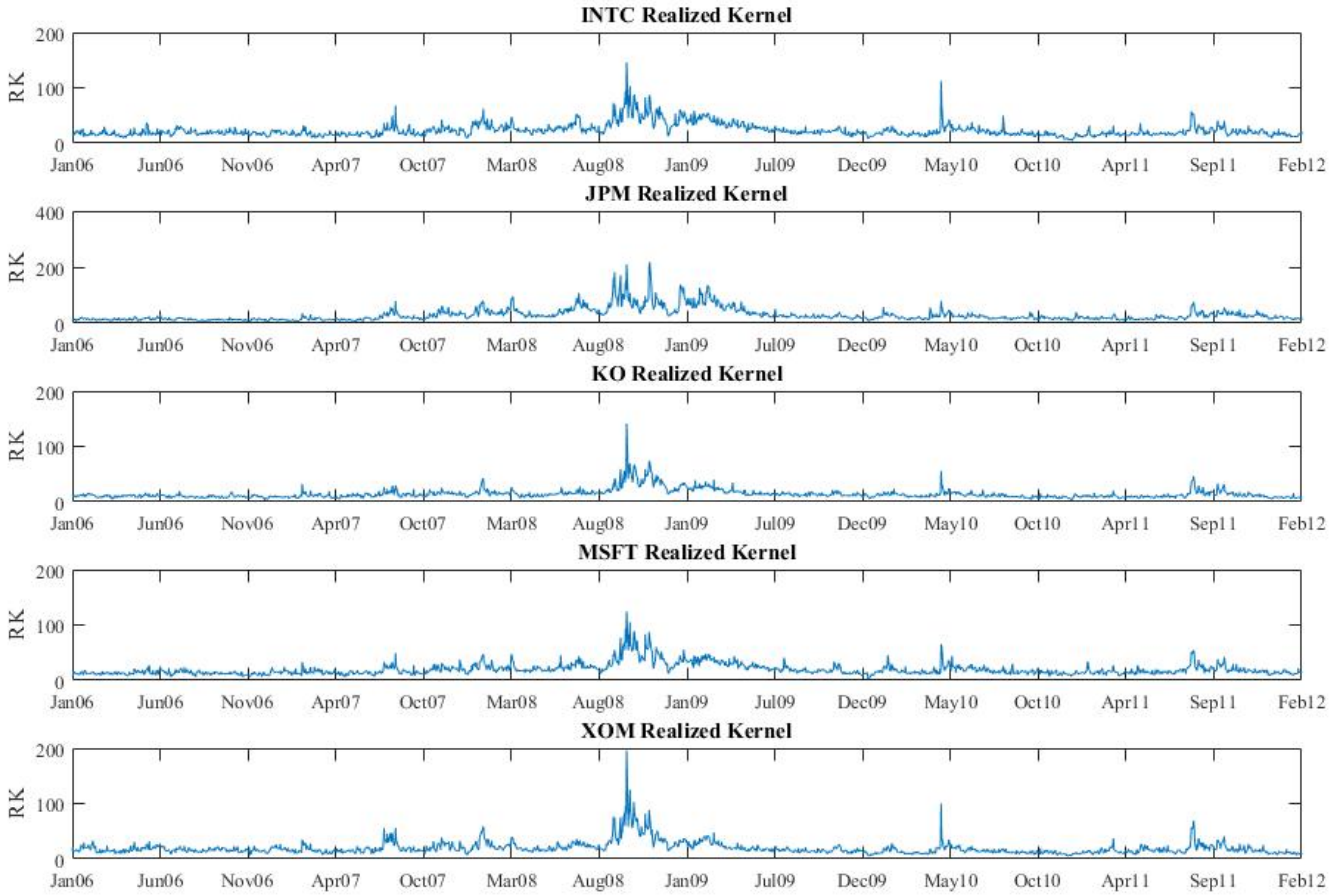

Figure A.4: The annualized realized volatility of INTC,JPM,KO,MSFT,XOM.