

Chen, Shi; Chen, Cathy Yi-Hsuan; Härdle, Wolfgang Karl; Lee, TM; Ong, Bobby

Working Paper

A first econometric analysis of the CRIX family

SFB 649 Discussion Paper, No. 2016-031

Provided in Cooperation with:

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Chen, Shi; Chen, Cathy Yi-Hsuan; Härdle, Wolfgang Karl; Lee, TM; Ong, Bobby (2016) : A first econometric analysis of the CRIX family, SFB 649 Discussion Paper, No. 2016-031, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/148867>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

SFB 649 Discussion Paper 2016-031

A first econometric analysis of the CRIX family

Shi Chen*
Cathy Yi-Hsuan Chen*
Wolfgang Karl Härdle*
TM Lee*²
Bobby Ong*²



* Humboldt-Universität zu Berlin,, Germany
*² CoinGecko, Singapore

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

A first econometric analysis of the CRIX family

Shi Chen¹, Cathy Yi-Hsuan Chen², Wolfgang Karl Härdle³,

TM Lee⁴, Bobby Ong⁵

August 30, 2016

¹Corresponding author. Humboldt-Universität zu Berlin, C.A.S.E.-Center of Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. Email: chenshiq@hu-berlin.de

²Humboldt-Universität zu Berlin, C.A.S.E.-Center of Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. Chung Hua University, department of Finance, 707 Sec.2 WuFu Rd., Hsinchu, Taiwan. Email: cathy1107@gmail.com

³Humboldt-Universität zu Berlin, C.A.S.E.-Center of Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. Email: haerdle@hu-berlin.de. Research fellow in Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, 6th Level, School of Economics, Singapore 178903.

⁴CoinGecko, 101 Upper Cross Street, No. 05-16 People's Park Centre, Singapore 058357. Email: tmlee@coingecko.com

⁵CoinGecko, 101 Upper Cross Street, No. 05-16 People's Park Centre, Singapore 058357. Email: bobby@coingecko.com

Contents

1	A first econometric analysis of the CRIX family	2
1.1	Econometric Review of CRIX	3
1.1.1	Introductionary Remarks	3
1.1.2	Statistical Analysis of CRIX Returns	8
1.2	ARIMA Models	11
1.2.1	Box-Jenkins Procedure	12
1.2.2	Lag Orders	12
1.2.3	ARIMA Model Estimation	14
1.3	Model with Stochastic Volatility	18
1.3.1	ARCH Model	19
1.3.2	GARCH Model	20
1.3.3	Variants of the GARCH Models	24
1.4	Multivariate GARCH Model	28
1.4.1	Formulations of MGARCH Model	31
1.4.2	DCC Model Estimation	32
1.4.3	DCC Model Diagnostics	35
1.5	Nutshell and Outlook	35

Chapter 1

A first econometric analysis of the CRIX family

The CRIX (CRyptocurrency IndeX) has been constructed based on approximately 30 cryptos and captures high coverage of available market capitalisation. The CRIX index family covers a range of cryptos based on different liquidity rules and various model selection criteria. Details of ECRIX (Exact CRIX), EFCRIX (Exact Full CRIX) and also intraday CRIX movements may be found on the webpage of hu.berlin/crix.


In order to price contingent claims one needs to first understand the dynamics of these indices. Here we provide a first econometric analysis of the CRIX family within a time-series framework. The key steps of our analysis include model selection, estimation and testing. Linear dependence is removed by an ARIMA model, the diagnostic checking resulted in an ARIMA(2,0,2) model for the available sample period from Aug 1st, 2014 to April 6th, 2016. The model residuals showed the well known phenomenon of volatility clustering. Therefore a further refinement lead us to an ARIMA(2,0,2)- t -GARCH(1,1) process. This specification conveniently takes care of fat-tail properties that are typical for financial markets. The multivariate GARCH models are implemented on the CRIX index family to explore the interaction. This chapter is practitioner oriented, four main questions are answered,


1. What's the dynamics of CRIX?
2. How to employ statistical methods to measure their changes over time?
3. How stable is the model used to estimate CRIX?
4. What do empirical findings imply for the econometric model?

A large literature can be reached for further study, for instance, Hamilton (1994), Franke et al. (2015), Box et al. (2015), Lütkepohl (2005), Rachev et al. (2007) etc. All numerical procedures are transparent and reproduced on www.quantlet.de.

1.1 Econometric Review of CRIX

1.1.1 Introductory Remarks

The CRyptocurrency IndeX  developed by Härdle and Trimborn (2015) is aimed to provide a market measure which consists of a selection of representative cryptos. The index fulfills the requirement of having a dynamic structure by relying on statistical time series techniques. The following table 1.1 are the 30 cryptocurrencies used in the construction of CRIX index.

The Research Data Center  supported by Collaborative Research Center (CRC) 649 provides access to the dataset. At time of writing, Bitcoins market capitalization as a percentage of CRIX total market capitalization is 83%.

No.	Cryptos	Symbol	Description
1	Bitcoin	BTC	Bitcoin is the first cryptocurrency. It was created by the anonymous person(s) named Satoshi Nakamoto in 2009 and has a limited supply of 21 million coins. It uses the SHA-256 Proof-of-Work hashing algorithm.
2	Ethereum	ETH	Ethereum is a Turing-completed cryptocurrency platform created by Vitalik Buterin. It raised US\$18 million worth of bitcoins during a crowdsale of ether tokens in 2014. Ethereum allows for token creation and smart contracts to be written on top of the platform. The DAO (No.30) and DigixDAO (No.15) are two tokens created on the Ethereum platform that is also used in the construction of CRIX.
3	Steem	STEEM	Steem is a social-media platform that rewards users for participation with tokens. Users can earn tokens by creating and curating content. The Steem whitepaper was co-authored by Daniel Larimer who is also the founder of BitShares (No.16).

4	Ripple	XRP	Ripple is a payment system created by Ripple Labs in San Francisco. It allows for banks worldwide to transact with each other without the need of a central correspondent. Banks such as Santander and UniCredit have begun experimenting on the Ripple platform. It was one of the earliest altcoin in the market and is not a copy of Bitcoin's source code.
5	Litecoin	LTC	Litecoin is branded the "silver to bitcoin's gold". It was created by Charles Lee, an ex-employee of Google and current employee of Coinbase. Charles modified Bitcoin's source code and made use of the Script Proof-of-Work hashing algorithm. There is a total of 84 million litecoin with a block time of 2.5 minutes. Initial reward was 50 LTC per block with rewards halving every 840,000 blocks.
6	NEM	NEM	NEM, short for New Economy Movement is a cryptocurrency platform launched in 2015 that is written from scratch on the Java platform. It provides many services on top of payments such as messaging, asset making and naming system.
7	Dash	DASH	Dash (previously known as Darkcoin and XCoin) is a privacy-centric cryptocurrency. It anonymizes transactions using PrivateSend (previously known as DarkSend), a concept that extends the idea of CoinJoin. PrivateSend achieves obfuscation by combining bitcoin transactions with another persons transactions using common denominations of 0.1DASH, 1DASH, 10DASH and 100DASH.
8	MaidSafeCoin	MAID	MaidSafeCoin is the cryptocurrency for the SAFE (Secure Access For Everyone) network. The network aims to do away with third-party central servers in order to enable privacy and anonymity for Internet users. It allows users to earn tokens by sharing their computing resources (storage space, CPU, bandwidth) with the network. MaidSafeCoin was released on the Omni Layer.

9	Lisk	LSK	Lisk is a Javascript platform for the creation of decentralized applications (DApps) and sidechains. Javascript was chosen because it is the most popular programming language on Github. It was created by Olivier Beddows and Max Kordek who were actively involved in the Crypti altcoin before this. Lisk conducted a crowdsale in early 2016 that raised about US\$6.15 million.
10	Dogecoin	DOGE	Dogecoin was created by Jackson Palmer and Billy Markus. It is based on the "doge", an Internet meme based on a Shiba Inu dog. Both the founders created Dogecoin for it to be fun so that it can appeal to a larger group of people beyond the core Bitcoin audience. Dogecoin found a niche as a tipping platform on Twitter and Reddit. It was merged-mined with Litecoin (No.5) on 11 September 2014.
11	NXT	NXT	NXT is the first 100% Proof-of-Stake cryptocurrency. It is a cryptocurrency platform that allows for the creation of tokens, messaging, domain name system and marketplace. There is a total of 1 billion coins created and it has a block time of 1 minute.
12	Monero	XMR	Monero is another privacy-centric altcoin that aims to anonymize transactions. It is based on the Cryptonote protocol which uses Ring Signatures to conceal sender identities. Many users, including the sender will sign a transaction thereby making it very difficult to trace the true sender of a transaction.
13	Synereo	AMP	Synereo is a decentralized and distributed social network service. It conducted its crowdsale in March 2015 on the Omni Layer where 18.5% of its tokens were sold.
14	Emercoin	EMC	Emercoin provides a key-value storage system, which allows for a Domain Name System (DNS) for .coin, .emc, .lib and .bazar domain extensions. It is inspired by Namecoin (No.26) DNS system which uses the .bit domain extension. It uses a Proof-of-Work/Proof-of-Stake hashing algorithm and allows for a maximum name length of 512.

15	DigixDAO	DGO	DigixDAO is a gold-backed token on the Etheruem (No.2) platform. Each token represents 1 gram of gold and each token is divisible to 0.001 gram. The tokens on the Ethereum platform are audited to ensure that the said amount of gold is held in reserves in Singapore.
16	BitShares	BTS	BitShares is a cryptocurrency platform that allows for many features such as a decentralized asset exchange, user-issued assets, price-stable cryptocurrencies, stakeholder approved project funding and transferable named accounts. It uses a Delegated Proof-of-Stake consensus algorithm.
17	Factom	FCT	Factom allows businesses and governments to record data on the Bitcoin blockchain. It does this by hashing entries before adding it onto a list. The entries can be viewed but not modified thus ensuring integrity of data records.
18	Siacoin	SC	Sia is a decentralized cloud storage platform where users can rent storage space from each other. The data is encrypted into many pieces and uploaded to different hosts for storage.
19	Stellar	STR	Stellar was created by Jed McCaleb, who was also the founder of Ripple (No.4) and Mt. Gox, the previously-largest bitcoin exchange which is now bankrupt. Stellar was created using a forked source code of Ripple. Stellar's mission is to expand financial access and literacy worldwide.
20	Bytecoin	BCN	Bytecoin is a privacy-centric cryptocurrency and is the first cryptocurrency created with the CryptoNote protocol. Its codebase is not a fork of Bitcoins.

21	Peercoin	PPC	Peercoin (previously known as PPCoin) was created by Sunny King. It was the first implementation of Proof-of-Stake. It uses a hybrid Proof-of-Work/Proof-of-Stake system. Proof-of-Stake is more efficient as it does not require any mining equipments to create blocks. Block creation is done via holding stake in the coin and therefore resistant to 51% mining attacks.
22	Tether	USDT	ether is backed 1-to-1 with traditional US Dollar in reserves so $1USDT = 1USD$. It is digital tokens formatted to work seamlessly on the Bitcoin blockchain. It exists as tokens on the Omni protocol.
23	Counterparty	XCP	Counterparty is the first cryptocurrency to make use of Proof-of-Burn as a method to distribute tokens. Proof-of-Burn works by having users send bitcoins to an unspendable address, in this case: $1CounterpartyXXXXXXXXXXXXXXXXXXXXXXXXXXXXWLpVr$. A total of 2,125 BTC were burnt in this manner, creating 2.6 million XCP tokens. The Proof-of-Burn method ensures that the Counterparty developers do not enjoy any privilege and allows for fair distribution of tokens. Counterparty is based on the Bitcoin platform and allows for creation of assets such as Storjcoin X (No.25).
24	Agoras	AGRS	Agoras is an application and smart currency market built on the Tau-Chain to feature intelligent personal agents, programming market, computational power market, and a futuristic search engine.
25	Storjcoin X	SJCX	Storjcoin X is used as a token to exchange cloud storage and bandwidth access. Users can obtain Storjcoin X by renting out resources to the network via DriveMiner and they will be able to rent space from other users by paying Storjcoin X using Metadisk. Storjcoin X is an asset created on the Counterparty platform (No.23).

26	Namecoin	NMC	Namecoin is one of the earliest altcoin that has been adapted from Bitcoins source code to allow for a different use case. It provides a decentralised key-value system that allows for the creation of an alternative Domain Name System that cannot be censored by governments. It uses the .bit domain extension. It was merge-mined with Bitcoin from September 2011.
27	Ybcoin	YBC	Ybcoin is a cryptocurrency from China that was created in June 2013. It uses the Proof-of-Stake hashing algorithm.
28	Nautiluscoin	NAUT	Nautiluscoin uses DigiShield difficulty retargeting system to safeguard against multi-pool miners. It has a Nautiluscoin Stabilization Fund (NSF) to reduce price volatility.
29	Fedoracoin	TIPS	Fedoracoin is based on the Tips Fedora Internet meme. Fedoracoin is also used as a tipping cryptocurrency.
30	The DAO	DAO	The DAO, short for Distributed Autonomous Organization ran one of the most successful crowdfunding campaign when it raised over US\$160 million. The DAO is a smart contract written on the Ethereum (No.2) platform. The DAO grants token holders voting rights to make decision in the organization based on proportion of tokens owned. In June 2016, a hack occurred resulting in the loss of about US\$60 million. The Ethereum Foundation decided the reverse the hack by conducting a hardfork of the Ethereum platform.

Table 1.1: 30 cryptocurrencies used in construction of CRIX.

1.1.2 Statistical Analysis of CRIX Returns

In the crypto market, the CRIX index was designed as a sample drawn from the pool of cryptos to represent the market performance of leading currencies. In order for an index to work as an investment benchmark, in this section we first focus on the stochastic properties of CRIX. The plots are often the first step in an exploratory analysis. Figure 1.1 shows the daily values from 01/08/2014 to



Figure 1.1: CRIX Daily Price from Aug 1st, 2014 to April 6th, 2016

 econ_crix

06/04/2016. We can observe that the values of CRIX fell down substantially until the mid of 2015, CRIX did poorly, perhaps as a result of the cool off of the cryptocurrency. After a few months moving up and down, the CRIX was, however, sloped up till now as a better year for crypto market. It is worthwhile to note here that the CRIX index were largely impacted and/or influenced by the crypto market, therefore, makes it a better indicator for the market performance.

To find out the dynamics of CRIX, we would first look closer to stationary time series. A stationary time series is one whose stochastic properties such as mean, variance etc are all constant over time. Most statistical forecasting methods are based on the stationary assumption, however the CRIX is far from stationary as observed in Figure 1.1. Therefore we need first to transform the original data into stationary time series through the use of mathematical transformations. Such transformations includes detrending, seasonal adjustment and etc, the most general class of models amongst them is ARIMA fitting, which will be explained in next section 1.2.

In practice, the difference between consecutive observations was generally computed to make a time series stationary. Such transformations can help stabilize the mean by removing the changes in the levels of a time series, therefore removing the trend and seasonality. Here the log returns of CRIX are computed

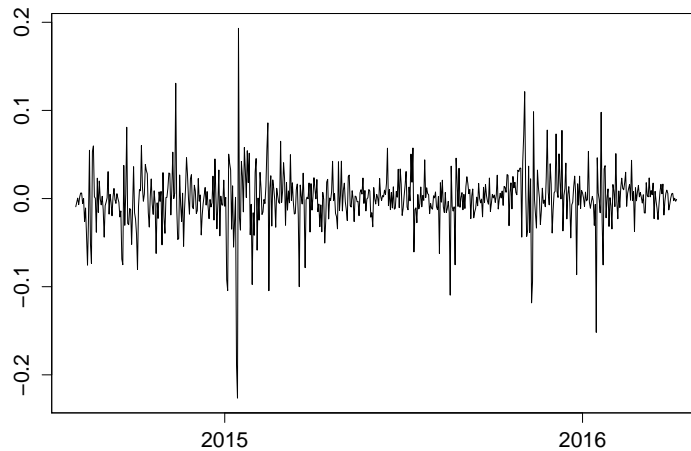


Figure 1.2: The log returns of CRIX index from Aug 2th, 2014 to April 6th, 2016

 econ_crix

for further analysis, we remove the unequal variances using the log of the data and take difference to get rid of the trend component. Figure 1.2 shows the time series plot of daily log returns of the CRIX index (henceafter CRIX returns), with the mean is -0.0004 and volatility is 0.0325 .

We continue to investigate distributional properties. We have the histogram of CRIX returns plotted in the left panel of Figure 1.3, compared with the normal density function plotted in blue. The right panel is QQ plot of CRIX daily returns. We can conclude that the CRIX returns is not normal distributed. Another approach widely used in density estimation is kernel density estimation. Furthermore, there are various methods to test if sample follows a specific distribution, for example Kolmogorov-Smirnoff test and Shapiro-Test.

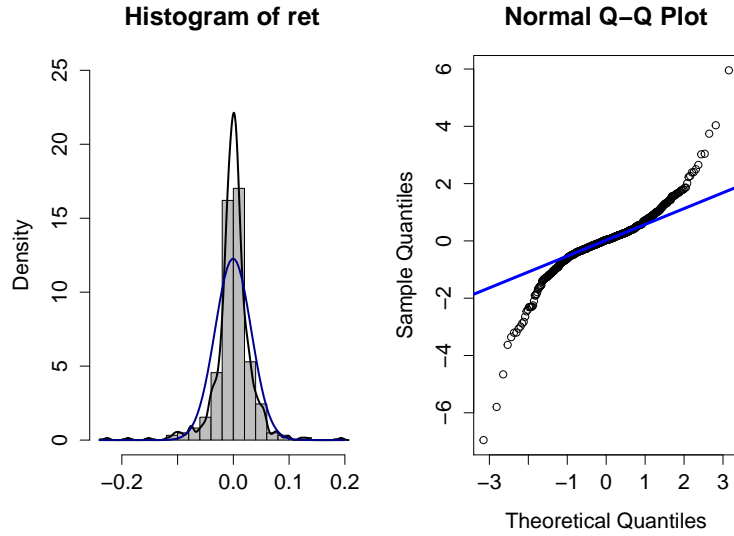


Figure 1.3: Histogram and QQ plot of CRIX returns.



1.2 ARIMA Models

The ARIMA(p, d, q) model with p standing for the lag order of the autoregressive model, d is the degree of differencing and q is the lag order of the moving average model, is given by (for $d = 1$)

$$\begin{aligned} \Delta y_t &= a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \dots + a_p \Delta y_{t-p} \\ &+ \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q} \end{aligned} \quad (1.1)$$

or

$$a(L)\Delta y_t = b_L \varepsilon_t \quad (1.2)$$

where $\Delta y_t = y_t - y_{t-1}$ is the differenced series and can be replaced by higher order differencing $\Delta^d y_t$ if necessary. L is the lag operator and $\varepsilon_t \sim N(0, \sigma^2)$.

There are two approaches to identify and fit an appropriate ARIMA(p, d, q) model. The first one is the Box-Jenkins procedure (subsection 1.2.1), another one to select models is selection criteria like Akaike information criterion (AIC) and Bayesian or Schwartz Information criterion (BIC), see subsection 1.2.2.

1.2.1 Box-Jenkins Procedure

The Box-Jenkins procedure comprises the following stages:

1. Identification of lag orders p, d and q .
2. Parameter estimation
3. Diagnostic checking

A detailed illustration of each stages can be found in the textbook of Box et al. (2015).

In the first identification stage, one needs first to determine the degree of integration d . Figure 1.2 shows that the CRIX returns are generally stationary over time. As well as looking at the time plot, the sample autocorrelation function (ACF) is also useful for identifying the non-stationary time series. The values of ACF will drop to zero relatively quickly compared to the non-stationary case. Furthermore, the unit root tests can be more objectively to determine if differencing is required. For instance, the augmented Dickey-Fuller (ADF) test and KPSS test, see Dickey and Fuller (1981) and Kwiatkowski et al. (1992) for more technical details.

Given d , one identifies the lag orders (p, q) by checking ACF plots to find the total correlation between different lag functions. In an MA context, there is no autocorrelation between y_t and y_{t-q-1} , the ACF dies out at q . A second insight one obtain is from the partial autocorrelation function (PACF). For an AR(p) process, when the effects of the lags $y_{t-1}, y_{t-2}, \dots, y_{t-p-1}$ are excluded, the autocorrelation between y_t and y_{t-p} is zero. Hence an PACF plot for $p = 1$ will drop at lag 1.

1.2.2 Lag Orders

We exhibit the discussion thus far by analyzing the daily log return of CRIX introduced in subsection 1.1.2. The stationarity of the return series is tested by ADF (null hypothesis: unit root) and KPSS (null hypothesis: stationary) tests. The p -values are 0.01 for ADF test, 0.1 for KPSS test. Hence one concludes stationarity on the level $d = 0$.

The next step is to choose the lag orders of p and q for the ARIMA model. The sample ACF and PACF are calculated and depicted in Figure 1.4, with blue dashed lines as 95% limits. The results suggest that the CRIX log returns are not random. The Ljung-Box test statistic for examining the null hypothesis of

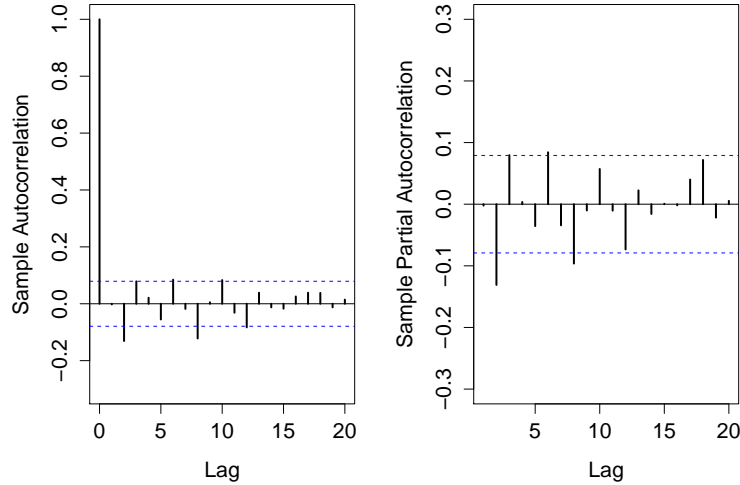


Figure 1.4: The sample ACF and PACF plots of daily CRIX returns from Aug 2th, 2014 to April 6th, 2016, with lags = 20.

 econ_arima

independence yields a p -value of 0.0017. Hence one rejects the null hypothesis and suggests that the CRIX return series has autocorrelation structure.

The ACF pattern in Figure 1.4 suggests that the existence of strong autocorrelations in lag 2 and 8, partial autocorrelation in lag 2, 6 and 8. These results suggest that the CRIX return series can be modeled by some ARIMA process, for example ARIMA(2, 0, 2).

In addition to ACF and PACF, several model selection criteria are widely used to overcome the problem of overparameterization. They are Akaike Information Criterion (AIC) from Akaike (1974) and Bayesian or Schwartz Information Criteria (BIC) from Schwarz et al. (1978), the formulas are given by,

$$AIC(\mathcal{M}) = -2 \log L(\mathcal{M}) + 2p(\mathcal{M}) \quad (1.3)$$

$$BIC(\mathcal{M}) = -2 \log L(\mathcal{M}) + p(\mathcal{M}) \log n \quad (1.4)$$

where n is the number of observations, $p(\mathcal{M})$ is the number of parameters in model \mathcal{M} and $L(\mathcal{M})$ represents the likelihood function of the parameters evaluated at the Maximum Likelihood Estimation (MLE).

The first terms $-2 \log L(\mathcal{M})$ in equation (1.3) and (1.4) reflect the goodness of fit for MLE, while the second terms stand for the model complexity. Therefore

AIC and BIC can be viewed as measures that combine fit and complexity. The main difference between two measures is the BIC is asymptotically consistent while AIC is not. Compared with BIC, AIC tends to overparameterize.

1.2.3 ARIMA Model Estimation

We start with ARIMA(1,0,1) as an example, fit the ARIMA(1,0,1) model derived from equation (1.1),

$$y_t = a_1 y_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1}$$

The estimated parameters are: $\hat{a}_1 = 0.5763$ with standard deviation of 0.5371, $\hat{b}_1 = -0.6116$ with standard deviation of 0.5205. y_t represents the CRIX returns.

In the third stage of Box-Jenkins procedure one evaluates the validity of the estimated model. The results of diagnostic checking is reported in the three diagnostic plots of Figure 1.5. The upper panel is the standardized residuals, the middle one is the ACF of residuals and the lower panel is the Ljung-Box test statistic for the null hypothesis of residual independence. One observes that the significant autocorrelations of the model residuals appear at lag of 2, 3, 6 and 8, and the low p -values of the Ljung-Box test statistic after lag 1. We cannot reject the null hypothesis at these lags, hence ARIMA(1,0,1) model is not the enough to get rid of the serial dependence. A more appropriate lag orders is needed for better model fitting.

Nevertheless, model diagnostic checking is often used together with model selection criteria. In practice, these two approaches complement each other. Based on the discussion results of Figure 1.4 in subsection 1.2.2, we select a combination of (p, d, q) with $d = \{0, 1\}$ and $p, q = \{0, 1, 2, 3, 4, 5\}$. A calculation of the AIC and BIC for each model find out the best six models listed in Table 1.2. In general, an ARIMA(2,0,2) model

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} \quad (1.5)$$

performs best. Its diagnostic plots are plotted in Figure 1.6 and look very good, the significant p -values of Ljung-Box test statistic suggest the independence structure of model residuals. Furthermore, the estimate of each element in equation (1.5) is reported in Table 1.3.

With the identified ARIMA model and its estimated parameters, we predict the CRIX returns for the next 30 days under the ARIMA(2,0,2) model. The out-of-sample prediction result is shown in Figure 1.7. The 95% confidence bands are computed using a rule of thumb of "prediction $\pm 2 * \text{standard deviation}$ ".

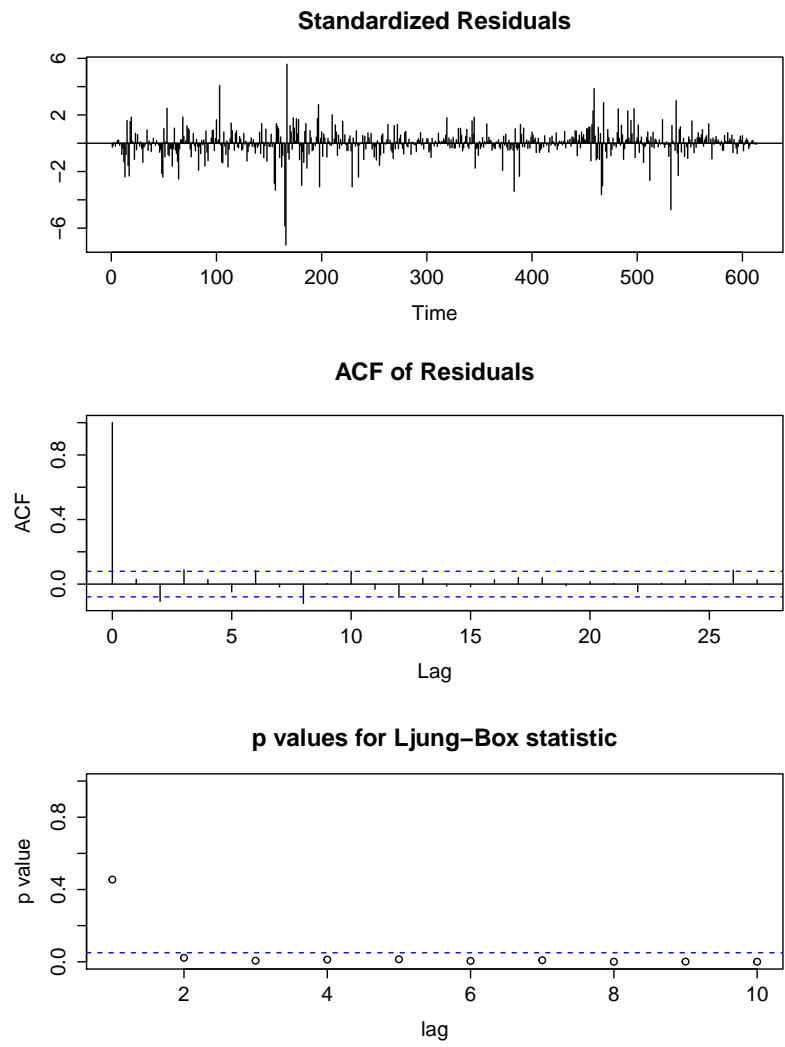



Figure 1.5: Diagnostic checking result of ARIMA(1,0,1).

 econ_arima

ARIMA model selected	AIC	BIC
ARIMA(2,0,0)	-2468.83	-2451.15
ARIMA(2,0,2)	-2474.25	-2447.73
ARIMA(2,0,3)	-2472.72	-2441.78
ARIMA(4,0,2)	-2476.35	-2440.99
ARIMA(2,1,1)	-2459.15	-2441.47
ARIMA(2,1,3)	-2464.14	-2437.62

Table 1.2: The ARIMA model selection with AIC and BIC.

 econ_arima

Coefficients	Estimate	Standard deviation
intercept c	-0.0004	0.0012
a_1	-0.6989	0.1124
a_2	-0.7508	0.1191
b_1	0.7024	0.1351
b_2	0.6426	0.1318
Log likelihood	1243.12	

Table 1.3: Estimation result of ARIMA(2,0,2) model.

 econ_arima

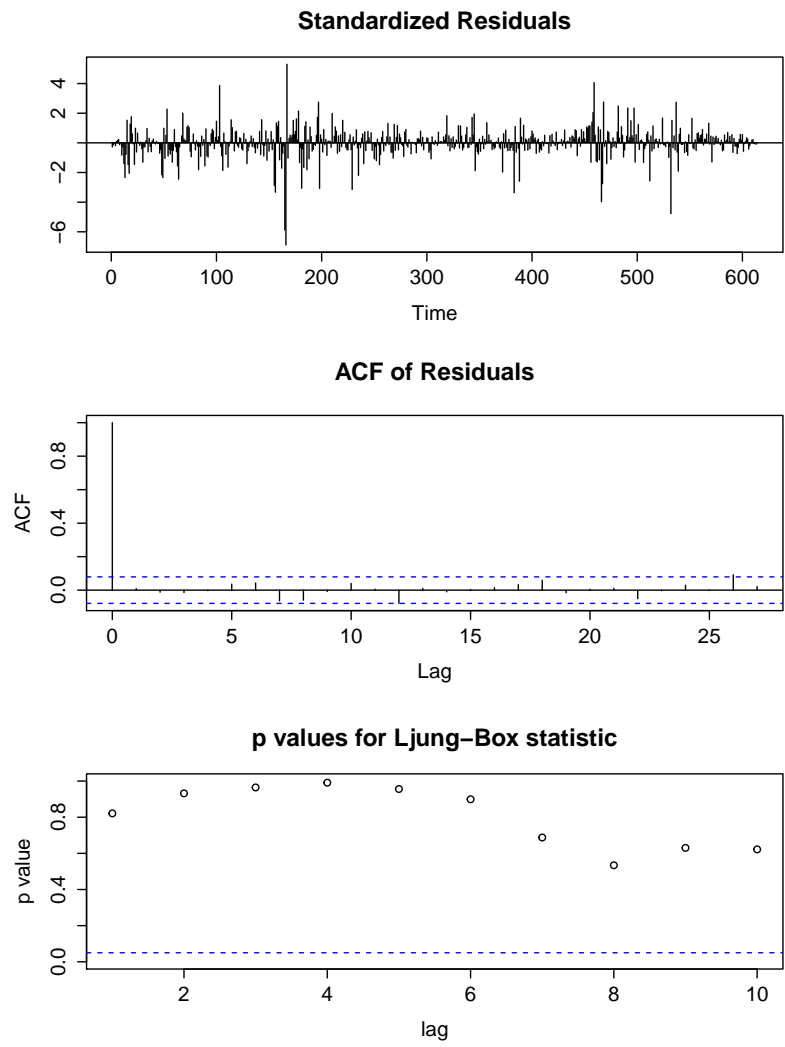


Figure 1.6: Diagnostic checking result of ARIMA(2,0,2).

 econ_arima

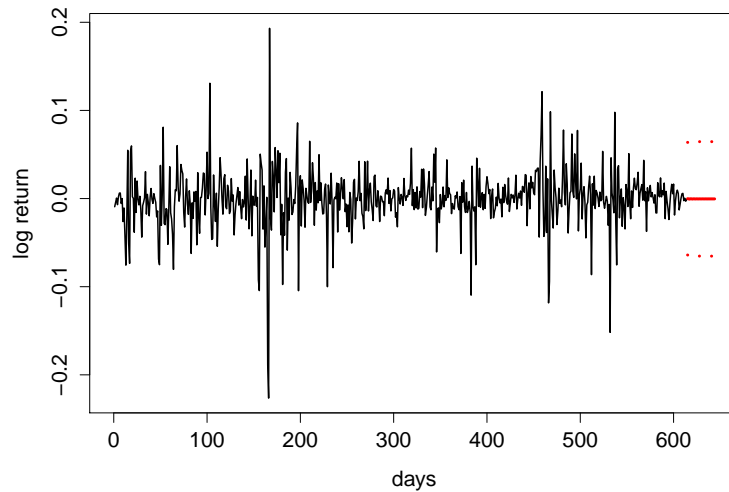


Figure 1.7: CRIX returns and predicted values. The confidence bands are red dashed lines.

 econ_arima

1.3 Model with Stochastic Volatility

Homoskedasticity is a frequently used assumption in the framework of time series analysis, that is, the variance of all squared error terms is assumed to be constant through time, see Brooks (2014). Nevertheless we can observe heteroskedasticity in many cases when the variances of the data are different over different periods.

In subsection 1.2.3 we have built an ARIMA model for the CRIX return series to model intertemporal dependence. Although the ACF of model residuals has no significant lags as evidenced by the large p -values for the Ljung-Box test in Figure 1.6, the time series plot of residuals shows some clusters of volatility. To be more specific, we display the squared residual plot of the selected ARIMA(2,0,2) model in Figure 1.8.

To incorporate the univariate heteroskedasticity, we first fit an ARCH (AutoRegressive Conditional Heteroskedasticity) model in subsection 1.3.1. In subsection 1.3.2, its generalization, the GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) model, provides even more flexible volatility pattern. In addition, a variety of extensions of the standard GARCH models will be explored in subsection 1.3.3.

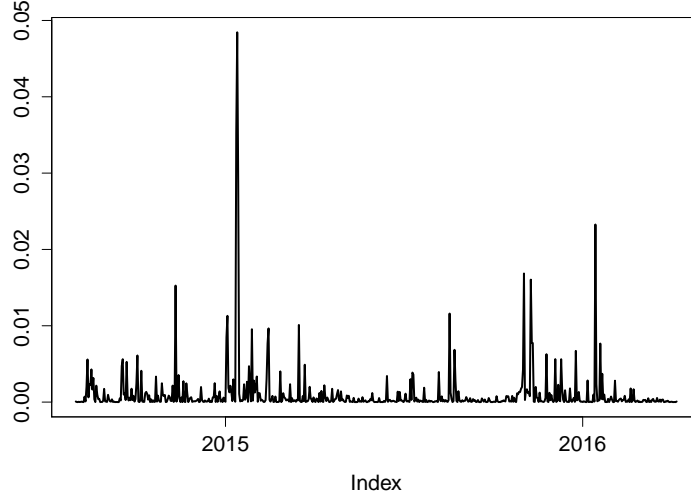


Figure 1.8: The squared ARIMA(2,0,2) residuals of CRIX returns.

 econ_vola

1.3.1 ARCH Model

The ARCH(q) model introduced by Engle (1982) is formulated as,

$$\begin{aligned}
 \varepsilon_t &= Z_t \sigma_t \\
 Z_t &\sim N(0, 1) \\
 \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2
 \end{aligned} \tag{1.6}$$

where ε_t is the model residual and σ_t^2 is the variance of ε_t conditional on the information available at time t . It should be noted that the parameters should satisfy $\alpha_i > 0, \forall i = 1, \dots, p$. The assumption of $\sum_i^p \alpha_i < 1$ is also imposed to assure the volatility term σ_t^2 is asymptotically stationary over time.

Based on the estimation results of subsection 1.2.3, we proceed to examine the heteroskedasticity effect observed in Figure 1.8. The model residual ε_t in equation (1.5) is used to test for ARCH effects using ARCH LM (Lagrange multiplier) test, the small p -value of $2.2e - 16$ cannot reject its null hypothesis of no ARCH effects. Another approach we can use is the Ljung-Box test for squared model residuals, see Tsay (2005). These two tests show similar result as the small p -value of Ljung-Box test statistic indicates the dependence structure of

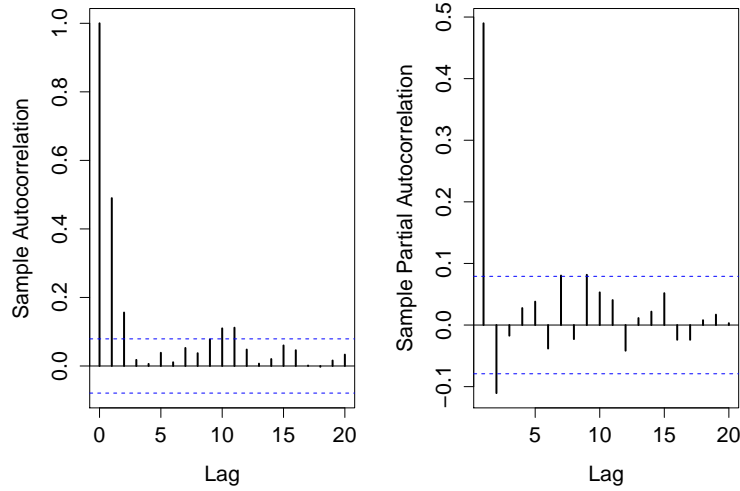



Figure 1.9: The ACF and PACF of squared residuals of ARIMA(2,0,2) model.  econ_vola

ε_t^2, \dots

To determine the lag orders of ARCH model, we display the ACF and PACF of squared residuals in Figure 1.9. The autocorrelations display a cutoff after the first two lags as well as some remaining lags are significant. The PACF plot in the right panel has a significant spike before lag 2. Therefore the lag orders of ARCH model should be at least 2.

We fit the ARCH models to the residuals using candidate values of q from 1 to 4, where all models are estimated by MLE based on the stochastic of equation (1.6). The results of model comparison are contained in Table 1.4. The Log likelihood and information criteria jointly select an ARCH(3) model, with the estimated parameters presented in Table 1.5. All the parameters except for the third one are significant at the 0.1% level.

1.3.2 GARCH Model

Bollerslev (1986) further extended ARCH model by adding the conditional heteroskedasticity moving average items in equation (1.6), the GARCH model indicates that the current volatility depends on past volatilities σ_{t-i}^2 and observa-

Model	Log Likelihood	AIC	BIC
ARCH(1)	1281.7	-2567.4	-2558.6
ARCH(2)	1283.4	-2560.8	-2547.6
ARCH(3)	1291.6	-2575.2	-2557.5
ARCH(4)	1288.8	-2567.5	-2545.4

Table 1.4: Estimation result of ARIMA-ARCH models.

 econ_arch

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	0.001	0.000	16.798*
α_1	0.195	0.042	4.589*
α_2	0.054	0.037	1.469
α_3	0.238	0.029	8.088*

Table 1.5: Estimation result of ARIMA(2,0,2)-ARCH(3) model, with significant level is 0.1%.

 econ_arch

GARCH models	Log likelihood	AIC	BIC
GARCH(1,1)	1305.355	-4.239	-4.210
GARCH(1,2)	1309.363	-4.249	-4.213
GARCH(2,1)	1305.142	-4.235	-4.199
GARCH(2,2)	1309.363	-4.245	-4.202

Table 1.6: Comparison of GARCH model, orders up to $p = q = 2$.



tions of model residual ε_{t-j}^2 .

The standard GARCH(p, q) is written as,

$$\begin{aligned}
 \varepsilon_t &= Z_t \sigma_t \\
 Z_t &\sim N(0, 1) \\
 \sigma_t^2 &= \omega + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2
 \end{aligned} \tag{1.7}$$

with the condition that,

$$\omega > 0; \quad \alpha_i \geq 0, \beta_i \geq 0; \quad \sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j < 1 \tag{1.8}$$

The conditions in equation (1.8) ensure that the GARCH model is strictly stationary with finite variance. Normally up to GARCH(2,2) model is used in practice. Particularly, the orders of $p = q = 1$ is sufficient in most cases.

The comparison of different GARCH models is reported in Table 1.6, the selection of lag orders up to $p = q = 2$. It shows that a GARCH(1,2) model performs slightly better than the other ones through the comparison of Log Likelihood and information criteria. Using the GARCH(1,2) model as selected,

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 \tag{1.9}$$

We obtain the estimation results presented in Table 1.7. The conditions $\omega > 0$

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	$9.906e - 05$	$4.753e - 05$	2.084*
α_1	$1.654e - 01$	$3.719e - 02$	4.448***
β_1	$8.074e - 02$	$8.244e - 02$	0.979
β_2	$6.513e - 01$	$8.202e - 02$	7.940***

Table 1.7: Estimation result of ARIMA(2,0,2)-GARCH(1,2) model. * represents significant level of 5% and *** of 0.1%.

 econ_garch

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	$5.324e - 05$	$2.251e - 05$	2.365*
α_1	$1.204e - 01$	$2.785e - 02$	4.324***
β_1	$8.322e - 02$	$3.992e - 02$	20.847***

Table 1.8: Estimation result of ARIMA(2,0,2)-GARCH(1,1) model. * represents significant level of 5% and *** of 0.1%.

 econ_garch

and $\alpha_1 + \beta_1 + \beta_2 = 0.897 < 1$ are fulfilled to obtain a strictly stationary solution. However β_1 is not significant using from the Ljung-Box test statistic.

Aforementioned, GARCH(1,1) is sufficient in most cases, we proceed further to fit the model residuals of ARIMA to the GARCH(1,1) model and present the estimation result in Table 1.8. The GARCH(1,1) outperforms the ARCH(3) model with all the estimated parameters are significant. The estimated parameters $\omega > 0$ and $\alpha_1 + \beta_1 = 0.953 < 1$ fulfill the stationary condition as well. Although the model performance of GARCH(1,2) is better than GARCH(1,1), all parameters of GARCH(1,1) are significant. Since the level of $\sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j$ reveals the persistence of volatility, we know that the GARCH(1,1) is more persistent in volatility compared than GARCH(1,2). Therefore for simplicity, GARCH(1,1) is suggested for further analysis in CRIX dynamics.

We have the model residuals of ARMA-GARCH process plotted in Figure

1.10. Figure 1.11 displays the ACF and PACF plots for model residuals of

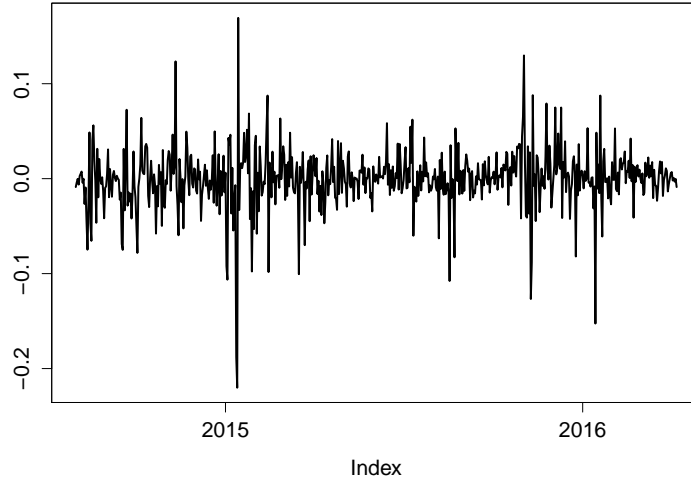


Figure 1.10: The ARIMA(2,0,2)-GARCH(1,1) residuals.

 econ_garch

ARIMA(2,0,2)-GARCH(1,1) process. We can see all the values are within the bands, which suggests that the model residuals have no dependence structure over different lags. Therefore GARCH(1,1) model is sufficient enough to explain the heteroskedasticity effect discussed in subsection 1.3.1.

1.3.3 Variants of the GARCH Models

As we observed in Figure 1.2, the return series of CRIX exhibits leptokurtosis. We further check the QQ-plot in Figure 1.12, which suggests the fat tail of model residuals using ARIMA(2,0,2)-GARCH(1,1) process. The Kolmogorov distance between residuals of the selected model and normal distribution is reported in Table 1.9. With the small p -value of Kolmogorov-Smirnov test statistic, we reject the null hypothesis that the model residuals are drawn from the normal distribution.

We impose the assumption on the residuals with student distribution, that is, applying the non-normal assumption on Z_t in equation (1.7). With $Z_t \sim t(d)$ to replace the normal assumption of Z_t in GARCH model, the MLE is implemented for model estimation. The results for ARIMA- t -GARCH process are

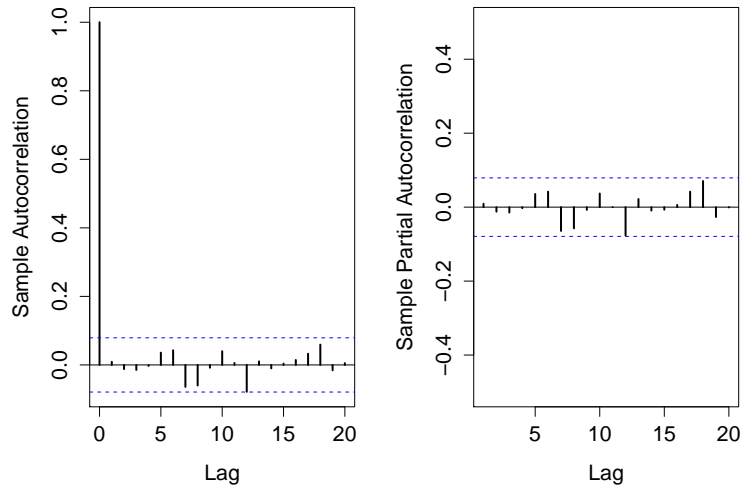


Figure 1.11: The ACF and PACF plots for model residuals of ARIMA(2,0,2)-GARCH(1,1) process.

 econ_garch

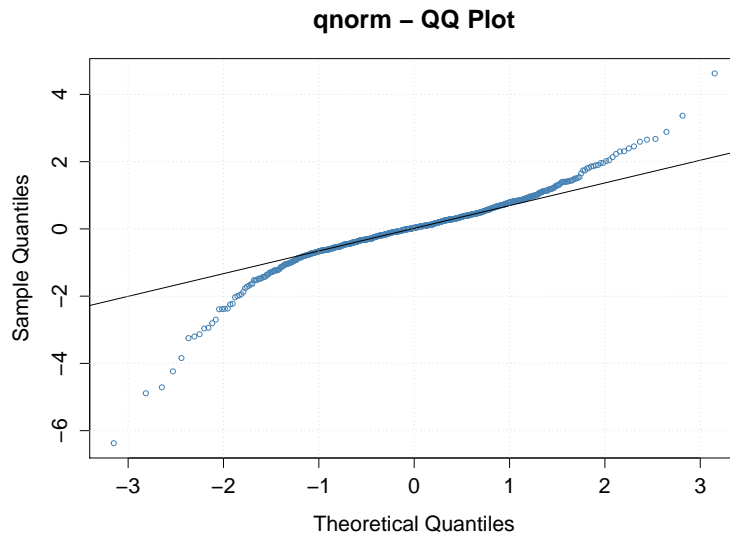



Figure 1.12: The QQ plots of model residuals of ARIMA-GARCH process.

 econ_garch

Model	Kolmogorov distance	P-value
ARIMA-GARCH	0.495	$2.861e - 10$

Table 1.9: Test of model residuals of ARIMA-GARCH process.

 econ_garch

Coefficients	Estimates	Standard deviation	t test
ω	$8.391e - 05$	$5.451e - 05$	1.539
α_1	$2.816e - 01$	$1.461e - 01$	1.928 [*]
β_1	$7.896e - 01$	$6.116e - 02$	12.910 ^{***}
ξ	$2.577e + 00$	$3.623e - 01$	7.113 ^{***}

Table 1.10: Estimation result of ARIMA(2,0,2)- t -GARCH(1,1) model. ^{*} represents significant level of 10% and ^{***} of 0.1%.

 econ_tgarch

represented in Table 1.10. The shape parameter ξ controls the height and fat-tail of density function, therefore different shape of distribution function. It is obvious that the shape parameter is significantly from zero. The QQ plot in Figure 1.13 indicates that the residuals are quite close to student- t distribution. The ACF and PACF plots for ARIMA- t -GARCH is following in Figure 1.14, with all values stay inside the bounds. Hence the residuals and their variance are uncorrelated.

In addition to the property of leptokurtosis, leverage effect is commonly observed in practice. According to a large literature, such as Engle and Ng (1993), the leverage effect refers to the volatility of an asset tends to respond asymmetrically with negative or positive shocks, declines in prices or returns are accompanied by larger increase in volatility compared with the decrease of volatility associated with rising asset market. Although the introduced GARCH model successfully solve the problem of volatility clustering, the σ_t^2 cannot capture the leverage effect.

To overcome this, the exponential GARCH (EGARCH) model with standard innovations proposed by Nelson (1991) can be expressed in the following nonlinear form,

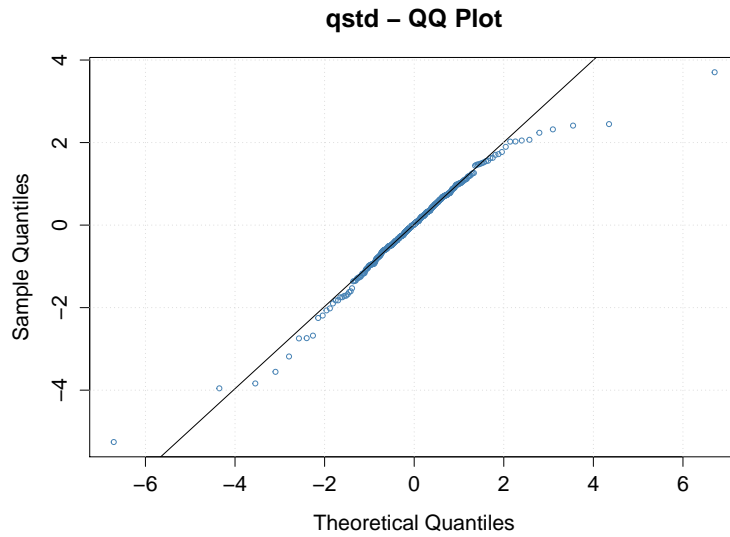


Figure 1.13: The QQ plot of t -GARCH(1,1) model.

 econ_tgarch

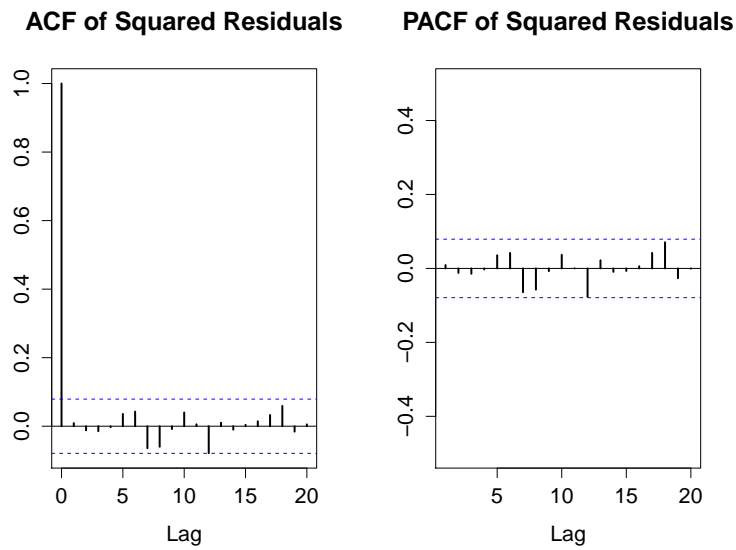



Figure 1.14: The ACF and PACF plots for model residuals of ARIMA(2,0,2)- t -GARCH(1,1) process.

 econ_tgarch

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	$9.906e - 05$	$4.753e - 05$	2.084*
α_1	$1.654e - 01$	$3.719e - 02$	4.448*
β_1	$8.074e - 02$	$8.244e - 02$	0.979
ϕ_1	$6.513e - 01$	$8.202e - 02$	7.940*

Table 1.11: Estimation result of ARIMA(2,0,2)- t -EGARCH(1,1) model. * represents significant level of 5% and *** of 0.1%.

 econ_tgarch

$$\begin{aligned}
\varepsilon_t &= Z_t \sigma_t \\
Z_t &\sim N(0, 1) \\
\log(\sigma_t^2) &= \omega + \sum_{i=1}^p \beta_i \log(\sigma_{t-i}^2) + \sum_{j=1}^q g_j(Z_{t-j}) \quad (1.10)
\end{aligned}$$

where $g_j(Z_t) = \alpha_j Z_t + \phi_j(|Z_{t-j}| - E|Z_{t-j}|)$ with $j = 1, 2, \dots, q$. When $\phi_j = 0$, we have the logarithmic GARCH (LGARCH) model from Geweke (1986) and Pantula (1986). However LGARCH is not popular due to the high value of the first few ACF of ε^2 .

Based on the results shown in Figure 1.12, we fit a EGARCH(1,1) model with student t distributed innovation term. The estimation results using the ARIMA(2,0,2)- t -EGARCH(1,1) model is reported in Table 1.11.

The ACF and PACF of ARIMA- t -EGARCH residuals are plotted in Figure 1.15. The small values indicate independent structure of model residuals. We further check the QQ plot in Figure 1.16, the model residuals fit better to student- t distribution compared with normal case of Figure 1.12.

We compare the model performance of selected GARCH models in Table 1.12, where the log likelihood and information criteria select the t -GARCH(1,1) model. With the selected ARIMA(2,0,2)- t -GARCH(1,1) model, we conduct a 30-step ahead forecast. The forecast performance is plotted in Figure 1.17 with the 95% confidence bands marked in blue.

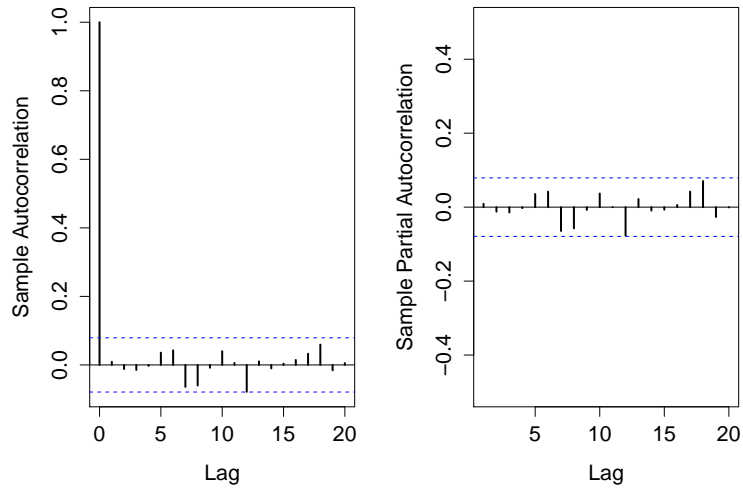



Figure 1.15: The ACF and PACF for model residuals of ARIMA- t -EGARCH process.

 econ_tgarch

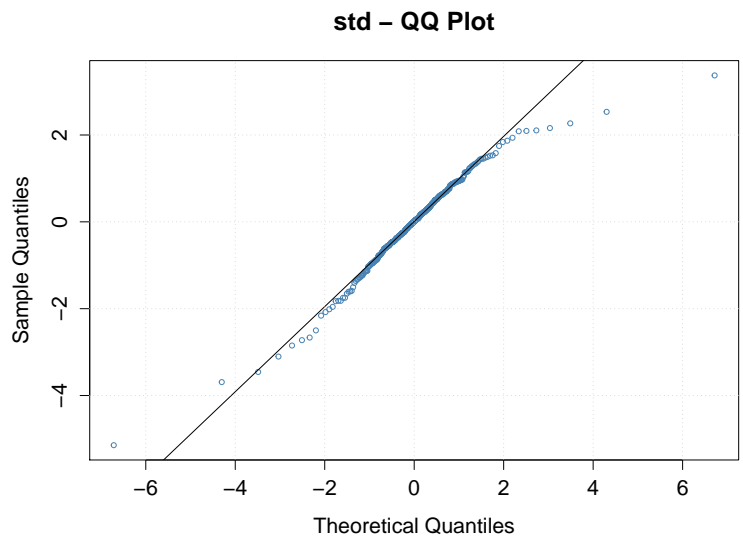



Figure 1.16: The QQ plot of t -EGARCH(1,1) model.

 econ_tgarch

GARCH models	Log likelihood	AIC	BIC
GARCH(1,1)	1305.355	-4.239	-4.210
<i>t</i> -GARCH(1,1)	1309.363	-4.249	-4.213
<i>t</i> -EGARCH(1,1)	1305.142	-4.235	-4.199

Table 1.12: Comparison of the variants of GARCH model.

 econ_tgarch

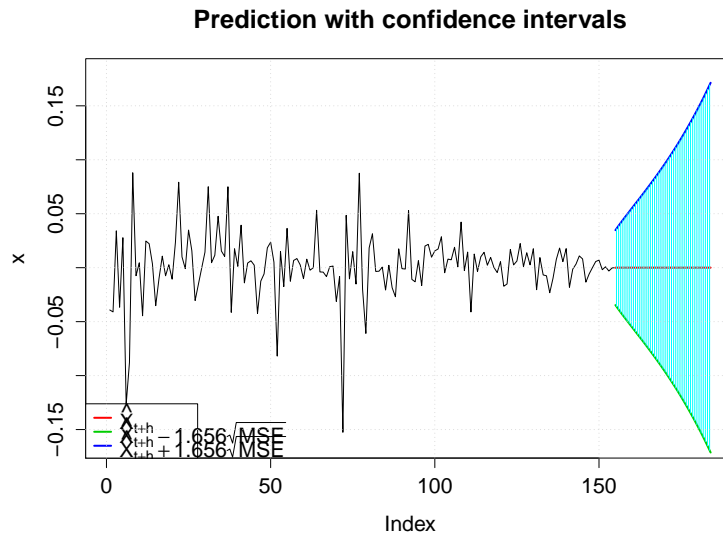



Figure 1.17: The 30-step ahead forecast using ARIMA-*t*-GARCH process.

 econ_tgarch

1.4 Multivariate GARCH Model

While modelling volatility of CRIX returns has been the main center of attention, understanding the co-movements of different indices in CRIX family are of great importance. In this subsection we proceed further to MGARCH (multivariate GARCH) model, whose model specification allows for a flexible dynamic structure. It provides us a tool to analyze the volatility and co-volatility dynamic of asset returns in a portfolio.

1.4.1 Formulations of MGARCH Model

Consider the error term ε_t with $\mathbf{E}(\varepsilon_t) = 0$ and the conditional covariance matrix given by the $(d \times d)$ positive definite matrix H_t , we assume that,

$$\varepsilon_t = H_t^{\frac{1}{2}} \eta_t \quad (1.11)$$

where $H_t^{\frac{1}{2}}$ can be obtained by Cholesky factorization of H_t . η_t is an iid innovation vector such that,

$$\begin{aligned} \mathbf{E}(\eta_t) &= 0 \\ \text{Var}(\eta_t) &= \mathbf{E}(\eta_t \eta_t^\top) = \mathcal{I}_d \end{aligned} \quad (1.12)$$

with \mathcal{I}_d is the identity matrix with order of d .

So far the standard MGARCH framework is defined, different specification of H_t yields various parametric formulations. The first MGARCH model was directly generalization of univariate GARCH model proposed by Bollerslev et al. (1988), which is called VEC model. Let $\text{vech}(\cdot)$ denotes an operator that stacks the columns of the lower triangular part of its argument square matrix. The VEC model is formulated as,

$$\text{vech}(H_t) = c + \sum_{j=1}^q A_j \text{vech}(\varepsilon_{t-j} \varepsilon_{t-j}^\top) + \sum_{i=1}^p B_i \text{vech}(H_{t-i}) \quad (1.13)$$

where A_j and B_i are parameter matrices and c is a vector of constant components.

However it is difficult to ensure the positive definiteness of H_t in VEC model without strong assumptions on parameter, Engle and Kroner (1995) proposed the BEKK specification (defined by Baba et al. (1990)) that easily imposes positive definite under weak assumption. The form is given by,

$$H_t = CC^\top + \sum_{k=1}^K \sum_{j=1}^q A_{kj}^\top \varepsilon_{t-j} \varepsilon_{t-j}^\top A_{kj} + \sum_{k=1}^K \sum_{i=1}^p B_{ki}^\top H_{t-i} B_{ki} \quad (1.14)$$

where C is a lower triangular parameter matrix.

Other than the direct generalization of GARCH models introduced above, the nonlinear combination of univariate GARCH models are more easily estimable. This kind of MGARCH model are based on the decomposition of the conditional covariance matrix into conditional standard deviations and correlations. The simplest is Constant Conditional Correlation (CCC) model introduced by Bollerslev (1990). The conditional correlation matrix of CCC model is time invariant, can be expressed as,

$$H_t = D_t P D_t \quad (1.15)$$

where D_t denotes the diagonal matrix with the conditional variances along the diagonal. Therefore $\{D_t\}_{ii} = \sigma_{it}^2$, with each σ_{it}^2 is a univariate GARCH model.

To overcome this limitation, Engle (2002) proposed a Dynamic Conditional Correlation (DCC) model that allows for dynamic conditional correlation structure. Rather than assuming that the conditional correlation ρ_{ij} between the i -th and j -th component is constant in P , it is now the ij -th element of the matrix P_t which is defined as,

$$\begin{aligned} H_t &= D_t P_t D_t \\ P_t &= (\mathcal{I} \odot \mathcal{Q}_t)^{-\frac{1}{2}} \mathcal{Q}_t (\mathcal{I} \odot \mathcal{Q}_t)^{-\frac{1}{2}} \end{aligned} \quad (1.16)$$

with

$$\mathcal{Q}_t = (1 - a - b)\mathcal{S} + a\varepsilon_{t-1}\varepsilon_{t-1}^\top + b\mathcal{Q}_{t-1} \quad (1.17)$$

where a is positive and b is a non-negative scalar such that $a + b < 1$. \mathcal{S} is unconditional matrix of ε_t , \mathcal{Q}_0 is positive definite.

1.4.2 DCC Model Estimation

Figure 1.18 presents the time path of price series for each indices of CRIX family. As observed, the price processes are slightly different after October of 2015. Before that, three indices present similar trend over time. This indicates that the ARIMA(2,0,2) model selected for CRIX return to remove the intertemporal dependence can be implemented to ECRIX and EFCRIX as well, the model selection and estimation procedure are similar to the way of CRIX. In this section, the ARIMA fitting residuals for each index are used for the following analysis .

The DCC-GARCH(1,1) model estimation is employed by the QMLE based on the stochastic process of equations (1.16) and (1.17). One of the assumptions is the iid innovation term of η_t in equation 1.11. We check the standard residuals



Figure 1.18: The price process of CRIX (black), ECRIX (blue) and EFCRIX (red).

 econ_ccgar

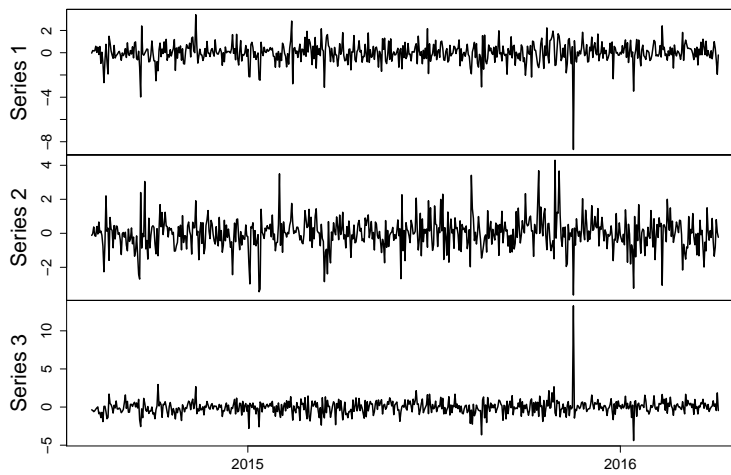


Figure 1.19: The standard error of DCC-GARCH model, with CRIX(upper), ECRIX (middle) and EFCRIX(lower).

 econ_ccgar

of DCC-GARCH(1,1) in Figure 1.19, which displays white noise pattern to some extent.

The estimation results are contained in Table 1.13.

Index type	Coef.	Estimates	Std Error	t test	<i>p</i> -value
CRIX	μ	0.000	0.000	0.759	0.448
	ω	0.000	0.000	0.874	0.382
	α_1	0.123	0.037	3.360	0.001
	β_1	0.832	0.091	9.155	0.000
ECRIX	μ	0.001	0.001	0.775	0.438
	ω	0.000	0.000	0.942	0.346
	α_1	0.123	0.044	2.807	0.004
	β_1	0.832	0.092	9.026	0.000
EFCRIX	μ	0.001	0.001	0.802	0.422
	ω	0.000	0.000	0.946	0.344
	α_1	0.124	0.042	2.960	0.003
	β_1	0.831	0.091	9.153	0.000
DCC	a	0.268	0.018	15.189	0.000
	b	0.571	0.015	38.966	0.000

Table 1.13: Estimation result of DCC-GARCH(1,1) model coefficients.

 econ_ccgar

All the estimated parameters are statistically significant except for the constant terms: mean μ and the constant ω from equation (1.7). Each σ_{it}^2 is a univariate

GARCH(1,1) model,

$$\begin{aligned}\sigma_{CRIX,t}^2 &= 0.123\varepsilon_{CRIX,t-1}^2 + 0.832\sigma_{CRIX,t-1}^2 \\ \sigma_{ECRIX,t}^2 &= 0.123\varepsilon_{ECRIX,t-1}^2 + 0.832\sigma_{ECRIX,t-1}^2 \\ \sigma_{EFCRIX,t}^2 &= 0.124\varepsilon_{EFCRIX,t-1}^2 + 0.831\sigma_{EFCRIX,t-1}^2\end{aligned}$$

The matrix \mathcal{Q}_t of equation (1.17) is,

$$\mathcal{Q}_t = (1 - 0.268 - 0.571)\mathcal{S} + 0.268\varepsilon_{t-1}\varepsilon_{t-1}^\top + 0.571\mathcal{Q}_{t-1}$$

with the unconditional covariance matrix \mathcal{S} ,

$$\mathcal{S} = \begin{pmatrix} 0.994 & 0.994 & 0.994 \\ 0.994 & 0.994 & 0.993 \\ 0.994 & 0.993 & 0.994 \end{pmatrix}$$

1.4.3 DCC Model Diagnostics

Based on the estimation of DCC-GARCH(1,1) model, the estimated and realized volatility are shown in Figure 1.20. The volatility clustering feature is seen graphically from the presence of the sustained periods of high or low volatility, the large changes tend to cluster together. In general, the DCC-GARCH(1,1) fitting is satisfactory as it captures almost all significant volatility changes.

Figure 1.21 presents the estimated autocorrelation dynamics for each of the following series (CRIX v.s. ECRIX, CRIX v.s. EFCRIX and ECRIX v.s. EFCRIX) respectively. We can observe that three autocorrelation dynamics are similar as we expect. To be more specific, three indices are highly positive correlated during the whole sample period. As evidenced in Figure 1.18, the time period after the third semester of 2015 is characterized by relatively lower correlation between three indices, which in turn explains the slightly declines in the autocorrelation dynamics.

To check the adequacy of MGARCH model, we compare the ACF and PACF plots between the premodel squared residual ε_t and the DCC-GARCH(1,1) squared residuals. Figure 1.22 and Figure 1.23 show the GARCH effect is largely eliminated by DCC-GARCH model. Most of the lags are within the 95% confidence bands marked in blue.

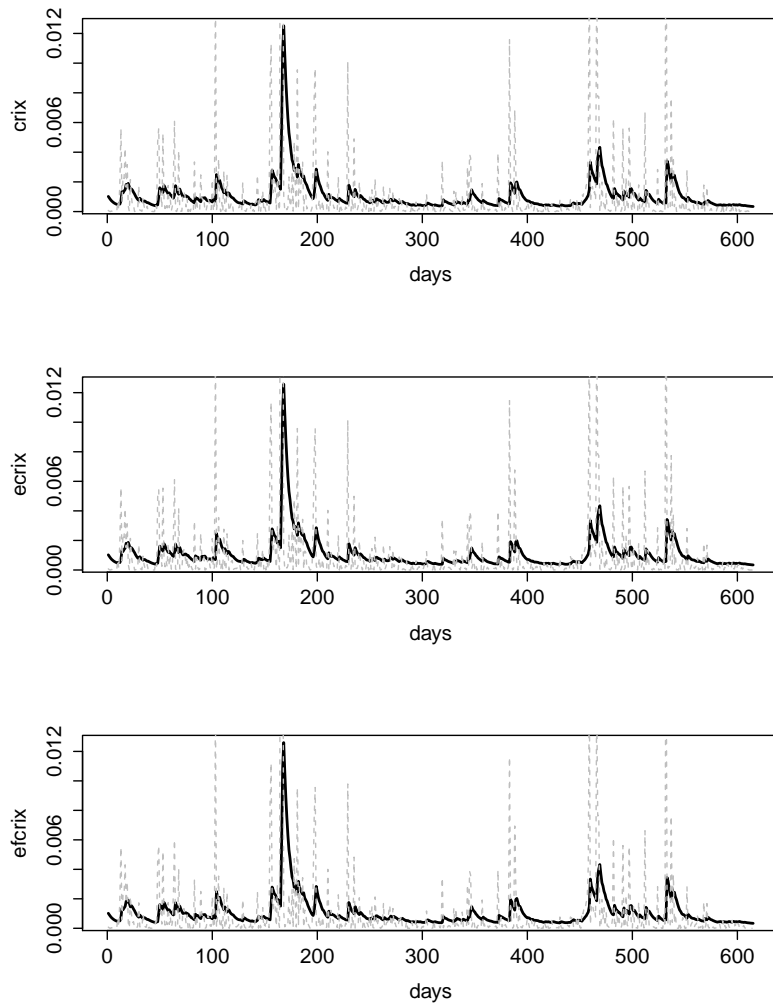


Figure 1.20: The estimated volatility (black) and realized volatility (grey) using DCC-GARCH model, with CRIX (upper), ECRIX(middle) and EFCRIX (lower).

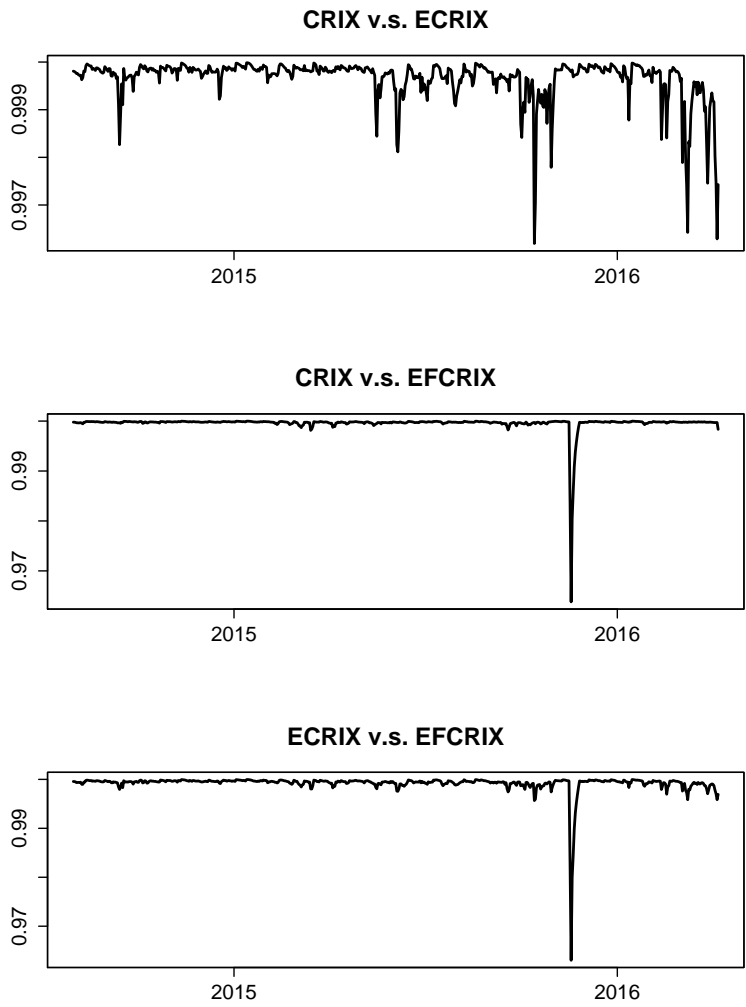


Figure 1.21: The dynamic autocorrelation between three CRIX indices: CRIX, ECRIX and EFCRIX estimated by DCC-GARCH model.

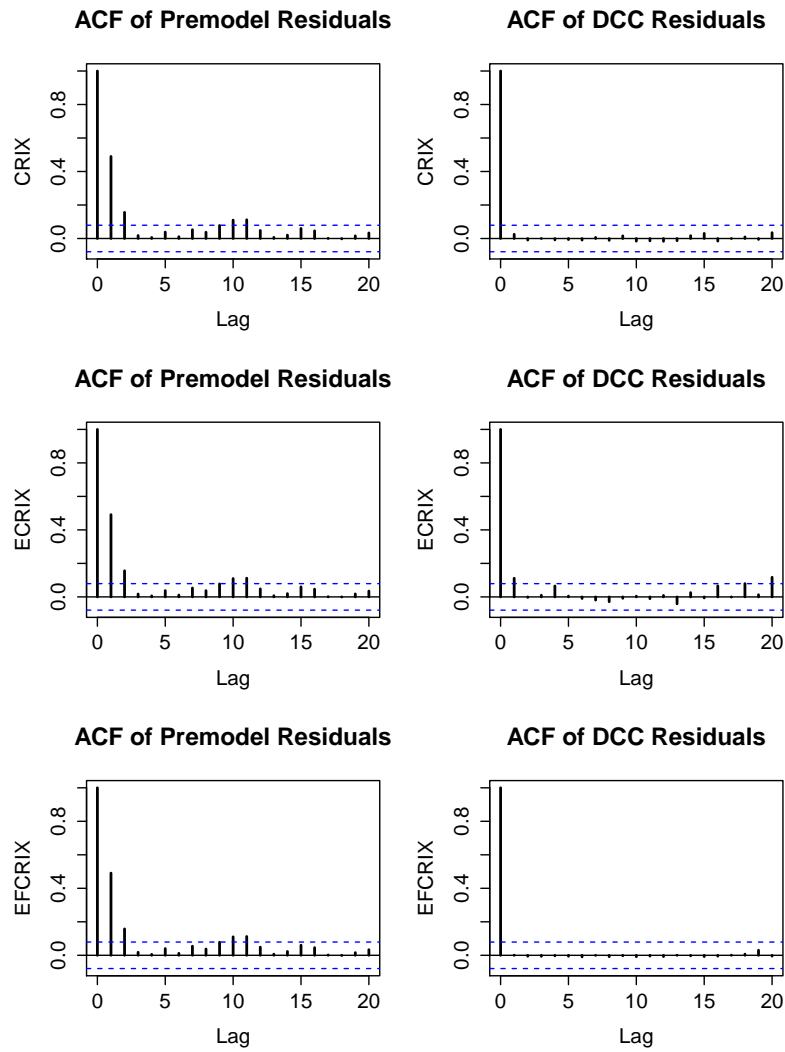


Figure 1.22: The comparison of ACF between premodel squared residuals and DCC squared residuals.

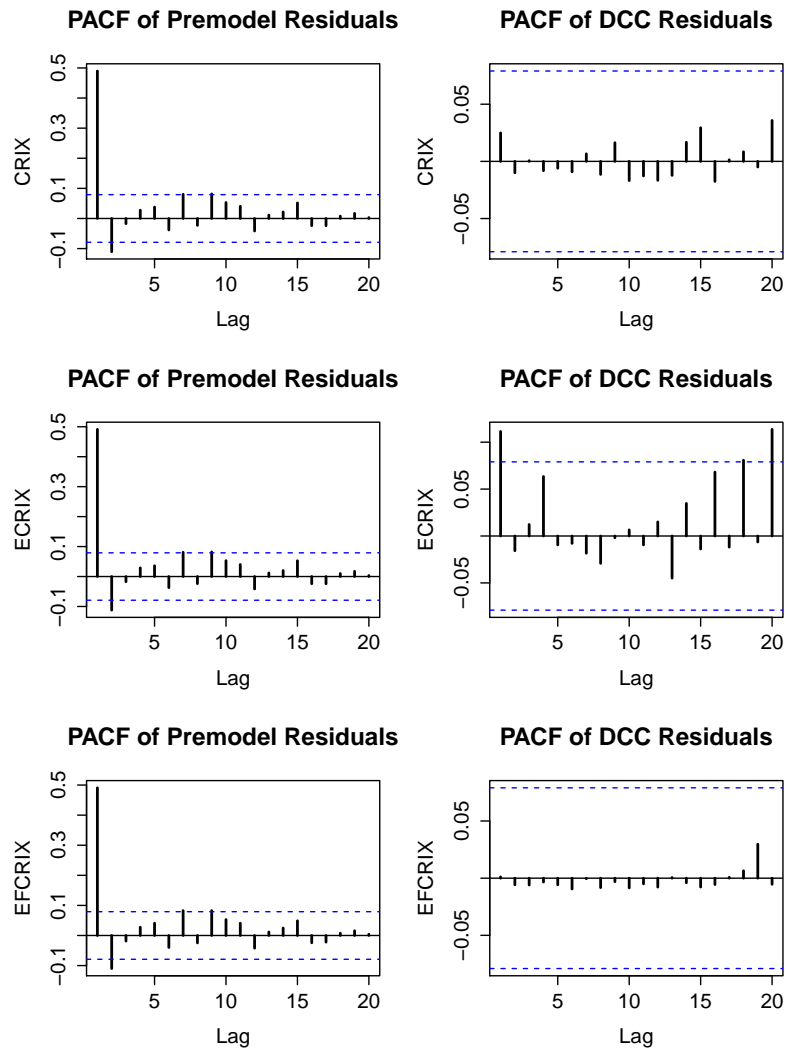


Figure 1.23: The comparison of PACF between premodel squared residuals and DCC squared residuals.

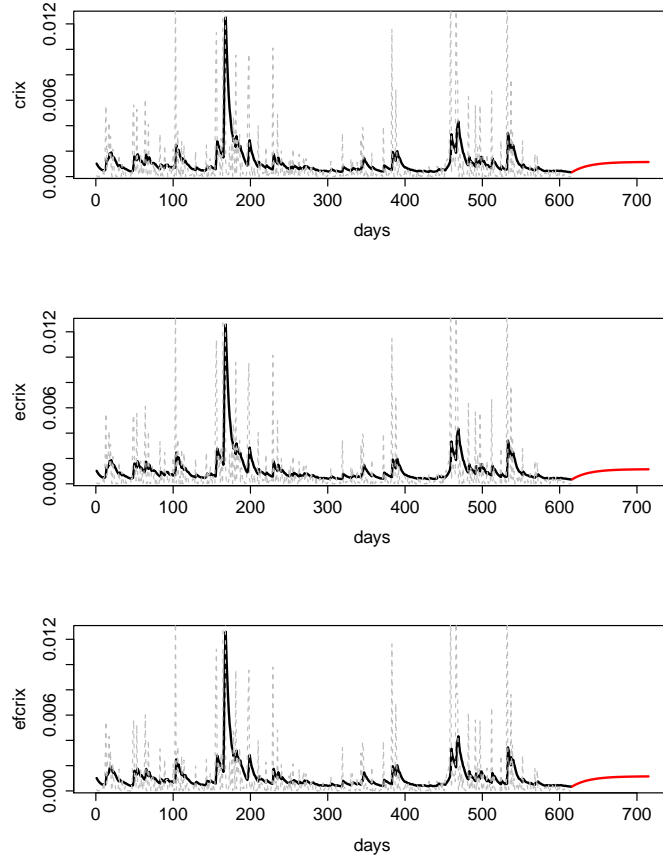


Figure 1.24: 100-step ahead forecasts of estimated volatility using DCC-GARCH(1,1) model.

Moreover, we conduct a 100-step ahead forecast of estimated volatility as illustrated in Figure 1.24, the forecast behavior generally follows the estimated dynamics (black line).

1.5 Nutshell and Outlook

Understanding the dynamics of asset returns is of great importance, it is the first step for practitioners go further with analysis of cryptocurrency markets, like volatility modelling, option pricing and forecasting etc. The motivation behind trying to identify the most accurate econometric model, to determine the

parameters that captures economic behavior arises from the desire to produce the dynamic modeling procedure.

In general it is difficult to model asset returns with basic time series model due to the features of heavy tail, correlated for different time period and volatility clustering. Here we provide a detailed step-by-step econometric analysis using the data of CRIX family: CRIX, ECRIX and EFCRIX. The time horizon for our data sample is from 01/08/2014 to 06/04/2016.

At first, an ARIMA model is implemented for removing the intertemporal dependence. The diagnostic checking stage helps to identify the most accurate econometric model. We then observe the well-known volatility clustering phenomenon from the estimated model residuals. Hence volatility models such as ARCH, GARCH and EGARCH are introduced to eliminate the effect of heteroskedasticity. Additionally, it is observed that the GARCH residuals shows fat-tail properties. We impose the assumption on the residuals with student- t distribution, t -GARCH(1,1) is selected as the best fitted model for all our sample of data based on measures of Log likelihood, AIC and BIC. Finally, a multivariate volatility model, DCC-GARCH(1,1), in order to show the volatility clustering and time varying covariances between three CRIX indices.

With the econometric model on the hand, it facilitates the practitioners to make financial decisions, especially in the context of pricing and hedging of derivative instruments.

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.
- Baba, Y., Engle, R., Kraft, D., and Kroner, K. (1990). Multivariate simultaneous generalized arch, department of economics, university of california at san diego. Technical report, Working Paper.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, pages 498–505.
- Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, pages 116–131.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Brooks, C. (2014). *Introductory econometrics for finance*. Cambridge university press.
- Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, pages 1057–1072.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.
- 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.
- Franke, J., Härdle, W. K., and Hafner, C. M. (2015). *Statistics of Financial Markets: An Introduction*. Springer.
- Geweke, J. (1986). Modelling the persistence of conditional variances: A comment. *Econometric Reviews*, 5(1):57–61.
- Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.
- Härdle, W. K. and Trimborn, S. (2015). Crix or evaluating blockchain based currencies. Technical report, SFB 649 Discussion Paper.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1):159–178.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.
- Pantula, S. G. (1986). Comment. *Econometric Reviews*, 5(1):71–74.
- Rachev, S. T., Mittnik, S., Fabozzi, F. J., Focardi, S. M., and JaĀsić, T. (2007). *Financial econometrics: from basics to advanced modeling techniques*, volume 150. John Wiley & Sons.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

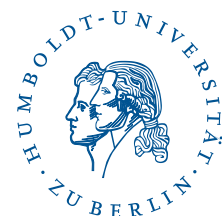
SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Downside risk and stock returns: An empirical analysis of the long-run and short-run dynamics from the G-7 Countries" by Cathy Yi-Hsuan Chen, Thomas C. Chiang and Wolfgang Karl Härdle, January 2016.
- 002 "Uncertainty and Employment Dynamics in the Euro Area and the US" by Aleksei Netsunajev and Katharina Glass, January 2016.
- 003 "College Admissions with Entrance Exams: Centralized versus Decentralized" by Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler and Morimitsu Kurino, January 2016.
- 004 "Leveraged ETF options implied volatility paradox: a statistical study" by Wolfgang Karl Härdle, Sergey Nasekin and Zhiwu Hong, February 2016.
- 005 "The German Labor Market Miracle, 2003 -2015: An Assessment" by Michael C. Burda, February 2016.
- 006 "What Derives the Bond Portfolio Value-at-Risk: Information Roles of Macroeconomic and Financial Stress Factors" by Anthony H. Tu and Cathy Yi-Hsuan Chen, February 2016.
- 007 "Budget-neutral fiscal rules targeting inflation differentials" by Maren Brede, February 2016.
- 008 "Measuring the benefit from reducing income inequality in terms of GDP" by Simon Voigts, February 2016.
- 009 "Solving DSGE Portfolio Choice Models with Asymmetric Countries" by Grzegorz R. Dlugoszek, February 2016.
- 010 "No Role for the Hartz Reforms? Demand and Supply Factors in the German Labor Market, 1993-2014" by Michael C. Burda and Stefanie Seele, February 2016.
- 011 "Cognitive Load Increases Risk Aversion" by Holger Gerhardt, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig, March 2016.
- 012 "Neighborhood Effects in Wind Farm Performance: An Econometric Approach" by Matthias Ritter, Simone Pieralli and Martin Odening, March 2016.
- 013 "The importance of time-varying parameters in new Keynesian models with zero lower bound" by Julien Albertini and Hong Lan, March 2016.
- 014 "Aggregate Employment, Job Polarization and Inequalities: A Transatlantic Perspective" by Julien Albertini and Jean Olivier Hairault, March 2016.
- 015 "The Anchoring of Inflation Expectations in the Short and in the Long Run" by Dieter Nautz, Aleksei Netsunajev and Till Strohsal, March 2016.
- 016 "Irrational Exuberance and Herding in Financial Markets" by Christopher Boortz, March 2016.
- 017 "Calculating Joint Confidence Bands for Impulse Response Functions using Highest Density Regions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, March 2016.
- 018 "Factorisable Sparse Tail Event Curves with Expectiles" by Wolfgang K. Härdle, Chen Huang and Shih-Kang Chao, March 2016.
- 019 "International dynamics of inflation expectations" by Aleksei Netsunajev and Lars Winkelmann, May 2016.
- 020 "Academic Ranking Scales in Economics: Prediction and Imputation" by Alona Zharova, Andrija Mihoci and Wolfgang Karl Härdle, May 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 021 "CRIX or evaluating blockchain based currencies" by Simon Trimborn and Wolfgang Karl Härdle, May 2016.
- 022 "Towards a national indicator for urban green space provision and environmental inequalities in Germany: Method and findings" by Henry Wüstemann, Dennis Kalisch, June 2016.
- 023 "A Mortality Model for Multi-populations: A Semi-Parametric Approach" by Lei Fang, Wolfgang K. Härdle and Juhyun Park, June 2016.
- 024 "Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors" by Kun Ho Kim, Shih-Kang Chao and Wolfgang K. Härdle, August 2016.
- 025 "Forecasting Limit Order Book Liquidity Supply-Demand Curves with Functional Autoregressive Dynamics" by Ying Chen, Wee Song Chua and Wolfgang K. Härdle, August 2016.
- 026 "VAT multipliers and pass-through dynamics" by Simon Voigts, August 2016.
- 027 "Can a Bonus Overcome Moral Hazard? An Experiment on Voluntary Payments, Competition, and Reputation in Markets for Expert Services" by Vera Angelova and Tobias Regner, August 2016.
- 028 "Relative Performance of Liability Rules: Experimental Evidence" by Vera Angelova, Giuseppe Attanasi, Yolande Hiriart, August 2016.
- 029 "What renders financial advisors less treacherous? On commissions and reciprocity" by Vera Angelova, August 2016.
- 030 "Do voluntary payments to advisors improve the quality of financial advice? An experimental sender-receiver game" by Vera Angelova and Tobias Regner, August 2016.
- 031 "A first econometric analysis of the CRIX family" by Shi Chen, Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle, TM Lee and Bobby Ong, August 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

