The Econometrics of CRIX

Shi Chen
Wolfgang Karl Härdle
Cathy Chen
TM Lee
Bobby Ong

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E.-Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Currencies - Cigarettes, USD, Cryptos

- Anything can be a currency

Figure 1: Cigarette trading in postwar Germany ([1])

- Anyone can offer a currency

Figure 2: Friedrich A. Hayek ([2])

Digital Economy

- Amazon
- Paypal
- Google Wallet
- Cryptocurrencies
- Ripple

Cryptocurrencies

- Decentralized, virtual, low transaction costs

- NYSE, Andreesen Horowitz, DFJ: Coinbase funding (75 M$)
- Nasdaq: company-wide utilization of blockchain technology
- Citigroup: own coin development
- PBOC: working on digital currency
- Switzerland Zug: first city accepts Bitcoin payments

Pokémon Go and Cryptocurrency

- Each creature could have an asset based crypto-tokens that could be traded in blockchain.
- Pokémon and BTC: PokéBits

Source: steemit, Bitcoin.com

Market Capitalization

CoinMarketCap

CRypto IndeX - CRIX

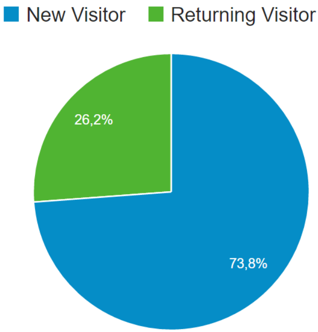
- high market capitalization
- covers approximately 30 cryptos
    - different liquidity rules
    - model selection criteria
- CRIX family
    - CRIX
    - ECRIX (Exact CRIX)
    - EFCRIX (Exact Full CRIX)

crix.hu-berlin.de

Reference: Trimborn, S. and Härdle, W. (2016)

CRypto IndeX - CRIX

- 290 cryptos
- Prices, capitalization, volume
- As of 20160815, overview of CRIX: hu.berlin/crix
  - Users: 1911
  - Page views: 3920
  - average time: 00:01:17



Challenge

1. What's the dynamics of CRIX?
2. How stable is the CRIX model over time?
3. Consequence for pricing derivatives.

The Econometrics of CRIX

Outline

1. Motivation ✓
2. Data
3. ARIMA Model
4. Stochastic Volatility Model
5. Multivariate GARCH Model
6. Nutshell

All QuantLets from www.quantlet.de

Three Indices

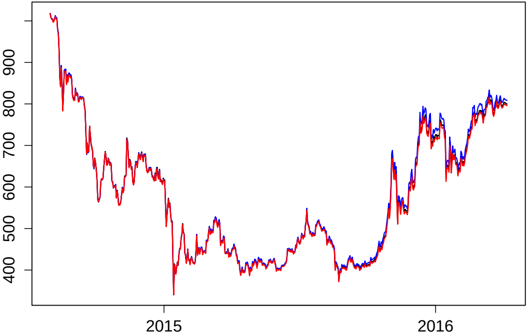


Figure 3: The daily value of indices in the CRIX family from 01/08/2014 to 06/04/2016: CRIX, ECRIX and EFCRIX.

econ_ccgar

Data Description

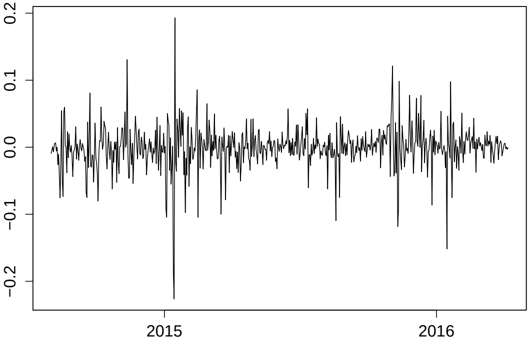


Figure 4: The log returns of CRIX index from 01/08/2014 to 06/04/2016.

econ_crix

Distributional Property

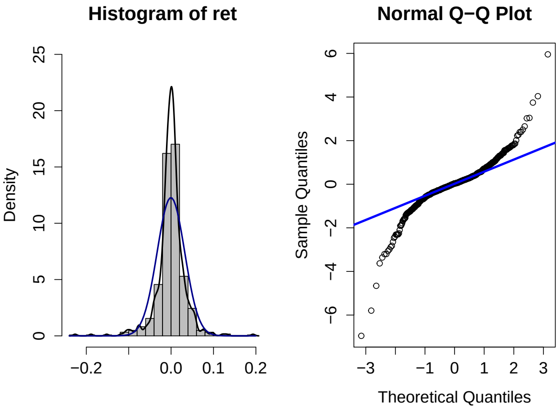


Figure 5: Histogram and QQ plot of CRIX returns from 01/08/2014 to 06/04/2016.

econ_crix

First Approach

The ARIMA(p, d, q) with $d = 1$ is,

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

or

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

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

Box-Jenkins Procedure

1. Identification of lag orders
2. Parameter estimation
3. Diagnostic checking

Step 1: Lag Orders

- p-value for stationarity tests: ADF test (null hypothesis: unit root) of 0.01; KPSS test (null hypothesis: stationary) of 0.1.

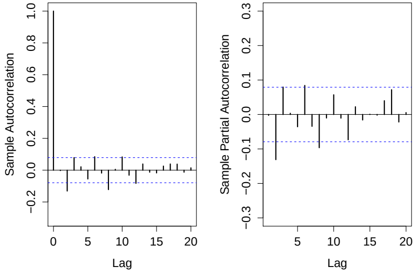


Figure 6: The sample ACF and PACF of CRIX returns from 01/08/2014 to 06/04/2016. econ_arima

Step 1: Lag Orders - ctd

ARIMA model selected	AIC	BIC
ARIMA(2,0,0)	-2469	-2451
ARIMA(2,0,2)	-2474	-2448
ARIMA(2,0,3)	-2473	-2442
ARIMA(4,0,2)	-2476	-2441
ARIMA(2,1,1)	-2459	-2441
ARIMA(2,1,3)	-2464	-2438

Table 1: The ARIMA model selection with AIC and BIC. econ_arima

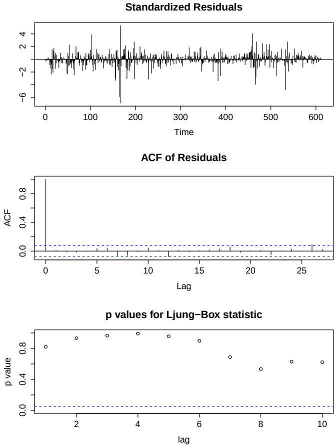
Step 2: Parameter Estimation

Coefficients	Estimate	Standard deviation
intercept c	-0.00	0.00
a_1	-0.70	0.11
a_2	-0.75	0.12
b_1	0.70	0.14
b_2	0.64	0.13
Log likelihood	1243.12	

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

Step 3: Diagnostic Checking

- Diagnostic plot of ARIMA(2,0,2) model
- significant *p*-values of Ljung-Box test statistic
- model residuals are independent



Q econ_arima

ARIMA Model Forecast

- With ARIMA(2,0,2) model, we predict CRIX returns for next 30 days.

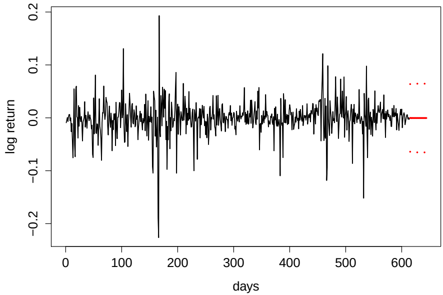


Figure 7: CRIX returns and predicted values. The confidence bands are red dashed lines. econ_arima

Discussion

- We build an ARIMA(2,0,2) model for the CRIX return series to model intertemporal dependence.
- ACF of model residuals has no significant lags as evidenced in Step 3: Diagnostic Checking.
- Further work: Homoskedasticity or Heteroskedasticity.

Volatility Clustering

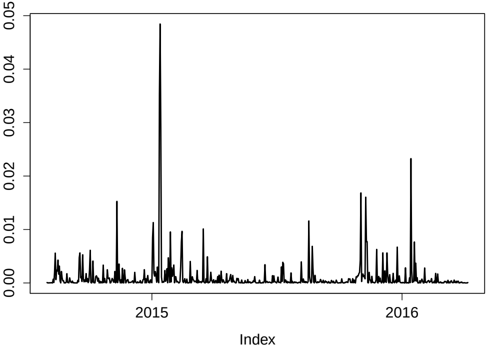


Figure 8: The squared ARIMA(2,0,2) residuals of CRIX returns.
econ_vola

ARCH Model

- ARCH(q) model,

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

  - ε_t is the ARIMA model residual
  - σ_t^2 is the variance of ε_t conditional on the information available at time t.

Heteroskedasticity effect

- Two approaches:
  - ARCH LM test (null hypothesis: no ARCH effects) of ε_t
  - Ljung-Box test for ε_t^2
- both $p-$values smaller than $2.2e - 16$.
- Next step: determine lag order q of ARCH model

Lag order q

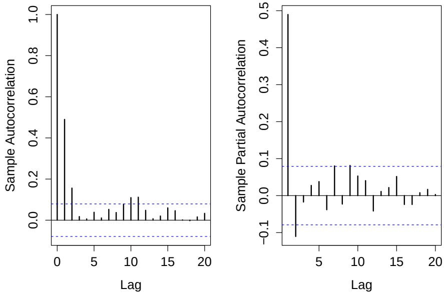


Figure 9: The ACF and PACF of squared ARIMA(2,0,2) residuals from 01/08/2014 to 06/04/2016. econ_vola

Lag Order q - ctd

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 3: The ARCH model selection with AIC and BIC. econ_arch

ARCH Estimation

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

GARCH Model

- The standard GARCH(p, q) model is,

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

with the condition that

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

- This ensures that the GARCH model is strictly stationary with finite variance. Continuous-time GARCH model

Lag Orders p, q

- Normally up to GARCH$(2, 2)$ model is used in practice.
- In particular, the orders of $p = q = 1$ is sufficient in most cases.

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 5: Comparison of GARCH model, orders up to $p = q = 2$.
econ_garch

GARCH Estimation I

- GARCH(1,2) model,

$$
\begin{aligned}
\varepsilon_t &= Z_t \sigma_t, \quad Z_t \sim N(0,1) \\
\sigma_t^2 &= \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2
\end{aligned}
$$

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	$9.91e-05$	$4.75e-05$	2.08*
α_1	$1.65e-01$	$3.72e-02$	4.45***
β_1	$8.07e-02$	$8.24e-02$	0.98
β_2	$6.51e-01$	$8.20e-02$	7.94***

Table 6: Estimation result of ARIMA(2,0,2)-GARCH(1,2) model. $*$ represents significant level of 5% and $***$ of 0.1%. econ_garch

GARCH Estimation II

- GARCH(1,1) model is sufficient in most cases,

$$\begin{aligned} \varepsilon_t &= Z_t \sigma_t, \quad Z_t \sim N(0,1) \\ \sigma_t^2 &= \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned}$$

- All parameters are significant:

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	$5.32e-05$	$2.25e-05$	2.37*
α_1	$1.20e-01$	$2.79e-02$	4.32***
β_1	$8.32e-02$	$3.99e-02$	20.85***

Table 7: Estimation result of ARIMA(2,0,2)-GARCH(1,1) model. $*$ represents significant level of 5% and $***$ of 0.1%. econ_garch

GARCH Estimation II - ctd

- With no significant correlations for any lag, GARCH(1,1) is sufficient enough to explain the heteroskedasticity effect.

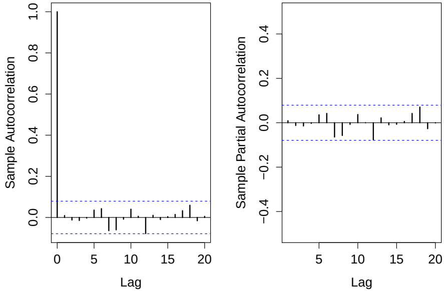


Figure 10: The ACF and PACF of squared ARIMA(2,0,2) residuals from 01/08/2014 to 06/04/2016. econ_garch

GARCH Model Residual

- Kolmogorov-Smirnov test of ARIMA-GARCH model residuals.
- The small p-value rejects the null hypothesis that the residuals are drawn from the normal distribution.
- Sample data exhibits leptokurtosis.

Model	Kolmogorov distance	p-value
ARIMA-GARCH	0.50	$2.86e-10$

Table 8: Test of model residuals of ARIMA(2,0,2)-GARCH(1,1) process.

econ_garch

GARCH Model Residual - ctd

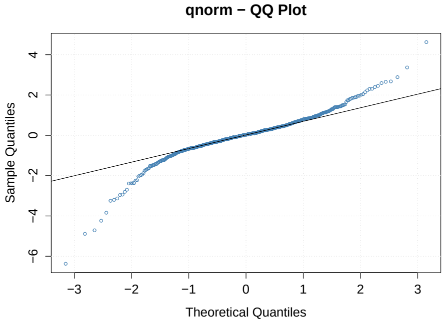


Figure 11: The QQ plots of model residuals of ARIMA-GARCH process.

econ_garch

t-GARCH Estimation

- Impose $Z_t \sim t(d)$ to replace the normal assumption of Z_t
- ξ controls the height and fat-tail of density function, therefore different shape of distribution function.

Coefficients	Estimates	Standard deviation	T test
ω	$8.39e-05$	$5.45e-05$	1.54
α_1	$2.82e-01$	$1.46e-01$	1.93$^{\cdot}$
β_1	$7.90e-01$	$6.12e-02$	12.91***
ξ	$2.58e+00$	$3.62e-01$	7.11***

Table 9: Estimation result of ARIMA(2,0,2)-t-GARCH(1,1) model. . represents significant level of 10% and $***$ of 0.1%. econ_tgarch

t-GARCH Model Estimation - ctd

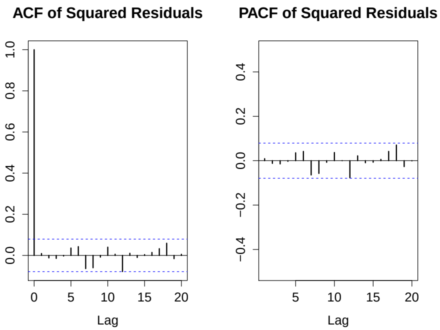


Figure 12: The ACF and PACF plots for model residuals of ARIMA(2,0,2)-t-GARCH(1,1) process. econ_tgarch

t-GARCH Model Residual

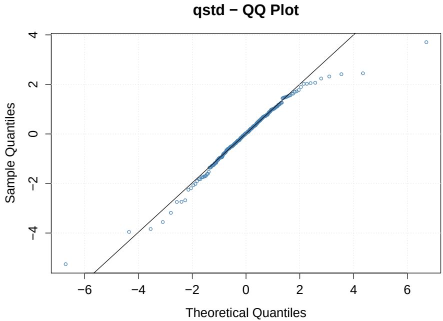


Figure 13: The QQ plots of model residuals of ARIMA-t-GARCH process.

econ_tgarch

EGARCH Model

- The introduced GARCH model successfully solve the problem of volatility clustering, but cannot capture the leverage effect.
- The exponential GARCH (EGARCH) model with standard innovations,

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

with the condition that

$$g_j(Z_t) = \alpha_j Z_t + \phi_j(|Z_{t-j}| - \mathsf{E}|Z_{t-j}|), \quad j = 1, 2, \ldots, q$$

t-EGARCH Estimation

- Fit a EGARCH(1,1) model with student *t* distributed innovation term.
- The estimation results of the ARIMA(2,0,2)-*t*-EGARCH(1,1) model is,

Coefficients	Estimates	Standard deviation	Ljung-Box test statistic
ω	$9.91e-05$	$4.75e-05$	2.08*
α_1	$1.65e-01$	$3.72e-02$	4.45*
β_1	$8.07e-02$	$8.24e-02$	0.98
ϕ_1	$6.51e-01$	$8.20e-02$	7.94*

Table 10: Estimation result of ARIMA(2,0,2)-*t*-EGARCH(1,1) model. $*$ represents significant level of 5% and $***$ of 0.1%. econ_tgarch

t-EGARCH Model Residual

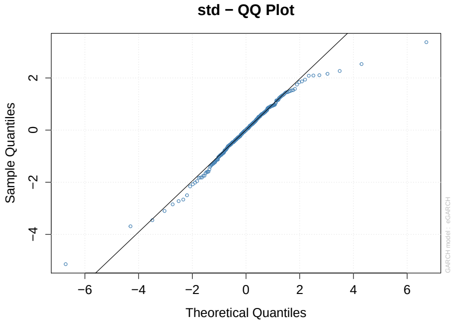


Figure 14: The QQ plots of model residuals of ARIMA-*t*-EGARCH process.

econ_tgarch

GARCH Model Selection

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

Table 11: Comparison of the variants of GARCH model. econ_tgarch

MGARCH Model

- Consider the error term ε_t with $\mathsf{E}(\varepsilon_t) = 0$, and conditional covariance matrix H_t is $(d \times d)$ positive definite,

$$\varepsilon_t = H_t^{\frac{1}{2}} \eta_t$$

$H_t^{\frac{1}{2}}$ can be obtained by Cholesky factorization of H_t.

- η_t is an iid innovation vector such that,

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

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

DCC-GARCH Model

- Different specification of H_t yields various parametric formulations: VEC, BEKK, CCC, DCC etc.
- Dynamic Conditional Correlation (DCC) model: conditional correlation ρ_{ij} between the i-th and j-th component is the ij-th element of the matrix P_t

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

with

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

  - The diagonal matrix D_t is the conditional variance matrix.
  - $\mathcal{S}$ is unconditional matrix of ε_t

DCC-GARCH Model Estimation

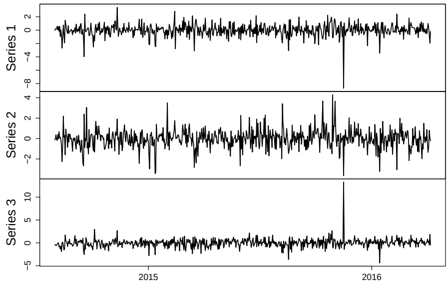


Figure 15: The standard error of DCC-GARCH model, with CRIX(upper), ECRIX (middle) and EFCRIX(lower). econ_ccgar

DCC-GARCH Model Estimation - ctd

- All the estimated parameters are statistically significant except for the constant terms,

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

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

- The unconditional covariance matrix $\mathcal{S}$,

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

DCC-GARCH Model Estimation - ctd

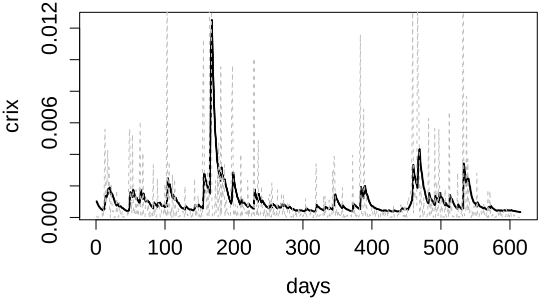


Figure 16: The estimated volatility (black) and realized volatility (grey) using DCC-GARCH model, for example CRIX. econ_ccgar

DCC-GARCH Model Estimation - ctd

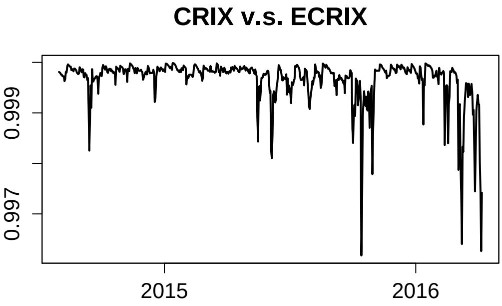


Figure 17: The dynamic autocorrelation between three CRIX indices: CRIX, ECRIX and EFCRIX estimated by DCC-GARCH model. econ_ccgar

DCC-GARCH Model Estimation - ctd

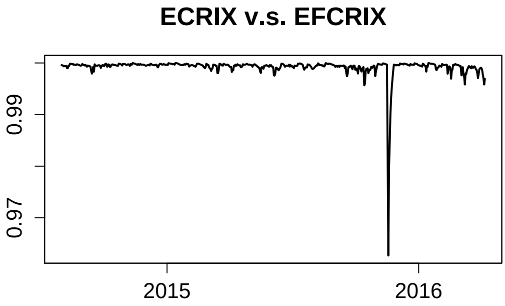


Figure 18: The dynamic autocorrelation between three CRIX indices: CRIX, ECRIX and EFCRIX estimated by DCC-GARCH model. econ_ccgar

DCC-GARCH Model Diagnostics

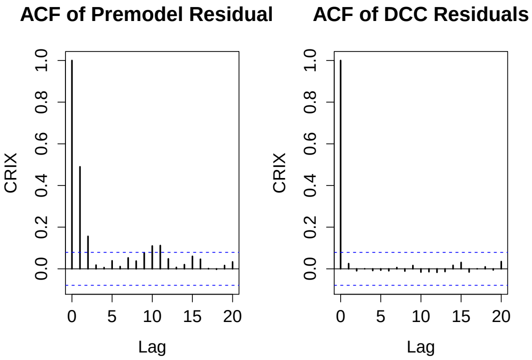


Figure 19: The comparison of ACF between premodel squared residuals and DCC squared residuals, for example CRIX. econ_ccgar

DCC-GARCH Model Diagnostics - ctd

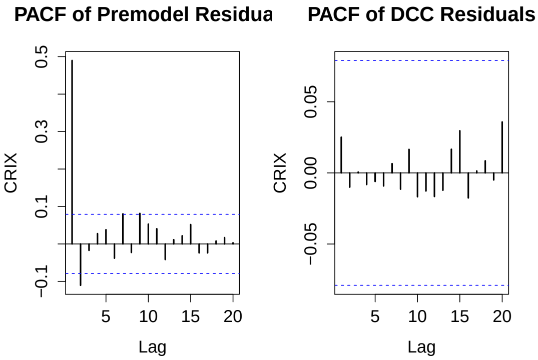


Figure 20: The comparison of PACF between premodel squared residuals and DCC squared residuals, for example CRIX. econ_ccgar

GARCH Option Pricing Model

- Option pricing models
  - Black-Scholes model
  - GARCH models: superior in describing asset return dynamics.
- For instance Heston and Nandi (2000), HN model in brief.
  - a closed form expression for European option prices
  - GARCH models with Gaussian innovations

HN model

- In the HN model ,the asset return dynamic under the risk neutral measure $\mathbb{Q}$ is,

$$\begin{aligned}
\log\left(\frac{S_t}{S_{t-1}}\right) &= r - \frac{\sigma_t^2}{2} + \sigma_t Z_t \\
\sigma_t^2 &= \omega_{hn} + \beta_{hn}\sigma_{t-1}^2 + \alpha_{hn}(Z_{t-1} - \gamma_{hn}\sigma_{t-1})^2
\end{aligned}$$

  - r is risk-free interest rate
  - Z_t is a standard Gaussian innovation
  - Risk neutral GARCH parameter: $\theta_{hn} = \{\omega_{hn}, \beta_{hn}, \alpha_{hn}, \gamma_{hn}\}$
  - S_t is the return to estimate.

HN model - ctd

- The call option C_t at time t, with strike price K and time to maturity τ is worth,

$$
\begin{aligned}
C_t &= \exp(-r\tau) f_{hn}(1) \left[\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathcal{R} \left\{ \frac{K^{-i\phi} f_{hn}(i\phi + 1)}{i\phi f_{hn}(1)} \right\} d\phi \right] \\
&- \exp(-r\tau) K \left[\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathcal{R} \left\{ \frac{K^{-i\phi} f_{hn}(i\phi)}{i\phi} \right\} d\phi \right]
\end{aligned}
$$

  - $\mathcal{R}\{\}$ denotes the real part of a complex number
  - $f_{hn}(\phi)$ is the conditional moment generating function at time t

$$
f_{hn}(\phi) = \mathrm{E}_{\mathbb{Q}} \left[\exp\left\{ \phi log(S_t) \right\} \mid \mathcal{F}_t \right] = S_t^{\phi} \exp(A_t + B_t \sigma_{t+1}^2)
$$

HN model - ctd

- The coefficients A_t and B_t are computed backward starting from the terminal condition $A_T = B_T = 0$.
- The HN model recursive equations are,

$$
\begin{aligned}
A_t &= A_{t+1} + \phi r + B_{t+1}\omega_{hn} - \frac{1}{2}\log(1 - 2\alpha_{hn}B_{t+1}) \\
B_t &= \phi\left(\gamma_{hn} - \frac{1}{2}\right) - \frac{\gamma_{hn}^2}{2} + \beta_{hn}B_{t+1} + \frac{1/2(\phi - \gamma_{hn})^2}{1 - 2\alpha_{hn}B_{t+1}}
\end{aligned}
$$

Nutshell

- ARIMA model is implemented for removing the intertemporal dependence.
- Volatility models such as ARCH, GARCH and EGARCH are applied to eliminate the effect of heteroskedasticity.
- The t-GARCH(1,1) is introduced to deal with the fat-tail properties of GARCH residuals.
- DCC-GARCH(1,1) exhibits time varying covariances between three CRIX indices.
- Outlook: GARCH option pricing model, eg. HN GARCH model.

The Econometrics of CRIX

Shi Chen
Wolfgang Karl Härdle
Cathy Chen
TM Lee
Bobby Ong

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E.-Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

References

Cigarette trading in postwar Germany, Bundesarchiv,
Bild 183-R79014 / CC-BY-SA.

The original uploader was DickClarkMises
English Wikipedia / CC BY-SA 3.0.

Franke, J., Härdle, W. K. and Hafner, C. M.
Statistics of Financial Markets: an Introduction.
4th ed., Springer Verlag, Berlin Heidelberg, 2015.

Barone-Adesi, G., Engle, R. F. and Mancini., L.
A GARCH option pricing model with filtered historical simulation
Review of Financial Studies 21.3 (2008): 1223-1258.

References

- Hamilton, J. D.
  Time series analysis.
  Princeton university press Princeton, 1994.

- Lütkepohl, H.
  New introduction to multiple time series analysis
  Springer Science & Business Media, 2005.

- Rachev, S. T., Mittnik, S., Fabozzi, F. J., Focardi, S. M., and JaÂsić, T.
  Financial econometrics: from basics to advanced modeling techniques
  John Wiley & Sons, volume 150, 2007.

COGARCH Model

- Irregularly spaced data: continuous-time GARCH model.
- The GARCH(1, 1) model diffusion limit satisfies,

$$
\begin{aligned}
dG_t &= \sigma_t dW_t^{(1)} \\
d\sigma_t^2 &= \theta(\gamma - \sigma_t^2) + \rho\sigma_t^2 dW_t^{(2)}
\end{aligned}
$$

  - G_t is the log return r_t to estimate.
  - $\left\{W_t^{(1)}\right\}_{t\geq 0}$ and $\left\{W_t^{(2)}\right\}_{t\geq 0}$ are two independent Brownian motions.
  - θ, γ and ρ are parameters.

Go Back