ARIMA-GARCH Model of the SKEW Index

Eddie Shim
7/6/16

TABLE OF CONTENTS

I. Introduction……………………………………………………………………..3
II. Model Discussion…………………………………………………………………4
III. Conclusion……………………………………………………………...……9
IV. Code……………………………………………………………...……....10

I. Introduction

This paper serves as an exercise for me to build an ARIMA-GARCH model on the CBOE Skew Index (SKEW) to see how a time series model forecasts this index to move in the near future.

SKEW is an option-based index which measures the perceived tail risk of the distribution of S&P 500 log returns. It measures how much traders in the options market are willing to pay to protect against fat tail risk. Fat tail risk is associated with rare outlier events (2-3+ standard deviations outside the mean). As seen below, S&P 500 log returns has a fat left tail. The SKEW quantifies the additional risk for assuming S&P 500 log returns are normal shaped.

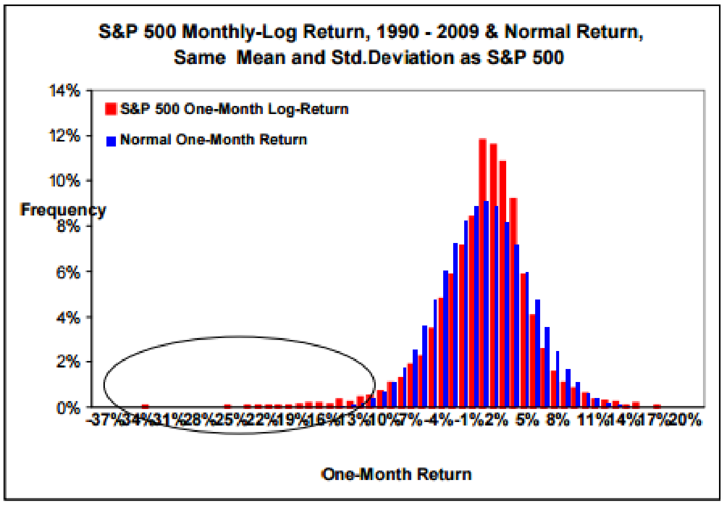


Figure 1 Source: CBOE

I chose to model SKEW because on June 28, 2016, the index reached its highest point since 1990. This observation was one of the many ripples of Brexit causing market uncertainty and pushing global bond yields to all time lows and pushing gold up. [1] In the last few days, the index has returned to normal levels.

Figure 2

[1] http://www.bloomberg.com/news/articles/2016-07-04/black-swans-and-game-theory-a-post-brexit-guide-to-bond-trading

II. Model Discussion

The data I am using is weekly closing price of SKEW from July 7th, 2015 until July 1st, 2016. Data was taken from Yahoo Finance. The raw data plot and logged data plot looks as follows:

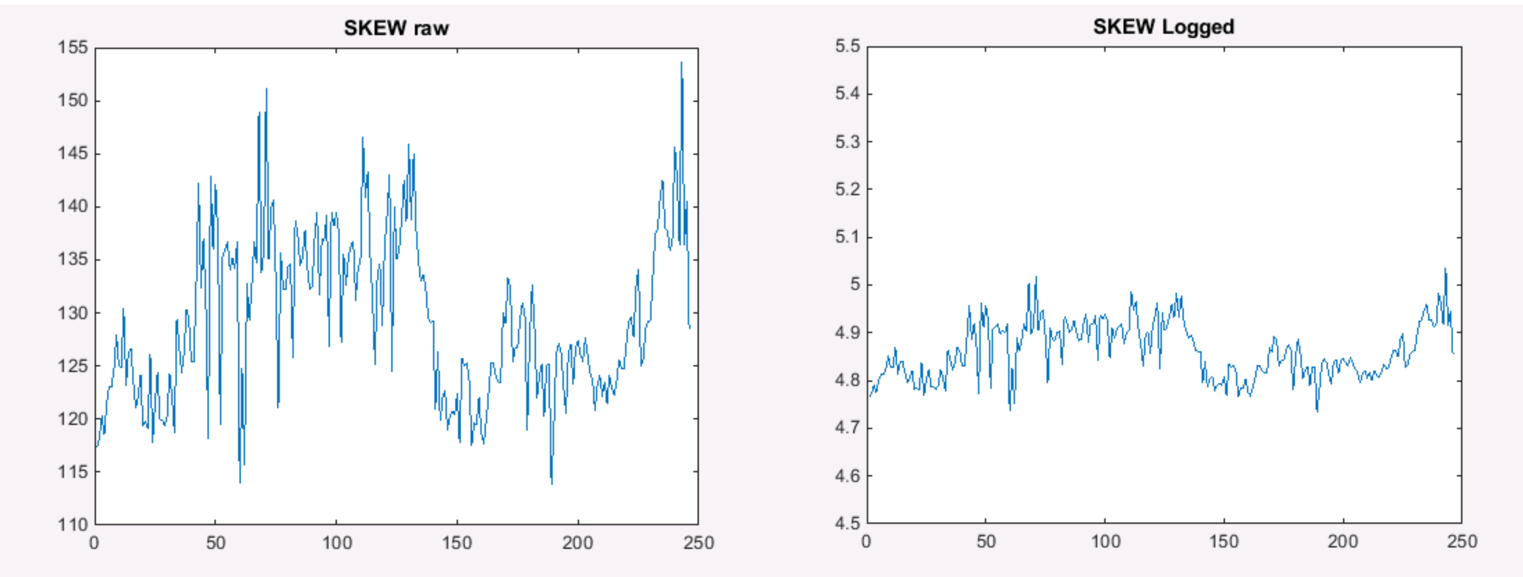


Figure 3

Using an Augmented Dickey Fuller test, I determined that a 1 lag difference transformation is required to make the dataset stationary. The data now looks more like white noise, but still exhibits heteroskedacity.

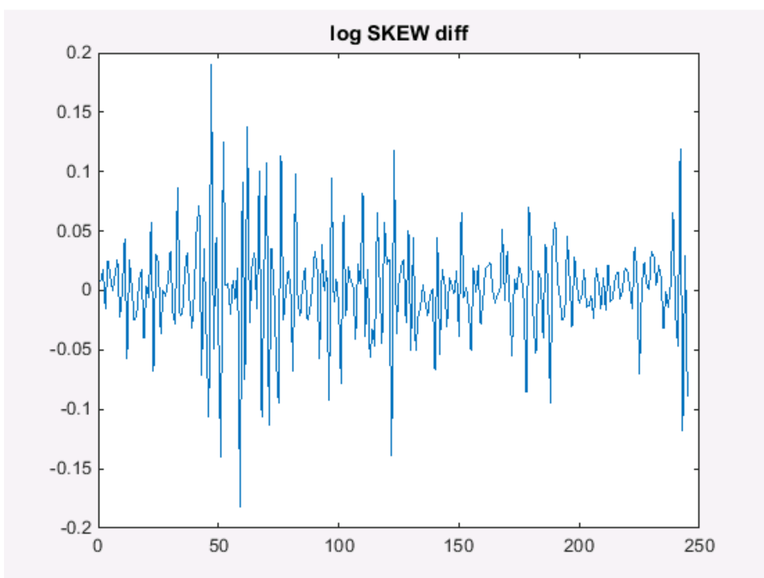
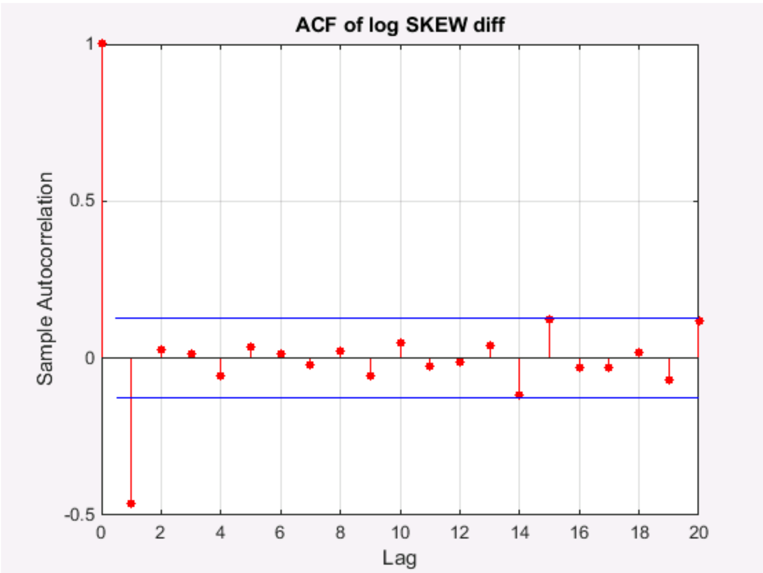


Figure 4

In this section, I will determine the (p,q) variables, the amount of lags incorporated in the model. The ACF plot shows significant autocorrelation at lag =1. The PACF plot shows significant autocorrelation at lag =1,2,3,4.



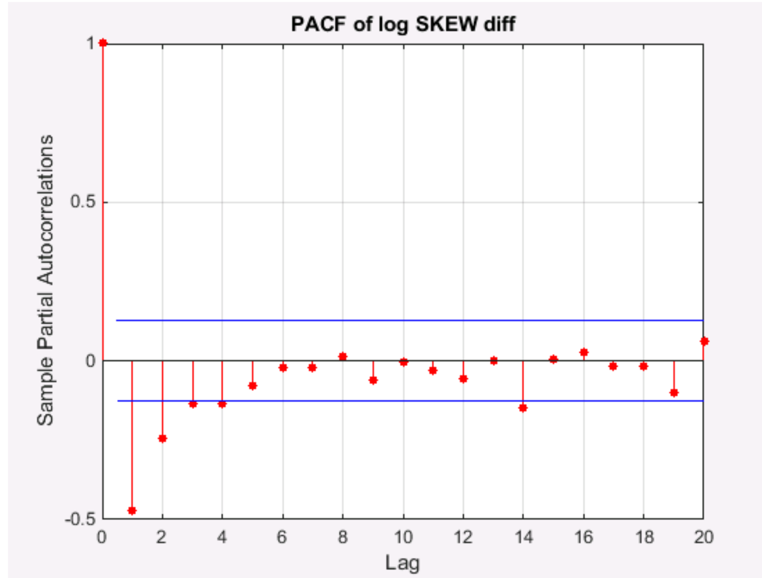


Figure 5

Using the Bayesian Information Criterion (BIC), I checked the loglikelihood values of ARIMA(p,1,q) models, spanning p=[1,…,4] and q = [1,…4]. Below is the output of loglikelihood values with degree p values on the rows and degree q values in the columns

```
BIC =

 -878.5382 -874.0548 -869.6655 -865.3727
 -874.2080 -876.5864 -868.2170 -871.1078
 -870.0188 -865.3575 -867.0345 -872.3628
 -865.2928 -867.3982 -864.0258 -867.7629
```

From this output, it seems ARIMA(1,1,1) should be our optimal model choice with regards to model accuracy and parsimony.

Below is the output for an ARIMA(1,1,1) model:

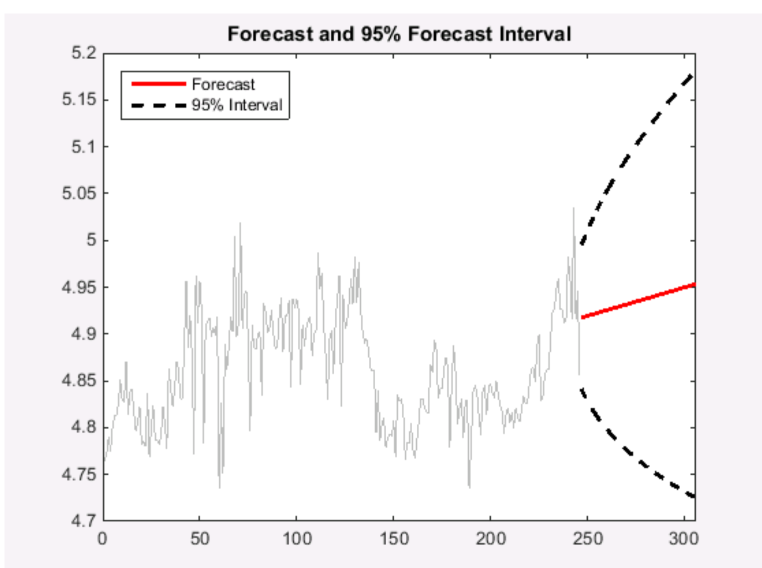


Figure 6

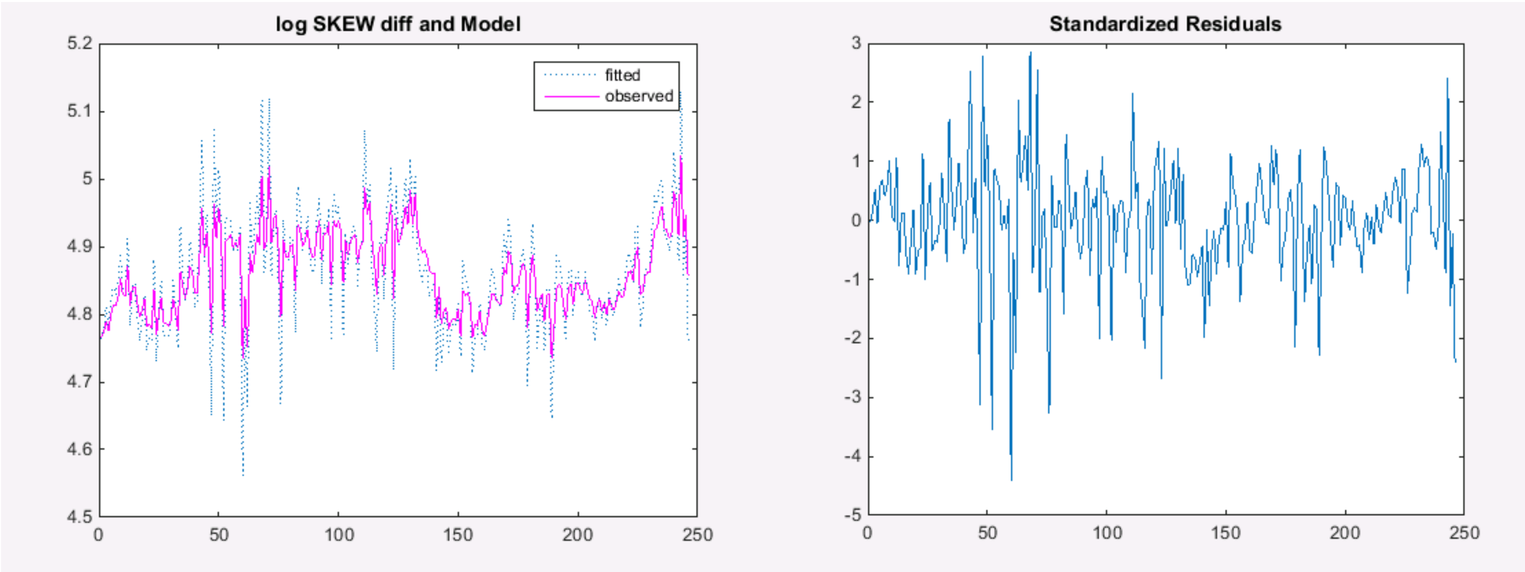


Figure 7

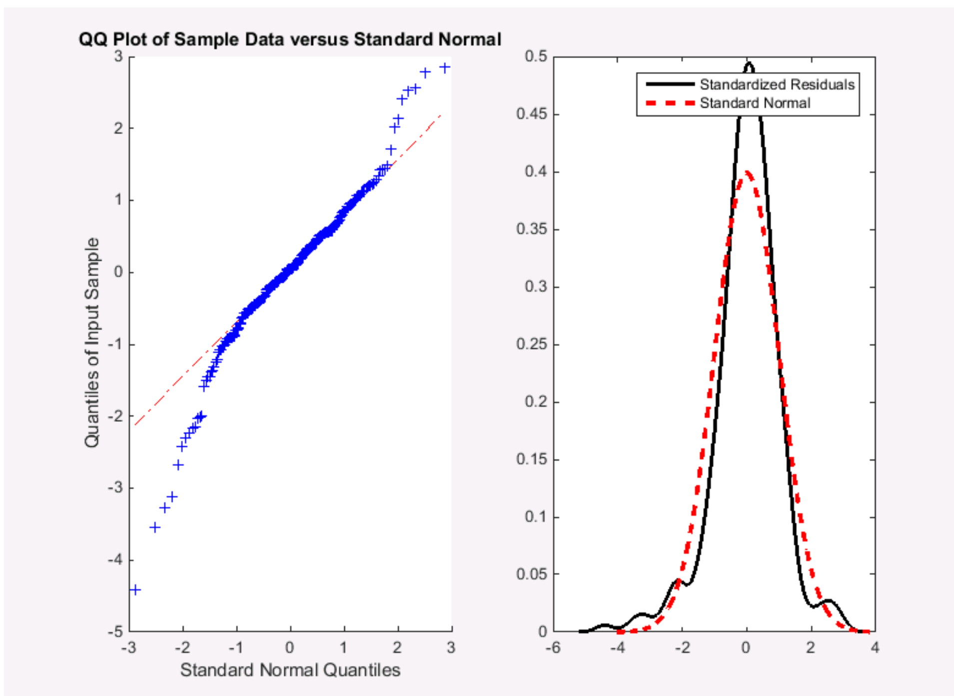


Figure 8

The model's residuals have fat tails – this is because of the data's heteroskedacity as mentioned earlier. The next attempt will be an ARIMA(1,1,1)-GARCH(1,1) to capture conditional variance.

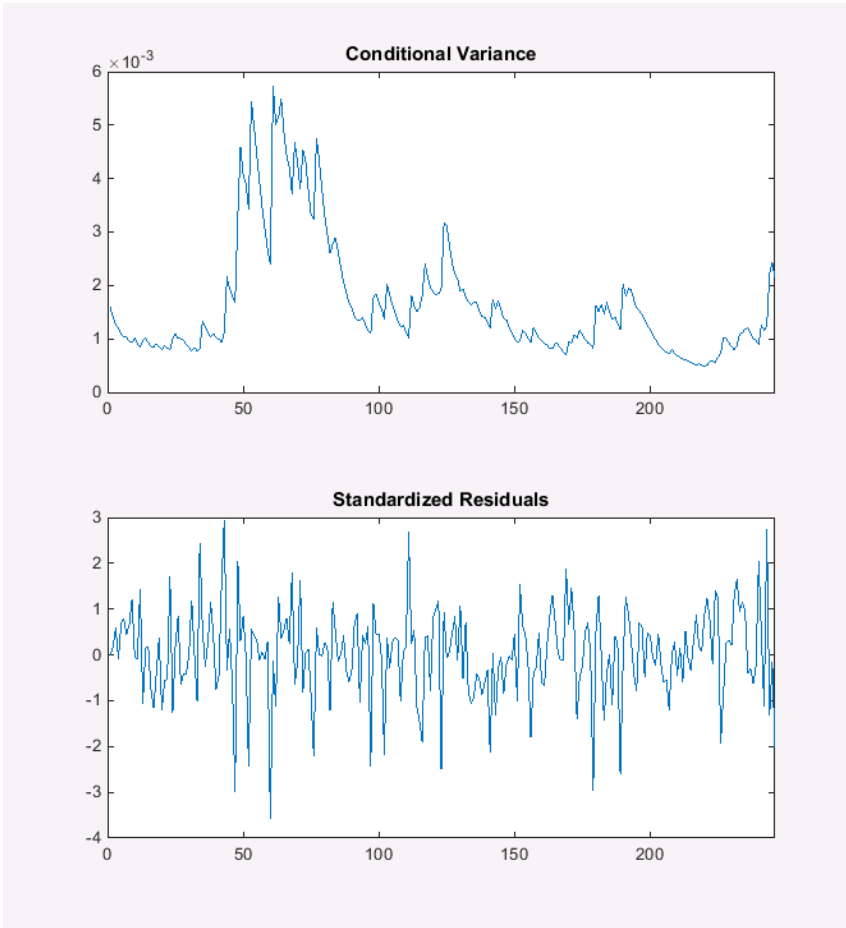


Figure 9

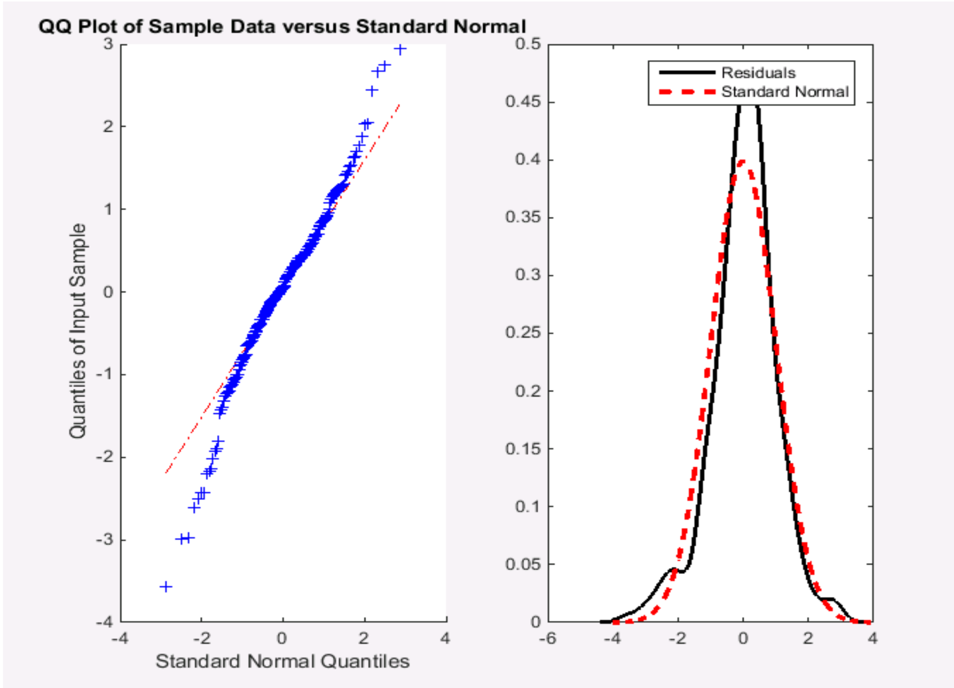


Figure 10

The GARCH model better in removing some of the conditional variance near t=50 and t=125. It's still not perfect, but it is an improvement.

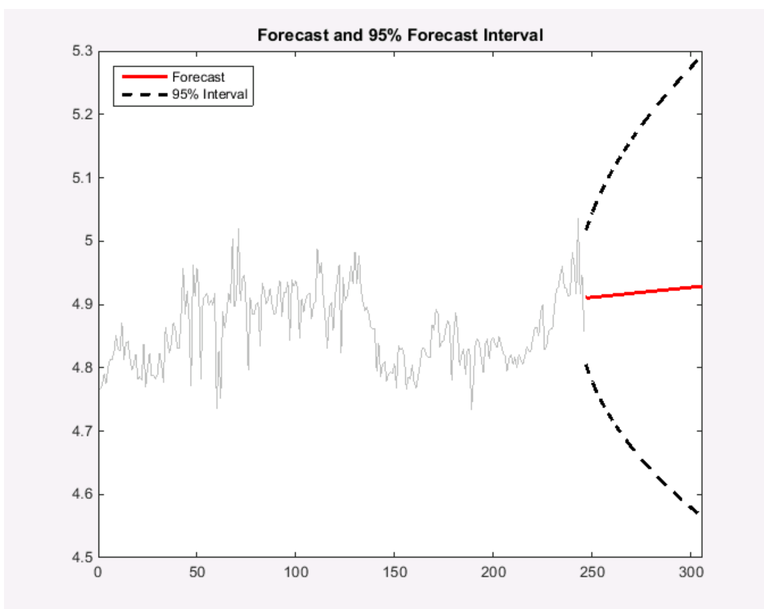


Figure 11

III. Conclusion

The ARIMA-GARCH model predicts a slight increase in the near future for SKEW. The predicted index value for July 2nd, 2016 is 135.802 with a 95% confidence interval of [122.242, 151.094].

The SKEW was at 125.35 on July 5th, 2016 (the next trading day after July 2nd). Though the drop was sharp, it was contained with the 95% confidence interval of the model. This drop may have been caused by market adjustments – traders settled down after the initial Brexit panic. The drop itself seems as steep of a difference as the climb on July 2nd, indicating that markets are quite volatile at the moment.

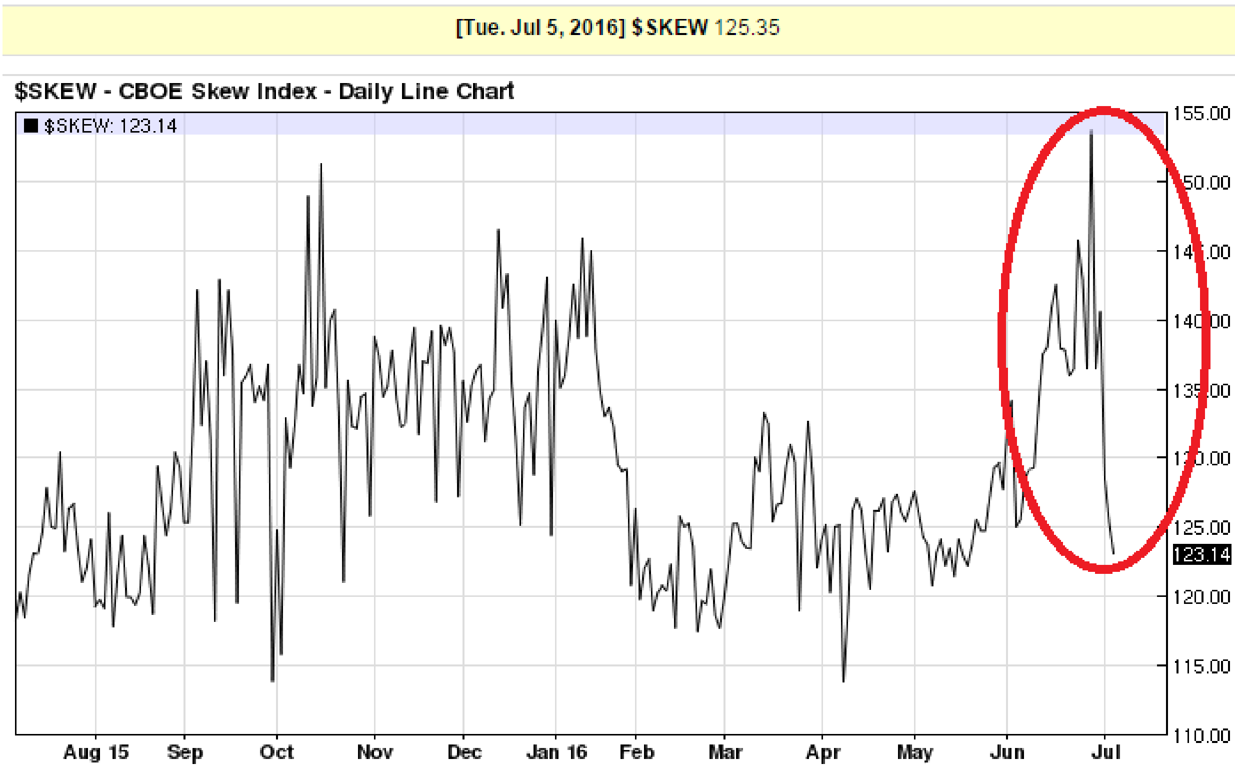


Figure 12

IV. Code

```
%% skew_arima
% Eddie Shim
% 7/5/16

clear

%load data

filename = 'CBOE_SKEW_INDEX_raw.csv';
data = load(filename);

y = log(data);
ydiff = diff(y);
T = length(data);

%Stationarity/Unit Root Tests
[h,pValue1] = adftest(y,'lags',1);
%[h2,pValue2] = kpsstest(ydiff,'lags',1, 'trend', false);
%[h3,pValue3] = lmctest(ydiff,'lags',1,'test','var2', 'trend', false);

%% ACF and PACF plots

figure
autocorr(ydiff)
title('ACF of log SKEW diff')
figure
parcorr(ydiff)
title('PACF of log SKEW diff')
hold off


%% Model Selection, BIC

for p = 1:4
    for q = 1:4
        mod = arima(p,1,q);
        [fit,~,logL] = estimate(mod,y,'print',false);
        LOGL(p,q) = logL;
        PQ(p,q) = p+q;
    end
end

LOGL = reshape (LOGL, 16,1);
PQ = reshape(PQ,16,1);
[~,bic] = aicbic(LOGL,PQ+1,100);
BIC = reshape(bic,4,4);
minBIC = min(BIC(:));
[p,q] = find(BIC == minBIC);    %finds row,column (p,q) of minimum loglikelihood value in matrix



%% create ARIMA model

Mdl = arima('ARLags',p,'D',1,'MALags', q, 'Variance', garch(1,1));
EstMdl = estimate(Mdl,y);

%plot

[yF,yMSE] = forecast(EstMdl,60,'Y0',y);
upper = yF + 1.96*sqrt(yMSE);
lower = yF - 1.96*sqrt(yMSE);

figure
plot(y,'Color',[.75,.75,.75])
hold on
h1 = plot(T+1:T+60,yF,'r','LineWidth',2);
h2 = plot(T+1:T+60,upper,'k--','LineWidth',1.5);
```

```
plot(T+1:T+60,lower,'k--','LineWidth',1.5)
xlim([0,T+60])
title('Forecast and 95% Forecast Interval')
legend([h1,h2],'Forecast','95% Interval','Location','NorthWest')
hold off



%% qqplot for ARIMA
% res = infer(EstMdl, y);
% stres = res/sqrt(EstMdl.Variance);
% figure
% plot(res)
% title('Standardized Residuals')
%
%
% figure
% subplot(1,2,1)
% qqplot(stres)
%
% x = -4:.05:4;
% [f,xi] = ksdensity(stres);
% subplot(1,2,2)
% plot(xi,f,'k','LineWidth',2);
% hold on
% plot(x,normpdf(x),'r--','LineWidth',2)
% legend('Residuals','Standard Normal')
% hold off



%% qqplot for ARIMA-GARCH
[res,v,logL] = infer(EstMdl,y);

figure
subplot(2,1,1)
plot(v)
xlim([0,T])
title('Conditional Variance')

subplot(2,1,2)
stres = res./sqrt(v);
plot(res./sqrt(v))
xlim([0,T])
title('Standardized Residuals')
hold off


figure
subplot(1,2,1)
qqplot(stres)

x = -4:.05:4;
[f,xi] = ksdensity(stres);
subplot(1,2,2)
plot(xi,f,'k','LineWidth',2);
hold on
plot(x,normpdf(x),'r--','LineWidth',2)
legend('Residuals','Standard Normal')
hold off

figure
plot(1:T,res+y,':', 1:T,y,'m')

%% Check residuals for autocorrelation
%
% figure
% subplot(2,1,1)
% autocorr(stres)
% subplot(2,1,2)
```

```
% parcorr(stres)
%
% [h,p] = lbqtest(stres,'lags',[5,10,15],'dof',[3,8,13])
```

MODEL OUTPUT:

```
    ARIMA(1,1,1) Model:
    --------------------
    Conditional Probability Distribution: Gaussian

                                  Standard          t
     Parameter       Value          Error       Statistic
    -----------   -----------   ------------   -----------
     Constant      0.00031986   0.000994793      0.321534
        AR{1}      -0.0200868      0.126318     -0.159018
        MA{1}       -0.563882      0.109876        -5.132

    GARCH(1,1) Conditional Variance Model:
    ----------------------------------------
    Conditional Probability Distribution: Gaussian

                                  Standard          t
     Parameter       Value          Error       Statistic
    -----------   -----------   ------------   -----------
     Constant     4.54461e-05   2.86123e-05       1.58834
     GARCH{1}        0.864417     0.0359348       24.0552
      ARCH{1}         0.11988     0.0386998       3.09769
```