NYU Stern: Forecasting Time Series Data
Professor Clifford Hurvich
Jasper Chang, 2020

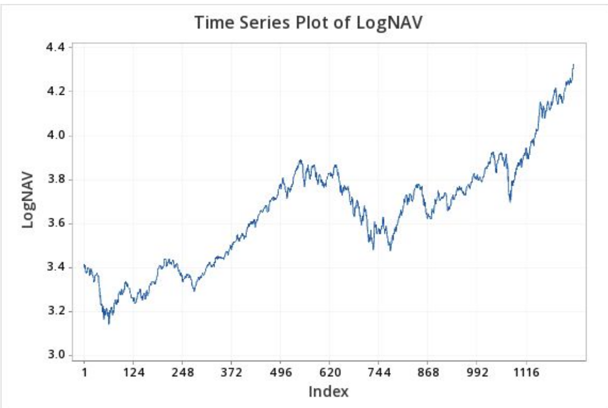
# JP Morgan China Fund

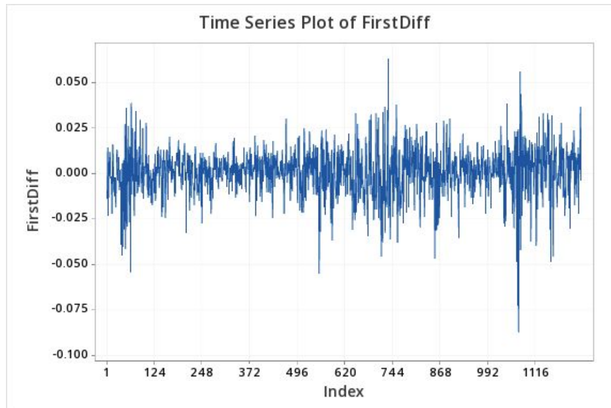
JPM (A) USD

# Forecasting Time Series: Final Project 2

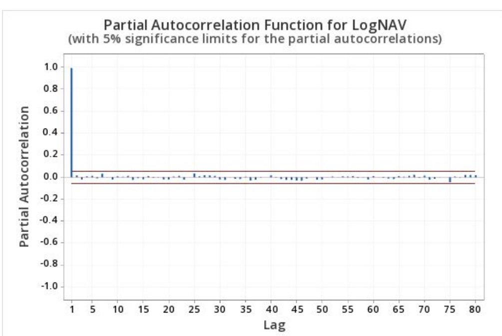
Data Set Used:
*JP Morgan (A) China Fund*

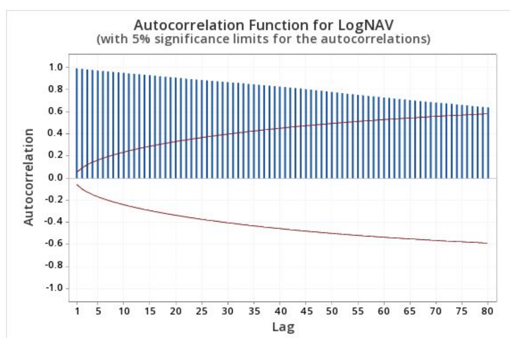
Data used is a 5 year time span, daily increments of closing NAVs from 11/9/2015 to 11/9/2020, data taken from JP Morgan, amounting to a total of 1236 data points: [Link]

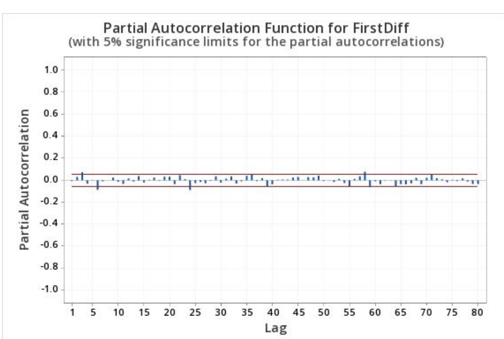


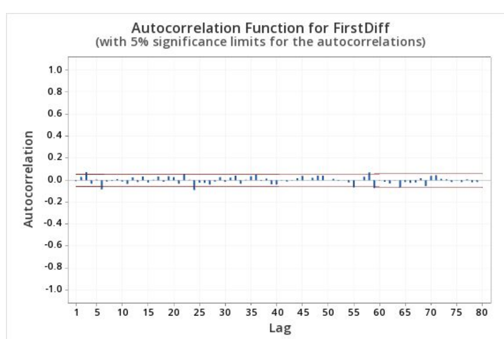




Preliminarily, the LogNAV is non-stationary and the FirstDiff (of the logs) is approximately stationary, with a lot of white noise but no clear signs for level-dependent volatility.



As per the ACFs and PACFs of both the first differences and the LogNAVs, a possible (3,1,3) ARIMA is predicted; this is despite the fact that there seems to exist significant spikes at lag 6.

| Without Constant | | | | | |
|---|---|---|---|---|---|
| N | p | d | q | SS | AICc |
| 1235 | 0 | 1 | 3 | 0.238019 | -10556.4428 |
| 1235 | 1 | 1 | 3 | 0.237679 | -10556.1919 |
| 1235 | 3 | 1 | 1 | 0.237693 | -10556.11916 |
| 1235 | 3 | 1 | 3 | 0.236991 | -10555.72953 |
| 1235 | 2 | 1 | 2 | 0.237825 | -10555.43351 |
| 1235 | 3 | 1 | 0 | 0.238282 | -10555.07893 |
| 1235 | 3 | 1 | 2 | 0.237646 | -10554.3438 |
| 1235 | 2 | 1 | 3 | 0.237706 | -10554.03203 |
| 1235 | 0 | 1 | 0 | 0.240018 | -10552.14325 |
| 1235 | 1 | 1 | 0 | 0.240014 | -10550.15734 |
| 1235 | 0 | 1 | 1 | 0.240015 | -10550.15219 |
| 1235 | 2 | 1 | 1 | 0.239301 | -10549.80878 |
| 1235 | 0 | 1 | 2 | 0.239693 | -10549.8004 |
| 1235 | 2 | 1 | 0 | 0.239726 | -10549.63038 |
| 1235 | 1 | 1 | 1 | 0.239844 | -10549.02263 |
| 1235 | 1 | 1 | 2 | 0.239512 | -10548.72032 |

| With Constant | | | | | |
|---|---|---|---|---|---|
| N | p | d | q | SS | AICc |
| 1235 | 0 | 1 | 3 | 0.237478 | -10559.25306 |
| 1235 | 1 | 1 | 3 | 0.237093 | -10559.24057 |
| 1235 | 3 | 1 | 1 | 0.237111 | -10559.14681 |
| 1235 | 3 | 1 | 3 | 0.236411 | -10558.75572 |
| 1235 | 2 | 1 | 2 | 0.237227 | -10558.54277 |
| 1235 | 3 | 1 | 0 | 0.237734 | -10557.92245 |
| 1235 | 2 | 1 | 3 | 0.237018 | -10557.61172 |
| 1235 | 3 | 1 | 2 | 0.237078 | -10557.29912 |
| 1235 | 0 | 1 | 1 | 0.239334 | -10553.66127 |
| 1235 | 1 | 1 | 0 | 0.239334 | -10553.66127 |
| 1235 | 1 | 1 | 2 | 0.238614 | -10553.3594 |
| 1235 | 0 | 1 | 2 | 0.239053 | -10553.10236 |
| 1235 | 2 | 1 | 0 | 0.23909 | -10552.91123 |
| 1235 | 2 | 1 | 1 | 0.238723 | -10552.79537 |
| 1235 | 1 | 1 | 1 | 0.239235 | -10552.16247 |
| 1235 | 0 | 1 | 0 | 0.243134 | -10536.21322 |

As per the AICc calculations, the best model is the (0,1,3) with a constant.

## Final Estimates of Parameters

| Type | Coef | SE Coef | T-Value | P-Value |
|---|---|---|---|---|
| MA 1 | 0.0045 | 0.0284 | 0.16 | 0.875 |
| MA 2 | -0.0328 | 0.0284 | -1.16 | 0.248 |
| MA 3 | -0.0898 | 0.0285 | -3.15 | 0.002 |
| Constant | 0.000742 | 0.000442 | 1.68 | 0.093 |

From this model, these are the final parameters:

$$x_t - x_{t-1} = \varepsilon_t - 0.0045\varepsilon_{t-1} + 0.0328\varepsilon_{t-2} + 0.0898\varepsilon_{t-3} + 0.000742$$

Assuming:

$X_t$ here is the price of the fund at time t.

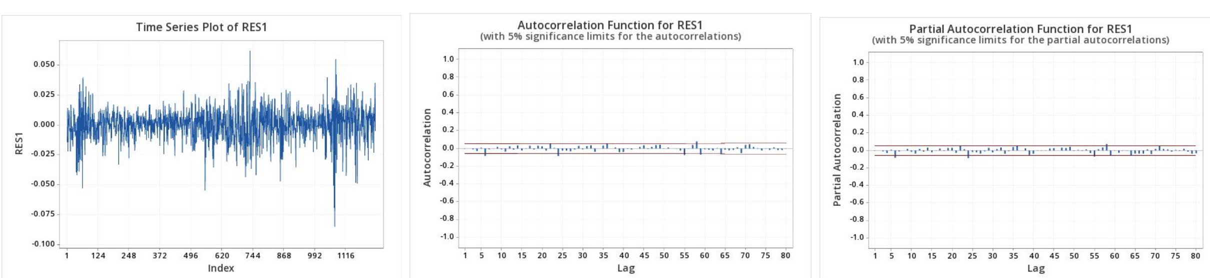
---

## Forecasts from period 1235

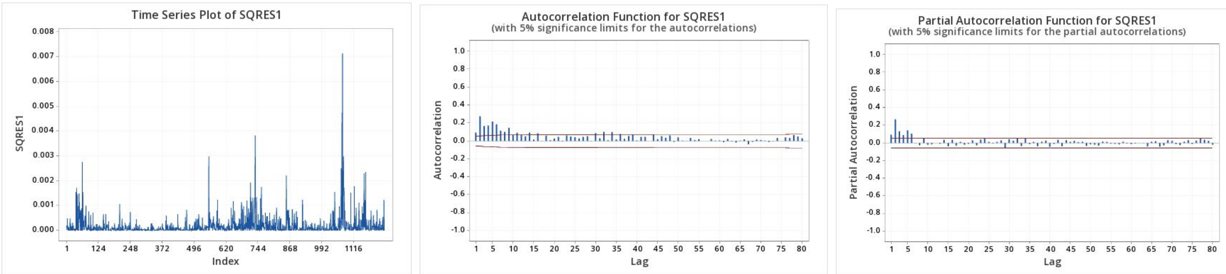
| | | 95% Limits | | |
|---|---|---|---|---|
| Period | Forecast | Lower | Upper | Actual |
| 1236 | 4.30541 | 4.27818 | 4.33264 | 4.32757 |

$f_{1235,1} = 4.30541$

As shown here, the forecast from 1235 for period 1236 is not entirely accurate, slightly understating the actual value (albeit falling within the 95% confidence interval).




^ Time Series/ACF/PACF of RES1




^ Time Series/ACF/PACF of SQRRES1 (squared)

Residuals are approximately uncorrelated and squared residuals are not uncorrelated with each other, all in all indicating that residuals are conditionally heteroskedastic.

| Model | q | LogLikelihood | AICc |
|---|---|---|---|
| ARCH(0) | 0 | 3531.254 | -7060.504756 |
| ARCH(1) | 1 | 3541.413 | -7078.81626 |
| ARCH(2) | 2 | 3565.955 | -7125.890504 |
| ARCH(3) | 3 | 3592.192 | -7176.35148 |
| ARCH(4) | 4 | 3605.441 | -7200.83318 |
| ARCH(5) | 5 | 3611.99 | -7211.911596 |
| ARCH(6) | 6 | 3610.565 | -7207.03872 |
| ARCH(7) | 7 | 3608.5 | -7200.882545 |
| ARCH(8) | 8 | 3607.18 | -7196.213061 |
| ARCH(9) | 9 | 3606.605 | -7193.030261 |
| ARCH(10) | 10 | 3604.074 | -7185.932137 |
| GARCH(1,1) | 2 | 3630.3 | -7254.580504 |

```
> model$coef
          a0           a1           b1
2.854161e-06 7.119571e-02 9.150364e-01
```

According to the AICc process of selecting the best model for the residuals (on R), the GARCH(1,1) model is the most suitable.

GARCH(1,1):
$h_{t+1}$=2.854e-06 + 7.12e-02$\varepsilon_t$ + 9.15e-01 $h_t$

Unconditional variance as per a0/[1-(a1+b1)]:
0.0002073

---

*The GARCH(1,1) summary are as follows:*

```
Call:
garch(x = x, order = c(1, 1), trace = F)

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max
-4.86662 -0.56496  0.04321  0.66558  3.10113

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 2.854e-06   9.629e-07    2.964  0.00304 **
a1 7.120e-02   9.584e-03    7.428  1.1e-13 ***
b1 9.150e-01   1.181e-02   77.502  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 134.73, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 2.5863, df = 1, p-value = 0.1078
```

The one-step forecast, aforementioned and calculated, was $f_{1235,1} = 4.30541$.

$H_{1236} = 2.854e\text{-}06 + 7.12e\text{-}02\varepsilon_t^2 + 9.15e\text{-}01h_{1235}$
$= 2.854e\text{-}06 + 7.12e\text{-}02*0.000742^2 + 9.15e\text{-}01*0.0002173352$
$= 0.0002017549$

Lower Bound $= f_{1235,1} - |z_{0.025}| * (h_{1236})^{0.5} = 4.30541 - 1.96 * 0.0002017549^{0.5} = 4.27757$
Upper Bound $= f_{1235,1} + |z_{0.025}| * (h_{1236})^{0.5} = 4.30541 + 1.96 * 0.0002017549^{0.5} = 4.33325$

Combined, this is a 95% confidence forecast interval of (4.27757,4.33325)

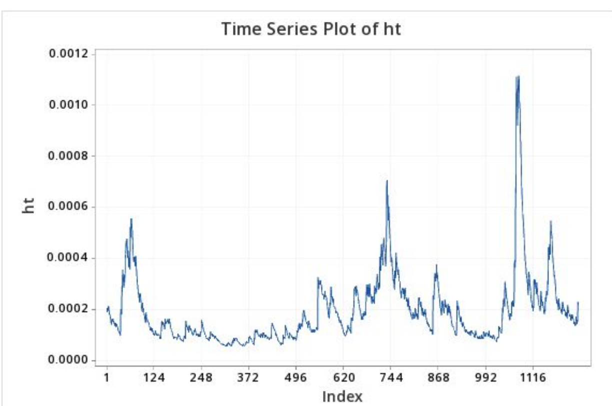
---

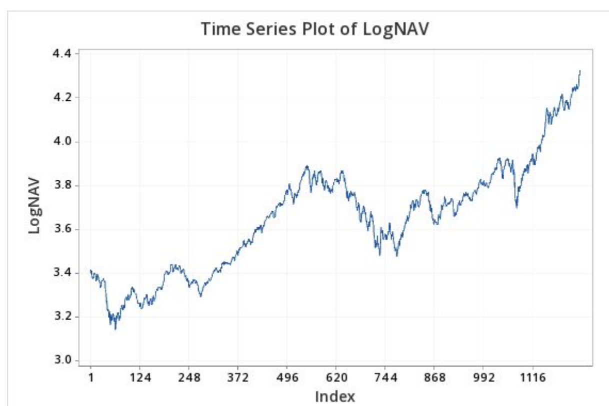
In comparison to the previous ARIMA(0,1,3) 95% prediction interval of (4.27818,4.33264), the one with an added GARCH(1,1) has a forecast interval of (4.27757,4.33325), which is actually wider than the ARIMA(0,1,3) by 0.00122. This means that the volatility now is slightly higher than the 5 year average.

The 5% percentile of the conditional distribution of the next period:
$Q_{5\%} = f_{1235,1} - |z_{0.05}| * (h_{1236})^{0.5} = 4.30541 - 1.64 * (0.0002017549)^{0.5} = 4.282115$

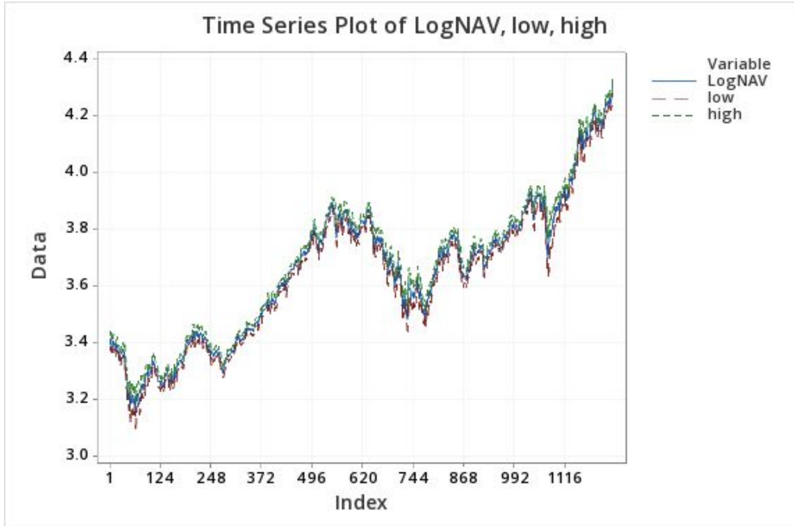
With the current price of 75.76 (NAV) and a natural log of 4.32757, it effectively falls within the 95th confidence prediction interval for both the ARIMA(0,1,3) and the ARIMA(0,1,3)+GARCH(1,1). This signifies that the range can fully encapsulate the data, but the volatility is an issue, as it falls close to the upper bound (might be a little too narrow).

After exporting the conditional variances into Minitab:

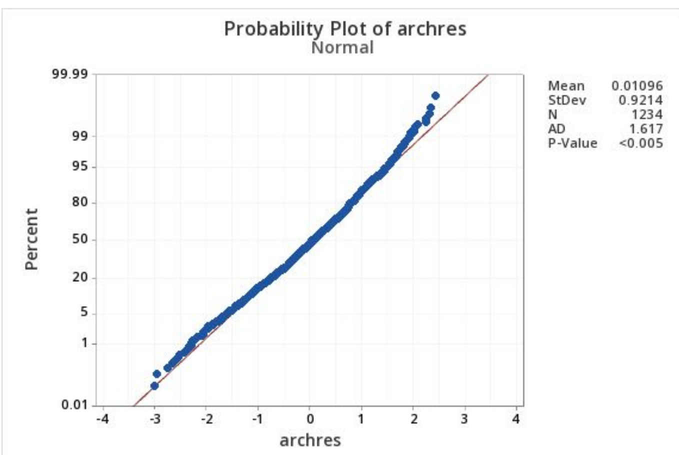



In comparison to the LogNAVs, the periods of higher volatility generally match for both time series plots, especially in regards to point 60, 744 and 1100. The volatilities do seem a bit more pronounced in the conditional variance graph, which is to be expected, as it measures the higher bursts of volatility.




^ Plot of LogNAV with ARIMA-GARCH one-step 95% forecast intervals




As shown by the shape of the probability plot for ht, the residuals for the ARIMA-ARCH model are almost entirely normal (with noticeable light tails at the top), indicating that leptokurtosis is explained well, to a certain degree.

The small p-value claims rejection against the null hypothesis.

There are a total of **41 errors**, amounting to a **3.3% probability of failure.** All in all, this maintains the hypothesis that the ARIMA-ARCH model is a good fit and predictor for this data set.