Abstract

The price of our daily commodities is closely related to our lives. This passage establishes two statistical models to fit pork and fish price series.

In the first part, I attempt to figure out the factors affecting pork price. Taken into consideration of PPI, CPI, fish price, chicken price, time and made variables selection, I build a regression model “Pork = 2896.41 – 1.45*Time + 13.35*Chicken + 0.45*PPI”. By analyzing the residuals, I fitted the autocorrelated error using ARIMA(p=2, d=0, q=1, P=2, D=0,Q=0)$_{12}$. Finally, the regression model became “Pork = 2973.27 -1.49*Time + 10.53*Chicken + 0.53*PPI + ARIMA(p=2, d=0, q=1, P=2, D=0,Q=0)$_{12}$”. I found the residual of the final model is stationary and it likes white noise by residual analysis.

In the second part, I established SARIMA model on the fish price series. After log transformation and differencing under the instruction of ACF and PACF plots, I finally got SARIMA(p=2, d=1,q=1,P=1,D=0,Q=0)$_{12}$ **[**$(1-0.18B^{12})(1-1.19B+0.36B^2)(1-B)\log(X_t) = 0.008 + (1-0.92B)W_t$**]** to forecast the fish price in the future. I found the residual of the final model tends to be stationary, however, not all the p-values for Ljung-Box Test are greater than 0.05. In general, it is often the case that we cannot perfectly fit a time series due to the complicated factors in real life. The model applied, forecasted prices of fish in next 5 month following October 2021 is $7.06, $7.14, $7.10, $7.07, $7.24 per kilogram.

Introduction

The meat industry is the most important economic activity in the agro-food sector. The pork meat is the most consumed. In this project, in the first part, my goal is to forecast the price of pork based on other economic indicators. The background of this idea is that I found the price of pork is related to PPI (Producer Price Index), the price of complementary products and the price of alternatives in China. Hence, I want to figure out whether it is the same case in the United States. In the second part, I fitted a SARIMA model to forecast the price of fish simply because I like eating fish very much and want to keep track of its price.

For data preparation, I downloaded PPI data from data achiever (https://data.bls.gov/cgi-bin/srgate) using “WPSID621” as input “Series ID”, which represents PPI Commodity data for seasonally adjusted Intermediate demand by commodity type-Unprocessed foodstuffs and feedstuffs; CPI data from https://data.bls.gov/PDQWeb/cu using “CUSR0000SAF1”, which represents Food in U.S. city average, all urban consumers, seasonally adjusted; monthly pork price from https://www.indexmundi.com/commodities/?commodity=pork; monthly chicken price from https://www.indexmundi.com/commodities/?commodity=chicken; monthly fish price from https://www.indexmundi.com/commodities/?commodity=fish. After achieving all the data, I truncated them all to the period from October 1991 to October 2021.

Statistical Methods

Part A: Regression analysis:

Whole display of different time series:

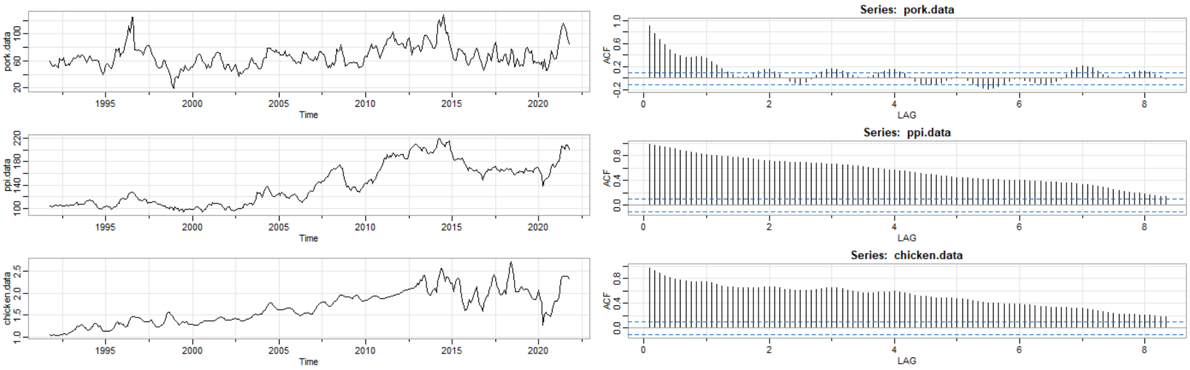


Fig. 1. Time series plot of pork price, PPI and chicken price (left) and ACF plots of them respectively(right)

The price of pork shows a cyclic pattern, the variance of the first half of data is 217.9466 and second half is 282.0113, hence, the variance increasing. So, the variance is not constant. Meanwhile, the ACF plot of pork price also shows that pork price is not stationary.

Time series of PPI data and chicken price show increasing trends before 2012 but fluctuate afterwards. The ACF plots imply there exists seasonal pattern. The variance of the first half of PPI data is 90.15 and second half is 531.0857; the variance of the first half of chicken price data is 0.02958304 and second half is 0.06247425. So, the variance increases. Therefore, the series is not stationary.

Here comes model fitting. At first, I came up with several possible models:

1. Pork ~ Time + PPI
2. Pork ~ Time + Chicken
3. Pork ~ Time + Fish
4. Pork ~ Time + Chicken + PPI
5. Pork ~ Time + Fish + PPI
6. Pork ~ Time + Chicken + PPI + Fish
7. Pork ~ Time + Chicken + PPI + Fish + CPI
8. Pork ~ Time + $(\text{Time})^2$ + Chicken + PPI + Fish + CPI

Table 1

Summary of basic fitting information of different models.

Models	1	2	3	4	5	6
R-square	0.504	0.3304	0.1223	0.5192	0.5043	0.5195
All coefficients are significant?	YES	YES	YES	YES	YES	NO
Models	7	8				
R-square	0.5296	0.5396				
All coefficients are significant?	NO	YES				

Comparing with model4 and model8, although model8 has higher R-square, the coefficient of fish is significant under significance level 0.05 and all the coefficients under model4 is significant under significance level 0. On top of that, model8 introduce extra three predictors compared with model4. Additionally, with too many predictors, there will be collinearity problems among predictors. In view of the R-square does not improve too much, I choose model4 at this step, namely, “Pork ~ Time + Chicken + PPI”.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2896.4123   282.0221  10.270  < 2e-16 ***
trend          -1.4531     0.1429 -10.170  < 2e-16 ***
chicken.data   13.3507     3.9723   3.361 0.000861 ***
ppi.data        0.4571     0.0386  11.840  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.73 on 357 degrees of freedom
Multiple R-squared:  0.5192,	Adjusted R-squared:  0.5152 
F-statistic: 128.5 on 3 and 357 DF,  p-value: < 2.2e-16
```

Fig. 2. Result of linear regression

With the coefficients obtained plugged in, I got “Pork = 2896.41 – 1.45*Time + 13.35*Chicken + 0.45*PPI”.

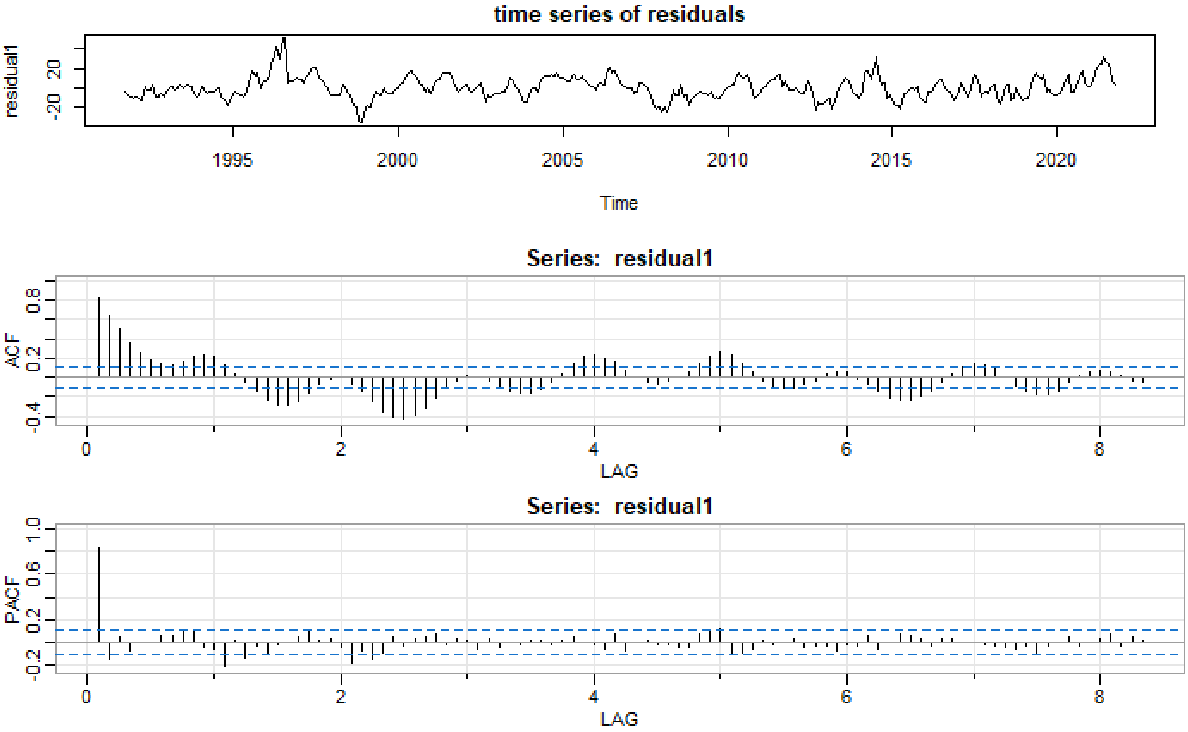


Fig. 3. Residual plot and its ACF and PACF

From the residual itself and its ACFs and PACFs, the previous model has obvious autocorrelated errors.

I use ARIMA model to fit the model. Since the ACFs display a seasonal pattern with cycle of 12.

Seasonal part: ACFs tail off and PACFs cut off.

Non-seasonal part: ACFs tail off and PACFs cut off/tail off.

I came up with four models to fit residuals:

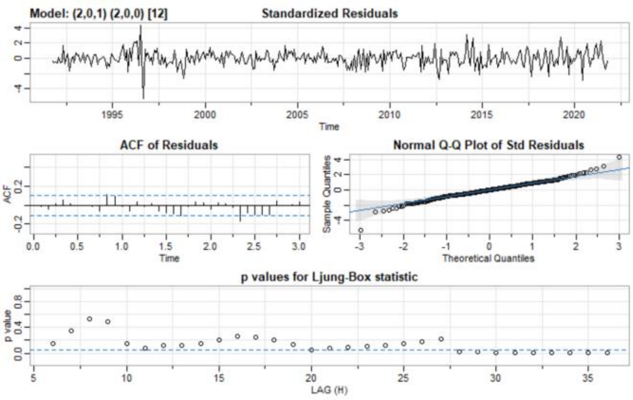
1. p=1, d=0, q=1, P=2, S=12
2. p=2, d=0, q=1, P=2, S=12
3. p=1, d=0, q=0, P=2, S=12
4. p=2, d=0, q=0, P=2, S=12

Table 2

Residual analysis and ttable of fitted models

Models	1	2	3	4
Significance of coefficients in ttable	Coefficient of chicken is not significant	Almost all the coefficients are significant under significance level of 0.05	Coefficient of chicken is not significant	Coefficients of chicken and AR2 are not significant
AIC	6.554975	6.549886	6.564892	6.559906
BIC	6.651927	6.657611	6.651072	6.656859
Standardized residuals	Like white noise	Like white noise	Like white noise	Like white noise
ACF of residuals	Inside blue lines	Inside blue lines	Inside blue lines	Inside blue lines
QQ plot	Normality	Normality	Normality	Normality
p-value for Ljung-Box statistic	Only some points are above the blue line.	Most of the points are above the blue line with some points far above blue line.	No point is above the blue line.	Only few points are above the blue line.

From the summary table above, in this step, I choose model2. By now, the overall model is “Pork ~ Time + Chicken + PPI + ARIMA(p=2, d=0, q=1, P=2, D=0,Q=0)$_{12}$”.




```
$ttable
              Estimate       SE t.value p.value
ar1             0.3467   0.1762  1.9670  0.0500
ar2             0.3719   0.1555  2.3919  0.0173
ma1             0.6155   0.1573  3.9139  0.0001
sar1            0.1813   0.0532  3.4048  0.0007
sar2            0.1687   0.0542  3.1133  0.0020
intercept    2973.2699 752.8503  3.9494  0.0001
trend          -1.4939   0.3783 -3.9494  0.0001
chicken.data   10.5304   5.4183  1.9435  0.0528
ppi.data        0.5266   0.0770  6.8419  0.0000
```

Fig. 4. Residual analysis and ttable of the ARIMA model for residuals

- The residual plot shows that after refit the correlated errors, the residuals are stationary.

- The ACF of residuals shows it is a white noise, since almost all spikes are within the blue line.

- Q-Q plot shows we cannot reject normality assumptions

- Ljung-Box statistic shows most of the p-values are above the blue line, which means it likes a white noise.

- The regression model combines with ARIMA model to fit the residuals is good.

Part B: SARIMA model

In this part, I want to determine the price of fish with time goes by.

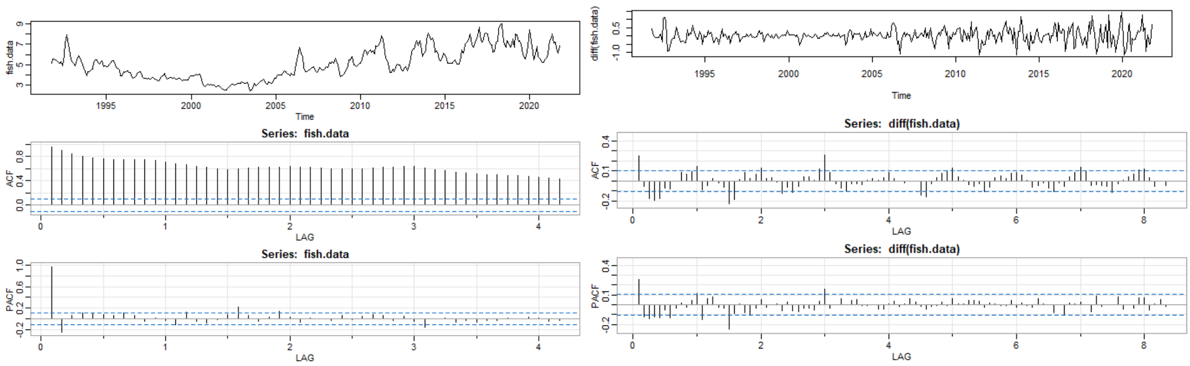


Fig. 5. Time series of fish and diff(fish) and their ACFs and PACFs

The fish price first went through a period of price going down and then had an increasing trend. The variance of the first half of data is 1.003637 and second half is 1.49142, hence, the variance increases a little bit. ACF plot of the fish price implies that there is a seasonal pattern. Since the ACF decays slowly, I take differencing once.

The series after differencing is more like a white noise, although the variance seems small between 1995 and 2005. To minimize the disparity in variance, I first took logarithm of fish price and then perform differencing.

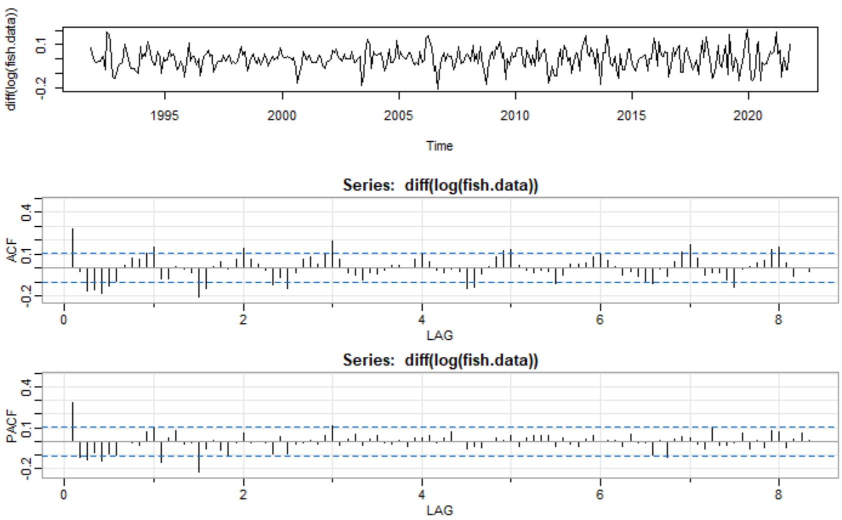


Fig. 6. Differencing log(fish price) and its ACF and PACF

After transforming the data, the series is stationary somehow. And according to ACF and PACF plots,
Seasonal part: ACF tails off with a cycle of 12, PACF cuts off
Non-seasonal part: ACF tails off, PACF cuts off/tails off
I came up with 6 model to fit log(fish) data:

1. SARIMA(p=1, d=1,q=0,P=1,D=0,Q=0,S=12)
2. SARIMA(p=2, d=1,q=0,P=1,D=0,Q=0,S=12)
3. SARIMA(p=1, d=1,q=1,P=1,D=0,Q=0,S=12)
4. SARIMA(p=1, d=1,q=1,P=2,D=0,Q=0,S=12)
5. SARIMA(p=2, d=1,q=1,P=1,D=0,Q=0,S=12)
6. SARIMA(p=2, d=1,q=1,P=2,D=0,Q=0,S=12)

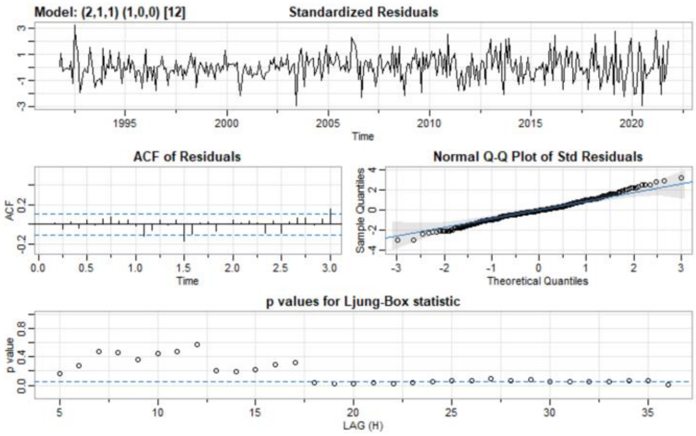
Table 3

Residual analysis and ttable of fitted models

Models	1	2	3	4	5	6
Significance of coefficients in ttable	All the coefficients are significant	All the coefficients are significant	AR1 is not significant	AR1, MA1, SAR2 are not significant	All the coefficients are significant	SAR2 is not significant
AIC	-2.553086	-2.562888	-2.555	-2.558583	-2.607335	-2.607947
BIC	-2.509907	-2.508914	-2.501026	-2.493815	-2.542567	-2.532384
Standardized residuals	Like white noise	Like white noise	Like white noise	Like white noise	Like white noise	Like white noise
ACF of residuals	Inside blue lines	Inside blue lines	Inside blue lines	Inside blue lines	Inside blue lines	Inside blue lines
QQ plot	Normality	Normality	Normality	Normality	Normality	Normality
p-value for Ljung-Box statistic	No points is above the blue line.	No points is above the blue line.	No points is above the blue line.	No points is above the blue line.	Some of the points at the beginning are above the blue line.	Some of the points at the beginning are above the blue line.

From the summary table above, there is no doubt that we will not choose model1 to model4 in consideration of no points above the bule line. As for model6, though its AIC is a little smaller than model5, however, not all its coefficients are significant and its p-value for Ljung-Box statistic is only a little above blue line.

Therefore, I picked model5, namely, SARIMA(p=2, d=1,q=1,P=1,D=0,Q=0)$_{12}$.




```
$ttable
         Estimate     SE  t.value p.value
ar1        1.1879 0.0601  19.7745  0.0000
ar2       -0.3648 0.0499  -7.3059  0.0000
ma1       -0.9228 0.0419 -22.0058  0.0000
sar1       0.1831 0.0564   3.2482  0.0013
constant   0.0008 0.0019   0.4280  0.6689
```

Fig. 7. Residual analysis and ttable of model5

Specifically, the model is $(1-0.18B^{12})(1-1.19B+0.36B^{2})(1-B)\log(X_t) = 0.008 + (1-0.92B)W_t$. Using this fitted model, I forecast future 5 values of fish price (round to 2 decimals):

Table 4

Future 5 values forecasted using SARIMA model

Month	Nov. 2021	Dec. 2021	Jan. 2022	Feb. 2022	Mar. 2022
price($/kilogram)	7.06	7.14	7.10	7.07	7.24

Results

- Pork = 2973.27 -1.49*Time + 10.53*Chicken + 0.53*PPI + ARIMA(p=2, d=0, q=1, P=2, D=0,Q=0)$_{12}$

In part A, I tried to fit pork price series using PPI, time, and chicken price. The regression model shows that the pork price is positively correlated with chicken price, which is against our intuitive. Since chicken and fish are alternatives, their price change should go in different directions. Moreover, the increase of PPI is a signal of price increasing, which shows the price will go up when producers are of high enthusiasm to produce. After fitting the regression model, I detected an autocorrelation in errors, therefore I fitted ARIMA model to the errors.

- SARIMA(p=2, d=1,q=1,P=1,D=0,Q=0)$_{12}$ **[**$(1-0.18B^{12})(1-1.19B+0.36B^{2})(1-B)\log(X_t) = 0.008 + (1-0.92B)W_t$**]**

In part B, I tried to fit fish price using SARIMA model. The SARIMA model shows the fish price has a seasonal pattern. Because the price will be influenced by both demand and supply. When the price goes up, more producers start raising fish, which increases supplies, and will in turn lower the price gradually.

The results I got can provide instructions to restaurants, each individual and whoever needs to buy pork and fish. They can make purchase when the price is low, which will reduce the cost a lot.

Discussion

This passage discussed the covariates of meat price and how we can use the models to predict future prices. And I also discuss how the price change seasonally. However, one improvement can be made through adding lagged term (e.g., lag (PPI, -1)). Since the lagged term might have higher correlation to pork price, in other words, some of the factors might change ahead of the change of price. Another improvements can be made by adding more predictors to explain more variation of the change of price. One big limitation of ARIMA model I constructed is that the residual in the model is not exactly the white noise, which means there is still some information hidden but I have not found them out.