HACETTEPE UNIVERSITY

FACULTY OF ECONOMICS & ADMINISTRATIVE SCIENCES

DEPARTMENT of ECONOMICS

ECO 346

Econometrics II

Term Project

S. Duru (2021), "Investigate of the Effect of Food Inflation on the Agricultural and Food Products Export Using Multiple Linear Regression Analysis in Turkey "

Student's Name- Student ID

Gözde Okuyucu-21941914

Doğaçhan Dülger- 21941597

ANKARA

MAY 2024

Duru S. (2021), "Investigate of the Effect of Food Inflation on the Agricultural and Food Products Export Using Multiple Linear Regression Analysis in Turkey "

Introduction

Our reference article examines the export of agricultural products and its determinants, such as the inflation rate in food, the Producer Price Index (PPI), the agricultural producer index, and the food and beverages retail sale index. In our homework, we aimed to apply some of the econometrics techniques we learned in class to USA data, using this article as a reference. First, we reviewed the reference article thoroughly. Then, we searched for relevant data for a new regression analysis, focusing on the USA from 1991 to 2023. Using Gretl, we applied the econometrics methods we had learned to the data we collected. We also included our results and commentary in our homework. Before detailing our applications, here is a review of the article “Determinants of Export of Agricultural Products in the USA.”

Review of Article

The term paper dives into how food prices affect farm exports, exploring economic factors shaping food costs and trade dynamics. It looks at data from 2010 to 2019, using a regression model to study how inflation affects farm and food exports.The paper stresses the importance of farm exports for underdeveloped countries but notes the challenge of rising food prices due to faster population growth than production.

It highlights various factors behind food inflation, like supply-demand shifts, natural disasters, and currency changes. It also mentions how poor organization and higher production costs contribute to inflation. The paper cites studies showing that while food costs can drive up inflation, farm exports tend to withstand changes in exchange rates and local prices.

In the material and method section of the study, it is stated that a multiple linear regression model was applied to determine the statistical relationship between variables. In this model, the relationship between a single dependent variable and more than one independent variable is examined. The multiple linear regression analysis model used in the model is given in equation 1:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \varepsilon \ \ldots\ldots (1)$$

The symbols used in the multiple linear regression model in Equation 1 are;

Y= Exports of agricultural and food products (dependent variable) (TGUI)

X_1= Year-end food and non-alcoholic beverages inflation rate (GAIEO)

X_2= Producer price index (PPI)

X_3= Agriculture PPI

X_4= Food, beverage and tobacco products retail sales volume index

(β_1, β_2, β_3 and β_4 = Parameters of the variables to be estimated, β_0= constant term, ε= error term (random))

In Formula 1, X_1, X_2, X_3 and X_4 were accepted as independent variables and the effects of these variables on the export of agricultural and food products were examined. The error term ε is random and represents the variables not included in the model. The exchange rate isn't included in the model because it's highly correlated with the producer price indices, which rely on imported goods and services.

Default assumptions of the multiple linear regression model: The error term (ε) should show a normal distribution. The expected mean value of the errors should be 0. The number of observations (n) should be at least 5 times the number of independent variables (ideally 20 times). There should be a linear relationship between the variables. There should not be a high correlation between error terms (ε) and explanatory variables. Variance Inflation Value (VIF) should be less than 10 to avoid multicollinearity problems.

The multiple correlation coefficient (R) indicates the strength of the relationship between Y and X's, while the coefficient of determination (R^2) shows the proportion of variance in the Y explained by theX's. Both coefficients range between 0 and +1, higher values stronger relationships and better explanatory power, approaching +1. (f^2) represents the ratio of explained variance to unexplained variance. It's calculated using Equation 2:

$$f^2 = R^2 (1 - R^2) -1$$

Values between 0.02 and 0.15 for f^2 indicate small effects, 0.15 to 0.35 suggest medium effects, and values ≥ 0.35 represent large effects. For the coefficient of determination (R^2), values between 0.02 and 0.13 are small, 0.13 to 0.26 are medium, and values ≥ 0.26 are large effect sizes.

Using data from TURKSTAT, the paper rigorously analyzes the relationship between different factors, boosting the credibility of its findings.In the discussion, it links its results to Turkey's economy, showing how high inflation hurts consumers and leads to policy changes like cutting agricultural subsidies and giving direct support to farmers.

The paper suggests ways to tackle food inflation, like boosting local food production and offering incentives for exports. These measures aim to stabilize food prices and support farm exports.

Article Results

In the paper author mentions that inflation of agricultural goods has affected consumers negatively. In order to stabilise these volatile prices government has taken some actions. Like signing a stand-by agreement with IMF. Then the committee of tracking and evaluating the food and agricultural products has been founded to assess the situation correctly and then create appropriate policies. Then the author mentions that the inflation of agricultural products has external factors. Such factors are the risk of volatile exchange rate, interest and inflation in economy as a whole.

During the research, the author used a data set from 2010-2019 and converted into a monthly data set and found that in the 10-year period the export of agricultural goods and food products has been increased by almost 50%. However, the producer price index (PPI) has increased more than the food products inflation. The reason for this may be the PPI gets effected by variety of other external factors. Also, agricultural PPI got doubled in the same period because of the fixated input prices to exchange rate.

Before making the analysis, the author tested for autocorrelation between variables and found that such problem does not exist. In the variance analysis, the author found that model as a whole is significant as in $F_{stat}>F_{crit}$ ($F_{stat}=13.248$; $p<0.001$) when tested individually with $p<0.10$, author finds that X_1 , X_2 and X_4 to be significant but X_3 to be not significant. After estimating the regression, the author finds the R^2 to be 0.32 which means the regression can only explain about 32% of variations in Y is explained by the variations in Xs. Which means data about inflation has a weak explanatory power about export data. In other words, the inflation is not a highly effective macroeconomic variable about explaining the export of agricultural and food products.

Our Regression Analysis

For our regression analysis, we tried to do the same regression analysis method into different data set. Our aim for changing the data set is to find if the data set has changed, can we get a better explanation about effects of inflation in export of agricultural and food products. So, we changed the data set to span 1991-2023 and made it annual so we can properly see the effect of inflation in agricultural and food products. Lastly, we chose USA as our country of analysis because it is one of the largest exporters and has a relatively stable inflation figures.

We defined our regression as:

$$Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 X_4 + u$$

Y= Export of Agricultural Products

X_1 = Inflation Rate in Food and Beverages

X_2 = Producer Price Index

X_3 = Agricultural Producer Price Index

X_4 = Food Beverages and Tobacco Goods Retail Sale Volume Index

u = error term

At first, we did an OLS estimation.

```
gretl: model 1
File  Edit  Tests  Save  Graphs  Analysis  LaTeX
Model 1: OLS, using observations 1991-2023 (T = 33)
Dependent variable: export_of_agricultural_products

                       coefficient    std. error   t-ratio  p-value
  --------------------------------------------------------------------
  const               -84.3956       13.0660       -6.459   5.37e-07 ***
  inflation_rate_i~     1.62478       0.644110      2.523   0.0176   **
  producer_price_i~     0.824946      0.316184      2.609   0.0144   **
  agricultural_pro~    -0.133625      0.264943     -0.5044  0.6180
  Food_beverages_a~    -0.000461183   0.000229751  -2.007   0.0545   *

Mean dependent var   100.2581   S.D. dependent var   48.73995
Sum squared resid    3380.523   S.E. of regression   10.98786
R-squared            0.955530   Adjusted R-squared   0.949178
F(4, 28)             150.4107   P-value(F)           1.70e-18
Log-likelihood      -123.2081   Akaike criterion     256.4161
Schwarz criterion    263.8987   Hannan-Quinn         258.9338
rho                  0.732325   Durbin-Watson        0.628479

Excluding the constant, p-value was highest for variable 4 (agricultural_producer_index)
```

Since the R^2 is 0.96, variations in Xs can explain about 96% of variations in Y. Which means model as a whole has a very high explanatory power. However, we stumbled the same insignificance problem in X_3, which is Agricultural Producer Price Index.

```
gretl: LM test (heteroskedasticity)
Breusch-Pagan test for heteroskedasticity
OLS, using observations 1991-2023 (T = 33)
Dependent variable: scaled uhat^2

                       coefficient   std. error   t-ratio  p-value
  -----------------------------------------------------------------
  const               -3.39941      1.57215      -2.162   0.0393   **
  inflation_rate_i~   -0.0118152    0.0775015    -0.1525  0.8799
  producer_price_i~   -0.0450840    0.0380443    -1.185   0.2460
  agricultural_pro~    0.0473995    0.0318788     1.487   0.1482
  Food_beverages_a~    5.18621e-05  2.76444e-05   1.876   0.0711   *

  Explained sum of squares = 23.5684

Test statistic: LM = 11.784212,
with p-value = P(Chi-square(4) > 11.784212) = 0.019030
```

Secondly, we checked for the heteroskedasticity problem in our model. We set the hypothesis as

H_0 = There isn't heteroskedasticity problem

H_1 = There is heteroskedasticity problem

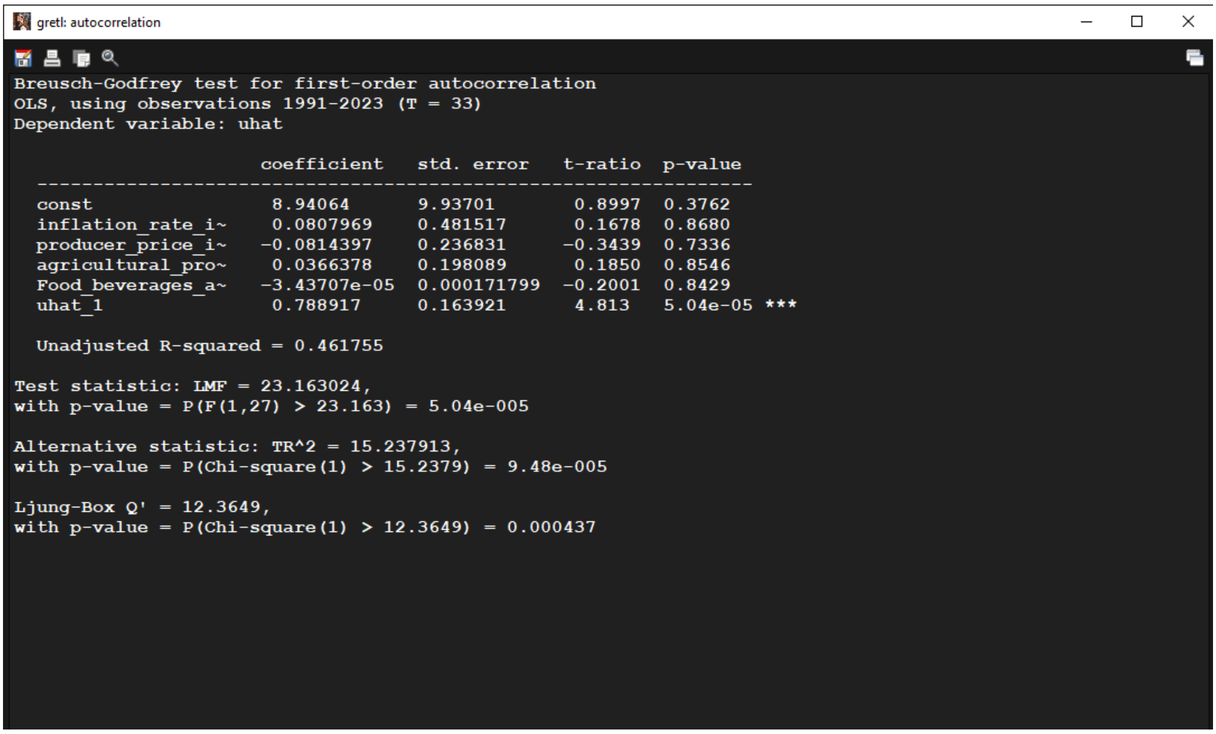
The results show a p value less than 0.10. So, we reject the null hypothesis. This means there is heteroskedasticity problem. Which in turn affects our efficiency while making further analysis. However, we can do a heteroskedasticity co rrected test and get these results.

```
gretl: model 4
File  Edit  Tests  Save  Graphs  Analysis  LaTeX
Model 4: Heteroskedasticity-corrected, using observations 1991-2023 (T = 33)
Dependent variable: export_of_agricultural_products

                       coefficient   std. error   t-ratio   p-value
  -----------------------------------------------------------------
  const               -69.1263       7.38232      -9.364    4.02e-010 ***
  inflation_rate_i~     1.09512      0.487839      2.245    0.0329    **
  producer_price_i~     0.521791     0.215569      2.421    0.0222    **
  agricultural_pro~     0.458860     0.174606      2.628    0.0138    **
  Food_beverages_a~    -0.000301169  0.000152591  -1.974    0.0584    *

Statistics based on the weighted data:

Sum squared resid    47.65212   S.E. of regression   1.304554
R-squared            0.981784   Adjusted R-squared   0.979182
F(4, 28)             377.2776   P-value(F)           6.53e-24
Log-likelihood      -52.88740   Akaike criterion     115.7748
Schwarz criterion    123.2573   Hannan-Quinn         118.2924
rho                  0.810423   Durbin-Watson        0.563077

Statistics based on the original data:

Mean dependent var   100.2581   S.D. dependent var   48.73995
Sum squared resid    4165.658   S.E. of regression   12.19727
```

Lastly, we checked if our model has an autocorrelation problem or not. So, we set our hypotheses as follows:

H_0 = There isn't autocorrelation problem

H_1 = There is autocorrelation problem



gretl: autocorrelation

```
Breusch-Godfrey test for first-order autocorrelation
OLS, using observations 1991-2023 (T = 33)
Dependent variable: uhat

                      coefficient    std. error    t-ratio   p-value
  ----------------------------------------------------------------
  const                8.94064       9.93701        0.8997   0.3762
  inflation_rate_i~    0.0807969     0.481517       0.1678   0.8680
  producer_price_i~   -0.0814397     0.236831      -0.3439   0.7336
  agricultural_pro~    0.0366378     0.198089       0.1850   0.8546
  Food_beverages_a~   -3.43707e-05   0.000171799   -0.2001   0.8429
  uhat_1               0.788917      0.163921       4.813    5.04e-05 ***

  Unadjusted R-squared = 0.461755

Test statistic: LMF = 23.163024,
with p-value = P(F(1,27) > 23.163) = 5.04e-005

Alternative statistic: TR^2 = 15.237913,
with p-value = P(Chi-square(1) > 15.2379) = 9.48e-005

Ljung-Box Q' = 12.3649,
with p-value = P(Chi-square(1) > 12.3649) = 0.000437
```

The results show a very low p-value. Since this number is less than the significance level of 0.10, we reject the null hypothesis. This means there is an autocorrelation problem.

In order to solve this autocorrelation problem, we did Hildreth-Lu test.

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
    rho          ESS      rho          ESS      rho          ESS
  -0.99      8949.40    -0.90      8301.85    -0.80      7617.94
  -0.70      6970.28    -0.60      6357.61    -0.50      5778.78
  -0.40      5232.96    -0.30      4719.77    -0.20      4239.40
  -0.10      3792.50     0.00      3379.92     0.10      3002.30
   0.20      2659.68     0.30      2351.36     0.40      2076.19
   0.50      1833.25     0.60      1622.49     0.70      1443.94
   0.80      1292.15     0.90      1141.26     0.99      1039.76

ESS is minimized for rho = 0.98


Model 3: Hildreth-Lu, using observations 1992-2023 (T = 32)
Dependent variable: export_of_agricultural_products
rho = 0.98

                       coefficient    std. error    t-ratio  p-value
  -------------------------------------------------------------------
  const                266.198        136.595         1.949   0.0618   *
  inflation_rate_i~     -1.42450        0.654591     -2.176   0.0385   **
  producer_price_i~      0.989431       0.158407      6.246   1.11e-06 ***
  agricultural_pro~      0.193365       0.144084      1.342   0.1908
  Food_beverages_a~     -0.000347619    0.000234187  -1.484   0.1493

Statistics based on the rho-differenced data:

Sum squared resid    1036.604   S.E. of regression   6.196186
R-squared            0.985693   Adjusted R-squared   0.983573
F(4, 27)             18.07309   P-value(F)           2.51e-07
rho                  0.310965   Durbin-Watson        1.342011

Statistics based on the original data:

Mean dependent var   102.1379   S.D. dependent var   48.28919

Excluding the constant, p-value was highest for variable 4 (agricultural_producer_index)
```

Nevertheless, we got a ρ=0.98 which means there is severe autocorrelation problem.

So, we removed the insignificant variable which is X_3 from regression and re-ran the tests.

```
gretl: model 14
File  Edit  Tests  Save  Graphs  Analysis  LaTeX
Model 14: OLS, using observations 1991-2023 (T = 33)
Dependent variable: export_of_agricultural_products

                       coefficient    std. error   t-ratio  p-value
  -----------------------------------------------------------------
  const               -83.2606       12.7042       -6.554   3.53e-07 ***
  inflation_rate_i~     1.59087       0.632304      2.516   0.0177   **
  producer_price_i~     0.714064      0.224299      3.184   0.0035   ***
  Food_beverages_a~    -0.000410451   0.000203889  -2.013   0.0535   *

Mean dependent var   100.2581   S.D. dependent var   48.73995
Sum squared resid    3411.235   S.E. of regression   10.84568
R-squared            0.955126   Adjusted R-squared   0.950484
F(3, 29)             205.7529   P-value(F)           1.23e-19
Log-likelihood      -123.3573   Akaike criterion     254.7146
Schwarz criterion    260.7006   Hannan-Quinn         256.7287
rho                  0.755598   Durbin-Watson        0.614174
```

Since the R^2 is 0.96, variations in Xs can explain about 96% of variations in Y. Which means model as a whole has a very high explanatory power. It still retains the high explanatory problem even with the omitted variable.

Secondly, we checked again for the heteroskedasticity problem in our model. We set the hypothesis as

H_0 = There isn't heteroskedasticity problem

H_1 = There is heteroskedasticity problem

We again ran into same heteroskedasticity problem. P-value is less than 0.10. Again, we did the heteroskedasticity corrected test and get these results.

gretl: model 15

File Edit Tests Save Graphs Analysis LaTeX

```
Model 15: Heteroskedasticity-corrected, using observations 1991-2023 (T = 33)
Dependent variable: export_of_agricultural_products

                      coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const              -74.9179       7.74755      -9.670    1.41e-010 ***
  inflation_rate_i~    0.679541     0.506603      1.341    0.1902
  producer_price_i~    1.01132      0.164880      6.134    1.10e-06  ***
  Food_beverages_a~   -0.000312848  0.000153048  -2.044    0.0501    *

Statistics based on the weighted data:

Sum squared resid    58.82572   S.E. of regression   1.424245
R-squared            0.978923   Adjusted R-squared   0.976743
F(3, 29)             448.9782   P-value(F)           2.16e-24
Log-likelihood      -56.36315   Akaike criterion     120.7263
Schwarz criterion    126.7123   Hannan-Quinn         122.7404
rho                  0.819783   Durbin-Watson        0.435902

Statistics based on the original data:

Mean dependent var   100.2581   S.D. dependent var   48.73995
Sum squared resid    3835.429   S.E. of regression   11.50027

Excluding the constant, p-value was highest for variable 2 (inflation_rate_in_food_and_beve)
```

Lastly, we checked if our model has an autocorrelation problem or not. So, we set our hypotheses as follows:

H_0 = There isn't autocorrelation problem

H_1 = There is autocorrelation problem gretl: autocorrelation

```
Breusch-Godfrey test for first-order autocorrelation
OLS, using observations 1991-2023 (T = 33)
Dependent variable: uhat

                      coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------------
  const                 9.49001     9.60594       0.9879   0.3316
  inflation_rate_i~     0.0255447   0.468626      0.05451  0.9569
  producer_price_i~    -0.0322203   0.166353     -0.1937   0.8478
  Food_beverages_a~    -4.07900e-05 0.000151323  -0.2696   0.7895
  uhat_1                0.824154    0.165488      4.980    2.93e-05 ***

  Unadjusted R-squared = 0.469716

Test statistic: LMF = 24.801857,
with p-value = P(F(1,28) > 24.8019) = 2.93e-005

Alternative statistic: TR^2 = 15.500615,
with p-value = P(Chi-square(1) > 15.5006) = 8.25e-005

Ljung-Box Q' = 11.7243,
with p-value = P(Chi-square(1) > 11.7243) = 0.000617
```

Again, we ran into same problem, the results show a very low p-value. Since this number is less than the significance level of 0.10, we reject the null hypothesis. This means there is an autocorrelation problem.

In order to solve this autocorrelation problem, we did Hildreth-Lu test.

```
gretl: model 16
File  Edit  Tests  Save  Graphs  Analysis  LaTeX
   rho          ESS       rho          ESS       rho          ESS
 -0.99      9157.82     -0.90      8452.86     -0.80      7715.42
 -0.70      7025.76     -0.60      6383.06     -0.50      5786.10
 -0.40      5233.20     -0.30      4722.10     -0.20      4250.02
 -0.10      3813.79      0.00      3410.30      0.10      3036.96
  0.20      2692.39      0.30      2376.72      0.40      2091.53
  0.50      1839.38      0.60      1623.14      0.70      1445.06
  0.80      1302.95      0.90      1179.66      0.99      1109.20

ESS is minimized for rho = 0.98


Model 16: Hildreth-Lu, using observations 1992-2023 (T = 32)
Dependent variable: export_of_agricultural_products
rho = 0.98

                      coefficient    std. error    t-ratio  p-value
  ------------------------------------------------------------------
  const               220.093        134.081         1.641   0.1119
  inflation_rate_i~    -1.04906        0.600203     -1.748   0.0915   *
  producer_price_i~     1.04277        0.155517      6.705   2.81e-07 ***
  Food_beverages_a~    -0.000330085    0.000237144  -1.392   0.1749

Statistics based on the rho-differenced data:

Sum squared resid    1105.751   S.E. of regression   6.284194
R-squared            0.984725   Adjusted R-squared   0.983089
F(3, 28)             22.84357   P-value(F)           1.10e-07
rho                  0.281970   Durbin-Watson        1.413171

Statistics based on the original data:

Mean dependent var   102.1379   S.D. dependent var   48.28919

Excluding the constant, p-value was highest for variable 5 (Food_beverages_and_tobacco_good)
```

Again, we got a ρ=0.98 which means there is severe autocorrelation problem.

Rerunning the tests did not yield efficient results. So, changing the data set did not yield efficient results. It may have a high explanatory power but it has many efficiency problems. With these many problems the estimated regression may not satisfy the efficiency criterion.

Yararlanılan Kaynaklar:

- Duru, S. (2021) Investigate of the Effect of Food Inflation on the Agricultural and Food Products Export Using Multiple Linear Regression Analysis in Turkey
- https://fred.stlouisfed.org/