**Statistical Analysis of Protein Demand**

Kevin Lim

7815073

University of Manitoba

ECON 4822

Gregory Mason

March 25, 2021

In the statistical analysis, will develop a forecast using an economic model to predict protein demand. The hypothesis for protein demand is that variables such as income, population, and prices of protein affect the demand. We will explore different types of regression models to determine the best fit for our study.

One of the issues of using time series data is autocorrelation, where the error terms of different time periods are correlated with each other. This is observed by plotting the residuals and conducting the Durbin-Watson test.



*Figure 1: Residual plot of regressing demand for chicken (top left), beef (top right), pork (bottom left), fish (bottom right).*

Looking at *Figure 1,* we see positive autocorrelation in our plots. By using the Durbin-Watson test, we can also test for autocorrelation. The test results in values between 0 and 4, where the closer the value is to 0, the closer it is to positive autocorrelation. If the value is closer to 4, it is a case of negative correlation. We list the resulting d-statistics in *Table 1*. All values are close to 0, confirming that there are cases of positive autocorrelation.

**Table 1**

|  |  |
| --- | --- |
| **Durbin-Watson Test Results** | |
| **Protein** | **D-statistic** |
| **Chicken** | 0.5569601 |
| **Beef** | 0.3737459 |
| **Pork** | 0.463584 |
| **Fish** | 0.4113023 |

To avoid autocorrelation, we use a robust regression will be used to account for this error. Using a robust regression helps in reducing errors in the regression. It does not change the coefficients of the variables, but it improves the t-values.

Seasonality is the yearly, reoccurring patterns within the data. We also want to run a regression to observe how regression models compare with and without seasonality. To do this, we add quarterly seasonal dummies to the regression model. In Table 2 and 3, we estimate regression with and without the dummy variables.

**Table 2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Linear Regression Without Dummies** | | | | |
| **Variable** | **D\_CHIC** | **D\_BEEF** | **D\_PORK** | **D\_FISH** |
| **MW** | .00002862 | -.00139634\*\*\* | -.00028158 | -.00042556 |
|  | 0.00046 | 0.00044 | 0.00072 | 0.00026 |
|  | 0.06 | -3.21 | -0.39 | -1.64 |
| **ARP\_CHIC** | -2.797066\*\*\* | -.9355548\*\*\* | -.93115733 | -1.1869417\*\*\* |
|  | 0.27364 | 0.28819 | 0.47429 | 0.17077 |
|  | -10.22 | -3.25 | -1.96 | -6.95 |
| **ARP\_SIRLOIN** | .37174683\*\*\* | .08457264 | .16468307 | -.05951998 |
|  | 0.05902 | 0.05563 | 0.10896 | 0.05152 |
|  | 6.30 | 1.52 | 1.51 | -1.16 |
| **ARP\_CHOPS** | -.08000099 | -.50995762\*\*\* | -.51325725 | .44743775\*\*\* |
|  | 0.15146 | 0.14308 | 0.32060 | 0.13081 |
|  | -0.53 | -3.56 | -1.60 | 3.42 |
| **ARP\_CANNED** | -.10907411 | -.2299257 | 2.002348\*\*\* | .30794731\*\*\* |
|  | 0.15929 | 0.14454 | 0.32235 | 0.06840 |
|  | -0.68 | -1.59 | 6.21 | 4.50 |
| **POP** | 1.423e-06\*\*\* | 6.608e-07\*\*\* | -6.860e-07 | 6.335e-07\*\*\* |
|  | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
|  | 3.81 | 2.15 | -1.21 | 3.60 |
| **\_cons** | -4.9235176 | 32.012386 | 50.471764 | -5.5502355 |
|  | 7.80565 | 6.12536 | 11.28474 | 3.55033 |
|  | -0.63 | 5.23 | 4.47 | -1.56 |
| **R2** | .89540379 | .95989719 | .90514804 | .7838323 |
| **F** | 91.630822 | 341.09088 | 141.65777 | 82.922615 |
| **N** | 77 | 77 | 77 | 77 |

*Note:* The results in each cell are the coefficients, standard errors, and t-values in descending order *(b/se/t)*.

\*\*\*p<.05

**Table 3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Linear Regressions with Dummies** | | | | |
| **Variable** | **D\_CHIC** | **D\_BEEF** | **D\_PORK** | **D\_FISH** |
| **MW** | .00007842 | -.00137979\*\*\* | -.00022471 | -.00043881 |
|  | 0.00049 | 0.00045 | 0.00071 | 0.00027 |
|  | 0.16 | -3.06 | -0.32 | -1.62 |
| **ARP\_CHIC** | -2.8517839\*\*\* | -.95117216\*\*\* | -1.0028965\*\*\* | -1.1803287\*\*\* |
|  | 0.26541 | 0.29620 | 0.45650 | 0.16572 |
|  | -10.74 | -3.21 | -2.20 | -7.12 |
| **ARP\_SIRLOIN** | .37752987\*\*\* | .08465554 | .17056221 | -.05867913 |
|  | 0.05460 | 0.05656 | 0.10364 | 0.05309 |
|  | 6.91 | 1.50 | 1.65 | -1.11 |
| **ARP\_CHOPS** | -.0935278 | -.51179807\*\*\* | -.53362844 | .44490732\*\*\* |
|  | 0.14936 | 0.14610 | 0.31462 | 0.13532 |
|  | -0.63 | -3.50 | -1.70 | 3.29 |
| **ARP\_CANNED** | -.10400499 | -.22768293 | 2.0152243\*\*\* | .3089751\*\*\* |
|  | 0.16505 | 0.14985 | 0.32150 | 0.06856 |
|  | -0.63 | -1.52 | 6.27 | 4.51 |
| **POP** | 1.408e-06\*\*\* | 6.560e-07\*\*\* | -6.952e-07 | 6.407e-07\*\*\* |
|  | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
|  | 3.60 | 2.07 | -1.26 | 3.50 |
| **Q1** | -.10480253 | -.05595779 | -.2394369 | .00495896 |
|  | 0.14972 | 0.17043 | 0.26026 | 0.12540 |
|  | -0.70 | -0.33 | -0.92 | 0.04 |
| **Q2** | -.2273976 | -.04963114 | -.3227638 | -.00612405 |
|  | 0.15021 | 0.17192 | 0.24395 | 0.12368 |
|  | -1.51 | -0.29 | -1.32 | -0.05 |
| **Q3** | -.13856494 | -.04136328 | -.29829246 | -.03322831 |
|  | 0.15277 | 0.16124 | 0.23789 | 0.11780 |
|  | -0.91 | -0.26 | -1.25 | -0.28 |
| **\_cons** | -4.4878476 | 32.138082 | 50.896958 | -5.6707849 |
|  | 8.13154 | 6.33993 | 10.94364 | 3.70548 |
|  | -0.55 | 5.07 | 4.65 | -1.53 |
| **R2** | .89834729 | .95996732 | .90740831 | .78431533 |
| **F** | 69.15542 | 218.46346 | 95.608791 | 62.9799 |
| **N** | 77 | 77 | 77 | 77 |

*Note:* The results in each cell are the coefficients, standard errors, and t-values in descending order *(b/se/t)*.

\*\*\*p<.05

The addition of dummy variables in Table 3 indicates that there is no significance at the five percent level, meaning that the variables are not statistically significant. Seasonality has no major effect on the dependent variables. Additionally, it does little to improve the R­­2. So we can drop the dummies and use the results from Table 2 to construct our regression models by using statistically significant variables. One thing to note from Table 2 is that the price variables *ARPSIRLOIN* in *DBEEF* and *ARPCHOPS* in *DPORK* are not significant. This suggests a result of mismatched price variables as sirloin steak and pork chops are more expensive products and a more common product from these proteins would be a better fit.

Another model that could be used to regress protein demand was by using seemingly unrelated regressions. This model shares the error term among the dependent variables. However, as we add more variables, the difference between ordinary least squares and seemingly unrelated regressions becomes smaller, making it difficult to justify this method over the other.

**Table 4**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Seemingly Unrelated Regressions** | | | | | | | | | | | | |
| **Equation** | **Obs** | | **Parms** | | **RMSE** | | **"R-sq"** | | **chi2** | | **P** | |
| **D\_CHIC** | 77 | | 6 | | .45593 | | 0.8954 | | 659.16 | | 0.0000 | |
| **D\_BEEF** | 77 | | 6 | | .4724211 | | 0.9599 | | 1843.07 | | 0.0000 | |
| **D\_PORK** | 77 | | 6 | | .7520334 | | 0.9051 | | 734.79 | | 0.0000 | |
| **D\_FISH** | 77 | | 6 | | .3140529 | | 0.7838 | | 279.20 | | 0.0000 | |
| **D\_CHIC** | | | | | | | | | | | | |
|  | | **Coef.** | | **Std. Err.** | | **z** | | **P>z** | | **[95% Conf. Interval]** | | |
| **MW** | | .0000286 | | .0004126 | | 0.07 | | 0.945 | | -.00078 | | .0008372 |
| **ARP\_CHIC** | | -2.797066 | | .2940496 | | -9.51 | | 0.000 | | -3.373393 | | -2.220739 |
| **ARP\_SIRLOIN** | | .3717468 | | .0619598 | | 6.00 | | 0.000 | | .2503079 | | .4931858 |
| **ARP\_CHOPS** | | -.080001 | | .1592231 | | -0.50 | | 0.615 | | -.3920726 | | .2320706 |
| **ARP\_CANNED** | | -.1090741 | | .1682157 | | -0.65 | | 0.517 | | -.4387707 | | .2206225 |
| **POP** | | 1.42e-06 | | 3.00e-07 | | 4.75 | | 0.000 | | 8.36e-07 | | 2.01e-06 |
| **\_cons** | | -4.923518 | | 6.10945 | | -0.81 | | 0.420 | | -16.89782 | | 7.050783 |
| **D\_BEEF** | | | | | | | | | | | | |
|  | | **Coef.** | | **Std. Err.** | | **z** | | **P>z** | | **[95% Conf. Interval]** | | |
| **MW** | | -.0013963 | | .0004275 | | -3.27 | | 0.001 | | -.0022342 | | -.0005585 |
| **ARP\_CHIC** | | -.9355548 | | .3046854 | | -3.07 | | 0.002 | | -1.532727 | | -.3383824 |
| **ARP\_SIRLOIN** | | .0845726 | | .0642009 | | 1.32 | | 0.188 | | -.0412588 | | .210404 |
| **ARP\_CHOPS** | | -.5099576 | | .1649823 | | -3.09 | | 0.002 | | -.8333169 | | -.1865983 |
| **ARP\_CANNED** | | -.2299257 | | .1743 | | -1.32 | | 0.187 | | -.5715475 | | .1116961 |
| **POP** | | 6.61e-07 | | 3.10e-07 | | 2.13 | | 0.033 | | 5.26e-08 | | 1.27e-06 |
| **\_cons** | | 32.01239 | | 6.330429 | | 5.06 | | 0.000 | | 19.60497 | | 44.4198 |
| **D\_PORK** | | | | | | | | | | | | |
|  | | **Coef.** | | **Std. Err.** | | **z** | | **P>z** | | **[95% Conf. Interval]** | | |
| **MW** | | -.0002816 | | .0006805 | | -0.41 | | 0.679 | | -.0016153 | | .0010521 |
| **ARP\_CHIC** | | -.9311573 | | .4850198 | | -1.92 | | 0.055 | | -1.881779 | | .0194639 |
| **ARP\_SIRLOIN** | | .1646831 | | .1021995 | | 1.61 | | 0.107 | | -.0356243 | | .3649904 |
| **ARP\_CHOPS** | | -.5132573 | | .2626304 | | -1.95 | | 0.051 | | -1.028003 | | .0014889 |
| **ARP\_CANNED** | | 2.002348 | | .2774632 | | 7.22 | | 0.000 | | 1.45853 | | 2.546166 |
| **POP** | | -6.86e-07 | | 4.94e-07 | | -1.39 | | 0.165 | | -1.65e-06 | | 2.82e-07 |
| **\_cons** | | 50.47176 | | 10.07723 | | 5.01 | | 0.000 | | 30.72076 | | 70.22276 |
| **D\_FISH** | | | | | | | | | | | | |
|  | | **Coef.** | | **Std. Err.** | | **z** | | **P>z** | | **[95% Conf. Interval]** | | |
| **MW** | | -.0004256 | | .0002842 | | -1.50 | | 0.134 | | -.0009825 | | .0001314 |
| **ARP\_CHIC** | | -1.186942 | | .2025467 | | -5.86 | | 0.000 | | -1.583926 | | -.7899575 |
| **ARP\_SIRLOIN** | | -.05952 | | .042679 | | -1.39 | | 0.163 | | -.1431693 | | .0241294 |
| **ARP\_CHOPS** | | .4474378 | | .1096758 | | 4.08 | | 0.000 | | .2324772 | | .6623984 |
| **ARP\_CANNED** | | .3079473 | | .11587 | | 2.66 | | 0.008 | | .0808463 | | .5350483 |
| **POP** | | 6.33e-07 | | 2.06e-07 | | 3.07 | | 0.002 | | 2.29e-07 | | 1.04e-06 |
| **\_cons** | | -5.550236 | | 4.2083 | | -1.32 | | 0.187 | | -13.79835 | | 2.697881 |

The two models suffer from multicollinearity, however. Multicollinearity occurs when independent variables are correlated with each other. By creating a matrix of the independent variables, we can determine which variables are correlated with each other. We find that there is a correlation between the money variables such as prices and wages as a result of inflation over time. To solve this, we try running the first difference regression. First differences measure changes over time compared to levels. By transforming the variables using differences, we no longer have an issue of multicollinearity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **First Difference Regression** | | | | |
| **Variable** | **FDChicken** | **FDBeef** | **FDPork** | **FDFish** |
| **DMW** | .00008438  0.00012  0.68 | -.0000327  0.00030  -0.11 | -.00021644  0.00017  -1.30 | .00004951  0.00010  0.51 |
| **DARPCHI** | -.14715039  0.18707  -0.79 | .33705236  0.46180  0.73 | -.27908685  0.25177  -1.11 | -.07893762  0.14630  -0.54 |
| **DARPS** | .02335943  0.03549  0.66 | -.05870785  0.08760  -0.67 | .06922742  0.04776  1.45 | .01733734  0.02775  0.62 |
| **DARPCHO** | .10487202  0.07814  1.34 | .16447824  0.19289  0.85 | -.33765647\*\*\*  0.10516  -3.21 | .0462116  0.06111  0.76 |
| **DARPCANNED** | .07302709  0.13758  0.53 | .47728914  0.33964  1.41 | -.00085847  0.18516  -0.00 | -.02290113  0.10759  -0.21 |
| **DPOP** | 3.874e-07  0.00000  0.66 | 9.056e-08  0.00000  0.06 | 1.094e-06  0.00000  1.37 | 4.912e-07  0.00000  1.06 |
| **\_cons** | .03376874  0.05645  0.60 | -.11138255  0.13934  -0.80 | -.15079319  0.07597  -1.98 | -.05369157  0.04414  -1.22 |
| **R2** | .05724304 | .0571014 | .16206752 | .03387679 |
| **F** | .6982658 | .69643338 | 2.2242562 | .40324373 |
| **N** | 76 | 76 | 76 | 76 |

*Note:* The results in each cell are the coefficients, standard errors, and t-values in descending order *(b/se/t)*.

\*\*\*p<.05

However, we now have an issue where most variables are not significant. This suggests that there is a complex multicollinearity problem and requires further research to solve this problem. Additionally, the R2 for each protein is unable to explain most of the variation in demand. Out of all the models tested, the linear regression model is the best fit from this analysis.

To conclude, using the results from Table 2, we find that the wage only affects the demand for beef. We also find that seasonality has no effect on any of the dependent variables. Population affects all proteins except for pork. Additionally, we find that some of the price variables are statistically significant such as the price of chicken and sirloin steak affecting the demand for chicken. However, the price of sirloin steak and pork chops do not affect their respective proteins, suggesting a case of mismatched variables. The model for demand in pork may need more research as only one variable is significant enough to affect demand. There is also the problem of multicollinearity in our model that would require further research to solve. A potential solution could be an instrumental variable approach where a variable affects the dependent only through the independent variable, but that is beyond the scope of our analysis.