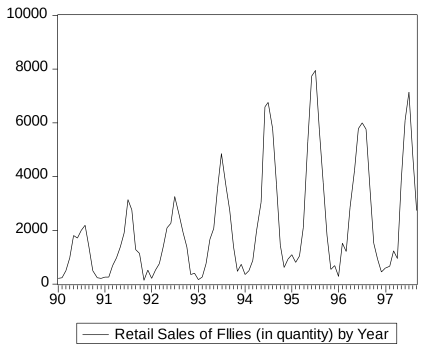
# **Artificial Fly Sales**

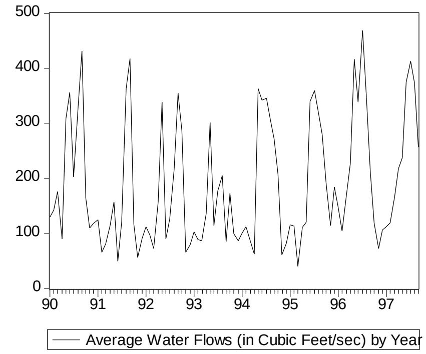
(an econometric look)

Joseph F. Adams
ECON6818

## General

The flyfishing industry in Colorado experienced a boom in growth during the early and mid 1990's. Much of this growth was attributed to the growth in the Colorado population base, commercial advertising, and popular films such as "A River Runs Through It". The increase not only occurred in the area of services such as guide trips and schools, but most notably in the retail sales arena. Unfortunately for the retail owners, this boom did not last through the decade and many flyfishing shops have closed as we near the end of the century. Discussions regarding the situation with current and previous owners indicate that many believe that the failure of stores is directly linked to the water flows of the South Platte River (below Cheeseman Canyon through Deckers to Chatfield). This water is categorized as "Blue Medal" fishing water (world class) and is the closest in proximity to metropolitan Denver. The Denver Water Board regulates the flow of water and bases the flow on consumption demands, as well as icepack thawing in the spring and early summer. According to shop owners late, heavy snows and precipitation result in reduced retail flyfishing sales and services. Shop owners also argue that water flows that are too low adversely impact retail sales. A review of sales records indicate that months of higher water flows actually correspond to higher retail sales (see chart below).




Central to the shop owners' belief is a discussion of why a person decides to go flyfishing. Simple queries of active flyfishermen seem to point to temperature and precipitation as partial determinants in the decision to fish or not. Water flows, however, was the principal deciding factor since it "really dictates the conditions of the river" and "relates to the probability of catching fish". For the purposes of this project, waterflows, temperature, and precipitation will be considered. Another potential variable that might impact the sales of flies is proximity of the fly shop to the body of water (and proximity of other fly shops). For this experiment, we will isolate one business that is the only shop within thirty miles of a body of water - and is located adjacent to the body of water.

This econometrics problem will investigate the hypothesis of the flyfishing shop owners. Specifically, can waterflows, temperature, and precipitation be used as a predictor of retail fly sales? In answering this question, we also analyze the issue of the shop owners regarding the affect of the waterflows on retail sales. This project will not prove causality, but rather, look at the statistical relationships. Gujarati's methodology for conducting econometric studies will be used.

## Data

In order to conduct this project, a fly shop located on the South Platte River (just below Cheeseman Dam) is used to gather sales information. This shop sits adjacent to the river and is the only shop within thirty miles of this area. Sales data consists of total monthly sales of artificial flies from 1990 to September of 1997 (see attachments). The United States Geological Survey provided the waterflow data from the measuring station just below Cheeseman Dam. This data consists of average monthly flows in cubic feet per second (CFS) for the same period as the fly sales. Monthly mean temperature in tenths of degrees Fahrenheit (at the same collection point as the water flow) comprises the temperature data provided by the National Weather Service. Total monthly precipitation also comes from the National Weather Service and is measured to the tenth of inches (also at the measuring station).

## Model

The mathematical model (without disturbance) that will used to predict the sale of flies will initially be of the form:

$$\text{Flies}_t = \quad + \quad \text{Water}_t + \quad \text{Precipitation}_t + \quad \text{Temp}_t \qquad (1)$$

where:

**Flies** is the monthly sales of artificial flies in quantity sold

**Water** is the monthly average waterflow

**Temp** is the monthly average temperature

**Precipitation** is the total monthly precipitation

, , , and , are parameters (regression coefficients)

Basically, the model consists of multiple explanatory variables, therefore, this type of regression is know as a multiple regression. In stochastic form, this type of model assumes that: the independent variables are not correlated with the disturbance term; the error term has a mean value (expected value) of zero; the variance of the error term is constant (homoscedasticity); the error terms are not correlated; no exact collinearity exists between the independent variables; and finally, for hypothesis testing the error term follows a normal distribution with mean zero and finite variance. If necessary, a time trend will also be included in the model to account for the fact that sales have seemingly increased over time (until 1995). This model is in a general form and will be tested (and subsequently updated) until the specification is satisfactory. The partial regression coefficients (which reflect the effect of one explanatory variable on the mean value of the dependent variable when the values of the other explanatory variables in the model are held constant) will have varying anticipated signs. As the temperature increases, one would anticipate a positive increase on the sale of flies (when it becomes colder, generally, fewer people desire to go fishing). Similarly, as the water level increases fly sales would also increase. On the contrary, an increase in precipitation would decrease the number of anticipated flyfishers (and thus decrease fly sales). Originally, waterflow was introduced into the model nonlinearly since it was thought that flows that are too high would be detrimental to the sport. Discussions with both the U.S. Forest Service and the Colorado Division of Wildlife indicated the opposite. Periodic high water flows are healthy for the overall conditions of a river, and are important in the life cycle of fish (as opposed to extremely low water levels, which are extremely dangerous to fish). Accordingly, waterflows is a linear explanatory variable in the model.

## Parameter Estimation

At this point, we conduct a multiple regression using the general model. This initial "iteration" provides OLS estimates, residuals, and enough information to evaluate the model goodness of fit and adequacy.

Dependent Variable: FLIES
Method: Least Squares
Date: 04/25/99 Time: 20:12
Sample: 1990:01 1997:09
Included observations: 93

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -2144.861 | 540.1163 | -3.971111 | 0.0001 |
| WATER | 5.203537 | 1.679575 | 3.098128 | 0.0026 |
| PRECIPITATION | -5.754819 | 24.27988 | -0.237020 | 0.8132 |
| TEMP | 78.18480 | 13.83364 | 5.651789 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.608260 | Mean dependent var | 2147.161 |
| Adjusted R-squared | 0.595055 | S.D. dependent var | 2009.251 |
| S.E. of regression | 1278.593 | Akaike info criterion | 17.18697 |
| Sum squared resid | 1.45E+08 | Schwarz criterion | 17.29590 |
| Log likelihood | -795.1939 | F-statistic | 46.06375 |
| Durbin-Watson stat | 0.565964 | Prob(F-statistic) | 0.000000 |

## Model Adequacy Check and Hypothesis Tests

For this regression, an analysis of the estimators indicate that, as predicted, increasing waterflows increases fly sales. Also, increasing temperature increases sales, however, increasing precipitation decreases sales. The intercept term, C, has a value of -2144.861 which would supposedly be the mean value of fly sales provided the coefficients on the explanatory variables were all zero. The residual sum of squares (RSS) is equal to:

$$e_t^2 = \quad (\text{Flies}_t - \quad - \quad \text{Water}_t - \quad \text{Precipitation}_t - \quad \text{Temp}_t)^2 = 1.45\text{E}+08,$$

which is extremely large. Testing hypotheses about individual partial regression coefficients (using the test of significance approach) indicates that we reject the null hypothesis that the coefficients on the explanatory variables waterflows and temperature are zero. On the other hand, at the 10% significance level, we fail to reject the null hypothesis for the precipitation variable coefficient. The multiple coefficient of determination (which gives the proportion of the total variation in Flies (fly sales) explained by waterflows, precipitation, and temperature jointly) is equal to $R^2 = 0.6082$. Generally, we would like to see a number as close to one as possible. An F-Test is used to measure the overall significance of the estimated regression line (and test the significance of $R^2$). In this instance,

$$F = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)}$$

where:

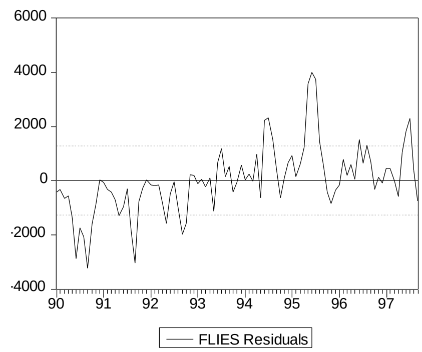
n = number of observations

k = number of explanatory variables including the intercept

Thus, F = [(.608260) / 3]/[(1 - .608260)/(93 - 4)] = 46.06375, and the null hypothesis is that $R^2 = 0$. At the 10% confidence level ( and 3, 89 DOF) the critical F value is between 2.13 and 2.18, therefore, we reject the null hypothesis that all "partial slopes" are simultaneously equal to zero or that $R^2$ is equal to zero.

Since the standard error of the regression is large, there exists the possibility of multicollinearity. Collinearity is high correlation among the explanatory variables therefore, it is first important to actually see if collinearity is a problem. Multicollinearity (according to Gujarati) is a question of degree of severity and kind. Auxiliary regressions of the explanatory variables will help detect/determine the nature of the multicollinearity and will be used later.

Of greater importance to this model is the issue of correlation (autocorrelation). It is not uncommon in time series regressions to discover that the residuals are correlated with their own lagged values. This correlation is inconsistent with the regression theory assumption that the error terms are not correlated with each other: $E(u_i u_j) = 0$ for $i \neq j$. While the least square estimators of the model are still linear and unbiased in the presence of autocorrelation, there are some significant undesirable consequences. Specifically: OLS estimators might not be efficient; estimated variances of the estimators could be biased (thus inflating *t* values); the usual *t* and F tests might not be generally reliable; and the conventionally computed R2 may be an unreliable measure of the true R2. Also, standard errors of the parameters could be larger which means that OLS will underestimate them if the autocorrelation is not addressed. The effect on the *t*-statistics is that they become larger (as in this model) than they really should be. An examination of the OLS residuals shows that the error terms seem to correlate positively.




To measure the association between adjacent residuals, we can use the Durbin-Watson Statistic. It is simply the ratio of the sum of squared differences in successive residuals to the RSS or:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

If there is no problem of association between adjacent residuals, the statistic will be around 2. With positive serial correlation, *d* will be less than 2 - and in the worst cases - be near 0. A *d* below 1.5 is usually a cause for concern. For the initial model, the regression yields a Durban-Watson Statistic (*d*) of 0.565964, which confirms what the plot of the residuals depicted - the presence of first order autocorrelation. We will now seek remedial measures.

Using the remedial procedures described by Gujarati (chapter 12), we assume the error terms follow the following scheme:

$$e_t = \quad e_{t-1} + \quad _t \qquad\qquad -1 \qquad\quad 1$$

where:

= coefficient of autocorrelation

= random term/satisfy the usual OLS assumptions

We can estimate  from our Durbin-Watson *d* statistic using the formula:

$$d \simeq 2(1 - \quad) \quad \text{or} \quad \simeq 1 - d/2$$

Substituting the value of *d* from the original regression gives:

$$\simeq 1 - (.565964/2) \simeq .717018$$

We can now use the estimated  to run the generalized difference equation presented by Gujarati. The next requirement is to transform the regression model so that the error term in the transformed model does not have first order autocorrelation. Starting with the original model (1), write the regression with a one-period time lag and multiply both sides of the equation by . This equates to:

$$\text{Flies}_{t-1} = \quad + \quad \text{Water}_{t-1} + \quad \text{Precipitation}_{t-1} + \quad \text{Temp}_{t-1} + \quad e_{t-1} \quad (2)$$

Subtracting equation (2) from the original regression equation (1) yields:

$$(Flies_t - \quad Flies_{t-1}) = \quad (1 - \quad ) + \quad (Water_t - \quad Water_{t-1}) + \quad (Precipitation_t - \quad Precipitation_{t-1})$$

$$+ \quad (Temp_t - \quad Temp_{t-1}) + \quad {}_t$$

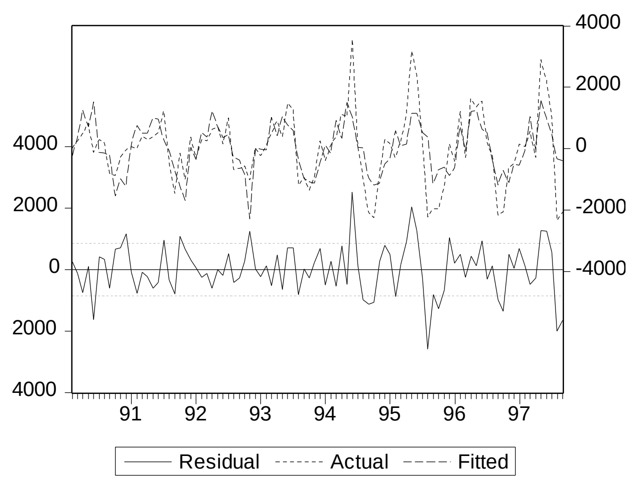
which is called the generalized difference equation. When Ordinary Least Squares (OLS) is applied to the transformed variables, the estimators obtained are the Generalized Least Squares (GLS) estimators. At this point the regression is conducted with the transformed variables (generalized difference equation).

Dependent Variable: DELTAFLIES
Method: Least Squares
Date: 05/03/99 Time: 19:38
Sample(adjusted): 1990:02 1997:09
Included observations: 92 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -3.820421 | 87.93035 | -0.043448 | 0.9654 |
| DELTAWATER | 0.272860 | 0.936539 | 0.291349 | 0.7715 |
| DELTAPRECIP | -14.30539 | 11.48402 | -1.245679 | 0.2162 |
| DELTATEMP | 100.3685 | 13.00551 | 7.717380 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.495393 | Mean dependent var | 27.55435 |
| Adjusted R-squared | 0.478191 | S.D. dependent var | 1166.580 |
| S.E. of regression | 842.6949 | Akaike info criterion | 16.35359 |
| Sum squared resid | 62491859 | Schwarz criterion | 16.46323 |
| Log likelihood | -748.2652 | F-statistic | 28.79772 |
| Durbin-Watson stat | 1.574791 | Prob(F-statistic) | 0.000000 |

The results of this regression depict an increase in the Durbin-Watson statistic, a reduction in the SSR, and a reduction in the standard error of the regression. A view of the residuals ( ) confirms that the significant, positive first-order autocorrelation of the original model has been corrected.



Since the standard error of the regression is still rather large, we will investigate if the samples have a collinearity problem. If the samples have high multicollinearity, the model might encounter insignificant *t* ratios, a higher $R^2$ value (but few significant *t* ratios), wrong signs for regression coefficients, unstable OLS estimators, and a difficulty in assessing the individual contributions of explanatory variables to the explained sum of squares or $R^2$. Auxiliary regressions and variance inflation factors (VIF) will help to determine if harmful collinearity is present. Since multicollinearity occurs because one or more of the explanatory variables are roughly linear combinations of other explanatory variables, regressing each variable on the remaining variables helps identify if these linear combinations exist. Starting with the first explanatory variable, waterflows, the auxiliary regression yields:

Dependent Variable: DELTAWATER
Method: Least Squares
Date: 05/03/99 Time: 19:41
Sample(adjusted): 1990:02 1997:09
Included observations: 92 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -0.351849 | 9.952107 | -0.035354 | 0.9719 |
| DELTAPRECIP | -1.045088 | 1.295060 | -0.806981 | 0.4218 |
| DELTATEMP | 5.589711 | 1.347481 | 4.148266 | 0.0001 |
| **R-squared** | **0.190903** | Mean dependent var | | 1.380435 |

The VIF is equal to the inverse of $(1 - R^2)$, therefore, for this regression the VIF is 1.2359. Peter Kennedy states in his Guide to Econometrics that a VIF of greater than ten depicts harmful conditions could be present (collinearity). Using the same procedure with the remaining explanatory variables, VIFs of 1.0796 (precipitation) and 1.2790 (temperature) indicate that multicollinearity is not a problem in this model.

Another area of concern centers around the assumption that the disturbance entering the regression has the same variance (homoscedastic). If the variance is not equal then heteroscedasticity could pose a problem for the model. While the OLS estimators will still be linear and unbiased, they will no longer be efficient and the usual confidence intervals and hypothesis tests based on *t* and F distributions will be unreliable. White's test will be used to test for the presence of heteroscedasticity. For this test, we first estimate the regression by OLS (obtaining the residuals), then run an auxiliary regression squaring the residuals and regressing them on all original variables, their squared values, and their cross products (to obtain $R^2$). Multiplying $R^2$ by the sample size (n) gives a value that follows a $^2$ distribution with degrees of freedom equal to the number of explanatory variable terms in the auxiliary regression. The null hypothesis for this test is that all of the slope coefficients are zero (no heteroscedasticity). In Eviews, the test looks as follows:

White Heteroskedasticity Test:

| | | | |
|---|---|---|---|
| F-statistic | 1.149902 | Probability | 0.338188 |
| Obs*R-squared | 10.30999 | Probability | 0.325977 |

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/28/99 Time: 07:58
Sample: 1990:02 1997:09
Included observations: 92

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 596642.8 | 190593.5 | 3.130446 | 0.0024 |
| Deltawater | -2233.524 | 1358.876 | -1.643656 | 0.1041 |
| Deltawater^2 | -10.46496 | 8.942012 | -1.170314 | 0.2453 |
| Deltawater*Deltapreci | 35.73782 | 265.0934 | 0.134812 | 0.8931 |
| Deltawater*Deltatemp | 75.11368 | 207.0480 | 0.362784 | 0.7177 |
| Deltaprecip | -21099.30 | 17212.10 | -1.225841 | 0.2238 |
| Deltaprecip^2 | -849.9323 | 903.7376 | -0.940464 | 0.3497 |
| Deltaprecip*Deltatem | -132.9135 | 2302.544 | -0.057725 | 0.9541 |
| Deltatemp | 26096.07 | 18651.60 | 1.399133 | 0.1655 |
| Deltatemp^2 | 3804.423 | 2654.957 | 1.432951 | 0.1557 |

| | | | |
|---|---|---|---|
| **R-squared** | **0.112065** | Mean dependent var | 679259.3 |
| Adjusted R-squared | 0.014609 | S.D. dependent var | 1170051. |
| S.E. of regression | 1161473. | Akaike info criterion | 30.87060 |
| Sum squared resid | 1.11E+14 | Schwarz criterion | 31.14470 |
| Log likelihood | -1410.047 | F-statistic | 1.149902 |
| Durbin-Watson stat | 1.840844 | Prob(F-statistic) | 0.338188 |

Taking the $R^2$ value of .112065 and multiplying by 92, we obtain the value 10.3099. With nine degrees of freedom, the $^2$ critical value (at 10% significance) is 14.6837 therefore, we fail to reject the null hypothesis of homoscedasticity.

So far, we have not detected the harmful effects of either multicollinearity or heteroscedasticity in the model. Autocorrelation has been corrected for, yet the results of the multiple regression indicate that some additional analysis is required. The signs of the partial regression coefficients are all as predicted, yet the intercept has a negative value of 3.820421. The standard error of the regression is still large (as is the sum of the squared residuals) and the $R^2$ value is only .495393 (an overall measure of the goodness of fit of the estimated regression line). In order to test the statistical significance of each of the partial regression coefficients, *t*-tests will be used. The null hypothesis is that each of the explanatory variables (individually) has no effect on the sale of flies (therefore the coefficients are zero). At the 10% significance level and (n - 4) = 88 degrees of freedom, the critical value is approximately 1.66. According to the individual *t*-tests, we would fail to reject the null hypothesis for the waterflow and precipitation variables, but fail to reject the null hypothesis for the temperature variable. This is not very satisfactory, therefore, joint tests are in order. The intent of the joint test is to determine if the explanatory variables together (and in pairs) have no influence on fly sales. The *t*-testing procedure is not valid for joint hypothesis testing, so F-tests are used.

The basic form of the F-test consists of:

$$\frac{(SSR_r - SSR_u)/m}{(SSRu)/(n-k)}$$

which has an F-distribution with m and n-k DOF and :

n = number of observations

k = number of parameters in the unrestricted regression (including intercept)

m = number of linear restrictions

To test if all three partial regression coefficients are equal to zero, the computed F statistic is 28.79772 and the critical value at m = 3 and (n-k) = 88 DOF is between 2.13 and 2.18. This means we would reject the null hypothesis that all three coefficients equal zero at the 10% significance level. The remaining joint tests consist of:

| | |
|---|---|
| coefficients of water and precipitation = 0 | F-stat = .85544 |
| coefficients of water and temperature = 0 | F-stat = 36.9705 |
| coefficients of precipitation and temperature = 0 | F-stat = 34.1541 |

With 2 and 88 DOF, the critical F value at a 10% significance level is between 2.35 and 2.39 therefore, we would fail to reject the null hypothesis for the coefficients of water and precipitation. This F-test corresponds to the answers achieved in the individual t-tests of the two variables (water and precipitation). This means that the explanatory power of the temperature variable is much more significant than that of the waterflows or precipitation variables in terms of artificial fly sales.

## Conclusion

Contrary to the thought of flyfishing shop owners, waterflows and precipitation do not have a statistically significant relationship to the sale of artificial flies. This study simply could not establish that such a significant relationship exists. Temperature, on the other hand, does have such a relationship when used as an explanatory variable. This means that changes in temperature have a greater (significant) explanatory power to indicate the nature of fly sales than does waterflows or precipitation.

## BIBLIOGRAPHY

Gujarati, Damodar. Essentials of Econometrics. Irwin, McGraw-Hill. Boston. 1999.

Enders, Walter. Applied Econometric Time Series. John Wiley and Sons, Inc. New York. 1995.

Econometric Views Handbook. Quantitative Micro Software. Irvine, CA. 1996.

Judd, Charles M. and Gary H. McClelland. Data Analysis a Model Comparison Approach. Harcourt, Brace, Jovanovich, Inc. New York. 1989.

Kennedy, Peter. A Guide to Econometrics. MIT Press. Cambridge, MA. 1998.