Project Overview:

An econometric analysis investigating nonlinear and multiple linear regressions between workers’ earnings and other relevant factors such as years of schooling, years of experience, year, whether a worker lives in a boom town, and various industry dummies.

**Visualizing Relationships**

Figure 1.2 Relationship between Earnings (Dependent Variable) and Workforce Experience (Independent Variable)

A graph with lines and a curve

AI-generated content may be incorrect.A graph of a graph of earnings

AI-generated content may be incorrect.

Figure 1.1 Relationship between Earnings (Dependent Variable) and Years of Education (Independent Variable)

**Figure 1.1: Relationship between Earnings (in $1000s) and Education (in Years)**

The relationship between both variables is positive, as observed by an increase in earnings when education increases and by the upward curved quadratic regression line. The curve suggests a non-constant growth in earnings, as the rate of earnings growth increases with time, thus displaying a nonlinear relationship.

**Figure 1.2: Relationship between Earnings (in $1000s) and Work Experience (in Years)**

A nonlinear relationship is observed between earnings and workforce experience. Earnings initially increase before they decline as the years of workforce experience increase. This is represented by the concave regression line observed in Figure 1.2, suggesting eventual diminishing returns relative to workforce experience. Thus, we observe a nonlinear relationship between annual earnings and workforce experience.

**Sequential Hypothesis Testing for Years Schooled**

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficients** | **Regression 1** | **Regression 2** | **Regression 3** |
| (Coefficient of Years Schooled Cubed) | -0.010  (0.017) | N/A | N/A |
| (Coefficient of Years Schooled Squared) | 0.549  (0.545) | 0.232\*\*\*  (0.052) | N/A |
| (Coefficient of Years Schooled) | -5.853  (5.529) | -2.613\*\*  (1.090) | 2.406\*\*\*  (0.166) |
| (Constant) | 41.313\*\*  (17.946) | 30.896\*\*\*  (5.626) | 4.611\*\*  (1.893) |
| Adjusted R2 | 0.100 | 0.100 | 0.092 |
| Number of Observations | 2,056 | 2,056 | 2,056 |

*Table 2.1 - Regression outputs for the 3 regressions listed above*

***Note for Remaining Document****: For tables with regression output, the numbers in parentheses beside coefficients are the standard errors (accounting for heteroskedasticity) corresponding to their respective estimates. The number of asterisks beside the coefficient estimates refers to the extent of the statistical significance of the hypothesis test. (\* means statistically significant at 10% statistical significance, \*\* means statistically significant at 5% statistical significance and \*\*\* means statistically significant at 1% statistical significance).*

We will begin the sequential hypothesis testing by testing the coefficient of the cubic term for statistical significance (refer to regression 1):

and

= -0.010 and SE() = 0.017 (from Table 2.1)

t-statistic = |-0.572| < 1.96 and p-value = 0.567 > 0.05

Since the p-value is greater than 0.05 and the t-statistic is less than 1.96, the critical level for a 95% confidence interval, in magnitude, we have insufficient statistical evidence that the coefficient of the cubic term is statistically different from zero at the 5% level of significance. This renders the cubic term insignificant in the polynomial regression. Hence, we will move down the order and test the coefficient of the quadratic term (see regression 2).

and

= 0.232 and SE () = 0.052 (from Table 2.1)

t-statistic = |4.448| > 1.96 and p-value < 0.001 < 0.05

Since the p-value is less than 0.05 and the t-statistic is greater than 1.96 in magnitude, we reject that the coefficient of the quadratic term is not statistically different from 0 at the 5% level of significance. Thus, the quadratic term is significant in this polynomial regression. The sequential hypothesis testing stops now since the coefficient of the highest degree term in regression 2 is statistically significant. Therefore, the quadratic nonlinear regression is chosen to be the most appropriate by the sequential hypothesis testing algorithm.

**Sequential Hypothesis Testing for Years Experience**

*Table 2.1 - Regression outputs for the 3 regressions listed above*

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficients** | **Regression 4** | **Regression 5** | **Regression 6** |
| (Coefficient of Years Experience Cubed) | 0.001  (0.005) | N/A | N/A |
| (Coefficient of Years Experience Squared) | -0.249\*\*  (0.126) | -0.220\*\*\*  (0.023) | N/A |
| (Coefficient of Years Experience) | 3.972\*\*\*  (0.921) | 3.774\*\*\*  (0.354) | 0.673\*\*\*  (0.106) |
| (Constant) | 18.965\*\*\*  (1.954) | 19.325\*\*\*  (1.211) | 28.280\*\*\* (0.740) |
| Adjusted R2 | 0.049 | 0.049 | 0.018 |
| Number of Observations | 2,056 | 2,056 | 2,056 |

We will begin the sequential hypothesis testing by testing the coefficient of the cubic term for statistical significance (refer to regression 4), this time using years of experience rather than school:

and

= 0.001 and SE () = 0.005 (from Table 3.1)

t-statistic = |0.242| < 1.96 and p-value = 0.809 > 0.05

Since the p-value is greater than 0.05 and the t-statistic is less than 1.96 in magnitude, we have insufficient statistical evidence that the coefficient of the cubic term is statistically different from zero at the 5% level of significance. This renders the cubic term insignificant in the polynomial regression. Hence, we will move down the order and test the coefficient of the quadratic term (see regression 5).

and

= -0.220 and SE () = 0.023 (from Table 3.1)

t-statistic = |-9.417| > 1.96 and p-value < 0.001 < 0.05

Since the p-value is less than 0.05 and the t-statistic is greater than 1.96 in magnitude, we reject that the coefficient of the quadratic term is not statistically different from zero at the 5% level of significance. Thus, the quadratic term is significant in this polynomial regression. The sequential hypothesis testing stops here since the coefficient of the highest degree term in regression 5 is statistically significant. Thus, the quadratic nonlinear regression is most appropriate by the sequential hypothesis testing algorithm for years of experience. This is similar to years of schooling as they both possess a quadratic term.