ELLER COLLEGE OF MANAGEMENT
HSLopez School of Business Analytics

ANALYTICS REPORT

TO: ECONOMIC RESEARCH TEAM

FROM: JAKE MOORE

SUBJECT: NONLINEAR SALARY GROWTH ANALYSIS

DATE: NOVEMBER 4^{TH}, 2025

Introduction
In this brief analysis, we examined data from 140 state universities to understand how starting salaries affect mid-career salaries among college graduates. The purpose of this analysis is to identify which nonlinear model best captures this relationship. This is useful as it allows students and organizations to better predict long-term salary outcomes based on early-career pay.

Four nonlinear regression models were tested including: Quadratic, Lin-Log, Log-Lin, and Log-Log. Each model provided a unique way to model the relationship between starting salary and mid-career salary. For each model, we estimated the regression equation, interpreted the coefficients, and evaluated the model fit using R^2, Adjusted R^2, and Standard Error.

The results concluded that while all four models did a decent job at explaining the relationship between both salary variables- the Log-Log model proved to be the most effective with the highest adjusted R^2 (0.7649), and the lowest standard error.

Based on our findings, we recommend using the Log-Log model when predicting mid-career salary outcomes from starting salary data as it provides both accuracy and the lowest variability.

Data Analysis
This section evaluates four nonlinear regression models including Quadratic, Lin-Log, Log-Lin, and Log-Log to determine which explains the relationship between starting salaries and mid-career salaries. Each model below includes an equation, interpretation, a goodness-of-fit evaluation and other insightful information. This section finishes with a selection of the best model based on their adjusted R^2 and standard error value. It concludes with an example case of a starting salary of $54,000 and a prediction using the best overall model.

<u>Quadratic Model:</u>
Regression Equation

$$\widehat{MCMS} = -3977.26 + 1.8294(SMS) + 1.2380E\text{-}06(SMS)^2 + \varepsilon$$

Quadratic Model Shape:
This model is concave up as the coefficient of the squared term is positive (1.24E-06). This means that mid-career salary increases as starting salary rises, although the effect size is relatively small.

Marginal Effect Equation

$$Marginal\ Effect = 1.83 + 2(1.24E\text{-}06)(SMS)$$

Marginal Effect Example
As an example, the marginal effect for someone with a starting salary of $54,000 would be

$$Marginal\ Effect = 1.83 + 2(1.24E\text{-}06)(54000) = 1.964$$

Coefficient Interpretation
As starting salary increases by $1,000 from $54,000 to $55,000, mid-career salary increases by $1,964, on average and all else constant. Because the slope of a quadratic function is not constant and can change across different ranges, the marginal effect is used to estimate the slope at a specific value of the independent variable. This equation provides the instantaneous rate of change in mid-career salary for a one-unit change in starting salary. In this model, b_1 represents the coefficient of the linear term (SMS), and b_2 represents the coefficient of the squared term (SMS^2). It is a point estimate, serving as an approximation of the slope at one specific value rather than representing a constant slope across all values.

<u>Lin-Log Model</u>
Regression Equation

$$\widehat{MCMS} = -858125.74 + 87676.16\ln(SMS) + \varepsilon$$

<u>Model Fit:</u>
Coefficient Interpretation
As starting salary increases by 1%, mid-career salary increases by $876.76, on average and all else constant.

R^2 Interpretation
An R^2 of 0.761 indicates that the model explains about 76.1% of the variation in mid-career salary, meaning we are 76.1% of the way toward perfectly predicting mid-career salary using this model.

Standard Error Interpretation
The standard error of $4,648.95 represents the average difference between the observed and predicted values of mid-career salary.

<u>Log-Lin Model</u>
Regression Equation

$$ln(\widehat{MCMS}) = 10.220 + 2.39E\text{-}05(SMS) + \varepsilon$$

<u>Model Fit:</u>
Coefficient Interpretation
As starting salaries increase by $1,000, mid-career salaries increase by 2.4%, on average and all else constant.

R^2 Interpretation

An R^2 of 0.750 indicates that the model explains about 75% of the variation in mid-career salary, meaning we are 75% of the way toward perfectly predicting mid-career salary using this model. Following a recalculation using unlogged mid-career salary values, the R^2 increased slightly to 0.761, meaning we are 76.1% of the way toward perfectly predicting mid-career salary using this model.

Standard Error Interpretation
The standard error of 0.059 represents the average deviation between observed and predicted values of mid-career salary. When converted to a percentage it corresponds to a 6% prediction error. This means the models predictions of mid-career salary are, on average, within 6% of the actual salary values.

<u>Log-Log Model</u>
Regression Equation

$$ln(\widehat{MCMS}) = -0.297 + 1.082ln(SMS) + \varepsilon$$

<u>Model Fit:</u>
Coefficient Interpretation
As starting salaries increase by 1%, mid-career salaries increase by 1.082% on average and all else constant.

R^2 Interpretation
An R^2 of 0.753 indicates that the model explains about 75.3% of the variation in mid-career salary, meaning we are 75.3% of the way toward perfectly predicting mid-career salary using this model.
Following a recalculation using unlogged mid-career salary values, the R^2 increased slightly to 0.766, meaning we are 76.66% of the way toward perfectly predicting mid-career salary using this model.

Standard Error Interpretation
The standard error of 0.059 represents the average deviation between observed and predicted values of mid-career salary. When converted to a percentage it corresponds to a 6% prediction error. This means the models predictions of mid-career salary are, on average, within 6% of the actual salary values.

<u>Best Overall Model</u>

Recalculating R^2 and Adjusted R^2
Adjusted R^2 was recalculated for both the Log-Lin and Log-Log so they could be directly compared to the Quadratic model. The original R^2 values could not be compared fairly because the models used different dependent variables. For example, MCMS for the Quadratic model and ln(MCMS) for the Log-Log and Log-Lin model. By exponentiating the predicted ln(MCMS) we converted it to MCMS and allowed us to put them both on the same scale. This recalculation allowed us to compute new R^2 and adjusted R^2 values for both Log-Lin and Log-Log models, which gives us a much fairer comparison across all four regression models.

Model Comparison and Selection

After calculating all the adjusted R^2 values, the Log-Log model achieved the highest score of 0.7649, slightly above the Quadratic model (0.7631) and both Log-Lin (0.7593) and Lin-Log (0.7592) models.
In addition, the Log-Log also had one of the lowest standard errors (0.0587), which indicates it is relatively accurate and consistent.
Due to all these factors, the Log-Log model is the best overall model. It offers the best statistical fit as it is the most accurate and has the strongest predicting power.

Overview of Log-Log Model
An R^2 of 0.7649 indicates that the model explains about 76.6% of the variation in mid-career salary, meaning we are 76.6% of the way toward perfectly predicting mid-career salary using this model. The standard error of 0.059 represents the average deviation between observed and predicted values of mid-career salary. When converted to a percentage it corresponds to a 6% prediction error. This means the models predictions of mid-career salary are, on average, within 6% of the actual salary values.

Example Prediction Using Log-Log Model
To see how the model could be used to predict, here is a prediction of the mid-career salary for someone with a starting salary of $54,000:

$$\widehat{MCMS} = e^{-0.30+1.082\ln(54{,}000)} = e^{11.49}$$
$$\widehat{MCMS} = e^{11.49+0.059^2/2} = e^{11.49+0.0017405}$$
$$\widehat{MCMS} = e^{11.49+0.00174} = e^{11.492}$$
$$\widehat{MCMS} = e^{11.492} = 97{,}930.41$$
$$\widehat{MCMS} = 97{,}930.41$$

Therefore, for a starting salary of $54,000, the predicted mid-career salary using the best fit Log-Log model is $97,930, on average and all else constant.

Conclusion
In this analysis, we examined the data from 140 state universities to determine which nonlinear model best predicts mid-career salary based on starting salary. The regression results proved that all four models were effective in explaining this relationship. However, the Log-Log model provided the best overall fit with the highest adjusted R^2 (0.7649) and the lowest standard error.

While the model explains roughly 76% of all variations in mid-career salary, there are still a lot of unmeasured factors that we did not take into consideration in this brief analysis. Variables like college, state, occupation, degree type, or even city could greatly influence these results and add a lot more variability. Despite these limitations, the Log-Log model provides a strong, reliable prediction, of mid-career salary based on the available variables.

Please feel free to contact me at jakemoore@arizona.edu if you have any questions or would like to discuss these recommendations in more detail.

Technical Appendix

Figure 1 – Quadratic Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8755142
R Square	0.766525115
Adjusted R Square	0.763116722
Standard Error	4610.513173
Observations	140

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	9561044055	4780522028	224.8934409	5.30215E-44
Residual	137	2912185945	21256831.71		
Total	139	12473230000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-3977.255185	27425.15691	-0.145022149	0.884906522	-58208.61647	50254.1061
Starting Median Salary	1.829422169	1.196048492	1.529555182	0.128432577	-0.535681451	4.194525789
SMS^2	1.23797E-06	1.29877E-05	0.095318356	0.924201243	-2.44443E-05	2.69203E-05

Jake Moore

Figure 2 – Lin-Log Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.872286031
R Square	0.76088292
Adjusted R Square	0.759150187
Standard Error	4648.953705
Observations	140

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	9490667664	9490667664	439.1231397	1.03209E-44
Residual	138	2982562336	21612770.55		
Total	139	12473230000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-858125.7437	44772.13067	-19.16651566	1.00644E-40	-946653.8363	-769597.651
ln(SMS)	87676.1647	4183.96827	20.9552652	1.03209E-44	79403.18942	95949.13998

Jake Moore

Figure 3 – Log-Lin Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.866026441
R Square	0.750001796
Adjusted R Square	0.748190215
Standard Error	0.058992988
Observations	140

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.440805293	1.440805293	414.0039654	2.24027E-43
Residual	138	0.480263831	0.003480173		
Total	139	1.921069123			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	10.22028945	0.052492575	194.6997145	2.7691E-170	10.11649569	10.32408321
Starting Median Sala	2.38533E-05	1.17232E-06	20.34708739	2.24027E-43	2.15352E-05	2.61713E-05

Jake Moore

Figure 4 – Log-Log Regression Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.867614096
R Square	0.75275422
Adjusted R Square	0.750962584
Standard Error	0.05866734
Observations	140

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.446092889	1.446092889	420.1490605	1.04176E-43
Residual	138	0.474976234	0.003441857		
Total	139	1.921069123			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-0.29714426	0.565000638	-0.525918451	0.599788996	-1.414322023	0.820033502
ln(SMS)	1.082259125	0.052799469	20.49753791	1.04176E-43	0.977858547	1.186659703

Jake Moore

Figure 5 – Adjusted R^2 Comparison

MCMS as y-variable	Adj. R^2
Quadratic	0.763
Lin-Log	0.759
Quadratic	0.763

ln(MCMS) as y-variable	Adj. R^2
Log-Log	0.751
Recalcuated	0.765
Difference	0.014
Log-Lin	0.748
Recalcuated	0.759
Difference	0.011

Adjusted. R^2 Values	
All Models MCMS Scale	
Quadratic	0.7631
Lin-Log	0.7592
Log-Lin	0.7593
Log-Log	0.7649

Jake Moore
Highest
0.7649
=MAX(E12:E15)
Log-Log Model