The analysis focuses on earnings among computer programmers, software developers, and web developers. These occupations are particularly interesting due to their high demand in the labor market, their strong association with technological advancements, and the varying skill levels and education requirements necessary for entry. Moreover, these fields exhibit notable gender disparities in employment and wage outcomes, making them valuable for studying the determinants of earnings.

To examine the factors influencing hourly earnings, several independent variables were selected. Gender is included to capture potential wage differences between male and female workers. Age and age squared are introduced to account for the non-linear relationship between experience and wages. Education variables, specifically the attainment of a master's (MA) or doctoral (PhD) degree, are incorporated to assess their impact on earnings. Interaction terms between gender and education levels help to determine whether advanced education benefits men and women differently. Finally, employment sector variables distinguish between private and government employment, as these sectors often exhibit different compensation structures.

The analysis includes four ordinary least squares (OLS) regression models with increasing complexity. The first model includes only gender and age as predictors, providing a basic understanding of earnings determinants. The second model introduces a quadratic term for age, improving the model's ability to capture non-linear wage growth. The third model further incorporates education and gender-education interactions, allowing for a more detailed exploration of human capital effects. Finally, the fourth model includes employment sector variables, accounting for structural differences in wage determination between private and government employment.

Model performance is assessed using root mean squared error (RMSE), and Bayesian Information Criterion (BIC). The results indicate that adding predictors improves model fit incrementally. RMSE decreases from 0.477 to 0.459, suggesting a reduction in prediction error. However, BIC, which penalizes model complexity, reaches its lowest value in Model 2 (2698) before slightly increasing in Models 3 and 4, implying that the additional predictors may contribute marginally to explanatory power relative to their complexity. This reflects the trade-off between bias and variance: simpler models may underfit the data (high bias), while more complex models risk overfitting (high variance).

Cross-validation results further reinforce this trend. The average RMSE across five folds stabilizes around 0.46 for Models 2, 3, and 4, showing diminishing returns in predictive accuracy despite additional variables. This suggests that while adding more predictors enhances explanatory power, it does not necessarily lead to substantial improvements in out-of-sample predictive performance.

In conclusion, the study highlights the trade-off between model complexity and performance. While including education and employment sector variables provides valuable insights into earnings determinants, the marginal improvements in prediction accuracy suggest that a more simplified model may be preferable (Occam's razor).