

A Regression Primer for Labor Practitioners

By John Johnson and Kristin Terris

Labor and employment litigation frequently involves complex quantitative issues, such as measuring wage differences between men and women or comparing promotion outcomes for different groups of workers. Econometrics provides techniques for reliably addressing these issues. This discipline combines statistical methods with economic theory to derive and test hypotheses and measure underlying relationships between variables. Powerful statements can result from correctly applying its basic principles. In fact, regression techniques are increasingly used in employment litigation as the courts begin to recognize the limited application of simple averages.

Consider a simple example from a stylized employment case. Assume we observe a company where the average wage rate of males is three dollars per hour

can help distinguish between the two hypotheses. If the wage gap is due to a higher percentage of men in management positions, then there should be no remaining wage gap once the management job characteristic is controlled for in the model.

To build a regression model for this simple example, an econometrician would perform the following tasks:

1. Formulate the question.

Are wages of women lower than wages of men after accounting for all other non-gender-related factors that determine wage rates?

2. Determine the foundational economic theory.

To develop the specific context for the analysis, or the model, for this specific company, the first step is to determine the broader context

Regression techniques are increasingly used in employment litigation as courts recognize that simple averages are less useful.

higher than the average wage rate of females. How can one interpret this observed wage gap?

The gap could be caused by any number of factors. For example, it could be that a higher percentage of men in the company are in management positions than women. Alternatively, women may be paid less than men, even when they hold the same jobs and have similar qualifications.

Differentiating between these two potential explanations with a comparison of simple averages is virtually impossible. Regression analysis, on the other hand, explicitly introduces control variables to explain variation in wages unrelated to gender. Regression

by answering questions like the following: What are the firm's personnel policies? How does compensation work? What are the job requirements for various positions? What are the differences between positions that men and women hold in the firm? Is the local labor market the same for the two groups? (At this point, with a more complex case than this simple example, one might determine that a single regression is not adequate to address the issue because the individuals in the group are too diverse to derive a meaningful result.)

3. Select variables that may affect the dependent variable.

In this example, after studying the

company and framing the regression, the analyst will decide which variables are likely to be important for understanding wages. In this stylized example, types of variables one might consider are years of education, experience, job title, specialized training courses taken, and the like.

4. Collect data on all the relevant variables.

Although companies keep extensive personnel databases, the data does not come "ready to use" for regression analysis. A significant effort is often required to collect and prepare the data so it can be analyzed.

5. Run the regression models.

Regression analysis is run using computer software; running a regression requires some basic computer programming.

6. Interpret the results.

The output of a regression includes both estimates of the effects of the explanatory variables on the dependent variable (called "coefficients") and measures of the statistical precision of the estimates (called "t-statistics"). A regression coefficient provides both relative size and direction of the effect (for example, in addition to testing differences between men and women, the regression equation could tell us that an extra year of education increases wages at this firm by \$10,000).

Econometrics also provides tools for interpreting results (such as hypothesis tests). An estimated effect is considered statistically significant when it is unlikely that the estimate could have been generated by random chance (statistical variation) rather than something systematic. Conversely, if an estimated effect is not statistically significant, it means there very well may be no true systematic effect, and instead the estimated effect was generated by random chance.

Thus, a regression is a powerful

methodology because it not only estimates the size and direction of a variable's effect, but it also determines whether this estimated effect is statistically significant.

Back to the simple example: assume the company had offered a special training course to five of its employees. When one measures the effect of this course in the regression model, one finds that wages of employees who took the course were \$25,000 higher than those who did not take it. However, this effect is not statistically significant, meaning that the \$25,000 difference might well have been generated by chance, so one cannot conclude with confidence that the course had any effect at all.

7. Test the specification of the model.

The process of determining whether one has a credible result involves testing the regression model. The specification of a model includes the variables, the form of the underlying equation, and other properties of the model. If the results of the model hold when small changes are made to the model, then it is robust. Additionally, there are statistical tests that can determine model reliability. It is important to conduct these tests to identify the strengths and weakness of the analysis.

Labor and employment law practitioners—including attorneys, judges, and juries—are increasingly confronting regression techniques. A careful regression analysis can provide value through both the testing of hypotheses and the measurement of effects. When reliably applied, econometrics provides objective and scientific methods for addressing quantitative issues. ■

John Johnson (john.johnson@nera.com) and **Kristin Terris** (kristin.terris@nera.com) are labor economists at NERA Economic Consulting in Washington, D.C., and Los Angeles, California, respectively.