

# Regression Analysis Applications in Litigation

By

*Robert Mills*  
Principal  
Micronomics, Inc.

*Dubravka Tasic, Ph.D.*  
Principal  
ERS Group

Practising Law Institute  
Pocket MBA: Finance for Lawyers

Summer 2011



# Regression Analysis Applications in Litigation

*Robert Mills*<sup>\*</sup>  
*Dubravka Tosic, Ph.D.*<sup>\*\*</sup>

March 2011

## I. Introduction to Regression Analysis

Regression analysis is a statistical tool used to examine relationships among variables. It provides a method for quantifying the impact of changes in one or more explanatory variables (known as independent variables) on a variable of interest (known as the dependent variable). Regression analysis is widely used in the field of econometrics, which is concerned with the application of statistical and mathematical methods to the analysis of economic data.<sup>1</sup> Useful applications also are found in finance, sociology, biology, psychology, pharmacology, and engineering, among other fields of study. In this paper, we provide an introduction to regression analysis and discuss a number of applications in the litigation context.

Regression analysis begins with a hypothesis. Suppose, for example, that we are interested in understanding factors that impact attendance at a sporting event. We might hypothesize that historical performance of the home team influences attendance.

---

<sup>\*</sup> Robert Mills is a Principal at Micronomics, Inc., an economic research and consulting firm in Los Angeles, California. Micronomics is a subsidiary of ERS Group, a national economic and statistical consulting firm.

<sup>\*\*</sup> Dubravka Tosic, Ph.D. is a Principal at ERS Group, and based in New York/New Jersey.

<sup>1</sup> Additional information can be found in an econometrics textbook such as James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 3rd ed. (Upper Saddle River: Prentice-Hall, 2010); William H. Greene, *Econometric Analysis*, 7th ed. (Upper Saddle River: Prentice-Hall, 2011); or Peter Kennedy, *A Guide to Econometrics*, 5th ed. (Cambridge: The MIT Press, 2003).

We might further believe that the relationship between historical performance and attendance is positive; that is, improvements in performance of the home team lead to greater attendance and declines in performance of the home team lead to lower attendance. Assuming historical attendance and home team performance data are available, we can estimate the following model:

$$A_i = \alpha + \beta P_i + \varepsilon_i$$

where

$A_i$  = attendance at game  $i$  (the dependent variable);

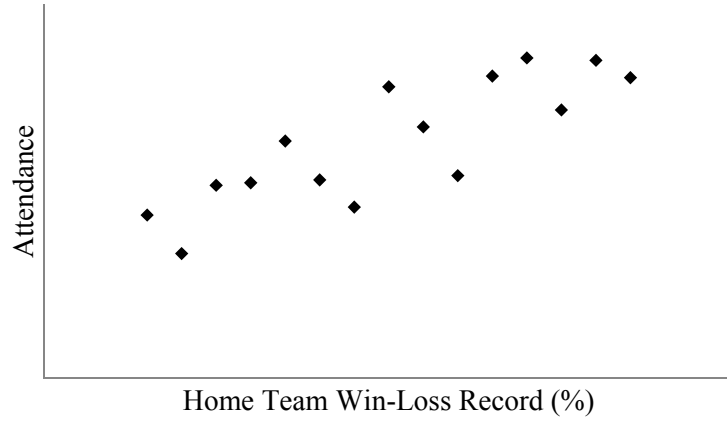
$P_i$  = home team performance as of game  $i$  measured by the win-loss record expressed as a percentage (the independent variable);

$\alpha$  = constant amount (interpreted as attendance given a win-loss record of zero percent);

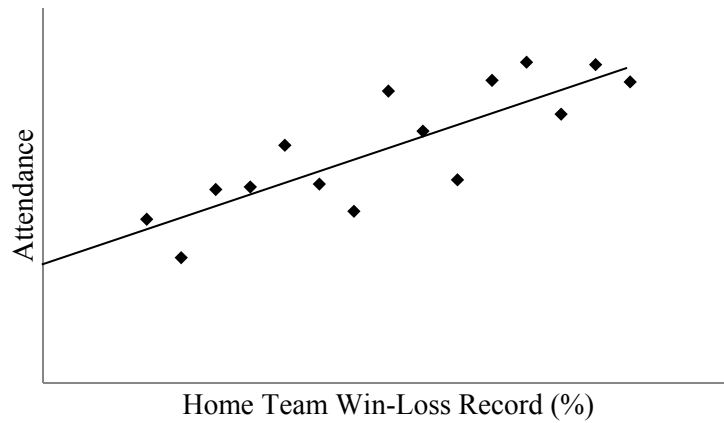
$\beta$  = the effect in attendance of each additional percentage in the home team win-loss record; and

$\varepsilon_i$  = a “disturbance” term reflecting other unmeasured factors that influence attendance.

Data for  $A$  and  $P$  are plotted in the following figure. The coefficients  $\alpha$  and  $\beta$  are not known. Regression analysis produces estimates for these coefficients, which customarily are denoted with a “hat” superscript (e.g.,  $\hat{\alpha}$  and  $\hat{\beta}$ ). The disturbance term,  $\varepsilon$ , also is unknown.



Graphically, estimation of the coefficients  $\alpha$  and  $\beta$  is tantamount to fitting a line to the attendance and home team win-loss record data, where  $\hat{\alpha}$  is the point at which the line intersects the vertical axis and  $\hat{\beta}$  is the slope of the line. The following figure depicts such a line.



This line appears to fit the data. Without an objective criterion, however, there is no guarantee that this line provides the best fit. Regression analysis provides a criterion. With regression analysis,

the intercept and slope of the line (i.e.,  $\alpha$  and  $\beta$ ) typically are estimated by minimizing the *sum of squared errors* (“SSE”).

First, an estimated error for each observation is measured as the vertical distance between the observed value of the variable and the estimated line. SSE is calculated by squaring this estimated error for each observation and summing across all observations. Estimates of the coefficients are chosen to minimize SSE. This is called the method of ordinary least squares. In practice, this estimation is carried out using regression software. With ordinary least squares the best fitted line for the data is estimated.

Common knowledge suggests that attendance at sporting events increases with improvements in home team performance. In other words, we expect a positive coefficient for home team win-loss record ( $\hat{\beta}$ ) indicating that attendance increases as performance improves and attendance decreases as performance declines, other things equal.

Estimating our model produces the following results.

Regression Output			
	Coefficient	Standard Error	t-Statistic
$R^2 = 0.70$			
<i>Intercept</i> ( $\hat{\alpha}$ )	25,419	4,913	5.17
<i>Win-Loss Record</i> ( $\hat{\beta}$ )	501	90	5.55

The estimated coefficient for the home team win-loss record is 501, which is interpreted as the estimated number of additional attendees for every one percent improvement in the home team win-loss record. This estimate is consistent with our expectation that the coefficient is positive. The intercept term is interpreted as the estimated number of attendees given a home team record of zero wins. Using these coefficient estimates, attendance can be predicted for any given home team win-loss record. For example,

if the win-loss record is 50% as of game  $i$ , estimated attendance at game  $i$  is

$$50,469 = 25,419 + (501 * 50).$$

The model suggests attendance would increase to 62,994 in the event that the home team win-loss record improved to 75%:

$$62,994 = 25,419 + (501 * 75).$$

The results of the regression analysis appear to confirm our *a priori* belief that attendance increases with improvements in home team performance. Using the  $t$ -statistic reported above, we can formally test the hypothesis that performance does not impact attendance. Operationally, this test involves comparing the reported  $t$ -statistic for the coefficient of interest to the critical value obtained from the  $t$  distribution. Courts have frequently adopted the concept of statistical “significance” when assessing the importance of a variable. Assuming a large sample size, the critical value is 1.96 (or approximately two standard deviations) at the five percent level of significance. Since the reported  $t$ -statistic of 5.55 exceeds the critical value of 1.96, we can reject the hypothesis that performance does not impact attendance at a five percent level of statistical significance.

Another useful statistic frequently reported with regression results is the coefficient of determination, or R-squared ( $R^2$ ).  $R^2$  reflects the proportion of total variation in the dependent variable explained by variation in the independent variable or variables. In other words, it provides a measure of the “explanatory power” of a model.

The value of  $R^2$  ranges from 0 to 1, with a value of 0 meaning that none of the variation in the dependent variable is explained by variation in the independent variables and a value of 1 signifying that all of the variation in the dependent variable is explained by variation in the independent variables. Roughly speaking, a high value of  $R^2$  often is associated with a good fit of the regression line whereas a low value of  $R^2$  is associated with a poor fit. This does not mean, however, that the relative strength of two competing

models can be assessed by their respective  $R^2$  values alone. Indeed, introducing additional independent variables into a model will tend to increase the value of  $R^2$  even where those variables have no hypothesized relationship with the dependent variable.

Specification of the regression model should be founded in theory. Explanatory variables should not be included in a model without a theoretical basis for inclusion. Similarly, explanatory variables that theory suggests are relevant should not be excluded without a sound basis for doing so. The exclusion of relevant explanatory variables from a model without basis is particularly problematic because it can lead to *omitted variable bias*.

Omitted variable bias arises when one or more variables that should be included in a model are excluded from the estimated model. In such cases, coefficient estimates for the included variables can be biased and the results of hypothesis tests rendered unreliable.<sup>2</sup>

Turning back to the sporting event attendance example, the only explanatory variable we have considered is home team performance as measured by the home team win-loss record. Clearly, this is an overly simplistic view of the determinants of attendance. Attendance likely is affected by a variety of factors in addition to home team performance. Economic theory suggests, for example, that ticket sales depend in part on ticket prices. The win-loss record of the visiting team, the number of games left in the season, the day of the week on which the event occurs, and game day rainfall (particularly for outdoor events) might also be relevant. Each of these variables is subject to measurement and could be included in the model. The problem of omitted variable bias is more troublesome when the omitted variable is not readily subject to measurement and therefore cannot be included.

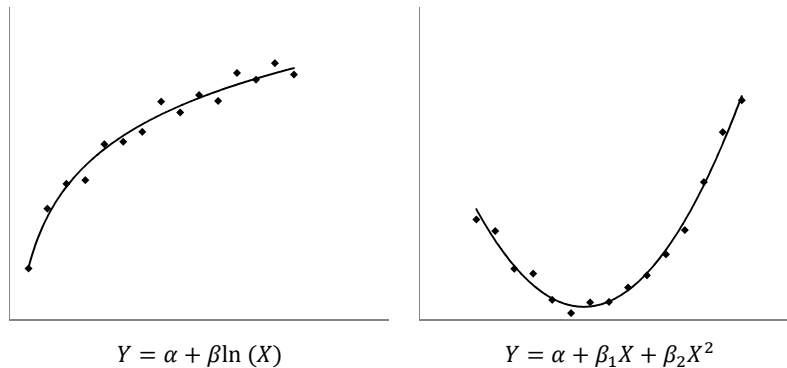
The problem of including irrelevant variables typically is less serious than the problem of omitting relevant variables because the

---

<sup>2</sup> Coefficient estimates for included variables will remain unbiased in the event that the omitted variable is uncorrelated with all of the included variables.

inclusion of irrelevant variables does not serve to bias estimates of the coefficients for relevant variables. This is not to say that the practice of including irrelevant variables in a model is without cost. The efficiency of the estimator is affected by including irrelevant variables, which can be problematic particularly when working with small sample sizes.

While identifying relevant explanatory variables is an essential aspect of model specification, the choice of functional form also is important. Thus far we have assumed that the relationship between the dependent variable and the independent variables is linear. Depending upon the application, theory may suggest that a nonlinear functional form is more appropriate. The left and right panels of the following figure illustrate nonlinear functional forms commonly used in practice. The left panel depicts a semi-log model and the right panel depicts a polynomial model.



Many nonlinear functional forms, including those shown above, can be estimated using standard linear regression techniques because they are linear in the coefficients. Nonlinear functional forms that are not subject to linear transformations require more sophisticated estimation techniques.

There are a number of assumptions that underlie the standard linear regression model. It is important to recognize situations

where these assumptions are violated so that alternative methods can be employed to produce sound results. It is beyond the scope of this paper to provide a comprehensive overview of all of the problems that can arise. Instead, we will focus on two common problems that are related to the disturbance term.

It typically is assumed that the disturbance term is composed of small, individually unimportant effects that are independently distributed from a normal population with an expected value of zero and constant variance.

Violations of the constant variance assumption are not uncommon in practice. When working with data from a cross section of firms in an industry, for example, a systematic difference between the disturbances for large and small firms may exist indicating that variance of the disturbance term is not constant. Disturbances are said to be *heteroscedastic* when they have different variances.

Violations of the independence assumption sometimes arise when working with time series data because the disturbance associated with observations for one period may carry over into future periods. Disturbances are said to be *serially correlated* when the disturbance terms for different observations are correlated.

In the presence of heteroscedasticity or serial correlation, the method of ordinary least squares produces coefficient estimates that are unbiased but not efficient. The loss of efficiency is undesirable because it can affect the results of hypothesis tests. Fortunately, procedures for identifying and correcting problems associated with heteroscedasticity and serial correlation are readily available.

## II. Examples of Practical Applications of Regression Analysis

The discussion thus far is intended to provide non-practitioners a brief introduction to regression analysis. We now introduce some practical applications of regression analysis in the litigation context. Specifically, we provide an overview of (A) the role of regression analysis in estimating price elasticity of demand in antitrust and intellectual property matters, (B) use of regression

analysis to conduct event studies designed to estimate the impact of specific events on the value of a firm, (C) the application of regression analysis to cost estimation in damages studies, and (D) applications of regression analysis in labor and employment disputes.

#### *A. Price Elasticity of Demand*

Demand refers to the quantity of a good or service consumers purchase at prevailing prices. Increases in the prevailing price of a good tend to result in reduced sales volume because some consumers will choose alternative products or refrain altogether from making a purchase as price increases. Conversely, decreases in the prevailing price tend to result in sales volume increases. The term price elasticity of demand refers to the extent to which sales volume is affected by price changes.

Own-price elasticity of demand measures the responsiveness of the quantity of a good demanded to changes in its price. Demand is said to be elastic if quantity demanded is highly sensitive to changes in price, and inelastic if price changes have little impact on quantity demanded. Cross-price elasticity of demand measures the responsiveness of quantity demanded for one good to changes in the price of another good.

Own-price elasticity of demand is negative since price increases lead to decreases in quantity demanded. This elasticity commonly is reported in terms of absolute value, however, and the negative sign can be assumed. Cross-price elasticity of demand can be positive or negative depending upon whether the goods are substitutes (positive cross-price elasticity) or compliments (negative cross-price elasticity). Together, own- and cross-price elasticity summarize anticipated substitution patterns among consumers faced with changes in price.

The concept of price elasticity of demand has been widely used in litigation, notably in assessing potential anticompetitive effects of mergers. Own- and cross-price elasticity are routinely used to define relevant antitrust markets, assess market power, and

simulate price increases resulting from mergers before they are consummated.

Use of price elasticity of demand also has emerged in patent infringement litigation, particularly in cases where price erosion is alleged to have occurred. An assessment of price erosion involves estimating the price that would have prevailed but for the infringement and then determining the amount of sales the patent owner would have made at that price. Although a patent owner may have been able to charge a higher price in the absence of the infringement, its sales might have been lower depending upon the price elasticity of demand.

Measures of price elasticity of demand commonly are derived by estimating one or more demand curves using regression analysis. Economic theory suggests the quantity of a good demanded depends upon its price, the price of substitutes and complements, and income, among other possible factors. In practice, data limitations may dictate which variables are included in a regression analysis, but the potential for omitted variable bias also should be considered when specifying models.

Suppose we have monthly data on the quantity of goods sold ( $Q_1$  and  $Q_2$ ) and corresponding price data ( $P_1$  and  $P_2$ ) for two substitute goods. We also have monthly income data ( $I$ ) for consumers that purchase the goods. We can estimate the following linear demand equations using regression analysis.

$$Q_1 = \alpha_1 + \beta_{1,1}P_1 + \beta_{1,2}P_2 + \beta_{1,3}I + \varepsilon_1$$

$$Q_2 = \alpha_2 + \beta_{2,1}P_1 + \beta_{2,2}P_2 + \beta_{2,3}I + \varepsilon_2$$

We use a linear demand model in this example for simplicity. Economic theory does not dictate an exact functional relationship between quantity demanded and the variables that impact demand. The properties of a specific functional form may lead the researcher to believe it superior for a given situation, but the choice is often somewhat arbitrary. If sufficient data are available, a variety of functional forms might be estimated to assess the sensitivity of the results to the choice of functional form. This

practice may lend credibility to the results if they are shown to be insensitive to the choice of functional form. Results that are extremely sensitive to functional form may prove difficult to defend.

Price elasticity for both goods can readily be estimated using the estimated coefficients from the linear demand model. Own-price elasticity is equal to the “first partial derivative” of the demand equation with respect to price times price divided by quantity.

$$\epsilon_1 = \frac{\partial Q_1}{\partial P_1} \cdot \frac{P_1}{Q_1} = \beta_{1,1} \frac{P_1}{Q_1}$$

$$\epsilon_2 = \frac{\partial Q_2}{\partial P_2} \cdot \frac{P_2}{Q_2} = \beta_{2,2} \frac{P_2}{Q_2}$$

In other words, own-price elasticity of demand is equal to the coefficient for the price variable multiplied by price which is divided by quantity. Cross-price elasticity of demand is calculated as the coefficient for the price of the other good multiplied by the price of the other good divided by quantity.

$$\epsilon_{1,2} = \frac{\partial Q_1}{\partial P_2} \cdot \frac{P_2}{Q_1} = \beta_{1,2} \frac{P_2}{Q_1}$$

$$\epsilon_{2,1} = \frac{\partial Q_2}{\partial P_1} \cdot \frac{P_1}{Q_2} = \beta_{2,1} \frac{P_1}{Q_2}$$

Price elasticity estimates can prove useful in the litigation context, particularly in cases where the interplay between price and quantity is an issue. In antitrust litigation, for example, elasticity and cross-price elasticity are often used to delineate relevant markets. Firms are likely to be grouped in the same market if the products they produce can be used interchangeably and where the products exhibit a high cross-price elasticity of demand. In cases where price allegedly would have been higher (or lower) in the absence of some conduct, elasticity estimates can be used to show the impact of that but-for price on quantity demanded.

### *B. Event Study Analysis*

Event studies measure the impact of specific events on the value of firms. There are many useful applications for event studies in litigation settings. For example, event studies are commonly used to estimate the impact of adverse information on movements in share prices in matters of alleged securities fraud. They also can provide insight into damages resulting from events such as product recalls, the loss of patent protection, credit facility constraints, and fraud.

The basic premise underlying an event study analysis is that given rational market participants, security prices will quickly adjust to reflect the announcement of an event. Roughly speaking, security price changes are attributable to company-specific information (such as the announcement of a new product) and industry or market-wide information (such as new regulation or changes in interest rates). Event study analysis provides a framework for isolating the impact of company-specific events on security prices. The total impact of an event can then be estimated by summing the company-specific impact across all of shares affected.

The first step in undertaking an event study analysis involves the identification of the event or events of interest. In the litigation context, the events of interest often are dictated by allegations in the complaint. Suppose, for example, that a publicly traded early-stage pharmaceutical company alleges that clinical trials for a potential new therapeutic drug were unsuccessful as a result of a failure on the part of its development partner to design a proper test protocol. In this example, the event of interest is the public announcement that the clinical trials were unsuccessful.

After the event of interest has been identified, it is necessary to determine the period of time over which the impact will be measured. This is called the *event window*. In practice, the event window typically is defined to include at least the day on which the event was announced and the following business day. Depending upon the circumstances, the event window may commence before the event is announced (e.g., if there is reason to believe that news of the event leaked before the official

announcement) and end days after the event is announced (e.g., if there is reason to believe that some market participants did not immediately learn of the event at the time it was announced). The event window ideally will be long enough to include any ongoing adjustment to news of the event in the market, but not so long as to capture effects of unrelated subsequent events.

A primary objective of event study analysis is to isolate the impact of the event in question from market-wide and industry-wide information that also impacts securities prices. The following model is often used in this context:

$$R_t = \alpha + \beta I_t + \varepsilon_t$$

where

$R_t$  = the security return on day  $t$  for the company of interest;

$I_t$  = the market index return on day  $t$ ;

$\alpha$  = the intercept coefficient;

$\beta$  = the market index coefficient; and

$\varepsilon_t$  = a disturbance term reflecting other factors that influence the security return for the company of interest.

Historical stock price data for the company in question are collected and daily returns are calculated. Market index data also are collected. This market index may be a widely available index such as the Standard and Poor's 500 or a custom index that includes peers of the company of interest. Returning to our early-stage pharmaceutical company example, a useful market index might be constructed to include other publicly traded early-stage companies involved in clinical trials for potential new therapeutics.

Regression analysis is employed to obtain estimates for  $\alpha$  and  $\beta$ . The results of the regression analysis are then used to calculate the predicted security return,  $\hat{R}_t$ , for each day in the event window:

$$\hat{R}_t = \hat{\alpha} + \hat{\beta}I_t.$$

The predicted security return is essentially an estimate of the security return but for the event in question. Predicted security returns are compared to actual returns to determine the impact of the event in question. The difference between the actual and predicted return on any particular day is called the *abnormal return*:

$$\text{abnormal return} = R_t - \hat{\alpha} - \hat{\beta}I_t.$$

Summing abnormal returns across all days in the event window yields *cumulative abnormal returns*:

$$\text{cumulative abnormal returns} = \sum_{t=j}^k R_t - \hat{\alpha} - \hat{\beta}I_t.$$

Cumulative abnormal returns (or CAR) speak to the magnitude of the event in question.

Event studies can also shed light on the materiality of events. Materiality is addressed using statistical testing. A common question in event studies is whether or not the hypothesis that the cumulative abnormal returns are zero can be rejected. Output obtained from the regression analysis provides the information necessary to conduct such a test.

Event study analysis has been used in a wide range of investigations. In the litigation context, it is used to estimate damages caused by securities fraud and other wrongful conduct. Event studies have also been used to understand the value created by mergers and acquisitions, the impact of corporate earnings restatements, and market reactions to jury verdicts.

### *C. Cost Estimation in Damages Studies*

In many cases, lost profits damages are calculated as the difference between profits that would have been generated but for some alleged conduct, such as a breach of contract, and actual profits generated given the conduct. Estimating but-for profits requires an understanding of the costs involved, and in particular those costs that were not incurred given the alleged conduct but would have been incurred in the absence of the alleged conduct. These costs are sometimes referred to as avoided costs. The estimation of avoided costs often requires an understanding of the distinction between those cost elements that are fixed and those that are variable.

Fixed costs do not vary with levels of output. Costs that frequently are fixed over moderate changes in output include rent, insurance premiums, business license fees, and salaries for permanent full time employees.

Variable costs are those that vary directly with the level of output. Depending upon the nature of the business, variable costs may include cost of goods sold, shipping charges, royalties, and sales commissions, among others.

Certain costs cannot be classified as strictly fixed or variable. These semi-variable costs include a mixture of fixed and variable components. Common examples of semi-variable costs include production labor (regular wages are fixed but overtime is variable), electricity, telephone bills, and postage.

An important consideration when assessing the nature of costs is that cost elements can be fixed over certain levels of output and variable over other levels of output. To illustrate this point, suppose a manufacturer has the capacity to increase production by ten percent without expanding its plant, but any increase in production above ten percent would require an expansion. In this example, the rent associated with the plant is fixed over relatively small increases in output. Increasing output by more than ten percent, however, would require an expansion of the plant and the

payment of additional rent. In other words, rent is a variable cost in this example over large increases in output.

Discussions with company management and accounting personnel can be helpful in understanding the fixed or variable nature of costs. Depending upon the availability of data, regression analysis may provide additional insight.

Regression analysis provides a means to examine and quantify relationships among variables. In the case of cost estimation, a common inquiry is “what is the relationship between changes in output and the cost of production?” Assuming sufficient data are available, the following model might be estimated to address this question:

$$C_t = \alpha + \beta P_t + \varepsilon_t$$

where

$C_t$  = cost of production during period  $t$ ;

$P_t$  = production during period  $t$ ;

$\alpha$  = the intercept coefficient;

$\beta$  = the production coefficient; and

$\varepsilon$  = a disturbance term reflecting other factors that influence the cost of production.<sup>3</sup>

The coefficient  $\alpha$  is interpreted as the cost of production when output falls to zero units. In other words, it provides an indication of the fixed cost of production. The coefficient  $\beta$  is interpreted as the cost of production for one additional unit of output. That is, it provides an indication of the variable cost of production. Together,

---

<sup>3</sup> Depending upon the situation, model specification might be more complicated in practice. Decisions concerning the variables to include, functional form, and data aggregation are driven by the specific facts and circumstances of the investigation.

these coefficients can be used to estimate the total cost of production for a given level of output.

In our example, the regression results might be used to calculate profit but for the alleged conduct:

$$\text{Profit}_{bf} = \text{Sales}_{bf} * (\text{Price}_{bf} - \hat{\beta}) - \hat{\alpha}.$$

Subtracting actual profits from but-for profits would yield an estimate of profits lost as a result of the alleged conduct.

#### *D. Labor and Employment Litigation*

Almost all employers face federal nondiscrimination requirements, and most states also have enacted employment laws specifically dealing with discrimination. These federal and state laws are intended to ensure that employers base employment practices (e.g., hiring, promotion, termination, discipline, compensation) on objective and fair measures, such as performance and merit. Employment discrimination allegations often charge employers with engaging in discrimination against a member or members of a protected class (legally protected characteristics include race, gender, ethnicity, national origin, religion, age, and disability). These allegations require plaintiffs to demonstrate that a pattern or practice of discrimination exists. Statistical analysis is commonly used to analyze such allegations. Various statistical tests can be performed utilizing human resources, payroll, and other business data. Regression analysis can also be employed to identify patterns in data that reflect employment decisions.

Regression analysis may be viewed as a tool that quantifies the relationship between a decision variable and other independent factors. For example, suppose a company faces an employment discrimination matter in which plaintiffs allege that women are being discriminated against in terms of base pay. The hypothesis we would want to test with regression analysis is that gender is not a significant factor in determining the base salary level of employees. The following multiple regression model could be estimated:

$$S_n = \alpha + \beta C_n + \gamma G_n + \varepsilon_n$$

where

$S_n$  = base salary for employee  $n$ ;

$C_n$  = characteristics of employee  $n$ ;

$G_n$  = gender of employee  $n$ ;

$\alpha$  = the intercept coefficient;

$\beta$  = the employee characteristics coefficients;

$\gamma$  = the gender coefficient; and

$\varepsilon_n$  = a disturbance term reflecting other factors that influence base salaries.

This model is referred to as a multiple regression model since multiple explanatory variables are considered. In our example, the dependent variable is base salary and the independent variables are various characteristics of employees that might influence base salary and for which data are available. The employer might contend that the following employee characteristics are important determinants of base salary, and as such should be included in the regression model: education, prior experience, tenure, special skills, department, and geographic region. To test the hypothesis that base salary for women is not different than the base salary for men after controlling for all of these factors, the regression model would also include a variable that reflects the gender of the employee, which is depicted in our model as  $G$ . The constant term,  $\alpha$ , is interpreted as the average base salary paid to a man who has a zero value in each independent variable (e.g., no education, no prior experience, and no tenure). The coefficients  $\beta$  and  $\gamma$  measure the influence of the independent variables on base salary.

Estimates of these coefficients are referred to as unbiased estimates of the influence of the independent variables on the dependent variable if the variables are independent of each other, no

important variables have been omitted, base salary is normally distributed, and other assumptions underlying the method of ordinary least squares hold.

The difference between average base salary for men and women is estimated by the coefficient  $Y$ . If this coefficient is statistically significant (i.e., it has a  $t$ -statistic of more than 1.96 assuming a five percent level of statistical significance), the difference between the base salary for men and women is said to be statistically significant after accounting for other factors included in the regression model. Assuming the regression model controls for factors influencing pay, this result would prompt us to reject the hypothesis that gender is not a significant factor in determining base salary.

Given the widespread availability of computing power and sophisticated computer software, it is possible to generate a wealth of information useful for identifying and examining outliers, testing the robustness of models, and analyzing the sensitivity of results to assumptions made. For instance, significant outliers are often examined to further evaluate the quality of the model and data. Using the base salary example provided above, data pertaining to employees that are identified as statistically significant positive or negative outliers (i.e., employees whose actual base salary is significantly higher or lower than their predicted base salary), could be reviewed to identify potential anomalies in the data. This process can provide information that might be used to further refine the model.

### III. Conclusion

Implementing regression analysis requires an appreciation for the statistical underpinnings of the analysis along with a well-designed model that is founded in theory. When used properly, regression analysis is a powerful tool with many practical applications in litigation. It has been widely accepted by courts as a reliable estimation framework.

*About the Authors*

ROBERT MILLS is an economist and Principal at Micronomics, Inc., an economic research and consulting firm located in Los Angeles, California. Micronomics, Inc. is a subsidiary of ERS Group, a national economic and statistical consulting firm.

Mr. Mills has been engaged in economic research and consulting for the past 15 years. A significant portion of his professional experience has involved the valuation of intellectual property and other assets, industrial organization, and the calculation of economic damages. His experience spans many industries, including software, semiconductors, health care, medical devices, pharmaceuticals, entertainment, telecommunications, real estate, apparel, manufacturing, retail sales, insurance, sporting goods, and energy, among others.

Mr. Mills has served as an expert witness or consultant in a wide range of matters, including patent, trademark and copyright infringement, theft of trade secrets, breach of contract, interference, conversion, fraud, predatory pricing, attempted monopolization, and labor disputes. He has testified as an economic expert in Federal District Court, state courts in multiple jurisdictions, and at arbitration.

Mr. Mills also engages in economic research and consulting outside the context of litigation. He has assessed the anticipated competitive effects of mergers and joint ventures on behalf of government regulatory agencies and merging parties; developed forecasts and strategic recommendations for government agencies and clients involved with real estate development; and assisted clients with the valuation of intangible assets and entire businesses.

Mr. Mills received a Bachelor of Science degree in economics and history from Portland State University and a Master of Arts degree in economics from the University of California at Santa Barbara.

*About the Authors*

DUBRAVKA K. TOSIC is an Economist and Principal at ERS Group, a national leader in economic and statistical consulting. She rejoined ERS Group in Spring 2010, after 12 years as a Director in the Dispute Analysis practice of PricewaterhouseCoopers, LLP in New York. Dr. Tosic has a wealth of experience in leading and managing projects for private and public sector clients, and their in-house and outside counsel, involving economic and quantitative analyses and damage calculations in a wide variety of complex disputes, litigation and arbitration matters, and pro-active risk management and compliance reviews.

Dr. Tosic's primary areas of expertise are labor and employment (employment discrimination litigation involving various employer actions (e.g., hiring, promotion, termination), compensation studies, reductions-in-force analyses, and wage and hour litigation) and complex commercial litigation and disputes. She has provided assistance as consulting expert or testifying expert to address issues of class certification, liability, and estimated damages. Additionally, Dr. Tosic has performed pro-active risk management and compliance reviews, and management consulting projects involving high-profile reform and transformational initiatives involving process reviews and data analytics.

Dr. Tosic received her Ph.D. in economics from Florida State University and her Bachelor's degree from University of Maryland, and previously worked at ERS Group from 1991-1996.

### *About ERS Group*

ERS Group is the preeminent economic and statistical consulting firm for analyses related to employment matters. Founded in 1981, with offices in Tallahassee, Washington, D.C., San Francisco, and Los Angeles, its statistically sound studies provide clients with a better understanding of their organizations and decision-making processes. Its research has been used by clients in high stakes employment litigation and regulatory matters involving allegations of discrimination in hiring, promotion and compensation. ERS Group's national reputation is founded on the unparalleled experience of its economists and testifying experts. Its reach extends to more than 3,000 clients, including Fortune 500 companies, prominent law firms, universities, government agencies, and industry trade associations. Its experts also have been asked to share their experience and knowledge with regulatory agencies such as the Office of Federal Contract Compliance and the Equal Employment Opportunity Commission.

### *About Micronomics*

Micronomics is an economic research and consulting firm located in Los Angeles, California. Founded in 1988, it is engaged in the application of price theory, analysis of issues relating to resource allocation, and assessment of real-world problems requiring practical and sound solutions. Micronomics focuses on industrial organization, antitrust, intellectual property, the calculation of economic damages, employment issues, and the collection, tabulation, and analysis of economic, financial and statistical data. Clients include law firms, publicly and privately held businesses, and government agencies. In January 2011, Micronomics joined ERS Group.