Applying regression analysis to valuation and forensic engagements

by Dan Werner, Ph.D., CPA

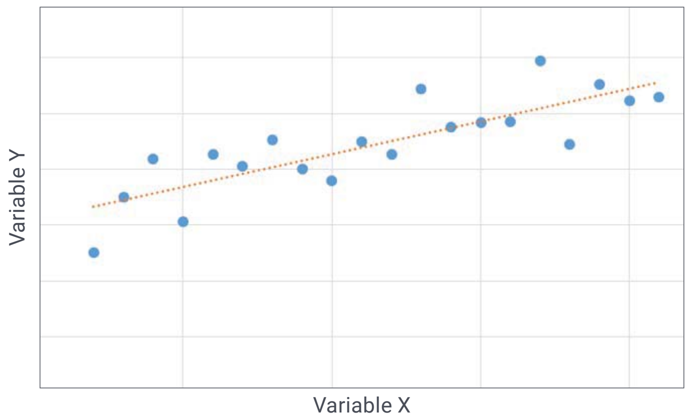
Despite being in an era of abundant data, advanced statistical tools are not always fully used by practitioners in valuation and forensic engagements. Well-established statistical tools, such as regression analysis, can add computational precision, enhance quantitative insights, and uncover new conclusions for valuation and forensic experts. Regression analysis has a variety of applications in forensic and valuation engagements, especially when projecting quantitative scenarios that would have occurred "but for" the alleged conduct.

This article introduces the concept of regression analysis in valuation and forensic engagements, before then discussing how it should be implemented correctly. This is the second of two articles I have written for this publication on regression analysis, with the prior article focusing on the use of regressions in consumer fraud and false advertising cases.[1] These articles are intended to provide a real-world introduction to such techniques in the context of valuation and forensic engagements.[2]

What is a regression?

A regression is a statistical technique used to estimate the relationship between variables. At its simplest, a regression can summarize the relationship between variable Y and variable X, as shown in figure 1 that follows. The blue dots provide a hypothetical scatter plot of data, and there is a clear positive relationship between variable Y and variable X. The orange line provides the best estimate of the relationship between the variables and is calculated using a regression. The general equation for the line in figure 1 is simply $y = mx + b$, where m represents the slope of the line and b represents where it intercepts the vertical y-axis. The slope of the regression line estimates the incremental impact of variable X on variable Y, meaning the amount that variable Y increases for each unit change in variable X.[3]

Figure 1: Hypothetical scatter plot with regression line



In this hypothetical example, consistent with real-world data analysis, the regression line does not fit the data perfectly and is not expected to. Although the line shown in figure 1 generally appears to fit the data well, a regression equation always includes an error term, which indicates the amount by which the actual observation differs from the regression line. Practitioners can use the standard deviation of this error term to quantify the uncertainty surrounding the estimated relationship between variables, in order to test if the estimated relationship is statistically different from having no relationship at all.[4]

There will always be a certain percentage of variation in the data that is unexplained by a regression model. Statisticians and economists regularly use the "coefficient of determination," denoted as R^2 (pronounced "R-squared"), to quantify the amount of variation that is explained by the regression model.[5] A regression model's R^2 can range between 0 and 1, with a higher R^2 generally meaning that the

continued on page 6

[1] Werner, Dan. "Using Advanced Data Analytics to Measure Economic Damages in Consumer Fraud and False Advertising Cases." *FVS Consulting Digest*, Issue 33. American Institute of Certified Public Accountants, January 2018.

[2] These articles are not meant to replace the advanced training and professional experience required to implement regression analysis. Thus, computational discussion is generally omitted from this article because regression analysis is commonly performed in statistical programs such as STATA, SAS, and R (simple linear regressions can also be done in Excel). There are many online resources covering introductory implementation of regression analysis in these programs (for example, see http://data.princeton.edu/stata/ or https://stats.idre.ucla.edu/sas/). For a full, detailed discussion on implementing regression analysis, see Wooldridge, Jeffrey M. Introductory Econometrics: A Modern Approach. Fourth Edition. Mason, OH: South-Western Cengage Learning, 2009; Greene, William H., *Econometric Analysis*, Seventh Edition. Boston: Prentice Hall, 2012.

[3] The equation for the slope of a regression line is $\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$ where $\bar{x}$ and $\bar{y}$ represent the average of the data points for the corresponding x-axis and y-axis.

[4] In certain cases, regression models may falsely report a relationship between variables purely by chance. However, using the output of the regression model, a practitioner can construct a confidence interval and measure the probability that the model's reported relationship is actually different from no relationship at all.

[5] For a detailed discussion of the coefficient of determination (R^2), see Greene, William H. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall, 2012, at pp. 39–46.

regression model better explains the dependent variable. Although what constitutes a "good" R^2 depends on the context of the model and the nature of the data involved, a lower R^2 can indicate that the regression model explains relatively little about the dependent variable. Practitioners should be cognizant of a model's R^2, but the final regression model specification should be driven by economic theory and the facts of the engagement, not by R^2 alone.

Although the previous simple example demonstrates the relationship between one explanatory variable (variable X) and a dependent variable (variable Y), regression analysis can also measure the effect of multiple explanatory variables (also called independent variables) on a dependent variable. In other words, regression analysis can help to explain the effect of one variable on another while "controlling" for other factors (that is, backing out the effect of variables we are not interested in). This article focuses on linear regression, which assumes that the dependent variable is a linear combination of the explanatory variables. Thus, the prior equation ($y = mx + b$) is generalized to $Y = \beta_0 + \beta_1 X_1 + \ldots \beta_n X_n + e$ for n number of explanatory variables, where β_n represents the marginal impact of X_n on Y, and e represents an error term. Regression analysis can also be used to model nonlinear relationships, although such discussion is beyond the scope of this article.[6]

Using regression analysis in valuation

Regression analysis can be used successfully in a variety of valuation engagements, especially in instances in which the market approach is used. In essence, the market approach to valuation estimates an asset's value by comparing it to sales of similar assets. The practitioner can analyze recent sales to establish a metric of value (for example, company valuation as a multiple of revenues), which can then be applied to the target asset in order to calculate its value. Two examples of using regression analysis in valuation follow.

Example 1: Although the market approach to valuation can be straightforward, there are many different financial measures that can theoretically relate to market valuation. For example, a multiple of revenues may be appropriate for a growing technology company with minimal profits, whereas a multiple of earnings before interest, taxes, depreciation and amortization (EBITDA) might be more appropriate for a stable company in a more mature industry. How does the practitioner know the appropriate metric to use? A valuation expert will often use his or her judgment to determine the appropriate metric of value consistent with industry standards, but this expertise can be supplemented using regression analysis.[7] A regression analysis, along with its associated diagnostics (for example, R^2 and tests of statistical significance), can help to determine which metric is the best, most accurate predictor of value.

To implement such analysis, a practitioner would first collect data on sales of comparable companies, for example, the price paid, annual revenues, EDTDA or other metrics. Next, a separate regression for each financial metric could be run and compared, or all metrics could be included as explanatory variables simultaneously in one regression. The appropriate approach depends on the nature of the data (for example, the sample size and the extent of variation in the variables) and the facts of the engagement. The results of such an analysis can indicate which financial metric is the best predictor of value in the industry subsector and, thus, provide a quantitative basis to choose the appropriate metric.

However, this approach is not without practical limitations. For example, when using statistical tests, the practitioner must be attuned to the sample size because there may not be a sufficient number of comparable company sales to properly perform meaningful statistical tests. Although there is not an agreed-upon threshold for the sample size required for regression analysis, including more explanatory variables may necessitate a larger sample size to prevent "Type II" error (that is, falsely concluding that there is no relationship when such relationship exists).[8]

continued on page 7

[6] For more detailed discussion on nonlinear regression models, see Greene, William H. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall, 2012, at pages. 181–218.

[7] For a step-by-step guide on implementing a similar hypothetical example, see Hawkins, George B. "Regression Analysis in Valuation Engagements." *Business Valuation Review* 27, No. 1 (2008): 30–37.

[8] A detailed discussion on hypothesis testing is beyond the scope of this article. For a more detailed discussion, see Greene, William H. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall, 2012, pages. 111–112 and 1062–1065.

Example 2: Regression analysis can also be particularly useful when valuing assets in development stages, when incremental value is added as development continues. Early stage projects and assets in progress can carry significantly more risk (thus, resulting in a lower "risk-adjusted" valuation when compared to operational assets). How does the practitioner know the relationship between the project life cycle and valuation? A 2016 Deloitte report on wind farm asset valuation provides a good example of using regression analysis to answer this question.[9]

First, the Deloitte team collected data on 278 onshore wind farm transactions at various stages in development. After categorizing the transactions into different stages of development, they performed a regression analysis to measure the impact of the project's megawatts (MW) on enterprise value (EV) at each stage of the project's life cycle. As a result, multiples of EV/MW can be estimated for the generation capacity in each stage of the project life cycle. With this information, a valuation practitioner can then provide a valuation of wind farm assets that are not yet fully installed and operational, thereby supplementing a valuation calculated via the discounted cash flow methodology.

Although the discounted cash flow methodology is theoretically feasible in valuing assets that are pre-cash flow, in some instances, there is substantial uncertainty over the appropriate discount rate to apply for early-stage projects in order to properly capture the risk of projected cash flows. A reproduction cost approach is also theoretically feasible for early-stage projects, but it may be difficult to acquire the necessary detailed data on construction costs along with purchase prices for comparable projects. Thus, depending on the facts of the engagement, it may be preferable to use a market approach to valuation while incorporating a well-designed regression analysis.

Using Regression Analysis in Forensic Engagements

Forensic engagements often require a computation of lost profits that involves comparing what actually occurred with what would have occurred in the absence of the alleged misconduct. Regression analysis can be particularly useful when there is a need to project quantitative scenarios that would have occurred "but for" the alleged conduct. Depending on the facts of the case, the regression model may be able to estimate the relationships between financial, economic, or accounting data during a benchmark period (for example, before the alleged misconduct occurred). After these relationships are established, the regression model can then be used to inform, or perhaps directly estimate, but-for scenarios. Two examples of using regression analysis in litigation follow.

Example 1: Consider a simple example in which a practitioner is hired to estimate lost profits due to some alleged misconduct, which allegedly resulted in lower production to the plaintiff firm. In addition to estimating but-for revenues, the practitioner may need to estimate the incremental costs associated with that but-for production using accounting data on total costs and production over time.[10] If the true production process requires fixed costs in addition to variable costs, then simply projecting but-for costs using the average cost per unit may yield incorrect but-for costs. Instead, a regression model can estimate variable costs using monthly accounting data in accordance with the following equation: Total Cost = Fixed Cost + (Variable Cost × Units Produced)

Using the estimated variable cost and fixed cost parameters, the practitioner can then more accurately predict but-for costs consistent with but-for production. Although this simple example considered only one variable cost component, a regression model can include many different explanatory variables, as previously discussed. In addition, although this example focuses on simple linear regression modeling, the approach can be generalized to model nonlinear relationships, as well, allowing the practitioner to account for issues such as economies of scale.

continued on page 8

[9] Deloitte. A Market Approach for Valuing Wind Farm Assets. April 2016.

[10] This example is adapted from Weil, Roman L., Peter B. Frank, Christian W. Hughes, and Michael J. Wagner. "Statistical Estimation of Incremental Cost from Accounting Data." Chapter 5 in *Litigation Services Handbook: The Role of the Financial Expert*. 4th Edition. John Wiley & Sons, 2007, pages. 10–14.

Example 2: Regression analysis can also directly estimate the overcharge to consumers related to alleged misconduct. Consider a hypothetical example of price-fixing allegations in which the practitioner is tasked with measuring the extent to which prices were elevated, if at all, because of the alleged conspiracy. Regression models are commonly deployed in antitrust cases to answer this question.[11] In this case, the dependent variable can be the price paid by consumers, whereas variables meant to capture supply and demand characteristics are included as explanatory variables, including a variable related to the allegations. If properly implemented, the regression model will identify the price impact from the alleged price-fixing and allow the practitioner to statistically test if this effect is greater than zero.

Implementing regression analysis correctly

Although regression analysis is a well-established, defensible statistical tool that has many tangible applications in valuation and forensic engagements, there are common potential pitfalls. In the text that follows, six examples of potential issues are described.

Data quality — At the onset, a practitioner should perform exploratory data analysis to ensure the data quality is sufficiently reliable for regression analysis. Any quantitative model is only as good as the data used to calibrate it. For example, if material measurement error is present in the data, this can create biased regression results and affect quantitative conclusions.[12] It is good practice to review summary statistics and ensure that the data conforms to reasonable expectations based on prior industry research. To the extent that a data point is an error or definitively an outlier, it should be removed prior to implementing a regression model. Similarly, it can be problematic if data for certain variables are systematically missing. Although it is not the focus of this article, it is worth noting that regression analysis can be used to estimate missing data, although there may be superior methods to impute missing data.[13]

Omitted variable bias — Before implementing a regression analysis, it is important to think critically about the explanatory variables that materially affect the dependent variable of interest. What outcome is being explained with the regression model, and what are the salient factors influencing that outcome? If there are multiple factors that considerably affect the dependent variable, omitting one or more important explanatory variables might result in statistically biased outcomes and result in false conclusions. However, a regression model is not expected to control for every miniscule variable that could theoretically affect outcomes. Remember that the regression "model is only a simplification of reality. It will include the salient features of the relationship of interest but will leave unaccounted for influences that might well be present but are regarded as unimportant."[14]

Model sensitivity — "The issue of robustness — whether regression results are sensitive to slight modifications in assumptions (for example, that the data are measured accurately) — is of vital importance."[15] When implementing a regression analysis, it is good practice to perform a sensitivity analysis to ensure the overall conclusions of the model are defensible. This can involve, for example, re-running the model while omitting potential outliers or omitting a small subset of the data to ensure that the model reaches similar conclusions without that group. Similarly, a sensitivity analysis may involve testing other "functional forms" because the relationship between variables is not always linear (for example, an explanatory variable may have an exponential impact on the dependent variable). However, any changes to model specification should be well-grounded in economic theory, the facts of the engagement, and the practitioner's own knowledge, education, expertise, and training. For example, if regression results change due to the unreasonable omission of a variable, then the practitioner's overall conclusions should not change.

Extrapolation — As the saying goes, "past performance is no guarantee of future results." In other words, practitioners should always be wary of making out-of-sample predictions (that is, extrapolation) from regression models. Although a regression model may accurately represent economic relationships in the past, the relationships between variables can change over longer periods of time. For example, a regression model may accurately predict outcomes in the near future, but the results can quickly become speculative if used to predict outcomes decades into the future.

continued on page 9

[11] For example, see Baker, Jonathan B., and Daniel L. Rubinfeld. "Empirical methods in antitrust litigation: Review and critique." *American Law and Economics Review 1*, No. 1/2 (1999): 386–435; Macartney, Gareth and Gordon Rausser. "Economic Principles for Class Certification Analysis in Antitrust Cases." *Antitrust* 30, No. 3 (Spring 2017): 60–67; Finkelstein, Michael O., and Hans Levenbach. "Regression estimates of damages in price-fixing cases." *Law and Contemporary Problems* 46, No. 4 (1983): 145–169.

[12] Rubinfeld, Daniel L. "Reference guide on multiple regression." *Reference Manual on Scientific Evidence* 179 (2000): 425–469, page. 200.

[13] For a detailed discussion of imputing data, see Enders, Craig K. *Applied Missing Data Analysis*. Guilford Press, 2010.

[14] Greene, William H. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall, 2012, page. 5.

[15] Rubinfeld, Daniel L. "Reference guide on multiple regression." *Reference Manual on Scientific Evidence* 179 (2000): 425–469, page. 195.

Correlation vs. causation — Finding a statistically significant correlation between variables does not necessarily mean that one variable causes changes in the other variable. When developing a regression model to explain how one variable causes another, practitioners should think critically about the direction of influence: Is the dependent variable being influenced by the independent variable, or vice versa? Feedback between variables and simultaneous causality can create significant problems in regression models if left unaddressed and may even result in an incorrect conclusion being reached.[16]

Economic vs. statistical significance — When performing regression results, practitioners should always consider regression results in context. There is rightfully a heavy focus on "statistical significance" when performing regression analysis because a regression model includes diagnostics that allow a practitioner to test if the measured effect between variables is statistically different than zero. However, "[t]oo much focus on statistical significance can lead to the false conclusion that a variable is 'important' for explaining [a dependent variable] even though its estimated effect is modest."[17] A variable can have an inconsequential economic effect on another variable but still be statistically significant.

Conclusion

Regression analysis is a flexible statistical tool used to model relationships between variables and has many tangible applications in forensic and valuation engagements. This article provided several introductory examples related to asset valuation and financial estimations, but more advanced regression techniques can be used more broadly. For example, a logit regression can measure how a variable affects the probability of an outcome, whereas an ordered probit regression can be used to predict categorical outcomes that have a regular order (for example, company credit ratings from poor to excellent).[18] Alternatively, the hedonic regression approach can be used to provide a valuation of product features, including the contested feature in labeling cases.[19] Regression analysis is a well-established, defensible statistical tool when properly applied. However, to avoid improper applications that could lead to incorrect conclusions, practitioners should be familiar with issues related to data quality, omitted variable bias, and extrapolation, among others, before implementing such an analysis.

Dan Werner, Ph.D., CPA, is an economist at *OnPoint Analytics*, an economic, financial, and statistical consulting firm specializing in expert testimony in litigation. Dr. Werner has worked on a variety of forensic accounting and economic damage issues across various industries, including, for example, matters related to valuation, lost profits, fraudulent inducement of investments, product labeling, anti-competitive behavior, and class certification, among others.

[16] Simultaneous causality, among other issues, can create a statistical problem known as endogeneity, which can result in biased regression results. This can be overcome using an "instrumental variables" technique that is beyond the scope of this article. For additional discussion on endogeneity and instrumental variable estimation, see Greene, William H. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall, 2012, page. 219–256.

[17] Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach*. Fourth Edition. Mason, OH: South-Western Cengage Learning, 2009, page. 135.

[18] For example, see Greene, William H. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall, 2012, pages. 681–832.

[19] Werner, Dan. "Using Advanced Data Analytics to Measure Economic Damages in Consumer Fraud and False Advertising Cases." *FVS Consulting Digest*, Issue 33. American Institute of Certified Public Accountants, January 2018.