Continued from page 89

predictive value provided by historically high or low dividend yields should be stripped away (for example, the correlation should be zero). All of the comments made regarding dividends are true, although indirectly, for earnings and book value. Since dividends ultimately are paid out of cash flow and earnings, buying stocks with historically high earnings yield has led to historically above-average subsequent total returns.

Correlation is a measure of strength and direction of linear dependence between two variables, and its value may not be sufficient to evaluate that dependence if the assumption of normality of the deviations from linearity is not valid. Therefore, it would be advisable to provide scatter plots of the data to inspect them for linearity, thus attesting the correlation suitableness.

We agree that it would have been preferable to obtain correlations that are closer to 1.00 in value. In fact, a correlation of 0.68 indicates that 46% of the variability in S&P returns is explained by the dividend/earning factors, so that 54% of the variance is unexplained by our model. (The 46% corresponds to the *r*-squared value in a simple linear regression of S&P returns on dividends/earning factors, which is calculated as the square of the correlation.)

While a higher correlation is always desirable, depending on the outcome, it can be difficult to achieve. Given the volatility and difficulty in predicting S&P returns, a model that explains almost half of the variability in these returns may be quite promising. This correlation for a single variable is in line with what other authors have found.

Marco Alves is adding a level of complexity to this analysis that I believe may create a less robust model of market returns than using an empirical observation stripped of any assumptions of either linearity or normal distribution or both.

Two variables can be correlated but in a nonlinear way, for example, age and height. The average six-year-old is taller than the average three-year-old, but this does not mean that the average 60-year-old is taller than the average 30-year-old, or that the ratio between the height of the three-year-old and the six-year-old tells you anything meaningful about the ratio of the height of the 60-year-old to the 30-year-old. In fact, as we age, we may lose height to the forces of osteoporosis, so the average 80-year-old may be shorter than the average 60-year-old. However, it is also incorrect to assume that because this relationship is nonlinear, it is useless. If we construct a table of age versus average height with distributions of height, we can make some determinations of height at some future date without making any assumptions about linearity, covariance, or correlation.

My attempt in the article was to provide a sort of growth chart for the stock market, dividing the data into buckets by quartile then averaging each of those buckets, making no assumptions about the distributions within or between those buckets.

The scatter plot in Figure 1 shows 12-month returns versus the dividend yield at the start of the period. As with any financial series, it is noisy with some high-return periods following low-dividend yields and some low returns following modest or moderately high-dividend yields, but the following is clear:

Continued on page 95