Beyond the Central Tendency: *Quantile Regression as a Tool in Quantitative Investing*

CHRIS GOWLAND, ZHIJIE XIAO, AND QI ZENG

CHRIS GOWLAND
is a vice president and senior quantitative analyst at Delaware Investments in Philadelphia, PA.
cgowlland@delinvest.com

ZHIJIE XIAO
is a professor in the Department of Economics at Boston College in Boston, MA.
xiaoz@bc.edu

QI ZENG
is a senior vice president and portfolio manager at Acadian Asset Management in Boston, MA.
qzeng@acadian-asset.com

Many quantitative investors make extensive use of regression-based analysis in analyzing factor performance, assessing the relative attractiveness of different firms, and monitoring the risks in their portfolios. Ordinary least squares (OLS) regression is highly effective for understanding the central tendencies within a dataset. But OLS analysis will generally be less useful for understanding the behavior of datapoints that are close to the upper or lower extremes within a population, even though such datapoints are often the most interesting from the perspective of an active investor or a risk manager.

To investigate the behavior of datapoints that are distant from the central tendency, a more appropriate tool is quantile regression. Quantile regression allows researchers to describe more precisely the behavior of points close to the tails of a distribution. Another advantage of quantile regression is its inherent robustness to heteroskedasticity, skewness, and leptokurtosis. Many financial datasets involve one or more of these statistical challenges; quantile regression may therefore be broadly useful to researchers working with such data.

In this article, we show how quantile regression can be interpreted as an extension of the familiar OLS approach to analyzing data. We also provide examples of how quantile regression can be used to analyze the effectiveness of individual factors for quantitative investment purposes. Our examples are all based on a universe of U.S. small-cap stocks, an asset class for which insight into heteroskedastic factor effectiveness may have considerable significance for quantitatively based investment strategies.

QUANTILE REGRESSION

Since the late 18th century, the most common form of regression in statistical analysis has been to find the "line of best fit"—in other words, the central tendency or conditional mean of a dataset—by minimizing the sum of squared errors. Such an approach is useful when the topic of interest is indeed the mean of the distribution, but in other contexts a regression based on least squares may not be the optimal tool. In the 1880s, Galton and Edgeworth both made use of alternative approaches to give a more nuanced understanding of particular datasets, and several other statisticians have also expressed concern about the extent to which conventional statistical analysis favors the use of least-squares methodologies (Koenker [2000]).

Many empirical researchers, in finance and other disciplines, make extensive use of the ordinary least squares approach, which is sufficiently widely accepted to be included in spreadsheet programs. Statistical theory shows that an OLS approach is optimal for finding the central tendency when the data are independent and identically distributed (i.i.d.) with errors that can be described as normal, or Gaussian. If the errors are non-normal, then OLS methods will still give results that are asymptotically valid, but will generally be less efficient than robust methods. In empirical analysis, financial time series tend to be heavy-tailed (or leptokurtic) and these features are usually accentuated when the data are sampled more frequently. Monte Carlo evidence has indicated that OLS estimators tend to be very sensitive to certain types of outliers and, consequently, that inference based on OLS estimation may have poor performance. Moreover, as pointed out previously, if a researcher is interested in the behavior of datapoints that are distant from the central tendency within the dataset, then OLS may not provide much enlightenment.

The modern form of quantile regression was introduced by Koenker and Bassett [1978] and has gradually gained popularity as a robust alternative to OLS methods in economics and finance. Perhaps the easiest way to understand quantile regression is to compare it with the familiar least-squares regression. One of the several possible formulations of the least-squares methodology is the following expression:

$$\min_{\beta \in R_p} \sum_{i=1}^{n} (y_i - \mu(x_i, \beta))^2$$

This expression shows that OLS analysis seeks to minimize the sum of squared differences between each of the n independent variables, y_i, and the expected value $\mu(x_i, \beta)$, which for linear regression will be $\beta_0 + \beta_1 x_i$.[1] Thus, OLS regression will produce an estimate of the conditional mean of the dependent variable,

$$\hat{E}(y_i | x_i) = \mu(x_i, \hat{\beta}) \tag{1}$$

such that the expected value of y_1 given x_1 will be determined on the basis of the central tendency μ and the (estimate of the) parameter β.

The basis of quantile regression is conceptually similar, with three important differences. First, OLS regression focuses on the central tendency μ, whereas quantile regression is concerned with the τth conditional quantile.[2] Second, OLS regression is principally concerned with deviations from the mean μ, whereas quantile regression is focused on the central tendency for the τth conditional quantile, which is conventionally notated as ξ. And finally, OLS regression aims to minimize squared deviations from the mean μ, whereas quantile regression seeks to minimize absolute deviations from the value ξ. Hence, quantile regression can be expressed as

$$\min_{\beta \in R_p} \sum_{i=1}^{n} \rho_\tau (y_i - \xi(x_i, \beta)) \tag{2}$$

where $\rho_\tau = u(\tau - I(u < 0))$ represents the check function as defined in Koenker and Bassett [1978]. A simple linear model illustrates the basic features of quantile regression. Define a $k \times 1$ vector of regressors, x_t, containing information that may have the power to predict future returns, and define $r_{(t+1)}$ as the return over the next period. The relevant linear quantile regression model will then be

$$\hat{\beta}(\tau) = \arg \min \rho_\tau (r_{t+1} - \beta' x_t) \tag{3}$$

In other words, finding the value of $\hat{\beta}(\tau)$ that will minimize the weighted sum of absolute deviations between $r_{(t+1)}$ and $\beta' x_t$, will provide an estimate for the τth quantile of the return $r_{(t+1)}$, conditional on the regressor x_t.[3] The conditional quantile function is the inverse of the conditional distribution function, and therefore quantile regression provides an estimate for the whole (conditional) distribution, instead of focusing solely on the mean of $r_{(t+1)}$ as OLS methods would do. This represents a relatively simple implementation of quantile regression; any practical application will require the estimation of a specific model.

Econometricians and statisticians have extended the basic framework of quantile regression in many different ways to enable quantile-based methods to be applied to a wide range of real-world problems.[4] The regression quantiles can be estimated using standard linear programming techniques.[5] The statistical packages R, S-Plus, Stata, and SAS all have quantile regression capabilities, either as part of their base distribution or as separate modules. These modules typically focus on linear quantile regression, but extensions to nonlinear applications are also feasible (Koenker and Hallock [2001]).

As mentioned previously, OLS and similar methods can be rather sensitive to the presence of non-normal errors in a dataset because they are based on mean-variance relationships, which can be strongly influenced by outliers. By contrast, quantile regression methods are generally quite robust to distributional assumptions because they are implicitly based on sample quantiles that are conceptually more similar to medians and median absolute deviations. These characteristics ensure that quantile regression is much less sensitive than OLS methods to outliers, extreme values, heteroskedastic errors, censored data, and/or truncated data. By comparison with OLS, the estimation of quantile-level parameters is influenced principally by the local behavior of the conditional distribution of the response near the specified quantile.

Consequently, researchers have begun using quantile regression in many contexts in medicine and economics (Yu, Lu, and Stander [2003]). Quantile regression methods are also being used in the field of ecology, where environmental heterogeneity often means that focusing solely on central tendencies can provide a misleading picture of the constraints facing different organisms (Cade and Noon [2003] and Vaz et al. [2008]). Quantile regression methods are now widely used in labor economics to analyze returns to education, effects of unionization, and labor market discrimination (Koenker and Hallock [2001]), and are also being used in development economics (Stasavage [2002] and Nguyen et al. [2007]), industrial organization (Görg and Strobl [2002], Reichstein et al. [2006], and Coad and Rao [2008]), and monetary economics (Lee and Yang [2007]).

Quantile regression has also attracted increasing attention in the discipline of finance. Taylor [1999] applied a quantile regression approach to estimating the distribution of multiperiod returns. Basset and Chen [2001] used quantile methods to analyze the investment style of institutional investors over time. Several authors have used these techniques in the estimation of value at risk (VaR) for financial institutions (Chernozhukov and Umantsev [2001], Wu and Xiao [2002], and Engle and Manganelli [2004]). Quantile approaches have also been used to analyze the motivations behind open market share repurchases (Billett and Xue [2008]).

Researchers have also begun applying quantile regression to time-series problems in the realm of finance. One recent article used quantile methods to analyze how Korean firms' capital structures have changed over time (Fattouh, Scaramozzino, and Harris [2005]). Quantile-based methods not only provide a more complete picture of the conditional relationship among financial variables, but can also test whether the quantile-specific parameters are stable over different quantiles and over time (Koenker and Xiao [2006]). As a consequence, quantile models can show how different variables affect the location, scale, and shape of the conditional distribution of the response. Such methods therefore constitute a significant extension of classical constant coefficient time-series models, in which the effect of conditioning is typically confined to a shift of the intercept and/or the slope of the central tendency.

USING QUANTILE REGRESSION TO ANALYZE FACTOR EFFECTIVENESS

Analysts and portfolio managers who make use of quantitative methods typically focus much of their attention on assessing the effectiveness of particular factors for investment purposes. In this context, a factor refers to some characteristic that can be measured across the constituents of a particular universe, such as the book-to-market ratio, total return over the previous six months, or measured sensitivity to changes in inflation (Chan, Karceski, and Lakonishok [1998]).

Most quantitative analysts backtest individual factors on a univariate basis to determine if investing on the basis of that factor would have been associated with positive or negative returns in the past. If a particular factor would have been effective at identifying future outperformers and/or future underperformers, then it may have value as part of a broader multifactor model, which can be used to rank individual stocks based on their overall attractiveness in terms of projected returns.

Two methods are commonly used for assessing factor effectiveness, namely regression-based analysis and synthetic factor-mimicking portfolios. Regardless of the method selected, the intent is to determine whether investing on the basis of a specific factor, or some particular combination of factors, would have been associated in the past with positive (or negative) returns. Regression-based analysis is widely used by academic researchers and by many practitioners, in part because it can readily be extended to a multivariate context. Many practitioners, and quite a few academics, also make use of factor-mimicking portfolios because they impose less structure on the data and thus offer greater robustness in the presence of outliers and similar problems. Most users of factor-mimicking portfolios, however, tend to focus on factor effectiveness at a univariate level because these techniques are not easily extensible to a multivariate context (Fama and French [2007]).

Quantile regression can provide a more complete picture of factor effectiveness because it is more robust than conventional OLS methods, while also being more easily adaptable to multivariate approaches than are factor-mimicking portfolios. As a concrete example, we consider the universe of listed U.S. small-cap stocks, which we define as those stocks ranked from 1,001 to 3,000 in terms of market cap at the end of a given month.[6] We analyze two commonly used factors: a value factor in the form of the book-to-price ratio, and a medium-term momentum factor, which we define as the total return over the prior 12 months, excluding the most recent month. After analyzing these two factors separately, we then consider them together. All data is taken from Compustat.

In this study, we examined the use of quantile regression on a universe-wide basis in order to mimic the actions of small-cap managers who are sector agnostic. It would also be possible to use a quantile approach to reflect the perspective of a manager who employs a sector-neutral strategy.

QUANTILE ANALYSIS OF BOOK-TO-PRICE

Numerous academic researchers have found that, over the long run, the book-to-price (also known as book-to-market) ratio has been positively associated with returns. In other words, stocks that are relatively cheap in terms of book-to-price have typically produced better returns over the subsequent period than their counterparts, which are more expensive in terms of this measure. This relationship has been widely accepted, at least since the work of Fama and French [1992]; subsequent work by Fama and French and by other researchers has confirmed that the book-to-price ratio appears to be quite effective across a wide range of markets and time periods. Indeed, one of the principal analytical puzzles in contemporary finance is to explain the existence of a value premium, which is often operationalized in terms of book-to-price.[7]

First, we consider the stocks in our U.S. small-cap universe at the end of April 2006. For each of these stocks, we calculated the book-to-price multiple. During the month of April 2006, the book-to-price ratios for our universe were mainly between 0.2 and 1, although many companies were above and below this range.[8] We then found the subsequent one-month total return for each of these stocks. The one-month total return is defined as change in price plus any dividend paid during May 2006. For this dataset, the one-month returns are generally between −20% and +20%, although again there are numerous outliers above and below this range. An OLS analysis of this dataset shows that the overall slope of the relationship was 3.38, indicating that stocks that were cheaper in terms of book-to-price at the end of April 2006 generally delivered higher returns during the subsequent month of May 2006.

As discussed earlier, OLS methods effectively focus on the central tendency within the data, but practitioners, such as risk managers, quantitative analysts, or portfolio managers, are often also interested in factor relationships toward the tails of the distribution. And where a dataset appears to suffer from outliers and/or heteroskedasticity, the estimated coefficients from OLS analysis may not be very useful guides to relationships that are relatively distant from the central tendency. For this type of analysis, quantile regression methods can be quite useful.

Exhibit 1 shows a scatter plot of the book-to-price multiples at the end of April 2006 and the percentage of total returns during the subsequent month of May 2006, for the U.S. small-cap universe. The scatter plot suggests heteroskedasticity in the dataset, given that the dispersion of results seems to be somewhat larger in the middle of the distribution. The estimated lines for the 10th, 50th, and 90th quantiles, also shown in the exhibit, indicate that for stocks that are relatively cheap in terms of this factor—in other words, stocks to the right of the distribution—the returns to the 10th and 90th quantiles are not very different. But among the "expensive" stocks, there is a very wide spectrum of returns between the 10th and 90th quantiles. In other words, the heteroskedasticity of the dataset for this month is concentrated among the names on the left of the distribution.

Exhibit 2 reports the estimated intercept and estimated slope for the nine quantiles—the 10th to the 90th.[9] The tabulated results suggest that the relationship between the book-to-price multiple and the subsequent one-month return is considerably stronger in the lower quantiles than in the higher ones. For the stocks whose returns are in the 80th and 90th quantiles, the relationship between valuation in terms of book-to-price and subsequent returns is quite feeble. By contrast, for stocks whose returns are in the 50th quantile and below, the estimated slope is considerably steeper.[10]

We conclude from these results that this factor would have provided quite useful signals for picking future outperformers, because stocks that are cheap in terms of book-to-price typically delivered strong returns over the next period. We would also infer, however, that the factor appears less reliable for choosing future underperformers,

EXHIBIT 1
Scatterplot of Book-to-Price Ratios in April 2006 vs. Return in May 2006

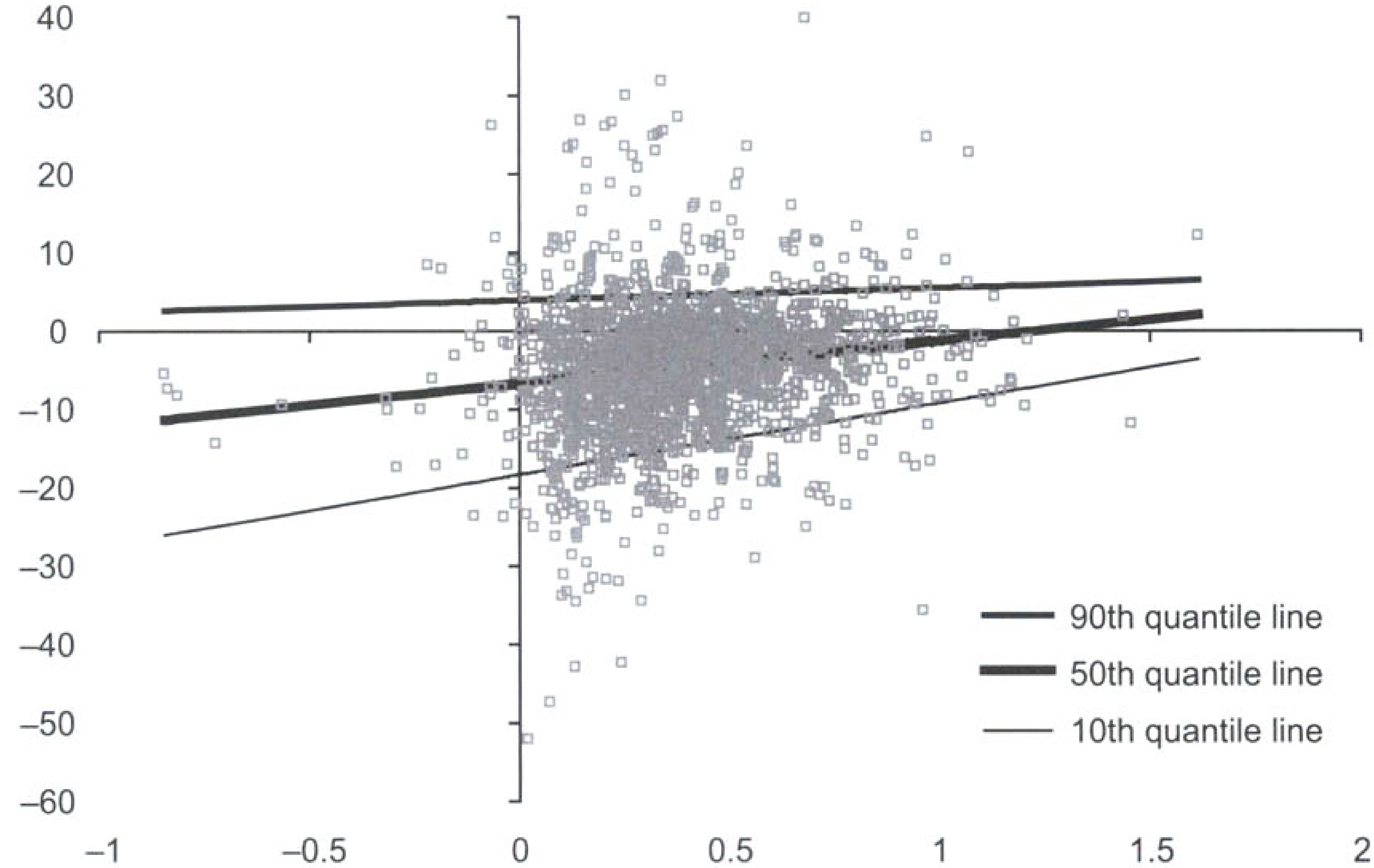


EXHIBIT 2
Quantile Regression Results: Book-to-Price Ratios in April 2006 vs. Return in May 2006

Quantile	Estimated Intercept	Standard Error	*t*-Statistic	Estimated Slope	Standard Error	*t*-Statistic
10th	−18.30	0.81	−22.54	9.12	1.17	7.80
20th	−13.64	0.46	−29.97	6.36	0.87	7.27
30th	−11.08	0.57	−19.33	6.83	1.20	5.69
40th	−8.56	0.46	−18.43	5.82	0.97	5.99
50th	−6.70	0.48	−13.85	5.46	0.98	5.55
60th	−4.51	0.48	−9.45	4.46	0.87	5.15
70th	−2.09	0.46	−4.55	2.78	0.80	3.49
80th	0.60	0.46	1.30	1.22	0.96	1.27
90th	3.93	1.14	3.34	1.57	2.74	0.57

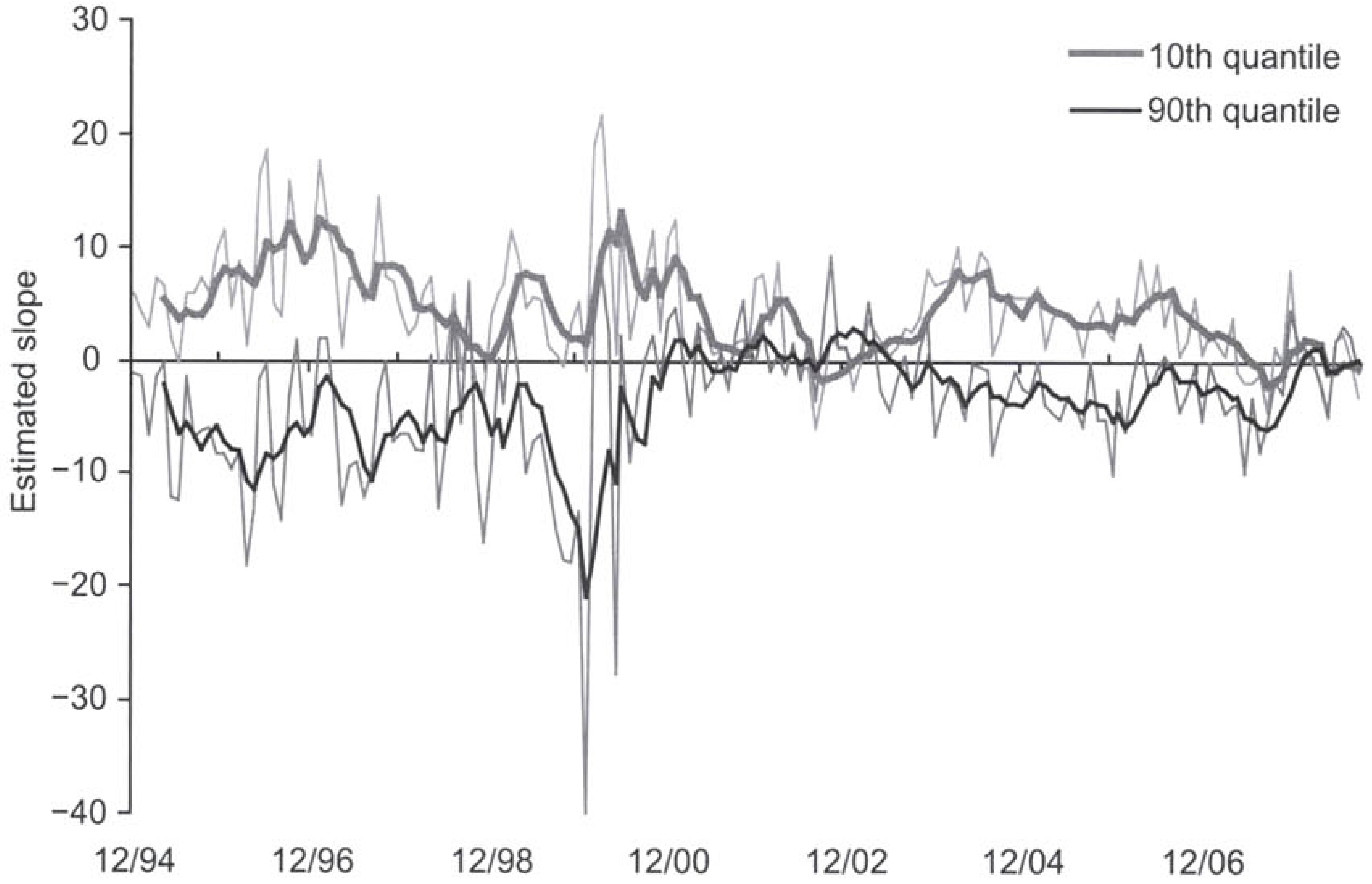
because there is a much wider range of returns for stocks that were expensive in terms of book-to-price.[11]

The same analysis can be repeated for successive months over a longer period, for instance, from December 1994 onward. The results are rather volatile, but can be smoothed by taking a five-month moving median. Exhibit 3 shows that quite consistently, the 10th quantile has been steeper than the 90th quantile. Indeed, for most of the period under analysis, the slope of the 90th quantile was negative, which suggests that the pattern seen in Exhibits 1 and 2 was not a fluke.

However, the slopes themselves are not very easily interpreted. One way to convert them into information directly useful to practitioners is to focus instead on the gap between the 10th and 90th quantiles for stocks with different characteristics. In the case of book-to-price, for

Exhibit 3

Raw and Smoothed Slopes from Quantile Regressions of Book-to-Price Ratios vs. Next Month Returns



instance, we might focus on the gap between the estimated 10th and 90th quantiles for companies that are considered expensive at a book-to-price of 0.1 versus those that are considered cheap at a book-to-price of 1.[12] We can then calculate the spread between the 10th and 90th quantiles for hypothetical companies trading at these two multiples, in order to better understand the dispersion of returns among stocks that look cheap or expensive in terms of this metric.

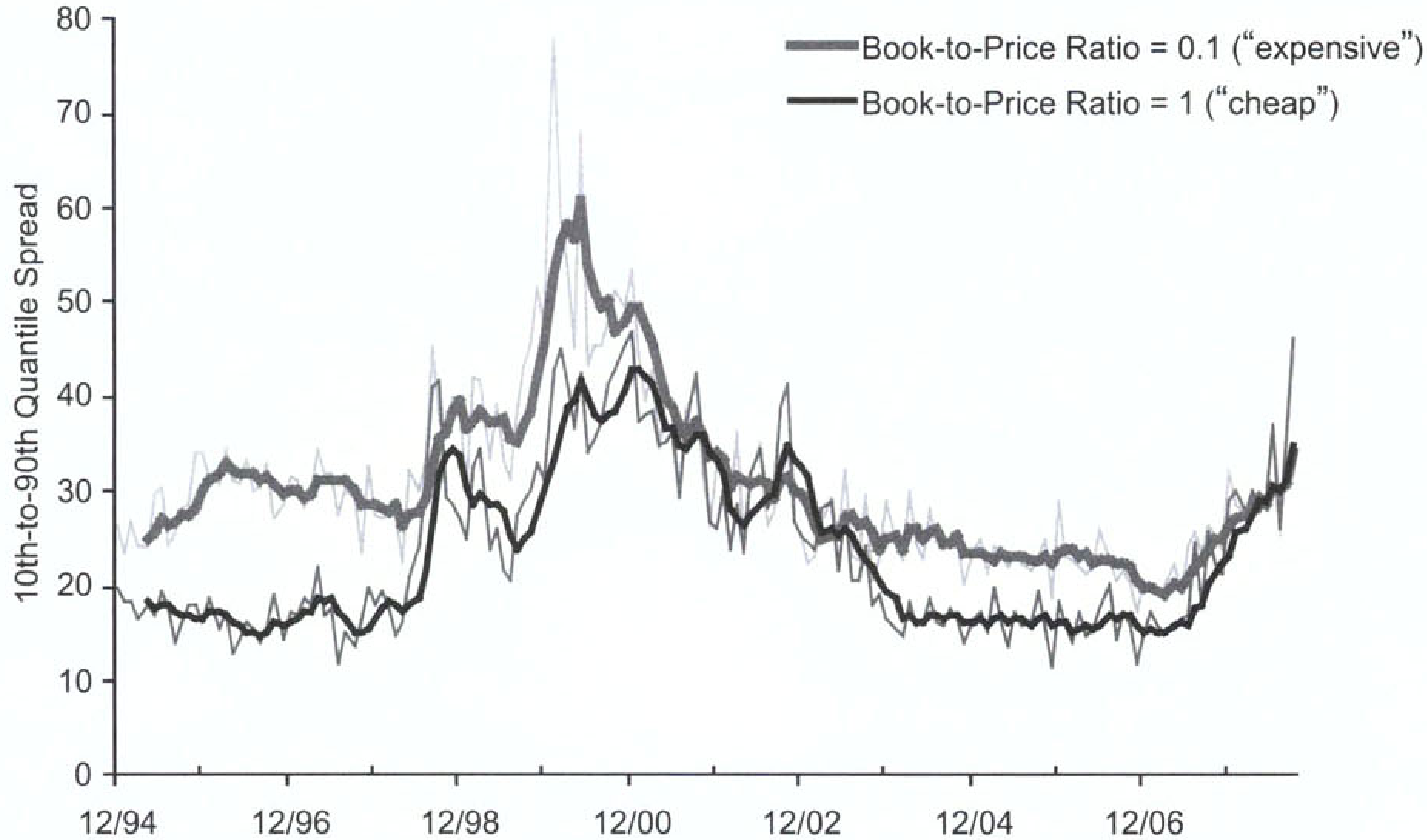
Exhibit 4 shows that the dispersion of returns between the 10th and 90th quantiles has not been constant, among either the stocks that look expensive in terms of book-to-price or those that look cheap by this criterion. In particular, the dispersion of returns for both categories rose considerably during the period from 1998 to 2002. Once again, we use a trailing five-month median in order to smooth the volatility of the month-by-month estimates.

Exhibit 4 also shows that the gap between the best and worst performers has generally been narrower for cheap stocks than for expensive ones. Indeed, over the past 14 years, the only two periods when the 10th to 90th quantile spread for these two categories was fairly close occurred from 2001 to 2003, and since the middle of 2007. During most other periods, the spread in predicted returns between the best and worst performers—among those stocks that are cheap in terms of book-to-price—has been considerably lower than the expected gap between best and worst performers among firms that appear expensive in terms of this metric.

The dispersion of expected results between the 10th and 90th quantiles can be interpreted as an indication of the likely spread between the lowest and highest quantiles for stocks falling into a particular category. For book-to-price, we found a relatively narrow spread of returns for stocks that are cheap in terms of this metric, suggesting that such firms are likely to deliver rather similar returns. By contrast, we observed a relatively wide spread of returns for stocks that are comparatively expensive by this criterion, indicating the likelihood of a wider gap between the best and worst performers.

These results are important for asset managers who are building portfolios on the basis of book-to-price, or on factors that have a similar profile when analyzed with quantile regression. All else equal, the spread of results between best and worst is likely to be greater among stocks that look expensive in terms of book-to-price. Consequently, if an investor is building a long/short portfolio

EXHIBIT 4
Estimated 10th to 90th Quantile Spread for High and Low Book-to-Price Ratios



on the basis of book-to-price, the short side will probably need to have more names than the long side.[13]

QUANTILE ANALYSIS OF MEDIUM-TERM MOMENTUM

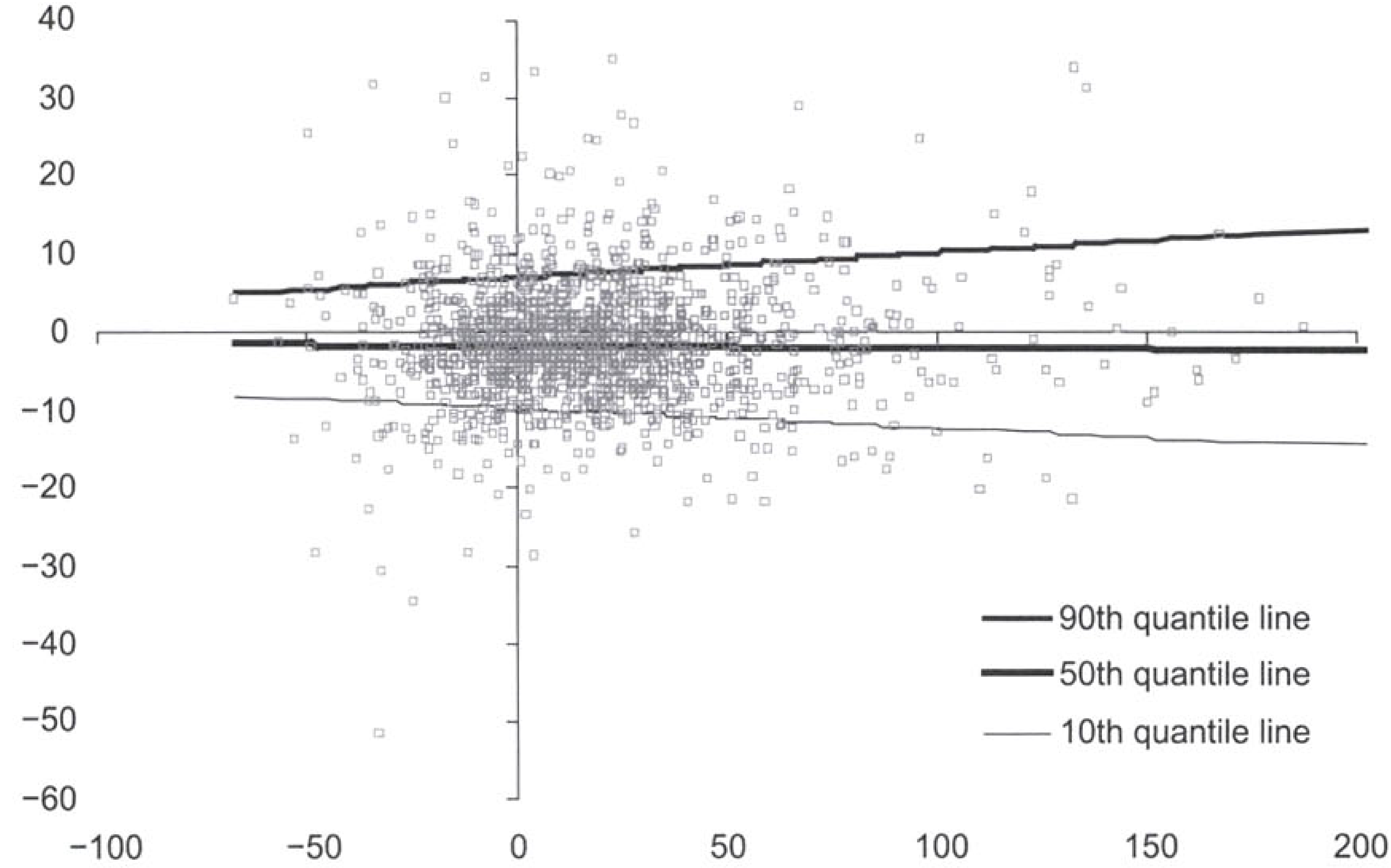
A topic of active research interest for both academics and practitioners is momentum, which can loosely be defined as past performance in terms of either prices or total returns. Investigation has generally found that short-term momentum reverses, medium-term momentum persists, and long-term momentum reverses. More concretely, this suggests that stocks that have delivered high returns over the past month tend to underperform over the next month; conversely, stocks that have performed well over the past year will generally continue to do well over the next year; and finally, stocks that have earned better returns than their peers over the last five years will generally struggle over the next five years. Most academic research has focused on specific lookback periods, although a few articles have compared performance across several different periods (Chan, Karceski, and Lakonishok [1998]).

With regard to medium-term momentum, the work of Jegadeesh and Titman [1993] has been very widely cited. A widely used base case in financial econometrics is the four-factor model (Carhart [1997]), which combines the three factors used in Fama and French [1992] with the medium-term momentum factor described by Jegadeesh and Titman. Most of these studies used a holding period of one year, but quant practitioners have often compressed the holding period to one month and still found that medium-term momentum tends to persist.[14]

Once again, we focus on a universe of U.S. small-cap stocks, namely those from 1,001 to 3,000 in terms of market cap as of a given month. For each of the stocks in this universe, Exhibit 5 shows a scatter plot of the trailing-month (12- minus 1-month) momentum at the end of May 2007 and the return during the subsequent month of June 2007. As in the case of book-to-price, the dispersion of results seems to be somewhat larger in the middle of the distribution. But unlike the case of book-to-price, the gap between the 10th and 90th quantiles is higher on the right side of the graph; in other words, among those stocks whose prior performance had been the strongest. The analysis suggests that for stocks that delivered relatively low total returns over the prior year, the dispersion of returns is relatively low, but among stocks that generated the highest total returns over the

EXHIBIT 5

Scatterplot of Medium-Term Momentum in May 2007 vs. Return in June 2007



prior year, the gap between the 10th and the 90th quantiles is wider in terms of subsequent performance.

We conclude from these results that, over this single month, investing on the basis of medium-term momentum would have provided quite useful signals for picking future underperformers because stocks that performed poorly in the prior year typically delivered rather weak returns over the subsequent month. However, the factor appears to have been less reliable for choosing future outperformers, in that the range of returns is much wider among stocks which looked most attractive in terms of medium-term momentum. Therefore, if a stock generated weak returns over the past year, it is also likely to deliver weak returns over the next month; but if a stock produced strong returns over the last year, its performance over the next month might be anything from very positive to very negative.

The same analysis can be repeated for successive months over a longer period, as we did in the case of book-to-price. Once again, the estimated slopes for each month are rather volatile, but can be smoothed using a five-month moving median. Exhibit 6 shows that since 1995, the slope of the 90th quantile has generally been steeper than the slope of the 10th quantile, implying that the dispersion of returns has been wider among the stocks that have shown the strongest returns (i.e., stocks similar to those on the right side of Exhibit 5).

These estimated monthly values of the intercept and slope were then used to generate an estimated return for companies falling into either the 10th or 90th quantile, depending on company characteristics. For convenience, we considered that stocks with a trailing total return of −50% would be considered relatively unappealing in terms of medium-term momentum, whereas those with a trailing total return of +100% would be deemed quite attractive. Using these two values, we then calculated the spread between the 10th and 90th quantiles for hypothetical companies with these characteristics, just as we did in the earlier case of book-to-price.

Exhibit 7 shows that the dispersion between the returns to the most and least attractive companies, in terms of momentum, has varied quite widely over the past 14 years. As in Exhibit 4, the dispersion of returns was considerably wider between 1998 and 2002 than either before or after. But contrary to the pattern illustrated in Exhibit 4, the gap between the 10th and 90th quantiles has typically been wider for companies considered to be attractive, namely, those which have delivered relatively strong

Exhibit 6

Raw and Smoothed Slopes from Quantile Regressions of Medium-Term Momentum vs. Next Month Returns

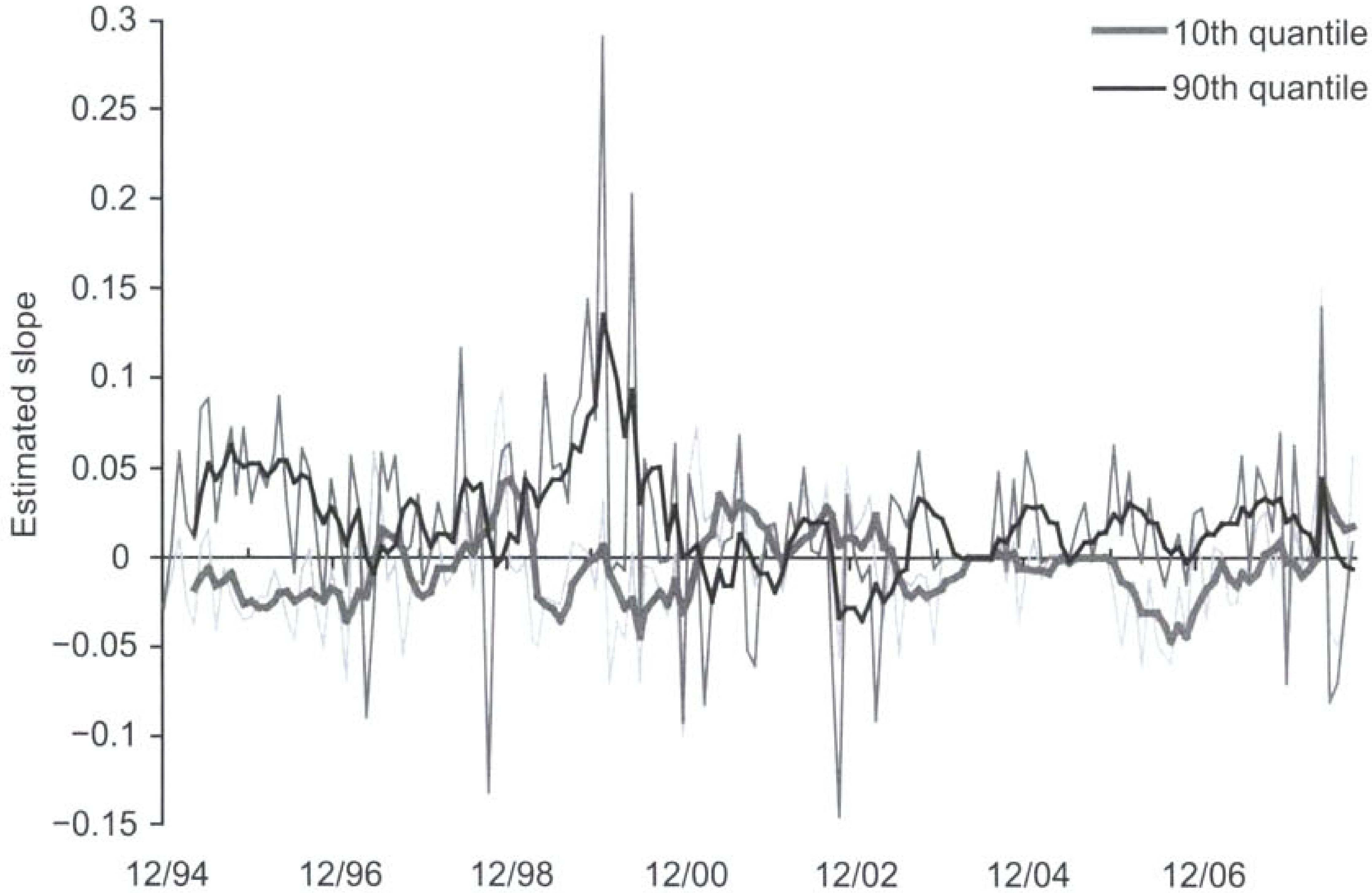
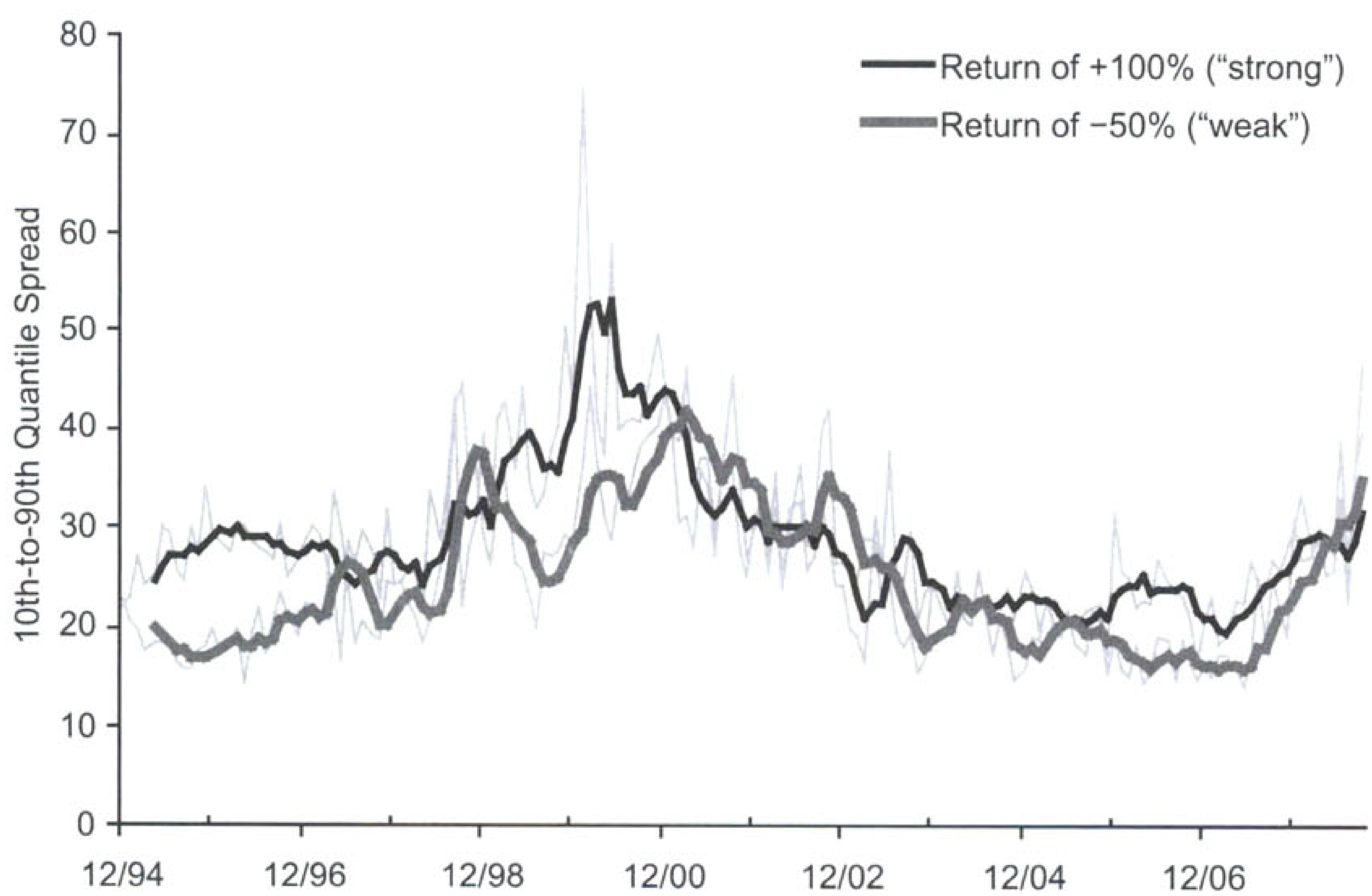


Exhibit 7

Estimated 10th to 90th Quantile Spread for Strong and Weak Medium-Term Momentum



performance over the past year. Admittedly, between 2001 and 2003, the dispersion of expected returns between the 10th and 90th quantiles for these two categories was quite similar, but at other times the gap between the "strong" and "weak" firms was generally fairly wide.

These results suggest that the gap between the best and worst performers among stocks that have delivered poor returns over the preceding year is relatively tight, in comparison to the gap between best and worst performers among stocks that have delivered strong returns over the same period.

The quantile regression analysis of medium-term momentum thus shows a pattern quite different from the one we found for book-to-price. Our results suggest that there is a wider gap between the 10th and 90th quantiles among those stocks that look most attractive in terms of momentum, in other words, those stocks that delivered strong returns over the previous year (excluding the most recent month).

The implications are thus the mirror image of what we proposed for book-to-price—all else equal, analysts and portfolio managers who rely heavily on medium-term momentum factors will probably want to have a wider range of names on the long side of the portfolio than on the short side of the portfolio. More generally, these results suggest that if a stock delivered weak returns over the last year (excluding the most recent month), the stock is rather unlikely to deliver very strong returns over the next month. Such a conclusion is consistent with research suggesting that medium-term momentum tends to persist, as found by Jegadeesh and Titman [1993, 2001].

QUANTILE MULTIVARIATE ANALYSIS OF VALUE AND MOMENTUM

As previously noted, quantile regression can readily be extended into multivariate analysis, just as conventional OLS methods can be used in a multivariate context. Moreover, quantile regression maintains the same attractive characteristics as in the univariate case, such that it is able to describe behavior away from the central tendency while not being excessively influenced by outliers. As a proof of concept, we used the two factors, whose univariate behavior we analyzed previously, as two independent variables to explain the subsequent month's returns.

Once again, it is convenient to show the results in terms of the spread between the 10th and 90th quantiles for companies with specific characteristics. Because two variables were considered, we needed to choose particular values for both. For consistency with the univariate results presented earlier, we used a combination of the same values; thus, a relatively attractive stock would have a book-to-price multiple of 1 with a trailing-month (12- minus 1-month) total return of 100%, whereas a relatively unattractive stock would have a book-to-price multiple of 0.1 with a trailing return of −50%.[15]

Exhibit 8 shows that over most of the past 14 years, the dispersion of returns was somewhat wider among the stocks that are attractive according to our hypothesis. Again, we observe the widening of dispersion between 1998 and 2002, and observe that from 2001 to 2003 the gap between the 10th to 90th quantile dispersion for the attractive and unattractive stocks narrowed considerably. But Exhibit 8 also shows that since the beginning of 2006, the gap between the dispersion of returns to the most and least attractive stocks has diminished. Moreover, since mid-2007 there has been a steady rise in the dispersion of returns between the "most attractive" and "least attractive" names, and as of late 2008 the 10th to 90th quantile spreads were roughly equal, although these dispersions were still somewhat below the levels observed during the late 1990s.

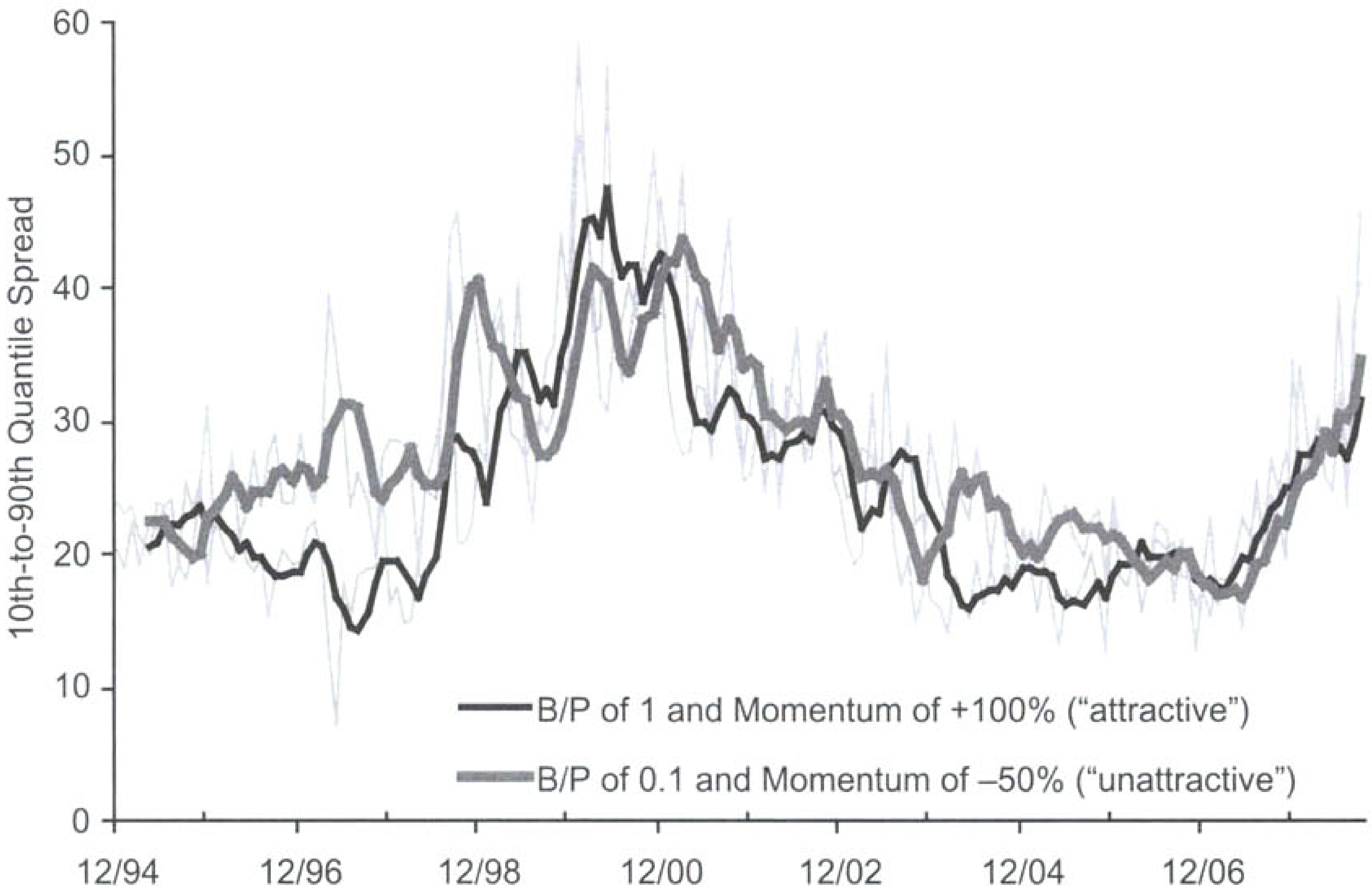
CONCLUSION

In this article, we have shown how quantile regression can be viewed as an extension of familiar methods, such as OLS regression. We have also presented the results of our analysis regarding the univariate effectiveness of two different factors and the effectiveness of a combination of those two factors for a universe of U.S. small-cap stocks. In this conclusion, we discuss the practical implications of our findings and possible extensions of this work.

Many analysts and portfolio managers focus on small-cap growth stocks, which typically look expensive in terms of book-to-price. The results presented in this article suggest that it is possible for a long-only portfolio of such stocks to deliver strong performance, provided that the stocks that are likely to end up in the higher quantiles can be identified. But investors and risk managers with responsibility for such portfolios should also be aware of the possibility for very poor returns among stocks that are expensive in terms of book-to-price. Essentially, an investor who is building a portfolio of such names should have strong convictions about each selection, and may also need to monitor the individual names more actively than might be true for a portfolio of small-cap value stocks.[16]

EXHIBIT 8

Estimated 10th to 90th Quantile Spread for Combination of Value and Momentum Factors



In the multivariate analysis that combined these two factors, and for the particular parameter values that we chose for expository purposes, we found that the two different patterns of dispersion appear to roughly offset each other. The result is that since the start of 2006 there has not been any clear disparity in the dispersion of results for the attractive versus the unattractive names. The levels of dispersion from early 2006 to mid-2007, however, were unusually low by historical standards, and the rise in dispersion over the ensuing period is quite striking. In general, if analysts and portfolio managers believe that the volatility of factor performance is likely to rise in the future, then it will typically be prudent to begin increasing the number of names in a portfolio in order to reduce concentration risk.[17] And based on the results shown in Exhibit 8, it would make sense to increase the number of names on the long and short sides of a portfolio by roughly equal amounts.

This article has deliberately adopted a relatively simple approach. There are numerous extensions and refinements that could readily be applied to the basic techniques discussed in the article. For instance, the analysis could be replicated at the level of individual sectors to reflect the behavior of sector-neutral investors who compare stocks with their immediate peers, rather than adopting a sector-agnostic approach that considers all stocks on a universe-wide basis.[18]

As a longer-term project, it would be feasible to construct artificial portfolios of different sizes in order to gain empirical estimates of the optimal number of names on the long and short sides under different assumptions about the future path of 10th to 90th quantile dispersion. And finally, all analysis presented in this article was based on an assumption of linearity, but it would also be possible to use the nonlinear methods discussed in the article as a means to determine if they might provide additional insights. We intend to investigate these issues in subsequent research.

ENDNOTES

Mr. Gowlland and Dr. Zeng wish to clarify that they wrote this article in their personal capacity, and that the opinions and methods outlined in the article should not be construed as representing the opinions and methods of Delaware Investments or of Acadian Asset Management, respectively.

[1]The form $\beta_0 + \beta_1 x_i$ involves the use of an intercept β_0 and a slope coefficient $\beta_1 x_i$. It is worth noting that x_i could be multivariate, and it is also worth noting that the formulation

of $\mu(x_i, \beta)$ is not restricted to linear estimation: an alternative, such as $\beta_0 + \beta_1 x_i + \beta_2 x_i^2$ or $\beta_0 + \beta_1 x_i^{-1}$, could also be fitted into the same conceptual framework.

[2]For example, setting $\tau = 0.5$ will generate the median, whereas setting $\tau = 0.1$ will find the appropriate values for the 10th quantile.

[3]The researcher is accepting that $\hat{\beta}(\tau) \rightarrow \beta(\tau)$, in probability, as $n \rightarrow \infty$. In other words, as the size of the dataset increases, it should be possible to obtain more and more precise estimates for $\beta(\tau)$.

[4]In addition, there are also quantile analogues to the "local estimate of slope" (or "loess") methods that are often used in exploratory data analysis and similar applications. Koenker [2005] offered an excellent overview of much of this work.

[5]For very large datasets, a simplex method is not always the most efficient method. Koenker and d'Orey [1987] took a modified simplex algorithm first proposed by Barrodale and Roberts [1974], and adapted it for quantile regression problems. Portnoy and Koenker [1997] described an approach that combines some statistical preprocessing with interior point methods, and which operates faster than the simplex method for very large problems.

[6]The universe created in this manner will therefore resemble the Russell 2000 small-cap index, although that index is subject to extensive rebalancing on an annual, rather than a monthly, basis.

[7]Barberis and Thaler [2003] provided a useful overview of the literature in this area, emphasizing the various explanations that have been offered by proponents of behavioral finance approaches.

[8]A book-to-price ratio of 0.2 would be the equivalent of a price-to-book ratio of 5. The book-to-price ratio is generally used in preference to the price-to-book ratio because it can readily deal with cases in which the book value is negative. As shown in Exhibit 1, our dataset includes quite a few such companies as of April 2006. We spot-checked some of these cases and found that their most recent book value as of that date was indeed negative. Such companies might technically be regarded as insolvent, because their total liabilities exceed their total assets, but they can continue trading if they have supportive lenders and auditors.

[9]Any quantile can be analyzed, but the 10th and 90th are particularly useful because they reflect the midpoint of the highest and lowest quintiles, and many quantitative investors and risk managers use quintiles in their empirical work.

[10]The table in Exhibit 2 also shows the estimated standard errors and resulting t-statistics associated with each quantile's estimated intercept and slope. Interpreting these indications of significance requires some care. The marginally positive slopes of the 80th and 90th quantiles indicate that, during this particular month, the best performers among the expensive stocks generated results that were slightly lower than the best performers among the cheap stocks. Moreover, as shown in Exhibit 3, it is quite common for the estimated slope of the 90th quantile to be negative, so that the best performers among the expensive stocks generate higher returns than the strongest performers among the cheap stocks. An investor will generally not know in advance, however, whether a particular stock will actually deliver a result in the 10th or 90th quantile.

[11]This result is consistent with the findings of Mohanram [2005], who also provided some useful insights into how investors can identify more and less attractive firms whose book-to-market valuations are low.

[12]Converting these figures into the price-to-book form that is more familiar to most fundamental analysts, a book-to-price ratio of 0.1 would be the equivalent of trading at a price-to-book multiple of 10. As shown in Exhibit 1, there were plenty of companies trading at such levels at the end of April 2006. Calculating the expected return to the 10th and 90th quantiles involves taking the estimated monthly intercept and the estimated monthly slope, and applying these to the selected values for cheap and expensive names.

[13]The chart in Exhibit 4 shows that, since mid-2007, there has been a noticeable increase in the spreads between the 10th and 90th quantiles for both cheap and expensive stocks, and also indicates that the 10th to 90th spreads have converged. The implication is that the return to a randomly selected portfolio of stocks that are cheap in terms of book-to-price has become rather less predictable than was true in 2005 and 2006. These shifts may well be linked to the unwinding of investment portfolios that had been constructed with an assumption that the value premium is persistent (see Fabozzi, Focardi and Jonas [2008]).

[14]Jegadeesh and Titman [2001] provided a survey of theoretical and empirical work based on their original article.

[15]Not many stocks may actually satisfy both of these criteria, either on the attractive or unattractive side. Assuming that book value does not change very much from year to year, one year ago stocks in the attractive category would presumably have had a book-to-price multiple of approximately 2, while those in the unattractive category would have had a book-to-price multiple of 0.05. These book-to-price multiples would be quite extreme, although Exhibits 1 and 5 show individual stocks that are even more extreme on one or the other dimension.

[16]As discussed in Endnote 12, these patterns have not always prevailed. Over the period from mid-2007 to late 2008, for instance, the dispersion of returns among cheap stocks and expensive stocks has been quite similar. To the extent that these categories mimic the behavior of small-cap value and small-cap growth stocks, respectively, one might infer that successful operation of a small-cap value portfolio during this period has been more challenging than usual.

[17]Some analysts and portfolio managers might believe that during turbulent times it is preferable to reduce the number of names in a portfolio in order to focus on those firms that they understand best. Such a stance may indeed be defensible if the analysts and portfolio managers truly can identify the firms best able to cope with such conditions. However, such a strategy will generally lead to an increase in idiosyncratic risk, which arguably would be more appropriate during periods of relative calm.

[18]The only sector for which this might be problematic would be telecoms, which has comparatively few names for reasons that are generally thought to involve network externalities (Economides [1996]). Focusing on a 25th to 75th quantile dispersion might be more reasonable for this particular sector.

REFERENCES

Barberis, N., and R. Thaler. "A Survey of Behavioral Finance" in *The Handbook of the Economics of Finance*, Vol. 1, No. 2, edited by M. Harris, R. Stulz, and G. Constantinides. New York: Elsevier Science, 2003, pp. 1051–1121.

Barrodale, I., and F.D.K. Roberts. "Solution of an Overdetermined System of Equations in the L_1 Norm." *Communications of the Association for Computing Machinery*, 17 (1974), pp. 319–320.

Bassett, G.W., Jr., and H.-L. Chen. "Portfolio Style: Return-based Attribution Using Quantile Regression." *Empirical Economics*, 26 (March 2001), pp. 293–305.

Billett, M.T., and H. Xue. "The Takeover Deterrent Effect of Open Market Share Repurchases." *Journal of Finance*, 62 (August 2008), pp. 1827–1850.

Cade, B.S., and B.R. Noon. "A Gentle Introduction to Quantile Regression for Ecologists." *Frontiers in Ecology and the Environment*, 1 (October 2003), pp. 412–420.

Carhart, M.M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (March 1997), pp. 57–82.

Chan, L.K.C., J. Karceski, and J. Lakonishok. "The Risk and Return from Factors." *Journal of Financial and Quantitative Analysis*, 33 (June 1998), pp. 159–188.

Chernozhukov, V., and L. Umantsev. "Conditional Value at Risk: Aspects of Modeling and Estimation." *Empirical Economics*, 26 (March 2001), pp. 271–292.

Coad, A., and R. Rao. "Innovation and Firm Growth in High-Tech Sectors: A Quantile Regression Approach." *Research Policy*, 37 (May 2008), pp. 633–648.

Economides, N. "The Economics of Networks." *International Journal of Industrial Organization*, 14 (March 1996), pp. 673–699.

Engle, R.F., and S. Manganelli. "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles." *Journal of Business & Economic Statistics*, 22 (October 2004), pp. 367–381.

Fabozzi, F., S.M. Focardi, and C. Jonas. *Challenges in Quantitative Equity Management*. Charlottesville, VA: CFA Institute. 2008.

Fama, E.F., and K.R. French. "The Cross Section of Expected Stock Returns." *Journal of Finance*, 47 (June 1992), pp. 427–465.

——. "Dissecting Anomalies." Center for Research in Securities Prices Working Paper No. 610, 2007.

Fattouh, B., P. Scaramozzino, and L. Harris. "Capital Structure in South Korea: A Quantile Regression Approach." *Journal of Development Economics*, 76 (February 2005), pp. 231–250.

Giacomini, R., and I. Komunjer. "Evaluation and Combination of Conditional Quantile Forecasts." *Journal of Business & Economic Statistics*, 23 (October 2005), pp. 416–431.

Görg, H., and E. Strobl. "Multinational Companies and Entrant Start-up Size: Evidence from Quantile Regressions." *Review of Industrial Organization*, 20 (February 2002), pp. 15–31.

Jegadeesh, N., and S. Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48 (March 1993), pp. 65–91.

——. "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations." *Journal of Finance*, 56 (April 2001), pp. 699–720.

Koenker, R. "Galton, Edgeworth, Frisch, and Prospects for Quantile Regression in Economics." *Journal of Econometrics*, 95 (April 2000), pp. 347–374.

——. *Quantile Regression*. New York: Cambridge University Press, 2005.

Koenker, R., and G. Bassett, Jr. "Regression Quantiles." *Econometrica*, 46 (January 1978), pp. 33–50.

Koenker, R., and V. d'Orey. "Computing Regression Quantiles." *Applied Statistics*, 36 (1987), pp. 383–393.

Koenker, R., and K.F. Hallock. "Quantile Regression: An Introduction." *Journal of Economic Perspectives*, 15 (Fall 2001), pp. 143–156.

Koenker, R., and Z. Xiao. "Quantile Autoregression." *Journal of the American Statistical Association*, 101 (September 2006), pp. 980–990.

Lee, T.-H., and W. Yang. "Money-Income Granger-Causality in Quantiles." Working Paper, University of California at Riverside, 2007.

Mohanram, P.S. "Separating Winners from Losers among Low Book-to-Market Stocks Using Financial Statement Analysis." *Review of Accounting Studies*, 10 (September 2005), pp. 133–170.

Nguyen, B.T., J.W. Albrecht, S.B. Vroman, and M.D. Westbrook. "A Quantile Regression Decomposition of Urban-Rural Inequality in Vietnam." *Journal of Development Economics*, 83 (July 2007), pp. 466–490.

Portnoy, S., and R. Koenker. "The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators, with discussion." *Statistical Science*, 12 (November 1997), pp. 279–300.

Reichstein, T., M.S. Dahl, B. Ebersberger, and M.B. Jensen. "The Devil Dwells in the Tails: A Quantile Regression Approach to Firm Growth." Danish Research Unit for Industrial Dynamics Working Paper No. 06–34, November 2006.

Stasavage, D. "Private Investment and Political Uncertainty." *Economics and Politics*, 14 (March 2002), pp. 41–63.

Taylor, J. "A Quantile Regression Approach to Estimating the Distribution of Multiperiod Returns." *Journal of Derivatives*, 24 (Fall 1999), pp. 64–78.

Wu, G., and Z. Xiao. "An Analysis of Risk Measures." *Journal of Risk*, 3 (2002), pp. 53–60.

Vaz, S., C.S. Martin, P.D. Eastwood, B. Ernande, A. Carpentier, G.J. Meaden, and F. Coppin. "Modelling Species Distributions Using Regression Quantiles." *Journal of Applied Ecology*, 45 (February 2008), pp. 204–217.

Yu, K., Z. Lu, and J. Stander. "Quantile Regression: Applications and Current Research Areas." *The Statistician*, 52 (October 2003), pp. 331–350.