

Metrick, Andrew

Article

Performance evaluation in financial economics

NBER Reporter Online

Provided in Cooperation with:

National Bureau of Economic Research (NBER), Cambridge, Mass.

Suggested Citation: Metrick, Andrew (2002) : Performance evaluation in financial economics, NBER Reporter Online, National Bureau of Economic Research (NBER), Cambridge, MA, Iss. Summer 2002, pp. 5-7

This Version is available at:

<https://hdl.handle.net/10419/61811>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Performance Evaluation in Financial Economics

Andrew Metrick*

A mutual-fund manager earns annualized returns of 20 percent per year for a five-year period. Over the same period, the stock market as a whole earns 10 percent per year. Was this manager smart, or just lucky?

Some companies engage in a lot of merger activity. Other companies do not. A researcher finds that the former group performs less well than the latter group in the stock market. Is this difference related to the merger activity, or does it simply reflect underlying differences between the two groups of firms?

While the questions just raised may seem quite different, they can be answered using similar methods. In both cases, it is necessary to define some appropriate “benchmark” return. This benchmark return then can be compared to the actual return earned by the mutual fund manager, group of merged firms, or group of non-merged firms. The difference between the actual and benchmark returns then can be defined as an “abnormal” return. Abnormal returns then can be tested for statistical and economic significance.

These are the key steps in performance evaluation (PE), a methodology central to the investigation of many questions in financial economics. The seminal PE study, Jensen (1968), uses the classic Capital Asset Pricing Model (CAPM) as its benchmark and analyzes mutual funds¹; for the next 25 years,

most PE studies followed this same strategy. In the last ten years, though, researchers have developed many new models of benchmark returns and demonstrated their usefulness in PE studies of both investor performance and corporate finance. In this article, I illustrate some of these diverse applications with recent examples from my own work and with studies of investment newsletters, insider trading, and corporate governance. I then discuss a new approach to PE that allows fresh insights into the canonical mutual-fund topic. I conclude with a discussion of future directions for PE-based research.

Applications

Investment newsletters have been around since the early 1900s, and the current industry of over 500 active letters has about 2 million subscribers. The typical newsletter is produced by a small staff and provides a wide range of advice targeted at the retail investor. Is any of this advice useful? Using PE methodology, I analyze the performance of newsletters’ equity recommendations using a dataset of 153 newsletters that spans 17 years.² In contrast to most PE studies, this study’s data contain information about every transaction, rather than just the periodic returns earned by these transactions. Thus, I can address two questions: First, do investment newsletters have stock-selection ability? Second, can transactions data be used to improve the precision of PE?

In response to the first question, I find that newsletters do not demonstrate significant abnormal performance: average abnormal returns are

close to zero; the best performing newsletter does not seem unusual given the sample size; and the number of extreme performers is not surprising. Taken together, these results imply that the average subscriber is not getting useful stock-selection advice.

To address the second question, I compare several methods. Most PE refinements involve adding additional benchmarks and forming multifactor extensions to the regression framework of the CAPM. These methods require only periodic return data. When transactions data are available, portfolios can be compared on a day-to-day basis, with each stock matched to an appropriate benchmark.³ Using a measure of precision defined in the paper, I find that the transactions-based approach yields a median improvement of 10 percent over an analogous multifactor model, with the former approach providing more precise estimates of abnormal performance for over 80 percent of the newsletters. This compares with a median improvement of less than one percent achieved by adding factors to the CAPM.

The increased precision of transactions data is also available for the trades made by corporate insiders, a group that includes most senior officers and all members of the board of directors. By law, insiders must file monthly SEC reports about their trades in their company’s stock, and these reports are quickly made public. They have been used by many authors, with most studies focused on attempts to build profitable trading strategies for non-insiders based on the disclosed insider-trading activity.⁴ Leslie Jeng, Richard Zeckhauser, and I take a different approach and use PE meth-

* Metrick is an NBER Faculty Research Fellow in the Asset Pricing Program and an Assistant Professor of Finance at the Wharton School of the University of Pennsylvania. His “Profile” appears later in this issue.

ods to compute the profits made by insiders themselves on all reported trades from 1975 to 1996.⁵ To do this, we place all insider purchases into a portfolio and hold them for exactly six months. This “purchase portfolio” is like a shadow mutual fund managed by the combination of all insiders. Similarly, we construct a “sale portfolio” comprised of all shares sold by insiders, with those shares held in the portfolio for exactly six months. The six-month holding period, while arbitrary, corresponds to the minimum time that an insider must hold a stock while still retaining profits from an off-setting transaction.⁶

We find that the purchase portfolio earns abnormal returns but that the sale portfolio does not. In raw returns, the purchase portfolio outperforms the market by 10.2 percent per year. Using several PE methods, the abnormal performance ranges between 50 and 67 basis points per month. About one quarter of these abnormal returns accrues within the first five days after the trade and one half accrues within the first month.

These results can be used to shed some light on the effectiveness of current insider-trading regulation. For example, despite the economically large abnormal returns to the purchase portfolio, non-insider counterparties have little to fear from these reported transactions, we find, because insider trades make up only a tiny portion of the market. We calculate that the expected loss to non-insiders attributable to the purchases of insiders is about 0.10 basis points over the subsequent six months. This translates into 10 cents for a \$10,000 transaction.

Studies of investment newsletters and insider trading are standard topics for PE, which traditionally has been used to analyze investor performance. The same tools, however, have also become important for corporate finance. Historically, many corporate-finance questions were analyzed using “event-study” methodology. In recent years, several authors have shown that event studies can have severe statistical problems when used to analyze long-horizon returns. One solution to these problems is a PE analysis conducted

on portfolios of event firms. Subsequently, some studies have used PE methods and, in several cases, reached conclusions differing from the event-study literature.⁷

Paul Gompers, Joy Ishii, and I take a PE approach to a corporate finance topic in a study of corporate governance.⁸ Corporate governance is defined by the set of rules, laws, and institutions that regulate the relationship between the shareholders and the managers of a corporation. Using the incidence of 24 governance rules at 1500 large firms, we construct an index to proxy for the level of shareholder rights at each firm during the 1990s. An investment strategy that bought firms in the lowest decile of the index (strongest rights) and sold firms in the highest decile of the index (weakest rights) would have earned abnormal returns of 8.5 percent per year between 1990 and 1999. Also, we find that firms with stronger shareholder rights had higher profits, higher sales growth, lower capital expenditures, and made fewer corporate acquisitions. We consider several explanations for the results, but the data do not allow strong conclusions about causality. There is some evidence, both in our sample and from other authors, that weak shareholder rights caused poor performance in the 1990s. It is also possible that the results are driven by some unobservable firm characteristic.

The abnormal returns to this investment strategy must be interpreted with care. When PE methods are used to evaluate a mutual fund manager, abnormal returns are sometimes thought to measure the investment “skill” of the manager. If a manager has skill, then one would expect abnormal returns to continue in future periods. For our governance study, the investment strategy is an artificial construct designed to isolate the relationship between governance and returns over some prior time period. We argue in the paper that there is no reason to expect that such abnormal returns would continue in future periods; rather, a more plausible explanation is that these abnormal returns reflect a slow adjustment, as investors learn about the impact of governance on

operating performance and agency costs.

Notwithstanding recent improvements in PE methodology, it is still very difficult to detect abnormal performance in most applications. For example, for typical portfolios of 100 stocks followed for ten years, the standard error for the abnormal-performance estimate would be about 25 basis points per month, or approximately 3 percent per year. In this case, a 95 percent confidence interval would include a range of abnormal performance of approximately 12 percent per year. For portfolios with fewer stocks or shorter histories, the range can be much larger. Thus, standard statistical tests often may fail to reject a null hypothesis of “no abnormal performance,” even when the true abnormal performance is economically large.

I first encountered the power limitations of PE in the investment newsletter study. There, it became clear to me that it would only be possible to make strong statements about average returns of all newsletters for the whole sample period, an analysis with a relatively low standard error for abnormal performance. In the studies of insider trading and corporate governance, the time periods were long enough and abnormal returns large enough to allow for statistical significance. But what if researchers want to provide guidance about investment strategies that have short histories and high volatility?

Consider the canonical PE topic of mutual funds. Most mutual funds are actively managed and charge fees averaging more than one percent per year. In contrast, passively managed index funds seek to replicate benchmark returns at a much lower cost. Since the seminal work of Jensen (1968), researchers have used a wide variety of PE models and datasets in hundreds of published analyses. A rough consensus of this literature is that the average actively managed mutual fund does not earn abnormal returns, and, while some funds may earn consistently positive abnormal returns, it is difficult to identify such funds, *ex ante*. But what does this mean for investors? Should investors only choose low-cost index funds?

Klaas Baks, Jessica Wachter, and I answer this question by explicitly taking an investor's perspective.⁹ We study the one-period portfolio allocation problem for an investor choosing from a riskless asset, benchmark assets (passively managed index funds), and non-benchmark assets (actively managed funds). We model the investor's decision in four steps. First, he states his belief about the distribution of investment skill in the population of all managers. (For this discussion, think of investment skill as equivalent to "expected abnormal returns of 3 percent per year.") Second, he observes and evaluates the history of returns for some group of managers. Third, he uses this history to update his beliefs about the skill of each manager in the group. Fourth, he makes an investment decision.

This "Bayesian" method of PE allows all investors to filter evidence through their own beliefs about managerial skill. Clearly, an investor who believes that no manager can possibly have skill would not choose to invest with active managers. Also, an investor with completely uninformative beliefs would lean towards investment after only a single period of good returns. We are interested in the vast middle ground; given the available statistical evidence, what prior beliefs would imply any investment in active managers? We find that an investment in active managers only requires a belief that at least one in 10,000 mutual fund managers has skill. From a frequentist statistical perspective, such beliefs are indistinguishable from a belief that "no manager has skill." We conclude that the case against investing in active managers cannot rely only on the return evidence. More generally, these results motivate the use of a Bayesian method of PE, where researchers can state the economic significance of their results as filtered through a range of plausible beliefs.

Future Directions

Innovations in PE methodology and applications to new problems are continuing at a rapid rate. In recent years, researchers have extended PE methods in several directions, including adjustments for predictable variation in benchmark expected returns, development of benchmarks that correspond to complex investment strategies used by hedge funds, and methods more closely tied to theoretical models of asset prices.¹⁰ While it will never be possible to specify a single "correct" model of benchmark expected returns, recent research demonstrates how to explicitly add model-based error into PE.¹¹ These methodological advances, when combined with the explosion of new data sources, will allow a fresh perspective on many topics in financial economics.

¹ M. C. Jensen, "The Performance of Mutual Funds in the Period 1945-1964," *The Journal of Finance*, 23 (May 1968), pp. 389-416.

² A. Metrick, "Performance Evaluation with Transactions Data: The Stock Selection of Investment Newsletters," NBER Working Paper No. 6648, July 1998, and *The Journal of Finance*, 54 (5) (October 1999), pp. 1743-75.

³ The most widely used multifactor model in PE is the four-factor model of M. Carhart, "On Persistence in Mutual Fund Performance," *The Journal of Finance*, 52 (March 1997), pp. 57-82. A transactions-based method that is its closest analogue is K. D. Daniel, M. Grinblatt, S. Titman, and R. Wermers, "Measuring Mutual Fund Performance with Characteristic Based Benchmarks," *The Journal of Finance*, 52 (August 1997), pp. 1035-58.

⁴ A thorough survey of these studies is given in H. N. Seyhun, *Investment Intelligence from Insider Trading*, Cambridge, MA: MIT Press, 1998.

⁵ L. A. Jeng, A. Metrick, and R. J. Zeckhauser, "The Profits to Insider Trading: A Performance-Evaluation Perspective," NBER Working Paper No. 6913, January 1999.

⁶ The six-month "short-swing" rule, SEC Rule 16(b), requires insiders to disgorge any profits made by offsetting transactions within a six-month window.

⁷ These statistical problems are documented by B. M. Barber and J. D. Lyon, "Detecting Long-Run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics," *Journal of Financial Economics*, 43 (March 1997), pp. 341-72; and S.P. Kothari and J. B. Warner, "Measuring Long-Horizon Security Price Performance," *Journal of Financial Economics*, 43 (March 1997), pp. 301-39. Several examples of differing conclusions between PE and event studies are given in M. L. Mitchell and E. Stafford, "Managerial Decisions and Long-Term Stock Price Performance," *The Journal of Business*, 73 (3) (July 2000), pp. 287-330.

⁸ P. A. Gompers, J. L. Ishii, and A. Metrick, "Corporate Governance and Equity Prices," NBER Working Paper No. 8449, August 2001, and *The Quarterly Journal of Economics*, forthcoming in February 2003.

⁹ K. Baks, A. Metrick, and J. A. Wachter, "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation," NBER Working Paper No. 7069, April 1999, and *The Journal of Finance*, 56 (1) (February 2001), pp. 45-86.

¹⁰ For examples of this work, see J. A. Christopherson, W. E. Ferson, and D. A. Glassman, "Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance," NBER Working Paper No. 5830, November 1996, and *Review of Financial Studies*, 11 (Spring 1998), pp. 111-42; W. Fung and D. A. Hsieh, "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers," *Review of Financial Studies*, 14 (Summer 2001), pp. 313-41; and H. Farnsworth, W. E. Ferson, D. Jackson, and S. Todd, "Performance Evaluation with Stochastic Discount Factors," NBER Working Paper No. 8791, February 2002.

¹¹ L. Pastor and R. F. Stambaugh, "Evaluating and Investing in Equity Mutual Funds," NBER Working Paper No. 7779, July 2000, and *Journal of Financial Economics*, 63 (3) (March 2002).