

Sources of Institutional Performance

Boris Kovtunen

Nathan Sosner¹

JOB MARKET PAPER

First Draft: June 2003

This Draft: November 2003

Abstract

An emerging body of empirical literature finds that institutions as a group are better investors than an average market participant. One part of the literature finds that institutions as a group outperform their benchmarks, before costs. This paper shows that outperformance of the aggregate institutional portfolio found in the previous literature can be explained by institutional preferences for stocks with high accounting profitability. In particular, risk-adjusted excess return on the institutional portfolio over the rest of the market during the 1982-2001 period can be explained by the fact that institutions avoided small stocks with low accounting profitability and overweighted large stocks with high accounting profitability. We do not find any evidence that this preference for accounting profitability is related to price momentum. The other part of the literature finds that fraction of institutional ownership positively predicts future stock returns in a cross-sectional regression. We reconcile these findings with our portfolio results by showing that the fraction of institutional ownership is not a statistically significant predictor of firm-level stock returns once we take into account book-to-market, return on equity, size, and biases caused by high positive correlation between the fraction of institutional ownership and stock size.

JEL classification: G14; G20.

Keywords: Financial institutions; Performance evaluation; Return on equity; Expected returns.

¹ Harvard University Department of Economics, Littauer Center, 1875 Cambridge Street, Cambridge MA 02138, USA; Email kovtunen@fas.harvard.edu and sosner@fas.harvard.edu. We would like to thank John Campbell, Ken Froot, Robin Greenwood, Andrew Metrick, Andrei Shleifer, Samuel Thompson, Tuomo Vuolteenaho, and participants of the Financial Economics seminar at Harvard for helpful comments and suggestions. All errors and omissions remain our responsibility.

1. Introduction

It is widely believed that financial institutions are better investors than the rest of the market, and with new data sets on institutional holdings becoming available, evidence supporting this view has started to accumulate. There are two reasons why institutions might be better investors. First, they might possess more information than the rest of the market or have better technology to process publicly available information. Second, even without any informational advantage, institutions might trade more rationally than other investors.² Empirical predictions are similar in both cases. First, the institutional portfolio should systematically outperform the rest of the market on a risk-adjusted basis. Second, in a regression predicting returns the fraction of institutional ownership should have predictive power beyond other commonly used predictors. In this paper we do not distinguish between the two explanations.³ Our main focus is to test their common empirical implications and to identify the sources of institutional performance.

The first part of the paper studies performance of the aggregate institutional portfolio relative to the rest of the market. We analyze the return on the institutional portfolio derived from quarterly holdings of SPECTRUM institutions. In other words, we examine performance of stocks held by institutions, which contrasts to the analysis of net returns on mutual funds. Although performance of this portfolio is not equivalent to the actual performance institutional portfolio because of intra-quarter trading, it can provide indirect evidence on the institutional ability to beat their benchmarks before costs. This approach was used in a number of other studies. Grinblatt and Titman (1989, 1993), Grinblatt, Titman, and Wermers (1995), Daniel et al. (1997), and Wermers (2000) find that mutual

² The existence of not-fully-rational investors, or noise traders, in the marketplace has become one of the central assumptions in modern financial modeling (e.g. see O'Hara (1995)). At the same time, the identity of noise traders remains unknown. One approach to identifying noise traders is to find the group of market participants who actively trade securities without possessing any superior information about expected returns. Lee et al. (1991) suggest that prices of stocks predominantly held by individuals might be considerably affected by noise trading. With new data on individual investment accounts becoming available, researchers have looked at whether individuals can be considered as noise traders (see Odean (1998, 1999) and Odean and Barber (2001)). An alternative approach is to focus on the performance of institutional investors who are generally not viewed as noise traders.

³ An important difference between the two explanations is that the former works even if markets are efficient, while the latter requires the existence of market inefficiencies.

funds have the ability to choose stocks that outperform their benchmarks, before any expenses are deducted. Pirinsky (2000) shows that SPECTRUM institutions statistically significantly outperform their benchmarks.

We show that the outperformance of the value-weighted aggregate portfolio held by SPECTRUM institutions can be attributed to institutional preferences towards stocks with high accounting profitability, measured by return on equity (ROE). To take the effect of these preferences into account, we construct an ROE factor, defined as a zero-investment portfolio that is long high ROE stocks and short low ROE stocks.⁴ Once the ROE factor is added to the Fama and French (1993) three-factor model assessing the relative performance of the institutional portfolio, the alpha becomes zero in the early sample period, 1982-1998, and negative but not statistically significant in the full sample period, 1982-2001. These findings suggest that an investor who takes into account institutional preference for high ROE stocks by using the ROE factor in the four-factor model, can improve the performance of her portfolio up to the point that she does no worse than the aggregate institutional portfolio.⁵

We also show that most of the explanatory power of the ROE factor comes from the fact that SPECTRUM institutions as a group avoid small stocks with low ROE and overweight large stocks with high ROE. In addition, we document a shift in the institutional investment strategy in the years 1999 to 2001 as compared to the earlier period from 1982 to 1998. Namely, in the period from 1982 to 1998 institutions as a group put more weight on large stocks with low ROE than the rest of the market, which was detrimental to the relative performance of the institutional portfolio. This effect is not observed in the full sample, 1982-2001, which partially explains the high relative performance of the institutional portfolio in the years 1999 to 2001.

⁴ The ROE factor is also balanced for size and book-to-market.

⁵ A number of studies demonstrate preferences of institutions for particular stock characteristics including size and book-to-market (e.g. Falkenstein (1996), Del Guersio (1996)). This paper concentrates on institutional preferences towards an additional characteristic not explored in these studies – accounting profitability measured by ROE.

The second part of the paper reconciles our portfolio findings with results from cross-sectional models of expected returns. A number of recent studies introduced the fraction of institutional ownership in a model of expected firm-level stock returns. Cross-sectional regressions in Brennan, Chordia, and Subrahmanyam (1997), Gompers and Metrick (2001), Cohen et al. (2002), and Bennett, Sias, and Starks (2002) suggest that, controlling for other predictors, higher institutional ownership predicts higher returns on individual stocks at quarterly and annual horizons. We show that the fraction of institutional ownership is neither a statistically nor an economically significant predictor of firm-level stock returns once book-to-market, return on equity, and size are taken into account. The significance of the fraction of institutional ownership found in the previous literature is a consequence of various biases caused by the fact that the fraction of institutional ownership is very highly positively correlated with stock size.

Expected return regressions can use either simple or log returns. Cohen et al. (2002) show that in cross-sectional regressions with log returns, fraction of institutional ownership is statistically and economically significant predictor of returns even after controlling for book-to-market and ROE. We argue that this result is driven by the fact that small stocks have much higher cross-sectional variance of returns than medium and large stocks and, as a result, the average log return is significantly lower for small stocks than for medium and large stocks, while average simple returns do not vary much across size groups. At the same time, the average fraction of institutional ownership is much lower in small stocks than in medium and large stocks. As a result, the fraction of institutional ownership is highly statistically and economically significant in a predictive regression with log returns not because it contains information about future returns beyond book-to-market and ROE, but because its cross-sectional means across size groups are highly positively correlated with cross-sectional variances of returns.

When simple returns are used in cross-sectional regressions, we show that the fraction of institutional ownership is a significant predictor of stock returns beyond book-to-market, ROE, and log market equity because of the positive bias introduced by the assumption that a one percent increase in market capitalization has the same effect on the expected

return of a small stock as on the expected return of a large stock. Once we take into account that a one percent increase in market capitalization has three to four times stronger effect on expected return among the smallest stocks than among the largest stocks, the fraction of institutional ownership becomes neither statistically nor economically significant predictor of firm-level stock returns.

The rest of the paper is organized as follows. Section 2 analyzes the performance of institutional portfolio. Section 3 considers fraction of institutional ownership in cross-sectional models of expected returns. Section 4 concludes.

2. Performance of Institutional Portfolio

2.1 Data

The data come from the CRSP-COMPUSTAT intersection linked to the SPECTRUM database of institutional holdings. Our final sample covers the period 1982-2001 (80 quarters) and consists of 311,209 firm-quarters. Appendix A provides details on data sources, variable definitions, and descriptive statistics of the data series.

2.2 Measuring performance of institutional portfolio relative to the rest of the market

Fraction of institutional ownership represents an investment decision – the demand of SPECTRUM portfolio managers. This allows us to look directly at performance of SPECTRUM institutions as a group relative to the rest of the market. For simplicity, hereafter we refer to the rest of the market as individuals. To assess the performance of institutions relative to individuals we use a simple relative performance measure. Every month from January 1982 to December 2001 we calculate the excess return of the value-weighted institutional portfolio over the value-weighted individual portfolio:

$$R_t^{Ex} = R_t^{Inst} - R_t^{Ind}.$$

Since institutional holdings are available only at quarterly frequency, we work under the assumption that institutional portfolio weights stay constant during the quarter. As a result, our measure of the relative performance of institutional portfolio does not directly

correspond to the performance of the actual institutional portfolio because we are missing portfolio rebalancing within a quarter. However, it allows us to assess the institutional performance before any expenses are deducted. Further details on calculation of institutional and individual portfolio returns are provided in Appendix B.1.

Figure 1 provides the first look at the relative performance of institutions: Institutional portfolio consistently outperforms individual portfolio over the sample period. During the twenty year period from 1982 to 2001 institutions outperformed individuals by 1.23 percent per year, on average (see Table 1, Panel A). As is clear from Figure 1, a substantial part of this outperformance comes from the year 2000 where institutions outperformed individuals by 10.17 percent (85 basis points a month, on average). Without the year 2000, in the remaining nineteen years, the performance of institutions becomes less spectacular, with an average annualized excess return of 76 basis points. Later in this section we will explore in detail institutional performance during the year 2000. Hereafter, we consider two alternative samples – 1982 to 1998 and 1982 to 2001. We refer to the former as the early sample and to the latter as the full sample.

Table 1, Panel A, shows annualized mean return for the early and the full samples. During the period from 1982 to 1998 the annualized mean excess return was 87 basis points with an annualized Sharpe ratio of 0.47. The t-statistic indicates that for the early sample we can reject the hypothesis that institutional portfolio did not outperform individual portfolio at a 5.3 percent confidence level.⁶ For the full sample the annualized mean excess return is 1.23 percent a year with an annualized Sharpe ratio of 0.59. The t-statistic for the full sample is 2.66, indicating that we can reject the hypothesis that institutions did not outperform individuals at the 1 percent confidence level. Notably, institutions outperformed individuals exactly at the time that market returns were low – the average excess return on the CRSP value weighted index from 1982 to 2001 was 8.90

⁶ T-test requires that excess returns are i.i.d. normal. The assumption of time-series independence is plausible since the series is monthly excess returns. Indeed, the autocorrelation coefficient of excess returns is negative 10 percent and is not statistically significant. Under the null of no autocorrelation the standard deviation of autocorrelation coefficient is $1/\sqrt{T}$, which is 6.5 percent for the full sample

percent a year, with a Sharpe ratio of 0.61, while the average excess return from 1982 to 1998 was 10.85 percent a year with an annualized Sharpe ratio of 0.74.

2.3 *The ROE factor*

Cohen et al. (2002) show that institutions prefer stocks with higher accounting profitability: The fraction of institutional ownership is higher for stocks with high return on equity (ROE). In addition, expected returns are strongly positively associated with accounting profitability (see e.g. Cohen et al. (2002) and Haugen and Baker (1996)). In order to take these effects into account, we construct an ROE factor, defined as a portfolio capturing stock returns associated with ROE.

We create three ROE sorted portfolios to ensure that our results are not an artifact of equally-weighted or value-weighted returns. The portfolios are size and book-to-market balanced and are based on quarterly sortings.⁷ Every quarter we divide stocks into six size groups (see Appendix A.3 for the definition of size groups). Then stocks in every size group are sorted into five book-to-market quintiles. We assume a two quarter lag in release of accounting information. For instance, if the fiscal year end of a firm is in quarter one, its book-to-market ratio based on the new book value of equity will be used for sorting only in the end of quarter three.

The size-book-to-market sorting yields thirty groups of stocks. We apply three alternative procedures to create ROE sorted portfolios. The first two procedures sort stocks by ROE within each of the thirty size-book-to-market cells. We define the top thirty percent of stocks in each cell as high ROE stocks and the bottom thirty percent as low ROE stocks. Similar to book-to-market, we assume that ROE becomes available after two quarters from the firm's fiscal year end. We end up with thirty high ROE groups and thirty low ROE groups – sixty groups in total. Each month we calculate average returns for the stocks in each group. Procedure one uses equally-weighted

consisting of 240 months (see Campbell, et al. (1997, Chapter 2) for distribution of autocorrelation statistic).

average return, while procedure two uses value-weighted average. Then we equally weight returns on the thirty high ROE portfolios and returns on the thirty low ROE portfolios, yielding two portfolios – high ROE and low ROE portfolio. Finally, we subtract the return of the low ROE portfolio from the return of the high ROE portfolio creating ROE portfolio return. We dub these ROE based portfolios the ROE factor mimicking portfolios, or for briefness the ROE factors. We refer to the portfolio obtained using equally-weighted group returns as the equally-weighted ROE factor and to the portfolio based on value weighted group returns as the value-weighted ROE factor.

Our third procedure uses ROE directly to determine portfolio weights. We create thirty ROE-weighted portfolios – one portfolio for each size-book-to-market cell. Portfolio weights are equal to ROE of a stock minus average ROE of the portfolio. In addition, we require that the sum of positive weights is equal to one. One can think of this portfolio as going one dollar long in stocks with ROE above average and one dollar short in stocks with ROE below average. Finally, we equally weight the thirty ROE portfolios. We refer to the resulting portfolio as the ROE-weighted ROE factor.

We created three different ROE factor portfolios to make sure that past ROE effect is robust to the method of calculation. Table 2 demonstrates performance of the three ROE factor portfolios and their relationships with each other and with the Fama-French factors – RMRF, SMB, and HML. Panel A shows raw performance of the three ROE factors in the early and the full samples. In the early sample, the equally-weighted, value-weighted and ROE-weighted ROE factor portfolios yield annualized return of 7.6, 8.1, and 8.5 percent, respectively. These excess returns are highly statistically significant with t-statistics above 5.⁸ In the full sample, the ROE factor portfolio returns become lower, with 7.0, 7.4, and 7.9 percent per annum, respectively, but still remain highly statistically

⁷ Our construction of accounting variables based on the actual fiscal year ends allows us to do book-to-market and ROE sortings at a higher than annual frequency. To be consistent with the frequency of institutional holdings data, we perform sortings at the quarterly frequency.

⁸ T-statistics presented in Table 2 take first order serial correlation into account. The ROE portfolios exhibit first order serial correlation of around 20 percent which is statistically significant. Higher order autocorrelations are negligible.

significant with t-statistics above 3. The results in Panel A show that there are expected returns associated with past accounting earnings.

Panel B of Table 2 shows bivariate correlations of the three ROE factor portfolios with each other and with Fama-French factor mimicking portfolios – RMRF, SMB, and HML. Returns of the three ROE factors are highly correlated with each other, with correlations above 90 percent in both the early and the full samples. This indicates that all three portfolios capture similar effect associated with past ROE. The correlations of the ROE factors with SMB and HML indicate that high ROE stocks tend to be larger and to have higher book-to-market ratios. We test whether Fama-French factor portfolios can replicate the ROE factor portfolios. Panel C of Table 2 presents the results of these tests. The main conclusion is that Fama-French factors cannot account for the returns of the ROE factor portfolios – all three ROE factors have large and highly significant alphas in both the early and the full samples.⁹ In addition, they do not have significant exposure to the market risk – their loadings on the market factor are not significantly different from zero. Also the ROE factors have positive loadings on larger stocks and, mostly in the later part of the sample, they load positively on value stocks. Panel D of Table 2 shows that the ROE factor is positively related to UMD factor that captures momentum. However, introduction of UMD factor has little effect on the alpha of the ROE factors in both the early and the full sample.

In sum, the results in Table 2 show that, although we use different procedures to construct the three ROE factors, their properties are a very similar. The remainder of the paper uses the value-weighted ROE factor, to which we will refer as the ROE factor. Figure 2 plots cumulated returns on the ROE factor and the Fama-French factors. ROE factor and HML exhibit similar trend in 1999-2001 time period – this increases the correlation between the ROE factor and HML when we move from the early to the full sample.

2.4 *Performance of institutional portfolio relative to the rest of the market*

Table 1, Panel B, shows the factor-adjusted performance of the institutional portfolio relative to the individual portfolio. In the early sample, institutions outperform individuals by 50 basis points per year controlling for the three Fama-French factors, although this result is not statistically significant. In the full sample, institutions outperform individuals by 67 basis points a year controlling for the three Fama-French factors, which is statistically significant at the 10 percent confidence level. In both samples, institutions tend to invest in larger stocks than individuals as the negative loading on SMB suggests. Also, in the early sample institutions seem to have a higher preference for growth stocks than individuals – the loading on HML is negative. This higher institutional preference for growth stocks is significantly affected by addition of the years 1999 to 2001 to the sample: The loading on HML factor is not statistically significant in the full sample.

Columns 2 and 4 of Table 1, Panel B include the ROE factor in addition to the Fama-French factors in the regression. The ROE factor is positive and highly statistically significant in both samples. In addition, the introduction of the ROE factor significantly decreases the alpha, making it zero in the early sample and slightly negative in the full sample. Thus, institutional preferences for high ROE stocks add to the relative performance of their portfolio: Relative to individuals, institutions overweight stocks with high accounting profitability and avoid stocks with low accounting profitability. The ROE factor plays an especially important role in the full sample. Except for statistically significantly positive market beta, the ROE factor is the only statistically significant factor explaining the relative performance of the institutional portfolio. Its inclusion in the regression makes SMB factor not statistically significant, while HML factor still remains not statistically significant. Moreover, ROE factor substantially improves the fit of the regression: The adjusted R-squared increases from 19.7 to 30.4 percent. In the early sample, the effect of the ROE factor is incremental. Its inclusion in

⁹ T-statistics presented in Panel C are based on heteroscedasticity and autocorrelation corrected standard errors. We only correct for the first order serial correlation since higher order autocorrelations of the residuals are negligible.

the regression does not significantly affect the coefficients on SMB and HML and improves the adjusted R-squared by only about 1.9 percent, from 38.2 to 40.1 percent.

When the Fama-French momentum factor, UMD, is included in any of the models in Table 1, Panel B, it is not statistically significant. In fact, inclusion of UMD factor reduces the adjusted R-squared in all of the regressions.¹⁰

The main conclusions that we draw from Table 1 are the following. For the early sample period from 1982 to 1998, institutions outperform individuals by 87 basis points a year. More than forty percent of this excess performance is explained by the three Fama-French factors – the alpha of the three-factor model is reduced to 50 basis points a year. However, institutions tend to invest in high ROE stocks relatively more than the rest of market, which explains the rest of their alpha: Introduction of the ROE factor portfolio in addition to the three Fama-French factors into the regression reduces the institutional alpha to zero. In the full period from 1982 to 2001, institutional portfolio outperforms the individual portfolio by 1.23 percent a year, which is statistically significant at the 1 percent confidence level. Almost fifty percent of this outperformance can be explained by the three-factor model – after controlling for the exposure to RMRF, SMB, and HML, alpha of the institutional portfolio is 67 basis points per year. If we further control for the exposure to the ROE factor, the alpha becomes slightly negative – individuals outperform institutions by 9 basis points a year, although this result is not statistically significant. Figure 3 illustrates the importance of the ROE factor in explaining the relative performance of the institutional portfolio. For both the early and the full sample, it plots cumulative excess returns on the institutional portfolio and fitted returns for the models with and without the ROE factor. In both samples, inclusion of the ROE factor shifts the fitted line uniformly upward. Thus, an investor who takes into account the institutional preferences for stocks with high accounting profitability can significantly improve her portfolio performance up to the point that she will do as well as the aggregate institutional portfolio.

¹⁰ Results are available from the authors upon request.

The year 2000 is characterized by an abnormal performance of the institutional portfolio that neither the three-factor nor the four factor model can adequately explain. An investor with one dollar long the institutional portfolio and one dollar short the individual portfolio in December 1999 would have earned 10.59 cents by the end of December 2000. Using the three-factor or the four-factor replicating strategy, she would lose 0.43 cents or gain 0.86 cents, respectively, over the same time period. Using the ROE factor improves the performance of the model in the year 2000 by 1.29 percent. However, it explains only 12 percent of the total relative return on the institutional portfolio in this year. Figure 4 further illustrates this point. It plots cumulated residuals – alpha plus the error term – for the models with and without the ROE factor over the full sample. The four factor model has consistently lower alpha, i.e. it consistently outperforms the three-factor model. However, both models have high positive residuals in the year 2000.

The comparison between the four-factor models in columns two and four of Table 1, Panel B, suggests that there was a shift in institutional strategy during the years 1999-2001 relative to the early period of 1982-1998. In the early sample, all four factors are significant in explaining the relative return on the institutional portfolio. In the full sample, the difference in performance of institutional and individual portfolio is fully explained by institutional preferences for higher beta and higher ROE stocks. Moreover, the inclusion of the ROE factor in the full sample seems to explain the institutional preferences for larger stocks – SMB becomes not statistically significant when we include the ROE factor in the regression. This suggests that the shift in the institutional strategy in 1999-2001 was driven by the interplay between institutional preferences for size and accounting profitability. Below we argue that this shift was driven by the change in institutional preferences towards large stocks with low accounting profitability.¹¹

¹¹ Figure 5 provides a graphic illustration of the point that there indeed was a shift in institutional strategy in 1999-2001. It presents the relative return on the institutional portfolio in every month from January 1999 to December 2001, together with the 95 percent confidence interval of the model that uses the estimated factor loadings from the 1982-1998 regressions to predict the relative institutional performance in 1999-2001 under the hypothesis of zero alpha. Further details on the construction of the confidence

This significance of the ROE factor in explaining the relative performance of the institutional portfolio can potentially come from different sources. Institutions may overweight stocks with high accounting profitability or, alternatively, they may avoid stocks with low accounting profitability. In addition, their preferences for profitable stocks might vary across size groups. To investigate these possibilities, we disaggregate the ROE factor into components corresponding to high or low ROE and to the six size groups. Then we use those portfolios in the regressions of relative return on the institutional portfolio, together with RMRF, SMB, and HML (see Appendix B.3 for further details). Table 3 shows the results.¹²

We draw several conclusions from Table 3. Column for all stocks shows that both in the early and in the full sample, the significance of the aggregate ROE factor in explaining the relative performance of the institutional portfolio comes from both overweighting stocks with high ROE (positive coefficient on the High ROE portfolio) and avoiding stocks with low ROE (negative coefficient on the Low ROE portfolio). In addition, in both samples the total effect of High ROE is mostly driven by overweighting the largest stock group (Quintile 5), while the total effect of Low ROE is mostly driven by avoiding small stocks (Quintile 1 Large).¹³ Rows for the net ROE effect show that the major difference in the institutional strategy between the early and the full sample is the following: In the early sample, the ROE factor significance is mostly driven by small stocks (Quintile 1 Large and, to some extent, Quintile 2). In the full sample the economic significance of the largest stocks (Quintile 5) dramatically increases because institutions stopped overweighting large stocks with low accounting profitability between 1999 and 2001 (Low ROE, Quintile 5).

The above findings can be used to further improve performance of the four-factor strategy. The two most important effects for the performance of the institutional portfolio

intervals are provided in Appendix B.2. As Figure 5 shows, factor models estimated for 1982-1998 period cannot explain the relative performance of the institutional portfolio in 2000 and early 2001.

¹² All the coefficients in Table 3 are standardized so that we can use their size to compare economic significance of different factors in the same regression.

are overweighting large stocks with high ROE and avoiding small stocks with low ROE. We focus on these two components of the ROE factor and construct the Partial ROE factor as a return on the portfolio that goes long the high ROE stocks in Quintile 5 and goes short the low ROE stocks in the large part of Quintile 1. Table 4 presents our factor regressions with the Partial ROE factor. Column one of Table 4 shows that for the early sample, alpha of the excess return on institutional portfolio drops by 22 basis points per year as compared to the regression with the total ROE factor (Table 1, Panel B, column two). At the same time alpha of the institutional portfolio for the full sample does not change at all (see column two of Table 4 and column four of Table 1, Panel B). This supports the conclusion that compared to the early sample, institutions stopped overweighting large unprofitable stocks in the years 1999-2001.¹⁴

In sum, after controlling for the risk characteristics of the institutional portfolio captured by the three Fama-French factors, we find weak evidence that institutions as a group outperformed the rest of the market. We argue that all this outperformance can be explained by institutional preferences for stocks with high accounting profitability (high ROE). We construct the ROE factor that, together with the three Fama-French factors, fully explains the relative performance of the aggregate institutional portfolio both in the early sample period from 1982 to 1998 and in the full sample period from 1982 to 2001. We also find that in both the early and the full sample, institutions avoided small stocks with low ROE and overloaded on large stocks with high ROE. However, in the years 1999-2001 there was a shift in the institutional strategy compared to the earlier sample of 1982-1998: Institutions stopped overweighting large stocks with low ROE, which was detrimental to the performance of their portfolio. This shift in strategy can partially explain the high relative performance of their portfolio in 2000 and early 2001. We extend this evidence to show that an investor who follows the institutional strategy only in avoiding small stocks with low ROE and overweighting large stocks with high ROE

¹³ Another common fact across both samples is that institutions overweight medium stocks with low accounting profitability (positive and significant loadings on Low ROE in Quintiles 3 and 4), which is detrimental to their relative performance.

¹⁴ Another difference between the total and the Partial ROE factor regression is the sign of the coefficient on SMB. The total ROE factor in Table 1, Panel B, is size balanced, while the Partial ROE factor goes

can further improve the performance of her portfolio in the earlier sample of 1982-1998 up to the point that she will slightly outperform the aggregate institutional portfolio.

2.5 *Discussion of the previous literature*

Our approach is similar to recent studies that look at the performance of the *stocks* held by institutions in their portfolios, which contrasts to studies that look at net returns on mutual fund portfolios after trading costs, fees, and “cash-drag” are taken into account. Grinblatt and Titman (1989, 1993) find that mutual fund managers have the ability to choose stocks that outperform their benchmarks, before any expenses are deducted. The evidence is especially strong among aggressive-growth and growth funds, which were able to outperform their benchmarks by an average of two to three percent per year. However, Daniel et al. (1997) and Grinblatt, Titman, and Wermers (1995) attribute much of this performance to the characteristics of the stocks held by funds. Daniel et al. (1997) consider mutual funds and work on an individual fund basis. They measure the abnormal performance of each fund using book-to-market, size, and momentum characteristic-based approach and find that during the period of 1975-1994 the average fund has a statistically significant positive excess return of 77 basis points per year. However, this performance is mainly concentrated in the first ten-year period, with the 1975-1984 excess return being 1.15 percent per year and highly statistically significant, while the 1985-1994 excess return being 39 basis points per year and not statistically significant. Wermers (2000) uses the same methodology as in Daniel et al. (1997) and finds that, on average, mutual funds hold stocks that outperform a broad market index (the CRSP value-weighted index) by 130 basis points per year during 1975-1994. He provides the following sources for this outperformance. First, funds held stocks having characteristics associated with average returns that were higher than the return on broad market indexes, which provided a boost for the funds returns of 55 to 60 basis points per year above the CRSP index. In addition, the funds chose stocks that outperformed their characteristic benchmarks by an average of 71 basis points per year.¹⁵

long large stocks and short small stocks. To compensate for this size imbalance, the coefficients on SMB become positive and statistically significant in the regressions with the Partial ROE.

¹⁵ In addition, Nofsinger and Sias (1999) find that extreme changes in institutional ownership predict returns – in the year following the change in the fraction of institutional ownership, the decile of firms

Daniel et al. (1997) and Wermers (2000) advocate book-to-market, size, and momentum characteristic-based performance measure against the factor approach that uses factor portfolios based on characteristic-sorted stocks because they believe that characteristic matching does a better job in explaining the cross-sectional pattern in returns on different funds. Since we work with the aggregate institutional portfolio, there is no benefit of using characteristic-based performance measures compared to characteristic-factor portfolios.

The methodology of Daniel et al. (1997) and Wermers (2000) essentially corresponds to evaluating the performance of the equally weighted portfolio of mutual funds, where the return on each fund gets the same weight. As a result, this methodology is not directly comparable with ours because, first, we use a broader set of institutions, and second, we use the value-weighted institutional portfolio to evaluate the performance of institutions as a group. However, a recent study by Pirinsky (2000) provides a link between the results in Daniel et al. (1997) and Wermers (2000) and our results. Pirinsky (2000) employs both characteristic-based and factor-regression methodologies to evaluate the performance of SPECTRUM institutions during the period 1982-1996. In his factor regressions, he uses both equally-weighted and value-weighted portfolios of institutions, which allows us to compare our results with the previous literature. Using the CS (Characteristic Selectivity) measure of Daniel et al. (1997), Pirinsky (2000) shows that SPECTRUM institutions statistically significantly outperform their benchmarks by 83 basis points per year on an equally-weighted basis and by 67 basis points per year on a value-weighted basis. He also estimates three-factor Fama-French and four-factor Carhart (1997) regressions on both equally weighted and value weighted portfolios of institutions. First, he shows that alphas in the equal weighted regressions are systematically higher than alphas in the value-weighted regressions: 1.02 percent equal

previously experiencing the largest increase in institutional ownership outperforms the decile of firms previously experiencing the largest decrease in institutional ownership by 5.43 percent. However, the methodology of Nofisnger and Sias cannot be directly compared with the above methodology that uses levels of institutional ownership to evaluate the performance of the average institution, or with our methodology that uses levels of institutional ownership to evaluate the performance of the aggregate institutional portfolio.

weighted vs. 0.79 percent value weighted in the three-factor regression and 0.68 percent equal weighted vs. 0.46 percent value weighted in the four-factor regressions. Another thing to notice is that alphas in the four-factor regressions are lower than corresponding CS measures, which is most likely driven by the fact that institutions tend to hold stocks with betas above one and characteristic-based approach does not take this fact into account.

Our methodology is different from the Pirinsky (2000) value-weighted factor regressions along two dimensions. First, our measure of institutional performance is different. Pirinsky uses the return on the institutional portfolio minus the risk-free rate in his factor regressions. Thus, he implicitly evaluates the performance of the institutional portfolio over the Fama-French market. In general, institutional investment decision can be separated in two steps. First, institutions need to decide which stocks they are going to follow in their investment strategy. Second, they need to determine how they are going to invest into the stocks they follow. Our paper only considers the stocks that were used by at least one SPECTRUM institution during the studied sample period. We call this set of stocks “the institutional universe.” The previous section of the paper studied how institutional investment performance is different from the individual performance in the stocks within the institutional universe. In addition to this, aggregate institutional portfolio can outperform the Fama-French market because of the difference between the return on the institutional universe and the Fama-French market return. Taking all these effects into account, we get

$$R_t^{Inst} - R_t^{FF Mkt} = (1 - w_{t-1}^{Inst})(R_t^{Inst} - R_t^{Ind}) + (R_t^{Inst Univ} - R_t^{FF Mkt}), \quad (1)$$

where w_t^{Inst} is the share of the total market capitalization of stocks within the institutional universe held by institutions at the end of period t . Since the share of the institutional universe controlled by SPECTRUM institutions steadily increased from 37% in the first quarter of 1982 to 56% in the last quarter of 1996 (and to 60% in the last quarter of 2001), Pirinsky’s factor regressions implicitly put more weight on the earlier years than our regressions do. In addition, our methodology allows us to separate relative institutional performance within the institutional universe of stocks from the difference in returns on the institutional universe portfolio and the commonly used market portfolio.

Another difference between Pirinsky (2000) factor regressions and our methodology is that we take into account institutional preferences for stocks with high accounting profitability (high ROE). Column one of Table 5, Panel B, essentially reproduces the results of Pirinsky (2000) value-weighted three-factor regressions. The alpha of this regression is 74 basis points per year and highly statistically significant with t-statistic 2.28.¹⁶ However, after we include the ROE factor in the regression, the alpha of the institutional portfolio drops to 7 basis points per year and becomes not statistically significant with t-statistic 0.20.

The difference between the return on the institutional universe and the Fama-French market return, $R_t^{Inst Univ} - R_t^{FF Mkt}$, is also an important part of the excess return on the institutional portfolio in (1). Panel A of Table 5 compares the institutional universe factor, which is the return on the institutional universe minus the risk-free rate, with the Fama-French RMRF. The correlation between the two factors is 99.9 percent, but the institutional universe portfolio outperforms the Fama-French market portfolio by 47 basis points per year during 1982-1996 and has a slightly higher annualized Sharpe ratio. Columns three and four of Table 5, Panel B, show the factor regressions with the institutional universe factor instead of the Fama-French RMRF. In the three factor model alpha drops by 34 basis points, which means that more than 70 percent of the difference between the institutional universe portfolio and the Fama-French market portfolio cannot be explained by size and book-to-market factors. Interestingly enough, column four of Table 5, Panel B, shows that after the ROE factor is included in the regression, alpha drops to 18 basis points and, similar to the regression with Fama-French market factor (column two of Table 5, Panel B), becomes not statistically significant. This suggests that more than 70 percent of the difference between returns on the institutional universe portfolio and the Fama-French market portfolio that cannot be explained by Fama-French size and book-to-market factors, can be explained by the ROE factor. Panel C of Table 5 directly confirms these findings: column one shows that 32 basis points per year of the

¹⁶ The value of the alpha is almost identical to Pirinsky's alpha of 79 basis points in the same regression setup.

difference between the returns on the institutional universe and Fama-French market portfolios cannot be explained by the three Fama-French factors, with alpha being statistically significant at 10 percent confidence level. When we introduce the ROE factor into the regression (column 2 of Table 5, Panel C), the alpha becomes negative and not statistically significant. Therefore, our results suggest that institutions prefer stocks with high accounting profitability not only when choosing which stocks to overweight in their portfolios but also in choosing which stocks to follow.

3. Fraction of Institutional Ownership in Cross-Sectional Models of Expected Returns

In Part 2 we showed that outperformance of the aggregate institutional portfolio found in the previous literature can be explained by institutional preferences for stocks with high accounting profitability (high ROE). At the same time a number of recent studies have introduced the fraction of institutional ownership in a model of expected firm-level stock returns. Cross-sectional regressions in Brennan, Chordia, and Subrahmanyam (1997), Gompers and Metrick (2001), Cohen et al. (2002), and Bennett, Sias, and Starks (2002) suggest that, controlling for other predictors, higher institutional ownership predicts higher returns on individual stocks at quarterly and annual horizons. This part of the paper reconciles this cross-sectional evidence with our portfolio results. We show that the fraction of institutional ownership is neither a statistically nor an economically significant predictor of firm-level stock returns once we take into account book-to-market, return on equity, and size.

3.1 Regression methodology

We follow Vuolteenaho (2002) and Cohen et al. (2002) and estimate a set of cross-sectional regressions as one system under the constraint that the variable coefficients are constant over time. If return and the fraction on institutional ownership regressions are estimated simultaneously, this system can be represented as VAR. The intercepts are allowed to vary over time, which is identical to introduction of time dummy variables

into the VAR. Although the coefficient estimates obtained through simultaneously estimating the system of cross-sectional regressions are similar to the estimates obtained with Fama-MacBeth (Fama and MacBeth, 1973) methodology,¹⁷ we prefer the former because it provides a stronger tool for evaluating the economic importance of IO for expected returns. While Fama-MacBeth regressions yield a one period forecast, VAR permits to study the dynamics of return-institutional ownership system at different horizons. If the fraction on institutional ownership (IO) is relevant for expected return, the arrival of new information about IO should make us revise the forecast of returns in every future period up to infinity. In particular, aggregation of changes in expected returns into infinity yields an expected permanent effect of IO on stock price. Revisions in the forecasts of future returns provide an intuitive way to judge whether IO is an economically important determinant of future returns. Appendix C provides details of the VAR estimation, statistical inference, and calculation of permanent effect of IO on stock prices.

As a model for conditional expected return, we consider a parsimonious VAR specification that uses simple stock return and the fraction of institutional ownership (IO) as state variables and log book-to-market and log ROE as exogenous explanatory variables.¹⁸ Lagged returns capture medium-term momentum (Jegadeesh and Titman, 1993, 2001). Book-to-market captures the “book-to-market anomaly” (Rosenberg et al. (1985), Fama and French (1992), and others) when high book-to-market-equity firms earn higher average stock return than low book-to-market-equity firms. Controlling for other characteristics, firms with higher accounting profitability have earned higher average stock returns (Haugen and Baker, 1996). Following Haugen and Baker (1996) and Cohen et al. (2002), we use ROE computed according to generally accepted accounting principles (GAAP). Our main variable of interest is the fraction of institutional ownership, IO. If institutions as a group have superior stock picking ability,

¹⁷ Results of Fama-MacBeth regressions are available from the authors upon request.

¹⁸ Although by definition VAR is a system consisting of state variables only, in the absence of a better term for the dynamic panel estimation with exogenous variables, we call our system VAR. We find this term appropriate because the estimation, statistical inference, and impulse response function analysis closely resemble the techniques used for VAR and require only a slight modification for inclusion of exogenous variables.

the level of IO should provide additional forecasting power for future stock returns in our VAR model. In particular, cross-sectional regressions in Brennan, Chordia, and Subrahmanyam (1997), Gompers and Metrick (2001), Cohen et al. (2002), and Bennett, Sias, and Starks (2002) suggest that, on individual stock level, higher institutional ownership predicts higher returns.

We work with quarterly return and IO data to capture both short-term and long-term interaction between the two series. However, book-to-market and ROE are available only at the annual frequency. Our way around this problem is to hold these accounting variables constant for the whole year after they become available and introduce them as exogenous variables in the VAR. This way they enter as variables predicting higher or lower level of return and are not directly incorporated in return-IO dynamics. Our treatment of accounting variables is similar to Brennan, Chordia, and Subrahmanyam (1997), who consider monthly returns and book-to-market from the previous calendar year, and Gompers and Metrick (2001), who consider quarterly returns and book-to-market calculated from the previous calendar year. While Brennan, Chordia, and Subrahmanyam (1997) and Gompers and Metrick (2001) consider only return regressions, our VAR approach allows us to study the joint dynamics of returns and IO, which is similar to Cohen et al. (2002). However, Cohen et al. (2002) work with annual horizons and, as a result, they include book-to-market and ROE in VAR as state variables.

We work with two alternative specifications of the VAR system – VAR(1) and VAR(4). VAR(4) captures the return-IO dynamics at longer horizons. In addition, our specification can capture potential seasonality in the return-IO dynamics and the return momentum effect observed for lags two to four. The problem with interpreting the estimated coefficients in VAR(4) is that the lags of IO are highly correlated and, as a result, effects of individual lags have little meaning. The inference should be done through techniques which capture the lags altogether, such as impulse response functions or Granger causality test. We find that estimation results produced by VAR(1) are qualitatively identical to VAR(4), and impulse response functions of the two

specifications are very similar. Therefore, we discuss results for VAR(1), which allow easier interpretation of the regression coefficients.

3.2 Regression results

Expected return regressions can use either simple or log returns. Cohen et al. (2002) use an equally-weighted VAR with log returns and find that IO is strongly statistically significant in predicting cross-sectional variation in the firm-level expected returns.¹⁹

When we estimate an equally-weighted VAR with log returns and IO as state variables and book-to-market and ROE as exogenous variables, we obtain the same result: In the first column of Table 6 lagged IO is strongly statistically and economically significant. We argue that an investor can arrive at misleading conclusions if she uses these results directly.²⁰ The cross-sectional return regression implicitly compares average log returns across size groups. However, smaller stocks have significantly higher cross-sectional variance of returns (see Table A.1, Panel A) and, as a result, significantly lower average log return while the average simple returns exhibit little variation across size groups. Indeed, under the assumptions that returns are cross-sectionally log-normal,

$$E[\log(R)] = \log(E[R]) - \frac{1}{2} \text{Var}[\log(R)]. \quad (2)$$

Together with the fact that the fraction of institutional ownership in small stocks is much lower than in medium and large capitalization stocks, the cross-sectional variation in log returns generates a high economic and statistical significance of lagged IO in the return regression. Virtually all this predictive power of IO is not dynamic but is driven by correlation between cross-sectional means of IO and cross-sectional variances of log returns across size groups. One way to control for across-size-group differences in cross-sectional variances is to introduce size dummy variables in the VAR. Column two of Table 6 shows that when size dummies are included in the return regression, lagged IO

¹⁹ Cohen et al. (2002) focus on how institutional investors react to news about future cash flows and positive and significant coefficient on lagged IO is not an important part of their analysis.

²⁰ If an investor tries to form an investment strategy based on the results from the return regression in column one Table 6, she might conclude that entering a long-short position in two stocks that have the same book-to-market and ROE but differ by 5.06% in the fraction of institutional ownership will earn 82.7 basis points annualized excess return in the subsequent quarter (20.7 basis points quarterly excess return). If she holds the position for one, two, or five years, she will on average earn 80.6, 77.6, or 69.2 basis points

becomes not statistically and economically significant. Figure 6 further illustrates this point. It presents average log returns, average log returns corrected for cross-sectional variance using formula (2), and log of average returns within our size groups.²¹ The diagram clearly shows that equation (2) gives a very close approximation for the log of average cross-sectional return in our sample. As opposed to average log returns, average simple returns do not increase dramatically as we move from the smallest stocks to the largest.

In sum, lagged IO is highly statistically and economically significant in log return regression not because it contains information about future returns beyond the commonly used predictors but because the average fraction of institutional ownership in small stocks is much lower than in large stocks and at the same time small stocks have much higher cross-sectional variance of returns than large stocks. In other words, lagged IO is highly statistically significant in the log-return regression merely because it explains across-size-group differences in cross-sectional variances of returns. If an investor wants to use the regression with log returns for trading purposes, she should correct the results for the differences in cross-sectional variances of returns across size groups. One way to do this is to use size dummy variables similar to column two of Table 6. Alternatively, she can use regressions with simple returns.

Brennan, Chordia, and Subrahmanyam (1997), Gompers and Metrick (2001), and Bennett, Sias, and Starks (2002) study cross-sectional regressions with simple returns and show that, controlling for other predictors, IO is a statistically significant predictor of firm-level stock returns. Columns one, three and five of Table 7 present the results for the equally-weighted VAR with simple returns and IO as state variables and book-to-market, ROE, and log market equity (Size) as explanatory variables. The first thing to notice is that, after controlling for book-to-market, ROE and Size, lagged IO is a

per year, respectively. The infinite horizon cumulative excess return is equal to 10.51% and is significant at the confidence level of 1 percent.

²¹ The numbers in Figure 2 are based on sample statistics from Table A.1, Panel A.

statistically significant predictor of returns in the first half of the sample.²² Although lagged IO is not statistically significant in predicting returns in the second part of the sample, the infinite horizon impulse response functions for returns in Panels B of Table 7 show that the economic significance of lagged IO in predicting returns is similar across the samples.²³ Thus, the cross-sectional regression evidence from columns one, three and five of Table 7 might suggest that fraction of institutional ownership is a significant predictor of firm-level stock returns even after controlling for book-to-market, return on equity, and size. This however would contradict our findings in Part 2 where excess return on the aggregate institutional portfolio relative to the rest of the market can be fully explained by the three Fama-French factors and the ROE factor.

We argue that the marginal statistical significance of lagged IO in predicting firm-level stock returns in columns one, three and five of Table 7 is mostly driven by the regression setup. Regressions in the odd columns of Table 7 assume that a one percent increase in book-to-market, ROE, or market equity has the same effect on expected return regardless of the firm characteristics. This is clearly a restrictive assumption that might create biases in the IO coefficient. There is a particular reason to expect this bias to be especially strong with respect to size. As Table A.1, Panel B, shows, IO is very highly positively correlated with size, with an equally-weighted correlation of 71 percent across the whole sample. At the same time correlations between IO and other variables are much less spectacular: Correlation between IO and book-to-market is negative 7 percent, while correlation between IO and ROE is 20 percent. If a one percent increase in firm market capitalization has a more negative impact on the expected return among small stocks than among large stocks, the IO coefficient will have a strong positive bias. Columns two, four and six of Table 7 where we allow the size coefficient to vary across different size groups, show that this is indeed the case: One percent increase in market

²² Brennan, Chordia, and Subrahmanyam (1997) consider 1977-1989 period, Gompers and Metrick (2001) consider 1980-1996 period, and Bennett, Sias, and Starks (2002) consider 1983-1997 period. Our results suggest that statistical significance of lagged IO in this regression setup is mostly driven by the subsample of the eighties and early nineties and does not generalize to the later years.

²³ Although our VAR methodology is similar in spirit to Fama-MacBeth cross-sectional regressions used by Brennan, Chordia, and Subrahmanyam (1997), Gompers and Metrick (2001), and Bennett, Sias, and Starks (2002), it gives us an additional benefit of using impulse response functions to assess economic significance of lagged IO in predicting returns.

capitalization has three to four times bigger influence on expected return among the smallest stocks than among the largest stocks. After we allow the regression setup to take this effect into account, lagged IO becomes not statistically significant in both halves of the sample, and its economic significance drops by 40 percent in either half of the sample and by more than 50 percent in the whole sample.

4. Conclusion

Prior research suggests that institutions as a group are better investors than the rest of the market. In particular, institutions outperform their benchmarks, before any expenses are deducted, and the fraction of institutional ownership positively predicts firm-level stock returns in cross-sectional regressions. This paper looks for the sources of superior institutional performance using both portfolio and cross-sectional regression methodology.

We find that although institutions as a group outperform their Fama-French benchmarks, this is fully explained by institutional preferences for stocks with high accounting profitability (high ROE). To take these preferences into account, we construct an ROE factor portfolio that is long high ROE stocks and short low ROE stocks. Together with the three Fama-French factors, the ROE factor fully explains the positive excess return on the aggregate institutional portfolio relative to the rest of the market.

We also show that most of the explanatory power of the ROE factor comes from the fact that institutions as a group avoid small stocks with low ROE and overweight large stocks with high ROE. Our results also suggest a shift in institutional investment strategy around the year 2000. Namely, between 1982 and 1998 institutions overweighted large stocks with low ROE relative to individuals, which was detrimental to the performance of their portfolio. This effect is much less pronounced once we include the years 1999 to 2001, which partially explains high relative performance of the institutional portfolio in the year 2000.

Finally, we reconcile our portfolio findings with the cross-sectional regression literature by showing that in cross-sectional regressions of expected returns, the fraction of institutional ownership is neither a statistically nor an economically significant predictor of firm-level stock returns once we take into account book-to-market, return on equity (ROE), size, and biases caused by high positive correlation between the fraction of institutional ownership and stock size.

Our findings concerning institutional preferences for stocks with high accounting profitability are similar in spirit to Cohen et al. (2002) results that institutions buy shares from (sell shares to) individuals in response to positive (negative) cash-flow news exploiting the underreaction phenomenon. However, we are not trying to answer the question whether institutions fully exploit this effect. Rather, we show that institutional preferences for profitable stocks, together with their preferences for size and book-to-market, can fully explain the return difference between institutional and individual portfolios.

To summarize, our results suggest that future empirical studies investigating institutional investment performance, or studying institutional trading behavior, should take into account institutional preferences towards accounting profitability and their interaction with the preferences for size and book-to-market.

Appendix A. Data

A.1 CRSP-COMPUSTAT data

The return, number of shares outstanding, and price data come from the Center for Research in Securities Prices (CRSP). The initial sample consists of all NYSE, AMEX, and NASDAQ stocks that appeared in CRSP monthly files from January 1980 to December 2001. For these stocks, the accounting ratios are computed using COMPUSTAT accounting data items, when available.

In order to be included in our sample, a security must satisfy the following criteria. First, it must be defined as a common stock by CRSP (share codes 10 and 11 by CRSP classification). Second, to be included in period t , the stock must have at least 12 months of CRSP return data. When accounting data is used, the stock is included in period t if it has at least eight quarters of accounting data on COMPUSTAT.

For quarterly returns, we use cumulative returns for the three month during the quarter. If a return in a given month is missing, it is assumed to be zero. Delisting returns are included when available on CRSP. If a delisting return is missing, it is substituted with negative 30% if the delisting is performance-related and with zero otherwise. The delisting return of -30% is used by Vuolteenaho (2002) and Cohen et al. (2002). Similar to them, this assumption is not important for our results.

In constructing accounting variables, we closely follow the methodology of Cohen et al. (2002). Market equity is calculated as a product of the number of shares outstanding and share price. For book equity, we use COMPUSTAT data item 60. If it is unavailable, we use item 235. Also, if short- and/or long-term deferred taxes are available (data items 35 and 71), we add them to book equity. If both data items 60 and 235 are unavailable, we proxy book equity by the last period book equity plus earnings (data item 172) minus dividends (data item 21). If neither earnings nor book equity is available, we assume that the book-to-market ratio has not changed from the last year and compute the book equity proxy from the last period's book-to-market and this period market equity. We treat

negative or zero book equity values as missing. Also, if the calculated book-to-market is more than 100 or less than 1/100, we drop it and assume that the observation is missing.

GAAP ROE is the earnings over the last period book equity, measure according to the US Generally Accepted Accounting Principles. We use the COMPUSTAT data item 172, earning available for common. When it is missing, earnings are computed as the change in book equity plus dividends (item 21). Also, we do not allow the firm to lose more than its book equity. That is, we define ROE as the maximum of the calculated ROE and -100%. However, when log ROE is used in the analysis, we set all ROE values below -98% equal to -98%.

A.2 SPECTRUM data

Our data on institutional ownership comes from The Thomson Financial Ownership database on 13F Institutional Holdings (CDA/Spectrum s34). A 1978 amendment to the Securities and Exchange Act of 1934 required all institutions with greater than \$100 million of securities under discretionary management to report their holdings to the SEC. Holdings are reported quarterly on the SEC's form 13F, where all common-stock positions greater than 10,000 shares or \$200,000 must be disclosed. These reports are available in electronic form back to 1980 from CDA/Spectrum, a firm hired by the SEC to process the 13F filings. Our sample includes the quarterly reports from the first quarter of 1980 through the fourth quarter of 2001 (88 quarters total). Throughout this paper, we use the term "institution" as a synonym for "an institution that files the SEC form 13F."

The SPECTRUM 13F holdings data for each institution contains date, CUSIP number, company name, ticker symbol, shares held at the end of the quarter, and net change in shares since prior report. All dates correspond to the ends of quarters (March 31, June 30, September 30, or December 31). First, we take the list of CRSP permanent numbers (PERMNOs) for stocks that appear in CRSP monthly files between January 1980 and December 2001 (20,525 stocks in total). For each CRSP PERMNO we record all corresponding CUSIPs. We download all the SPECTRUM data available for those

CUSIPs. Our initial SPECTRUM sample consists of 19,323 stocks (identified by CRSP PERMNO) and 3,239 institutions (identified by SPECTRUM MGRNO), for whom we have 22,901,934 records total.

Before using SPECTRUM data in our analysis, we do a substantial amount of data cleaning. First, we identify all the inconsistent records. A record is considered to be inconsistent if the number of shares held by an institution in a particular stock at the end of quarter $t-1$ is not equal to the number of shares held at the end of quarter t minus the net change in shares since prior quarter (reported at the end of quarter t):

$$holdings(t-1) \neq holdings(t) - change(t) . \quad (B1)$$

Out of 22,901,934 records, 1,282,347 (5.6%) are inconsistent according to the above definition. 260,343 inconsistencies can be explained by rounding to a 1,000 shares in June 1988, 655,754 records are consistent with stock splits as reported by CRSP; 54,681 records are consistent with late filings of up to one quarter;²⁴ 88,511 inconsistent records can be explained by rounding errors other than in June 1988. As a result, we are left with 223,058 inconsistent records (0.97% of the initial sample) that we can not explain. For these records, we assume that the holdings data is correct, i.e. the net change data is incorrect.

The second step in our data cleaning process is to fill in the report gaps, or missing records. For each institution and each stock, missing records are defined as missing quarters in a time series of holdings. For the whole sample, we have 1,743,861 missing records, out of which 842,934 correspond to only one quarter missing in an otherwise continuous time series of institutional holdings in a particular stock. Our general rule is that, if a stock has return on CRSP but does not have holdings on SPECTRUM in a given quarter, we assume that holdings are 0. However, some of the missing records are inconsistent with this assumption because holdings at the end of quarter t are above the reported net change from the previous quarter:

²⁴ When an institution reports later than it was supposed to report, CDA/Spectrum applies the split ratios up to the date when the data was recorded. For example, if an institution reports 100,000 shares held and 50,000 shares net change since previous quarter in XYZ stock as of March 31, 1997, but CDA/Spectrum

$$holdings(t) > change(t). \quad (B2)$$

In this case we fill in the holdings for the end of quarter $t-1$ as

$$holdings(t-1) = \frac{holdings(t) - change(t)}{SplitRatio_{CRSP}(t-1, t)}. \quad (B3)$$

713,695 missing records (3.1% of the initial sample) are filled, with 572,917 records filling one missing quarter between two reported quarters.²⁵

Third, we apply the same methodology to add the missing first records. For a particular institution and stock, the first record is considered to be missing if we have reported holdings above reported net change for the first available quarter t , the same as in (B2). In this case we use formula (B3) to fill in the missing first record. 574,546 first records are added.²⁶ As a result of the whole data cleaning process, we have 24,190,175 records for 19,323 CRSP PERMNOs and 3,239 SPECTRUM MGRNOs.

To get aggregate institutional holdings for each security at the end of each quarter, we sum up the shares held by all institutions in the sample for each PERMNO and date. Firms with no SPECTRUM data in a particular quarter are assumed to have zero institutional holdings. The fraction of institutional ownership is calculated as a ratio of aggregate institutional holdings to the total number of shares outstanding for that security at the end the quarter as reported by CRSP. If the aggregate institutional holdings in a particular quarter calculated from SPECTRUM data are greater than the number of shares outstanding as reported by CRSP at the end of this quarter, we assume that the fraction of institutional ownership is not available for this stock in this quarter.

Out of the 19,323 stocks for which SPECTRUM data is available, we use only 16,876 stocks that are common stocks by CRSP classification. In our cross-sectional regressions, we drop stocks with less than 8 quarters of SPECTRUM data. Our final

records this data in June after 2:1 stock split has happened in May, the data will be recorded as 200,000 shares held with 100,000 net change from the previous quarter.

²⁵ We have 77 missing records with $holdings(t) < change(t)$ that we don't fill in because of the obvious data inconsistency.

²⁶ We have 57 reported first records with $holdings(t) < change(t)$. In this case we don't add the first record because of the inconsistency in the data.

sample covers the 20 year period from January 1982 to December 2001 (80 quarters) and consists of 311,209 firm-quarters.

A.3 Definition of robust size groups

Fama-French NYSE market capitalization breakpoints available from Kenneth French website are used to allocate stocks into size quintiles. Since the breakpoints are taken with respect to the NYSE universe which consists of relatively large stocks, the smallest size quintile, Quintile 1, includes more than half of the stocks in CRSP database. We split Quintile 1 into two groups: Quintile 1 Small, including all stocks with market capitalization below the 5th percentile of the NYSE by market cap, and Quintile 1 Large, including all stocks with market capitalization above the 5th percentile. An additional reason for separating Quintile 1 into Quintile 1 Small and Quintile 1 Large is that all summary statistics, except for simple returns, are significantly different for these groups (see Table A.1).

In the end of every month within a quarter stocks are assigned to the six size groups (five size quintiles with the smallest quintile split in two by the 5th percentile breakpoint). Stocks which stayed in the same size group in the end of all three months within the quarter are assigned to this size group. Stocks which changed size group assignment during the quarter are not considered in the next quarter. Therefore, size group definitions are “robust” with respect to changes in size group during the quarter.

A.4 Descriptive statistics

Table A.1 provides descriptive statistics for the data series used in this paper. Two alternative methods are used to calculate the sample moments. The first method applies equal weights to data points in every cross-section; the second method applies weights proportional to the market capitalization to data points in the cross-sections. Both methods then assign equal weights to all cross-sections.

Appendix B. Relative Performance of Institutional Portfolio

B.1 Construction of portfolios

In the end of every quarter from 1981, Q4, to 2001, Q3, institutional portfolio weights are calculated as a ratio of institutional dollar holdings in a stock to the total value of institutional holdings aggregated across all stocks. To calculate individual portfolio weights only the stocks in the institutional universe – stocks matched in SPECTRUM database – are used. For every stock in the institutional universe individual holdings are defined as a difference between the stock's market capitalization and institutional holdings in the stock. In the end of every quarter individual weights are then calculated as a ratio of individual dollar holdings in a given stock to the aggregate individual dollar holdings. Both institutional and individual weights are held constant during the quarter. Returns on institutional and individual portfolios for every month in a given quarter are calculated using the institutional and individual portfolio weights in the end of the previous quarter.

B.2 Variance of prediction error

The prediction error is the deviation of the actual value realized at $T + s$ from the value forecasted by the model: $y_{T+s} - \hat{y}_{T+s}$. Expectation of the prediction error is given by

$$E(y_{T+s} - \hat{y}_{T+s}) = E(x'_{T+s}\beta + \varepsilon_{T+s} - x'_{T+s}\hat{\beta}) = x'_{T+s}E(\beta - \hat{\beta}) + E(\varepsilon_{T+s}) = 0$$

Variance of the prediction error is

$$\begin{aligned} Var(y_{T+s} - \hat{y}_{T+s}) &= E\left((y_{T+s} - \hat{y}_{T+s})(y_{T+s} - \hat{y}_{T+s})'\right) = \\ &= E\left(\left([y_{T+s} - E(y_{T+s})] - [\hat{y}_{T+s} - E(\hat{y}_{T+s})]\right)\left([y_{T+s} - E(y_{T+s})] - [\hat{y}_{T+s} - E(\hat{y}_{T+s})]\right)'\right) = \\ &= E\left(\left(\varepsilon_{T+s} - x'_{T+s}(X'X)^{-1}X'\varepsilon\right)\left(\varepsilon_{T+s} - x'_{T+s}(X'X)^{-1}X'\varepsilon\right)'\right) \end{aligned}$$

Under the assumption that the error terms are homoscedastic and are not serially correlated this becomes

$$\begin{aligned} Var(y_{T+s} - \hat{y}_{T+s}) &= E(\varepsilon_{T+s}^2) + E\left(x'_{T+s}(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}x_{T+s}\right) = \\ &= \sigma^2\left(1 + x'_{T+s}(X'X)^{-1}x_{T+s}\right) \end{aligned}$$

B.3 Regressions with the disaggregated ROE factor

The factor regression for the excess return on institutional portfolio over the rest of the market is

$$R_t^{Inst} - R_t^{Ind} = \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 ROE_t + \varepsilon_t \quad (B1)$$

where $RMRF$, SMB , and HML are the Fama-French market, size, and value factors and ROE is the factor based on ROE that we propose in this paper. The ROE factor contains twelve portfolios – a low ROE and a high ROE portfolio for each of the six size groups. Definition of the size groups is provided in Appendix A.3. The ROE factor in (B1) is calculated under the constraint that the weights of the six high ROE portfolios are equal 1/6 and the weights of the six low ROE portfolios are equal negative 1/6:

$$ROE_t = \frac{1}{6} \sum_{j=1}^6 R_{j,t}^{HighROE} - \frac{1}{6} \sum_{j=1}^6 R_{j,t}^{LowROE}$$

This constraint can be relaxed in a number of ways. First, all twelve coefficients of the ROE portfolio components can be unconstrained, which yields a regression of the following form

$$\begin{aligned} R_t^{Inst} - R_t^{Ind} = & \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t \\ & + \sum_{j=1}^6 \gamma_j R_{j,t}^{HighROE} + \sum_{j=1}^6 \delta_j R_{j,t}^{LowROE} + \varepsilon_t \end{aligned} \quad (B2)$$

Alternatively, the six excess return ROE portfolios can be calculated for each size group, but the coefficients of the size groups allowed to vary

$$\begin{aligned} R_t^{Inst} - R_t^{Ind} = & \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t \\ & + \sum_{j=1}^6 \gamma_j (R_{j,t}^{HighROE} - R_{j,t}^{LowROE}) + \varepsilon_t \end{aligned} \quad (B3)$$

Finally, a high ROE and a low ROE portfolios can be defined through equal weighting of size portfolios, but the coefficients of the aggregate high ROE and low ROE portfolios allowed to vary

$$\begin{aligned} R_t^{Inst} - R_t^{Ind} = & \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t \\ & + \gamma \frac{1}{6} \sum_{j=1}^6 R_{j,t}^{HighROE} + \delta \frac{1}{6} \sum_{j=1}^6 R_{j,t}^{LowROE} + \varepsilon_t \end{aligned} \quad (B4)$$

To assess the economic significance of the regression coefficients in (B1)-(B4), the coefficients are scaled by the ratio of standard deviation of the respective explanatory variable to standard deviation of the dependent variable. Namely, if the regression is defined as

$$y_t = \alpha + \sum_{j=1}^K \beta_j x_{j,t} + \varepsilon_t ,$$

we report $\beta_j \frac{\sigma_j}{\sigma_y}$, where σ_j and σ_y are the time-series standard deviations of $x_{j,t}$ and y_t , respectively,

Appendix C. Regression Methodology

C.1 Regression setup

A system of cross-sectional regressions under the constraint that the coefficients are equal across time can be represented by VAR. This approach was previously used by Vuolteenaho (2002) and Cohen et al. (2002). Estimation of the set of cross-sectional regressions as one system allows to fully exploit the time-series and cross-sectional dynamics of the data. The return and institutional ownership processes are specified as follows

$$\begin{pmatrix} r_{it} \\ IO_{it} \end{pmatrix} = \alpha_t + \sum_{j=1}^p B_j \begin{pmatrix} r_{i,t-j} \\ IO_{i,t-j} \end{pmatrix} + \Gamma x_{i,t-1} + \varepsilon_{it} ,$$

where r_{it} is log return of stock i during quarter t , IO_{it} is the fraction of institutional ownership in stock i in the end of quarter t , and $x_{i,t-1}$ is a vector of predictive variables known before period t . IO_{it} is calculated at the end of quarter t as the number of shares held by institutions in stock i divided by the total number of shares outstanding. The intercepts of the regressions, α_t , are allowed to vary over time. Since time dummies are introduced, the results are cross-sectional and similar in spirit to the Fama-MacBeth (1973) procedure.

Interpretation of the return regression is as follows. Given past returns and predictors of returns we ask whether a higher fraction of institutional ownership predicts a higher return in a cross-section. Time fixed effects stand for the time-varying average cross-sectional return in a given quarter. Lagged returns $r_{i,t-1}, \dots, r_{i,t-p}$ and the predictive variables $x_{i,t-1}$ capture the expected deviations of firm-level returns from the average return in the cross section. Finally, $IO_{i,t-1}, \dots, IO_{i,t-p}$ capture expected deviations of firm-level returns due to higher or lower levels of institutional ownership. If institutions possess information about expected returns beyond the information incorporated into the model, one would expect the coefficients on past IO to be positive and significant. In the institutional ownership regression, the fixed effects stand for the average fraction of institutional ownership in a given period. Past returns capture the response of IO to past performance, predictive variables capture response of IO to predictors of firm-level returns, and past levels of IO controls for persistence in IO.

C.2 *Infinite-horizon return*

VAR system above describes the joint dynamics of returns and IO. The estimated coefficients can be used to calculate impulse response functions showing how shocks to IO affect the future path of returns. Cumulating the impulse response function into infinity allows to assess the permanent effect of a shock to IO on stock price. The example below illustrates the calculation of this permanent price effect for a VAR with one lag.²⁷ Suppose in period t IO of stock i was ξ percent above the expected level of IO predicted by the VAR model, while the return was equal to the conditional mean. As a result of this shock to IO, in period $t + 1$ the return will be $e1'B \begin{pmatrix} 0 \\ \xi \end{pmatrix}$ percent in excess of the return expected before the shock, where $e1' = (1 \ 0)$. Similarly, the excess return in period $t + 2$ will be $e1'B^2 \begin{pmatrix} 0 \\ \xi \end{pmatrix}$ percent, the cumulative excess return between period t and

²⁷ Since VAR of any order can be transformed into a VAR with one lag, these results are extendable to higher order VARs.

$t + 2$ will be $e1'(0 + B + B^2) \begin{pmatrix} 0 \\ \xi \end{pmatrix}$, etc. If we calculate the cumulative excess return for the infinite horizon, we obtain the permanent effect of the change in IO on the stock price, $e1'(I - B) \begin{pmatrix} 0 \\ \xi \end{pmatrix}$. This statistic becomes especially convenient for VARs with more than one lag, since in that case it is hard to interpret individual VAR coefficients.

The following part provides a detailed explanation of the methodology used to estimate the VAR parameter and their standard errors.

C.3 Estimation methodology

Consider a linear model

$$y_t = X_t b + \varepsilon_t \quad (C1)$$

where ε_t is a cross-sectional error and b is a $K \times 1$ vector of parameters. Under the assumption of exogeneity of the regressors, this generates a set of K orthogonality conditions for every cross-section:

$$E_t(X'_{it}(y_{it} - X_{it}b)) = 0 \quad (C2)$$

where the expectations are taken within cross-sections. That is, subscript t denotes expectation taken over data points $i = 1, 2, \dots, N_t$ in the cross-section of period t . Let $g_t(b)$ be a vector of sample moments corresponding to (C2), then

$$g_t(b) = \frac{1}{N_t} \sum_{i=1}^{N_t} (X'_{it}(y_{it} - X_{it}b))$$

Combining the system for T time periods, we obtain

$$g(b) = \begin{pmatrix} g_1(b) \\ g_2(b) \\ \dots \\ g_T(b) \end{pmatrix} = \begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} (X'_{i1}(y_{i1} - X_{i1}b)) \\ \frac{1}{N_2} \sum_{i=1}^{N_2} (X'_{i2}(y_{i2} - X_{i2}b)) \\ \dots \\ \frac{1}{N_T} \sum_{i=1}^{N_T} (X'_{iT}(y_{iT} - X_{iT}b)) \end{pmatrix} \quad (C3)$$

GMM prescribes to minimize an objective function

$$g(b)'Wg(b)$$

The set of first order conditions is

$$\frac{\partial g'(b)}{\partial b}Wg(b) = 0 \quad (C4)$$

Differentiating (C3) with respect to b yields a $K \times KT$ matrix

$$\frac{\partial g'(b)}{\partial b} = - \begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} X'_{i1} X_{i1} \\ \frac{1}{N_2} \sum_{i=1}^{N_2} X'_{i2} X_{i2} \\ \dots \\ \frac{1}{N_T} \sum_{i=1}^{N_T} X'_{iT} X_{iT} \end{pmatrix}' \quad (C5)$$

Substituting (C3) and (C5) into (C4), we obtain

$$\begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} X'_{i1} X_{i1} \\ \frac{1}{N_2} \sum_{i=1}^{N_2} X'_{i2} X_{i2} \\ \dots \\ \frac{1}{N_T} \sum_{i=1}^{N_T} X'_{iT} X_{iT} \end{pmatrix}' W \begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} (X'_{i1} (y_{i1} - X_{i1}b)) \\ \frac{1}{N_2} \sum_{i=1}^{N_2} (X'_{i2} (y_{i2} - X_{i2}b)) \\ \dots \\ \frac{1}{N_T} \sum_{i=1}^{N_T} (X'_{iT} (y_{iT} - X_{iT}b)) \end{pmatrix} = 0 \quad (C6)$$

where

$$\begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} X'_{i1} X_{i1} \\ \frac{1}{N_2} \sum_{i=1}^{N_2} X'_{i2} X_{i2} \\ \dots \\ \frac{1}{N_T} \sum_{i=1}^{N_T} X'_{iT} X_{iT} \end{pmatrix}' W$$

is a $K \times KT$ matrix of weights applied to the sample moments corresponding to the orthogonality conditions defined in (C2).

Define $u_t = y_t - X_t b$. Let

$$h_t(b) = X_t' u_t \quad (C7)$$

then the optimal weighting matrix under the assumption of time series and cross-sectional independence of the residuals is given by $W = S^{-1}$, where

$$S = E_t \left(E \left(h(b) h(b)' \right) \right) \quad (C8)$$

with the corresponding sample moments

$$S = \begin{pmatrix} \frac{1}{N_1^2} \sum_{i=1}^{N_1} E(u_{i1}^2 X_{i1}' X_{i1}) & 0 & \dots & 0 \\ 0 & \frac{1}{N_2^2} \sum_{i=1}^{N_2} E(u_{i2}^2 X_{i2}' X_{i2}) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{N_T^2} \sum_{i=1}^{N_T} E(u_{iT}^2 X_{iT}' X_{iT}) \end{pmatrix} \quad (C9)$$

Assuming further that the residuals are homoscedastic within cross-sections and over time, we can rewrite (C9) as

$$S = \begin{pmatrix} \sigma^2 \frac{1}{N_1^2} \sum_{i=1}^{N_1} E(X_{i1}' X_{i1}) & 0 & \dots & 0 \\ 0 & \sigma^2 \frac{1}{N_2^2} \sum_{i=1}^{N_2} E(X_{i2}' X_{i2}) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \frac{1}{N_T^2} \sum_{i=1}^{N_T} E(X_{iT}' X_{iT}) \end{pmatrix} \quad (C10)$$

Substituting S given by (C10) into (C6) and simplifying, we obtain a pooled regression OLS estimator

$$\tilde{b} = \left(\sum_{t=1}^T (X_t' X_t) \right)^{-1} \left(\sum_{t=1}^T (X_t' y_t) \right)$$

This OLS estimator is consistent under standard assumptions as is shown in Baltagi (2001, pp.159-162) and Wooldridge (2002, pp. 578-581).

The problem with cross-sectional return regressions is that it is virtually impossible to control for all common factors in returns and, therefore, regression residuals tend to be

cross-sectionally correlated. In this case weighing matrix S is not block diagonal. The GMM estimator can be obtained by iterative procedure.²⁸ As an alternative, Vuolteenaho (2002) proposes a simple and robust approach. He suggests to put equal weight on each cross-sectional orthogonality condition. This captures the idea of cross-sectional correlation – the amount of information does not increase proportionally to the number of data points in the cross-section. In particular, when we assign an equal weight to every cross-section, we implicitly assume that every cross-section contains the same amount of information. In this case (C6) reduces to

$$(\mathbf{1}'_T \otimes I_K) \begin{pmatrix} \frac{1}{N_1} \sum_{i=1}^{N_1} (X'_{i1} (y_{i1} - X_{i1} b)) \\ \frac{1}{N_2} \sum_{i=1}^{N_2} (X'_{i2} (y_{i2} - X_{i2} b)) \\ \dots \\ \frac{1}{N_T} \sum_{i=1}^{N_T} (X'_{iT} (y_{iT} - X_{iT} b)) \end{pmatrix} = 0 \quad (C11)$$

where $\mathbf{1}_T$ is a vector of ones of length T . Rewriting (C11) in matrix notation, we obtain:

$$(\mathbf{1}'_T \otimes I_K) \begin{pmatrix} \frac{1}{N_1} X'_1 (y_1 - X_1 b) \\ \frac{1}{N_2} X'_2 (y_2 - X_2 b) \\ \dots \\ \frac{1}{N_T} X'_T (y_T - X_T b) \end{pmatrix} = (\mathbf{1}'_T \otimes I_K) \begin{pmatrix} X'_1 \Omega_1 (y_1 - X_1 b) \\ X'_2 \Omega_2 (y_2 - X_2 b) \\ \dots \\ X'_T \Omega_T (y_T - X_T b) \end{pmatrix} \quad (C12)$$

$$= \sum_{t=1}^T X'_t \Omega_t (y_t - X_t b) = 0$$

where $\Omega_t = \frac{1}{N_t} I_{N_t}$. Or

$$X \Omega (y - Xb) = 0 \quad (C13)$$

²⁸ In principal, a one-iteration procedure already yields an asymptotically efficient estimator.

where

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_T \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{pmatrix}, \Omega = \text{diag} \begin{pmatrix} \Omega_1 \\ \Omega_2 \\ \dots \\ \Omega_T \end{pmatrix}$$

Expression (C12) yields the estimator

$$\hat{b} = (X' \Omega X)^{-1} (X' \Omega y) \quad (\text{C14})$$

Or

$$\hat{b} = \left(\sum_{t=1}^T \frac{1}{N_t} (X_t' X_t) \right)^{-1} \left(\sum_{t=1}^T \frac{1}{N_t} (X_t' y_t) \right).$$

We refer to this estimator as an equally-weighted estimator because every data point in a cross-section gets the same weight.

Now consider a fixed effects estimator with a general weighting matrix W

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\begin{pmatrix} D & X \end{pmatrix}' \Omega \begin{pmatrix} D & X \end{pmatrix} \right)^{-1} \begin{pmatrix} D & X \end{pmatrix}' \Omega y \quad (\text{C15})$$

Here $\hat{\alpha}$ is the estimator of the fixed effects, and $\hat{\beta}$ is the within estimator. We can rewrite (C15) as

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} D' \Omega D & D' \Omega X \\ X' \Omega D & X' \Omega X \end{pmatrix}^{-1} \begin{pmatrix} D' \Omega y \\ X' \Omega y \end{pmatrix} \quad (\text{C16})$$

Using the results for partitioned matrices, we can write

$$\begin{aligned} \hat{\beta} &= -F(X' \Omega D)(X' \Omega X)^{-1} D' \Omega y + F X' \Omega y \\ F &= (X' \Omega X - (X' \Omega D)(D' \Omega D)^{-1} (D' \Omega X))^{-1} \end{aligned} \quad (\text{C17})$$

Let the Cholesky factorization of Ω be $\Omega = P'P$. Then

$$\begin{aligned} \hat{\beta} &= F X' P' P y - F (X' P' P D) (X' \Omega X)^{-1} D' P' P y = \\ &= F (P X)' \left(I - P D (D' \Omega D)^{-1} (P D)' \right) P y = F (P X)' M_{\text{od}} P y \end{aligned} \quad (\text{C18})$$

where

$$M_{\text{od}} = I - P D (D' \Omega D)^{-1} (P D)' \quad (\text{C19})$$

Reexpressing F in a similar way, we obtain

$$\hat{\beta} = \left((PX)' M_{\text{od}} (PX) \right)^{-1} (PX)' M_{\text{od}} Py \quad (\text{C20})$$

Introducing transformed variables

$$\begin{aligned} \tilde{y} &= M_{\text{od}} Py = \left(I - PD(D'\Omega D)^{-1} (PD)' \right) Py = \\ &P \left(I - D(D'\Omega D)^{-1} D'W \right) y = PM_{\text{od}} y \\ \tilde{X} &= M_{\text{od}} PX = \left(I - PD(D'\Omega D)^{-1} (PD)' \right) PX = \\ &P \left(I - D(D'\Omega D)^{-1} D'\Omega \right) X = PM_{\text{od}} X \end{aligned} \quad (\text{C21})$$

we can rewrite (C20) as

$$\hat{\beta} = (\tilde{X}\tilde{X})^{-1} \tilde{X}\tilde{y}$$

Suppose

$$P = \begin{pmatrix} \frac{1}{\sqrt{N_1}} & \dots & 0 & \dots & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \frac{1}{\sqrt{N_1}} & \dots & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \frac{1}{\sqrt{N_T}} & \dots & \dots \\ 0 & \dots & 0 & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & \dots & 0 & \dots & \frac{1}{\sqrt{N_T}} \end{pmatrix} \quad (\text{C22})$$

where N_t is a number of observations in cross-section t .

Using the definition in (C21), it can be shown that the elements of the transformed variables are then

$$\tilde{y}_{it} = \frac{1}{\sqrt{N_t}} (y_{it} - \bar{y}_t), \quad \tilde{x}_{it} = \frac{1}{\sqrt{N_t}} (x_{it} - \bar{x}_t) \quad (\text{C23})$$

where \bar{x}_t and \bar{y}_t are cross-sectional means which include only those i s for which all elements in (y_{it}, x'_{it}) are observed.

For example, if y_t is a return, then $(y_{it} - \bar{y}_t)$ is an excess return over the equally weighted portfolio. Therefore, fixed effects estimator allows to switch from raw returns to excess returns. The same is true for other variables.

Consider the estimator in (C20):

$$\hat{\beta} = \left((PX)' M_{wd} (PX) \right)^{-1} (PX)' M_{wd} Py = \left((PM_{wd} X)' (PM_{wd} X) \right)^{-1} (PM_{wd} X)' PM_{wd} y$$

The conditional covariance matrix of $\hat{\beta}$ is then

$$\begin{aligned} Var(\hat{\beta} | X) &= E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' | X \right) = \\ & \left((PM_{wd} X)' (PM_{wd} X) \right)^{-1} (PM_{wd} X)' PM_{wd} E(\varepsilon \varepsilon' | X) M_{wd} P' (PM_{wd} X) \left((PM_{wd} X)' (PM_{wd} X) \right)^{-1} \end{aligned}$$

Assume that the residuals are not autocorrelated but can be cross-sectionally correlated. In this case matrix PM_{wd} is block-diagonal with blocks corresponding to cross-sections.

Therefore under the assumption of cross-sectional correlation and no autocorrelation, the covariance matrix can be estimated by

$$\hat{Var}(\hat{\beta} | X) = (\tilde{X}' \tilde{X})^{-1} \sum_{t=1}^T \left[\tilde{X}_t' \tilde{u}_t \tilde{u}_t' \tilde{X}_t \right] (\tilde{X}' \tilde{X})^{-1} \quad (C24)$$

where \tilde{X} are transformed variables and \tilde{u} are residuals obtained from the linear regressions with transformed variables. This form of robust coefficient covariance matrix was suggested by Arellano (1987), and it is consistent in T .

References

- Arellano, M., 1987, Computing robust standard errors for within-groups estimators, *Oxford Bulletin of Economics and Statistics* 49, 431-434.
- Baltagi, B.H., 2001, *Econometric Analysis of Panel Data*, 2nd Ed., John Wiley & Sons, Ltd.
- Bennett, J., R. Sias, and L. Starks, 2002, Greener Pastures and the Impact of Dynamic Institutional Preferences, working paper.
- Brennan, M.J., T. Chordia, and A. Subrahmanyam, 1997, Cross-Sectional Determinants of Expected Returns, FMRC working paper No.96-25, Owen Graduate School of Management, Vanderbilt University.
- Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press.
- Cohen R.B., P.A. Gompers, and T. Vuolteenaho, 2002, Who underreacts to cash-flow news? Evidence from Trading between Individuals and Institutions, *Journal of Financial Economics* 66, 409-462.
- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring Mutual Fund Performance with Characteristic-Based Benchmarks, *Journal of Finance* 52, 1035-1058.
- Del Guercio, D., 1996, The Distorting Effect of the Prudent-Man Laws on Institutional Equity Investments, *Journal of Financial Economics* 40, 31-62.
- Falkenstein, E.G., 1996, Preferences for Stock Characteristics As Revealed by Mutual Fund Portfolio Holdings, *Journal of Finance* 51, 111-135.
- Fama, E., and J. MacBeth, 1973, Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy* 81, 607-636.
- Fama, E.F., and K.R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427-465.
- Fama, E.F., and K.R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.
- Gompers, P.A. and A. Metrick, 2001, Institutional Ownership and Equity Prices, *Quarterly Journal of Economics* 116, 229-260.
- Grinblatt, M., and S. Titman, 1989, Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, *Journal of Business* 62, 393-416.
- Grinblatt, M., and S. Titman, 1993, Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns, *Journal of Business* 66, 47-68.

- Grinblatt, M., S. Titman, and R. Wermers, 1995, Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, *American Economic Review* 85, 1088-1105.
- Haugen, R., and N. Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401-439.
- Jegadeesh, N., and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65-91.
- Jegadeesh, N., and S. Titman, 2001, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations, *Journal of Finance* 56, 699-720.
- Lee, C.M.C., A. Shleifer, and R.H. Thaler, 1991, Investor Sentiment and the Closed-End Fund Puzzle, *Journal of Finance* 46, 75-109.
- Nofsinger, J.R., and R.W. Sias, 1999, Herding and Feedback Trading by Institutional and Individual Investors, *Journal of Finance* 54, 2263-2295.
- Odean, T., 1998, Are Investors Reluctant to Realize Their Losses?, *Journal of Finance* 53, 1775-1798.
- Odean, T., 1999, Do Investors Trade Too Much?, *American Economic Review* 89, 1279-1298.
- Odean, T. and B. Barber, 2001, Boys Will Be Boys: Gender, Overconfidence and Common Stock Investment, *Quarterly Journal of Economics* 116, 261-292.
- O'Hara, M., 1995, *Market Microstructure Theory*, Blackwell Publishers, Cambridge, Mass.
- Pirinsky, C., 2000, Are Financial Institutions Better Investors?, working paper, Fisher College of Business, The Ohio State University.
- Rosenberg, B., K. Reid, and R. Lanstein, 1985, Persuasive Evidence of Market Inefficiency, *Journal of Portfolio Management* 11, 9-17.
- Vuolteenaho, T., 2002, What Drives Firm-Level Stock Returns?, *Journal of Finance* 57, 233-264.
- Wermers, R., 2000, Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses, *Journal of Finance* 55, 1655-1695.
- Wooldridge, J.M., 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Table 1. Performance of institutional portfolio.

Panel A reports the annualized mean, standard deviation, and Sharpe ratio of excess return of the institutional portfolio over the rest of the market. All statistics are annualized. Appendix B.1 contains the details of portfolio calculation. Panel B measures performance of the institutional portfolio relative to three- and four-factor models. RMRF, SMB, and HML are Fama-French factors obtained from Kenneth French's website. ROE factor is the return on a size-boo-to-market balanced portfolio going long high return on equity stocks and short low return on equity stocks. The factor regression alphas are annualized. T-statistics are in square brackets. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

Sample period	01/1982 - 12/1998		01/1982 - 12/2001	
Number of months	204		240	
Panel A. Performance of institutional portfolio				
Annualized mean excess return	0.0087		0.0123	
Annualized standard deviation	0.0185		0.0207	
Annualized Sharpe ratio	0.4685		0.5944	
T-statistic	1.9315		2.6582	
Panel B. Performance of institutional portfolio controlling for factors				
Annualized alpha	0.0050	0.0000	0.0067 *	-0.0009
	[1.33]	[0.01]	[1.69]	[-0.21]
RMRF	0.0498 ***	0.0513 ***	0.0570 ***	0.0566 ***
	[7.16]	[7.28]	[6.93]	[7.05]
SMB	-0.0579 ***	-0.0464 **	-0.0471 ***	-0.0162
	[-4.23]	[-3.51]	[-3.24]	[-1.07]
HML	-0.0690 ***	-0.0668 ***	0.0106	-0.0157
	[-4.42]	[-4.3]	[0.52]	[-0.97]
ROE factor		0.0610 **		0.1217 ***
		[2.37]		[4.97]
R squared	39.13%	41.28%	20.67%	31.56%
R squared adjusted	38.21%	40.10%	19.66%	30.39%

Table 2. ROE factor portfolios and the Fama-French factors.

Panel A reports the annualized mean, standard deviation, and Sharpe ratio of the ROE factors – portfolios going long high return on equity stocks and short low return on equity stocks. ROE portfolios are size-book-to-market balanced. All statistics are annualized. T-statistics are obtained using standard errors corrected for the first-order serial correlation of the residuals. Panel B reports correlations between the three ROE factors and the Fama-French factors, RMRF, SMB, and HML, obtained from Kenneth French's website. Panel C measures performance of the ROE factors relative to the Fama-French three-factor model. Panel D measures performance of the ROE factors relative to the four-factor model including UMD, the momentum factor, in addition to the three Fama-French factors. UMD is obtained from Kenneth French's website. The factor regression alphas are annualized. T-statistics in square brackets are obtained using standard errors corrected for the first-order serial correlation of the residuals. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

Panel A. Performance of ROE factors						
Sample period	01/1982 - 12/1998			01/1982 - 12/2001		
Number of months	204			240		
	Equally-weighted	Value-weighted	ROE-weighted	Equally-weighted	Value-weighted	ROE-weighted
Annualized excess return	0.0758	0.0806	0.0840	0.0698	0.0742	0.0794
Annualized standard deviation	0.0473	0.0474	0.0567	0.0723	0.0732	0.0907
Annualized Sharpe ratio	1.6027	1.6997	1.4802	0.9654	1.0126	0.8754
T-statistic	5.5489	5.9467	5.1257	3.4281	3.5754	3.1547
Panel B. Correlation coefficients between ROE factor portfolios and Fama-French factors						
Sample period: 01/1982 - 12/1998						
	Equally-weighted	Value-weighted	ROE-weighted	RMRF	SMB	HML
Equally-weighted	1.000	0.984	0.937	-0.070	-0.301	0.067
Value-weighted	0.984	1.000	0.927	-0.102	-0.337	0.074
ROE-weighted	0.937	0.927	1.000	-0.061	-0.346	0.075
RMRF	-0.070	-0.102	-0.061	1.000	0.165	-0.523
SMB	-0.301	-0.337	-0.346	0.165	1.000	-0.284
HML	0.067	0.074	0.075	-0.523	-0.284	1.000
Sample period: 01/1982 - 12/2001						
	Equally-weighted	Value-weighted	ROE-weighted	RMRF	SMB	HML
Equally-weighted	1.000	0.993	0.966	-0.227	-0.551	0.523
Value-weighted	0.993	1.000	0.959	-0.250	-0.570	0.533
ROE-weighted	0.966	0.959	1.000	-0.234	-0.561	0.535
RMRF	-0.227	-0.250	-0.234	1.000	0.175	-0.541
SMB	-0.551	-0.570	-0.561	0.175	1.000	-0.477
HML	0.523	0.533	0.535	-0.541	-0.477	1.000

Table 2 (continued)

Panel C. Performance of ROE factors controlling for Fama-French factors						
Sample period	01/1982 - 12/1998			01/1982 - 12/2001		
Number of months	204			240		
	Equally-weighted	Value-weighted	ROE-weighted	Equally-weighted	Value-weighted	ROE-weighted
Annualized alpha	0.0744 *** [5.24]	0.0806 *** [5.72]	0.0807 *** [5.09]	0.0569 *** [3.25]	0.0627 *** [3.59]	0.0631 *** [3.14]
RMRF	-0.0132 [-0.44]	-0.0254 [-0.86]	-0.0085 [-0.27]	0.0157 [0.48]	0.0032 [0.1]	0.0186 [0.48]
SMB	-0.1684 *** [-4.67]	-0.1893 *** [-4.96]	-0.2328 *** [-5.68]	-0.2368 *** [-5.37]	-0.2536 *** [-5.62]	-0.3021 *** [-6.09]
HML	-0.0228 [-0.49]	-0.0358 [-0.77]	-0.0243 [-0.45]	0.2238 *** [3.32]	0.2167 *** [3.05]	0.2862 *** [3.42]
R squared	9.22%	11.91%	12.05%	39.22%	41.33%	40.81%
R squared adjusted	7.86%	10.58%	10.73%	38.45%	40.59%	40.06%
Panel D. Performance of ROE factors controlling for Fama-French factors and UMD						
Sample period	01/1982 - 12/1998			01/1982 - 12/2001		
Number of months	204			240		
	Equally-weighted	Value-weighted	ROE-weighted	Equally-weighted	Value-weighted	ROE-weighted
Annualized alpha	0.0655 *** [4.33]	0.0718 *** [4.77]	0.0679 *** [4.07]	0.0524 *** [3.03]	0.0588 *** [3.44]	0.0525 *** [2.59]
RMRF	-0.0141 [-0.44]	-0.0263 [-0.82]	-0.0098 [-0.30]	0.0195 [0.6]	0.0064 [0.2]	0.0273 [0.73]
SMB	-0.1526 *** [-4.18]	-0.1737 *** [-4.58]	-0.2101 *** [-5.19]	-0.2396 *** [-5.23]	-0.2561 *** [-5.53]	-0.3088 *** [-5.71]
HML	-0.0107 [-0.23]	-0.0238 [-0.51]	-0.0070 [-0.13]	0.2310 *** [3.5]	0.2229 *** [3.22]	0.3031 *** [3.76]
UMD	0.0868 ** [2.07]	0.0860 ** [2.00]	0.1245 *** [2.72]	0.0359 [1.15]	0.0305 [0.96]	0.0833 ** [2.07]
R squared	12.40%	15.00%	16.58%	39.73%	41.69%	42.53%
R squared adjusted	10.64%	13.29%	14.91%	38.70%	40.70%	41.56%

Table 3. Loadings on the components of the value-weighted ROE factor.

The table reports loadings of the excess return on the institutional portfolio over the rest of market on the components of the value-weighted ROE factor. Variables in the regression are standardized in order for the magnitude of the coefficients to be comparable. The regressions include the three Fama-French factors and a particular decomposition of the value-weighted ROE factor. Appendix B.3 provides details on the regression estimation. T-statistics are in square brackets. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

	Quintile 1 Small	Quintile 1 Large	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All Stocks
<u>Sample period: 01/1982 - 12/1998</u>							
High ROE	-0.1746 [-1.00]	0.0807 [0.30]	0.3152 [1.38]	0.6546 *** [2.76]	0.0105 [0.05]	0.8739 *** [3.76]	0.5644 [1.37]
Low ROE	0.2477 * [1.70]	-0.9640 *** [-3.89]	0.1546 [0.77]	0.5263 *** [2.62]	0.4502 ** [2.05]	0.6538 ** [2.57]	-0.5835 * [-1.74]
Net	-0.0021 [-0.03]	0.2659 *** [3.69]	0.0971 [1.50]	-0.0798 [-1.29]	-0.0858 [-1.24]	0.0348 [0.49]	0.1564 ** [2.37]
<u>Sample period: 01/1982 - 12/2001</u>							
High ROE	-0.0824 [-0.44]	-0.0002 [0.00]	0.1357 [0.64]	0.3648 * [1.69]	0.3265 * [1.67]	0.7025 *** [2.71]	1.0857 *** [3.78]
Low ROE	0.0266 [0.16]	-0.8080 *** [-3.89]	-0.1117 [-0.58]	0.5633 *** [3.34]	0.4397 ** [2.29]	0.1395 [0.59]	-1.0061 *** [-3.48]
Net	0.0954 [1.05]	0.2729 *** [3.07]	0.2198 ** [2.51]	-0.0914 [-1.24]	-0.0910 [-1.17]	0.1660 ** [2.10]	0.4308 *** [4.97]

Table 4. Partial ROE factor in performance of institutional portfolio.

The table measures performance of the institutional portfolio relative to a four-factor model. RMRF, SMB, and HML are Fama-French factors obtained from Kenneth French's website. Partial ROE factor is the return on a portfolio going long large (Quintile 5) high return on equity stocks and short small (Quintile 1 Large) low return on equity stocks. The ROE portfolio is book-to-market balanced. The factor regression alphas are annualized. T-statistics are in square brackets. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

Sample period	01/1982 - 12/1998	01/1982 - 12/2001
Number of months	204	240
Annualized alpha	-0.0022 [-0.53]	-0.0009 [-0.23]
RMRF	0.0476 *** [7.06]	0.0548 *** [8.05]
SMB	0.0570 ** [2.31]	0.0671 ** [2.52]
HML	-0.0625 *** [-4.03]	0.0131 [0.75]
Partial ROE factor	0.0750 *** [4.69]	0.0843 *** [5.49]
R squared	47.85%	35.57%
R squared adjusted	46.81%	34.47%

Table 5. Factor regression models for the period 1982-1996.

Panel A reports the annualized mean, standard deviation, and Sharpe ratio of the Fama-French market factor, RMRF, and the institutional universe factor, return on stocks in the institutional (SPECTRUM) universe minus the risk-free rate. Correlation is the correlation between the two-market factor returns. Mean, standard deviation, and Sharpe ratio are annualized. Panel B measures performance of the institutional portfolio over the rest of the market relative to the three- and four-factor models. RMRF is the Fama-French market factor in columns one and two and the institutional universe market factor in columns three and four, respectively. SMB and HML are Fama-French factors obtained from Kenneth French's website. ROE factor is the return on a portfolio going long high return on equity stocks and short low return on equity stocks. ROE portfolio is size-book-to-market balanced. The factor regression alphas are annualized. Panel C reports the estimates of the factor models for the excess return on the institutional universe market factor over the Fama-French market factor. RMRF, SMB, and HML and ROE factor are as in panel B. The factor regression alphas are annualized. T-statistics are in square brackets. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

Panel A. Fama-French market factor and institutional universe factor				
	Fama-French market factor		Institutional universe factor	
Number of months	180		180	
Annualized mean return	0.0940		0.0987	
Annualized standard deviation	0.1444		0.1468	
Annualized Sharpe ratio	0.6508		0.6726	
Correlation	99.90%			
Panel B. Performance of institutional portfolio				
	Fama-French market factor		Institutional universe factor	
Annualized alpha	0.0074 ** [2.28]	0.0007 [0.20]	0.0040 * [1.86]	0.0018 [0.69]
RMRF	1.0436 *** [146.88]	1.0464 *** [164.68]	1.0286 *** [256.07]	1.0294 *** [253.42]
SMB	-0.0274 ** [-2.26]	-0.0127 [-1.12]	-0.0364 *** [-4.03]	-0.0315 *** [-3.78]
HML	-0.0378 *** [-3.20]	-0.0318 *** [-2.72]	-0.0404 *** [-4.41]	-0.0386 *** [-4.19]
ROE factor		0.0796 *** [4.23]		0.0266 * [1.91]
R-squared	99.43%	99.49%	99.73%	99.74%
Adjusted R-squared	99.42%	99.48%	99.73%	99.73%
Panel C. Excess return of institutional universe portfolio over Fama-French market portfolio				
Annualized alpha	0.0032 ** [1.68]	-0.0011 [-0.58]		
RMRF	0.0153 *** [3.29]	0.0171 *** [4.52]		
SMB	0.0087 [1.32]	0.0183 *** [2.81]		
HML	0.0031 [0.52]	0.0070 [1.21]		
ROE factor		0.0516 *** [4.60]		
R-squared	11.04%	22.22%		
Adjusted R-squared	9.52%	20.44%		

Table 6. VAR parameter estimates with log returns.

Panel A reports VAR parameters estimated from quarterly panel. Log return and fraction of institutional ownership are state variables. Log book-to-market, log return on equity, and size-group dummies are introduced as exogenous variables. Size groups are defined in Appendix A.3. Panel B reports the one standard deviation shock to the fraction of institutional ownership and the infinite horizon return – permanent effect of a one standard deviation shock to the fraction of institutional ownership on the stock price. T-statistics in square brackets are calculated using standard errors corrected for cross-sectional correlation of the residuals. Appendix A provides further details on estimation methodology and calculation of standard errors. Stocks – is the time-series average number of stocks in a cross-section; observations – is total number of firm-quarters. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

<i>Panel A. VAR estimates for 1982-2001 sample</i>		
Stocks	3,939	3,521
Observations	311,208	278,172
	<u>Return regression</u>	
Return(-1)	0.0058 [0.42]	-0.0030 [-0.22]
IO(-1)	0.0409 *** [2.83]	0.0036 [0.30]
B/M	0.0182 *** [5.89]	0.0201 *** [6.60]
ROE	0.0258 *** [5.88]	0.0246 *** [6.02]
Dum Q1L		0.0127 *** [2.95]
Dum Q2		0.0240 *** [3.96]
Dum Q3		0.0281 *** [3.95]
Dum Q4		0.0315 *** [3.91]
Dum Q5		0.0350 *** [3.58]
R-squared	1.36%	1.48%
	<u>IO regression</u>	
Return(-1)	0.0178 *** [20.59]	0.0134 *** [17.22]
IO(-1)	0.9795 *** [649.55]	0.9627 *** [599.33]
B/M	-0.0010 *** [-4.36]	0.0002 *** [0.84]
ROE	0.0018 *** [7.68]	0.0010 *** [4.80]
Dum Q1L		0.0095 *** [20.72]
Dum Q2		0.0136 *** [21.38]
Dum Q3		0.0147 *** [22.03]
Dum Q4		0.0150 *** [21.81]
Dum Q5		0.0150 *** [17.58]
R-squared	95.25%	95.55%
<i>Panel B. Infinite horizon returns for 1982-2001 sample</i>		
Initial shock to IO	0.0506	0.0492
Infinite horizon return	0.1051 *** [2.58]	0.0048 [0.30]

Table 7. VAR parameter estimates with simple returns.

Panel A reports VAR parameters estimated from quarterly panel. Simple return and fraction of institutional ownership are state variables. Log book-to-market, log return on equity, log size, and interactions between size-group dummies and log size are introduced as exogenous variables. Regressions in columns one, three and five constraint the coefficient of log size to be the same for all size groups. Regressions in columns two, four and six allow the coefficient of log size to vary across size groups. Size groups are defined in Appendix A.3. Panel B reports the one standard deviation shock to the fraction of institutional ownership and the infinite horizon return – permanent effect of a one standard deviation shock to the fraction of institutional ownership on the stock price. T-statistics in square brackets are calculated using standard errors corrected for cross-sectional correlation of the residuals. Appendix A provides further details on estimation methodology and calculation of standard errors. Stocks – is the time-series average number of stocks in a cross-section; observations – is total number of firm-quarters. ***, **, and * denote significance at the 1, 5, and 10 percent, respectively.

Panel A. VAR parameter estimates						
Sample period	1982-2001		1982-1991		1992-2001	
Stocks	3,939	3,521	3,533	3,203	4,349	3,842
Observations	311,209	278,173	137,792	124,918	169,621	149,834
	<u>Return regression</u>					
Return(-1)	-0.0026 [-0.20]	-0.0034 [-0.29]	-0.0048 [-0.39]	-0.0042 [-0.33]	0.0026 [0.13]	0.0015 [0.08]
IO(-1)	0.0181 [1.54]	0.0090 [0.86]	0.0223 ** [2.09]	0.0145 [1.28]	0.0204 [1.21]	0.0131 [0.94]
B/M	0.0136 *** [4.28]	0.0131 *** [4.35]	0.0134 *** [3.79]	0.0122 *** [3.43]	0.0113 ** [2.19]	0.0106 ** [2.28]
ROE	0.0097 ** [2.42]	0.0110 *** [2.84]	0.0105 *** [2.89]	0.0127 *** [3.50]	0.0108 ** [2.00]	0.0118 ** [2.28]
Size	-0.0029 [-1.53]		-0.0010 [-0.41]		-0.0039 [-1.52]	
Size * Dum Q1S		-0.0233 *** [-5.55]		-0.0182 *** [-4.33]		-0.0284 *** [-4.19]
Size * Dum Q1L		-0.0144 *** [-4.45]		-0.0100 *** [-2.98]		-0.0186 *** [-3.53]
Size * Dum Q2		-0.0107 *** [-3.60]		-0.0066 ** [-2.11]		-0.0144 *** [-2.97]
Size * Dum Q3		-0.0093 *** [-3.58]		-0.0057 * [-1.92]		-0.0130 *** [-3.12]
Size * Dum Q4		-0.0082 *** [-3.43]		-0.0050 * [-1.82]		-0.0113 *** [-3.03]
Size * Dum Q5		-0.0069 *** [-3.27]		-0.0043 * [-1.71]		-0.0094 *** [-2.95]
R-squared	0.34%	0.53%	0.37%	0.54%	0.32%	0.53%
	<u>IO regression</u>					
Return(-1)	0.0122 *** [14.58]	0.0098 *** [12.41]	0.0094 *** [7.92]	0.0083 *** [7.55]	0.0141 *** [12.31]	0.0108 *** [9.48]
IO(-1)	0.9619 *** [596.10]	0.9603 *** [594.79]	0.9622 *** [348.42]	0.9619 *** [345.87]	0.9614 *** [465.02]	0.9587 *** [463.27]
B/M	0.0005 *** [2.70]	0.0006 *** [3.15]	0.0001 [0.44]	0.0002 [0.80]	0.0010 *** [3.55]	0.0011 *** [4.08]
ROE	0.0009 *** [4.51]	0.0008 *** [4.27]	0.0014 *** [5.71]	0.0012 *** [5.02]	0.0006 ** [2.06]	0.0005 * [1.95]
Size	0.0031 *** [26.00]		0.0029 *** [21.38]		0.0033 *** [18.15]	
Size * Dum Q1S		0.0031 *** [11.23]		0.0028 *** [8.74]		0.0037 *** [11.50]
Size * Dum Q1L		0.0042 *** [19.72]		0.0039 *** [15.42]		0.0047 *** [17.84]
Size * Dum Q2		0.0042 *** [21.50]		0.0037 *** [16.14]		0.0048 *** [20.36]
Size * Dum Q3		0.0038 *** [23.37]		0.0033 *** [19.45]		0.0044 *** [22.01]
Size * Dum Q4		0.0033 *** [22.99]		0.0031 *** [21.18]		0.0038 *** [19.08]
Size * Dum Q5		0.0028 *** [21.93]		0.0025 *** [17.32]		0.0032 *** [18.53]
R-squared	95.28%	95.55%	95.46%	95.71%	95.14%	95.43%
Panel B. Infinite horizon returns						
Initial shock to IO	0.0505	0.0493	0.0443	0.0432	0.0560	0.0547
Infinite horizon return	0.0241 [1.51]	0.0112 [0.85]	0.0261 ** [2.04]	0.0164 [1.28]	0.0299 [1.19]	0.0175 [0.94]

Table A.1. Descriptive statistics.

Panel A reports means, medians, standard deviations, percentiles, and first-order autocorrelations of return, log return, fraction of institutional ownership, log book-to-market, log US GAAP return on equity, and log market equity. Appendix A provides the details of data sources and variable definitions. Descriptive statistics are estimated from quarterly panel data using equally- and value-weighted schemes. Equally-weighted scheme assigns equal weights to all data points in a cross-section; value-weighted scheme assigns weights proportional to the stock's market capitalization; both schemes assign equal weights to all cross-sections. Stocks – is the time-series average number of stocks in a cross-section; observations – is total number of firm-quarters; sample weight – is the time-series average weight of a quintile. Sample period is from 1982, Q1, to 2001, Q4, 80 quarters in total. Panel B reports correlations between the data series for the sample period 1982, Q1, to 2001, Q4. Correlations are calculated using the equally-weighted and the value-weighted schemes explained in description of Panel A.

Panel A. Data series descriptive statistics							
	All stocks	Quintile 1 Small	Quintile 1 Large	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Stocks	3,939	1,267	791	508	356	303	296
Observations	311,209	100,126	62,486	40,165	28,096	23,943	23,357
Sample weight (%)	100.00	0.81	2.30	4.12	6.72	13.64	64.57
<u>Return</u>							
EW Means	0.0411	0.0451	0.0378	0.0433	0.0421	0.0424	0.0413
VW Means	0.0409	0.0366	0.0387	0.0431	0.0418	0.0424	0.0415
EW Std	0.2823	0.3555	0.2724	0.2346	0.2053	0.1863	0.1630
VW Std	0.1748	0.3238	0.2694	0.2328	0.2035	0.1858	0.1562
Min	-0.6800	-0.6800	-0.6800	-0.6800	-0.6800	-0.6800	-0.6800
25 Per	-0.1135	-0.1667	-0.1186	-0.0884	-0.0715	-0.0631	-0.0512
50 Per	0.0161	0.0000	0.0115	0.0289	0.0340	0.0364	0.0395
75 Per	0.1541	0.1702	0.1570	0.1568	0.1461	0.1376	0.1280
Max	1.5417	1.5417	1.5417	1.5417	1.5417	1.5417	1.5417
EW AR(1)	-0.0172	-0.0306	0.0064	-0.0245	-0.0429	-0.0527	-0.0006
VW AR(1)	-0.0104	0.0009	0.0068	-0.0263	-0.0416	-0.0506	0.0013
<u>Log Return</u>							
EW Means	0.0051	-0.0088	0.0037	0.0168	0.0211	0.0252	0.0277
VW Means	0.0255	-0.0095	0.0053	0.0169	0.0211	0.0254	0.0290
EW Std	0.2682	0.3260	0.2619	0.2314	0.2061	0.1838	0.1631
VW Std	0.1749	0.3031	0.2590	0.2297	0.2044	0.1830	0.1568
Min	-1.1394	-1.1394	-1.1394	-1.1394	-1.1394	-1.1394	-1.1394
25 Per	-0.1205	-0.1823	-0.1263	-0.0926	-0.0742	-0.0652	-0.0526
50 Per	0.0160	0.0000	0.0114	0.0285	0.0335	0.0357	0.0387
75 Per	0.1433	0.1572	0.1458	0.1456	0.1364	0.1289	0.1205
Max	0.9328	0.9328	0.9328	0.9328	0.9328	0.9328	0.9328
EW AR(1)	-0.0037	-0.0185	0.0153	-0.0174	-0.0410	-0.0559	0.0003
VW AR(1)	-0.0101	0.0073	0.0087	-0.0243	-0.0445	-0.0584	0.0038
<u>Fraction of Institutional Ownership (IO)</u>							
EW Means	0.2900	0.0930	0.2599	0.3834	0.4576	0.5179	0.5500
VW Means	0.5122	0.1203	0.2736	0.3895	0.4610	0.5227	0.5338
EW Std	0.2399	0.1167	0.1825	0.2128	0.2171	0.2004	0.1680
VW Std	0.1834	0.1286	0.1868	0.2139	0.2170	0.1987	0.1584
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
25 Per	0.0787	0.0088	0.1202	0.2155	0.2905	0.3771	0.4551
50 Per	0.2461	0.0526	0.2377	0.3729	0.4548	0.5350	0.5704
75 Per	0.4803	0.1400	0.3868	0.5493	0.6297	0.6805	0.6753
Max	0.9041	0.9041	0.9041	0.9041	0.9041	0.9041	0.9041
EW AR(1)	0.9826	0.9444	0.9635	0.9700	0.9716	0.9693	0.9698
VW AR(1)	0.9705	0.9477	0.9657	0.9706	0.9721	0.9685	0.9708

Table A.1 (continued)

	All stocks	Quintile 1 Small	Quintile 1 Large	Quintile 2	Quintile 3	Quintile 4	Quintile 5
<u>Log Book-to-Market (B/M)</u>							
EW Means	-0.4473	-0.2814	-0.4270	-0.4891	-0.5406	-0.5650	-0.6690
VW Means	-0.6898	-0.3274	-0.4287	-0.4931	-0.5396	-0.5635	-0.7455
EW Std	0.8302	0.9198	0.8177	0.7140	0.6779	0.7056	0.7606
VW Std	0.7718	0.9000	0.8051	0.7105	0.6787	0.7059	0.7809
Min	-3.6803	-3.6803	-3.6803	-3.6803	-3.6803	-3.6803	-3.6803
25 Per	-0.9155	-0.7640	-0.8398	-0.8816	-0.9450	-1.0144	-1.1919
50 Per	-0.3743	-0.1664	-0.3484	-0.4258	-0.4935	-0.5301	-0.6563
75 Per	0.0788	0.2986	0.0767	-0.0438	-0.0856	-0.0878	-0.1420
Max	1.6356	1.6356	1.6356	1.6356	1.6356	1.6356	1.5575
EW AR(1)	0.9508	0.9422	0.9475	0.9432	0.9525	0.9692	0.9810
VW AR(1)	0.9785	0.9453	0.9472	0.9440	0.9533	0.9686	0.9849
<u>Log Return-on-Equity (ROE)</u>							
EW Means	-0.0650	-0.2379	-0.0588	0.0404	0.0945	0.1148	0.1351
VW Means	0.1241	-0.1845	-0.0484	0.0451	0.0964	0.1159	0.1427
EW Std	0.6639	0.8445	0.6579	0.4760	0.3282	0.2282	0.2023
VW Std	0.2664	0.8099	0.6454	0.4654	0.3191	0.2222	0.2000
Min	-3.9120	-3.9120	-3.9120	-3.9120	-3.9120	-3.9120	-3.9120
25 Per	-0.0141	-0.2055	-0.0075	0.0560	0.0775	0.0791	0.0907
50 Per	0.0899	0.0214	0.0860	0.1139	0.1269	0.1278	0.1429
75 Per	0.1555	0.1050	0.1466	0.1650	0.1744	0.1787	0.1913
Max	0.9088	0.9088	0.9088	0.9088	0.9088	0.9088	0.9088
EW AR(1)	0.9215	0.9161	0.9172	0.9085	0.8884	0.8806	0.8729
VW AR(1)	0.8950	0.9189	0.9185	0.9109	0.8844	0.8910	0.8687
<u>Log Market Equity (Size)</u>							
EW Means	4.6242	2.3504	4.2252	5.3746	6.3133	7.2486	8.6886
VW Means	8.3747	2.8302	4.3454	5.4420	6.3679	7.3220	9.1855
EW Std	2.1295	0.9568	0.5767	0.5200	0.4939	0.5259	0.8429
VW Std	1.5665	0.7657	0.5683	0.5170	0.4902	0.5333	0.8572
Min	-0.0449	-0.0449	-0.0449	2.0510	3.9578	5.1861	5.9662
25 Per	3.0950	1.7703	3.8746	5.0243	5.9692	6.9219	8.1362
50 Per	4.5286	2.4982	4.2703	5.3899	6.3338	7.2950	8.6672
75 Per	6.1366	3.1025	4.6655	5.7803	6.6905	7.6572	9.3688
Max	10.1782	6.1592	7.4668	8.2022	8.2467	9.6692	10.1782
EW AR(1)	0.9996	0.9705	0.9697	0.9599	0.9542	0.9682	0.9831
VW AR(1)	0.9948	0.9784	0.9693	0.9596	0.9525	0.9696	0.9829

Table A.1 (continued)

Panel B. Data series correlations

<u>Equally-Weighted Correlation Matrix</u>						
	Return	Log Return	IO	B/M	ROE	Size
Return	1.0000	0.9617	0.0207	0.0686	0.0406	0.0997
Log Return	0.9617	1.0000	0.0592	0.0973	0.0999	0.1549
IO	0.0207	0.0592	1.0000	-0.0735	0.2015	0.7107
B/M	0.0686	0.0973	-0.0735	1.0000	0.3073	-0.1780
ROE	0.0406	0.0999	0.2015	0.3073	1.0000	0.2127
Size	0.0997	0.1549	0.7107	-0.1780	0.2127	1.0000

<u>Value-Weighted Correlation Matrix</u>						
	Return	Log Return	IO	B/M	ROE	Size
Return	1.0000	0.9765	0.0285	0.0512	0.0444	0.0766
Log Return	0.9765	1.0000	0.0252	0.0853	0.0679	0.0989
IO	0.0285	0.0252	1.0000	-0.2312	0.0853	0.2931
B/M	0.0512	0.0853	-0.2312	1.0000	-0.1062	-0.2809
ROE	0.0444	0.0679	0.0853	-0.1062	1.0000	0.1455
Size	0.0766	0.0989	0.2931	-0.2809	0.1455	1.0000

Figure 1. Cumulated institutional and individual portfolio returns and cumulated excess return.

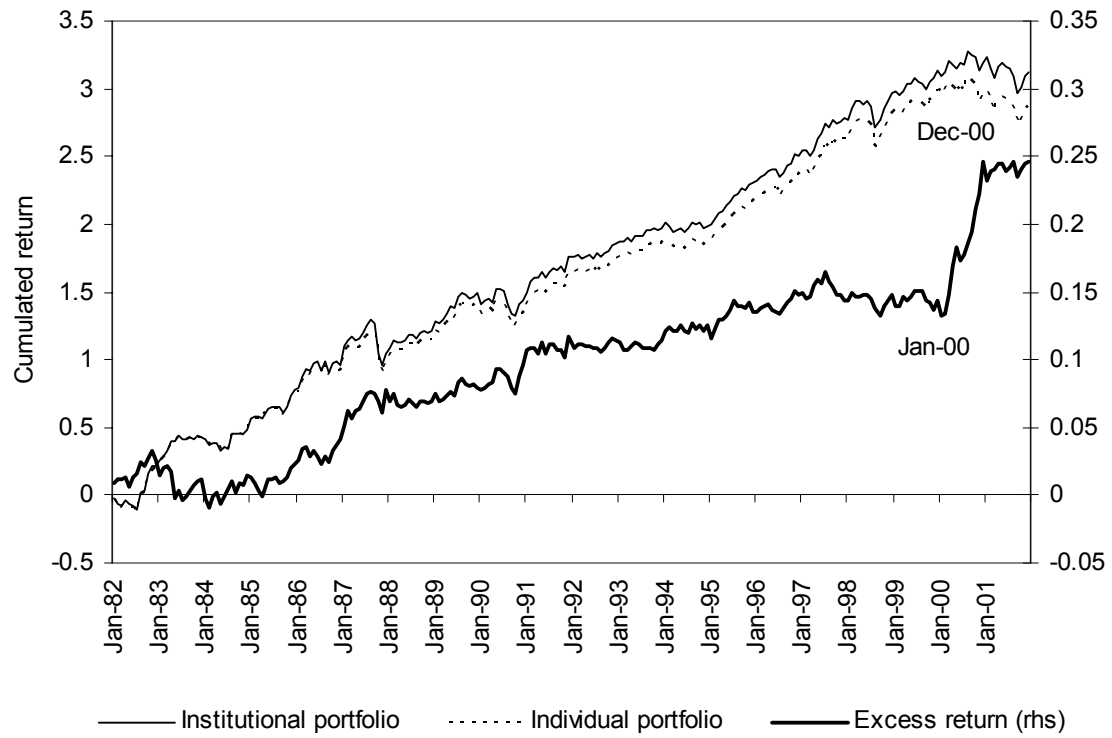


Figure 2. Cumulated returns on ROE factor and Fama-French factors.

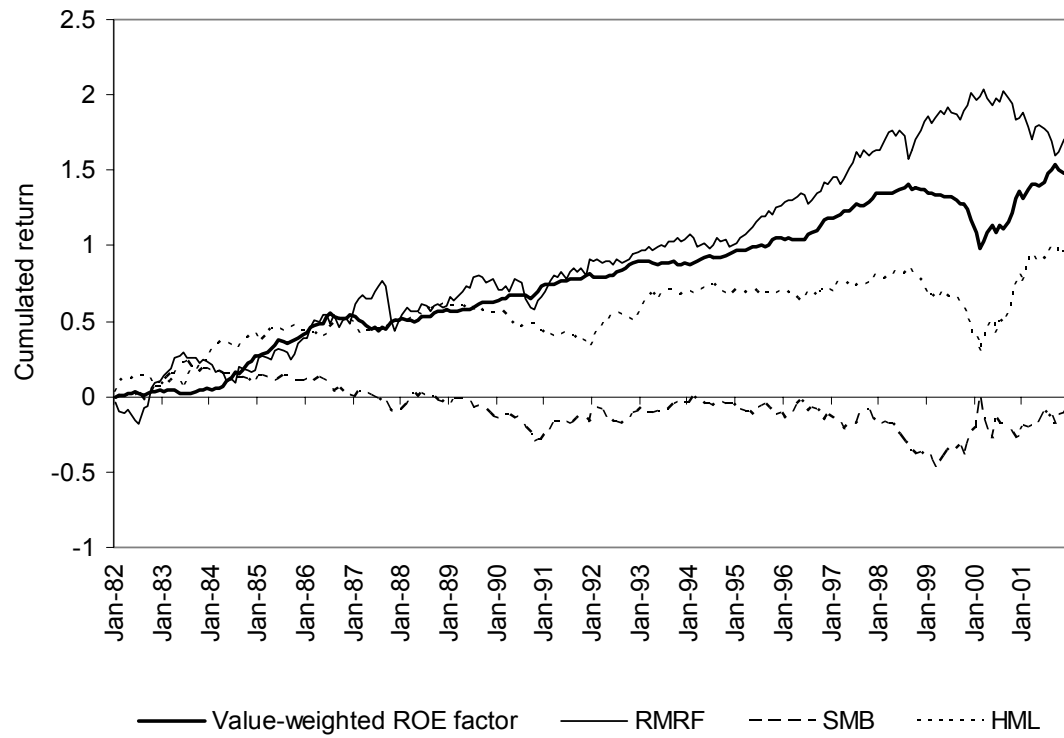


Figure 3.a. Actual and fitted returns of excess return of institutional portfolio over the individual portfolio. Early sample: 01/1982-12/1998.



Figure 3.b. Actual and fitted returns of excess return of institutional portfolio over the individual portfolio.
Full sample: 01/1982-12/2001.

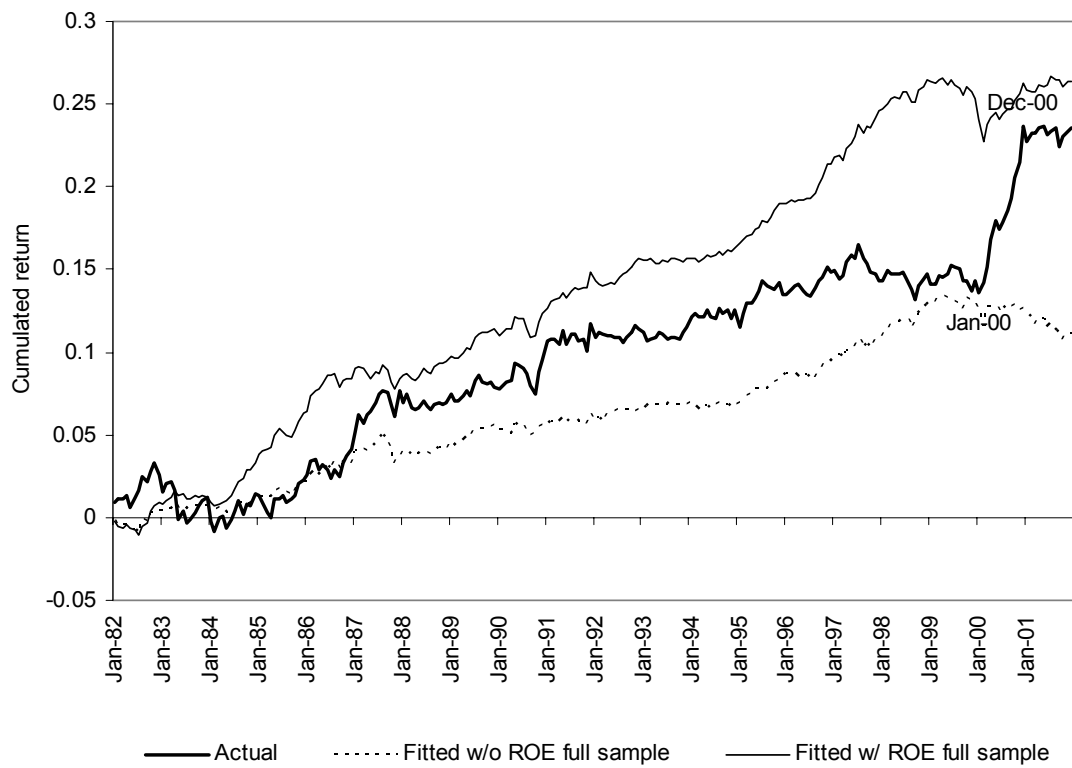


Figure 4. Residuals from the models of excess returns. Early sample: 01/1982-12/2001.

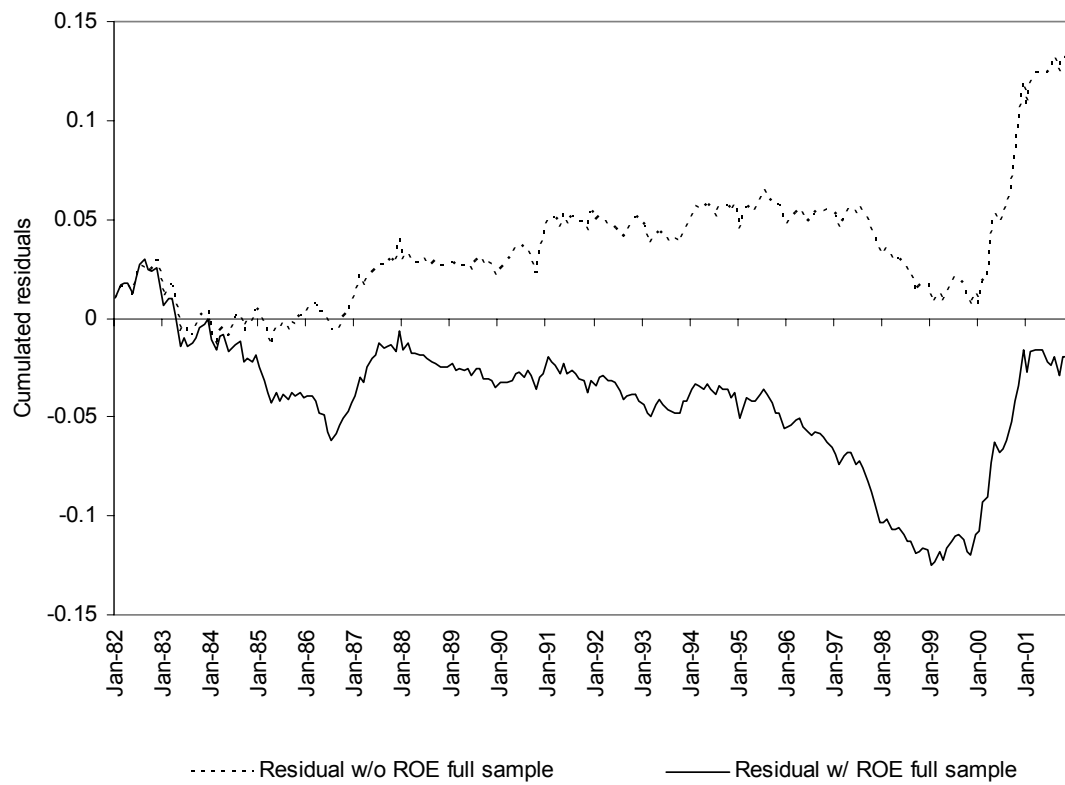


Figure 5. Predicted and actual excess returns on the institutional portfolio.

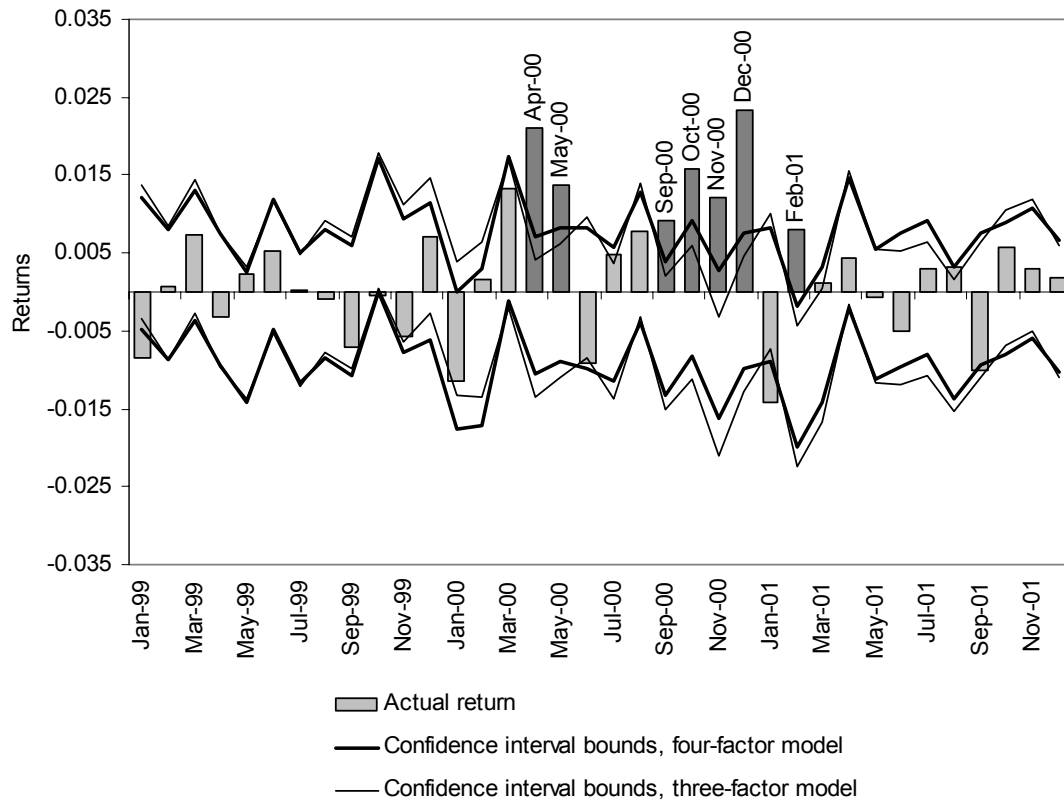


Figure 6. Annualized log returns across size groups.

