Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

__________________________________ __________

Jonathan W. Weaver Date

Future Security Returns and Factor Model Information Content

By

Jonathan W. Weaver
Master of Business Studies

In Business

____
Jay A. Shanken, Ph.D.
Advisor

____
Narasimhan Jegadeesh, Ph.D.
Advisor

Accepted:

____
Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

____
Date

Future Security Returns and Factor Model Information Content

By

Jonathan W. Weaver
Master of Business Administration, 2000
Bachelor of Arts, 1993

Advisor: Jay A. Shanken, Ph.D., Carnegie-Mellon University, 1983

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of

Master of Business Studies

In Business

2010

Abstract

Future Security Returns and Factor Model Information Content

By

Jonathan W. Weaver

In this paper I use the R-squared of a factor model estimated over the prior five years to examine whether future returns can be explained by a stock's past ratio of explained volatility to total volatility. I argue that the lack of success of a factor model in providing useful output creates an ambiguity premium that increases future expected returns. I find evidence that this is in fact the case, and for the 41 years from 1968 to 2008, the R-squared from the factor loading estimation period is superior to both idiosyncratic and total risk in sorting future returns. Considering what is known relative to what is unknown for a stock's risk in a portfolio is a more precise measure to examine a premium for lack of information, or ambiguity. I find that R-squared also is related to future risk and is informative in predicting systematic risk estimation error. The R-squared retains predictive power while holding idiosyncratic, systematic, or total risk quintiles constant.

Future Security Returns and Factor Model Information Content

By

Jonathan W. Weaver
Master Business Administration, 2000
Bachelor of Arts, 1993

Advisor: Jay A. Shanken, Ph.D., Carnegie-Mellon University, 1983

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of

Master of Business Studies

In Business

2010

Table of Contents

List of Figures

List of Tables

1. Introduction

Traditional asset pricing models argue that higher returns are compensation for higher systematic risk. For example, the capital asset pricing model (CAPM) suggests that the expected return on a stock is a linear function of its beta, a measure of covariance between its return and the market return. Empirical shortcomings of the CAPM discovered by Basu (1977), Fama and French (1992) and Jegadeesh and Titman (1993) have motivated other asset pricing models such as the Fama-French 3 factor model and the Fama-French 3 factor model augmented by the winners-minus-losers portfolio, a proxy for momentum. While the models above all differ in the determination of specific risk factors, they do all agree that risk should be measured as a sensitivity, or covariance, to an appropriate risk factor.

The CAPM argues that only systematic risk matters and that any volatility that is not correlated to the market return can be diversified away. The Fama French 3-factor model uses similar logic to extend the list of rationally priced covariates to 3, and momentum, or a winners-minus-losers portfolio, adds a fourth. Despite the theoretical and historical prominence of these arguments, there are numerous reasons to believe that it is not an accurate depiction of how investors actually perceive risk, particularly in situations where covariance is measured with error or unknown. It seems plausible that investors view a risk free asset with constant payout as less risky than a wildly volatile asset that happens to have zero covariance with the market return. This of course requires the behaviorally based argument that securities are at least in part judged in isolation rather than in a portfolio context. Moreover, Barberis and Huang (2001) show that if investors are loss averse over fluctuations in individual stocks within their portfolio, then investors are willing to pay a premium for stocks with low idiosyncratic risk.

Nevertheless, behavioral biases such as loss aversion and mental accounting are not necessary for factor loading estimation risk to be priced. Rational reasons such as transaction costs, knowledge of private information, and employee compensation that give workers company equity but place restrictions on selling the company equity, could also consequently result in limited diversification, making the precision of estimating the portfolio risk, or risk factor loadings, a reasonable concern. In fact, there is substantial empirical evidence that most individual investors do not hold well diversified portfolios. Barber and Odean (2000) report that the mean household's portfolio contains only 4.3 stocks and that the median household invests in only 2.6 stocks. Benartzi (2001) and Benartzi and Thaler (2001) document that individuals hold a disproportionate amount of their pension plan funds in the stock of the company they work for. Huberman (2001) surveys evidence that investors are prone to investing in familiar stocks, and ignore portfolio diversification. Malkiel and Xu (2002) argue that when a subset of investors does not choose to hold the market portfolio, the resulting investors must then consider both idiosyncratic and total risk in addition to market risk. This is due to "constrained" investors and "free" investors together making up the entire market.

It is therefore not surprising that there is a mass of empirical literature devoted to deciphering the exact relationship between idiosyncratic volatility and expected returns. Ultimately, the literature has been mixed. Litner (1965) shows that idiosyncratic volatility carries a positive coefficient in cross sectional regressions. Lehmann (1990) finds similar results over a longer sample period. Similarly, Tinic and West (1986) and Malkiel and Xu (2002) find that portfolios with high idiosyncratic volatility risk have higher average returns. Green and Rydqvist (1997) study Swedish lottery bonds and find that these bonds command a return premium for a risk that is idiosyncratic by construction. However, Longstaff (1989) finds that a cross sectional regression coefficient on total variance for size sorted portfolios carries an insignificant negative sign. Moreover, using daily data Ang, Hodrick, Xing, and Zhang (2006) find that stocks with high average idiosyncratic volatility have significantly lower average returns.

One implication in using a factor model to price assets is that estimation risk is not priced, which in turn relies on assumptions such as mean-variance maximizing investors with costless diversification. However, evidence from psychology suggests that most investors are averse to ambiguous risks. The Ellsberg (1961) paradox outlines that people have a preference for known risks rather than unknown risks. Klein and Bawa (1976) and Klein and Bawa (1977) conclude that optimal portfolio choice is affected by parameter uncertainty, and that increasing the precision of the parameter estimates ultimately leads to an increased allocation to the risky asset due to the redeuction in parameter uncertainty. The situation where firms face differing levels of parameter uncertainty is addressed in Barry and Brown (1985). They find the difference in the CAPM betas versus the true betas is related to the level of parameter information. A different situation is examined in Merton (1987) where not every investor has equal breadth of knowledge, but there is complete agreement on parameter levels. In this case the level of demand, thus the required rate of return in equilibrium, is affected by a firm's visibility. Meanhout (1999) concludes that if investors are concerned that their model is misspecified, they will charge a premium for the perceived ambiguity of the unknown probability distribution. In a differential information setting, Easley and O'Hara (2004) conclude that those investors who realize that they are at an informational disadvantage, such as a in security that has high levels of private information potential, will demand a premium to compensate for this disadvantage. Given that some of the assumptions required for a factor model to correctly price assets seem untenable, coupled with the empirical evidence referenced above clearly showing that most investors do not hold well-diversified portfolios, it is only natural to challenge the conclusions that follow from those assumptions. The approach in this paper is to examine if there is information in the results of a factor structure model estimation that provide evidence that the model is incomplete, or fails to correctly price expected returns.

This paper should contribute to this extensive literature by examining a possible alternative explanation to the idiosyncratic risk literature. If the inability to understand the risk of a stock in a portfolio is in fact a priced risk premium itself, the R-squared is a superior measure by which to consider how observed idiosyncratic risk can affect asset pricing because it essentially contains information content about the level of pricing model ambiguity. Rather than idiosyncratic risk in and of itself being a determinant of expected returns, perhaps it is the unwanted side effect of the imprecise or ambiguous systematic risk component estimates that excessive idiosyncratic risk itself creates that is ultimately priced. All else being equal, it seems both plausible and rational to prefer a stock with a known covariance to a given risk factor to a stock with an unknown covariance. Having stocks with known covariances to risk factors would facilitate the portfolio construction process undertaken to give the investor the highest utility risk reward tradeoff in expectation, under the assumption that covariances accurately describe expected returns. Adding the additional layer of ambiguity created by estimation risk requires costly diversification to render it meaningless, adding relative importance to the cost of holding undiversified portfolios.

Thus, the main idea of this paper is to explore stock returns under different levels of factor model explanatory power. I endeavor to quantify this preference in terms of a return premium afforded these stocks for which the systematic risk component is small relative to a stock's overall risk. I also explore two alternative proxies for the model information content: the absolute value of the sum of the T-statistics of the model coefficients and the sum of the standard errors of these coefficients. In this scenario where the proxies for model success is low (high in the case of sum of standard errors) investors may find it difficult to gauge a stock's merits in a portfolio, and the resulting factor loading coefficient estimates can suffer from bias and imprecision, essentially signaling the ineffectiveness of the results. The main engine, which delivers the importance of an ambiguity premium, requires that investors care about individual components of a portfolio; an implication of holding undiversified portfolios where accurate parameter estimates or model results can lead to a higher expected future utility.

Following the work of Brennan, Chordia, and Subrahmanyam (1998), I first ran the Fama-French 3-factor market model regressions to compute 5 year rolling estimates of the factor loadings. I then multiplied the coefficient estimates by the factor realizations for the following period and added the risk free rate to compute an expected return for each stock given the factor returns and the coefficient estimates. Next, I subtracted this calculated expected return from the stock's actual return to compute an unexpected, or abnormal, return for the month. I then examined the behavior of the resulting unexpected returns on calculated R-squared. Finding a mechanical relationship between abnormal returns and R-squared was natural resulting from sorting on estimation error. Consequently I turn the focus to examining raw stock returns and what information the calculated R-squared can convey about expected returns when the systematic risk is low relative to the idiosyncratic risk.

This asset pricing framework uses single securities in order to examine the relationship between relative risk components and returns in the cross-section. Starting with individual regressions avoids the loss of information when stocks are sorted into portfolios (Litzenberger and Ramaswamy (1979)). In this context I am considering potential estimation error and model misspecification potential as proxied by R-squared a firm specific characteristic. Thus, this interaction between perceived sources of risk appears more robust than looking at either idiosyncratic or systematic risk's role in expected returns, specifically through its channel of increasing noise or bias, or general lack of confidence in the estimate of a stock's systematic risk component(s), and potentially decreasing the pricing model's role in explaining returns.

Using my measure, I find the following interesting results. The baseline factor model for calculating the R-squared in these results is the Fama French 3-factor model (1993). I find that future period returns decrease monotonically with the R-squared of the estimation period. The average stock in the lowest decile of R-squared compared to the highest decile stocks earns a statistically significant 47 basis point higher return in 1^{st} post-regression month, and a statistically significant 8.95% higher return in 1^{st} post-regression year. Sorting stock by other risk characteristic deciles show patterns consistent with higher idiosyncratic and higher total risk earning higher returns, but the spread and monotonicity is less pronounced. I also find that R-squared predicts the bias in the Fama French 3-factor coefficient estimates relative to future periods. Further, I find that R-squared retains its predictive ability holding whether idiosyncratic, systematic, or total risk is held constant. This measure is superior due to its additional ability to predict when the pricing model is most accurately related to future returns and when it is least accurate. The results are qualitatively similar using the two alternative proxies previously mentioned.

This result is generally consistent with and is somewhat parallel to the rational models of investors demanding a return premium for idiosyncratic risk due to diversification costs (see Malkiel and Xu (2002) and Jones and Rhodes-Kropf (2003)). Moreover, there may be a plausible preference argument that excessive idiosyncratic volatility disturbs the clarity of estimating the loading on a rationally priced covariate, and that little systematic volatility relative to total risk creates pricing model risk. The results are inconsistent with the findings of Ang, Hodrick, Xing, and Zhang (2004) who are baffled by their results. They conclude, “the cross-sectional expected return patterns found by sorting on idiosyncratic volatility presents something of a puzzle.” I present a different take on the literature related to idiosyncratic volatility which could shed light on the role of both total volatility and idiosyncratic volatility, by considering the amount of noise they introduce in determining a stock’s risk in a portfolio. I open the question as to whether for a subset of stocks, the proportion of calculated systematic risk can give information as to whether stocks are priced according to the model which calculated that proportion. If it is unlikely that a pricing model contains useful information going forward, this model may simply be replaced by a more behavioral pricing model or heuristic. I find that the inability of a pricing model to yield useable results corresponds to higher stock returns going forward.

This paper will proceed as follows. Section 2 discusses the mechanical relationship between R-squared and estimation risk. Section 3 discusses the data and describes the methodology in greater detail. Section 4 presents the empirical findings. Section 5 explores two alternative measures of parameter uncertainty, and Section 6 concludes.

2. A Simulation of Calculated R-Squared and Beta

When the risk of a stock over a time period is decomposed into its explained components and unexplained, the decomposition naturally takes an observed total volatility, which may or may not be representative of that stock’s true volatility, and divides it into two components which are completely orthogonal by construction. Thus estimated idiosyncratic and systematic risk are assumed to have had zero correlation on average in the sample. For a single regression, if the true realizations of the two components of risk happened to have something other than zero correlation during the estimation period, the resulting estimates of both volatility components will have measurement error in relation to the magnitude of this chance realized correlation.

To gain insight into how this can affect the coefficient estimates, I performed the following simulation. For 60 months, I generated a random market return from a normal distribution with 1% mean and 4.6189% standard deviation (equivalent to a 16% annual standard deviation). I then generated returns for 1000 identical stocks for the same 60 months. For each stock, I introduced a 16% annualized idiosyncratic risk component, and forced a true exact beta of 1. So for each month, the stock return was generated as the exact market return plus a random realization from an i.i.d. normal distribution of mean 0, and a 4.6189% standard deviation. I then ran regressions of these simulated stock's returns on the market return. In expectation, every stock should have a zero constant, a beta of 1, and an R-squared of 50%. Obviously the random nature of the idiosyncratic risk will make some realizations of the idiosyncratic component of a stock's return correlated with the market return, and some realizations inversely correlated over the 60 month period. I then sorted the stocks on the R-squared of the regressions, and have presented summary statistics in Table 1. Recall that every stock has a true beta of 1 by construction, and therefore a true expected return of 1% by construction.

Table 1 shows a representation of what the potential spread in calculated betas over a 60 month period is for 1,000 stocks. It is easy to see that sorting by R-squared and looking at the beta coefficient estimates, that on average, using these coefficients to estimate next period's abnormal return will result in the perception of a strong relationship between R-squared and abnormal return. This is the aforementioned mechanical relationship. The calculated betas for each decile have a range of 0.45 and the R-squared calculations have a range of 0.31 as a natural result of sorting and forming decile averages. The simulation shows that an abnormal return spread of 45 basis points per month could be observed simply as a function of identical risk and identical betas among 1,000 stocks, and the result of an idiosyncratic component of risk whose random realizations also happen to have a random correlation structure with the market return, which affects the estimated beta. The fourth column shows that for a 60-month period, the range in coincidental correlation can be substantial. This in turn affects the perception of how much idiosyncratic risk is actually truly idiosyncratic, and consequently how much systematic risk explains the past returns. Columns 5-7 show how this sorting can also affect the three components of the risk decomposition. Again, we know since the true risk of each stock is identical in expectation by construction, stocks with low calculated betas appear to have high idiosyncratic risk that appears related to positive abnormal returns, and vice versa. The sorting process also shows a relationship between abnormal returns and Total risk. In essence, there are two sources from which to observe a lower R-squared in the simulations. The first, assuming the realization of the idiosyncratic component is not different from 16% annualized volatility, is by spurious negative correlation of the idiosyncratic component to the market return. The second is a stock that has a higher than expected volatility of the realized idiosyncratic component, say calculated at 20%, with no spurious correlation to the market component, thus the observed total risk in relation to the explained risk is higher. Either will result in lower R-squared and higher standard error of the coefficient, but only the first should create a bias in the coefficients, therefore a bias in expected return. Of course both effects are likely to be present to some degree in each 60-month stock sample. A coincidental positive correlation between the market return and an average volatility realization of the idiosyncratic component will mask itself as a higher total risk realization, and result in a higher market beta calculation, creating a perceived negative abnormal return next period. This simple simulation may shed some light on how the different conclusions in the literature could show mixed results, depending on the interaction of the systematic, idiosyncratic, or total component of risk being observed, and the methodology used.

3. Data and Methodology

I use three data sets in this paper. The first is used to obtain monthly risk factors for the Market excess return, the Small-Minus-Big and High-Minus-Low porfolios, and the risk free rate, acquired from Ken French's website. The second data set is from the COMPUSTAT tape, used to calculate the Book Values of individual equities. The third data set is from the CRSP (Center for Research in Security Prices) database is includes returns for all stocks in the database from August 1963 to December 2008. I merged these three data sets, eliminating all stocks not listed on either the NYSE, the AMEX, or

NASDAQ. I also dropped returns greater that 2400% (results are materially unaffected if I had chosen 500%) and kept only common ordinary shares. Following the majority of the literature on asset pricing anomalies, I dropped financial firms as well.

The following is the specification of the Fama French 3-factor model used in the paper to obtain coefficient estimates, risk decomposition estimates, and R-squared calculations. I used the risk free rate data to calculate an excess return for each stock in each month. In a tradeoff between the timeliness of information and the amount of information contained in the time series, and to achieve some level of consistency with other research, I decided I would compute factor loading estimates from 5 year monthly rolling time series regressions. Consequently, I dropped all stock that did not have at least 60 months of data. The following regression was run over the prior 60 months for each current month data for stock i:

$$R_{it}\text{-}R_{ft} = \alpha_{it} + \beta_{i1}(market) + \beta_{i2}(smb) + \beta_{i3}(hml) + \upsilon_{it} \qquad (1)$$

This resulted in monthly estimates of β_i , systematic volatility, idiosyncratic volatility, and monthly R-squared results, (and other regression statistics) for each stock which were stored for later use.

I then used the rolling factor loading coefficient estimates from the past five years, known at the end of period *t-1*, multiplied them by the current period factor realizations in period *t*, and added the risk free rate to get an expected return for period *t* for each stock. I then subtracted this expected return from the realized return to get unexpected return for period *t*.

$$Unexpected\ Return_{it} = Realized\ Return_{it} - Expected\ Return_{it} \qquad (2)$$

(*using estimates from period t-1*)

From the simulation results in section 2, the Unexpected return as expected shows that abnormal returns are related to calculated R-squared, and are significant. Using this as justification that not only does R-squared relate to the future dispersion of coefficient estimates, it also relates to bias. For this reason the primary focus of the paper is whether R-squared has explanatory power for future raw returns, and how using two components of risk can lead to more information than analyzing idiosyncratic risk alone.

The first step, recognizing that R-squared carries information about the bias is to examine future changes in calculated risk coefficients and volatility components. I then sort the observations into deciles starting with R-squared, then moving to idiosyncratic risk deciles, followed by systematic risk deciles, and finaly Total risk deciles. To differentiate the ability of R-squared to predict future return differences, I form three sets of 25 portfolios, each with R-squared quintiles holding the other risk measure quintile constant.

Finally, I examine whether a heurisitic pricing model performs better for stocks with low R-squared. If the factor covariance structure is unknown, let us assume that a stock is simply priced relative to its volatility. Using past individual stock and market volatility, I calculate the ratio of the two, and assume that an investor requires a higher return for more volatile stocks. The intuition is that the level observed total variance is a signal for the *potential* of undiversifiable risk. This absence of covariance in the pricing of the required rate of return violates theory, but let us assume that the volatility ratio proxies for some level of ambiguous systematic risk. I use the following specification for the heuristic pricing model to calculate expected returns. The coefficient is chosen somewhat ad hoc to scale the observed spread in volatilities to the observed spread in returns when sorted on volatility. I chose total risk as the heuristic variable of choice because it has no components of risk decomposition invoked by a factor model. Obviously a more formal approach would yield a more rigorous model, but for the sake of exploration, let the following suffice:

$$E(R_{it}) = R_{mt} + 0.002\ (Volatility_{i,t\text{-}1\ to\ t\text{-}60} / Volatility_{m,t\text{-}1\ to\ t\text{-}60}) \qquad (3)$$

I compare the relative ability of the Fama-French 3 factor model to the heuristic model to explain the following month's stock return. To further the investigation, I also allow for the possibility of a combined model in that stocks may be priced using a factor pricing model in combination with the heuristic model, with the relative importance of each model being determined by the past R-squared. This would serve to accommodate a "partial" adjustment to reflect more granularly the effect of ambiguity and expected return adjustment made in response to the estimation error. I use the following regression specification:

$$R_{it} = \alpha_{it} + \beta_{i1}(FF\text{-}3\ model\ exp\ ret) + \beta_{i2}(heuristic\ model\ exp\ ret) + \upsilon_{it} \qquad (4)$$

I then look at how the coefficients change over various quintile sorting methods.

Note that the entire methodology revolves around using estimates known at time *t* to predict future returns only with realizations of factors in the case of the Fama French 3-factor model, and the realization of the market return in the case of the heuristic model. This is in contrast to traditional asset pricing work that uses a formation period, holding period, and estimation or ranking period. In this paper returns are predicted by information known *a priori,* due to the focus of pricing estimation and ambiguity levels that are observed ex ante.

I expect that deviations of the coefficient estimates for either model from their true value are not likely to be corrected in the subsequent rolling 60-month regression. The error terms for a given stock are likely to exhibit some degree of serial correlation in addition to the cross-sectional correlation the Fama-Macbeth (1973) procedure usually suffices to correct. In panel data cases where there is likely to be both time-series and cross-sectional correlation, clustering the standard errors on two dimensions can lead to more accurate standard errors (Peterson, 2006). The double cluster by firm and month takes into account both the lack of serial and cross sectional independence and also adjusts for heteroskedasticity. As Petersen (2006) notes, "standard errors clustered on multiple dimensions are unbiased and produce correctly sized confidence intervals whether the firm effect is permanent or temporary." Results based on the Fama-MacBeth (1973) methodology produce qualitatively similar results.

4. Empirical Findings

4.1 Understanding What R-Squared Implies for Model Uncertainty

I begin the empirical investigation by examining the information content of observed past 60 month R-squared with respect to factor bias and uncertainty. The first goal is to highlight how the market beta observed changes from its calculated value at time t to its calculated value at time $t+60$, a period chosen to have 60 independent months used in the future factor loading calculation. Stocks were first grouped by original calculated market beta, in increments of 0.1, starting from a market beta of 0.5 and moving to 1.5, creating 10 original market beta groups. Table 2 contains two rows for each beta grouping, one corresponding to the stocks classified originally in the lowest R-squared decile, and one for the highest R-squared decile. I dropped all observations for which a beta estimate and corresponding R-squared did not have a future estimate within each stock. This essentially eliminates all stocks for which there is less than 10 years of monthly data. Two important insights come from the table. The first is that as expected, the dispersion of the future betas is roughly twice as much for the low R-squared stock as the dispersion in future betas for the high R-squared stocks. The weighted interquartile range and standard deviation for future betas is 0.80 and 0.72 respectively for the low R-squared decile, and 0.47 and 0.39 respectively for the high R-squared decile. This is the general focus on how precision can affect portfolio formation. A high R-squared in the original estimation should increase the probability that the future realized factor loading will be closer to the original estimated factor loading when the portfolio is formed. This can, in preference terms, raise expected utility when an investor does not hold a well-diversified portfolio.

The second insight from Table 2 is seeing a pattern in either the increase or decrease in the future calculated market beta, depending on the R-squared. For all but the highest original market beta group, stocks in the low R-squared deciles have, on average, higher market betas 60-months in the future. The weighted average increase in market beta is 0.27 for this group. The high R-squared decile in each and every original market beta grouping shows a lower average market beta 60-months forward. The weighted average decrease in this group's market beta is -0.22.

This finding illustrates how R-squared is mechanically related to next month's abnormal return, through a simple example. If a high R-squared stock reports a coefficient estimate on its past 60-month regression of 1.22 for market beta, but the *future* market beta is 1.00 similar to the average Table 2 findings, and the average market excess return is some positive number in general in period $t+1$, then this high R-squared stock is likely to report a negative *unexpected* return next period due to its *over*estimated market beta. If a low R-squared stock reports a coefficient estimate on market beta from the past 60 month regression that is 0.73, but the *future* market beta is 1.00 similar to the average Table 2 findings, and again the average market excess return is some positive number in general in period *t+1*, then this high R-squared stock is likely to report a positive *unexpected* return next period due to its *under*estimated market beta. Not only is this mechanical relationship the driving force behind this paper choosing to focus on analyzing future *raw* returns, it highlights the theme of using R-squared as a proxy for imprecision and bias, creating a metric to gauge the accuracy, or usefulness of the pricing model in explaining returns.

Figure 1 shows future calculated factor loadings and risk components. Two Panels for each of the highest and lowest R-squared quintiles highlight the differences in future patterns of both factor loadings and risk components, averaged by future month conditional on a stock in month *t* belonging to either the first (lowest) R-squared quintile or the fifth (highest) R-squared quintile. For the lowest R-squared quintile, Panel A shows that the future calculated systematic risk component increases over time and both the idiosyncratic and total risk decline. Panel B shows that for the lowest R-squared quintile each of the 3-factor sources of systematic risk increases, highlighting that market beta is not the only source of bias created when grouping into R-squared quintiles. As seen in Panel C, the quintile 5 (highest) R-squared stocks have declining systematic risk driving the decline in total, while idiosyncratic risk remains relatively constant. Panel D shows that high R-squared stocks show a reduction in their calculated market and Small-minus-Big coefficients, while their High-minus-Low coefficient tends to increase. Again, this shows the bias is not limited to the market beta phenomenon illustrated in the previous table. R-squared appears to have a high level of information content for the interaction of various sources of calculated risk.

4.2 The Power R-Squared as a More Robust Measure of Risk

Table 3 presents four panels of summary statistics for the sample, by sorting stocks into deciles according to various calculated risk components. Since the data starts in August 1963, and 60 months are required to calculate the factor loading coefficient estimates and R-squared data, the summary statistics in Table 3 start in August of 1968. I first sort the regression results into R-squared deciles, and present summary statistics as the average of that statistic for that decile. For example, stock *j* in November of 1983 has 60 months of data up to and including November of 1983. Those 60 months are used to calculate stock *j's* coefficient estimates on the market, small-minus-big, and high-minus-low factors for those same 60 months. The factor loading coefficient estimates, Idiosyncratic risk, Systematic risk, Total risk, and R-squared of that 60 month regression are captured as data relevant (known as of the end of November 1983) for predicting the return for December of 1983 and the 3,6, and 12 months forward. The size and book-to-market are reported as of November 1983, and the previous compounded 12-month return is reported as the return from the November 1982 to October of 1983, essentially skipping November of 1983 to avoid any bias potentially introduced by the bid ask bounce. Columns 2-5 report the future raw returns, and Low-Minus-High R-squared decile differences and t-statistics. Columns 6-11 report the calculated factor loadings and annualized volatility estimates of the various risk components. Column 12 reports the previous 12-month return with one-month lag. The investor would know all the information before the future returns are realized with the exception of the factor loading realizations for future months. These realizations create the expected return from the Fama French 3-factor model, given the coefficient estimates and following factor realizations. The data set includes 1,037,169 firm-months, or just over 100,000 observations per decile.

As shown in Table 3 Panel A, the average Low R-squared decile future returns are 1.55% next month, and 22.77% for the next year. Conversely, the average High R-squared decile future returns are 1.08% next month, and 13.82% for the next year. The differences three months and beyond are significant. Looking at the following 12-month period from column 5, the range is 8.95% and the sort is monotonic with the sample mean in the middle of deciles 5 and 6, as expected. The average stock in the full sample has an average R-squared of 0.26 (not shown), or about one-quarter of its past monthly variation explained by systematic factors. The highest R-squared stocks have roughly half of the variation explained by the systematic component, while the lowest R-squared decile stocks have roughly one-sixteenth of their overall variation explained by this covariance with systematic risk factors. There is a strong relationship between the calculated market

coefficient and the R-squared, but in a curiously absent similar relationship among the other two Fama French risk factors. It would be difficult to make an argument that stocks in the Low R-squared decile have the highest returns because they have the highest risk coefficients. In fact the Fama French 3-factor model would suggest these stocks should have the lowest expected returns.

The lowest R-squared deciles have the highest Idiosyncratic and Total risk.
There are patterns in both size and book-to-market (not shown) when sorted into R-squared deciles. The natural log of size increases monotonically with R-squared, indicating that the small minus big factor does not preclude a more granular relationship between size and R-squared. In addition, there is also a prevalent monotonic relationship between book-to-market and R-squared, with high R-squared stocks having lower book-to-market values, though significantly less pronounced than a sort based solely on book-to-market. On a final note, there is a comparable pattern in previous 12-month return to the pattern observed in total risk. There is a minor decline in past returns as the R-squared decile increases, again negligible as compared to a sort on past return, which would have produced a spread of 175% between highest and lowest deciles, rather than 4%.

Panel B shows identical summary statistics but stocks are grouped into idiosyncratic risk deciles. It appears there is a relationship between idiosyncratic risk and future return, with an insignificant difference between the highest and lowest deciles. Both the range and monotonicity are far less prominent than when stocks are sorted by R-squared. It is interesting to note, in contrast to sorts by R-squared, that the systematic coefficients for market beta and smb show a pattern consistent with increasing return being related to systematic risk. The range of idiosyncratic risk is quite large, from 17% to 91%. The high idiosyncratic risk deciles also show superior past 12-month returns.

Moving to Panel C, the factor loadings illustrate the pattern consistent with sorting on systematic risk, with the exception of the hml factor. Again, the highest systematic risk deciles also have the highest idiosyncratic and total risk averages. The most salient finding is that there is little pattern in future returns when sorting on systematic risk. Theory would intimate that sorting on purportedly undiversifiable risk would have produced the largest and most consistent spread in future returns, but the other three sorts are far superior in establishing a relationship between future returns and decile increments. I would note the peculiar decline in hml loadings in the highest systematic risk deciles, which may be tempering the observed future return spread.

A pattern emerges when sorting stocks by Total risk, which is by construction a sorting method insulated from a factor model and thus the corresponding calculations from Panels A trough C, that higher risk is associated with higher future returns. The return pattern peaks around deciles 7 and 8, then declines slightly moving to the highest deciles. The expected patterns between the factor loadings and risk components are consistent with intuition. This sort also shows the strongest monotonic relationship between Past 12-month return and Total risk. Once again, note the inverse relationship between sorting on Total risk and the calculated High-minus-Low coefficients. R-squared appears to be the most reliable risk-based sort related to future returns. Total Risk and Idiosyncratic Risk have increasing return patterns, and Systematic risk appears to perform poorly isolationto predict differences in future returns. To gain further insight into the quality of information contained in a calculated R-squared, I formed quintile portfolios and looked at future 12-month returns of R-squared quintiles within each of the three single risk measure quintiles. In Table 4, future 12-month returns for each R-squared quintile, and the corresponding Low-Minus-High R-squared quintile portfolios are compared with respect to idiosyncratic, systematic, and total risk quintiles (Panels A-C, respectively.)Focusing initially on Panel A, R-squared portfolio averages (standard errors in parentheses) decrease moving to higher R-squared quintiles, regardless of the idiosyncratic risk quintile. The highest two idiosyncratic quintiles have higher future 12 month return averages than the lowest two quintiles. Within the idiosyncratic risk quintiles, R-squared spreads increase with more idiosyncratic risk. One interpretation is that the role of systematic risk to reduce ambiguity becomes more highly valued as idiosyncratic risk rises. If we hold idiosyncratic risk constant, low R-squared means more Total Risk, and the two-way sorts would suggest that on average, higher Total risk is consistent with higher future returns. The highest two idiosyncratic Low-minus High portfolios have significantly positive results, and within the highest quintile of idiosyncratic risk, the annual spread is 12%.

Panel B examines R-squared quintiles holding systematic risk quintiles constant. R-squared portfolio averages (standard errors in parentheses) again decrease moving to higher R-squared quintiles, with the exception of the highest systematic risk quintile, and the idiosyncratic quintile Low-minus-High r-squared portfolios show significantly positive returns. Within the systematic risk quintiles, R-squared spreads get larger with lower absolute levels of systematic risk. Holding systematic risk constant, a lower R-squared again implies more Total risk. The systematic quintile averages do not show higher portfolio average returns with increasing systematic risk, but a closer look reveals that this is primarily driven by the lowest R-squared quintile. This is precisely where the systematic risk estimates are most unreliable (note the extremely large standard errors compared to other portfolio averages). The other four quintiles are consistent with an investor earning a return premium for bearing systematic risk.

The Total Risk quintiles in Panel C highlight the relationship between the relative proportion of the Total Risk that is explained by the past 60-month regression. Similar to the results from Panel A, this effect is most prominent at higher risk levels. The various sorts shed light on whether the R-squared relationship to the following 12-month's return is the result of spuriously sorting on another characteristic previously shown to have a relationship to expected returns. There is evidence that R-squared contains additional information related to future returns that is missed simply by looking at one risk measurement in isolation. Of the 15 single-risk based quintiles in the three panels of Table 4, 14 show a positive return for the Low-minus-High R-squared portfolio, of which 11 are significantly positive. If the interpretation is that R-squared proxies for the amount of confidence an investor can assign to understanding a stock's covariance structure, or potential risk within a portfolio, this understanding becomes more highly valued as overall past observed risk increases.

4.3 An Example of Abandoning a Pricing Model When its Output is Unreliable

In the situation where a pricing model creates an ambiguous output for portfolio creation, an investor must look elsewhere to assess a stock's return and risk potential. We could assume that at some level of uncertainty, a stock would be unattractive to an investor when its risk cannot be assessed, unless that investor is compensated in the form of a higher expected return. Unlike most asset pricing studies, this could easily be a reality for an individual investor or professional portfolio manager who holds less than a fully diversified portfolio, as the benefit of looking forward in time to gain additional information does not exist. For comparison purposes, let us assume that a stock is simply priced relative to its volatility. As mentioned previously, the intuition is that the level of observed total variance is a signal for the *potential* of undiversifiable risk, and proxies for some level of ambiguous systematic risk. The total risk heuristic pricing model has no components of risk decomposition invoked by a factor model. In addition to calculating an expected return for each stock using the Fama-French coefficient estimates known ahead of time to the investor, I also calculated an expected return for each stock using risk estimates known ahead of time to the investor. I then ran three sets of cross-sectional regressions for the entire sample and each R-squared quintile. The dependent variable in each specification is the stock's raw monthly return, and the two independent variables are the expected return from the Fama-French 3-factor model and the heuristic model. This allows for a relative performance comparison of each specification to explain monthly stock returns, and by using separate regressions within each R-squared quintile, we can see if the informativeness, or success of the Fama-French 3-factor model is related to its future ability to explain returns relative to the heuristic model.

For the pooled cross-section I use the Peterson (2006) double cluster methodology to calculate standard errors for the coefficients. It is reasonable to assume there is significant monthly cross-sectional heteroskedascticity, given within industry correlations and other factors leading to clustering of stocks' return behavior within any given month. Examining the Fama-MacBeth standard errors provides evidence of this cross-sectional heteroskedascticity. In addition, there is also reason to assume that the error terms would have a degree of serial correlation, in that any firm specific effect leading to estimation error is unlikely to decay immediately, yet it is also unlikely to last as the rolling regression adjusts its coefficients thereby adjusting the unexpected return over time. I used the test for serial correlation in the idiosyncratic errors of a linear panel-data model discussed by Wooldridge (2002). Drukker (2003) presents simulation evidence that this test has good size and power properties in reasonable sample sizes. In this case when the duration of firm specific effects is unknown, Peterson argues that clustering on two dimensions with a sufficient number of time and firms leads to the most accurate standard errors.

Table 5 shows that for the entire sample, the Fama-French 3-factor model expected return known to the investor in advance does a better job of explaining raw returns. On average. 12.44% of next month's return is explained by the Fama-French 3-factor model expected return, while 10.13% is explained by the heuristic model expected return. In instances where the past 60-month regression had low explanatory power, such as the lowest two quintiles of R-squared, the heuristic model actually outperforms the Fama-French 3-factor model in explaining expected returns. In the highest two quintiles, the opposite is shown. In all cases, combining the two models explains more of next month's return than either model used in isolation. By looking at both models and examining the relative coefficients by quintile, the lower quintiles load more heavily on the heuristic model expected return, and the higher quintiles load more heavily on the Fama-French 3-factor model expected return. This is consistent with some weight or consideration placed on the relative informativeness of past data and past covariance structure. The more potential for ambiguity, represented by the lack of explanatory power, the more past risk seems to be used as a benchmark for expected returns. The same intuition is confirmed when examining loadings based on the other three risk variables (not shown). In all cases, the story is confirmed that when the historical regression does not provide meaningful information, some "adjustment" takes place in assessing a stock's risk, and a corresponding premium for lack of understanding the risk seems to be apparent in realized future returns.

5. Alternative Measures of Model Uncertainty

As alternative measures of model uncertainty, I calculate two statistics for each month. The first is the sum of the absolute value of the T-statistics for the three Fama-French (1993) model coefficients run for the previous 60 months. The second is the sum of the standard errors of those coefficients. The first obviously is scaled by the point estimates of each coefficient. In Table 6, the three panels shows portfolios created at time *t* based on R-squared decile, Sum of T-statistic decile, and sum of standard errors decile, respectively. The variable of interest is the average alpha of these portfolios relative to the Fama-French (1993) 3-factor model for the following five years. This was chosen to have the property of 60 new observations for a 60-month regression.

Panel A highlights that the current R-squared decile predicts future alphas and the results are both statistically and economically significant. The average future alpha is 25 basis points, and the spread between the highest and lowest R-squared deciles is 57 basis points per month, or almost 7% per year. All but the middle two deciles are statistically significantly different from the average and conform to the hypothesized relationship.

Panels B and C repeat the same exercise with the initial portfolio formation based on the Sum of the absolute values of T-statistics and Sum of standard errors. The T-statistic decile sort produces virtually identical results for future alpha prediction. This holds for both economic significance and the statistical significance of the deciles. The opposite relationship holds for the Sum of standard errors deciles. Higher sums proxy for higher uncertainty, and lead to higher future alphas. Again, the economic magnitude in Panel C is similar to Panels A and B. Overall the various proxies for model uncertainty all tell the same story, that unreliable model output may be a priced premium in security pricing.

Table 7 attempts to compare the model information content proxies to idiosyncratic risk, to determine if the observed effect is independent of idiosyncratic risk. The regression is observed R-squared today predicting the Fama-French (1993) constant term 60 months from now for a 60-month time series model. Panel A shows that R-squared explains almost as much as idiosyncratic volatility in predicting future alphas. The first specification is R-squared alone, the second is idiosyncratic risk alone, and the third is both together. Each is significant in its own specification with similar explanatory power. Each retains its significance when combined and together they explain about 50% more variation in future alphas than used alone. Panel B shows a similar but slightly less pronounced effect of the Sum of T-statistics when compared to and used in conjunction

with idiosyncratic volatility. In Panel C the effect of the Sum of Standard errors outperforms idiosyncratic volatility in explanatory power, and renders is insignificant when used together in a regression. This is intuitive because the sum of standard errors does not control for the amount of variation that *is* explained. It is for this reason that R-squared appears to be the superior proxy for model success, and the most relevant measure for determining whether a pricing premium exists for the ambiguity created by relative lack of success.

6. Conclusion

R-squared provides a different set of information than either idiosyncratic risk or total risk arising from its ability to assess what is known *relative* to what is not known. This is not to say that high R-squared regression results are not prone to error, but to intimate that on average, the level of confidence in the model output, and therefore understanding the risk of a stock going forward, increases with that model's past success. This is particularly important at higher levels of overall stock risk because errors in the covariance structure can significantly affect portfolio results in less than fully diversified portfolios. Using R-squared as a proxy for the level of knowledge of a stock's risk in any given portfolio, I find that low levels of knowledge correspond to higher levels of return within various other risk measure sorts. In addition, in the absence of good past information, an overall ambiguity premium seems to be present when investors have little information on which to evaluate a stocks future risk. The findings are robust to alternative proxies for model uncertainty or ambiguity.

The general results I find are consistent with behavioral models that argue that idiosyncratic risk should demand a return premium, and consistent with rational models that argue costs related to diversification should result in a return premium associated with idiosyncratic risk. I present an alternative explanation that imprecision and relative uninformativeness of past data factor loading estimates may be the rational impetus driving a return premium due to preference arguments and an ambiguity premium. The ability to form portfolios with desired risk characteristics without an additional layer of ambiguity as to whether the risk and reward of a stock has been accurately assessed should be desirable to investors. Those stocks which make that process easier should have lower expected returns for their role in removing that ambiguity. Both explanations have similar conclusions, in that idiosyncratic risk could command a premium in the aggregate if the investing public does not diversify to the required extent, but I argue that the *relative* potential for idiosyncratic risk to create undesired portfolio outcomes is a more accurate metric by looking at what is known relative to what is unknown. I argue that R-squared is a stronger way to look at the components of risk that drive future stock returns. I find that R-squared outperforms both idiosyncratic and Total risk in predicting future returns.

References

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross section of volatility and expected returns, *Journal of Finance* 61, 259-299.

Avramov, Doron and Tarun Chordia, 2006, Asset pricing models and financial market anomalies, *Review of Financial Studies* 19, 1001-1040.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth, *Journal of Finance* 55, 773-806.

Barberis, Nicholas and Ming Huang, 2001, Mental accounting, loss aversion, and individual stock returns, *Journal of Finance* 56, 1247-1292.

Barry, Christopher B., and Stephen J. Brown, 1985, Differential information and security market equilibrium, *Journal of Financial and Quantitative Analysis* 20, 407-422.

Basu, Sanjoy, 1977, Investment performance of common stocks in relation to their price earnings ratios: a test of the efficient market hypothesis, *Journal of Finance* 32, 663-682.

Benartzi, Shlomo, 2001, Excessive extrapolation and the allocation of 401(k) accounts to company stock, *Journal of Finance* 56, 1747-1764.

Benartzi, Shlomo, and Richard H. Thaler, 2001, Naïve diversification strategies in defined contribution saving plan, *American Economic Review* 91, 79-98.

Black, Fisher and Myron Scholes, 1974, The effects of dividend yield and dividend policy on common stock prices and returns, *Journal of Financial Economics* 1, 1-22.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross section of expected stock returns, *Journal of Financial Economics* 49, 345-373.

Daniel, Kent and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns" *Journal of Financial Economics* 52, 1-33.

Easley, David. And Maureen O'Hara, 2004, Information and the cost of capital, *Journal of Finance* 59, 1553-1583.

Fama, Eugene and Kenneth R. French, 1992, The cross section of expected stock returns, *Journal of Finance* 47, 427-465.

Fama, Eugene and James D. MacBeth, 1973, Risk return, and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607-636.

Goyal, Amit and Pedro Santa-Clara, 2003, Idiosyncratic Risk Matters! *Journal of Finance* 58, 975-1007.

Green, Richard C., and Kristian Rydqvist, 1997, The valuation of nonsystematic risk and the pricing of Swedish lottery bonds, *Review of Financial Studies* 10, 447-480.

Hur, Gubernamn, 2001. Familiarity breeds investment, *Review of Financial Studies* 14, 659-680.

Jagannathan, Ravi, and Zhenyu Wang, 1998, An asymptotic theory for estimating beta pricing models using cross-sectional regressions, *Journal of Finance* 48, 65-91.

Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-92.

Jones, Charles M., and Matthew Rhodes-Kropf, 2003, The price of diversifiable risk in venture capital and private equity, Working Paper, Columbia University.

Klein, R, and V Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics.*

Klein, R, and V Bawa, 1977, The effect of limited information and estimation risk on optimal portfolio diversification, *Journal of Financial Economics.*

Lehmann, Bruce N., 1990, Residual Risk revisited, J*ournal of Econometrics* 105, 161-184.

Litner, John, 1965, Security prices, risk, and maximal gains from diversification, *Journal of Finance* 20, 587-615.

Litzenberger, Robert and Krishna Ramaswamy, 1979, Dividends, short-selling restrictions, tax-induced investor clienteles and market equilibrium, *Journal of Finance* 7, 163-196.

Longstaff, Francis A., 1989, Temporal aggregation and the continuous-time capital asset pricing model, *Journal of Finance* 44, 871-887.

Lo, Andrew and A. Craig McKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431-468.

Malkiel, Burton G., and Yexiao Xu, 2002, Idiosyncratic risk and security returns, Working paper, University of Texas at Dallas.

Merton, R, 1987, A simple model of capital market equilibrium with incomplete information, *The Journal of Finance.*

Pastor, Lubos and Robert Stambough, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642-685.

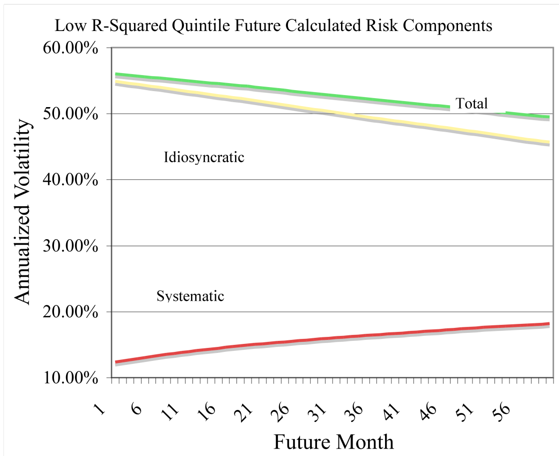
Petersen, Mitchell A., 2006, Estimating standard errors in finance panel data sets: comparing approaches, Working Paper, Northwestern University.

Shanken, Jay, 1992, On the estimation of beta pricing models, *Review of Financial Studies* 5, 1-33.

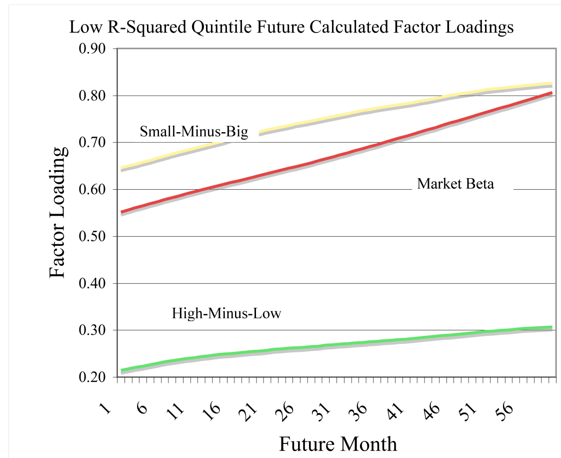
Tinic, Seha M., and Richard R. West, 1986, Risk, return, and equilibrium: A revisit, *Journal of Political Economy* 94, 126-147.

Figure 1:

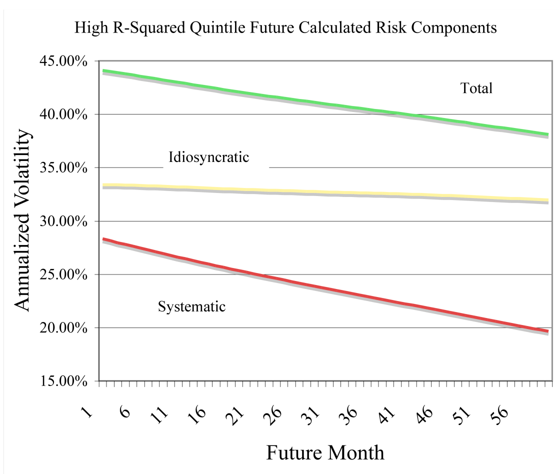
PANEL A:



PANEL B:



PANEL C:



PANEL D:

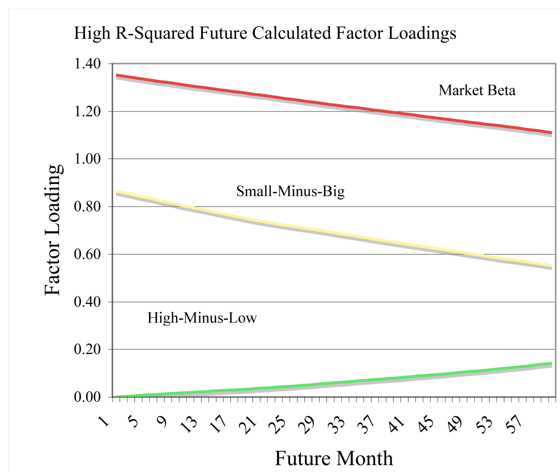


TABLE 1:

Simulations Based on equal Systematic and Unsystematic risk

R-squared Decile	Mean R-Squared	Average Decile Beta	Average Correlation to Market	Calculated Idiosyncratic Volatility	Calculated Systematic Volatility	Calculated Total Volatility	Average Expected Return	True Average Expected Return	Abnormal Return Caused By random Idiosyncratic Risk and Sorting
1 (Low)	0.277	0.76	-0.20	0.17	0.11	0.20	0.76%	1.00%	0.24%
2	0.345	0.87	-0.11	0.17	0.12	0.21	0.87%	1.00%	0.13%
3	0.380	0.92	-0.08	0.16	0.13	0.21	0.92%	1.00%	0.08%
4	0.408	0.94	-0.05	0.16	0.13	0.21	0.94%	1.00%	0.06%
5	0.431	0.98	-0.02	0.16	0.14	0.21	0.98%	1.00%	0.02%
6	0.451	1.02	0.01	0.16	0.14	0.21	1.02%	1.00%	-0.02%
7	0.471	1.04	0.03	0.15	0.15	0.21	1.04%	1.00%	-0.04%
8	0.496	1.09	0.08	0.15	0.15	0.22	1.09%	1.00%	-0.09%
9	0.531	1.15	0.13	0.15	0.16	0.22	1.15%	1.00%	-0.15%
10 (High)	0.588	1.21	0.20	0.14	0.17	0.22	1.21%	1.00%	-0.21%
TOTAL	**0.438**	**0.998**	**-0.002**	**0.158**	**0.140**	**0.212**	**1.00%**	**1.00%**	**0.00%**

*Based on 1,000 stocks for 60 months, generated randomly from identical iid distributions

TABLE 2:

Market Beta time t compared with Market Beta time t+60

	Time t		Time t+60				
	Obs	Original Beta grouping	25%ile	50%ile	75%ile	Inter Quartile Range	Standard Deviation
Low R-squared	2,764	0.55	0.526	0.912	1.320	0.794	0.723
High R-squared	139	0.55	0.278	0.329	0.399	0.121	0.139
Low	2,358	0.65	0.629	1.029	1.386	0.757	0.668
High	270	0.65	0.172	0.370	0.610	0.438	0.304
Low	2,121	0.75	0.677	1.097	1.519	0.842	0.771
High	918	0.75	0.349	0.532	0.754	0.405	0.306
Low	1,768	0.85	0.706	1.053	1.512	0.806	0.733
High	2,137	0.85	0.456	0.701	0.936	0.480	0.349
Low	1,271	0.95	0.710	1.119	1.495	0.785	0.702
High	3,775	0.95	0.555	0.810	1.073	0.518	0.381
Low	923	1.05	0.824	1.255	1.581	0.757	0.724
High	5,512	1.05	0.660	0.918	1.122	0.462	0.377
Low	684	1.15	0.782	1.141	1.618	0.836	0.721
High	7,496	1.15	0.742	0.977	1.203	0.461	0.382
Low	529	1.25	0.955	1.309	1.747	0.792	0.650
High	7,818	1.25	0.776	1.007	1.249	0.473	0.387
Low	334	1.35	1.162	1.482	1.865	0.703	0.782
High	7,073	1.35	0.827	1.076	1.313	0.486	0.400
Low R-squared	249	1.45	0.793	1.302	1.716	0.923	0.995
High R-squared	6,062	1.45	0.888	1.122	1.354	0.466	0.420

TABLE 3:

PANEL A: Descriptive Statistics of Deciles Sorted by R-squared

R-squared Decile	Next Month Raw Return	Next 3 Months Raw Return	Next 6 Months Raw Return	Next 12 Months Raw Return	Average Market Coeff	Average SMB Coeff	Average HML Coeff	Average Idiosyncratic Risk	Average Systematic Risk	Average Total Risk	Average Past 12 Month Return
1	1.55%	4.98%	10.48%	22.77%	0.57	0.74	0.18	51.51%	10.78%	53.03%	20.45%
2	1.40%	3.98%	8.54%	19.40%	0.69	0.77	0.25	51.59%	12.72%	53.58%	17.88%
3	1.37%	4.12%	8.62%	18.57%	0.83	0.85	0.29	47.96%	15.84%	50.89%	18.07%
4	1.37%	4.07%	8.29%	17.80%	0.92	0.90	0.29	44.68%	17.97%	48.48%	16.61%
5	1.29%	3.90%	8.17%	17.45%	1.00	0.93	0.29	42.47%	19.80%	47.16%	16.75%
6	1.34%	3.95%	8.11%	16.86%	1.06	0.94	0.26	40.23%	21.23%	45.77%	17.03%
7	1.23%	3.68%	7.74%	17.14%	1.12	0.95	0.23	38.28%	22.67%	44.75%	15.57%
8	1.19%	3.52%	7.26%	15.81%	1.20	0.95	0.19	36.78%	24.41%	44.39%	14.57%
9	1.17%	3.52%	7.13%	14.80%	1.28	0.89	0.09	34.50%	26.21%	43.54%	15.34%
10	1.08%	3.20%	6.50%	13.82%	1.43	0.85	-0.10	32.29%	30.76%	44.85%	16.44%
Low-High	0.47%	1.78%	3.98%	8.95%							
T-stat	1.03	1.97	2.80	3.73							
TOTAL	**1.28%**	**3.82%**	**7.92%**	**17.06%**	**1.04**	**0.89**	**0.20**	**41.39%**	**20.87%**	**47.27%**	**16.64%**

PANEL B: Descriptive Statistics of Deciles Sorted by Idiosyncratic Risk

Idiosyncratic Risk Decile	Next Month Raw Return	Next 3 Months Raw Return	Next 6 Months Raw Return	Next 12 Months Raw Return	Average Market Coeff	Average SMB Coeff	Average HML Coeff	Average Idiosyncratic Risk	Average Systematic Risk	Average Total Risk	Average Past 12 Month Return
1	1.09%	3.29%	6.79%	14.32%	0.72	-0.01	0.34	16.85%	11.03%	20.55%	13.44%
2	1.15%	3.40%	6.94%	14.43%	0.89	0.29	0.27	22.45%	14.35%	27.16%	13.65%
3	1.22%	3.65%	7.45%	15.28%	0.95	0.50	0.27	26.59%	16.02%	31.63%	14.37%
4	1.23%	3.74%	7.61%	16.02%	0.99	0.66	0.25	30.66%	17.49%	35.97%	14.64%
5	1.28%	3.77%	7.65%	16.47%	1.05	0.82	0.24	35.03%	19.36%	40.76%	14.74%
6	1.28%	3.90%	8.14%	17.63%	1.10	0.96	0.19	39.81%	21.37%	46.00%	15.80%
7	1.43%	4.38%	9.25%	20.20%	1.14	1.12	0.15	45.39%	23.41%	51.97%	15.97%
8	1.42%	4.31%	8.82%	19.37%	1.19	1.29	0.10	52.41%	25.57%	59.33%	17.03%
9	1.45%	4.23%	8.99%	19.98%	1.19	1.50	0.08	62.50%	27.98%	69.63%	19.75%
10	1.30%	3.61%	7.81%	17.97%	1.22	1.97	0.06	90.55%	34.66%	98.55%	29.08%
High-Low	0.21%	0.33%	1.02%	3.64%							
T-stat	0.43	0.34	0.68	1.52							
Total	**1.28%**	**3.82%**	**7.92%**	**17.06%**	**1.04**	**0.89**	**0.20**	**41.39%**	**20.87%**	**47.27%**	**16.64%**

TABLE 3: (continued)

PANEL C: Descriptive Statistics of Deciles Sorted by Systematic Risk

Systematic Risk Decile	Next Month Raw Return	Next 3 Months Raw Return	Next 6 Months Raw Return	Next 12 Months Raw Return	Average Market Coeff	Average SMB Coeff	Average HML Coeff	Average Idiosyncratic Risk	Average Systematic Risk	Average Total Risk	Average Past 12 Month Return
1	1.18%	3.71%	7.73%	16.79%	0.48	0.26	0.29	29.68%	6.38%	30.75%	14.62%
2	1.35%	3.88%	8.04%	16.99%	0.66	0.35	0.33	30.80%	10.36%	32.98%	14.84%
3	1.24%	3.73%	7.69%	16.58%	0.78	0.49	0.31	32.81%	13.08%	35.86%	15.61%
4	1.29%	3.86%	7.96%	17.19%	0.90	0.59	0.31	34.24%	15.40%	38.14%	15.70%
5	1.34%	4.07%	8.36%	17.45%	1.00	0.69	0.30	36.04%	17.64%	40.79%	15.91%
6	1.38%	4.02%	8.32%	17.60%	1.10	0.81	0.30	38.66%	20.02%	44.25%	16.85%
7	1.26%	3.84%	8.02%	17.00%	1.19	0.96	0.24	42.69%	22.78%	49.16%	15.99%
8	1.30%	3.78%	7.73%	17.10%	1.28	1.16	0.14	47.33%	26.38%	54.99%	15.71%
9	1.32%	3.85%	8.09%	18.12%	1.41	1.43	-0.03	53.17%	31.82%	62.84%	17.16%
10	1.20%	3.48%	7.27%	15.75%	1.62	2.14	-0.20	68.52%	44.89%	83.09%	24.07%
High-Low	0.02%	-0.23%	-0.46%	-1.04%							
T-stat	0.04	-0.23	-0.31	-0.45							
Total	**1.28%**	**3.82%**	**7.92%**	**17.06%**	**1.04**	**0.89**	**0.20**	**41.39%**	**20.87%**	**47.27%**	**16.64%**

PANEL D: Descriptive Statistics of Deciles Sorted by Total Risk

Total Risk Decile	Next Month Raw Return	Next 3 Months Raw Return	Next 6 Months Raw Return	Next 12 Months Raw Return	Average Market Coeff	Average SMB Coeff	Average HML Coeff	Average Idiosyncratic Risk	Average Systematic Risk	Average Total Risk	Average Past 12 Month Return
1	1.09%	3.28%	6.77%	14.43%	0.63	0.01	0.37	17.19%	9.34%	19.89%	13.36%
2	1.18%	3.52%	7.20%	14.98%	0.83	0.27	0.25	22.63%	13.19%	26.69%	13.75%
3	1.22%	3.67%	7.46%	15.32%	0.91	0.46	0.29	26.65%	15.09%	31.22%	14.40%
4	1.26%	3.70%	7.57%	15.88%	0.98	0.62	0.27	30.61%	16.80%	35.60%	15.03%
5	1.21%	3.79%	7.80%	16.90%	1.04	0.77	0.26	34.92%	18.64%	40.33%	14.94%
6	1.35%	4.00%	8.24%	17.72%	1.10	0.92	0.23	39.69%	20.78%	45.63%	15.18%
7	1.43%	4.38%	9.32%	20.31%	1.15	1.10	0.15	45.13%	23.21%	51.71%	16.25%
8	1.46%	4.27%	8.77%	18.85%	1.22	1.27	0.07	51.92%	25.89%	59.12%	17.25%
9	1.37%	4.17%	8.72%	19.84%	1.26	1.53	0.03	61.54%	29.57%	69.61%	19.03%
10	1.26%	3.47%	7.50%	17.01%	1.32	2.09	0.03	89.16%	38.29%	99.01%	28.69%
High-Low	0.17%	0.19%	0.73%	2.58%							
T-stat	0.33	0.19	0.47	1.07							
Total	**1.28%**	**3.82%**	**7.92%**	**17.06%**	**1.04**	**0.89**	**0.20**	**41.39%**	**20.87%**	**47.27%**	**16.64%**

TABLE 4:

PANEL A: Next 12 Month Return Two Way Sorts Between Idiosyncratic Risk and R-Squared

	R-Squared Quintile							
Idiosyncratic Risk Quintile	1	2	3	4	5	Total	1 Minus 5	T-Stat
1	15.86%	15.71%	14.25%	14.12%	13.76%	14.38%	2.10%	1.57
	1.08%	0.86%	0.81%	0.77%	0.79%	0.71%		
2	17.75%	16.22%	15.66%	15.63%	14.55%	15.64%	3.19%	1.76
	1.44%	1.10%	0.97%	1.00%	1.10%	0.92%		
3	18.38%	17.60%	17.84%	16.53%	15.39%	17.05%	2.99%	1.30
	1.53%	1.31%	1.27%	1.24%	1.71%	1.14%		
4	22.61%	20.96%	19.56%	19.74%	15.51%	19.79%	7.10%	2.39
	1.76%	1.56%	1.71%	2.15%	2.39%	1.53%		
5	22.52%	19.00%	18.09%	17.74%	10.91%	19.03%	11.61%	2.63
	2.30%	2.13%	2.69%	3.11%	3.76%	2.09%		
TOTAL	20.45%	18.18%	17.15%	16.47%	14.31%	17.06%	6.15%	3.33
	1.42%	1.16%	1.15%	1.13%	1.18%	1.11%		

*Standard errors in parentheses

PANEL B: Next 12 Month Return Two Way Sorts Between Systematic Risk and R-Squared

	R-Squared Quintile							
Systematic Risk Quintile	1	2	3	4	5	Total	1 Minus 5	T-Stat
1	19.83%	15.67%	14.17%	12.93%	11.88%	16.89%	7.95%	4.85
	1.24%	0.87%	0.79%	0.79%	1.07%	0.86%		
2	23.14%	17.94%	15.40%	14.81%	13.93%	16.88%	9.21%	4.06
	2.07%	1.20%	0.93%	0.88%	0.93%	0.96%		
3	20.49%	21.62%	18.07%	16.06%	14.32%	17.52%	6.16%	2.43
	2.37%	1.67%	1.28%	1.07%	0.90%	1.07%		
4	20.60%	17.70%	19.92%	16.95%	13.93%	17.05%	6.67%	1.91
	3.31%	1.99%	1.80%	1.37%	1.09%	1.32%		
5	14.55%	20.70%	17.42%	18.70%	14.83%	16.95%	-0.28%	-0.05
	5.31%	2.78%	2.42%	2.18%	2.12%	1.97%		
Total	20.45%	18.18%	17.15%	16.47%	14.31%	17.06%	6.15%	3.33
	1.42%	1.16%	1.15%	1.13%	1.18%	1.11%		

*Standard errors in parentheses

PANEL C: Next 12 Month Return Two Way Sorts Between Total Risk and R-Squared

	R-Squared Quintile							
Total Risk Quintile	1	2	3	4	5	Total	1 Minus 5	T-Stat
1	16.71%	15.79%	14.68%	14.07%	13.58%	14.71%	3.13%	2.34
	1.09%	0.84%	0.80%	0.76%	0.77%	0.71%		
2	18.28%	16.68%	15.35%	15.52%	14.04%	15.60%	4.24%	2.42
	1.43%	1.11%	0.99%	1.02%	1.01%	0.91%		
3	19.70%	18.84%	18.21%	16.18%	14.80%	17.31%	4.90%	2.31
	1.63%	1.43%	1.31%	1.21%	1.36%	1.13%		
4	23.83%	20.43%	19.63%	19.27%	15.85%	19.59%	7.98%	2.86
	2.01%	1.64%	1.74%	1.91%	1.94%	1.52%		
5	21.96%	18.89%	18.00%	18.53%	13.40%	18.49%	8.56%	2.01
	2.51%	2.22%	2.63%	2.84%	3.44%	2.14%		
Total	20.45%	18.18%	17.15%	16.47%	14.31%	17.06%	6.15%	3.33
	1.42%	1.16%	1.15%	1.13%	1.18%	1.11%		

*Standard errors in parentheses

TABLE 5:

Comparison of Fama French 3-factor and heuristic model performance and relative loadings in explaining cross section of returns

	Constant	Fama French 3-Factor Expected Return	Heuristic Expected Return	R-Squared
Full Sample	0.42% (4.13)	0.72 (30.20)		12.44%
	-0.43% (-3.20)		1.11 (30.74)	10.13%
	-0.15% (-1.79)	0.52 (17.01)	0.53 (14.38)	13.77%
R-squared quintile 1	1.07% (6.50)	0.42 (11.62)		1.93%
	0.41% (2.51)		0.71 (15.53)	3.20%
	0.43% (2.91)	0.16 (5.65)	0.59 (13.48)	3.39%
R-squared quintile 2	0.68% (5.16)	0.59 (17.50)		5.72%
	-0.09% (-0.59)		0.91 (22.14)	6.09%
	0.03% (0.24)	0.34 (10.01)	0.59 (14.29)	7.28%
R-squared quintile 3	0.47% (4.26)	0.70 (24.03)		10.59%
	-0.31% (-2.11)		1.07 (26.49)	9.09%
	-0.09% (-0.89)	0.48 (14.76)	0.54 (13.18)	11.89%
R-squared quintile 4	0.16% (1.55)	0.78 (23.69)		19.15%
	-0.74% (-5.18)		1.25 (31.77)	15.49%
	-0.39% (-4.22)	0.57 (12.85)	0.55 (9.97)	20.69%
R-squared quintile 5	0.07% (0.58)	0.85 (17.31)		31.53%
	-1.12% (-7.21)		1.51 (32.89)	22.72%
	-0.50% (-3.51)	0.67 (8.28)	0.53 (4.73)	33.00%

* Standard errors in parentheses double cluster Petersen (2006)

TABLE 6:

PANEL A: Next 5 Year Alpha based on Portfolios Sorted by R-Squared Decile

R-Squared Decile	Average T+60 3-factor Alpha	Standard Deviation	T-statistic (from Average)
1	0.585%	0.047%	7.05
2	0.496%	0.039%	6.26
3	0.392%	0.033%	4.33
4	0.311%	0.030%	2.06
5	0.286%	0.028%	1.29
6	0.245%	0.026%	-0.21
7	0.165%	0.025%	-3.45
8	0.129%	0.026%	-4.67
9	0.070%	0.026%	-6.83
10	0.010%	0.031%	-7.71
Sample Average	0.250%		

PANEL B: Next 5 Year Alpha based on Portfolios Sorted by Sum of Absolute Value of the T-Statistics Decile

Sum of Absolute Value of T-Stats Decile	Average T+60 3-factor Alpha	Standard Deviation	T-statistic (from Average)
1	0.611%	0.055%	6.55
2	0.458%	0.037%	5.61
3	0.361%	0.035%	3.17
4	0.331%	0.030%	2.68
5	0.258%	0.027%	0.30
6	0.210%	0.025%	-1.62
7	0.191%	0.025%	-2.38
8	0.117%	0.024%	-5.54
9	0.098%	0.024%	-6.35
10	0.033%	0.024%	-9.06
Total	0.250%		

PANEL C: Next 5 Year Alpha based on Portfolios Sorted by Sum of Coefficient Standard Errors Decile

Sum of Standard Errors Decile	Average T+60 3-factor Alpha	Standard Deviation	T-statistic (from Average)
1	0.094%	0.020%	-7.84
2	0.053%	0.023%	-8.58
3	0.070%	0.025%	-7.23
4	0.104%	0.028%	-5.22
5	0.171%	0.030%	-2.63
6	0.248%	0.034%	-0.06
7	0.417%	0.041%	4.06
8	0.458%	0.050%	4.14
9	0.541%	0.062%	4.69
10	0.653%	0.085%	4.73
Total	0.250%		

TABLE 7:

PANEL A: Comparing R-Squared with Idiosyncratic Volatility

Dependent Variable is 60 Month Future Alpha

	1	2	3
R-Squared	-0.0160		-0.0112
StdDev is 0.156	-16.90		-11.85
Idiosyncratic Volatility		0.0114	0.0084
StdDev is 0.225		11.63	8.48
Constant	0.712%	-0.181%	0.257%
	18.70	-5.57	4.92
R-Sqr	2.23%	2.41%	3.37%

PANEL B: Comparing Absolute Sum of T-Statistics with Idiosyncratic Volatility

Dependent Variable is 60 Month Future Alpha

	1	2	3
Absolute Sum of T-Stats	-0.0008		-0.0005
StdDev is 2.508	-14.95		-9.51
Idiosyncratic Volatility		0.0114	0.0312
0.225		11.63	9.02
Constant	0.764%		0.222%
	16.57	-5.57	3.91
R-Sqr	1.64%	2.41%	2.91%

PANEL C: Comparing Sum of Standard Errors with Idiosyncratic Volatility

Dependent Variable is 60 Month Future Alpha

	1	2	3
Sum of Std Errors	0.0030		0.0032
0.932	12.58		5.45
Idiosyncratic Volatility		0.0114	-0.0039
0.225		11.63	-0.48
Constant	-0.215%	-0.181%	-0.213%
	-6.58	-5.57	-6.53
R-Sqr	2.79%	2.41%	2.79%