Spring 2011

Essays in asset pricing: on testing asset-pricing anomalies and modeling stock returns using model pools

Michael Shane O'Doherty
University of Iowa

ESSAYS IN ASSET PRICING:

ON TESTING ASSET-PRICING ANOMALIES AND MODELING STOCK RETURNS USING MODEL POOLS

by

Michael Shane O'Doherty

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration
in the Graduate College of
The University of Iowa

May 2011

Thesis Supervisor: Associate Professor Ashish Tiwari

ABSTRACT

In this dissertation, I consider a range of topics in cross-sectional asset pricing. The primary research focus is twofold. First, I provide new insights on analyzing and testing capital market "anomalies" or patterns in equity returns that are not well explained by the traditional models used in the finance literature. Second, I propose and examine a methodology for pooling asset-pricing models to better characterize the cross section of stock returns.

The first chapter offers an explanation for the financial distress anomaly, i.e., the previously documented poor stock-price performance for financially distressed firms. I first show that market betas for distressed firms are highly volatile and tend to be low during bad economic times. After properly controlling for exposure to market risk, the low historical returns on these stocks are consistent with the conditional Capital Asset Pricing Model (CAPM). I then explain these findings through a theoretical model in which a levered firm's equity beta is negatively related to uncertainty about the unobserved value of its underlying assets. Empirical tests support the main predictions of the theory.

The second chapter proposes a hierarchical Bayes approach for evaluating and testing asset-pricing anomalies using individual firms as test assets. The empirical results indicate that much of the anomaly-based evidence against the CAPM is overstated. Anomalies are primarily confined to small stocks, few characteristics are robustly associated with CAPM alphas out of sample, and most firm characteristics do not contain unique information about abnormal returns.

Lastly, the third chapter proposes a new econometric methodology to combine predictive densities from a set of competing asset-pricing models to better characterize the cross section of stock returns. Using a variety of test portfolios, the optimal pool of models consistently outperforms the best individual model on both statistical and economic grounds.

Abstract Approved: ______________________________
Thesis Supervisor

____
Title and Department

____
Date

ESSAYS IN ASSET PRICING:

ON TESTING ASSET-PRICING ANOMALIES AND MODELING STOCK RETURNS USING MODEL POOLS

by

Michael Shane O'Doherty

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration
in the Graduate College of
The University of Iowa

May 2011

Thesis Supervisor: Associate Professor Ashish Tiwari

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Michael Shane O'Doherty

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Business Administration at the May 2011 graduation.

Thesis Committee: ______________________________
Ashish Tiwari, Thesis Supervisor

____
David Bates

____
Matthew Billett

____
N. Eugene Savin

____
Paul Weller

ACKNOWLEDGEMENTS

I would like to thank my advisor, Ashish Tiwari, for his guidance and encouragement. I have learned a great deal from him during my time at Iowa and could never repay his generosity. In addition, I would like to thank the other members of my thesis committee: David Bates, Matthew Billett, N. Eugene Savin, and Paul Weller. They have all provided invaluable support throughout my graduate studies.

The faculty and graduate students at the University of Iowa have also been responsible for a large part of my academic development, and I am grateful to everyone. In particular, I would like to thank my two coauthors, Scott Cederburg and Phil Davies, both of whom are great friends.

Finally, this dissertation would not have been possible without the love and support of my parents, Pat and Julie, and my sister Trish.

ABSTRACT

In this dissertation, I consider a range of topics in cross-sectional asset pricing. The primary research focus is twofold. First, I provide new insights on analyzing and testing capital market "anomalies" or patterns in equity returns that are not well explained by the traditional models used in the finance literature. Second, I propose and examine a methodology for pooling asset-pricing models to better characterize the cross section of stock returns.

The first chapter offers an explanation for the financial distress anomaly, i.e., the previously documented poor stock-price performance for financially distressed firms. I first show that market betas for distressed firms are highly volatile and tend to be low during bad economic times. After properly controlling for exposure to market risk, the low historical returns on these stocks are consistent with the conditional Capital Asset Pricing Model (CAPM). I then explain these findings through a theoretical model in which a levered firm's equity beta is negatively related to uncertainty about the unobserved value of its underlying assets. Empirical tests support the main predictions of the theory.

The second chapter proposes a hierarchical Bayes approach for evaluating and testing asset-pricing anomalies using individual firms as test assets. The empirical results indicate that much of the anomaly-based evidence against the CAPM is overstated. Anomalies are primarily confined to small stocks, few characteristics are robustly associated with CAPM alphas out of sample, and most firm characteristics do not contain unique information about abnormal returns.

Lastly, the third chapter proposes a new econometric methodology to combine predictive densities from a set of competing asset-pricing models to better characterize the cross section of stock returns. Using a variety of test portfolios, the optimal pool of models consistently outperforms the best individual model on both statistical and economic grounds.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INFORMATION RISK, CONDITIONAL BETAS, AND THE FINANCIAL DISTRESS ANOMALY

1.1 Introduction

In a recent article, Campbell, Hilscher, and Szilagyi (2008) show firms with a high probability of bankruptcy or default earn lower average returns than those with a low probability of financial distress in the post-1980 period.[1] Furthermore, this poor performance of distressed stocks is not explained by standard asset-pricing models, including the unconditional Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) and the three-factor model of Fama and French (1993). Portfolios of distressed stocks have high market betas, load heavily on SMB and HML, and have significantly negative unconditional alphas. These empirical findings have received considerable attention in the finance literature for at least two reasons. First, they contradict the notion that many cross-sectional anomalies are linked to a *premium* required by investors for exposure to non-diversifiable distress risk (e.g., Fama and French (1996)). Second, they imply financial distress should be included in the growing list of characteristic-based CAPM anomalies.

This paper shows these results are consistent with an intuitive, risk-based

[1]See Dichev (1998) and Griffin and Lemmon (2002) for additional evidence. Dichev (1998) uses Altman's (1968) Z-score and Ohlson's (1980) O-score to identify firms with a high likelihood of bankruptcy. He finds that portfolios of distressed stocks earn significantly lower than average returns. Griffin and Lemmon (2002) show that Dichev's (1998) results are driven by the exceptionally poor performance of stocks with high bankruptcy risk and low book-to-market ratios. More recently, Campbell, Hilscher, and Szilagyi (2008) estimate a dynamic panel model that includes both market and accounting data to measure the probability a firm enters bankruptcy, is delisted for financial reasons, or defaults over a given period. They then sort firms into portfolios based on this failure measure and find a negative relation between the probability of financial distress and average returns.

explanation. As indicated above, prior studies documenting the financial distress anomaly only control for risk in an unconditional framework and do not consider the potential for time variation in factor loadings or risk premiums. Jagannathan and Wang (1996), Cochrane (2001), and Lewellen and Nagel (2006), among others, demonstrate that stocks can exhibit large pricing errors relative to unconditional asset-pricing models even when a conditional version of the CAPM holds perfectly. In particular, a stock's conditional alpha might be zero, when its unconditional alpha is not, if its beta changes through time and is correlated with the equity premium.

Building on these arguments, I find that conditional market betas for distressed stocks are highly volatile, suggesting the unconditional risk measures used in prior studies are inappropriate. After properly controlling for the conditional risk-return relation, the low average returns for distressed stocks are consistent with the conditional CAPM. I also propose a simple economic mechanism that explains the observed time variation in market risk for distressed firms. The theoretical analysis and empirical evidence suggest that 'information risk,' that is, investor uncertainty about the underlying value of a firm's assets, has a pronounced impact on equity betas and required returns for distressed stocks.

I conduct the analysis in this paper in three steps. First, I revisit the empirical evidence on the financial distress anomaly. Using both Campbell, Hilscher, and Szilagyi's (2008) failure probability (CHS) and Ohlson's (1980) O-score to proxy for distress risk, I find portfolios of distressed stocks have low average returns. Moreover, a trading strategy that goes long the quintile of stocks with highest distress risk and short the quintile with lowest distress risk earns a significantly negative unconditional alpha regardless of the proxy used for financial distress. While these findings are consistent with the existing evidence in the literature, I further show the poor unconditional performance of distressed stocks is explained by the conditional CAPM. There is no significant difference in the CAPM alphas for high- and low-distress stocks once we allow for time variation in market risk. The results are robust across proxies for distress risk and several alternative estimation methodologies for the conditional CAPM.

I provide additional insight on these results by considering a decomposition of unconditional alpha in terms of conditional betas and risk premiums. As discussed in Lewellen and Nagel (2006), the conditional CAPM can only explain cross-sectional anomalies if long-short portfolio betas covary with the equity premium. More formally, if the conditional CAPM holds, we should observe an unconditional CAPM alpha for a given asset of $\alpha_i^u \approx \mathrm{cov}(\beta_{i,t}, \gamma_t)$, where $\beta_{i,t}$ is the asset's conditional beta and γ_t is the conditional market risk premium. I find that when examining the financial distress anomaly, long-short portfolio betas exhibit significant negative covariance with the equity premium. As such, conditional betas tend to adjust over time in a manner that can explain the negative unconditional alphas for distressed stocks. These observed patterns in portfolio betas turn out to be key to the economic story in this paper. The asset-pricing literature generally argues that the market risk premium is countercyclical.[2] The above result therefore implies that distressed stocks must have relatively low systematic risk during bad economic times when the

[2]See Fama and French (1989), Constantinides and Duffie (1996), Campbell and Cochrane (1999), Petkova and Zhang (2005), and Rapach, Strauss, and Zhou (2010).

price of risk is high. While these findings may initially seem counterintuitive, they end up having a simple explanation.

Second, I build on the model of Johnson (2004) to explain the observed variation in market risk for distressed stocks. The analysis highlights the importance of information risk in determining equity betas and expected returns. In the model, asset values are unobservable and investors receive noisy signals about the level of fundamentals. Within this theoretical environment, I show that increases in uncertainty about asset values lead to lower market betas and lower expected returns. These effects are magnified in firms with high leverage. Given that distressed firms also tend to be highly levered, the model suggests that increases in investor uncertainty about fundamentals during bad economic times can potentially account for the observed negative correlation between betas for distressed firms and the market risk premium.

Third, I formally test the implications of the model in the context of the financial distress anomaly. Following Johnson (2004), I use dispersion in analysts' earnings estimates as a measure of unpriced information risk. I hypothesize that (i) proxies for information risk for distressed stocks are positively related to estimates of the market risk premium; (ii) portfolio betas for stocks with a high probability of failure are negatively related to measures of investor uncertainty in the time series; and (iii) portfolio betas for stocks with a low probability of failure are unrelated to measures of investor uncertainty. The empirical evidence is largely consistent with the model's predictions. The results speak to the importance of considering information uncertainty when assessing the systematic risk of levered firms. More importantly, the model successfully reconciles the potentially counterintuitive fact that market risk for distressed stocks declines in bad economic times. I show that these observed declines in beta are simply an artifact of sharp increases in investor uncertainty about the fundamentals of distressed firms.

This paper contributes to the literature attempting to reconcile the financial distress anomaly. While Griffin and Lemmon (2002) interpret the low returns on distressed stocks as evidence of market mispricing, recent papers have proposed rational explanations. Chen, Novy-Marx, and Zhang (2010) develop an unconditional three-factor model that includes mimicking portfolios based on investment and productivity. They show this model can explain the low average returns for distressed firms. Chava and Purnanandam (2010) argue the poor performance of firms with high distress risk in the post-1980 period is simply the result of bad luck. They show the evidence of underperformance weakens when they extend the sample period back to 1953. George and Hwang (2010) propose a model in which firms that have high exposures to systematic risk and high distress costs choose lower levels of financial leverage as a safeguard and, thus, also have lower default probabilities. This choice results in a cross section of expected returns that is negatively related to leverage and default probabilities.[3] Garlappi and Yan (2010) also propose an equity valuation model that incorporates financial leverage and potential shareholder recovery upon the resolution of financial distress. In this model, the possibility of debt renegotiation drives a negative relation between leverage and equity betas in

[3]See Johnson, Chebonenko, Cunha, D'Almeida, and Spencer (2010) for a comment on George and Hwang (2010) and additional discussion of the impact of endogenous debt choice on expected stock returns.

firms with high default probabilities. The primary advantage of my approach is its simplicity. I show that the distress effect is completely consistent with the CAPM once we allow for sufficient time variation in market risk and propose a mechanism to account for the documented changes in equity betas.[4]

This article also contributes to a substantial literature on the effect of parameter uncertainty on asset prices and expected returns. Of particular relevance to this paper, Johnson (2004) offers a theoretical explanation for the negative relation between analysts' forecast dispersion and average returns based on the interpretation of dispersion as a proxy for uncertainty about asset valuation. Pástor and Veronesi (2003) show asset prices should increase with uncertainty about firm profitability. Cremers and Yan (2009) simplify the model in Pástor and Veronesi (2003) to analyze the impact of parameter uncertainty on bond prices. Finally, Korteweg and Polson (2009) calibrate the Leland (1994) model to assess the effect of uncertainty about firm asset value and asset volatility on corporate bond credit spreads. Consistent with my paper, they also find uncertainty about asset values increases significantly during times of market stress.

The paper is organized as follows. Section 1.2 introduces the conditional CAPM, outlines the empirical methodology, and describes the data. Section 1.3 presents results on the performance of distress-sorted portfolios relative to the conditional CAPM. Section 1.4 derives closed-form solutions for equity betas in the context of a theoretical valuation model. Section 1.5 presents formal tests of the

[4]Other relevant papers include Vassalou and Xing (2004), Garlappi, Shu, and Yan (2008), Avramov, Chordia, Jostova, and Phillipov (2009), Avramov, Cederburg, and Hore (2010), and Da and Gao (2010).

model. Section 1.6 concludes.

1.2 The conditional CAPM

Section 1.2.1 briefly introduces the conditional CAPM and presents a general formula for the unconditional CAPM pricing error when the conditional CAPM holds, Section 1.2.2 outlines the empirical methodology, Section 1.2.3 discusses the proxies for distress risk used in the paper, and Section 1.2.4 describes the sample.

1.2.1 Key features of the conditional CAPM

The unconditional CAPM is given by

$$\mathrm{E}[r_{i,t}] = \beta_i^u \gamma, \tag{1.1}$$

where $r_{i,t}$ is the excess return on asset i in period t, β_i^u is the asset's unconditional beta, and γ is the unconditional market risk premium. The unconditional alpha for asset i is then

$$\alpha_i^u = \mathrm{E}[r_{i,t}] - \beta_i^u \gamma. \tag{1.2}$$

The conditional version of the CAPM allows betas and the market risk premium to vary over time. Thus, expected excess returns in period t depend on information available at the end of period $t-1$:

$$\mathrm{E}_{t-1}[r_{i,t}] = \beta_{i,t}\gamma_t, \tag{1.3}$$

where $\beta_{i,t}$ is the conditional beta for asset i, and γ_t is the conditional market risk premium. Taking unconditional expectations of equation (1.3) yields

$$\mathrm{E}[r_{i,t}] = \mathrm{E}[\beta_{i,t}]\gamma + \mathrm{cov}(\beta_{i,t}, \gamma_t). \tag{1.4}$$

Equation (1.4) shows that the average excess return on asset i depends on the asset's average beta and on the covariance of the asset's conditional beta with the market risk premium. All else equal, stocks or portfolios that have higher betas in times when the price of risk is high earn higher average returns. Moreover, the general consensus in the asset-pricing literature is that the market risk premium tends to be high in recessionary periods that are associated with heightened macroeconomic risks and/or risk aversion. The conditional CAPM implies investors will require higher average returns for holding securities with betas that rise during these times when they particularly dislike risk.

In this paper, we are primarily concerned with explaining negative unconditional alphas for zero-cost portfolios that are long stocks with high bankruptcy risk and short stocks with low bankruptcy risk. Thus, we would like an expression for the unconditional CAPM alpha for asset i assuming the conditional version of the model holds. Substituting equation (1.4) into equation (1.2) yields

$$\alpha_i^u = \gamma\,(\mathrm{E}[\beta_{i,t}] - \beta_i^u) + \mathrm{cov}(\beta_{i,t}, \gamma_t). \tag{1.5}$$

Lewellen and Nagel (2006) derive an expression for β_i^u and substitute into equation (1.5) to obtain

$$\alpha_i^u = \left[1 - \frac{\gamma^2}{\sigma_m^2}\right]\mathrm{cov}(\beta_{i,t}, \gamma_t) - \frac{\gamma}{\sigma_m^2}\mathrm{cov}\left[\beta_{i,t}, (\gamma_t - \gamma)^2\right] - \frac{\gamma}{\sigma_m^2}\mathrm{cov}(\beta_{i,t}, \sigma_t^2), \tag{1.6}$$

where σ_t^2 is the conditional variance of the market risk premium, and σ_m^2 is the unconditional variance of the market risk premium. They then show that, for reasonable monthly parameter values, the squared Sharpe ratio for the market, γ^2/σ_m^2, and the second covariance term in equation (1.6) are negligible. Thus, we can approximate the unconditional CAPM alpha for asset i as

$$\alpha_i^u \approx \text{cov}(\beta_{i,t}, \gamma_t) - \frac{\gamma}{\sigma_m^2}\text{cov}(\beta_{i,t}, \sigma_t^2). \tag{1.7}$$

Equation (1.7) suggests the observed negative unconditional alphas for long-short distress portfolios are consistent with the conditional CAPM if conditional portfolio betas covary sufficiently negatively with the market risk premium and/or positively with market volatility.

1.2.2 Empirical methodology

I test the conditional CAPM on distress-sorted stock portfolios following the short-window regression methodology in Lewellen and Nagel (2006). These tests directly estimate conditional portfolio alphas and betas using a sequence of time-series CAPM regressions. Specifically, I estimate a separate CAPM regression each month, quarter, or half year using daily return data to obtain a time series of non-overlapping conditional portfolio alphas that spans the entire sample period. The regression model is

$$r_{i,t} = \alpha_i + \beta_{i,0} r_{m,t} + \beta_{i,1} r_{m,t-1} + \beta_{i,2}\left[(r_{m,t-2} + r_{m,t-3} + r_{m,t-4})/3\right] + \epsilon_{i,t}, \tag{1.8}$$

where $r_{i,t}$ is the excess return on portfolio i and $r_{m,t}$ is the excess market return on day t. The portfolio beta estimate is

$$\hat{\beta}_i = \hat{\beta}_{i,0} + \hat{\beta}_{i,1} + \hat{\beta}_{i,2}. \tag{1.9}$$

The regressions include lags of the market return to control for nonsynchronous returns, and the slopes on lags two through four are constrained to be equal. If the conditional CAPM holds for portfolio i, the average of the conditional alphas should be zero. I also consider whether the time-series variation in conditional portfolio betas is consistent with the relation in equation (1.7). That is, I test if the estimated portfolio betas covary with the market risk premium or with market volatility in a manner that might explain the large negative unconditional alphas for high-distress portfolios.

For robustness, I use three different window lengths (i.e., monthly, quarterly, and semiannual). This methodology only requires that CAPM regression parameters do not vary too much within each estimation window. The shorter the window length, the more confident we can be that portfolio betas are relatively stable. There is a tradeoff, however, as shorter window lengths can result in less precise parameter estimates.

More traditional time-series tests of the conditional CAPM model conditional betas and/or the market risk premium as functions of state variables. These methods require the econometrician to take a stance on an appropriate set of conditioning variables. This decision can be problematic because investors' information sets are inherently unobservable. As an example, Cooper and Gubellini (2009) show the choice of state variables can critically affect inferences about the conditional CAPM.[5] The primary advantage of Lewellen and Nagel's (2006) approach is that the econometrician does not have to choose any conditioning variables.

Several recent papers, however, pose challenges to the Lewellen and Nagel (2006) methodology. Li and Yang (2008) and Ang and Kristensen (2010) argue the

[5] Also see Ghysels (1998).

estimation windows should not be specified arbitrarily by the researcher and propose kernel regression techniques to use the data more efficiently. Similarly, Boguth, Carlson, Fisher, and Simutin (2009) point out that using contemporaneous estimates of portfolio betas when estimating conditional alphas can lead to an overconditioning bias. For robustness, I consider alternative estimation methodologies for the conditional CAPM, including a specification in which portfolio betas are modeled as a linear function of state variables and several specifications that address concerns raised in the recent literature. The results, reported in Appendix B, are qualitatively similar to those obtained using the Lewellen and Nagel (2006) approach.

1.2.3 Measures of financial distress

To ensure the results are robust to alternative proxies for distress risk, I consider two separate measures. The first proxy for distress is Campbell, Hilscher, and Szilagyi's (2008) failure probability (hereafter CHS). See Appendix A for details on the model and data construction. Using a dynamic panel specification, Campbell, Hilscher, and Szilagyi (2008) model the probability a firm files for bankruptcy, is delisted from an exchange for financial reasons, or receives a D rating from a leading credit rating agency over the next twelve months as a function of firm specific covariates. Using their model, Campbell, Hilscher, and Szilagyi (2008) document a significantly negative association between distress risk and average stock returns.

The CHS measure represents the state of the art in reduced-form bankruptcy forecast accuracy and is, thus, a logical choice for characterizing financial distress. CHS also has several advantages over potential alternatives. Notably, Shumway

(2001) shows several widely used static bankruptcy prediction models, including those of Altman (1968), Ohlson (1980), and Zmijewski (1984), introduce a selection bias because they are estimated with only one observation for each sample firm. In contrast, CHS is estimated from a panel that classifies each firm month as a separate observation. Other advantages of CHS are the inclusion of market-based predictor variables (e.g., Shumway (2001)) and the use of Chava and Jarrow's (2004) proprietary bankruptcy data.

One disadvantage of CHS is that this distress measure is estimated from firm failures that cover the period 1963 to 2003. This sample period has some obvious overlap with the period considered in my paper. To alleviate any concerns about look-ahead bias, I also use Ohlson's (1980) bankruptcy measure which becomes available in 1980. Despite its vintage, O-score remains a common proxy for financial distress in papers relating distress risk to equity returns, as well as in the general finance literature.[6]

1.2.4 Sample construction

Dichev (1998) and Campbell, Hilscher, and Szilagyi (2008) document the financial distress anomaly in the post-1980 period. I therefore restrict the sample period in this paper to January 1981 through December 2009. The sample includes all NYSE, Amex, and NASDAQ ordinary common stocks with (i) return data available on the CRSP daily file and (ii) data available to compute either CHS or O-score. Ohlson's (1980) O-score was developed explicitly for industrial companies,

[6] See Dichev (1998), Griffin and Lemmon (2002), Fama and French (2006a), Chava and Purnanandam (2010), and Chen, Novy-Marx, and Zhang (2010).

so I exclude financial firms (SIC codes between 6000 and 6999) from the sample.

The empirical tests use daily returns on distress-sorted portfolios. Each year at the beginning of January, I sort firms into five groups based on either CHS or O-score. The value-weighted portfolios are held for twelve months and then rebalanced. In many cases, I focus on the performance of a long-short hedge portfolio that is long the top quintile and short the bottom quintile of stocks based on a particular measure of distress risk. Following Campbell, Hilscher, and Szilagyi (2008), I exclude stocks with a share price below one dollar at the portfolio formation date. For cases in which a firm is delisted from an exchange during a given month, I replace any missing returns with the delisting returns provided by CRSP.[7] Data on the daily market return and risk-free rate are from Kenneth French's website.[8]

Depending on data availability, I also construct firm-level measures of size ($RSIZE$), book-to-market equity (BM), leverage (LEV), and information risk ($DISP1$ and $DISP2$). $RSIZE$ is the log ratio of market capitalization to the market value of the S&P 500 Index. BM is the ratio of the book value of equity to the market value of equity, where the book value of equity is defined as shareholders' equity if it is available, and as the difference between total assets and total liabilities, otherwise. LEV is the ratio of total liabilities to the market value of total assets, where the market value of assets is the book value of debt plus the market value of equity. $DISP1$ is analysts' forecast dispersion divided by the absolute value of the

[7]See Shumway (1997) for a discussion of delisting bias.

[8]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. I thank Kenneth French for making this data available.

mean forecast, and $DISP2$ is analysts' forecast dispersion divided by the market value of total assets. Forecast dispersion is computed from the IBES database as the month-end standard deviation of current-fiscal-year earnings estimates across analysts tracked by IBES. See Appendix A for additional details.

Finally, several of the empirical tests in this paper also require data on macroeconomic state variables. I use the default premium (DEF), dividend-to-price ratio (DP), dividend yield (DY), net equity expansion ($NTIS$), short-term interest rate (TB), and term premium ($TERM$). These variables are described in Appendix A.

1.3 Results

The next section revisits the prior literature on the unconditional performance of portfolios sorted on either CHS or O-score and then presents the paper's main results on the performance of these portfolios relative to the conditional CAPM.

1.3.1 Conditional CAPM alphas

Panel A of Table 1.1 reports the average returns from January 1981 to December 2009 for portfolios sorted on financial distress. I report estimates in percent per month (i.e., the daily average returns, unconditional alphas, and average conditional alphas are multiplied by 21). Using either proxy for distress, there is an inverse relation between average portfolio returns and distress risk. A zero-cost portfolio that is long the high-distress quintile and short the low-distress quintile earns an average return of -0.31% per month for the CHS sample and -0.27% per month for the O-score sample.

Panel B shows the financial distress anomaly relative to the unconditional CAPM.[9] Unconditional portfolio betas are monotonically increasing in distress risk for the CHS portfolios. Thus, the performance of distressed stocks appears even worse after adjusting for risk in an unconditional framework. The long-short CHS portfolio has an unconditional alpha of -0.67% per month (-8.04% per year), which is statistically significant at the 5% level (t-statistic of -2.19).

For the O-score portfolios, the relation between beta and financial distress is not strictly monotonic, but the long-short portfolio does have a positive unconditional beta. The unconditional alpha for this portfolio is -0.48% per month (-5.76% per year), which is also significant at the 5% level (t-statistic of -2.25). Thus, the financial distress anomaly is not sensitive to the proxy for distress risk. Campbell, Hilscher, and Szilagyi (2008) show portfolios of distressed stocks also have significant underperformance relative to other unconditional factor models, including the three-factor model of Fama and French (1993) and the four-factor model of Carhart (1997). I confirm these results in my sample (not reported).

Panel C of Table 1.1 shows the results for the conditional CAPM. As described above, I estimate a separate CAPM regression each month (M), quarter (Q), or half year (SA) to obtain a series of conditional alphas for each portfolio. Panel C reports the averages of these conditional alpha estimates. I test for statistical significance by using the time-series variability of the conditional alpha estimates

---

[9]The unconditional CAPM regressions are also estimated following equations (1.8) and (1.9).

to compute standard errors. None of the average alphas for the long-short portfolios is significantly different from zero at the 5% level. The average conditional alpha for the CHS hedge portfolio ranges from -0.37% to -0.17% per month, depending on the rolling estimation window length (t-statistics are between -1.07 and -0.55). Each of these estimates is noticeably smaller in magnitude than the -0.67% unconditional alpha reported in Panel B. The results for the long-short O-score portfolio are even more striking. The average conditional alphas range from -0.13% to 0.25% per month (t-statistics between -0.64 and 0.97) and are considerably larger than the -0.48% unconditional estimate in Panel B.

Figure 1.1 shows the time series of conditional (quarterly) alpha estimates for the long-short distress portfolios. The estimates are noisy, but seem to fluctuate around zero without any obvious time-series trends. Thus, the figure alleviates one potential concern related to the Lewellen and Nagel (2006) approach for testing the CAPM. If the conditional CAPM holds, then average conditional alphas are close to zero. The converse of this statement, however, is not true. For example, conditional alphas for a portfolio could be significantly positive over the first half of the sample period and significantly negative over the second half. The average conditional alpha would be close to zero, but we would not say the conditional CAPM provides an adequate description of portfolio returns. Figure 1.1 shows there are no such patterns in conditional alphas over time.

The results in Table 1.1 suggest the conditional CAPM is able to explain the poor unconditional performance of financially distressed stocks.[10] In Section 1.3.2,

[10]The main empirical findings in this paper are robust to using decile, rather than

I present more detail on why the conditional CAPM works so effectively.

1.3.2 Decomposing unconditional alphas

Panel A of Figure 1.2 presents a plot of the conditional (quarterly) beta for the long-short CHS portfolio, as well as a plot of the individual betas for the high- and low-distress portfolios. The long-short portfolio beta is remarkably volatile, ranging from a low of -1.06 in 1995 to a high of 4.09 in 2001. There are substantial periods of time in the late 1980s and mid 1990s in which the portfolio beta is less than zero, suggesting high-distress stocks often have less market risk than low-distress stocks. The plots of the individual portfolio betas reveal that most of the variation in the long-short beta is driven by the high-distress portfolio. The beta for the low-distress group is generally close to one. Panel B of Figure 1.2 plots the conditional betas for the O-score portfolios. The plots are qualitatively similar to those in Panel A. The long-short beta has a low of -0.92 in 1997 and a high of 1.55 in 2000.

The instability in portfolio betas seen in Figure 1.2 clearly suggests that the unconditional models previously used to characterize the performance of distressed firms are inappropriate. In the context of the conditional CAPM, however, volatility in beta only matters if there is also some meaningful correlation with either the expected market return or market volatility. More formally, equation (1.7) shows that the negative unconditional alphas reported in Table 1.1 for long-short distress portfolios are entirely consistent with the conditional CAPM if the portfolio betas

---

quintile, portfolios and considering the performance of a zero-cost portfolio that is long the high-distress decile and short the low-distress decile.

covary sufficiently negatively with the market risk premium and/or positively with market volatility. These quantities, $\text{cov}(\beta_{i,t}, \gamma_t)$ and $\text{cov}(\beta_{i,t}, \sigma_t^2)$, can be estimated directly from the data. The results are reported in Table 1.2.

Panel A presents covariance estimates that rely on estimated betas for the long-short distress portfolios. That is, I estimate $\text{cov}(\beta_{i,t}, \gamma_t)$ as $\text{cov}(\hat{\beta}_{i,t}, r_{m,t})$, where $\hat{\beta}_{i,t}$ is the estimated conditional beta for portfolio i and $r_{m,t}$ is the realized excess market return over the same estimation window as the beta. Similarly, I estimate $\text{cov}(\beta_{i,t}, \sigma_t^2)$ as $\text{cov}(\hat{\beta}_{i,t}, \hat{\sigma}_t^2)$, where $\hat{\sigma}_t^2$ is the realized variance of the market risk premium, calculated from daily returns over the same estimation window as the conditional beta. I then multiply this covariance estimate by $\hat{\gamma}/\hat{\sigma}_m^2$ to approximate the volatility effect in equation (1.7).

The results in Panel A show that the covariance between betas and the market risk premium can explain a large proportion of the negative unconditional alphas for long-short distress portfolios. The covariance estimates range from -0.40% to -0.20% per month for the CHS portfolio (although only one of the three is statistically significant at the 5% level) and from -0.62% to -0.30% for the O-score portfolio (all three are statistically significant). However, with the exception of the monthly window O-score estimate (-0.62%), none of the covariance estimates can fully explain the corresponding unconditional alpha reported in Table 1.1 (-0.67% for the CHS portfolio and -0.48% for the O-score portfolio). Moving to the covariances between conditional betas and market volatility reported in Panel A, all of the estimates are positive and significant at the 5% level. The estimated effects on unconditional alphas range from -0.16% to -0.13% per month for the CHS portfolio and from -0.10% to -0.09% for the O-score portfolio. Thus, the positive covariance between conditional betas and market volatility also appears to explain a portion of the negative unconditional alphas for the long-short portfolios.

Summing the risk premium effect and the volatility effect in Panel A, I find that the unconditional alpha should be between -0.56% and -0.36% per month for the long-short CHS portfolio and between -0.72% and -0.39% per month for the O-score portfolio (the numbers vary depending on the rolling estimation window length). Thus, the conditional betas for distressed stocks vary over time in a way that makes the unconditional CAPM alpha estimates of -0.67% for the CHS portfolio and -0.48% per month for the O-score portfolio seem quite reasonable.

For robustness, Panel B of Table 1.2 repeats the analysis using predicted portfolio betas. I estimate $\text{cov}(\beta_{i,t}, \gamma_t)$ as $\text{cov}(\hat{\beta}^*_{i,t}, r_{m,t})$ and $\text{cov}(\beta_{i,t}, \sigma^2_t)$ as $\text{cov}(\hat{\beta}^*_{i,t}, \hat{\sigma}^2_t)$, where $\hat{\beta}^*_{i,t}$ is the fitted value from a regression of $\hat{\beta}_{i,t}$ on its own lag, DEF_{t-1}, and DP_{t-1}. The results in Panel B using predicted betas are qualitatively similar to those in Panel A. The combined risk premium and volatility effects for the CHS portfolio range from -0.44% to -0.31% per month. The covariance estimates for the O-score portfolio imply an unconditional CAPM alpha between -0.30% and -0.27% per month.

1.3.3 Conditional betas and the market risk premium

The results in Table 1.2 suggest that the success of the conditional CAPM in explaining the financial distress anomaly is largely driven by the covariance of conditional betas with the market risk premium. Stocks with high measures of financial distress tend to have low market risk in states of the world in which the market risk premium is high. These findings may be counterintuitive under the assumption that a high market risk premium is associated with recessionary periods and relatively high investor risk aversion. However, as we will see below, these results are integral to the economic story in this paper.

Also note that the covariance estimates in Table 1.2 all rely on realized market returns to proxy for risk premiums. To solidify the relation between conditional betas and expected returns, we require a more formal characterization of the market risk premium. I build such a model following the forecast combination methodology in Rapach, Strauss, and Zhou (2010). Specifically, I estimate six separate regressions of the form

$$r_{m,t} = \delta_0 + \delta_1 Z_{t-1} + \epsilon_{m,t}, \tag{1.10}$$

where $r_{m,t}$ is the quarterly excess return on the CRSP stock market index and Z_{t-1} is a single lagged state variable known at the beginning of the period over which the excess market return is measured. The six Z_{t-1} variables are DEF_{t-1}, DP_{t-1}, DY_{t-1}, $NTIS_{t-1}$, TB_{t-1}, and $TERM_{t-1}$, each of which is an established predictor variable in the literature on forecasting the equity premium. The fitted values from the above regressions take the following form:

$$\hat{r}_{m,t} = \hat{\delta}_0 + \hat{\delta}_1 Z_{t-1}. \tag{1.11}$$

The estimate of market risk premium, $\hat{\gamma}$, is the average of the six fitted values associated with the six individual predictor variables.

Rapach, Strauss, and Zhou (2010) discuss the advantages of the forecast combination approach. Notably, this methodology combines information from several predictor variables that provide potentially unique signals about macroeconomic conditions. While individual predictive regressions often exhibit structural instability, performing well in some time periods and poorly in others, the combination forecast approach provides diversification benefits and produces estimates of the market risk premium that are much more economically plausible. The combination forecasts also perform better statistically in terms of mean squared prediction error relative to other alternatives in the literature. Finally, Rapach, Strauss, and Zhou (2010) show combination estimates of the market risk premium have strong linkages to the real economy. These forecasts tend to have local maxima (minima) during NBER business-cycle troughs (peaks) and are significantly correlated with future growth in several macroeconomic time series, including real GDP.

I use a long time series of data on the six predictor variables, from 1927 to 2009, to construct the combination estimate of the market risk premium. This resulting series, shown in Figure 1.3, appears economically reasonable. The market risk premium is always positive, ranging from 0.37% to 5.20% per quarter. The series shows pronounced peaks during the two recessionary periods of the Great Depression in the 1930s, as well as during the current global financial crisis. There are clear links to macroeconomic conditions, with the risk premium typically rising sharply during recessions as defined by the NBER.

To better characterize the dynamics of conditional betas for the distress-sorted portfolios, I examine average portfolio betas across economic states. I define economic states based on the combination estimates of the market risk premium in

Figure 1.3. Specifically, I divide the 1981 to 2009 sample period into three states: '$\hat{\gamma}$ Low,' '$\hat{\gamma}$ Intermediate,' and '$\hat{\gamma}$ High,' which correspond to the lowest third, middle third, and highest third of observations of the market risk premium, respectively.

Table 1.3 reports the average (quarterly) conditional betas for the distress-sorted portfolios in each state. The table also presents the p-value for a test of the null hypotheses that the average beta for a given portfolio is the same in states '$\hat{\gamma}$ Low' and '$\hat{\gamma}$ High.' Focusing on the CHS sample, the average beta for the high-distress portfolio in state '$\hat{\gamma}$ Low' is 1.46, but falls to 1.17 in state '$\hat{\gamma}$ High.' Thus, distressed stocks tend to be relatively riskier in good times when the price of risk is low. By comparison, the average beta for the low-distress portfolio rises from 0.87 in state '$\hat{\gamma}$ Low' to 0.95 in state '$\hat{\gamma}$ High.' Consequently, the long-short CHS portfolio is relatively risky in good times (average beta of 0.59), but much less so in bad times (average beta of 0.22). The numbers for the O-score sample are similar. Financially distressed stocks tend to have high market risk in good economic times (average long-short portfolio beta is 0.31), but relatively low betas in bad times (average long-short portfolio beta is 0.00). The test of the null hypothesis that average long-short betas are equal in states '$\hat{\gamma}$ Low' and '$\hat{\gamma}$ High' is rejected for both samples.

1.3.4 Discussion

In an influential paper, Lewellen and Nagel (2006) examine the size, value, and momentum effects and argue the conditional CAPM can not explain these anomalous cross-sectional patterns in stock returns. They then present formal estimates of covariances between long-short portfolio betas and both the market risk premium and market volatility, similar to the ones provided in Table 1.2. These estimates are either too close to zero or of the wrong sign to explain any of the three anomalies. The reader is left with a sense that time-varying exposure to market risk is not a sufficient mechanism to explain cross-sectional deviations from the unconditional CAPM.

The results in Tables 1.1 through 1.3 provide a stark contrast with those in Lewellen and Nagel (2006) on the size, value, and momentum anomalies. So what makes the financial distress effect different from the other CAPM anomalies previously considered in the literature? First, in contrast with any of the anomalies examined in Lewellen and Nagel (2006), the betas for the long-short distress portfolios covary positively with market volatility. That is, the second covariance term in equation (1.7) explains a small portion of the distress effect. Second and most importantly, however, the conditional betas for the long-short distress-sorted portfolios are highly volatile and negatively correlated with the market risk premium. Betas for zero-cost portfolios for the other anomalies do not fluctuate in any meaningful way with the expected market return.

The remainder of the paper provides a reasonable economic interpretation for the observed variation in equity betas for distressed firms. I start by considering the characteristics of the firms comprising the high- and low-distress portfolios. Table 1.4 presents summary statistics on firm size ($RSIZE$), book-to-market equity (BM), market leverage (LEV), and analysts' forecast dispersion ($DISP1$) for the distress-sorted portfolios. Panel A reports time-series averages of cross-sectional medians for each characteristic within the high- and low-distress groups. Consistent with the conjecture in Fama and French (1993) that book-to-market equity proxies for sensitivity to a systematic distress factor, the high-CHS group tends to be more tilted toward value stocks relative to the low-CHS portfolio. Interestingly, however, this association is not robust across proxies for financial distress. There is no apparent relation between BM and distress risk for the O-score sample. These results suggest that theories linking conditional equity beta to BM (e.g., Gomes, Kogan, and Zhang (2003) and Zhang (2005)) are unlikely to provide a sufficient explanation for the findings in Tables 1.1 to 1.3.

The summary statistics for the other three variables, however, motivate the proposed theoretical explanation for the dynamics in betas for distressed stocks. Following Johnson (2004), I interpret scaled dispersion in analysts' forecasts, $DISP1$, as a proxy for unpriced information risk, reflecting investor uncertainty about the fundamental value of a firm. More formally, consider an environment in which the value of a firm's assets follows some unobservable process. In this environment, there are two sources of risk facing investors: (i) systematic risk which is reflected in the covariance between changes in the value of the firm and changes in a stochastic discount factor process and (ii) information risk or investor uncertainty about the current value of the firm. Johnson (2004) argues dispersion in analysts' forecasts is a measure of the latter and presents a valuation model linking information risk and expected returns.

In the next section, I derive a closed-form relation between equity beta and information risk. In the model, equity betas and expected returns are inversely related to investor uncertainty about fundamentals and these effects are magnified with increases in firm leverage. Intuitively, when equity is viewed as a levered claim on the underlying value of the firm, raising uncertainty about fundamental value raises the option value of the claim and also lowers exposure to priced risk.

Panel A of Table 1.4 reveals distressed stocks tend to be much smaller than those with a low probability of failure and have higher market leverage. The table also shows that distressed stocks tend to have substantially higher $DISP1$. Following the interpretation outlined above, distressed stocks tend to be smaller firms that are considerably more difficult to assess and far less transparent than firms with a low probability of bankruptcy or default. More importantly, any sufficient explanation for the success of the conditional CAPM must capture the time-series variation in portfolio betas shown in Tables 1.2 and 1.3. Panel B of Table 1.4 suggests distressed stocks show substantial variation in $DISP1$ across economic states. Investor uncertainty about fundamentals for distressed stocks appears to increases sharply in times of high risk premiums. I argue in the next section that these fluctuations in information risk across the business cycle can reasonably account for the observed patterns in portfolio betas.

1.4 Model

Johnson (2004) presents a straightforward valuation model based on an environment in which a firm's asset value is unobservable and investors receive noisy signals about the level of fundamentals. The parameter governing the quality of these signals is analogous to dispersion in analysts' earnings forecasts. The model relies on the theory of unobserved state variables and the pricing of levered claims along the lines of Merton (1974). Johnson (2004) applies this model to analyze the cross-sectional relation between forecast dispersion and expected returns.

In this section, I extend the theory to explicitly consider a CAPM economy and the relation between information risk and equity betas. I briefly review the basic elements of the model in Section 1.4.1 and present closed-form expressions for asset prices, expected returns, and equity betas in Section 1.4.2. I consider several testable implications of the model in the context of the financial distress anomaly in Section 1.4.3.

1.4.1 Information environment

The model considers a single firm whose asset value, V_t, follows an unobservable geometric Brownian motion

$$\frac{dV_t}{V_t} = \epsilon dt + \sigma_V dW_t^V, \tag{1.12}$$

where ϵ is the known expected instantaneous rate of return on assets, σ_V is the instantaneous volatility, and W_t^V is a standard Brownian motion. Investors do not observe the true level of fundamentals, but are able to aggregate the information contained in analysts' forecasts of the firm's earnings rate to make inferences about V_t. That is, investors receive a signal, U_t, of the true value process corrupted by a stationary noise process:

$$U_t = V_t e^{\eta_t}. \tag{1.13}$$

Taking logs,

$$dv_t = \bar{\epsilon} dt + \sigma_V dW_t^V, \tag{1.14}$$

$$du_t = dv_t + d\eta_t, \tag{1.15}$$

where $\bar{\epsilon} \equiv \epsilon - \frac{1}{2}\sigma_V^2$. The noise process, η_t, is assumed to follow an Ornstein–Uhlenbeck process:

$$d\eta_t = -\kappa \eta_t dt + \sigma_\eta dW_t^\eta. \tag{1.16}$$

The parameters κ and σ_η determine the information quality of the signal, U_t, but the primary focus of the model is the impact of σ_η on asset prices and expected returns. Intuitively, we should see substantial heterogeneity both across firms and over time in terms of investor familiarity, predictability of operations, and information transparency. The empirical proxy for this uncertainty about fundamentals is dispersion in analysts' forecasts.[11]

In the model, the rate of return on assets also has a known systematic component and investors are able to use information on the aggregate state of the economy to make inferences about V_t. That is, the earnings process in equation (1.12) has a known correlation with the stochastic discount factor. While Johnson (2004) simply assumes a generic pricing kernel, I specify the following stochastic discount factor process that is consistent with the CAPM:

$$\frac{d\Lambda_t}{\Lambda_t} = -r dt - \sigma_M dW_t^M, \tag{1.17}$$

[11]The signal process presented in equation (1.15) is a simplified version of a model in which an investor receives N distinct analyst signals of the value process. Johnson (2004) shows that the considering a single aggregate signal, U_t, is sufficient. Moreover, if the innovations to the individual signals are independent, σ_η can be viewed as the dispersion across the forecasts.

where r is the risk-free rate and σ_M is market volatility. A derivation of the stochastic discount factor specified in equation (1.17) is provided in Appendix A. Finally, it is assumed that the variances and covariances of the processes specified above are known and the signal noise, dW_t^η, is uncorrelated with both dW_t^V and dW_t^M. The model solution follows the approach in Johnson (2004).

In summary, investors observe the processes U_t and Λ_t and use this information to update their beliefs about the unobserved state variable V_t. Let $m_t \equiv \mathrm{E}_t[v_t]$ and $\omega_t \equiv \mathrm{E}_t[(v_t - m_t)^2]$. Applying results on optimal nonlinear filtering, m_t follows the stochastic process given below

$$dm_t = \bar{\epsilon}dt + \tilde{h}_t d\tilde{W}_t, \tag{1.18}$$

where $\tilde{W}_t$ is a Brownian motion and $\tilde{h}_t$ converges to σ_V in the steady state. The posterior variance, ω_t, also converges to a steady-state value, ω, given by

$$\omega = \frac{\sigma_V^2}{\kappa}(1 - \rho_{VM}^2)\left[\sqrt{1 + \frac{\sigma_\eta^2}{\sigma_V^2(1 - \rho_{VM}^2)}} - 1\right], \tag{1.19}$$

where ρ_{VM} is the correlation between dW_t^V and dW_t^M. Assuming $\tilde{h}_t$ and ω_t have reached their steady-state values, the conditional distribution at time t about future values v_T is

$$N(m_t + \bar{\epsilon}\tau, \omega + \sigma_V^2\tau), \tag{1.20}$$

where $\tau \equiv T - t$. Thus, uncertainty about future asset values depends on both the true volatility of fundamentals and on the level of parameter uncertainty. Parameter uncertainty, in turn, is a function of the volatility of the noise process, σ_η, and the mean reversion parameter, κ.

1.4.2 Asset prices and equity betas

The asset pricing results assume that the value of the firm is paid out to equityholders at some future time T. The price, S_t, of an unlevered claim to the firm is

$$\begin{aligned} S_t &= \mathrm{E}_t[V_T \Lambda_T]/\Lambda_t \\ &= e^{-r\tau} e^{m_t + (1/2)\omega} e^{(\epsilon - \rho_{VM}\sigma_V\sigma_M)\tau}. \end{aligned} \tag{1.21}$$

For a levered firm with face value of debt K, equity is valued as a call option on the value of the firm along the lines of Merton (1974). The price, P_t, of equity for a levered firm is

$$P_t = S_t \Phi(d_1) - e^{-r\tau} K \Phi(d_2), \tag{1.22}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function and

$$\begin{aligned} d_1 &= \frac{\log(S_t/K) + r\tau + (\omega + \sigma_V^2 \tau)/2}{\sqrt{\omega + \sigma_V^2 \tau}}, \\ d_2 &= \frac{\log(S_t/K) + r\tau - (\omega + \sigma_V^2 \tau)/2}{\sqrt{\omega + \sigma_V^2 \tau}}. \end{aligned}$$

Applying Itô's lemma to equations (1.21) and (1.22), we can solve for the dynamics of S_t and P_t as follows:

$$\frac{dS_t}{S_t} = (r + \rho_{VM}\sigma_V\sigma_M)dt + \sigma_V d\tilde{W}_t, \tag{1.23}$$

$$\frac{dP_t}{P_t} = (r + \rho_{VM}\sigma_V\sigma_M \Phi(d_1)\frac{S_t}{P_t})dt + (\sigma_V \Phi(d_1)\frac{S_t}{P_t})d\tilde{W}_t. \tag{1.24}$$

From equations (1.23) and (1.24), we can readily derive the primary quantities of interest for this paper. The equity beta for the unlevered claim can be solved for as

$$\beta_S = \frac{\rho_{VM}\sigma_V}{\sigma_M}. \tag{1.25}$$

Note that for an unlevered firm, the parameter risk term, ω, has absolutely no effect on either equity beta or expected returns. Parameter risk, however, does have a pronounced impact on the beta for a levered claim

$$\begin{aligned} \beta_P &= \beta_S \frac{S_t}{P_t} \Phi(d_1) \\ &= \beta_S \frac{S_t}{P_t} \Phi\left(\frac{\log(S_t/K) + r\tau + (\omega + \sigma_V^2 \tau)/2}{\sqrt{\omega + \sigma_V^2 \tau}} \right). \end{aligned} \tag{1.26}$$

The relevant result for this paper is that β_P is *decreasing* in ω (and also σ_η). Increasing parameter risk raises investor uncertainty about the future value of the firm, V_T, that is to be paid out to investors at time T. Since parameter risk is unpriced, there is no impact on the asset beta or the required return for an unlevered claim. For a levered firm, however, an increase in parameter risk increases the option value of the claim and lowers the covariance between equity payoffs and the stochastic discount factor.

These effects are displayed graphically in Figure 1.4. I plot the theoretical relation between the volatility of the noise process, σ_η, and equity beta for various levels of firm leverage. In constructing the plots, I set the value of an unlevered claim, S, equal to 100 and follow the discussion in Johnson (2004) in selecting other reasonable parameter values. These values are provided in Table 1.5. In particular, Johnson (2004) argues that the range of σ_η displayed in Figure 1.4 is a realistic representation of investor uncertainty observed in the IBES data. The plots are intended to show qualitative patterns which are robust across a wide range of parameter choices.

Figure 1.4 shows that when the firm has no debt (i.e., $K = 0$), there is no relation between parameter risk and equity beta. For firms with leverage (i.e, $K > 0$), we see that equity beta decreases monotonically with parameter risk. These effects are significantly magnified as leverage increases. The simple theoretical relations in Figure 1.4 form the basis for the economic argument in this paper. In the next section, I formalize the testable implications of the model in the context of the financial distress anomaly.

1.4.3 Testable implications

The model presented above does not explicitly allow for time variation in σ_η. We would, however, like to apply the predictions of the model in a time-series setting to relate patterns in equity betas for distressed stocks to changes in parameter uncertainty. To some extent, this exercise is consistent with the model if we view the conditional CAPM as a model that holds period-by-period. At the same time, the closed-form solutions for asset prices and equity betas rely on the assumption that enough time has elapsed for certain model parameters to reach their steady-state values. With this caveat in mind, I build on the model's qualitative implications on the relation between equity beta, information risk, and leverage.

Finally, I also note that we can not make any causality statements about the relation between distress risk and equity beta. Intuitively, investor uncertainty about fundamentals could affect the firm's chosen level of debt which, in turn, would impact its probability of default. The model takes the level of debt, K, as given, and I analyze the impact of information risk on equity betas and returns in a partial equilibrium setting.

Figure 1.5 presents density histograms of market leverage, LEV, for the high- and low-distress portfolios sorted on both CHS and O-score. Given that both models of distress risk explicitly include a measure of firm leverage as a predictor variable, it is not surprising to see that firms with low probabilities of bankruptcy or default tend to have relatively low levels of leverage. While the distribution of leverage is relatively more dispersed within the high-distress groups, it is still quite common to see firms with market leverage of 0.60 or higher. According to the model in the previous section, these firms are exactly the ones for which parameter uncertainty should have the greatest impact on equity betas. Moreover, we saw some preliminary evidence in Table 1.4 that information risk for distressed stocks tends to peak during times of high expected market returns. Holding asset betas constant, these increases in investor uncertainty should be associated with decreases in measured equity betas. The analysis so far leads to the following testable empirical predictions:

1. Hypothesis 1: Proxies for information risk for distressed stocks should be positively related to estimates of the market risk premium. That is, investor uncertainty about the fundamentals of distressed firms should be higher in states of the world with high required returns.

2. Hypothesis 2: The portfolio betas for stocks with a high probability of failure should be negatively related to measures of investor uncertainty in the time series.

3. Hypothesis 3: The portfolio betas for stocks with a low probability of failure

should be unrelated to measures of investor uncertainty in the time series.

Hypothesis 1 is a relatively indirect test of the theory. If changes in information risk drive the observed time-series variation in the betas for distressed stocks, then according to the model it must be the case that information risk is more pronounced at times when equity betas for distressed firms are relatively low. Hypotheses 2 and 3 reflect the model's predicted relation between equity beta, parameter uncertainty, and leverage. Distressed firms tend to have high leverage and, as such, should exhibit an inverse association between equity beta and information risk over time. In contrast, firms with low distress risk have low leverage. For these firms, the theory suggests changes in information risk should have little, if any, impact on observed betas.

1.5 Empirical analysis

In Section 1.5.1, I describe the measures of information risk used in the empirical analysis. I then present formal tests of Hypotheses 1 to 3 in Sections 1.5.2 and 1.5.3.

1.5.1 Measures of information risk

I construct firm-level measures of information risk based on dispersion in analysts' earnings forecasts. Johnson (2004) discusses in detail the motivation for using dispersion as a proxy for nonsystematic risk related to investor uncertainty about underlying value of a firm's assets. He specifically cites the common practice in the social sciences to use disagreement across a survey of respondents as a measure of uncertainty in the underlying environment. Barron, Stanford, and Yu (2009) provide additional support for the argument that dispersion levels reflect idiosyncratic uncertainty that increases the option value of the firm. Their results directly contradict other potential interpretations, including the notion that forecast dispersion proxies for information asymmetry or disagreement among investors (e.g., Diether, Malloy, and Scherbina (2002)). Similarly, Güntay and Hackbarth (2010) examine the relation between dispersion of analysts' earnings forecasts and corporate bond credit spreads and conclude dispersion primarily reflects future cash flow uncertainty.

As described in Section 1.2.4 and Appendix A, I compute $DISP1$ and $DISP2$ for each firm-month with data available in the IBES database, where $DISP1$ is analysts' forecast dispersion divided by the absolute value of the mean forecast and $DISP2$ is analysts' forecast dispersion divided by the market value of total assets. Table 1.6 provides summary statistics on analysts' forecast dispersion and firm size over the period 1983 to 2009. Each year I compute the percentage of sample stocks with at least one valid monthly observation of dispersion (i.e., stocks covered by at least two analysts during one month of the year). Panel A reports the average, minimum, and maximum of these percentages across all sample years. In a typical year, roughly half of sample stocks have dispersion estimates, although the percentage of covered stocks ranges from 35.5% to 79.8%. Panel A also shows that the IBES database is heavily tilted toward big firms. Panels B and C of Table 1.6 report summary statistics for the CHS and O-score portfolios, respectively. Consistent with the evidence in Diether, Malloy, and Scherbina (2002), I find financially distressed firms tend to be under-represented in the IBES database. On average, only 25.0% (26.8%) of stocks in the highest CHS (O-score) group are covered. Within any given portfolio, there is also evidence that the covered firms have larger market capitalization.

Testing the hypotheses outlined in Section 1.4.3 requires us to convert the firm-level proxies for information risk into corresponding portfolio-level measures. Given the evidence in Table 1.6 that forecast dispersion is often unavailable for distressed firms, one potential concern is that any particular proxy for the high-distress group may not adequately reflect investor uncertainty about stocks in that portfolio. For robustness, I consider several alternatives.

For each distress-sorted portfolio, I compute the following four measures of investor uncertainty: $DISP1^{VW}$, $DISP1^{M}$, $DISP2^{VW}$, and $DISP2^{M}$. The measure $DISP1^{VW}$ ($DISP1^{M}$) for a given month is the value-weighted average (median) value of $DISP1$ across all stocks in the specified portfolio with a valid observation for $DISP1$. Similarly, $DISP2^{VW}$ ($DISP2^{M}$) is the value-weighted average (median) value of $DISP2$ across all stocks in the specified portfolio. I also construct two aggregate proxies for information risk: $DISP1^{AGG}$ and $DISP2^{AGG}$, where $DISP1^{AGG}$ ($DISP2^{AGG}$) for a given month is the median value of $DISP1$ ($DISP2$) across all sample stocks.

1.5.2 Time-varying information risk

Hypothesis 1 predicts a positive relation between information risk for distressed stocks and the market risk premium. Intuitively, if the model outlined above is to account for the inverse relation between market risk for distressed firms and the equity premium established in Section 1.3, it must be that investor uncertainty about fundamentals for distressed firms increases during times of high risk premiums. To test this prediction, I compare the portfolio-level measures of investor uncertainty across economic states. Economic states are again based on the combination estimates of the market risk premium defined in Section 1.3.3 (i.e., '$\hat{\gamma}$ Low,' '$\hat{\gamma}$ Intermediate,' and '$\hat{\gamma}$ High,' correspond to the lowest third, middle third, and highest third of observations of the market risk premium, respectively).

Panel A of Table 1.7 presents results for the high-distress CHS portfolio. The table shows the average of each information risk measure within each economic state. Regardless of the proxy for investor uncertainty, there is a pronounced increase in the value of the measure when moving from '$\hat{\gamma}$ Low' to '$\hat{\gamma}$ High.' For example, $DISP1^{VW}$ goes from 0.252 in times of low risk premiums to 0.948 in times of high premiums. The table also reports a p-value for each measure of information risk for the test of the null hypotheses that average is the same in states '$\hat{\gamma}$ Low' and '$\hat{\gamma}$ High.' The null hypothesis is strongly rejected in each case.

Panel B of Table 1.7 reports results corresponding to the high-distress O-score portfolio, and Panel C presents figures for the two aggregate uncertainty measures. In each case, the results are analogous to those in Panel A. Information risk tends to rise sharply with corresponding increases in the market risk premium. For additional insight, Figure 1.6 plots the relation between between $DISP1^{VW}$ and the equity premium for each of the high-distress portfolios.

Taken as a whole, the results in Table 1.7 provide strong support for Hypothesis 1. Uncertainty about fundamentals appears to rise substantially during times of market stress when investors are also demanding high risk premiums. I note these results are consistent with those based on alternative proxies in the literature. For example, Korteweg and Polson (2009) use Markov chain Monte Carlo methods to uncover the posterior distribution of a firm's asset value given observed prices, accounting data, and a structural model. They also find parameter uncertainty rises during bad economic times, particularly during the recent credit crisis.

1.5.3 Determinants of conditional betas

Hypotheses 2 and 3 are direct predictions based on the theoretical relation between equity beta, parameter uncertainty, and leverage. Given that firms in the high-distress portfolios are often highly levered, Hypothesis 2 conjectures an inverse relation between proxies for information risk and betas for distressed stocks. In contrast, Hypothesis 3 suggests we should see no relation between uncertainty about fundamentals and equity betas for the non-distressed portfolio.

Table 1.8 provides a time-series regression analysis of these predictions. In each model, the dependent variable is the quarterly conditional beta for either the high- or low-distress quintile portfolio. The explanatory variables are the portfolio-level proxies for information risk, median portfolio leverage, and lagged beta. Each of these variables is known at the beginning of the quarter over which beta is estimated.

Panel A presents results for the high-distress CHS portfolio. Models 1 to 4 are univariate regressions of conditional beta on the individual measures of uncertainty. The results support Hypothesis 2. Each of the four proxies for information risk is negatively associated with beta in the time series, and the coefficients for $DISP1^{VW}$, $DISP1^{M}$, and $DISP2^{VW}$ are significant at the 5% level using a one-tailed test. Models 5 and 6 add controls for leverage and/or lagged beta, which the theory implies could be important determinants of beta. Information risk remains significantly negatively related to equity beta for the high-CHS portfolio.[12] Moreover, the positive coefficients on LEV are also consistent with the theoretical relation in equation (1.26).[13] Models 7 to 9 present results for the low-distress portfolio. In each case, there is no significant association between information risk and conditional beta. These findings are consistent with Hypothesis 3.

Panel B of Table 1.8 presents corresponding regression results for the O-score portfolios. The conclusions are consistent with those in Panel A. In Models 1 to 6, information risk is significantly negatively related to conditional beta at the 5% level in all cases. Models 7 to 9 again reveal no significant association between investor uncertainty and equity betas for firms with a low probability of bankruptcy. Thus, the results in Table 1.8 provide strong empirical support for the model's predicted time-series relation between equity betas for distress-sorted portfolios and information risk.

1.6 Conclusion

In this paper, I show that a conditional version of the CAPM can explain the apparent underperformance of financially distressed stocks. Using the empirical

[12]For brevity, I only present results for $DISP1^{VW}$ in Models 5 to 9. The findings are qualitatively similar using the other three proxies for information risk.

[13]Also see Figure 1.4.

methodology of Lewellen and Nagel (2006), I find the average conditional alphas for long-short portfolios sorted on either Campbell, Hilscher, and Szilagyi's (2008) failure probability or Ohlson's (1980) O-score are not significantly different from zero. The conditional CAPM successfully explains the distress anomaly because conditional betas for financially distressed stocks are negatively correlated with the market risk premium and positively correlated with market volatility. In particular, betas for distressed stocks tend to decline in times when business conditions are weak and investors demand a large premium for bearing market risk.

I also show these results are consistent with a simple equity valuation model that incorporates investor uncertainty about underlying firm fundamentals. Specifically, building on the model of Johnson (2004), I demonstrate that equity betas for levered firms are negatively related to information risk. The predicted effects are economically significant. Moreover, the empirical results suggest the model goes a long way in explaining the observed time-series patterns in systematic risk for distressed stocks. Thus, this study not only helps to establish a rational explanation for the financial distress anomaly, but also further highlights the importance of parameter uncertainty in understanding the determinants of equity beta and the performance of levered firms.

Table 1.1: Average returns, unconditional CAPM regressions, and average conditional alphas, 1981–2009

	CHS portfolios						*O*-score portfolios					
Portfolio	L	2	3	4	H	H–L	L	2	3	4	H	H–L
Panel A: Average portfolio returns												
Mean	1.08	0.98	1.03	0.94	0.78	-0.31	0.99	1.09	1.02	1.11	0.72	-0.27
	(4.19)	(3.81)	(3.49)	(2.49)	(1.79)	(-0.94)	(3.29)	(4.23)	(4.12)	(3.81)	(1.95)	(-1.25)
Panel B: Unconditional CAPM regressions												
$\hat{\alpha}^u$	0.21	0.09	0.04	-0.21	-0.46	-0.67	0.04	0.20	0.11	0.11	-0.44	-0.48
	(2.56)	(1.26)	(0.50)	(-1.20)	(-1.77)	(-2.19)	(0.38)	(2.86)	(1.46)	(1.05)	(-2.19)	(-2.25)
$\hat{\beta}^u$	0.86	0.89	1.07	1.39	1.56	0.70	1.02	0.89	0.92	1.09	1.41	0.39
	(104.3)	(126.2)	(120.0)	(80.5)	(60.6)	(23.2)	(111.8)	(130.7)	(123.2)	(103.3)	(60.7)	(18.7)
Panel C: Average conditional CAPM alphas												
$\bar{\hat{\alpha}}^c$ (M)	0.19	-0.01	0.05	-0.10	-0.13	-0.32	-0.18	0.12	0.28	0.29	0.07	0.25
	(2.18)	(-0.13)	(0.49)	(-0.51)	(-0.42)	(-0.88)	(-1.72)	(1.55)	(3.05)	(2.59)	(0.27)	(0.97)
$\bar{\hat{\alpha}}^c$ (Q)	0.15	0.06	0.08	-0.14	-0.22	-0.37	-0.09	0.13	0.23	0.28	-0.18	-0.08
	(2.08)	(1.02)	(0.95)	(-0.88)	(-0.70)	(-1.07)	(-1.01)	(1.97)	(2.58)	(3.03)	(-0.76)	(-0.36)
$\bar{\hat{\alpha}}^c$ (SA)	0.14	0.06	0.09	0.01	-0.03	-0.17	-0.02	0.16	0.20	0.25	-0.15	-0.13
	(2.10)	(1.00)	(0.89)	(0.03)	(-0.11)	(-0.55)	(-0.18)	(2.55)	(2.33)	(2.29)	(-0.74)	(-0.64)

Note: The table presents average returns (Panel A), unconditional CAPM regression coefficients (Panel B), and average conditional CAPM regression alphas (Panel C) for portfolios sorted on financial distress. *CHS* is Campbell, Hilscher, and Szilagyi's (2008) failure probability. *O*-score is Ohlson's (1980) bankruptcy probability. The portfolios are value weighted and rebalanced annually. The regressions use daily returns and correct for nonsynchronous trading as discussed in the the text. Average returns and alphas are expressed in percent per month (i.e., the daily alpha estimates and average returns are multiplied by 21). The conditional alphas used to construct the averages in Panel C are estimated monthly (M), quarterly (Q), and semiannually (SA) using daily returns. The numbers in parentheses are *t*-statistics. 'H' is the high-distress quintile, 'L' is the low-distress quintile, and 'H–L' is their difference.

Table 1.2: Decomposing unconditional alphas, 1981–2009

	CHS portfolios						*O*-score portfolios					
	Monthly		Quarterly		Semiannual		Monthly		Quarterly		Semiannual	
Panel A: Covariances using estimated betas												
	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect
$\text{cov}(\beta, \gamma)$	-0.29	-0.29	-0.20	-0.20	-0.40	-0.40	-0.62	-0.62	-0.33	-0.33	-0.30	-0.30
	(-1.24)		(-0.97)		(-2.53)		(-3.31)		(-2.43)		(-2.68)	
$\text{cov}(\beta, \sigma^2)$	6.81	-0.13	7.55	-0.16	7.58	-0.16	4.95	-0.10	4.64	-0.10	4.32	-0.09
	(2.56)		(2.58)		(2.64)		(2.32)		(2.28)		(2.08)	
Total		-0.42		-0.36		-0.56		-0.72		-0.43		-0.39
Panel B: Covariances using predicted betas												
	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect
$\text{cov}(\beta, \gamma)$	-0.28	-0.28	-0.23	-0.23	-0.17	-0.17	-0.19	-0.19	-0.17	-0.17	-0.18	-0.18
	(-2.39)		(-1.89)		(-1.31)		(-1.98)		(-1.78)		(-1.91)	
$\text{cov}(\beta, \sigma^2)$	7.99	-0.16	8.25	-0.17	6.63	-0.14	5.43	-0.11	4.84	-0.10	4.47	-0.09
	(6.37)		(5.03)		(3.04)		(5.28)		(3.64)		(2.71)	
Total		-0.44		-0.40		-0.31		-0.30		-0.27		-0.27

Note: The table presents estimates of the covariance between excess market returns and conditional betas and the covariance between market volatility and conditional betas for portfolios sorted on financial distress. The estimates are for value-weighted hedge portfolios that are long the highest qunitile of stocks and short the lowest quintile for each variable. The covariance estimates in Panel A use estimated betas from short-window regressions. The covariance estimates in Panel B use fitted betas, which are based on regressions of estimated betas on state variables. The estimated covariances between excess market returns and conditional betas are reported in percent per month (i.e., the quarterly covariance estimate is divided by three and the semiannual estimate is divided by six). For the covariances between market volatility and conditional betas, 'Effect' is an estimate of the quantity $-\frac{\gamma}{\sigma_m^2}\text{cov}(\beta_{i,t}, \sigma_t^2)$. The numbers in parentheses are t-statistics.

Table 1.3: Average conditional betas across economic states, 1981–2009

		CHS portfolios			*O*-score portfolios		
Economic state	Risk premium	L	H	H–L	L	H	H–L
$\hat{\gamma}$ Low	1.07	0.87	1.46	0.59	1.11	1.41	0.31
		(0.03)	(0.12)	(0.14)	(0.03)	(0.08)	(0.08)
$\hat{\gamma}$ Intermediate	1.67	0.97	1.31	0.34	1.07	1.40	0.33
		(0.02)	(0.10)	(0.11)	(0.03)	(0.08)	(0.09)
$\hat{\gamma}$ High	2.43	0.95	1.17	0.22	1.04	1.04	0.00
		(0.03)	(0.08)	(0.10)	(0.03)	(0.05)	(0.07)
p-value		0.071	0.047	0.040	0.103	0.000	0.004

Note: The table presents average conditional betas in different economic states for portfolios sorted on financial distress. Economic states are determined by sorting on the average fitted value from six separate conditional market regressions of the form $r_{m,t} = \delta_0 + \delta_1 Z_{t-1} + \epsilon_{m,t}$, where $r_{m,t}$ is the excess market return and Z_{t-1} is a state variable (i.e., DEF_{t-1}, DP_{t-1}, DY_{t-1}, $NTIS_{t-1}$, TB_{t-1}, or $TERM_{t-1}$) known at the beginning of the period over which the excess market return is measured. The market regressions use quarterly data, and the state variables are defined in the text. '$\hat{\gamma}$ Low' corresponds to the lowest third of observations of the market risk premium, '$\hat{\gamma}$ Intermediate' corresponds to the middle third, and '$\hat{\gamma}$ High' corresponds to the highest third. The betas are estimated quarterly using daily data. The last line reports the p-value for the test of the null hypotheses that the average beta for a given portfolio is the same in states '$\hat{\gamma}$ Low' and '$\hat{\gamma}$ High.' *CHS* is Campbell, Hilscher, and Szilagyi's (2008) failure probability. *O*-score is Ohlson's (1980) bankruptcy probability. The portfolios are value weighted and rebalanced annually. 'H' is the high-distress quintile, 'L' is the low-distress quintile, and 'H–L' is their difference.

Table 1.4: Characteristics of distress-sorted portfolios, 1981–2009

	CHS portfolios								*O*-score portfolios							
	RSIZE		*BM*		*LEV*		*DISP*1		*RSIZE*		*BM*		*LEV*		*DISP*1	
Panel A: Unconditional characteristics																
	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H
	-9.36	-11.89	0.48	0.72	0.20	0.50	0.042	0.267	-9.62	-11.65	0.46	0.45	0.14	0.42	0.050	0.189
Panel B: Characteristics by economic state																
	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H
$\hat{\gamma}$ Low	-9.74	-12.33	0.43	0.60	0.19	0.41	0.034	0.160	-10.06	-11.97	0.40	0.34	0.11	0.30	0.043	0.124
$\hat{\gamma}$ Intermediate	-9.30	-11.88	0.47	0.69	0.19	0.49	0.036	0.227	-9.62	-11.61	0.45	0.41	0.13	0.41	0.041	0.155
$\hat{\gamma}$ High	-9.05	-11.48	0.55	0.87	0.22	0.60	0.055	0.420	-9.18	-11.39	0.53	0.59	0.17	0.56	0.068	0.294

Note: The table presents summary statistics for the following variables: the log ratio of market capitalization to the market value of the S&P 500 Index (*RSIZE*), the ratio of total liabilities to the market value of total assets (*LEV*), the book-to-market ratio (*BM*), and the standard deviation of current-fiscal-year earnings forecasts divided by the absolute value of the mean forecast (*DISP*1). The figures reported are time-series means of cross-sectional medians within portfolios sorted on financial distress. Panel A reports averages across all sample months. Panel B reports averages for separate economic states based on the estimated market risk premium. *CHS* is Campbell, Hilscher, and Szilagyi's (2008) failure probability. *O*-score is Ohlson's (1980) bankruptcy probability. 'H' is the high-distress quintile, and 'L' is the low-distress quintile. The sample period for *DISP*1 is restricted to 1983–2009 based on data availability.

Table 1.5: Parameter values

Parameter	Symbol	Value
Price of unlevered claim to V	S	100
Beta for unlevered firm	β_S	0.70
Risk-free rate	r	0.04
Market volatility	σ_M	0.25
Volatility of fundamentals	σ_V	0.20
Mean reversion parameter in noise process	κ	0.10
Cash-flow horizon	τ	3.00
Face value of debt	K	Variable
Volatility of noise process	σ_η	Variable

Note: The table reports the parameter values used to construct the plots of equity beta in Figure 4. The model parameters are discussed in detail in Section 1.4.

Table 1.6: Summary statistics on analysts' forecast dispersion, 1983–2009

	Avg. % of stocks with estimates	Minimum %	Maximum %	$RSIZE$ stocks with estimates	$RSIZE$ stocks in portfolio
Panel A: Statistics for full sample					
	50.6	35.5	79.8	-9.446	-10.487
Panel B: Statistics for CHS portfolios					
L	63.8	49.7	79.6	-8.914	-9.458
2	65.6	48.9	87.5	-8.999	-9.391
3	57.0	40.5	84.7	-9.521	-10.152
4	43.5	29.0	76.9	-10.141	-11.034
H	25.0	12.2	71.1	-10.795	-11.974
Panel C: Statistics for O-score portfolios					
L	66.0	54.2	84.9	-9.173	-9.710
2	62.3	51.1	85.8	-9.013	-9.646
3	55.4	38.4	85.1	-9.285	-9.987
4	44.8	27.6	76.5	-9.889	-10.885
H	26.8	10.4	71.5	-10.550	-11.739

Note: The table presents descriptive statistics on analyst coverage. Panel A reports statistics for all sample stocks, i.e., all NYSE, Amex, and NASDAQ ordinary common stocks (excluding financial firms) with data available to compute either CHS or O-score and a share price of at least one dollar at the beginning of January in a given year. Panel B reports statistics for portfolios sorted on CHS, and Panel C reports statistics for portfolios sorted on O-score. Analysts' forecast dispersion is the standard deviation of current-fiscal-year earnings estimates across analysts tracked by IBES. A firm must be covered by at least two analysts to have a valid observation of dispersion. $RISZE$ is the log ratio of market capitalization to the market value of the S&P 500 Index. For each group in each calendar year, I compute the percentage of stocks with at least one valid monthly observation of analysts' forecast dispersion. The table reports the average, minimum, and maximum of these percentages across all sample years. The table also reports the time-series averages of the cross-sectional medians of $RISZE$ (measured at the portfolio formation date) across firms with dispersion estimates and across all firms in the indicated group. CHS is Campbell, Hilscher, and Szilagyi's (2008) failure probability. O-score is Ohlson's (1980) bankruptcy probability. The distress-sorted portfolios are formed at the beginning of January and rebalanced annually. 'H' is the high-distress quintile, and 'L' is the low-distress quintile.

Table 1.7: Time-varying information risk, 1983–2009

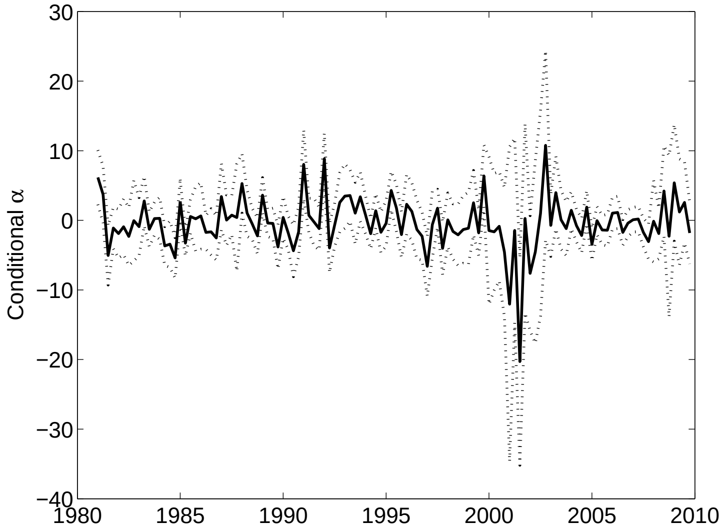
	$DISP1^{VW}$	$DISP1^{M}$	$DISP1^{AGG}$	$DISP2^{VW}$	$DISP2^{M}$	$DISP2^{AGG}$
Panel A: Information risk for CHS, H portfolio						
$\hat{\gamma}$ Low	0.252	0.160		0.0048	0.0056	
$\hat{\gamma}$ Intermediate	0.501	0.228		0.0077	0.0062	
$\hat{\gamma}$ High	0.948	0.420		0.0157	0.0106	
p-value	0.000	0.000		0.000	0.000	
Panel B: Information risk for O-score, H portfolio						
$\hat{\gamma}$ Low	0.269	0.124		0.0064	0.0055	
$\hat{\gamma}$ Intermediate	0.358	0.155		0.0066	0.0049	
$\hat{\gamma}$ High	0.601	0.294		0.0108	0.0079	
p-value	0.000	0.000		0.000	0.000	
Panel C: Information risk across all stocks						
$\hat{\gamma}$ Low			0.050			0.0018
$\hat{\gamma}$ Intermediate			0.056			0.0021
$\hat{\gamma}$ High			0.089			0.0038
p-value			0.000			0.000

Note: The table presents average measures of information risk in different economic states. Each proxy for information risk is a monthly, portfolio-level time series. $DISP1$ is the standard deviation of current-fiscal-year earnings forecasts divided by the absolute value of the mean forecast. The portfolio-level measure, $DISP1^{VW}$ ($DISP1^{M}$), for a given month is the value-weighted average (median) value of $DISP1$ across all stocks in the specified portfolio with a valid observation for $DISP1$. $DISP2$ is the standard deviation of current-fiscal-year earnings forecasts divided by the market value of assets. The portfolio-level measure, $DISP2^{VW}$ ($DISP2^{M}$), for a given month is the value-weighted average (median) value of $DISP2$ across all stocks in the specified portfolio. Similarly, $DISP1^{AGG}$ ($DISP2^{AGG}$) for a given month is the median value of $DISP1$ ($DISP2$) across all sample stocks. Economic states are determined by sorting on the average fitted value from six separate conditional market regressions of the form $r_{m,t} = \delta_0 + \delta_1 Z_{t-1} + \epsilon_{m,t}$, where $r_{m,t}$ is the excess market return and Z_{t-1} is a state variable (i.e., DEF_{t-1}, DP_{t-1}, DY_{t-1}, $NTIS_{t-1}$, TB_{t-1}, or $TERM_{t-1}$) known at the beginning of the period over which the excess market return is measured. The market regressions use quarterly data, and the state variables are defined in the text. '$\hat{\gamma}$ Low' corresponds to the lowest third of observations of the market risk premium, '$\hat{\gamma}$ Intermediate' corresponds to the middle third, and '$\hat{\gamma}$ High' corresponds to the highest third. Panel A reports average monthly measures of information risk in different economic states for the high-distress quintile of stocks sorted on CHS. Panel B reports average monthly measures of information risk in different economic states for the high-distress quintile of stocks sorted on O-score. Panel C reports average aggregate monthly measures of information risk in different economic states. The table also reports a p-value for each measure of information risk for the test of the null hypotheses that average is the same in states '$\hat{\gamma}$ Low' and '$\hat{\gamma}$ High.'

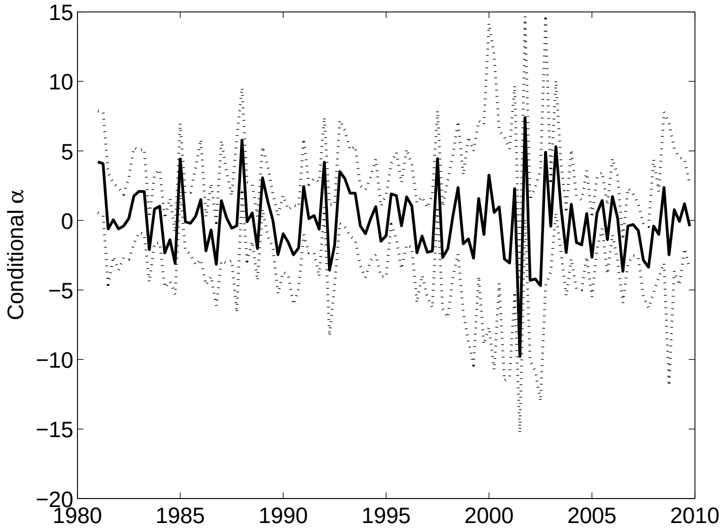
Table 1.8: Determinants of conditional betas, 1983–2009

	$\hat{\beta}_H$						$\hat{\beta}_L$		
Model	1	2	3	4	5	6	7	8	9
Panel A: CHS portfolios									
Intercept	1.60	1.71	1.50	1.53	1.11	0.56	0.93	1.15	0.56
	(14.8)	(14.3)	(13.5)	(9.97)	(4.06)	(2.04)	(25.5)	(12.7)	(4.77)
$DISP1^{VW}$	-0.48				-0.69	-0.52	0.01	0.15	0.03
	(-3.02)				(-3.63)	(-2.91)	(0.02)	(0.34)	(0.07)
$DISP1^{M}$		-1.36							
		(-3.70)							
$DISP2^{VW}$			-17.6						
			(-1.87)						
$DISP2^{M}$				-26.2					
				(-1.47)					
LEV					1.26	1.08		-1.17	-0.70
					(1.96)	(1.84)		(-2.58)	(-1.78)
$\hat{\beta}_{i,t-1}$						0.40			0.54
						(4.68)			(6.57)
Panel B: O-score portfolios									
Intercept	1.63	1.60	1.63	1.55	1.95	1.24	1.05	1.11	0.78
	(19.5)	(19.2)	(14.6)	(11.7)	(15.9)	(5.97)	(24.6)	(13.6)	(5.80)
$DISP1^{VW}$	-0.80				-0.43	-0.33	0.40	0.48	0.30
	(-4.77)				(-2.25)	(-1.85)	(0.87)	(1.03)	(0.66)
$DISP1^{M}$		-1.55							
		(-4.41)							
$DISP2^{VW}$			-40.8						
			(-3.32)						
$DISP2^{M}$				-41.0					
				(-2.10)					
LEV					-1.19	-0.70		-0.55	-0.30
					(-3.43)	(-2.03)		(-0.88)	(-0.49)
$\hat{\beta}_{i,t-1}$						0.37			0.29
						(4.15)			(3.06)

Note: The table presents time-series regressions of conditional betas on explanatory variables. The conditional betas are estimated quarterly using daily data. $\hat{\beta}_H$ is the conditional beta for the high-distress quintile of stocks sorted on either CHS (Panel A) or O-score (Panel B). $\hat{\beta}_L$ is the conditional beta for the low-distress quintile. The explanatory variables are all know at the beginning of the quarter over which the betas are estimated. $DISP1^{VW}$ ($DISP1^{M}$) is the value-weighted average (median) value of $DISP1$ across all stocks in the specified portfolio, $DISP2^{VW}$ ($DISP2^{M}$) is the value-weighted average (median) value of $DISP2$ across all stocks in the specified portfolio, LEV is the median market leverage across all stocks in the specified portfolio, and $\hat{\beta}_{i,t-1}$ is the conditional beta from the prior quarter. The numbers in parentheses are t-statistics.



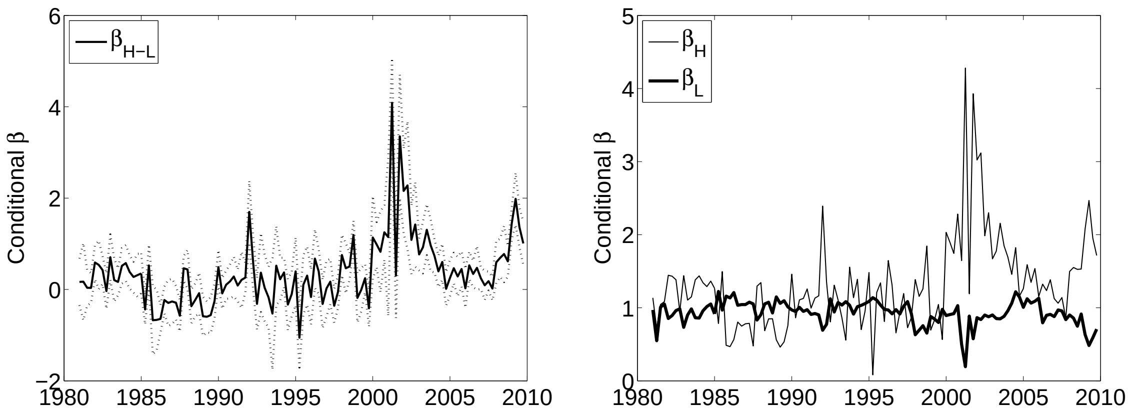
Panel A: CHS portfolios



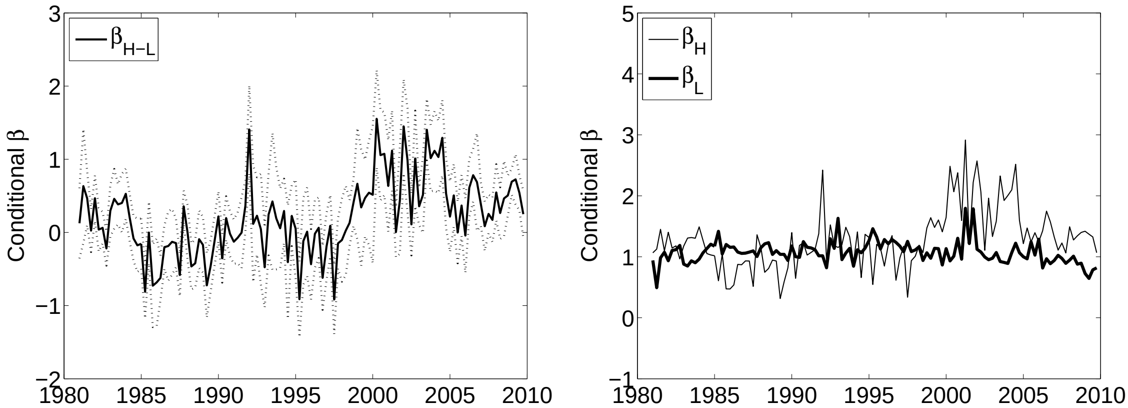
Panel B: O-score portfolios

Figure 1.1: Conditional alphas, 1981–2009

The figure presents conditional CAPM regression alphas (in percent per month) for portfolios sorted on financial distress. Panel A plots the conditional alphas for the hedge portfolio that is long the top quintile and short the bottom quintile of firms sorted on Campbell, Hilscher, and Szilagyi's (2008) failure probability. Panel B plots the conditional alphas for the hedge portfolio that is long the top quintile and short the bottom quintile of firms sorted on Ohlson's (1980) bankruptcy probability. The alphas are estimated quarterly using daily returns. The dotted lines indicate a two-standard-error confidence interval.



Panel A: CHS portfolios



Panel B: O-score portfolios

Figure 1.2: Conditional betas, 1981–2009

The figure presents conditional CAPM regression betas for portfolios sorted on Campbell, Hilscher, and Szilagyi's (2008) failure probability (Panel A) and Ohlson's (1980) O-score (Panel B). The betas are estimated quarterly using daily returns. Within each panel, the plot on the left is the conditional beta for the hedge portfolio that is long the top quintile and short the bottom quintile of firms sorted on the given measure of financial distress. The dotted lines indicate a two-standard-error confidence interval. The plot on the right shows conditional betas for the high- and low-distress quintiles.

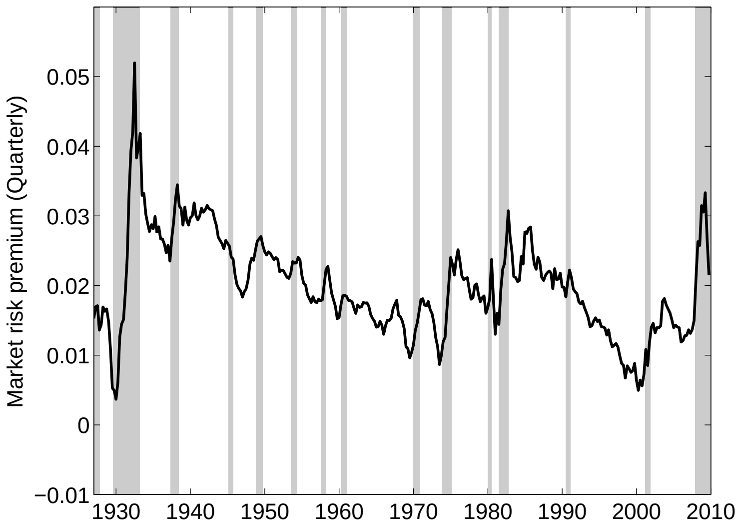


Figure 1.3: Market risk premium, 1927–2009

The figure shows the quarterly market risk premium estimated as the average fitted value from six separate conditional market regressions of the form $r_{m,t} = \delta_0 + \delta_1 Z_{t-1} + \epsilon_{m,t}$, where $r_{m,t}$ is the excess market return and Z_{t-1} is a state variable (i.e., DEF_{t-1}, DP_{t-1}, DY_{t-1}, $NTIS_{t-1}$, TB_{t-1}, or $TERM_{t-1}$) known at the beginning of the period over which the excess market return is measured. The market regressions use quarterly data, and the state variables are defined in the text. The shaded areas indicate recessions as defined by the National Bureau of Economic Research.

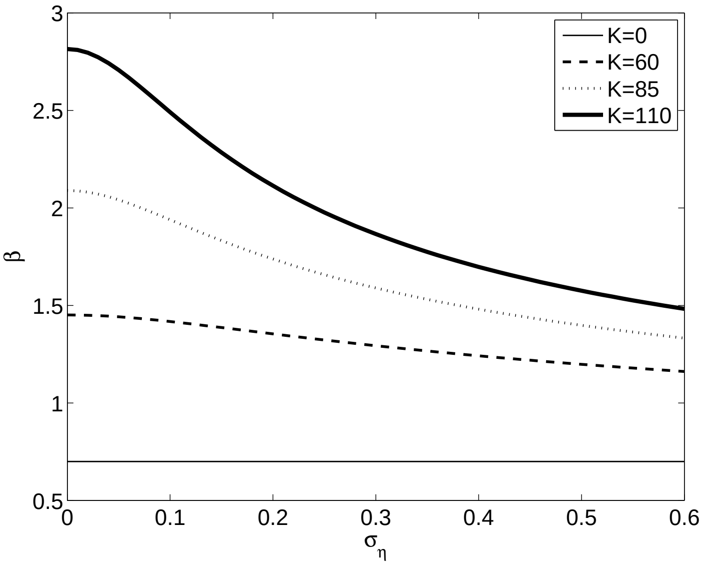
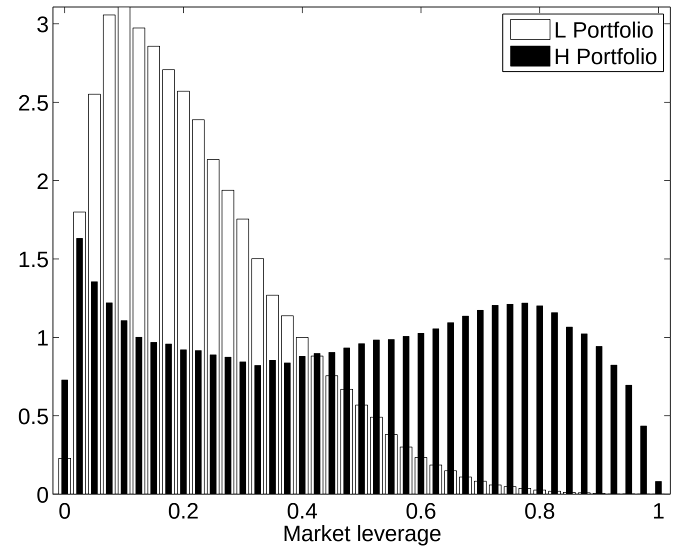
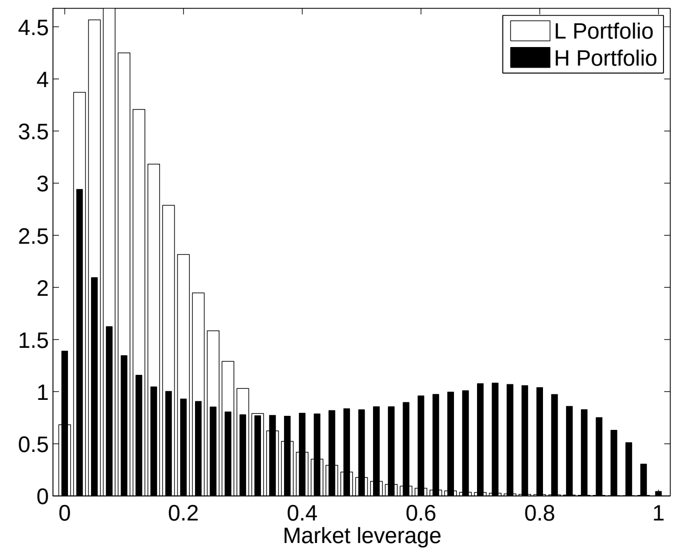


Figure 1.4: Theoretical relation between information risk and equity beta
The figure shows equity beta as a function of the volatility of the noise process (σ_η) and the face value of debt (K) based on the model outlined in Section 1.4. The plots correspond to unlevered firm value $S = 100$, unlevered beta $\beta_S = 0.7$, and cash-flow horizon $\tau = 3$.

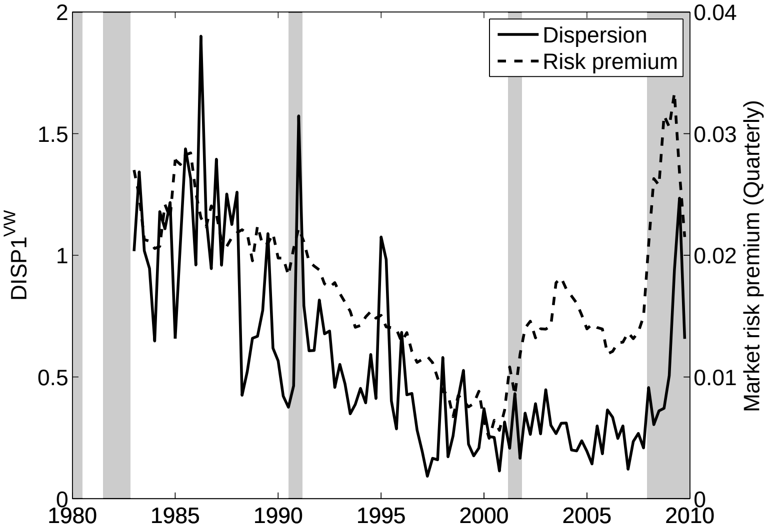


Panel A: CHS portfolios

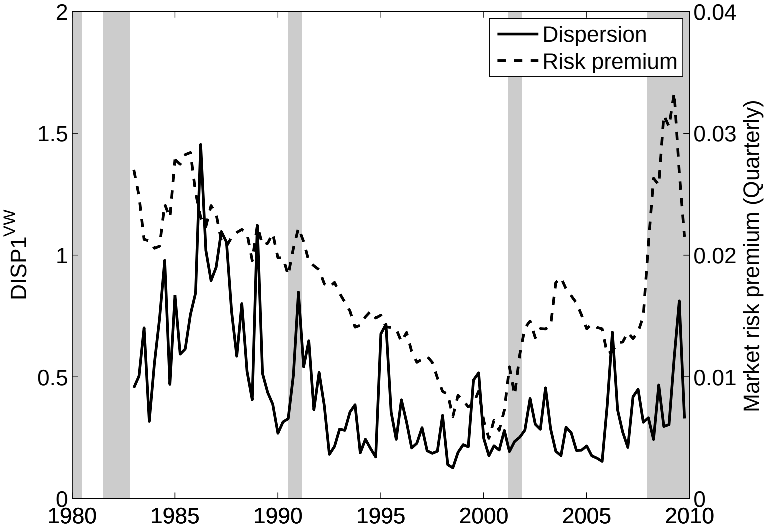


Panel B: O-score portfolios

Figure 1.5: Market leverage, 1981–2009
The figure presents plots of the density histogram of market leverage for portfolios sorted on Campbell, Hilscher, and Szilagyi's (2008) failure probability (Panel A) and Ohlson's (1980) O-score (Panel B). Market leverage is the ratio of total liabilities to the market value of total assets (LEV). Within each panel, the histograms are for the lowest and highest quintile of firms sorted on the given measure of financial distress. The histograms use all firm-month observations.



Panel A: Dispersion for CHS, H portfolio



Panel B: Dispersion for O-score, H portfolio

Figure 1.6: Portfolio-level measures of information risk, 1983–2009

The figure shows the time series of $DISP1^{VW}$ for the portfolio that is long the top quintile of firms sorted on Campbell, Hilscher, and Szilagyi's (2008) failure probability (Panel A) and for the portfolio that is long the top quintile of firms sorted on Ohlson's (1980) bankruptcy probability (Panel B). The portfolio-level measure, $DISP1^{VW}$, for a given month is the value-weighted average value of $DISP1$ across all stocks in the specified portfolio with a valid observation for $DISP1$. The dashed line is the quarterly market risk premium estimated as the average fitted value from six separate conditional market regressions of the form $r_{m,t} = \delta_0 + \delta_1 Z_{t-1} + \epsilon_{m,t}$, where $r_{m,t}$ is the excess market return and Z_{t-1} is a state variable (i.e., DEF_{t-1}, DP_{t-1}, DY_{t-1}, $NTIS_{t-1}$, TB_{t-1}, or $TERM_{t-1}$) known at the beginning of the period over which the excess market return is measured. The shaded areas indicate recessions as defined by the National Bureau of Economic Research.

CHAPTER 2
ASSET-PRICING ANOMALIES AT THE FIRM LEVEL

2.1 Introduction

An anomaly is a pattern in average stock returns that is inconsistent with the predictions of the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965). Anomalies are commonly identified using a portfolio-based approach. The researcher sorts stocks on a firm characteristic and constructs a zero-cost hedge portfolio by taking long and short positions in the extreme groups. If the hedge portfolio earns abnormal returns relative to the CAPM, the sorting characteristic is classified as an anomaly. Over the past three decades, a large number of anomalies have been uncovered, suggesting the CAPM is unable to explain much of the cross-sectional variation in average stock returns.

There are, however, growing concerns in the literature about the use of portfolios to identify anomalies and, more generally, to test asset-pricing models. These arguments are centered around the idea that grouping firms into portfolios and aggregating returns wastes and potentially distorts valuable information about cross-sectional patterns in abnormal returns.[1] One way to avoid the concerns with portfolios is to use firm-level data. To examine anomalies at the firm level, however, the

---

[1]For example, Litzenberger and Ramaswamy (1979) and Ang, Liu, and Schwarz (2010) consider the loss in efficiency from using portfolios rather than individual firms in asset-pricing tests, while Roll (1977), Kandel and Stambaugh (1995), and Fama and French (2008) discuss how patterns in firm-level pricing errors can be distorted at the portfolio level. Lo and MacKinlay (1990) highlight the data-snooping biases inherent in portfolio-based asset-pricing tests. Ahn, Conrad, and Dittmar (2009) and Lewellen, Nagel, and Shanken (2010) show inferences in asset-pricing tests are remarkably sensitive to the choice of test portfolios. For other issues, see Conrad, Cooper, and Kaul (2003), Kan (2004), and Daniel and Titman (2006b).

researcher has to relate firm characteristics to abnormal returns. Abnormal returns are not directly observable, so the researcher must model and estimate the evolution of betas – a challenging problem, especially at the firm-level.

Two recent firm-level studies adopt contrasting approaches to control for market risk. Avramov and Chordia (2006) model market risk as an exact linear function of firm size, book-to-market, and macroeconomic variables, while Fama and French (2008) argue that market risk should not be related to firm characteristics removing the need to examine abnormal returns. Both approaches are problematic. Specifying betas as an exact linear function of covariates is only valid if the researcher knows the full set of variables associated with variation in betas, while tests examining the relation between firm characteristics and raw returns will overstate the CAPM's failings if firm-level betas are associated with firm characteristics.

In this paper, we develop a hierarchical Bayes approach to explore anomalies at the firm level. Specifically, we simultaneously estimate (1) conditional CAPM model parameters for each firm using an approach similar to Lewellen and Nagel (2006) which specifies short time periods and avoids the need for conditioning information, (2) the cross-sectional relation between conditional alphas and firm characteristics in each time period, and (3) the systematic association between alphas and firm characteristics across the entire sample period. Our approach has several desirable features relative to the prior literature. We put little structure on the dynamics of conditional betas, thereby minimizing potential model mis-specification. Our one-step methodology eliminates a measurement error problem encountered in traditional two-step approaches (e.g., Brennan, Chordia, and Subrahmanyam (1998) and Avramov and Chordia (2006)). We also implicitly control for cross-sectional heteroskedasticity and cross-correlations among stocks.

We use this approach to examine nine anomalies over the period 1963 to 2008: size, book-to-market, momentum, reversal, profitability, asset growth, net stock issues, accruals, and financial distress.[2] Studying each anomaly separately, we find that firm-level associations are distorted at the portfolio level for four of the nine anomalies. For example, the traditional portfolio approach suggests size and reversal are associated with abnormal returns, but using information from the entire cross section of stocks there is no evidence of a robust relation between either of these variables and firm alphas. Further analysis suggests the portfolio-level results for size and reversal are driven by a small subset of stocks with extreme values for these characteristics. Nevertheless, the initial firm-level evidence still paints a bleak picture for the CAPM. Seven of the nine characteristics are significantly associated with alphas, suggesting that the CAPM does indeed fail across multiple dimensions. These results, however, may be misleading for three reasons.

First, a disadvantage of using the entire cross section of firms to study anoma-

[2]Previous papers show a positive relation between average returns and book-to-market equity (Rosenberg, Reid, and Lanstein (1985), Chan, Hamao, and Lakonishok (1991), and Fama and French (1992)), stock return momentum (Jegadeesh (1990) and Jegadeesh and Titman (1993, 2001)), and profitability (Haugen and Baker (1996) and Cohen, Gompers, and Vuolteenaho (2002)). There is a negative relation between average returns and size (Banz (1981) and Fama and French (1992)), stock return reversal (DeBondt and Thaler (1985) and Chopra, Lakonishok, and Ritter (1992)), asset growth (Fairfield, Whisenant, and Yohn (2003), Titman, Wei, and Xie (2004), and Cooper, Gulen, and Schill (2008)), net stock issues (Loughran and Ritter (1995), Ikenberry, Lakonishok, and Vermaelen (1995), Daniel and Titman (2006a), and Pontiff and Woodgate (2008)), accruals (Sloan (1996), Collins and Hribar (2000), and Xie (2001)), and financial distress (Dichev (1998) and Campbell, Hilscher, and Szilagyi (2008)).

lies is that inferences may be heavily influenced by small, illiquid stocks. It is possible that the anomalous patterns are being driven primarily by firms which represent only a tiny fraction of the total market capitalization. We investigate this issue by allowing associations between alphas and firm characteristics to vary across micro, small, and big stocks.[3] We find the associations are strongest in terms of statistical and economic magnitude for micro and small stocks. For big stocks, alphas are significantly associated with only three of the nine characteristics - asset growth, net stock issues, and accruals.

Second, anomalies could be the result of temporary market mis-pricing or data snooping by researchers. These explanations suggest established deviations from the CAPM should weaken over time. To examine this issue, we consider whether the relation between alphas and each firm characteristic attenuates or persists after the anomaly is established in the asset-pricing literature. Anomalies that persist post publication are more likely to reflect a fundamental failure of the CAPM. Of the seven anomaly variables with sufficient post-publication sample periods only two – book-to-market and accruals – are significantly related to abnormal returns after publication. Further, these relations are driven by micro stocks for which transaction costs and liquidity concerns diminish investors' ability to exploit anomalies and correct mis-pricings (e.g., Jensen (1978)). Among big stocks, no firm characteristic is significantly associated with abnormal returns post publication.

Third, firm characteristics could be correlated with each other and offer lit-

[3]Following Fama and French (2008), we classify stocks into three size groups - micro, small, and big. The breakpoints are based on the 20th and 50th percentiles of market capitalization for NYSE stocks at the end of June each year.

tle unique information about abnormal returns. Asset-pricing tests that consider each firm characteristic in isolation are likely to suffer from an omitted variable bias that will result in the importance of an anomaly being overstated. Traditional portfolio approaches are unable to adequately address this omitted variable problem. Researchers typically rely on multi-dimensional sorts to isolate the effects of a particular characteristic, but controlling for more than one or two characteristics simultaneously is infeasible. In contrast, our approach is particularly well suited to assess which anomalies contain unique information; we simply specify conditional alphas as a function of multiple firm characteristics. Our results suggest that univariate tests do indeed suffer from a pronounced omitted variable bias. Considering all characteristics simultaneously we find that size, momentum, reversal, asset growth, and financial distress do not contain significant incremental information about abnormal returns, in contrast to the corresponding portfolio-level results.

Taken together, the results suggest that while the CAPM does not perfectly explain firm returns, the anomaly-based evidence against the CAPM is generally overstated. Relations between firm characteristics and conditional firm-level alphas are primarily focused among micro and small stocks and tend not to persist after the anomalies are first documented. Furthermore, few of the firm characteristics associated with alphas actually contain unique information.

The paper is organized as follows. Section 2.2 develops our econometric model for testing asset-pricing anomalies and discusses the advantages and disadvantages of the proposed approach. Section 2.3 describes the data. Section 2.4 presents the empirical results. Section 2.5 concludes.

2.2 Methodology

This section develops our firm-level approach for identifying anomalies relative to the CAPM. The Sharpe–Lintner version of the CAPM states that

$$E\left[r_{i,t}\right] = \beta_i E\left[r_{m,t}\right], \tag{2.1}$$

where $E\left[r_{i,t}\right]$ denotes the expected excess return on stock i at time t, $E\left[r_{m,t}\right]$ is the market risk premium, and $\beta_i = \frac{Cov(r_{i,t}, r_{m,t})}{Var(r_{m,t})}$ captures stock i's exposure to market risk. The Sharpe–Lintner CAPM relates the unconditional expectations of firm and market returns. In reality, as a firm grows and evolves, its exposure to market risk will change. Similarly, the market risk premium is likely to vary depending on the state of the economy and the risk tolerance of investors. In the presence of time-varying risk exposures and risk premiums, a conditional version of the CAPM,

$$E_{t-1}\left[r_{i,t}\right] = \beta_{i,t} E_{t-1}\left[r_{m,t}\right], \tag{2.2}$$

may hold even if the unconditional CAPM does not (Jagannathan and Wang (1996)).

The conditional CAPM implies that the expected conditional alpha, defined as

$$E_{t-1}[\alpha_{i,t}] = E_{t-1}\left[r_{i,t}\right] - \beta_{i,t} E_{t-1}\left[r_{m,t}\right], \tag{2.3}$$

should equal zero for all stocks. A common way of testing this prediction is to examine whether alphas can be forecasted by firm characteristics. Many existing tests in the literature rely on portfolio-based approaches. However, grouping firms into portfolios and aggregating returns has adverse effects. Specifically, valuable information is discarded while averaging across firms and cross-sectional patterns in firm returns may be distorted as a result of the portfolio formation procedure. An alternative approach involves testing the CAPM's prediction that alphas are not forecastable using the full cross section of firm returns by examining the cross-sectional relation,

$$\alpha_{i,t} = \delta_0 + \delta_x x_{i,t-1} + \epsilon_{i,t}, \tag{2.4}$$

where $x_{i,t-1}$ is a firm characteristic that is observable at time $t-1$. The conditional CAPM implies that $\delta_x = 0$ in a cross-sectional regression based on equation (2.4).

Analysis of this cross-sectional regression is complicated by the fact that the dependent variable, $\alpha_{i,t}$, is a latent variable. As such, a model for the latent alphas is necessary to examine the relation in equation (2.4). A test using such a model would ideally have two features. First, the specification should not introduce a spurious relation between $\alpha_{i,t}$ and $x_{i,t-1}$ through the model for the latent alphas. Second, given the structure of the problem, the posterior precision of δ_x should be maximized.

With these considerations in mind, we develop a firm-level test of the CAPM's implication that alphas are not predictable. Specifically, we estimate a system of simultaneous equations,

$$r_{i,t,y} = \alpha_{i,y} + \beta_{i,y} r_{m,t,y} + \epsilon_{i,t,y}, \ \epsilon_{i,t,y} \sim N\left(0, \sigma^2_{i,y}\right), \tag{2.5}$$

$$\alpha_{i,y} = X_{i,y}\delta_y + \eta_{i,y}, \ \eta_{i,y} \sim N\left(0, \sigma^2_{\alpha,y}\right), \tag{2.6}$$

$$\delta_y = \overline{\delta} + \nu_y, \ \nu_y \sim MVN\left(0, \mathbf{V}\right), \tag{2.7}$$

where $r_{i,t,y}$ denotes the excess return on stock i in subperiod t of time period y, $r_{m,t,y}$ is the excess market return, and $X_{i,y}$ is a matrix including a constant and firm characteristics observable at the beginning of period y. In the primary model specification, we use monthly subperiods (t) and annual periods (y). We therefore allow firm alphas and betas to change each year, utilizing the short-window regression approach of Lewellen and Nagel (2006) to test the conditional CAPM. In equation (2.6), δ_y measures the year-by-year relations between alphas and firm characteristics. In a given year, however, abnormal returns may be related to characteristics purely by chance. To examine whether there is a systematic relation between firm characteristics and alphas throughout the entire sample period, we assume that the parameter vectors, $\{\delta_y\}_{y=1}^{Y}$, in equation (2.6) are drawn from the multivariate normal distribution specified in equation (2.7). If an element of $\overline{\delta}$ is focused away from zero, there is evidence of an anomaly that persists through time. In our empirical analysis, we analyze $\overline{\delta}$ when assessing the importance of firm characteristics in forecasting alphas.

We estimate the system of equations (2.5) to (2.7) simultaneously as a hierarchical Bayes model.[4,5] The model structure and estimation technique provides important benefits when examining the relations between alphas and characteristics. In particular, we minimize the potential for specification issues while modeling the

[4]Several papers have used Bayesian techniques to examine asset-pricing models. McCulloch and Rossi (1991) and Geweke and Zhou (1996) develop Bayesian analyses of the Arbitrage Pricing Theory (APT), while Shanken (1987), Harvey and Zhou (1990), Kandel, McCulloch, and Stambaugh (1995), and Cremers (2006) propose Bayesian tests for the mean-variance efficiency of a given portfolio. Ang and Chen (2007) use Bayesian methods to examine whether the conditional CAPM can explain the value premium. Davies (2010) and Cederburg (2010) test the CAPM and the ICAPM, respectively, using Bayesian approaches.

[5]See Rossi, Allenby, and McCulloch (2005) for a discussion of hierarchical Bayes models.

latent alphas and maximize the precision of $\overline{\delta}$.

Relative to existing approaches, our methodology is unlikely to find spurious relations between alphas and characteristics. We make limited assumptions about the evolution of betas over time, only assuming that betas are relatively stable within each year. In contrast, the conditional CAPM is often tested by allowing betas to vary as a function of state variables. Avramov and Chordia (2006) take this approach and model firm betas as an exact linear function of size, book-to-market, and macroeconomic variables. However, such an approach requires the econometrician to know the "right" state variables (e.g., Harvey (1989), Shanken (1990), Jagannathan and Wang (1996), and Lettau and Ludvigson (2001)). Further, misspecification of the process for betas may introduce spurious relations between measured alphas and firm characteristics. If betas are related to other firm characteristics, such as profitability or leverage, that are not included in the model, firm betas will be systematically mismeasured and a researcher may generate incorrect inferences about $\overline{\delta}_x$.

Rather than taking an approach which relies on conditioning information, we directly examine firm alphas and betas within each year, which Fama and French (2006b) note is less vulnerable to specification issues. Fama and French (2008) also avoid complex dynamics for betas by regressing raw returns on firm characteristics to examine anomalies, implicitly assuming that all stocks have betas of one. However, even in the absence of a relation between alphas and firm characteristics, this approach will find $\overline{\delta}_x \neq 0$ if $Corr(\beta_{i,t}, x_{i,t-1}) \neq 0$. There is ample theoretical and empirical evidence that betas are related to firm characteristics (e.g., Karolyi (1992), Gomes, Kogan, and Zhang (2003), and Avramov and Chordia (2006)), so properly adjusting for market risk is important while testing whether alphas are forecastable.

Given that our model design is unlikely to produce spurious relations between alphas and characteristics, we turn to developing estimates of $\overline{\delta}$ which are as precise as possible conditional on the data. Our approach of simultaneously estimating equations (2.5) to (2.7) maximizes the precision of $\overline{\delta}$. In contrast, a common approach in the literature is to estimate the relations between alphas and characteristics in two steps (e.g., Brennan, Chordia, and Subrahmanyam (1998) and Avramov and Chordia (2006)). In the first step, the latent alphas are estimated for each firm. The second step involves a cross-sectional regression of estimated alphas on the firm characteristics,

$$\hat{\alpha}_{i,t} = \delta_0 + \delta_x x_{i,t-1} + \epsilon_{i,t}. \tag{2.8}$$

However, a two-step approach introduces a measurement error problem which leads to an understatement of the evidence against the conditional CAPM. Alphas are measured with error in the first step, and the variance of each firm's estimated alpha is greater than the posterior variance of the firm alpha. The variance of $\overline{\delta}$ is increasing in the variance of the alphas used as dependent variables, so the measurement error problem decreases the precision of $\overline{\delta}$. As a result, $\overline{\delta}_x$ may not be different from zero and alphas may appear to be unforecastable even when a significant relation exists in the data. Our methodology eliminates this measurement error problem by simultaneously estimating equations (2.5) to (2.7) to maximize the precision of $\overline{\delta}$. In particular, while a two-step approach uses only time-series information about the latent alphas, the simultaneous estimation methodology utilizes both time-series and cross-sectional information to make inferences about alphas.

An additional feature of the model specified in equations (2.5) to (2.7) is that cross-sectional heteroskedasticity and cross-correlations among firms are implicitly taken into account. Cross-sectional heteroskedasticity and cross-correlations will influence the precision of δ_y in each period. By allowing the relation between firm characteristics and alphas to vary over time, these features of the cross section of returns will be reflected in the posteriors of $\overline{\delta}$ and $\mathbf{V}$ (Shanken and Zhou (2007)). Thus, a large number of test assets can be considered without requiring the estimation of a variance-covariance matrix.

Estimating equations (2.5) to (2.7) simultaneously is, however, a challenging problem. The model involves a high-dimensional parameter space since firm-specific parameters must be estimated for thousands of firms in each year. Moreover estimation is further complicated by the fact that the latent variables $\alpha_{i,y}$ and δ_y appear in multiple equations within the system. Fortunately, the problem can be greatly simplified by recognizing the hierarchical structure of the model. Equation (2.7) is a hierarchical prior for δ_y in equation (2.6), while equation (2.6) is a hierarchical prior for $\alpha_{i,y}$ in equation (2.5). Thus, we adopt a hierarchical Bayes approach to estimate equations (2.5) to (2.7) simultaneously. In addition to greatly reducing the computational burden relative to using maximum likelihood estimation or the generalized method of moments, the Bayesian approach provides a complete accounting of parameter uncertainty and exact finite sample inference.

The Bayesian approach does require the researcher to specify explicit priors and hyperparameters for all model parameters. We specify the prior for the parameter vector of interest, $\overline{\delta}$, to be

$$\overline{\delta} \sim MVN\left(0, 100\mathbf{I}\right). \tag{2.9}$$

The prior mean of zero implies that firm-level alphas are not associated with firm characteristics, which is not consistent with the considerable empirical evidence to the contrary. However, the informativeness of the prior depends on the prior variance. We specify a large prior variance indicating that we have little prior information about $\overline{\delta}$, so our prior has little effect on the posterior distribution of $\overline{\delta}$. In unreported results, we considered non-zero prior means for each firm characteristic based on the evidence in the asset-pricing literature, but the impact on the posterior distributions was minimal due to the large prior variance.

We specify the prior for firm-level betas as

$$\beta_{i,y} \sim N\left(1, 10\right). \tag{2.10}$$

We use a prior mean equal to one because the average beta of firms in the market must equal one. We set the prior variance at 10, so the prior mean should have little impact on the posterior distribution of betas for most firms. For comparison, Vasicek (1973) recommends a prior variance of 0.25, which has a much stronger effect of shrinking firm betas toward one.[6]

It is also necessary to specify priors for $\left\{\sigma^2_{i,y}\right\}$, $\left\{\sigma^2_{\alpha,y}\right\}$, and $\mathbf{V}$. We model $\left\{\sigma^2_{i,y}\right\}$ and $\left\{\sigma^2_{\alpha,y}\right\}$ using the Inverse Gamma distribution and $\mathbf{V}$ with the Inverse

[6]We also considered a hierarchical model structure for firm betas, similar to the model specified in equations (2.6) to (2.7) for firm-level alphas. However, we found that the posterior distributions for the parameter vector of interest, $\overline{\delta}$, are almost identical using either the hierarchical prior or the prior specified above so we opt for the more parsimonious specification.

Wishart distribution. The hyperparameters for these distributions are chosen to ensure that they have minimal influence on the posterior distributions. Our results are not sensitive to either doubling or halving the hyperparameter values.

We estimate the model specified in equations (2.5) to (2.7) using standard Markov chain Monte Carlo (MCMC) techniques. We draw directly from the conditional posterior distributions for all model parameters using a Gibbs sampler. The algorithm converges quickly. For our empirical applications, we run the chain for 5,000 iterations and discard the first 2,500 as a burn-in period. To test whether the algorithm has converged, we initially ran the chain for 20,000 iterations and found that the posterior distributions characterized using iterations 2,500 to 5,000 were nearly identical to those based on iterations 17,500 to 20,000.

A detailed description of the estimation algorithm and the prior distributions and associated hyperparameters is provided in Appendix C. We also conduct a series of simulation experiments to demonstrate the validity of the estimation approach as well as the robustness of inferences to various features of the cross section of firm returns. A summary of these results is also provided in Appendix C.

2.3 Data

This section outlines the sample construction and data requirements for estimating the model described in equations (2.5) to (2.7). We obtain accounting data from the Compustat Fundamentals Annual files and stock return data from CRSP. The sample includes all NYSE, Amex, and NASDAQ ordinary common stocks with the data required to compute at least one of the following firm characteristics: size

(M), book-to-market (BM), momentum (MOM), reversal (REV), profitability (ROA), asset growth (AG), net stock issues (NS), accruals (ACC), and financial distress (OS).

Following Fama and French (1992), year y runs from July of calendar year y through June of calendar year $y+1$. The characteristics are measured at the end of June in each calendar year y. The variables are matched to monthly returns from July of calendar year y to June of calendar year $y+1$. We exclude financial firms (SIC codes between 6000 and 6999) and firms with negative book equity. Based on Fama and French (2008), we classify firms into micro, small, and big categories using the 20th and 50th percentiles of market capitalization for NYSE stocks at the end of June of calendar year y.

The model described in Section 2.2 requires alphas and betas to be estimated for each firm-year observation. For a firm to be included in the estimation sample in a given year, we require 12 months of return data during that year. The final sample includes 163,603 firm-years of data from July 1963 to June 2008. We use the CRSP value-weighted stock market index as the proxy for the unobserved market portfolio. Monthly excess returns on the CRSP value-weighted stock market index, the risk-free rate, and size breakpoints are from Kenneth French's website.[7] See Appendix C for a detailed description of variable definitions and data construction.

[7]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. We thank Kenneth French for making this data available.

2.4 Results

This section applies the methodology developed in Section 2.2 to explore cross-sectional anomalies. Section 2.4.1 presents the main firm-level results from the estimation of the model described in equations (2.5) to (2.7) and contrasts these results with those from traditional portfolio-level tests. Section 2.4.2 takes a more detailed look at CAPM anomalies at the firm level.

2.4.1 Firm-level tests

Panel A of Table 2.1 summarizes the posterior distribution of $\overline{\delta}$ in equation (2.7), which measures the systematic relation between alphas and firm characteristics over the entire sample period. Initially we examine each firm characteristic in isolation. Following Avramov and Chordia (2006) and Fama and French (2008) we assume a linear relation between conditional alphas and firm characteristics.

Panel A shows that seven of the nine firm characteristics are significantly associated with firm-level alphas. Alphas are positively associated with book-to-market, momentum, and profitability and negatively associated with asset growth, net stock issues, accruals, and financial distress. In terms of economic significance, a one-standard-deviation change in any of the seven variables is associated with a change in alpha ranging in magnitude from 16 basis points (bps) per month for momentum (0.51 x 0.32) to in excess of 20 bps per month for book-to-market, profitability, asset growth, and net stock issues.

For comparison, in Panel B of Table 2.1 we report results based on the traditional portfolio approach that is commonly used to identify anomalies. For each firm characteristic, we sort stocks into deciles each year at the end of June and then form hedge portfolios that are long the highest decile and short the lowest decile of stocks. The portfolios are equally weighted and rebalanced annually. Panel B presents the average conditional CAPM alphas. The conditional alphas are computed following the short-window regression methodology in Lewellen and Nagel (2006). Specifically, we estimate a separate CAPM regression each year using monthly data to obtain a time series of non-overlapping conditional portfolio alphas. The standard errors reported in Panel B are based on the time-series variability of the estimated conditional alphas.

The hedge portfolios formed from sorts on size, book-to-market, momentum, reversal, asset growth, net stock issues, and accruals have CAPM alphas that are significantly different from zero at the 1% level. We find no evidence of significant abnormal returns for the hedge portfolios formed on profitability and financial distress.[8] The results in Table 2.1 provide evidence that underlying firm-level associations can be obscured at the portfolio level. Comparing the firm-level results in Panel A to the portfolio-based tests in Panel B, we find that inferences differ for four of the nine firm characteristics: size, reversal, profitability, and financial distress.

One clear difference between the firm-level and portfolio-level tests is that the portfolio approach only considers firms in deciles one and ten, ignoring information contained in the remaining 80% of stocks. The firm-level approach, on the other

[8]The results for profitability are consistent with those of Fama and French (2008) for portfolio sorts. Chava and Purnanandam (2010) document that the financial distress anomaly is specific to the post-1980 period used by Dichev (1998) and Campbell, Hilscher, and Szilagyi (2008).

hand, utilizes information from the entire cross section. Thus, the portfolio-level analysis could be unduly influenced by a small number of outlier observations in the extreme deciles. To investigate this possibility, in Panel A of Table 2.2 we specify alphas as a function of a constant, the firm characteristic, and two dummy variables identifying whether a particular firm lies in the top or bottom decile for that characteristic. The results suggest that the portfolio-level associations for size and reversal are driven primarily by the extremes. For example, there is no linear relation between alphas and size, but firms in the smallest decile earn alphas that are nearly 0.4% per month higher than firms in the largest decile. For all other firm characteristics, inferences are not substantially altered by the introduction of dummy variables.[9]

When conducting firm-level tests of the CAPM it is also important to consider the potential impact of non-synchronous returns. Our initial model specification uses monthly returns and assumes that all stocks are traded frequently. If trading is infrequent, betas measured by relating firm returns to contemporaneous market returns will tend to understate exposure to market risk. This issue is particularly relevant for our analysis if the extent to which a firm has non-synchronous returns is associated with a given firm characteristic. To control for non-synchronicities we follow Dimson (1979) and include the lagged excess market return as an additional factor in equation (2.5) to correct for any downward bias in measured betas. Panel B of Table 2.2 shows that allowing for non-synchronicities has little impact on the

[9]In results not reported, we considered other non-linear specifications, including the addition of squared and cubed terms for each characteristic, but our inferences were unchanged.

relations between alphas and firm characteristics.

Although there is evidence in Tables 2.1 and 2.2 that firm-level associations between alphas and firm characteristics are distorted at the portfolio level for four out of nine characteristics, the firm-level analysis nonetheless finds substantial evidence against the conditional CAPM. Seven of the nine firm characteristics are significantly associated with conditional CAPM alphas even after allowing for the possibility of non-linearities and non-synchronous returns. In the next section we take a more detailed look at the empirical shortcomings of the conditional CAPM.

2.4.2 A closer look at CAPM anomalies

Given our main results in Panel A of Table 2.1, it is tempting to conclude that the CAPM provides a poor characterization of stock returns. However, in order to properly evaluate the performance of the CAPM we must consider the performance of the model across three dimensions. First, from an economic perspective, it is important to know whether anomalous patterns in returns are market-wide or limited to illiquid stocks that represent a small portion of the total market capitalization. Second, anomalies could arise due to temporary mis-pricing or data snooping by researchers and, as such, would be unlikely to persist over time. Third, it is important to examine to what extent firm characteristics identified as anomalies contain unique information about abnormal returns. If multiple firm characteristics contain the same information then tests that consider each firm characteristic in isolation are likely to suffer from an omitted variable bias that will result in the importance of an anomaly being overstated.

To examine whether anomalies are pervasive across size groups, we repeat the firm-level analysis from Table 2.1, but allow $\overline{\delta}$ to vary across micro, small, and big stocks.[10] The posterior distributions are presented in Figure 2.1 for each firm characteristic. Of the nine characteristics considered, seven are significantly related to the conditional CAPM alphas of micro stocks based on 95% credible intervals. In contrast, only three anomaly variables – asset growth, net stock issues, and accruals – are significantly associated with the abnormal returns of big stocks. Moreover, the economic magnitude of the relations is greatly reduced among big stocks relative to micro stocks. For example, a one-standard-deviation shock in asset growth has a 32 bps per month impact on micro stocks compared to just 12 bps for big stocks. The results in Figure 2.1 suggest that the CAPM provides a much more effective characterization of the returns of big stocks, which constitute over 90% of the total market capitalization.

In Table 2.3 we examine the extent to which firm-level relations between firm characteristics and alphas persist after each firm characteristic is first documented as an anomaly.[11] We re-estimate the model in equations (2.5) to (2.7), but unlike

[10]The robustness of anomalies across size subgroups is an active area of interest. For example, Loughran (1997) argues that the value effect is restricted to small stocks, while Fama and French (2006b) show Loughran's (1997) results are specific to the value/growth indicator, the sample period, and US stocks. Several other papers documenting individual anomalies conduct double sorts on size and a particular anomaly variable, with mixed results. Fama and French (2008) take a more comprehensive approach by analyzing the relations between returns and several firm characteristics within size subgroups.

[11]In prior research regarding the persistence of anomalies, Schwert (2003) finds that the size and book-to-market effects appear to have attenuated after the anomalies were documented, while the momentum anomaly has persisted. Jegadeesh and Titman (2001) also find that the momentum anomaly appears to have persisted throughout the 1990s.

Panel A of Table 2.1, in which $\overline{\delta}$ is constant across the whole sample period, we allow $\overline{\delta}$ to vary across the pre- and post-publication periods.[12] In pre-publication periods the results in Table 2.3 show that alphas are positively related to book-to-market, momentum, and profitability, and negatively related to accruals and financial distress. In the post-publication periods, only book-to-market and accruals remain significantly associated with firm alphas. Moreover, the results for book-to-market and accruals are driven by micro stocks. Among big stocks, there is no evidence of any robust relations between firm characteristics and conditional alphas post publication.

In Figure 2.2 we highlight the relation between accruals and firm alphas before and after the initial publication by Sloan in 1996. Pre-publication there is a robust relation between accruals and alphas across stocks of all sizes. Post-publication, the negative relation persists among micro stocks, diminishes for small stocks, and disappears in big stocks. This pattern is consistent with market participants attempting to exploit the anomaly to earn abnormal returns. Among big stocks, where transaction costs are lowest and there are few, if any, short selling constraints, deviations from CAPM pricing are quickly eliminated. In contrast, investors appear to be unable to trade away the anomaly among micro stocks, where transaction costs are high, liquidity is low, and short selling is often difficult to

[12] We use publication dates based on the following papers: Banz (1981) (size), Rosenberg, Reid, and Lanstein (1985) (book-to-market), Jegadeesh (1990) (momentum), DeBondt and Thaler (1985) (reversal), Haugen and Baker (1996) (profitability), Sloan (1996) (accruals), and Dichev (1998) (financial distress). The asset growth (Cooper, Gulen, and Schill (2008)) and net stock issues (Daniel and Titman (2006a)) anomalies were only recently uncovered so we do not include these characteristics in our analysis.

implement (e.g., Jensen (1978)).

The pre-post analysis in Table 2.3 and Figure 2.2 provides little evidence that anomalies persist after they are first documented, especially among big firms. As such, our evidence is more consistent with the hypothesis that anomalies arise in the data either due to market participants making a mistake which they later correct or due to data snooping by researchers.

Thus far our analysis has focused on the relation between conditional alphas and each firm characteristic in isolation. If firm characteristics are correlated with each other and offer little unique information about alphas then studying each characteristic in isolation will overstate the failings of the conditional CAPM. The traditional portfolio approach is unable to adequately address this omitted variable problem. Researchers typically rely on two- or possibly three-dimensional sorts to isolate the effects of a particular characteristic. Controlling for more than one or two characteristics simultaneously, however, is infeasible and inferences are sensitive to both the sorting technique and the sorting sequence (e.g., Conrad, Cooper, and Kaul (2003)). In contrast, our approach is particularly well suited to assess which anomalies contain unique information; we simply specify firm-year alphas in equation (2.6) as a function of all nine firm characteristics in Table 2.1.

In Figure 2.3 we compare the posterior distributions from two analyses – one in which each firm characteristic is considered in isolation and one in which all characteristics are considered simultaneously.[13] Momentum, asset growth, and financial

[13]In results not reported we also considered a model specification in which conditional alphas were modeled as a function of multiple firm characteristics and the relations were allowed to vary across micro, small, and big stocks. As in Figure 2.1 the relations between

distress are significantly associated with CAPM alphas when considered in isolation, but as Figure 2.3 highlights, none of these characteristics contain significant incremental information when all characteristics are considered simultaneously. The only firm characteristics that are significantly related to firm-level alphas when multiple characteristics are considered simultaneously are book-to-market, profitability, net stock issues, and accruals. Our analysis therefore suggests that univariate tests provide a low hurdle for firm characteristics to be classified as anomalies.

2.5 Conclusion

In this paper, we use a hierarchical Bayes framework to examine asset-pricing anomalies, modeling firm-year alphas as a function of one or more firm characteristics. We investigate nine anomalies – size, book-to-market, momentum, reversal, profitability, asset growth, net stock issues, accruals, and financial distress – over the period 1963 to 2008. Studying each anomaly separately we find robust evidence that CAPM alphas are positively associated with book-to-market, momentum, and profitability. Alphas are negatively associated with asset growth, net stock issues, accruals, and financial distress.

These initial results imply the failings of the CAPM are widespread. A deeper investigation of anomalies, however, suggests that while the CAPM may not perfectly explain firm returns, the anomaly-based evidence against the CAPM is greatly overstated. Relations between firm characteristics and conditional firm-level alphas are primarily focused among micro and small stocks and tend not to persist

characteristics and alphas are generally driven by micro and small stocks.

after the anomaly is first documented. Among large firms there is no evidence of any persistent anomalies. Furthermore, few of the firm characteristics associated with alphas actually contain unique information.

Table 2.1: Firm characteristics and CAPM alphas, 1963-2008

M	BM	MOM	REV	ROA	AG	NS	ACC	OS
Panel A: Base specification								
Posterior mean for the aggregate-level parameters, $\overline{\delta}$								
0.05	0.24**	0.51**	-0.06	1.79**	-0.45**	-1.36**	-1.27**	-1.20*
(0.06)	(0.08)	(0.19)	(0.08)	(0.51)	(0.09)	(0.21)	(0.22)	(0.46)
Average cross-sectional standard deviation of firm characteristics								
1.89	0.86	0.32	0.78	0.15	0.58	0.16	0.13	0.14
Panel B: Performance of hedge portfolios								
Average conditional CAPM alpha, $\hat{\alpha}^{CAPM}$								
-1.08**	1.37**	0.61**	-0.78**	0.17	-1.23**	-1.14**	-0.54**	-0.18
(0.33)	(0.20)	(0.20)	(0.23)	(0.29)	(0.17)	(0.16)	(0.12)	(0.27)

Note: Panel A presents the results from the estimation of the model described in equations (2.5) to (2.7) examining the cross-sectional relation between firm alphas and each firm characteristic separately. We report the posterior mean and standard deviation for the aggregate-level parameters, $\overline{\delta}$, which provide information about the relation between alphas and firm characteristics across the entire sample period. An * (**) indicates that the 95% (99%) credible interval of the posterior distribution does not include zero. Panel B reports average conditional alphas for hedge portfolios that are long the highest decile of stocks and short the lowest decile for each variable. Following Lewellen and Nagel (2006), the conditional CAPM alphas are estimated annually using monthly data. Standard errors are in parentheses. An * (**) indicates significance at the 5% (1%) level using a two-tailed test. The firm characteristics are described in Appendix C.

Table 2.2: Alternative model specifications, 1963-2008

	M	BM	MOM	REV	ROA	AG	NS	ACC	OS
Panel A: Nonlinear specification									
Posterior Mean for the aggregate-level parameters, $\overline{\delta}$									
$\overline{\delta}$, Linear	0.08	0.23**	0.50*	-0.01	1.74**	-0.42**	-0.91**	-1.44**	-0.79*
	(0.07)	(0.09)	(0.24)	(0.09)	(0.51)	(0.10)	(0.23)	(0.28)	(0.33)
$\overline{\delta}$, Decile 1	0.19	-0.07	-0.42**	-0.15	-0.15	-0.20	0.17*	-0.46**	-0.01
	(0.12)	(0.11)	(0.14)	(0.15)	(0.17)	(0.14)	(0.08)	(0.13)	(0.09)
$\overline{\delta}$, Decile 10	-0.18*	-0.02	-0.31*	-0.27*	-0.15	-0.08	-0.16	-0.21*	-0.33*
	(0.08)	(0.11)	(0.13)	(0.11)	(0.10)	(0.12)	(0.11)	(0.10)	(0.16)
Panel B: Sum betas									
Posterior mean for the aggregate-level parameters, $\overline{\delta}$									
$\overline{\delta}$	0.09	0.23**	0.48*	-0.07	2.03**	-0.47**	-1.49**	-1.42**	-1.48**
	(0.06)	(0.08)	(0.22)	(0.09)	(0.54)	(0.09)	(0.24)	(0.23)	(0.47)

Note: The table presents the results from the estimation of the model described in equations (2.5) to (2.7) examining the cross-sectional relation between firm alphas and each firm characteristic separately. We report the posterior mean and standard deviation for the aggregate-level parameters, $\overline{\delta}$, which provide information about the relation between alphas and firm characteristics across the entire sample period. Panel A shows estimates from a nonlinear specification including a linear component and dummy variables for firms with characteristic values in the top or bottom deciles. Panel B shows estimates using sum betas. An * (**) indicates that the 95% (99%) credible interval of the posterior distribution does not include zero.

Table 2.3: Firm characteristics and CAPM alphas pre- and post-publication, 1963-2008

	M	BM	MOM	REV	ROA	ACC	OS
Publication date	1981	1985	1990	1985	1996	1996	1998
Pre-publication – Posterior means for the aggregate-level parameters, $\overline{\delta}$							
All	-0.00	0.24*	0.64**	-0.19	1.91**	-1.32**	-1.18*
	(0.09)	(0.11)	(0.24)	(0.11)	(0.58)	(0.25)	(0.52)
Micro	-0.03	0.31*	0.52*	-0.17	1.48**	-1.31**	-1.18**
	(0.12)	(0.13)	(0.20)	(0.13)	(0.46)	(0.22)	(0.40)
Small	0.12	0.24	0.53	-0.19	2.53**	-1.35**	-2.25**
	(0.13)	(0.15)	(0.29)	(0.14)	(0.60)	(0.38)	(0.60)
Big	-0.08	0.09	0.57	-0.12	1.16	-1.31**	0.39
	(0.11)	(0.14)	(0.37)	(0.15)	(0.93)	(0.39)	(0.84)
Post-publication – Posterior means for the aggregate-level parameters, $\overline{\delta}$							
All	0.08	0.23*	0.33	0.07	1.29	-0.97*	-1.19
	(0.08)	(0.11)	(0.33)	(0.11)	(0.98)	(0.43)	(0.94)
Micro	0.01	0.41**	0.22	-0.02	0.99	-1.35**	-0.92
	(0.10)	(0.13)	(0.26)	(0.13)	(0.75)	(0.40)	(0.72)
Small	0.11	0.24	0.39	0.04	1.67	-0.50	-1.91**
	(0.11)	(0.15)	(0.37)	(0.14)	(0.95)	(0.64)	(0.74)
Big	-0.01	0.12	0.36	-0.01	0.49	0.11	-1.80
	(0.09)	(0.14)	(0.47)	(0.16)	(1.58)	(0.70)	(1.05)

Note: The table presents the results from the estimation of the model described in equations (2.5) to (2.7) examining the cross-sectional relation between firm alphas and each firm characteristic separately. We report the posterior mean and standard deviation for the aggregate level parameters, $\overline{\delta}$, which provide information about the relation between alphas and firm characteristics across time. We allow for different aggregate-level parameters in the pre- and post-publication periods. We estimate two models for each anomaly, one in which $\overline{\delta}$ is restricted to be the same across all firms (All) and one in which $\overline{\delta}$ varies across micro, small, and big stocks. An * (**) indicates that the 95% (99%) credible interval of the posterior distribution does not include zero.

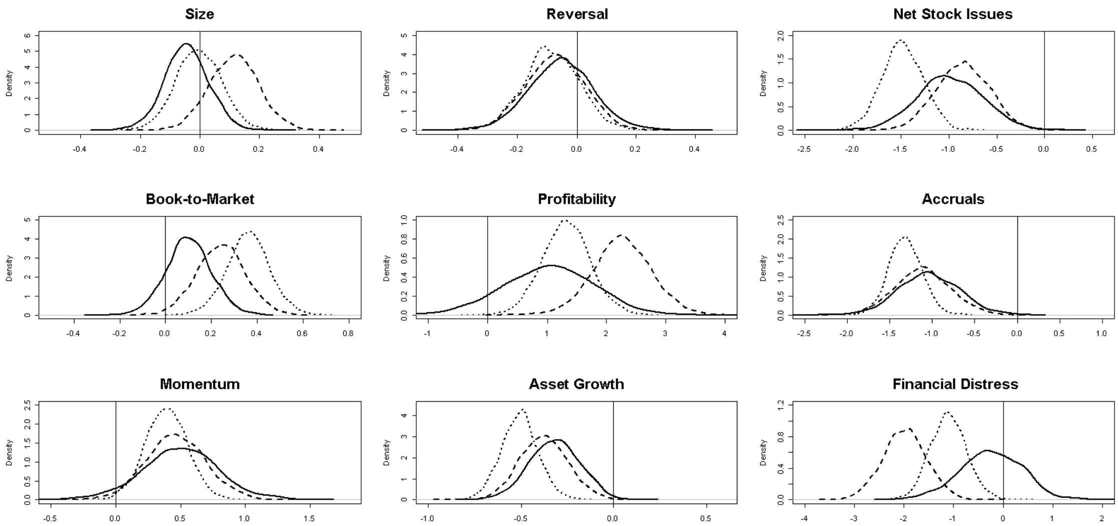


Figure 2.1: Firm characteristics and CAPM alphas by size group
The figure presents the results from the estimation of the model described in equations (2.5) to (2.7) examining the cross-sectional relation between firm alphas and each firm characteristic separately. We report the posterior distributions for the aggregate-level parameters, $\overline{\delta}$, which provide information about the relation between alphas and firm characteristics across the entire sample period. We estimate a model for each anomaly in which the aggregate-level parameters ($\overline{\delta}$) vary across micro (dotted), small (dashed), and big (line) stocks.

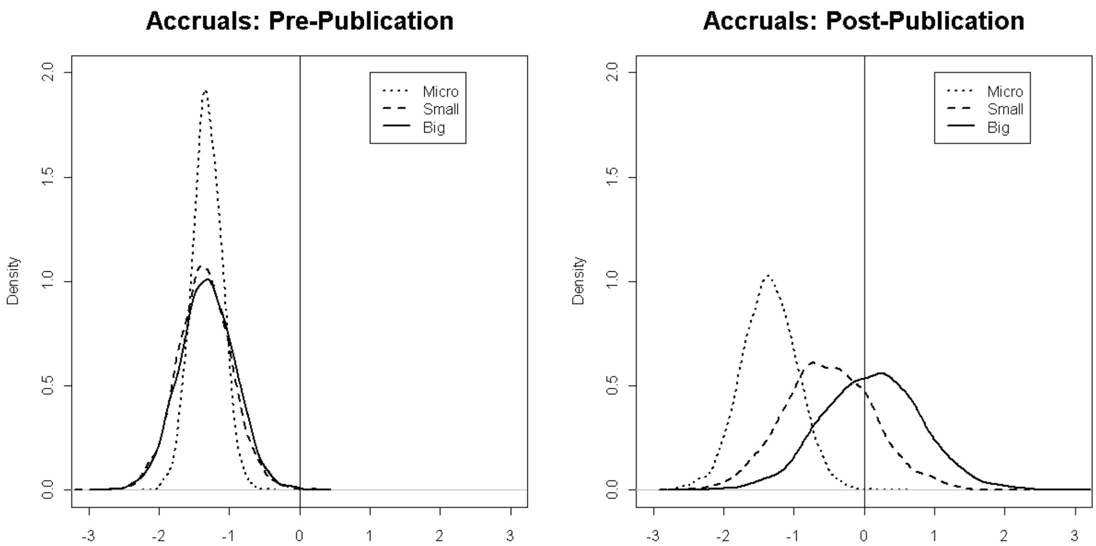


Figure 2.2: The accruals anomaly pre- and post-publication
The figure presents the results from the estimation of the model described in equations (2.5) to (2.7) examining the cross-sectional relation between firm alphas and accruals. We report the posterior distributions for the aggregate-level parameters, $\overline{\delta}$, which provide information about the relation between alphas and firm characteristics across time. We allow for different aggregate-level parameters in the pre- and post-publication periods and also allow the aggregate-level parameters to vary across micro (dotted), small (dashed), and big (line) stocks.

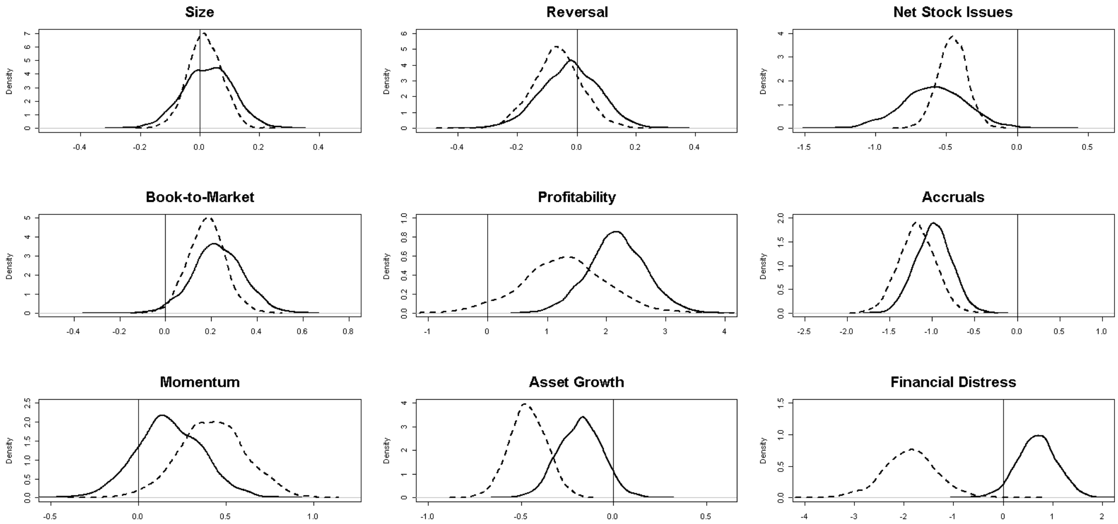


Figure 2.3: Individual and multiple anomaly variables
The figure presents the results from the estimation of the model described in equations (2.5) to (2.7) examining the cross-sectional relation between firm alphas and multiple firm characteristics simultaneously. We report the posterior distributions (line) for the aggregate-level parameters, $\overline{\delta}$, which provide information about the relation between alphas and firm characteristics across the entire sample period. For comparison, for each anomaly variable we also present the posterior distribution (dashed) of $\overline{\delta}$ from estimation of the model described in equations (2.5) to (2.7) for each characteristic in isolation using the same data sample.

CHAPTER 3
MODELING THE CROSS SECTION OF STOCK RETURNS: A MODEL POOLING APPROACH

3.1 Introduction

Over the past three decades, finance academics have discovered a number of empirical regularities in the cross section of stock returns that pose a challenge to models such as the CAPM. The discovery of these anomalies has in turn led to a vigorous search for the 'true' model that explains the cross section of asset returns. This effort has given rise to a number of asset pricing models that purport to improve upon the CAPM or the Fama–French (1993) three-factor model. The search for such models is not without criticism, however. For example, in a recent critique, Lewellen, Nagel, and Shanken (2010) argue that the apparently strong explanatory power of many such models, as evidenced by high cross-sectional R^2s, in fact provides quite weak support for the models.

This paper proceeds in the spirit of Box (1980) that all models are false, but some are useful. Most would agree with the dictum of Box when applied to the case of asset pricing models. More formally, this implies that the space of asset pricing models is incomplete. A model space is said to be complete when it contains the true model. When faced with competing models in a decision context such as performance evaluation or cost of capital estimation, a commonly used strategy is to implement model selection, that is, to choose one of the competing models to the exclusion of the others. However, model selection is a flawed strategy when the model space is incomplete.

The purpose of this paper is to illustrate the advantages of a model pooling approach in contrast to model selection in the context of asset pricing models. One implementation of the model pooling approach is to combine point forecasts of individual models. A more comprehensive approach, and the one adopted in this paper, is to combine the predictive distributions produced by the models under consideration. This approach is warranted because decision problems involving asset returns often require knowledge of the entire probability distribution of returns. Combining distributions has received relatively limited treatment in the econometrics literature. Papers adopting this methodology include Wallis (2005), Hall and Mitchell (2007), and Geweke and Amisano (2010).

Following the framework in Geweke and Amisano (2010), the predictive distributions are combined linearly and the resulting combinations are evaluated using the log predictive score function. The log predictive score function is a well-known measure of the out-of-sample prediction performance track record of the model. The log score rule was introduced by Good (1952), and the connection between the log score rule and the Kullback–Leibler directed distance is studied in Hall and Mitchell (2007). The use of the log score rule typically results in several of the models in the pool under consideration receiving positive weights, a desirable feature in contexts where the model space is incomplete in the sense noted above. As noted by Hendry and Clements (2002), model combination is particularly valuable in applications where individual models are likely to be misspecified due to periodic shifts in the underlying data generating process.

This paper considers optimal model pools of some well-known asset pricing models for the purpose of forming predictions of the one-step-ahead cross section of stock portfolio returns. The weights for the optimal model pools are obtained by maximizing the log predictive score criterion. The properties of the resulting optimal model pools are illustrated for several alternative sets of test assets, including industry-based stock portfolios and portfolios formed on size and book-to-market equity. The illustrations use monthly data for the period 1927–2008.

The initial analysis considers the properties of two-model pools using three of the best-known asset pricing models, namely, the CAPM, the Fama–French three-factor model, and the Carhart (1997) four-factor model. The optimal two-model pools are shown to perform better than any of the individual constituent models when judged by the log predictive score criterion. A similar finding holds for the three-model pool. For both the two- and three-model pools, the model weights of the optimal pools vary over time confirming the well-known instability in the performance of the individual models. It is noteworthy that all the models, including the CAPM, are typically assigned positive weights under the model pooling approach. Although the CAPM is often inferior to other models in the pool, it can help prediction by being included in the pool rather than by being discarded. This application illustrates that treating the model space as incomplete can lead to improved predictions compared to an approach that discards a model like the CAPM on model selection grounds.

Next the performance of two recently proposed models is compared to the original three-model pool. These are the models introduced by Pástor and Stambaugh (2003) and Chen and Zhang (2009). On a stand-alone basis these models are unable to beat the predictive performance of the original optimal three-model pool. This suggests that the three-model pool is a more appropriate benchmark against which to compare new candidate models rather than using the pre-existing individual models on a stand-alone basis as benchmarks, as is typically done in the literature. Finally, the performance of the five-model pool is considered. As expected, the five-model pool, which contains the two recent models, exhibits better performance than the original three model pool, which in turn is the revised benchmark for model performance.

The advantage of the model pooling approach is confirmed when judged by a metric motivated by Fama and French (1996), namely the cross-sectional average absolute pricing error. Specifically, consider comparing the performance of two researchers interested in modeling the cross section of returns. Suppose one researcher uses model pooling to address model uncertainty and the other uses model selection. We show that the model pooling strategy consistently delivers lower pricing errors compared to several alternative strategies based on model selection. For completeness, the performance of a Bayesian model averaging strategy is also considered. Superficially Bayesian model averaging appears to be model pooling. In practice, this procedure typically results in model selection since one model tends to dominate as the sample size increases. As such the model pooling strategy also outperforms the Bayesian model averaging strategy.

This study contributes to the literature on forecast combination pioneered by Bates and Granger (1969) and more recently surveyed by Clemen (1989), Diebold and Lopez (1996), Newbold and Harvey (2001), and Timmermann (2006). The primary contribution of the paper is to illustrate the advantage of model pooling based on the log predictive score criterion in an asset pricing context. Our results complement earlier results on the advantage of pooling forecasts of macroeconomic variables (see, e.g., Stock and Watson (2003, 2004) and Guidolin and Timmermann (2009)). Applications of forecast combination methodologies remain relatively scarce in the finance literature. Mamaysky, Spiegel, and Zhang (2007) and Rapach, Strauss, and Zhou (2010) are recent examples in the context of mutual fund performance evaluation and equity premium prediction, respectively. A distinguishing feature of the present paper compared to the earlier work is the focus on combining predictive densities rather than point forecasts.

The rest of the paper is organized as follows. Section 3.2 describes the class of asset pricing models considered in the paper and the relevant predictive distributions for asset returns. Section 3.3 reviews the theoretical framework for model pooling based on the log predictive score criterion and establishes the basis for the choice of predictive distributions employed in this paper. Section 3.4 explores the characteristics of two-model pools while Section 3.5 examines the multi-model pools. Section 3.6 compares model pooling to model selection and Bayesian model averaging using an economic metric, namely the out of sample cross-sectional average pricing errors. Section 3.7 contains the concluding comments.

3.2 Predictive distributions

The focus of this paper is on constructing optimal pools of asset pricing models for the purpose of forming predictions of the cross section of stock portfolio returns. The choice of the models to be used in the pool is, of course, a matter of judgment on the part of the researcher who may be guided by theory or empirical evidence. The model weights for the optimal pools are obtained by maximizing the log predictive score criterion, which is formally introduced in Section 3.3. The log predictive score function incorporates the predictive densities for the individual models included in the pool. This section specifies the class of asset pricing models considered in this paper and develops the corresponding predictive densities for asset returns. This class of models includes three well-known asset pricing models: the CAPM, the Fama–French three-factor model, and the Carhart four-factor model.

Consider asset pricing models that can be specified as multivariate normal linear regression models with random regressors. Assume that the conditional expectation function is linear in the factors and the disturbance vector is independently and identically normally distributed with mean zero and positive definite variance matrix $\mathbf{\Sigma}$. Then the excess return vector at time t for a typical asset pricing model is

$$\mathbf{r}_t = \alpha + \beta\mathbf{f}_t + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } N\left(0, \mathbf{\Sigma}\right),$$

where $\mathbf{r}_t$ is an $m \times 1$ vector of excess returns at time t for a set of m assets, $\mathbf{f}_t$ is a $k \times 1$ vector of k factors at time t and ϵ_t is an $m \times 1$ vector of disturbances at time t. The parameters α and β have dimensions $m \times 1$ and $m \times k$, respectively. As stated, the return vector is conditional on the contemporaneous factor vector. The model rewritten in terms of a sample of length τ is

$$\mathbf{R}_{t-\tau,t-1} = \mathbf{F}_{t-\tau,t-1}\mathbf{B} + \mathbf{U}_{t-\tau,t-1}, \tag{3.1}$$

where the subscript $t-\tau, t-1$ denotes a sample that extends from month $t-\tau$ to month $t-1$. In equation (3.1) the i-th row of the $\tau \times m$ matrix $\mathbf{R}_{t-\tau,t-1}$ is $\mathbf{r}_{t-\tau-1+i}$, the i-th row of the $\tau \times (k+1)$ matrix $\mathbf{F}_{t-\tau,t-1}$ is $(1, \mathbf{f}'_{t-\tau-1+i})$, the i-th row of the $\tau \times m$ matrix $\mathbf{U}_{t-\tau,t-1}$ is $\epsilon'_{t-\tau-1+i}$ and $\mathbf{B} = [\alpha, \beta]'$ is a $(k+1) \times m$ matrix of regression parameters. It is assumed that $\text{rank}(\mathbf{F}_{t-\tau,t-1}) = k+1$. By the earlier assumption on the ϵ_t vectors, the rows of $\mathbf{U}_{t-\tau,t-1}$ are i.i.d. $N(0, \mathbf{\Sigma})$.

A frequentist or Bayesian approach may be used to obtain the predictive density of $\mathbf{r}_t$ conditional on $\mathbf{f}_t$. In the frequentist approach, the parameter estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\mathbf{\Sigma}}$ are obtained by least squares or maximum likelihood. At time $t-1$ the one-step-ahead prediction for $\mathbf{r}_t$ is obtained by replacing the unknown parameters with estimates and conditioning on $\mathbf{f}_t$, the one-step-ahead factor realizations. Note that in the context of asset pricing models the relevant past history for estimation also includes the time series of the factor vectors as well as the return vectors. Accordingly, the resulting complete conditional predictive density is

$$p(\mathbf{r}_t; \mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, \mathbf{f}_t) = p(\mathbf{r}_t | \mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, \mathbf{f}_t, \hat{\beta}, \hat{\mathbf{\Sigma}}).$$

A special feature of prediction with factor models is that α is set equal to zero.

This paper implements a Bayesian approach to obtain the conditional predictive density for asset returns. Assume that the investor has the standard uninformative prior on $\mathbf{B}$:

$$p(\mathbf{B}, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-\frac{m+1}{2}}. \tag{3.2}$$

The posterior distribution of the parameters of interest is then given by

$$\mathbf{\Sigma}^{-1} \sim Wishart(\tau - k, \mathbf{S}^{-1}),$$

$$Vec(\mathbf{B}) \sim N\left(vec(\hat{\mathbf{B}}), \mathbf{\Sigma} \otimes (\mathbf{F}'_{t-\tau,t-1}\mathbf{F}_{t-\tau,t-1})^{-1}\right),$$

where $\mathbf{S} = (\mathbf{R}_{t-\tau,t-1} - \mathbf{F}_{t-\tau,t-1}\hat{\mathbf{B}})'(\mathbf{R}_{t-\tau,t-1} - \mathbf{F}_{t-\tau,t-1}\hat{\mathbf{B}})$, and $\hat{\mathbf{B}} = [\hat{\alpha}, \hat{\beta}]'$ is the matrix of maximum likelihood (ML) parameter estimates.

Note that the marginal posterior probability density function (pdf) for $\mathbf{\Sigma}$ is in the form of the inverse-Wishart distribution and the conditional posterior pdf for $\mathbf{B}$ is multivariate normal. Following Zellner (1971) the expression for the one-step-ahead conditional predictive density for $\mathbf{r}_t$ has the form of the multivariate Student t distribution:

$$\begin{aligned} p(\mathbf{r_t}|\mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, \mathbf{f}_t) = & \frac{v^{1/2}\Gamma\left[(v+m)/2\right]|\mathbf{V}|^{1/2}}{\pi^{m/2}\Gamma\left(v/2\right)} \times \\ & \left[v + \left(\mathbf{r}_t - \hat{\alpha} - \hat{\beta}\mathbf{f}_t\right)' \mathbf{V} \left(\mathbf{r}_t - \hat{\alpha} - \hat{\beta}\mathbf{f}_t\right)\right]^{-(v+m)/2}, \end{aligned} \quad (3.3)$$

where $\mathbf{f}_t$ is the vector of one-step-ahead factor realizations, $\mathbf{V} = gv\mathbf{S}^{-1}$, $g = 1 - \mathbf{f}'_t(\bar{\mathbf{F}}'\bar{\mathbf{F}} + \mathbf{f}'_t\mathbf{f}_t)^{-1}\mathbf{f}_t$, $\bar{\mathbf{F}} = \{\mathbf{f}_{t-\tau}, \ldots, \mathbf{f}_{t-1}\}$ and $v = \tau - (k+1) - (m-1)$.

For the investor with complete confidence in the factor model, the posterior predictive distribution incorporates the dogmatic point prior $\alpha = 0$. This implies that $\hat{\alpha}$ is set equal to zero in the above conditional predictive density. This convention is followed here for the purpose of model performance evaluation.

Also note that the predictive density of $\mathbf{r}_t$ is conditional on $\mathbf{f}_t$. This framework is appropriate since the class of asset pricing models considered here specifies a relation between the realized asset returns and the contemporaneous realization of the factor returns. While this density can be used to compare the in-sample prediction performance of alternative asset pricing models, the predictive distribution of $\mathbf{r}_t$ conditional on $\mathbf{f}_t$ is not apparently useful for comparing models when the evaluation is based on true out-of-sample predictive performance. This is because the factor return realized in period t is unknown at time $t-1$. In other words, a predictive distribution of asset returns that conditions on contemporaneous factor returns does not represent the outcome of a realistic prediction exercise. We show below, however, that under the assumption the asset pricing models share a common prediction model for the factor returns, the density in (3.3) can be used to evaluate the true out-of-sample predictive performance for a set of competing models. Thus, conditioning on observed factor returns does not detract from the predictive nature of the exercise.

There are two approaches to obtaining a predictive distribution that is conditional only on the history available at $t-1$. The first approach is based on the joint prediction of asset and factor returns. This assumes a model for forecasting the factor returns. The joint one-step-ahead predictive distribution of $\mathbf{r}_t$ and $\mathbf{f}_t$ can be written as the product of two predictive distributions, the one-step-ahead predictive distribution of $\mathbf{r}_t$ conditional on $\mathbf{f}_t$, which is provided by model A, an asset pricing model, and the one-step-ahead predictive distribution of $\mathbf{f}_t$ provided by model B, a factor returns model. The product is

$$\begin{aligned} p(\mathbf{r}_t, \mathbf{f}_t | \mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, A, B) = & p(\mathbf{r}_t | \mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, \mathbf{f}_t, A) \times \\ & p(\mathbf{f}_t | \mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, B). \end{aligned}$$

In the second approach the one-step-ahead predictive distribution of $\mathbf{r}_t$ is the marginal distribution of $\mathbf{r}_t$ derived from the joint one-step-ahead predictive distribution. This marginal is obtained by integrating out $\mathbf{f}_t$ from the joint distribution:

$$p(\mathbf{r}_t|\mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, A) = \int p(\mathbf{r}_t|\mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, \mathbf{f}_t, A) \times$$
$$p(\mathbf{f}_t|\mathbf{R}_{t-\tau,t-1}, \mathbf{F}_{t-\tau,t-1}, B) d\mathbf{f}_t.$$

With either of these approaches the prediction of the asset returns relies only on the history available at time $t-1$, not on the contemporaneous factor returns.

The focus in this paper is on the first approach because it provides the basis for constructing a realistic prediction exercise that can be implemented in practice. The joint predictive distribution of $\mathbf{r}_t$ and $\mathbf{f}_t$ can be thought of as the result of a two-stage prediction procedure where the first stage involves predicting the factor returns and the second stage predicts the asset returns conditional on the forecast of the factor returns. Indeed, this interpretation is often implicit in the specification of asset pricing models. As we show below, under the assumption that the prediction model B for the factors is the same in each case, the evaluation of the various asset pricing models does not require the explicit specification of a prediction model for the factors.

3.3 Optimal model pools

This section reviews the construction of optimal model pools using the framework in Geweke and Amisano (2010). The section begins by introducing the log scoring rule to evaluate linear combinations of predictive probability distributions implied by individual asset pricing models. Linear combinations of predictive distributions are known as linear prediction pools. An optimal prediction pool is one where the model weights are chosen to maximize the log predictive score function. The linear prediction pools used in this paper are linear combinations of the joint predictive distributions of the asset and factor returns. This section establishes the rationale for using these joint predictive distributions to construct optimal prediction pools.

As noted above, the log predictive score function can be based on either the joint predictive distributions of asset and factor returns or the marginal predictive distribution of asset returns. In this paper the log score is based on the joint predictive distribution. A key reason for basing the log predictive score function on the joint distribution of asset and factor returns is that the prediction model for the factors does not have to be specified. As in Geweke and Amisano (2010), the primitives are the predictive densities and the realizations of the time series $\mathbf{r}_t$ and $\mathbf{f}_t$, where the latter are denoted by $\mathbf{r}_t^o$ and $\mathbf{f}_t^o$ (o for observed) in situations where the distinction between the random vector and its realization are important. For a sample consisting of $\mathbf{r}_t^o$, $\mathbf{f}_t^o$, $\mathbf{R}_{t-\tau,t-1}^o$, and $\mathbf{F}_{t-\tau,t-1}^o$, the log predictive score function of the joint predictive distribution is

$$LS(\mathbf{r}_t^o, \mathbf{f}_t^o, \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, A, B) = \sum_{t=\tau+1}^{T} \log\left[p(\mathbf{r}_t^o, \mathbf{f}_t^o | \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, A, B)\right].$$

The log predictive score function is intuitively appealing as it gives a high score to the model that ex ante assigns a high probability to the value of $\mathbf{r}_t$ that materializes, that is, $\mathbf{r}_t^o$. As such the log predictive score function reflects the out-of-sample prediction performance of a given asset pricing model.

This paper considers n alternative asset pricing models and hence alternative prediction models $A_1, \ldots, A_n$ for $\mathbf{r}_t$ conditional on $\mathbf{f}_t$. The log scoring rule is used to evaluate linear combinations of the predictive densities of the form

$$\sum_{i=1}^{n} w_i p(\mathbf{r}_t^o, \mathbf{f}_t^o | \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, A_i, B); \quad \sum_{i=1}^{n} w_i = 1; \quad w_i \geq 0 \quad (i = 1, \ldots, n).$$

This formulation assumes that the prediction model for the factors is the same for all the asset pricing models, namely model B is the same for all $A_1, \ldots, A_n$. The restrictions on the weights w_i are necessary and sufficient to assure that the linear combination of density functions is a density function for all values of the weights and all arguments of any density function. Note that in our context the predictive density produced by each model as shown in (3.3) is heavy-tailed. The combination of these densities is also heavy-tailed, a desirable feature when modeling asset returns.

The assumption that the prediction model for the factors is the same for all A_i implies that the log predictive score function can be expressed as the sum of two terms where the first term involves only the prediction model B and the second term only the prediction models $A_1, \ldots, A_n$. The expression is

$$\begin{aligned} f_T(\mathbf{w}) &= \sum_{t=\tau+1}^{T} \log \left[\sum_{i=1}^{n} w_i p(\mathbf{r}_t^o, \mathbf{f}_t^o | \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, A_i, B) \right] \\ &= \sum_{t=\tau+1}^{T} \log \left[p(\mathbf{f}_t^o | \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, B) \right] + \\ &\quad \sum_{t=\tau+1}^{T} \log \left[\sum_{i=1}^{n} w_i p(\mathbf{r}_t^o | \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_i) \right]. \end{aligned} \tag{3.4}$$

The derivation of this result in the case of a two-model pool is the following:

$$
\begin{aligned}
f_T(\mathbf{w}) = \sum_{t=\tau+1}^{T} & \log[wp(\mathbf{f}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, B)p(\mathbf{r}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_1)+ \\
& (1-w)p(\mathbf{f}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, B)p(\mathbf{r}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_2)] \\
= \sum_{t=\tau+1}^{T} & \log\{p(\mathbf{f}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, B)\times \\
& [wp(\mathbf{r}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_1)+ \\
& (1-w)p(\mathbf{r}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_2)]\} \\
= \sum_{t=\tau+1}^{T} & \log[p(\mathbf{f}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, B)]+ \\
\sum_{t=\tau+1}^{T} & \log[wp(\mathbf{r}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_1)+ \\
& (1-w)p(\mathbf{r}_t^o|\mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_2)].
\end{aligned}
$$

Observe that the weights only enter into one term of expression (3.4), namely the right-hand side term in the last line of (3.4). This term involves the linear combination of the predictive distributions of $\mathbf{r}_t$ conditional on $\mathbf{f}_t$ for asset pricing models $A_1, \ldots, A_n$. Similarly, in the case of the two-model pool the weights are associated only with the asset pricing models A_1 and A_2, not with the factor model B. The relevance of this result will become apparent once the choice of the weights is addressed.

Hall and Mitchell (2007) have proposed combining density forecasts using optimal nonnegative weights. This choice of weights is motivated by the desire to obtain the most accurate density forecast in a statistical sense. An optimal prediction pool is one where the weights are chosen to maximize $f_T(\mathbf{w})$ subject to the restrictions noted above. Accordingly, the optimal prediction pool corresponds to

$$\mathbf{w}_T^* = \arg\max_{\mathbf{w}} f_T(\mathbf{w}).$$

The motivation in Hall and Mitchell (2007) is asymptotic: as $T \to \infty$, the weights that maximize the log predictive score function minimize the distance between the combined forecast density and the true but unknown density as measured by the Kullback–Leibler information criterion (KLIC). See Geweke and Amisano (2010) for further discussion.

It is clear from the above discussion that the rationale for basing the log predictive score function on the joint distribution of asset and factor returns is that the prediction model for the factors does not have to be specified. This rationale hinges on the assumption that the prediction model B for the factors is the same for all A_i. As shown in (3.4), with this assumption, the log predictive score function $f_T(\mathbf{w})$ can be expressed as the sum of two terms where the weights only enter into one of the terms. The term with weights is the log score function of the linear combination of the predictive distributions of $\mathbf{r}_t$ conditional on $\mathbf{f}_t$ for asset pricing models $A_1, \ldots, A_n$. This result implies that the objective function for calculating the weights of the optimal prediction pool reduces to the last term in (3.4). In other words, the optimal prediction pool corresponds to $\mathbf{w}_T^* = \arg\max_{\mathbf{w}} \tilde{f}_T(\mathbf{w})$, where

$$\tilde{f}_T(\mathbf{w}) = \sum_{t=\tau+1}^{T} \log \left[\sum_{i=1}^{n} w_i p(\mathbf{r}_t^o | \mathbf{R}_{t-\tau,t-1}^o, \mathbf{F}_{t-\tau,t-1}^o, \mathbf{f}_t^o, A_i) \right]. \tag{3.5}$$

The function expressed in (3.5) is referred to as the reduced objective function in this paper. The reduced objective function does not involve the prediction model B for the factor returns, and hence the optimal prediction pool can be constructed without specifying model B. This provides the rationale for using the joint one-step-ahead predictive distribution of $\mathbf{r}_t$ and $\mathbf{f}_t$ instead of the marginal distribution of $\mathbf{r}_t$ derived from the joint one-step-ahead predictive distribution of $\mathbf{r}_t$ and $\mathbf{f}_t$. By contrast, the marginal distribution of the asset returns involves model B. The model pool weights $\mathbf{w}_T^*$, determined from the optimization of the objective function expressed in (3.5), are based on the entire sample. From the perspective of an investor making decisions in real-time, the model weights may be determined recursively at each date t using data available through date $t-1$. This latter approach is used in the empirical implementation in this paper.

The fact that the prediction model for the factor returns does not have to be specified is important in practical applications. This is because the factor prediction model is typically unspecified in empirical studies of asset pricing models. In the empirical sections of this paper the weights corresponding to the optimal prediction pool are calculated by maximizing $\tilde{f}_T(\mathbf{w})$, and the value of the objective function reported for an optimal pool is the maximized value of $\tilde{f}_T(\mathbf{w})$. What is lost is the value of the full objective function, $f_T(\mathbf{w})$, but that does not matter for the purpose of comparing the performance of prediction models and model prediction pools.

Geweke and Amisano (2010) point out that the features of optimal prediction pools tend to be strikingly different from those constructed by means of Bayesian model averaging as well as those that result from conventional frequentist testing. Given a data generating process D that produces ergodic $\mathbf{r}_t$, a limiting optimal prediction pool exists, and unless one of the models A_i coincides with D, several of the weights in an n-model pool typically are positive. In the case of a two-model pool, if $\mathbf{w}^* \in (0,1)$, then for a sufficiently large sample size the optimal pool will have a log predictive score superior to that of either A_1 or A_2 alone, and as the sample size increases, $\mathbf{w}_T^* \stackrel{\text{a.s.}}{\to} \mathbf{w}^*$. In contrast, both conventional Bayesain model comparisons and frequentist tests tend to exclude one model or the other as the sample size increases. For the Bayesian approach, the contrast is due to the fact that the conventional setup conditions on one of either $D = A_1$ or $D = A_2$ being true. However, an incomplete model space implies $D \neq A_1$ and $D \neq A_2$. Models that are inferior, as measured by Bayes factors, can substantially improve predictions from the superior model as measured by a log scoring rule. For non-Bayesian testing the explanation is similar. A frequentist test that rejects one model and accepts the other must also condition on one of either $D = A_1$ or $D = A_2$ being true.

This paper uses the past performance of the pool to select the weights of the optimal prediction pool; the past constitutes the training sample for the present. Following a long standing tradition in finance starting with Fama and MacBeth (1973), the predictive densities for the asset pricing models use a five-year rolling window of monthly time-series data, that is, $\tau = 60$ to calculate $\hat{\beta}$ and $\mathbf{S}$. Accordingly, the first density is calculated for date $t = 61$, which is the month of January of the sixth year of the time series. As noted previously, for a pool of models, the optimal weights are calculated from the predictive densities based on the maximization of the reduced objective function (3.5).

The model weights are obtained for three different designs: the full sample, expanding windows and rolling five-year windows. In the case of the full sample design, the log score function value is calculated for date T using monthly predictive densities from date $t = 61$ to date $t = T$ where T is the last December of the time series. In the case of expanding windows, the first log score function value is calculated using predictive densities from date $t = 61$ to date $t = 120$, which is the period from January of the sixth year of the time series to December of the tenth year; the subsequent log score values are calculated using densities up to and including the subsequent Decembers. In other words, the first log score value is computed using an initial five-year window and the subsequent values are updated annually based on expanding windows. Finally, for the five-year rolling window design, the first log score function value is calculated using predictive densities from date $t = 61$ to date $t = 120$; the subsequent log score values are calculated for subsequent Decembers using only the previous five years of monthly predictive densities up to and including each December. Note that the notation for equation (3.5) corresponds to the full sample design case. Modifications to accommodate the expanding and rolling windows are straightforward. In each case the model weights are determined at date t using only data available through date $t - 1$.

For the purpose of comparison, in some parts of this study the model weights are also calculated under a Bayesian model averaging (BMA) approach. Under BMA the individual model weights are the Bayesian posterior probabilities that the given model is the true model, conditional on the data (see, e.g., Hoeting, Madigan, Raftery, and Volinsky (1998) for a review). Given diffuse priors and equal prior probabilities for the n models under consideration, the BMA weight for model A_i

based on a sample of size T is approximately

$$w_i = \frac{\exp\left(-\frac{1}{2}BIC_i\right)}{\sum_{j=1}^{n}\exp\left(-\frac{1}{2}BIC_i\right)},$$

where $BIC_i = 2L_i + k_i \log(T)$ is the Bayesian information criterion for model A_i, L_i is the negative log likelihood, and k_i is the number of parameters in model A_i.

3.4 Examples of two-model pools

This section illustrates the main features of two-model pools with asset pricing models. The individual models considered are three well-known asset pricing models: the CAPM, the Fama–French three-factor model, and the Carhart four-factor model. The illustration uses monthly data for the period January 1927 to December 2008. The factor returns for the CAPM, the Fama–French three-factor model, and the Carhart four-factor model and the test asset returns are taken from Kenneth French's (2009) website. The primary set of test assets considered in this paper is a collection of 50 stock portfolios consisting of 30 value-weighted industry portfolios, 10 value-weighted portfolios formed on market capitalization, and 10 value-weighted portfolios formed on book-to-market equity. This choice of test assets follows the suggestion in Lewellen, Nagel, and Shanken (2010) to augment the standard size and book-to-market portfolios with industry-sorted portfolios when evaluating asset pricing models. In a few cases, we also present results for a subset of portfolios based solely on size, book-to-market, or industry. All raw portfolio returns are converted to excess returns by subtracting the one-month U.S. T-bill return.

First consider the log predictive score as a function of the weight $\mathbf{w}$ in a two-model pool using 30 industry portfolios together with 10 size portfolios and 10 book-to-market portfolios for the entire sample period, January 1932 to December 2008. Figure 3.1 shows for each of the two-model pools that $\tilde{f}_T(\mathbf{w})$ is a concave function and that w is between 0 and 1. In other words, the optimal two-model pool in each case involves a positive weight on both models in the pool. For example, as seen in the upper left panel of Figure 3.1, the optimal two-model pool consisting of the CAPM and the Fama–French model assigns a weight of $w = 0.48$ to the CAPM with the balance to the Fama–French model. A similar pattern emerges in the other two panels of Figure 3.1.

Table 3.1 reports the log predictive scores for each of the individual models as well as for the optimal model combinations using the full sample. Also shown are the corresponding model weights for the optimal combinations. Panel A of the table presents results for the case with 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios as test assets. First consider the performance of each of the three models individually. Over the entire sample period, the CAPM performed the best as evidenced by the model's log score function value of -126,699, which is the highest (least negative) of the three models. By contrast, the Carhart model was the worst performer during this period with the lowest log score function value.

Next, consider the two-model pools in Panel A. Several facts are noteworthy here. One, the log score function value for each of the optimal two-model pools is higher than the corresponding values for either of the individual constituent models. For example, the log score value for the two-model pool containing the CAPM and the Fama–French model is -125,530, which is higher than the log score values of -126,699 and -128,766 for the CAPM and Fama–French model, respectively. Two, the best performing two-model pool includes the CAPM and the Fama–French model. Moreover, even though the CAPM outperforms the Fama–French model on a stand-alone basis based on log score, the Fama–French model is assigned a higher weight (0.52) in the optimal pool. Finally, note that not all the two-model pools dominate the CAPM, in particular, the model pool that includes the Fama–French and Carhart models.

For comparison, Panel B of Table 3.1 reports results using only 30 industry portfolios as test assets. With this reduced set of test portfolios, the Fama–French model exhibits the best out-of-sample predictive ability as measured by the model's log score function value of -74,867. Similar to Panel A, the CAPM and the Fama–French model are featured in the best performing two-model pool. Interestingly, even though the CAPM is the worst performing individual model, it is in fact included in the best performing two-model pool. This last observation is important as it highlights a key feature of model pooling, namely, that models that appear to be inferior on a stand-alone basis may in fact be assigned a positive and significant weight when optimally pooled with other models.

Finally, Panels C and D of Table 3.1 present results based on 10 size-sorted portfolios and 10 book-to-market portfolios, respectively. In these cases we would expect the Fama–French and Carhart models to outperform the CAPM given that both of these models include factors designed to 'explain' returns on the cross section of portfolios sorted by size and book-to-market equity. In each panel, we see the

Fama–French model has the highest log predictive score, while the CAPM performs the worst. In Panels C and D, the two-model pools typically assign only a small weight to the CAPM. These results also highlight an important feature of optimal pooling. Specifically, the log predictive score criterion does not mechanically allocate a substantial weight to all models in a given pool. Models that fail to improve out-of-sample prediction are largely excluded from the optimal pool.

The model weights reported in Table 3.1 are all based on the full sample period. An alternative design is to iteratively update the weights each year using only available data. Figure 3.2 shows the evolution of model weights with an expanding window design for the two-model pools using 30 industry portfolios together with 10 size portfolios and 10 book-to-market portfolios over the period 1936–2008. With the expanding window design, the first weight for a two-model pool is calculated at the end of December 1936 using predictive densities for January 1932 to December 1936. The subsequent weights are calculated at the end of December of each of the following years using an expanding window of data beginning in January 1932. As seen in the figure, the individual model weights vary moderately over time suggesting instability in the relative model performance. Further note that in every model pool, each of the constituent models has a significant weight over the entire period.

In this context it is instructive to compare the optimal model pool weights in Figure 3.2 to the model weights obtained under BMA when considering two-model pairs. Figure 3.3 depicts the evolution of model weights under the latter approach using an expanding window design. As seen from Figure 3.3, BMA effectively results in model selection. For example, when the two models under consideration are the

CAPM and the Fama–French model, the CAPM is assigned a zero weight almost throughout the sample period. A similar result obtains for the case when the two models under consideration are the CAPM and the Carhart model. Once again the CAPM is assigned a zero weight.

An investor relying on BMA to address model uncertainty would have effectively discarded the CAPM and relied on a single competing model. However, as is clear from the earlier discussion, this strategy would have been sub-optimal in terms of out-of-sample forecasting since an optimal model combination that includes the CAPM in fact outperforms the best performing single model.

For the purpose of comparison, Figure 3.4 shows the evolution of the model weights for the optimal two-model pools under a five-year rolling window design. As expected, with the rolling window design, the individual model weights tend to be considerably more volatile compared to the corresponding weights under the expanding window design shown in Figure 3.2. In other words, the variation over time in the composition of the optimal two-model pools is more pronounced under the rolling window design. Nevertheless, in nearly every case, each model in the pool is consistently assigned a substantial and positive weight.

3.5 Multiple-model pools

In this section the characteristics of multiple-model pools are examined. Initially the three-model pool consisting of the CAPM, the Fama–French model and the Carhart model is considered. Then two recent factor models are introduced: the Pástor and Stambaugh (2003) and the Chen and Zhang (2009) models. The

Pástor and Stambaugh model includes the market factor, SMB, HML, and a liquidity factor designed to capture the systematic effect of cross-sectional differences in liquidity. The Chen and Zhang model includes the market factor, and two new factors based on firm-level profitability and investment. The Pástor and Stambaugh traded liquidity factor is taken from the Wharton Research Data Services (WRDS), and the factors for the Chen and Zhang three-factor model are taken from Long Chen's (2009) website.

Turning again to Table 3.1 we see the optimal model weights and the corresponding log predictive score values for the three-model pools. These results are also based on the full sample period, January 1932 to December 2008. When the optimal model pool is based on the 30 industry portfolios, 10 size portfolios, and 10 book-to-market portfolios (last row of Panel A), the CAPM and Fama–French model dominate with weights of 0.47 and 0.44, respectively. The residual weight of 0.09 is assigned to the Carhart model. Panel B shows that with 30 industry portfolios as test assets, the weights are roughly comparable across the three models. Not surprisingly, the three-model pools in Panels C and D allocate very little weight to the CAPM. Furthermore, as expected, the log predictive score values reported in Table 3.1 confirm that the optimal three-model pool for each set of assets outperforms the corresponding best performing two-model pool and the best individual model.

Comparing the results for the three-model pool with those for a two-model pool that omits a particular model provides an indication of the omitted model's contribution to the performance of the complete pool. For example, in the three-model pool with industry portfolios (Table 3.1, Panel B) note that the CAPM is assigned the smallest weight. However, this does not reflect the contribution of the CAPM to the performance of the pool. We can see that the two-model pool which omits the CAPM has the lowest log predictive score. In contrast, the cost of excluding either the Fama–French or the Carhart model is noticeably less. This result highlights the notion that a large model weight is not necessarily a reflection of the model's overall value to the pool.

The optimal weights and the corresponding log predictive score values for three-, four-, and five-model pools are reported in Table 3.2 using a sample restricted to the period January 1977 to December 2008, where the restriction is due to the data availability for the Chen and Zhang model. The test assets for Table 3.2 are restricted to the base case of 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios. On a stand-alone basis, the CAPM outperforms the other four models. By contrast, the Chen and Zhang model performs the worst. Turning to the three-model pools, the best performing pool includes the CAPM, the Fama–French model, and the Chen and Zhang model. Note that the best three-model pool is only slightly better than the pool that includes only the benchmark models, namely the CAPM, Fama–French model, and Carhart model. As expected, the best four-model pool and the five-model pool have higher log score function values than the corresponding best three-model pool. Note, however, that not every four-model pool outperforms the best three-model pool.

The contribution of each individual model to the performance of the complete pool may be inferred by contrasting the results for the four- and five-model pools.

For example, note that the assigned weight on the Fama–French model slightly exceeds the CAPM's weight in the five-model pool. Interestingly, however, the omission of the CAPM from the complete pool reduces the log predictive score value by 405 points. In contrast, the cost of excluding the Fama–French model is only 96 points. Similarly, the weight on the Carhart model exceeds that on the Chen and Zhang model in the five-model pool, but omitting the Chen and Zhang model from the complete pool is more detrimental to the resulting log score. Thus, we see that a model's relative weight is not necessarily synonymous with the model's overall value to the pool.

To summarize, the results in this section confirm the advantage of a model pooling approach over model selection as the optimal multi-model pool is shown to outperform the best performing individual model. In this sense the performance of the optimal model pools shown in Tables 3.1 and 3.2 is impressive. Of course, the composition of the optimal model pool is not known to a decision maker ex ante. Similarly, the identity of the single best performing model over the entire sample period is also not known ex ante. A question of interest is whether model pooling is an attractive strategy for a decision maker in real time. This question is examined in the following section.

3.6 An economic interpretation

As a way of motivating an economic interpretation of the previous results, this section compares the performance of two researchers interested in modeling the cross section of stock returns. One researcher uses model pooling to address model uncertainty and the other uses model selection. Following Fama and French (1996) the measure of performance is based on model pricing errors. The examples in this section illustrate that the researcher using model pooling outperforms the researcher using model selection when performance is judged by model pricing errors. Of course, the model pricing errors are subject to sampling variations. In the spirit of Cochrane (2006), our objective is to provide the reader with a sense of the economic value of model pooling versus model selection.

The researcher using model pooling employs the log predictive score function to determine model pooling weights. At the start of each year, the researcher chooses the model pooling weights by optimizing the log score function using predictive densities from the past sixty months. These weights are used by the researcher to model returns for the next twelve months. Under model pooling the predicted returns are the weighted average of predicted returns of each model using the optimal weights, where the predicted return based on a model for period t is $\hat{\beta}\mathbf{f}_t$. As noted previously, the one-step-ahead predictions are conditional on the one-step-ahead realizations of the factor returns.

By contrast the researcher using model selection chooses one model from the model space at the start of each year using one of four well-known model selection criteria applied to the data from the past sixty months. The criteria are the average absolute time-series regression alpha, the adjusted R^2, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The single model selected is the basis for modeling returns over the next twelve months.

The prediction or model pricing errors for month t are the difference between the realized and predicted returns on each portfolio. The average model pricing error for each portfolio is defined as the time-series average of these monthly out-of-sample errors from 1982 to 2008. The model pricing errors are calculated for 50 portfolios. The measure of economic performance for the researcher is the average pricing error criterion, namely, the across-portfolio average of the absolute value of the average pricing error. For completeness, the performance of a researcher who relies on BMA is also examined. Note that BMA is not equivalent to model selection when only five years of data are used for model evaluation.

Figure 3.5 shows the evolution of the model weights over the period December 1981 to December 2008 for the three-model pool that includes the CAPM, the Fama–French model, and the Carhart model based on 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios and a rolling five-year window design. The first set of model weights are calculated at the end of December 1981 and the weights are subsequently updated on an annual basis. As noted above, the researcher who relies on a model pooling strategy based on the log predictive score criterion uses these weights to make predictions for the next twelve months. As can be seen, while the model weights vary through time, each model typically receives a positive weight in the pool.

Table 3.3 presents the value of the average pricing error criterion under the model pooling strategy and the corresponding errors under the alternative model selection strategies or under BMA. As a starting point, first consider the case in which the researcher relies on a single pre-determined model to form expectations for the portfolio returns. The Carhart model is then seen to be the best performing model over 1982–2008 on a stand-alone basis as evidenced by the lowest average pricing error of 14.5 basis points per month.

Next consider the performance of the researcher that relies on the three-model pool based on the log predictive score criterion. The three models in the pool are the CAPM, the Fama–French model and the Carhart model. For this investor, the average pricing error criterion value is 13.0 basis points per month. This model pooling strategy improves upon the performance of the Carhart model-based strategy. This result highlights that even if a researcher, at the start of the sample period, had perfect foresight about the identity of the best individual model, she would still do better by adopting a model pooling approach. Moreover, the reported pricing error for a researcher relying on the log score criterion is lower than the corresponding measure for a BMA strategy (15.3 basis points per month) that relies on the same three models. The average pricing error criterion value for the model pooling approach (based on the log predictive score criterion) is also lower than for each of the four model selection strategies.

Turning next to the five-model pools, the model pooling strategy based on the log predictive score criterion again outperforms the BMA strategy with an average pricing error criterion value of 13.1 basis points per month compared to an average pricing error criterion value of 15.9 basis points per month for the BMA strategy. The model pooling strategy also outperforms all of the model selection strategies. The evolution of the model weights used by the five-model pooling strategy is shown in Figure 3.6. With a few exceptions, each of the five models has a consistently positive weight in the optimal pool.

In summary, the results in this section confirm that the model pooling approach offers advantages over model selection to a researcher making decisions in real time. Both the three- and the five-model pools based on the log predictive score criterion outperform the best individual model which, of course, is not known ex ante. In contrast, both the BMA strategy and the model selection strategies lead to higher realized pricing errors compared to the best individual model.

3.7 Conclusion

The choice of an appropriate asset pricing model is a key problem faced by researchers and investors alike in applications such as performance evaluation or cost of capital estimation. The conventional approach is to implement a model selection procedure to choose one model from among a set of competing models. Such an approach is justified when the space of models being considered is complete in the sense that the true model is included in the set of models being evaluated. However, this approach is inherently misguided in the much more likely scenario when the true model is not available to the researcher.

This paper considers optimal linear pools of some well-known asset pricing models for the purpose of forming expectations (i.e., predictions) of the one-step-ahead cross section of stock portfolio returns. The optimal model pool weights are based on the optimization of the log predictive score criterion, a measure of the out-of-sample predictive ability of a model. The properties of the resulting optimal model pools are illustrated in the context of stock portfolios formed on industry, size, and book-to-market equity. The study first examines the performance of two-model pools based on three asset pricing models, namely, the CAPM, the Fama–French three-factor model, and the Carhart four-factor model. The optimal pool of two models is shown to perform better than any of the individual models in a two-model pool when judged by the log predictive score criterion. A similar finding holds for the optimal three-model pool that includes all of the models. A key aspect of the model pooling approach in the present context is that all the models under consideration, including the CAPM, are assigned positive weights in the optimal pool. A model that may appear to be inferior to other models on a stand-alone basis can in fact help prediction by being included in the pool rather than by being discarded. This application illustrates that treating the model space as incomplete can lead to improved predictions compared to an approach that discards a model like the CAPM on model selection grounds.

Next the study compares the performance of the recently proposed alternative models of Pástor and Stambaugh (2003) and Chen and Zhang (2009) to the original three-model pool. It turns out that these models are unable to beat the predictive performance of the optimal original three-model pool. This suggests that the three-model pool is a more appropriate benchmark against which to compare new candidate models rather than using the pre-existing individual models as benchmarks as is typically done in the literature. The advantage of the model pooling approach is further confirmed when evaluating model performance using an economic metric, namely the average out-of-sample model pricing error criterion. Again, the optimal model pools consistently outperform the best individual model(s).

The central message of this paper is that researchers would be well served to explicitly recognize in their research design what they implicitly understand, namely, that all models involve approximations to reality. In other words, the model space is incomplete. An explicit acknowledgement of this fact dictates the use of a model pooling approach, as opposed to model selection, when dealing with the problem of forming expectations of stock returns. This paper illustrates the advantages of the former approach in the context of various stock portfolios. Future extensions of this work include implementing the model pooling framework in applications involving performance evaluation, e.g., in the context of assessment of portfolio management skill or the calculation of 'abnormal performance' in corporate event studies.

Table 3.1: Log predictive scores and optimal two- and three-model pools, 1932–2008

	Model weights					Model weights			
Pool size	CAPM	Fama–French	Carhart	Log score	Pool size	CAPM	Fama–French	Carhart	Log score
Panel A: 30 Industry + 10 Size + 10 B/M portfolios					Panel B: 30 Industry portfolios				
1	1.000	–	–	-126,699	1	1.000	–	–	-75,187
	–	1.000	–	-128,766		–	1.000	–	-74,867
	–	–	1.000	-132,154		–	–	1.000	-75,145
2	0.477	0.523	–	-125,530	2	0.273	0.727	–	-74,547
	0.671	–	0.329	-126,004		0.383	–	0.617	-74,560
	–	0.833	0.167	-128,734		–	0.534	0.466	-74,767
3	0.474	0.441	0.085	-125,517	3	0.260	0.355	0.386	-74,466
Panel C: 10 Size portfolios					Panel D: 10 B/M portfolios				
1	1.000	–	–	-13,881	1	1.000	–	–	-17,848
	–	1.000	–	-12,136		–	1.000	–	-16,812
	–	–	1.000	-12,176		–	–	1.000	-16,860
2	0.007	0.993	–	-12,129	2	0.035	0.965	–	-16,781
	0.014	–	0.986	-12,160		0.063	–	0.937	-16,807
	–	0.571	0.429	-12,094		–	0.543	0.457	-16,765
3	0.006	0.566	0.428	-12,087	3	0.032	0.529	0.439	-16,738

Note: The table reports log predictive scores for individual asset pricing models and optimal weights and log predictive scores for various two- and three-model pools. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, and Carhart (1997) four-factor model. The pools are based on four distinct sets of test assets: 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios (Panel A), 30 industry portfolios (Panel B), 10 portfolios formed on market capitalization (Panel C), and 10 portfolios formed on book-to-market equity (Panel D).

Table 3.2: Log predictive scores and optimal N-model pools, 1977–2008

	Model weights					
Pool size	CAPM	FF	Carhart	CZ	PS	Log predictive score
1	1.000	–	–	–	–	-53,156
	–	1.000	–	–	–	-53,776
	–	–	1.000	–	–	-55,030
	–	–	–	1.000	–	-55,319
	–	–	–	–	1.000	-55,306
3	0.440	0.432	0.128	–	–	-52,588
	0.373	0.550	–	0.077	–	-52,584
	0.443	0.487	–	–	0.070	-52,596
	0.521	–	0.369	0.109	–	-52,774
	0.512	–	0.272	–	0.216	-52,687
	0.530	–	–	0.126	0.344	-52,819
	–	0.543	0.168	0.288	–	-52,978
	–	0.648	0.222	–	0.129	-53,742
	–	0.612	–	0.288	0.099	-52,991
	–	–	0.356	0.340	0.304	-53,155
4	0.368	0.426	0.128	0.077	–	-52,572
	0.439	0.375	0.125	–	0.062	-52,585
	0.371	0.482	–	0.077	0.069	-52,580
	0.422	–	0.268	0.098	0.213	-52,665
	–	0.471	0.161	0.288	0.081	-52,974
5	0.367	0.370	0.125	0.077	0.060	-52,569

Note: The table reports optimal weights and log predictive scores for various N-model pools. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model (FF), Carhart (1997) four-factor model (Carhart), Chen and Zhang (2009) three-factor model (CZ), and Pástor and Stambaugh (2003) four-factor model (PS). For a given pool, the log score and optimal weights are computed from predictive densities based on the maximization of the reduced objective function in equation (5) using the entire sample. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios.

Table 3.3: Average absolute pricing errors, 1982–2008

		Model Combination		Model Selection			
Pool size	Models	Log score	BMA	$\|\alpha\|$	R^2	AIC	BIC
1	CAPM	16.0					
	FF	17.1					
	Car	14.5					
	CZ	23.3					
	PS	18.3					
3	CAPM, FF, Car	13.0	15.3	15.1	14.5	14.6	14.6
5	CAPM, FF, Car, CZ, PS	13.1	15.9	14.7	14.9	15.0	15.0

Note: The table reports out-of-sample average cross-sectional absolute pricing errors in basis points per month realized by decision makers relying on individual asset pricing models, model combination, or model selection to form one-step-ahead expectations of excess returns. The weights for the model combination and model selection approaches are updated once a year at the end of December using a rolling five-year window of data, as described in the text. The log-score methodology, Bayesian model averaging, and the model selection criteria are outlined in detail in the text. The pricing errors are computed as the difference between realized returns and the one-step-ahead expected returns. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model (FF), Carhart (1997) four-factor model (Car), Chen and Zhang (2009) three-factor model (CZ), and Pástor and Stambaugh (2003) four-factor model (PS).

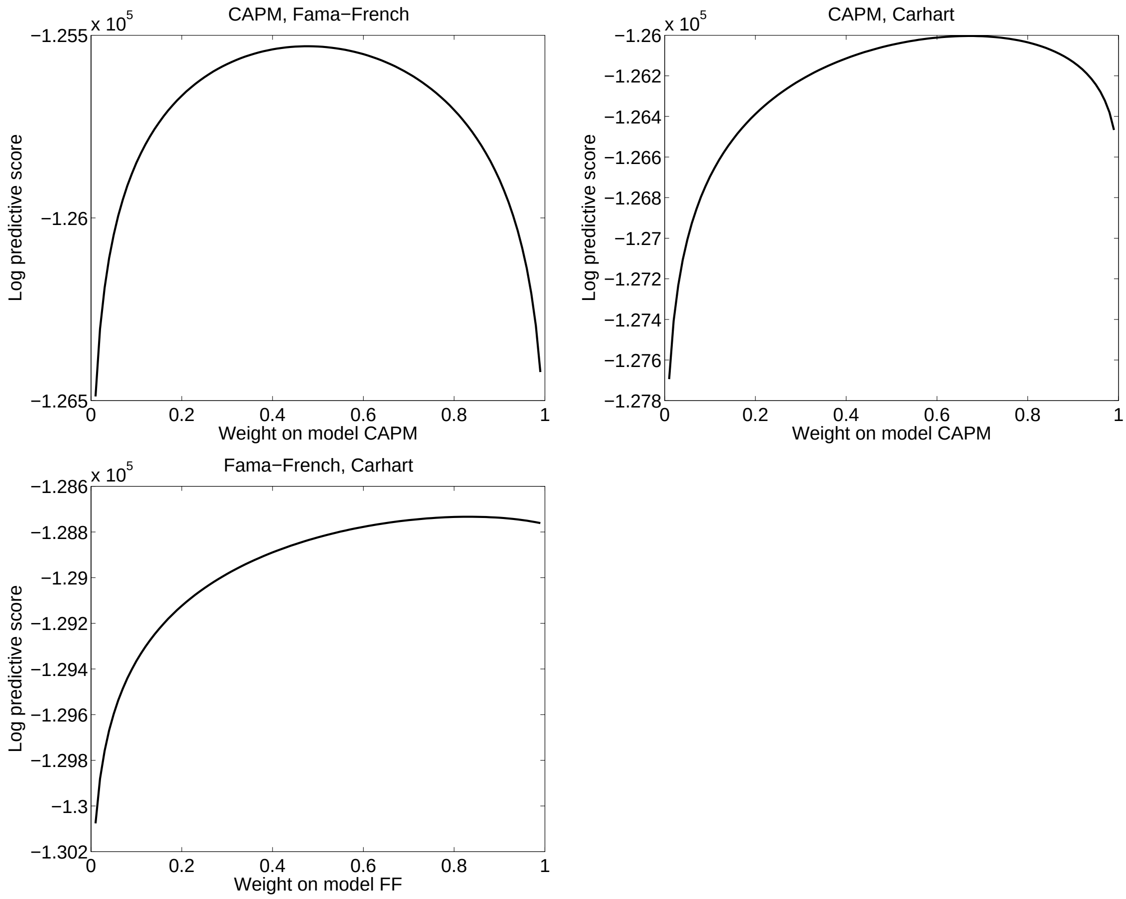


Figure 3.1: Log predictive scores as a function of model weight in two-model pools, 1932–2008

The figure shows log predictive scores for various two-model pools as a function of model weight. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, and Carhart (1997) four-factor model. For a given pool, the log score is computed from predictive densities based on the reduced objective function in equation (5) using the entire sample. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios.

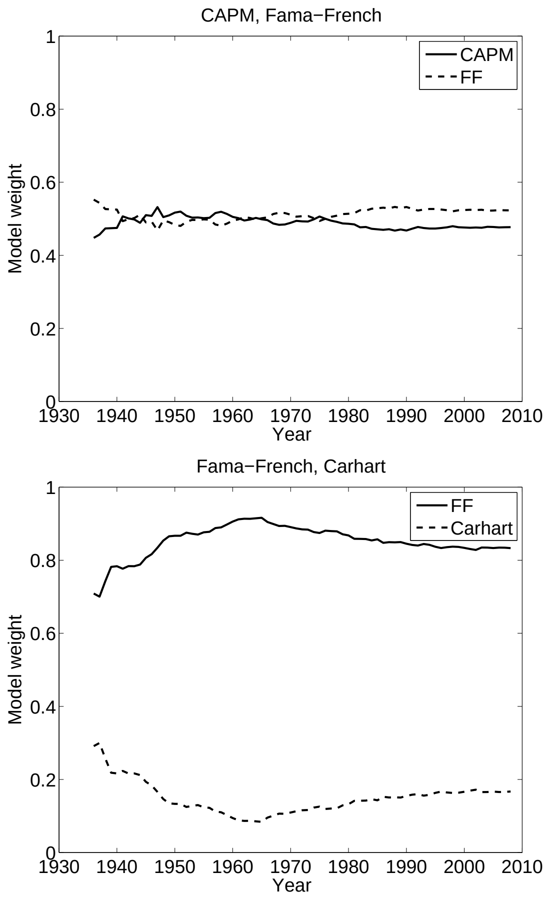


Figure 3.2: Evolution of model weights with expanding windows in two-model pools, 1936–2008

The figure shows the evolution of model weights for various two-model pools of predictive densities. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, and Carhart (1997) four-factor model.

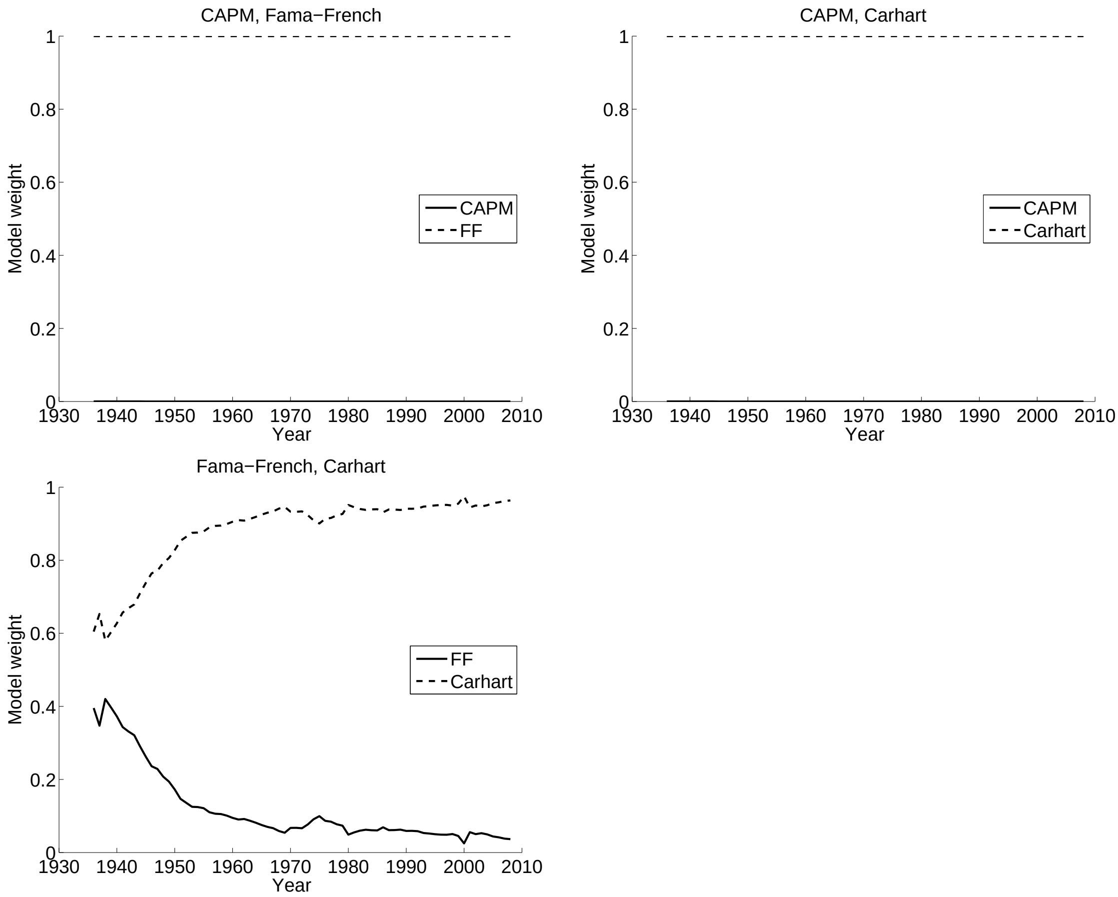


Figure 3.3: Evolution of model weights with expanding windows in two-model pools under Bayesian model averaging, 1936–2008

The figure shows the evolution of model weights for various two-model pools. The optimal weights are based on Bayesian model averaging, as described in the text. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, and Carhart (1997) four-factor model.

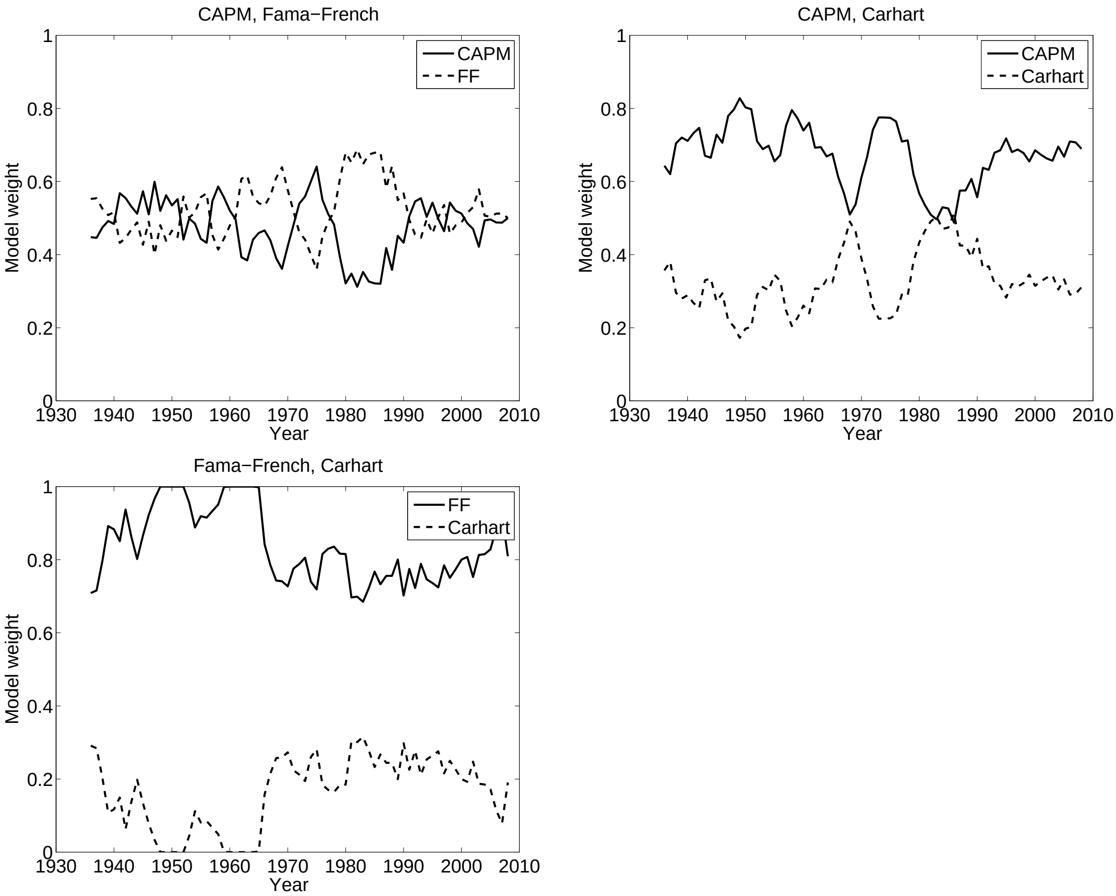


Figure 3.4: Evolution of model weights with five-year rolling windows in two-model pools, 1936–2008

The figure shows the evolution of model weights for various two-model pools of predictive densities. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, and Carhart (1997) four-factor model. Optimal weights are based on the maximization of the reduced objective function in equation (5) using a five-year rolling window.

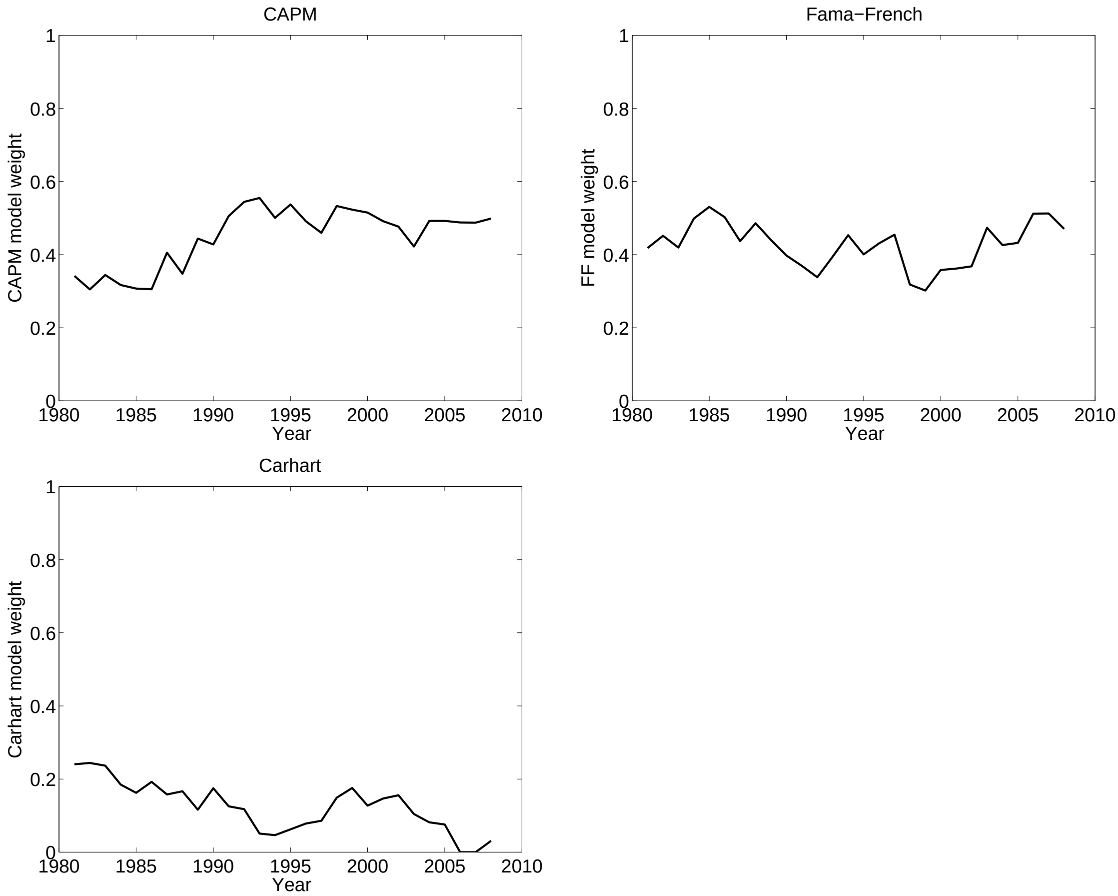


Figure 3.5: Evolution of model weights with five-year rolling windows in three-model pool, 1981–2008
The figure shows the evolution of model weights for a three-model pool of predictive densities. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, and Carhart (1997) four-factor model. Optimal weights are based on the maximization of the reduced objective function in equation (5) using a five-year rolling window.

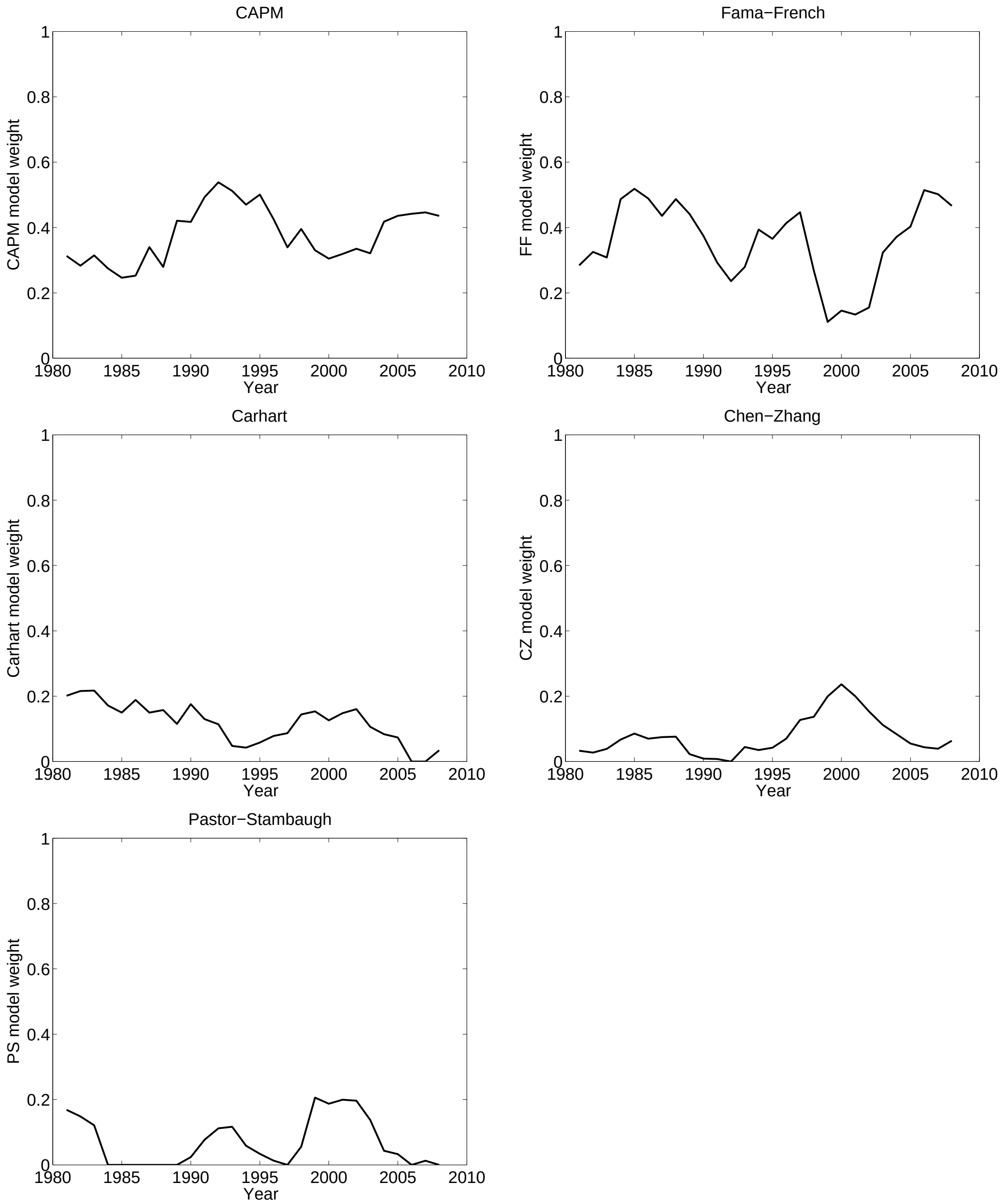


Figure 3.6: Evolution of model weights with five-year rolling windows in five-model pool, 1981–2008

The figure shows the evolution of model weights for a five-model pool of predictive densities. The test assets are 30 industry portfolios used together with 10 size portfolios and 10 book-to-market portfolios. The asset pricing models considered are the CAPM, Fama–French (1993) three-factor model, Carhart (1997) four-factor model, Chen–Zhang (2009) three-factor model, and Pástor–Stambaugh (2003) four-factor model. Optimal weights are based on the maximization of the reduced objective function in equation (5) using a five-year rolling window.

APPENDIX A
DATA AND MODEL APPENDIX FOR CHAPTER 1

This appendix provides details on Campbell, Hilscher, and Szilagyi's (2008) failure probability, Ohlson's (1980) O-score, analysts' forecast dispersion, the macroeconomic state variables used in the empirical analysis, and the stochastic discount factor process specified in equation (1.17). In constructing the proxies for financial distress, I merge the CRSP daily file, CRSP monthly file, and Compustat quarterly file. I lag all accounting data by four months to ensure it is available to investors at the time of portfolio formation. If an accounting variable for a given firm is missing in the Compustat file, I replace the missing variable with the most recent observation for that firm. To alleviate the influence of extreme outliers, I winsorize each of the explanatory variables in the CHS and O-score models at the 1st and 99th percentiles of their monthly cross-sectional distributions.

A.1 Campbell et al. (2008) failure probability

Campbell, Hilscher, and Szilagyi (2008) model the probability a firm files for bankruptcy, is delisted from an exchange for financial reasons, or receives a D rating from a leading credit rating agency over the next twelve months as a function of accounting and market variables. This proxy for financial distress is given by

$$\begin{aligned} CHS = & -9.164 - 20.264\,(NIMTAAVG) + 1.416\,(TLMTA) \\ & -7.129\,(EXRETAVG) + 1.411\,(SIGMA) - 0.045\,(RSIZE) \\ & -2.132\,(CASHMTA) + 0.075\,(MB) - 0.058\,(PRICE)\,. \end{aligned} \tag{A.1}$$

The variable $NIMTAAVG$ is a geometrically declining average of past measures of firm-level profitability:

$$NIMTAAVG_{t-1,t-12} = \frac{1-\phi^3}{1-\phi^{12}}(NIMTA_{t-1,t-3} + \cdots + \phi^9 NIMTA_{t-10,t-12}),$$

where $NIMTA_{t-1,t-3}$ is the ratio of net income to the market value of total assets for the most recent quarter and $\phi = 2^{\frac{1}{3}}$. Similarly,

$$EXRETAVG_{t-1,t-12} = \frac{1-\phi}{1-\phi^{12}}(EXRET_{t-1} + \cdots + \phi^{11} EXRET_{t-12}),$$

where $EXRET_{t-1}$ is the log excess return in the previous month relative to the S&P 500 index. Other model variables are defined as follows: $TLMTA$ is the ratio of total liabilities to the market value of total assets, $SIGMA$ is the standard deviation of daily stock returns over the previous three months, $RSIZE$ is the log ratio of market capitalization to the market value of the S&P 500 index, $CASHMTA$ is the ratio of cash to the market value of total assets, MB is the market-to-book ratio, and $PRICE$ is the log price per share truncated from above at $15. In constructing the market-to-book ratio, MB, I define the book value of equity as shareholders' equity if it is available, and as the difference between total assets and total liabilities, otherwise.[1]

[1]See Campbell, Hilscher, and Szilagyi (2008) for additional details on variable definitions. CHS is the third model from Table IV in their paper.

A.2 Ohlson's (1980) bankruptcy probability

Ohlson's (1980) bankruptcy measure is given below

$$
\begin{aligned}
O\text{-score} = & -1.32 - 0.407\,(SIZE) + 6.03\,(TLTA) - 1.43\,(WCTA) \\
& + 0.076\,(CLCA) - 1.72\,(OENEG) - 2.37\,(NITA) \\
& - 1.83\,(FUTL) + 0.285\,(INTWO) - 0.521\,(CHIN)\,,
\end{aligned} \tag{A.2}
$$

where $SIZE$ is the log of the ratio of total assets to the GNP price-level index, $TLTA$ is the ratio of total liabilities to total assets, $WCTA$ is the ratio of working capital to total assets, $CLCA$ is the ratio of current liabilities to current assets, $OENEG$ is a dummy variable equal to one if total liabilities exceeds total assets and zero otherwise, $NITA$ is the ratio of net income to total assets, $FUTL$ is the ratio of funds from operations to total liabilities, $INTWO$ is a dummy variable equal to one if net income was negative for the past two years and zero otherwise, and $CHIN$ is the following measure of the change in net income:

$$CHIN = \left(\frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|} \right),$$

where NI_t is net income for the most recent year.[2]

A.3 Analysts' forecast dispersion

I compute analysts' forecast dispersion from the IBES database as the month-end standard deviation of current-fiscal-year earnings estimates across analysts tracked by IBES. I compute forecast dispersion from the unadjusted detailed history file using only the most recent forecast made by a particular analyst. I eliminate forecasts

[2] See Ohlson (1980) for exact variable definitions. O-score is the first model from Table IV in his paper.

that are over six months old or are related to fiscal periods that have already ended. I construct two proxies for dispersion of beliefs. The measure $DISP1$ is the standard deviation of current-fiscal-year earnings forecasts divided by the absolute value of the mean forecast, and $DISP2$ is the standard deviation of current-fiscal-year earnings forecasts divided by the market value of assets, where the market value of assets is the book value of debt plus the market value of equity. These firm-level measures of information risk are available at a monthly frequency from January 1983 to December 2009. I also winsorize both $DISP1$ and $DISP2$ at the 1st and 99th percentiles of their monthly cross-sectional distributions.

A.4 State variables

The state variables described below are used in various parts of the empirical analysis. Each of the variables is available at a monthly frequency from January 1927 to December 2009. The data construction largely follows Welch and Goyal (2008).

1. Default premium (DEF): the yield spread between Moody's Baa and Aaa corporate bonds. The bond yields are from the Federal Reserve Bank of St. Louis website.[3]

2. Dividend-to-price ratio (DP): the difference between the log of the sum of dividends accruing to the CRSP value-weighted market portfolio over the previous 12 months and the log of the current index level.

[3] http://research.stlouisfed.org/fred2/

3. Dividend yield (DY): the difference between the log of the sum of dividends accruing to the CRSP value-weighted market portfolio over the previous 12 months and the log of the lagged index level from 12 months prior.

4. Net equity expansion ($NTIS$): the sum of net issues by NYSE listed stocks over the previous 12 months divided by the current market capitalization of NYSE stocks. The net issuing activity for month t is computed as

$$Net\ Issue_t = MCAP_t - MCAP_{t-1} \times (1 + VWRETX_t),$$

where $MCAP_t$ is the total market capitalization at the end of month t, and $VWRETX_t$ is the value-weighted return (excluding dividends) on the NYSE index. The data are from CRSP.

5. Short-term interest rate (TB): The short-term interest rate from 1927 to 1933 is the U.S. yield on short-term United States securities, three-six month Treasury notes and certificates, three month Treasury series in the NBER Macrohistory database.[4] The short-term interest rate from 1934 to 2009 is the yield on the three-month Treasury bill from the Federal Reserve Bank of St. Louis website.

6. Term premium ($TERM$): the yield spread between long-term government bonds and Treasury bills (TB). The long-term government bond yield from January 1927 to March 1953 is the U.S. yield on long-term United States Bonds from the NBER Macrohistory database. The long-term government bond yield

[4]http://www.nber.org/macrohistory/

from April 1953 to December 2009 is the 10-year Treasury constant maturity rate from the Federal Reserve Bank of St. Louis website.

A.5 Stochastic discount factor process

This section shows that the stochastic discount factor process in equation (1.17) is consistent with the CAPM. Following Chapter 9 in Cochrane (2001), assume investors have log utility over consumption, $u(c_t) = \log(c_t)$, and define the discount factor in continuous time as

$$\Lambda_t \equiv e^{-\delta t} u'(c_t).$$

Assume the price, p_t^W, of the aggregate wealth portfolio follows a diffusion as specified below

$$\frac{dp_t^W}{p_t^W} = \mu dt + \sigma_M dW_t^M.$$

The first-order condition for a security with price p_t^W that pays a dividend stream equal to aggregate consumption is given by

$$u'(c_t)p_t^W = \mathrm{E}_t \left[\int_0^\infty e^{-\delta s} u'(c_{t+s}) c_{t+s} ds \right].$$

With $u'(c_t) = 1/c_t$ we have

$$\frac{p_t^W}{c_t} = \int_0^\infty e^{-\delta s} ds = \frac{1}{\delta}.$$

The equation above reflects the well-known result that log utility leads to a constant consumption-to-wealth ratio. Log utility also implies

$$\Lambda_t = \frac{e^{-\delta t}}{c_t} = \frac{e^{-\delta t}}{\delta p_t^W}.$$

Applying Itô's lemma and substituting for the aggregate wealth process, the stochastic discount factor process is

$$\begin{aligned}\frac{d\Lambda_t}{\Lambda_t} &= -\delta dt - \frac{dp_t^W}{p_t^W} + \frac{1}{2}\frac{dp_t^{W\,2}}{p_t^{W\,2}} \\ &= \left(-\delta - \mu + \frac{1}{2}\sigma_M^2\right) dt - \sigma_M dW_t^M \\ &= -r dt - \sigma_M dW_t^M.\end{aligned}$$

APPENDIX B
ROBUSTNESS TESTS FOR CHAPTER 1

In this appendix, I return to the results in Table 1.1 on the performance of distress-sorted portfolios relative to the conditional CAPM. The primary advantage of the Lewellen and Nagel (2006) methodology is that the econometrician is not forced to model conditional betas and/or expected market returns as a function of observable state variables. Recent studies such as Li and Yang (2008), Boguth, Carlson, Fisher, and Simutin (2009), and Ang and Kristensen (2010), however, have raised concerns with this empirical approach and present alternative procedures for estimating the conditional CAPM with high frequency data. Below I show the main results in Table 1.1 are robust to these methodologies as well as more traditional approaches to estimating the conditional CAPM.

B.1 The lagged portfolio approach

Boguth, Carlson, Fisher, and Simutin (2009) argue the Lewellen and Nagel (2006) methodology, which relies on contemporaneous estimates of portfolio betas when estimating conditional alphas, can be problematic. Given that these betas are not in the investor information set, the econometrician is essentially 'overconditioning.' Boguth, Carlson, Fisher, and Simutin (2009) show this approach can lead to large biases in reported alphas when an asset's payoffs are nonlinear in market returns. This overconditioning problem is also most relevant when the rolling regression windows are divided too finely. One method proposed by Boguth, Carlson, Fisher, and Simutin (2009) to address overconditioning is the lagged portfolio risk adjustment approach.

Let $j = 1, \ldots, J$ index the short-window regression intervals in the Lewellen and Nagel (2006) approach (e.g., months, quarters, or half years), where J is the total number of intervals. The vector of Lewellen and Nagel (2006) factor loadings for portfolio i in interval j is

$$\hat{B}_{i,j} = \begin{bmatrix} \hat{\beta}_{i,0,j} & \hat{\beta}_{i,1,j} & \hat{\beta}_{i,2,j} \end{bmatrix}'$$

The lagged portfolio estimate of alpha is given by

$$\hat{\alpha}_i^{BCFS} = \frac{1}{T} \sum_{j=1}^{J} \sum_{t \in j} r_{i,t} - \hat{B}'_{i,j-1} R_{m,t}, \tag{B.1}$$

where

$$R_{m,t} = \begin{bmatrix} r_{m,t} & r_{m,t-1} & (r_{m,t-2} + r_{m,t-3} + r_{m,t-4})/3 \end{bmatrix}',$$

and T is the total number of days in the estimation sample. In short, the lagged portfolio approach simply estimates conditional portfolio alphas in interval j using estimates of beta from interval $j-1$. These factor loadings are available to investors at the start of interval j and, as such, this approach alleviates any potential bias in estimated alphas.

Table B.1 (Method 2) reports estimates of average long-short conditional alphas for the CHS and O-score portfolios using the lagged portfolio methodology of Boguth, Carlson, Fisher, and Simutin (2009). The results are provided for monthly, quarterly, and semiannual periods. For comparision, I also reproduce the corresponding conditional CAPM results from Table 1.1 for the Lewellen and Nagel (2006) approach (Method 1). The average BCFS alpha for the long-short CHS

portfolio ranges from -0.48% to 0.03% per month depending on the rolling estimation window length. None of these estimates is statistically significant at the 5% level. Moreover, the BCFS alphas are similar in magnitude to the Lewellen and Nagel (2006) alphas, suggesting that any overconditioning bias is likely small. Similarly, none of the BCFS alpha estimates for the O-score portfolio is significantly different from zero. The monthly BCFS estimate of -0.23, however, is considerably smaller than the corresponding LN monthly alpha of 0.25. This result suggests overconditioning could be an issue when the estimation windows are divided very finely. In contrast, the quarterly and semiannual alphas for the O-score portfolio are relatively similar across the LN and BCFS approaches.

B.2 State variables approach

A second approach, also advocated by Boguth, Carlson, Fisher, and Simutin (2009), to address potential problems with overconditioning is to estimate the conditional CAPM using lagged conditioning variables. This methodology follows Ferson and Schadt (1996), Ferson and Harvey (1999), Petkova and Zhang (2005), and others and is the traditional empirical approach for implementing the conditional CAPM. In this spirit, I estimate the following time-series regression with daily portfolio return data:

$$r_{i,t} = \alpha_i + (b_i Z_{t-1})' R_{m,t} + \epsilon_{i,t}, \tag{B.2}$$

where Z_{t-1} is a $(k+1) \times 1$ vector of a constant and k state variables that are known at time $t-1$, and b_i is a $3 \times (k+1)$ vector of coefficients. I model the conditional portfolio betas as linear functions of two state variables: DEF_{t-1} and DP_{t-1}. The state variables are defined in Appendix A and updated monthly. The portfolio beta estimate is then

$$\hat{\beta}_{i,t}^{SV} = (\hat{b}_i Z_{t-1})' \iota,$$

where ι is a vector of ones. This regression approach allows beta to vary over time with the state variables and also incorporates an adjustment for nonsynchronous returns.

Table B.1 (Method 3) presents the regression results. The long-short CHS portfolio has a conditional CAPM alpha of -0.44% per month (t-statistic of -1.49), and the long-short O-score portfolio has a conditional alpha of -0.32% (t-statistic of -1.56). Neither estimate is significant at the 5% level, suggesting this version of the conditional CAPM also explains the distress-risk anomaly. The table also reports factor loadings that are summed across lags. These parameter estimates allow us to see how conditional betas respond to the specific state variables. The long-short portfolio betas tend to increase with the default premium and decrease with the dividend-to-price ratio.

B.3 Kernel regression approach

A problem with the empirical methodologies outlined in Sections B.1 and B.2 is that both require the econometrician to know the 'right' state variables. In contrast, the Lewellen and Nagel (2006) approach avoids the problem of using incorrect state variables but is subject to potential overconditioning bias from using a contemporaneous realized beta that is not entirely within the investor information set. Given that overconditioning is typically a result of dividing the rolling-regression windows too finely, recent papers attempt to compute an optimal window size that minimizes overconditioning bias but still allows for sufficient time variation in factor loadings. Li and Yang (2008) and Ang and Kristensen (2010) apply nonparametric methods to estimate the conditional CAPM in a manner that is similar to Lewellen and Nagel (2006) but also includes a data-driven optimal window size. Essentially, the window size is chosen such that the window contains sufficient data to achieve estimation efficiency but is still short enough that the information structure does not substantially change.

Let z characterize the estimation window length. The objective of the kernel regression approach is to choose, at every time t, parameters $\alpha_{i,t}$ and $B_{i,t}$ to minimize the following quantity:

$$\min_{\alpha,B} \sum_{s=t-z}^{t+z} \left(r_{i,s} - \alpha_{i,t} - B'_{i,t} R_{m,s}\right)^2 k_{s,t}, \tag{B.3}$$

where

$$\hat{B}_{i,t} = \begin{bmatrix} \hat{\beta}_{i,0,t} & \hat{\beta}_{i,1,t} & \hat{\beta}_{i,2,t} \end{bmatrix}',$$
$$k_{s,t} = \frac{T}{z} k\left(\frac{s-t}{z}\right),$$

and $k(\cdot)$ is the Epanechnikov kernel:

$$k(u) = \frac{3}{4}\left(1-u^2\right) 1\left(|u| \leq 1\right).$$

Thus, this estimation approach uses data in the window $t-z$ to $t+z$ to estimate alpha and beta at time t. The functional form of $k(\cdot)$ implies that observations closer to time t receive more weight in estimating $\alpha_{i,t}$ and $B_{i,t}$. The researcher can either specify z or estimate an optimal window size as described below. Following the methodology outlined in Li and Yang (2008), I use the reflection method to create pseudodata for estimating alpha and beta close to the boundary areas. The optimal window size is obtained from the cross-validation method. Define $\hat{\alpha}_{-i,t}$ and $\hat{B}_{-i,t}$ as the 'leave one out' estimates from the following:

$$\min_{\alpha,B} \sum_{s=t-z,\ s\neq t}^{t+z} \left(r_{i,s} - \alpha_{-i,t} - B'_{-i,t} R_{m,s}\right)^2 k_{s,t}. \tag{B.4}$$

These estimates will be functions of the chosen window size, z. The optimal window size minimizes the following quantity:

$$Q(z) = \sum_{t=1}^{T} \left(r_{i,t} - \hat{\alpha}_{-i,t} - \hat{B}'_{-i,t} R_{m,t}\right)^2. \tag{B.5}$$

This approach is based on the idea that data in the vicinity of time t can be used to predict the observation at time t.

Solving the optimization problem in (B.3) yields a time series of daily alpha estimates. To assess the performance of the long-short distress portfolios relative to the conditional CAPM, I report the average of these daily pricing errors:

$$\hat{\alpha}_i^{KR} = \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_{i,t}. \tag{B.6}$$

Li and Yang (2008) show $\hat{\alpha}_i^{KR}$ is distributed asymptotically normal with variance equal to the $(1,1)$th element of

$$\frac{1}{T}\Omega_0^{-1} \sum_{l=-\infty}^{\infty} E(X_t X'_{t+l} \epsilon_t \epsilon_{t+l}) \Omega_0^{-1},$$

where $X_t = \left[1 \ R'_{m,t}\right]'$, $\Omega_0 = E(X_t X'_t)$, and l is the lag order.

Table B.1 (Method 4) reports the results. For each long-short distress portfolio, I compute average conditional alphas corresponding to monthly ($z = 11$ days), quarterly (31 days), and semiannual (63 days) windows, as well as an optimal window length determined from equation (B.5). For the CHS portfolio, the average pricing errors are quite similar across the various estimation windows, ranging from -0.25% to -0.18% per month. These estimates are also close to those obtained from the Lewellen and Nagel (2006) method. None is statistically significant at the 5% level. The optimal z is 52 days. The nonparametric alpha estimates for the O-score portfolio are similar to those from the Lewellen and Nagel (2006) approach. The optimal z for this portfolio is 60 days, corresponding to an average pricing error of -0.14% per month (t-statistic of -0.72).

B.4 Dummy variables approach

Finally, I estimate conditional CAPM regressions for the distress-sorted portfolios in the spirit of Fama and French (2006b). These regressions also follow equations (1.8) and (1.9), but include slope dummy variables to allow for periodic changes in betas. I allow for monthly, quarterly, or semianual changes in beta and estimate one portfolio intercept for the entire sample period. Table B.1 (Method 5) reports the estimated alphas in percent per month. The zero-cost CHS portfolio has conditional CAPM alphas ranging from -0.39% to -0.18% per month, and the long-short O-score portfolio alpha estimates are between -0.15% and 0.08% per month. None of these estimates is statistically different from zero.

Figure B.1 plots the conditional betas for the long-short CHS portfolio corresponding to Methods 1, 3, 4, and 5.[1] The pattern in portfolio betas is generally

[1]The conditional betas for Method 2 are the same as those for Method 1 with a lag of

similar across the estimation methodologies. Betas tend to be low in the late 1980s and high in the early 2000s.

In summary, the findings in this paper appear robust to reasonable alternative estimation methodologies. In particular, the insignificant alphas reported in Panel C of Table 1.1 are unlikely a result of the overconditioning problem outlined in Boguth, Carlson, Fisher, and Simutin (2009) or selected window lengths that are divided too finely. Moreover, the results hold when more traditional estimation methodologies based on conditioning information or dummy variables are applied.

---

one period.

Table B.1: Alternative estimation methodologies, 1981–2009

		LN		BCFS		State Vars.		Kernel Reg.		FF	
Parm.	Horizon	*CHS*	*O*-score	*CHS*	*O*-score	*CHS*	*O*-score	*CHS*	*O*-score	*CHS*	*O*-score
Conditional CAPM alphas (% monthly)											
	All					-0.44	-0.32				
						(-1.49)	(-1.56)				
$\uparrow$	M	-0.32	0.25	0.03	-0.23			-0.18	0.26	-0.32	0.08
		(-0.88)	(0.97)	(0.09)	(-1.04)			(-0.88)	(1.64)	(-1.18)	(0.39)
$\hat{\alpha}$	Q	-0.37	-0.08	-0.23	-0.24			-0.25	-0.04	-0.39	-0.11
		(-1.07)	(-0.36)	(-0.76)	(-1.12)			(-1.05)	(-0.21)	(-1.49)	(-0.56)
$\downarrow$	SA	-0.17	-0.13	-0.48	-0.28			-0.22	-0.14	-0.18	-0.15
		(-0.55)	(-0.64)	(-1.56)	(-1.31)			(-0.88)	(-0.74)	(-0.70)	(-0.73)
	Optimal							-0.22	-0.14		
								(-0.90)	(-0.72)		
Other parameter estimates											
$\hat{b}_{i,Const.}$						-4.00	-2.78				
						(-11.6)	(-11.6)				
$\hat{b}_{i,DEF}$						45.4	29.3				
						(9.89)	(9.17)				
$\hat{b}_{i,DP}$						-1.08	-0.73				
						(-13.4)	(-13.0)				

Note: The table presents conditional CAPM regressions for long-short portfolios sorted on financial distress. The portfolios are value weighted and rebalanced annually. The estimation methodologies are described in the text. ‘LN’ is the Lewellen and Nagel (2006) rolling-regression approach. ‘BCFS’ is the lagged portfolio approach outlined in Boguth, Carlson, Fisher, and Simutin (2009). ‘State Vars.’ is based on the following time-series regression: $r_{i,t} = \alpha_i + (b_i Z_{t-1})' R_{m,t} + \epsilon_{i,t}$. ‘Kernel Reg.’ is the kernel regression approach of Li and Yang (2008). ‘FF’ is the Fama–French (2006b) regression approach. Alphas are expressed in percent per month. The numbers in parentheses are t-statistics.

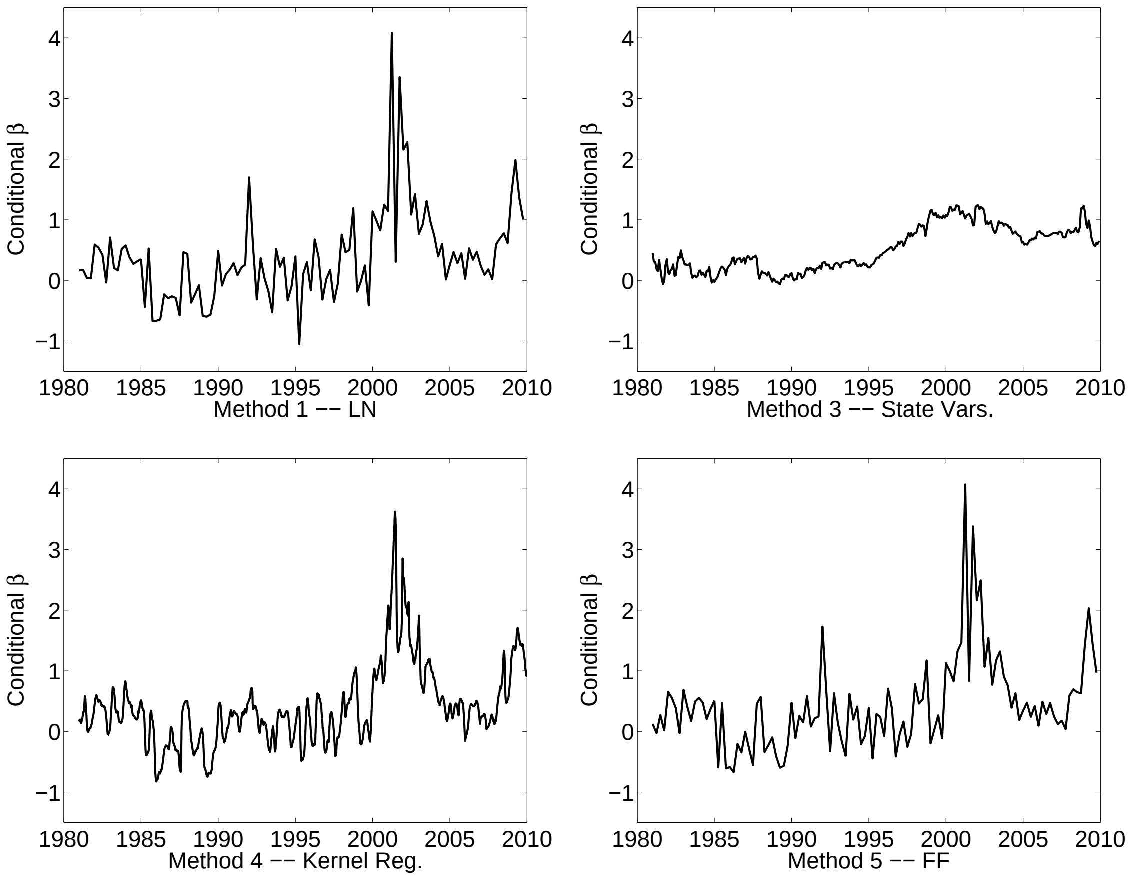


Figure B.1: Conditional beta for portfolio sorted on financial distress: Alternative estimation methodologies, 1981–2009

The figure presents conditional CAPM regression betas for the hedge portfolio that is long the top quintile and short the bottom quintile of firms sorted on Campbell, Hilscher, and Szilagyi's (2008) failure probability. Each plot corresponds to a different estimation approach as described in the text. 'LN' is the Lewellen and Nagel (2006) rolling-regression approach with quarterly windows. 'State Vars.' is based on the following time-series regression: $r_{i,t} = \alpha_i + (b_i Z_{t-1})' R_{m,t} + \epsilon_{i,t}$. 'Kernel Reg.' is the kernel regression approach of Li and Yang (2008) with an optimally selected window size. 'FF' is the Fama–French (2006b) regression approach with quarterly changes in beta.

APPENDIX C
DATA AND MODEL APPENDIX FOR CHAPTER 2

Section C.1 provides details about the MCMC estimation algorithm, and Section C.2 presents a simulation study that demonstrates the ability of the algorithm to accurately recover parameters. Section C.3 describes the construction of the anomaly variables used in the empirical analysis.

C.1 Estimation methodology

The model outlined in equations (2.5) to (2.7) can be estimated by repeatedly cycling through steps 1 to 6 below. As discussed in the text, we place a hierarchical structure on alphas, but not on betas. Instead we impose a proper, but diffuse, prior directly on betas in the base specification, $\beta_{i,y} \sim N\left(\mu = 1, \sigma_\beta^2 = 10\right)$. Let $r_{i,t,y}^e$ denote the excess return on stock i in month t of year y and $r_{m,t,y}^e$ the excess return on the market portfolio. Further, let Z denote a matrix in which the first column is a vector of ones and the second column is the excess returns on the market portfolio, and let X denote a matrix of a constant and firm-year characteristics associated with anomalies.

1. Draw $\alpha_{i,y}, \beta_{i,y} | \sigma_{i,y}^2, \delta_y, \sigma_{\alpha,y}^2$ for each stock $i = 1, ..., N$, in each year $y = 1, ..., Y$. We obtain a draw from the marginal posterior distribution of $\alpha_{i,y}$ and $\beta_{i,y}$ as follows:

$$\begin{bmatrix} \alpha_{i,y} \\ \beta_{i,y} \end{bmatrix} \sim N\left(\overline{\lambda}_i, (\sigma_{i,y}^{-2} Z_{i,y}' Z_{i,y} + \mathbf{V}_\lambda^{-1})^{-1}\right), \tag{C.1}$$

where

$$\overline{\lambda}_i = (\sigma_{i,y}^{-2} Z_{i,y}' Z_{i,y} + \mathbf{V}_\lambda^{-1})^{-1} (\sigma_{i,y}^{-2} Z_{i,y}' Z_{i,y} \widehat{\lambda}_i + \mathbf{V}_\lambda^{-1} \overline{\overline{\lambda}}_{i,y}), \tag{C.2}$$

$$\widehat{\lambda}_i = \left(Z_{i,y}' Z_{i,y}\right)^{-1} Z_{i,y}' r_{i,y}^e, \tag{C.3}$$

$$\overline{\overline{\lambda}}_{i,y} = \begin{bmatrix} X_{i,y}\delta_y \\ 1 \end{bmatrix}, \tag{C.4}$$

and

$$\mathbf{V}_\lambda = \begin{bmatrix} \sigma_{\alpha,y}^2 & 0 \\ 0 & 10 \end{bmatrix}. \tag{C.5}$$

2. Draw $\sigma_{i,y}^2 \,|\alpha_{i,y}, \beta_{i,y}$ for each stock $i = 1, ..., N$, in each year $y = 1, ..., Y$. We obtain a draw from the marginal posterior distribution of $\sigma_{i,y}^2$ as follows:

$$\sigma_{i,y}^2 \sim \text{Inverse Gamma}\left(\frac{v_1 s_1^2}{2}, \frac{v_1}{2}\right), \tag{C.6}$$

$$v_1 = v_0 + M, \tag{C.7}$$

and

$$s_1^2 = \frac{v_0 s_0^2 + s^2}{v_0 + M}, \tag{C.8}$$

where s^2 is the sample sum of squared errors and M denotes the number of observations. The priors, v_0 and s_0^2, are determined by the researcher. We set v_0 equal to 3 and s_0^2 equal to the variance of the monthly returns for stock i in year y.

3. Draw $\delta_y | \left\{\alpha_{i,y}\right\}, \sigma_{\alpha,y}^2, \overline{\delta}, \mathbf{V}$ for each year $y = 1, ..., Y$. Let α denote a column vector composed of draws of $\alpha_{i,y}$ for all firms i in the dataset in year y. We obtain a draw from the marginal posterior distribution of δ_y as follows:

$$\delta_y \sim N\left(\overline{\overline{\delta}}_y, (\sigma_{\alpha,y}^{-2} X_y' X_y + \mathbf{V}^{-1})^{-1}\right), \tag{C.9}$$

where

$$\bar{\bar{\delta}}_y = (\sigma_{\alpha,y}^{-2} X_y' X_y + \mathbf{V}^{-1})^{-1} (\sigma_{\alpha,y}^{-2} X_y' X_y \widehat{\delta}_y + \mathbf{V}^{-1} \bar{\delta}), \tag{C.10}$$

and

$$\widehat{\delta}_y = \left(X_y' X_y\right)^{-1} X_y' \alpha. \tag{C.11}$$

4. Draw $\sigma_{\alpha,y}^2 \mid \{\alpha_{i,y}\}, \delta_y$ for each year $y = 1, ..., Y$. We obtain a draw from the marginal posterior distribution of $\sigma_{\alpha,y}^2$ as follows:

$$\sigma_{\alpha,y}^2 \sim \text{Inverse Gamma}\left(\frac{v_1 s_1^2}{2}, \frac{v_1}{2}\right), \tag{C.12}$$

$$v_1 = v_0 + M, \tag{C.13}$$

and

$$s_1^2 = \frac{v_0 s_0^2 + s^2}{v_0 + M}, \tag{C.14}$$

where s^2 is the sample sum of squared errors and M denotes the number of observations. The priors, v_0 and s_0^2, are determined by the researcher. We set v_0 equal to 3. We elicit priors for s_0^2 in the following manner. For each stock in year y we estimate equation (2.5) using OLS and store $\widehat{\alpha}$. We set s_0^2 equal to the variance of $\widehat{\alpha}$ across all firms in year y.

Having drawn the firm- and year-level coefficients we proceed to draw the aggregate-level parameters. Let P denote a $Y \times nvar$ matrix comprised of a draw of $\{\delta_y\}_{y=1}^{Y}$, where $nvar$ denotes the number of columns in X, and let H be a matrix of covariates the researcher believes to be associated with the evolution of the parameter vector δ_y over time. In our specification, H is a column vector of ones, but could easily be extended, for example, to include macroeconomic variables.

5. Draw $\mathbf{V}|\left\{\delta_y\right\}$. We obtain a draw from the marginal posterior distribution of $\mathbf{V}$ as follows:

$$\mathbf{V} \sim \text{Inverse Wishart}\left(nvar + Nu + Y, \mathbf{V}_0 + \mathbf{S}\right), \tag{C.15}$$

where

$$\mathbf{S} = \left(P - H\widetilde{\mathbf{\Gamma}}\right)'\left(P - H\widetilde{\mathbf{\Gamma}}\right) + \left(\widetilde{\mathbf{\Gamma}} - \overline{\mathbf{\Gamma}}\right)'\mathbf{A}\left(\widetilde{\mathbf{\Gamma}} - \overline{\mathbf{\Gamma}}\right), \tag{C.16}$$

$$\widetilde{\mathbf{\Gamma}} = \left(\left(H'H + \mathbf{A}\right)^{-1}\left(H'H\widehat{\mathbf{\Gamma}} + \mathbf{A}\overline{\mathbf{\Gamma}}\right)\right), \tag{C.17}$$

and

$$\widehat{\mathbf{\Gamma}} = \left(H'H\right)^{-1}\left(H'P\right). \tag{C.18}$$

$\mathbf{A}$, $\overline{\mathbf{\Gamma}}, Nu$ and $\mathbf{V}_0$ are priors specified by the researcher. We set $\mathbf{A}^{-1} = 100\mathbf{I}$ and define $\overline{\mathbf{\Gamma}}$ to be an $n_H \times nvar$ matrix of zeros, where n_H denotes the number of columns in H. Nu is set to $nvar + 3$, and $\mathbf{V}_0 = Nu\mathbf{I}$. $\mathbf{I}$ denotes an appropriately dimensioned identity matrix.

6. Draw $\gamma|\left\{\delta_y\right\}, \mathbf{V}$. We obtain a draw from the marginal posterior distribution of γ as follows:

$$\gamma \sim N\left(\widetilde{\gamma}, \mathbf{V} \otimes \left(H'H + \mathbf{A}\right)^{-1}\right), \tag{C.19}$$

where $\widetilde{\gamma} = vec\left(\widetilde{\mathbf{\Gamma}}\right)$. Given that H is a vector of ones, $\overline{\delta} = \gamma$.

C.2 Model simulation

In this section, we conduct a simulation exercise and show our estimation algorithm successfully recovers the parameters of interest. Data are created for 1,000 firms over a 45-year period. The length of each time period, y, is set to 12 months. We assume there are two firm characteristics associated with firm-level alphas, x_1 and x_2, which are both uniformly distributed over the range -0.5 to +0.5. The parameters in the simulation are set to ensure that the simulated firm-level returns, alphas, betas, and market returns are consistent with the actual values observed using the CRSP return data.

1. Draw $\delta_y \sim \text{MVN}\left(\mu = \overline{\delta}, \sigma^2 = \mathbf{V}\right)$ for each 12-month time period, y. We set

$$\overline{\delta} = \begin{bmatrix} \overline{\delta}_0 = 0 \\ \overline{\delta}_1 = 1 \\ \overline{\delta}_2 = 1 \end{bmatrix}, \text{ and } \mathbf{V} = \begin{bmatrix} 1.5 & 0.5 & 0.5 \\ 0.5 & 1.5 & 0.5 \\ 0.5 & 0.5 & 1.5 \end{bmatrix}.$$

2. Draw $\alpha_y \sim \text{MVN}\left(\mu = \delta_{0,y} + \delta_{1,y}x_1 + \delta_{2,y}x_2, \sigma^2 = \mathbf{\Sigma}_\alpha\right)$, where α_y is a column vectors of firm-specific alphas in time period y. We consider two specifications for the variance-covariance matrix, $\mathbf{\Sigma}_\alpha$, one in which the error terms are independent across firms, and one in which the error terms are correlated across firms. We examine two different levels of correlations, low to medium with correlations ranging from -0.5 to $+0.5$, and medium to high with correlations ranging from -0.9 to $+0.9$. The diagonal elements of $\mathbf{\Sigma}_\alpha$ are set equal to $\sigma_\alpha^2 = 2$.[1]

3. Draw $\beta_{i,y} \sim \text{N}\left(\mu = 1, \sigma^2 = 4\right)$ for each firm i in each time period y.

---

[1]We use the following procedure to create a 1,000 $\times$ 1,000 variance-covariance matrix. First, create a column vector, u, with 1,000 draws from the Uniform(-1,1) distribution. Second, calculate $\kappa uu'$ where $\kappa = p\sigma_\alpha^2$. The parameter, p is a scaling factor, between 0 and 1, for the maximum level of correlation in the error terms across firms. If $p = 0$, firm-level alphas are independent. If $p = 1$, $\kappa uu'$ correlations range from -1 to $+1$. For low to medium correlations we set $p = 0.5$, while for medium to high correlations we set $p = 0.9$. Finally, set $\mathbf{\Sigma}_\alpha = \kappa uu'$ and replace the diagonal elements with $\sigma_\alpha^2 = 2$.

4. Generate excess monthly returns on the market: $r^e_{m,t,y} \sim \mathrm{N}\left(\mu = 0.5, \sigma^2 = 25\right)$.

5. Generate monthly excess returns for each firm in each month of each time period: $r^e_{t,y} = \alpha_y + \beta_y r^e_{m,t,y} + \epsilon_{t,y}$, where $\epsilon_{t,y} \sim \mathrm{MVN}\left(\mu = 0, \sigma^2 = \mathbf{\Sigma}_{ret}\right)$ and $r^e_{t,y}$ denotes a column vector of excess returns for all firms in month t of time period y. α_y and β_y are column vectors of firm-specific alphas and betas. The specifications for the variance-covariance matrix, $\mathbf{\Sigma}_{ret}$, are constructed in a similar manner to those for $\mathbf{\Sigma}_{\alpha}$. The only difference is that the diagonal elements of $\mathbf{\Sigma}_{ret}$, σ^2_{ret}, are set equal to 169.

We examine seven different scenarios to investigate the sensitivity of our model to different correlation structures in the error terms of equations (2.5) and (2.6). The MCMC algorithm is run for 1,000 iterations for each scenario. The algorithm converges quickly. The posterior distributions are characterized using the final 500 iterations. We use the same seed for the random number generator for each scenario. Table C.1 reports the results from the simulation study. Regardless of the correlation structure in the error terms of equations (2.5) and (2.6), the estimation algorithm is able to accurately recover the aggregate-level model parameters, $\overline{\delta}$ and $\mathbf{V}$, indicating that the approach is not sensitive to the possibility of cross-correlations across firms.

C.3 Data description

We obtain accounting data from the Compustat Fundamentals Annual files and stock return data from the CRSP monthly return files. Each of the anomaly variables is measured once a year at the end of June in calendar year j. The variables are matched to returns from July of calendar year j to June of calendar year $j+1$. To ensure that the accounting data are known prior to the returns they are used to forecast, we lag all accounting variables by six months. The sample includes all NYSE, Amex, and NASDAQ ordinary common stocks with the data required to compute at least one of the following anomaly variables:

1. Size (M): The natural log of price per share times the number of shares outstanding at the end of June of year j.

2. Book-to-market (BM): The natural log of the ratio of book value of equity to market value of equity. Following Fama and French (2008), we define the book value of equity as total assets (at), minus total liabilities (lt), plus balance sheet deferred taxes and investment tax credits (txditc) if available, minus the book value of preferred stock if available. Depending on availability, we use liquidating value (pstkl), redemption value (pstkrv), or carrying value (upstk) for the the book value of preferred stock. The market value of equity is price per share times the number of shares outstanding at the end of December of year $j-1$.

3. Momentum (MOM): The continuously compounded stock return from January to June of year j. We require a firm to have a price for the end of December of year $j-1$ and a good return for June of year j.

4. Reversal (REV): The continuously compounded stock return from July of year $j-5$ to June of year $j-1$. We require a firm to have a price for the end of June of year $j-5$ and a good return for June of year $j-1$.

5. Profitability (ROA): Income before extraordinary items (ib), minus dividends on preferred (dvp) if available, plus income statement deferred taxes (txdi) if available divided by total assets (at).

6. Asset growth (AG): Total assets (at) at the fiscal year end in year $j-1$, minus total assets at the fiscal year end in year $j-2$ divided by total assets at the fiscal year end in year $j-2$. We also require a firm to have non-zero total assets in both year $j-1$ and $j-2$.

7. Net stock issues (NS): The natural log of the ratio of split-adjusted shares at the fiscal year end in year $j-1$ divided by split-adjusted shares at the fiscal year end in year $j-2$. The number of split-adjusted shares outstanding is common shares outstanding from Compustat (csho) times the cumulative adjustment factor by ex-date (adjex_f).

8. Accruals (ACC): The change in current assets (act) from the fiscal year end in year $j-2$ to $j-1$, minus the change in current liabilities (lct), minus the change in cash and short-term investments (che), plus the change in debt in current liabilities (dlc), minus depreciation (dp) in fiscal year $j-1$ divided by total assets (at) from the fiscal year end in year $j-2$.

9. Financial distress (OS): Ohlson's (1980) O-score:

$$O\text{-score} = \frac{1}{1+\exp(-x)},$$

where

$$
\begin{aligned}
x = & -1.32 - 0.407\,(SIZE) + 6.03\,(TLTA) - 1.43\,(WCTA) \\
& + 0.076\,(CLCA) - 1.72\,(OENEG) - 2.37\,(NITA) - 1.83\,(FUTL) \\
& + 0.285\,(INTWO) - 0.521\,(CHIN)\,,
\end{aligned}
$$

where $SIZE$ is the log of the ratio of total assets (at) to the GNP price-level index, $TLTA$ is the ratio of total liabilities (lt) to total assets, $WCTA$ is the ratio of working capital (act – lct) to total assets, $CLCA$ is the ratio of current liabilities (lct) to current assets (act), $OENEG$ is a dummy variable equal to one if total liabilities exceeds total assets and zero otherwise, $NITA$ is the ratio of net income (ni) to total assets, $FUTL$ is the ratio of funds from operations (pi) to total liabilities, $INTWO$ is a dummy variable equal to one if total net income was negative for the past two years and zero otherwise, and $CHIN$ is the change in net income from fiscal year $j-2$ to $j-1$ divided by the sum of the absolute values of net income in fiscal years $j-2$ and $j-1$. Data on the GNP price-level index are from the Federal Reserve Bank of St. Louis website.[2] Following Ohlson (1980), we assign the index a value of 100 in 1968, and the index year is as of the year prior to the year of the balance sheet date.

We exclude financial firms (SIC codes between 6000 and 6999) and firms with negative book equity. The sample period is July 1963 to June 2008. To alleviate the influence of outliers, we winsorize ROA, AG, NS, and ACC at the 1st and 99th

[2] http://research.stlouisfed.org/fred2/.

percentiles. For cases in which a firm is delisted from an exchange during a given month, we replace any missing returns with the delisting returns provided by CRSP.

Table C.1: Model estimation on simulated data

Case	Cross-Correlation in Return Error (Equation (2.5))	Cross-Correlation in Alpha Error (Equation (2.6))	$\overline{\delta}_0$	$\overline{\delta}_1$	$\overline{\delta}_2$	$\mathbf{V}_{11}$	$\mathbf{V}_{22}$	$\mathbf{V}_{33}$	$\mathbf{V}_{12}$	$\mathbf{V}_{13}$	$\mathbf{V}_{23}$
True Values			0.00	1.00	1.00	1.50	1.50	1.50	0.50	0.50	0.50
			Posterior Means for the Aggregate-Level Parameters, $\overline{\delta}$ and $\mathbf{V}$								
Case 1	None	None	-0.22	0.94	1.07	1.86	1.74	1.67	0.57	0.68	0.75
			(0.20)	(0.20)	(0.21)	(0.39)	(0.38)	(0.40)	(0.29)	(0.31)	(0.28)
Case 2	Low	None	-0.21	0.87	1.06	1.86	1.79	1.65	0.51	0.67	0.78
			(0.20)	(0.20)	(0.21)	(0.39)	(0.39)	(0.40)	(0.29)	(0.31)	(0.29)
Case 3	High	None	-0.20	0.84	1.08	1.87	1.79	1.65	0.49	0.66	0.78
			(0.20)	(0.20)	(0.21)	(0.39)	(0.39)	(0.40)	(0.29)	(0.31)	(0.29)
Case 4	None	Low	-0.22	0.93	1.07	1.86	1.74	1.72	0.56	0.68	0.79
			(0.20)	(0.20)	(0.21)	(0.39)	(0.38)	(0.41)	(0.29)	(0.31)	(0.29)
Case 5	None	High	-0.22	0.93	1.07	1.86	1.74	1.72	0.56	0.67	0.80
			(0.20)	(0.20)	(0.21)	(0.39)	(0.38)	(0.41)	(0.29)	(0.31)	(0.29)
Case 6	Low	Low	-0.21	0.86	1.06	1.86	1.82	1.70	0.51	0.66	0.82
			(0.20)	(0.20)	(0.21)	(0.39)	(0.40)	(0.41)	(0.30)	(0.31)	(0.30)
Case 7	High	High	-0.20	0.83	1.08	1.88	1.82	1.69	0.49	0.64	0.84
			(0.20)	(0.20)	(0.21)	(0.39)	(0.40)	(0.41)	(0.30)	(0.31)	(0.30)

Note: The table presents the results from the estimation of the model described in equations (2.5) to (2.7) for simulated data. We report the posterior mean and standard deviation for the aggregate-level parameters $\overline{\delta}$ and $\mathbf{V}$. We simulate data for 1,000 firms over a 45-year period. We create seven different sets of data using the same seed for the random number generator in each scenario. Each set of data differs only with respect to the assumptions about cross-correlations in the error terms of equation (2.5) (monthly firm returns) and/or equation (2.6) (firm-year alphas). Specifically, we allow cross-correlations in each equation to take on one of three levels: zero, low, or high. The low level allows cross-correlations to range between ∓ 0.5, while the high level allows cross-correlations to range between ∓ 0.9. We run the Gibbs sampler for 1,000 iterations and discard the first 500 as a burn-in period.

REFERENCES

Ahn, D.-H., J. Conrad, and R. F. Dittmar (2009): "Basis assets," *Review of Financial Studies*, 22, 5133–5174.

Altman, E. I. (1968): "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *Journal of Finance*, 23, 589–609.

Ang, A., and J. Chen (2007): "CAPM over the long run: 1926–2001," *Journal of Empirical Finance*, 14, 1–40.

Ang, A., and D. Kristensen (2010): "Testing conditional factor models," Working paper, Columbia University.

Ang, A., J. Liu, and K. Schwarz (2010): "Using stocks or portfolios in tests of factor models," Working paper, Columbia University.

Avramov, D., S. Cederburg, and S. Hore (2010): "Cross-sectional asset pricing puzzles: An equilibrium perspective," Working paper, University of Maryland.

Avramov, D., and T. Chordia (2006): "Asset pricing models and financial market anomalies," *Review of Financial Studies*, 19, 1001–1040.

Avramov, D., T. Chordia, G. Jostova, and A. Phillipov (2009): "Credit ratings and the cross-section of stock returns," *Journal of Financial Markets*, 12, 469–499.

Banz, R. W. (1981): "The relationship between return and market value of common stocks," *Journal of Financial Economics*, 9, 3–18.

Barron, O. E., M. H. Stanford, and Y. Yu (2009): "Further evidence on the relation between analysts' forecast dispersion and stock returns," *Contemporary Accounting Research*, 26, 329–357.

Bates, J., and C. Granger (1969): "The combination of forecasts," *Operational Research Quarterly*, 20, 451–468.

Boguth, O., M. Carlson, A. Fisher, and M. Simutin (2009): "Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas," Working paper, University of British Columbia.

Box, G. (1980): "Sampling and Bayes inference in scientific modeling and robustness," *Journal of the Royal Statistical Society Series A*, 143, 383–430.

Brennan, M. J., T. Chordia, and A. Subrahmanyam (1998): "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, 49, 345–373.

Campbell, J. Y., and J. H. Cochrane (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 102, 205–251.

Campbell, J. Y., J. Hilscher, and J. Szilagyi (2008): "In search of distress risk," *Journal of Finance*, 63, 2899–2939.

Carhart, M. (1997): "On persistence in mutual fund performance," *Journal of Finance*, 52, 57–82.

Cederburg, S. (2010): "Intertemporal risk and the cross section of expected stock returns," Working paper, University of Iowa.

Chan, L. K., Y. Hamao, and J. Lakonishok (1991): "Fundamentals and stock returns in Japan," *Journal of Finance*, 46, 1739–1764.

Chava, S., and R. A. Jarrow (2004): "Bankruptcy prediction with industry effects," *Review of Finance*, 8, 537–569.

Chava, S., and A. K. Purnanandam (2010): "Is default risk negatively related to stock returns?," *Review of Financial Studies*, 23, 2523–2559.

Chen, L., R. Novy-Marx, and L. Zhang (2010): "An alternative three-factor model," Working paper, Washington University in St. Louis.

Chen, L., and L. Zhang (2009): "A better three-factor model that explains more anomalies," Forthcoming in *Journal of Finance*.

Chopra, N., J. Lakonishok, and J. R. Ritter (1992): "Measuring abnormal performance: Do stocks overreact?," *Journal of Financial Economics*, 31, 235–268.

Clemen, R. (1989): "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5, 559–581.

Cochrane, J. (2006): "Financial markets and the real economy," Working paper, University of Chicago.

Cochrane, J. H. (2001): *Asset Pricing*. Princeton University Press, Princeton, NJ.

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.

Collins, D. W., and P. Hribar (2000): "Earnings-based and accrual-based market anomalies: One effect or two?," *Journal of Accounting and Economics*, 29, 101–123.

Conrad, J., M. Cooper, and G. Kaul (2003): "Value versus glamour," *Journal of Finance*, 58, 1969–1996.

Constantinides, G. M., and D. Duffie (1996): "Asset pricing with heterogeneous consumers," *Journal of Political Economy*, 104, 219–240.

Cooper, M. J., and S. Gubellini (2009): "The critical role of conditioning information in determining if value is riskier than growth," Working paper, University of Utah.

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset growth and the cross-section of stock returns," *Journal of Finance*, 63, 1609–1651.

Cremers, K. J. M. (2006): "Multifactor efficiency and Bayesian inference," *Journal of Business*, 79, 2951–2998.

Cremers, M., and H. Yan (2009): "Uncertainty and valuations," Working paper, Yale University.

Da, Z., and P. Gao (2010): "Clientele change, liquidity shock, and the return on financially distressed stocks," *Journal of Financial and Quantitative Analysis*, 45, 27–48.

Daniel, K., and S. Titman (2006a): "Market reactions to tangible and intangible information," *Journal of Finance*, 61, 1605–1643.

——— (2006b): "Testing factor-model explanations of market anomalies," Working paper, Northwestern University.

Davies, P. (2010): "A firm-level test of the CAPM," Working paper, University of Iowa.

DeBondt, W. F. M., and R. Thaler (1985): "Does the stock market overreact?," *Journal of Finance*, 40, 793–805.

Dichev, I. D. (1998): "Is the risk of bankruptcy a systematic risk?," *Journal of Finance*, 53, 1131–1147.

Diebold, F., and J. Lopez (1996): "Forecast evaluation and combination," in *Statistical Methods in Finance*, ed. by G. Maddala, and C. Rao, vol. 14, pp. 241–268. Elsevier Science B.V., Amsterdam.

Diether, K. B., C. J. Malloy, and A. Scherbina (2002): "Differences of opinion and the cross section of stock returns," *Journal of Finance*, 57, 2113–2141.

Dimson, E. (1979): "Risk measurement when shares are subject to infrequent trading," *Journal of Financial Economics*, 7, 197–226.

Fairfield, P. M., J. S. Whisenant, and T. L. Yohn (2003): "Accrued earnings and growth: Implications for future profitability and market mispricing," *The Accounting Review*, 78, 353–371.

Fama, E. F., and K. R. French (1989): "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 25, 23–49.

——— (1992): "The cross-section of expected stock returns," *Journal of Finance*, 47, 427–465.

——— (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

——— (1996): "Mutlifactor explanations of asset pricing anomalies," *Journal of Finance*, 51, 55–84.

——— (2006a): "Profitability, investment, and average returns," *Journal of Financial Economics*, 82, 491–518.

——— (2006b): "The value premium and the CAPM," *Journal of Finance*, 61, 2163–2185.

——— (2008): "Dissecting anomalies," *Journal of Finance*, 63, 1653–1678.

Fama, E. F., and J. MacBeth (1973): "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 81, 607–636.

Ferson, W. E., and C. R. Harvey (1999): "Conditioning variables and the cross section of stock returns," *Journal of Finance*, 54, 1325–1360.

Ferson, W. E., and R. W. Schadt (1996): "Measuring fund strategy and performance in changing economic conditions," *Journal of Finance*, 51, 425–462.

Garlappi, L., T. Shu, and H. Yan (2008): "Default risk, shareholder advantage, and stock returns," *Review of Financial Studies*, 21, 2743–2778.

Garlappi, L., and H. Yan (2010): "Financial distress and the cross-section of equity returns," Forthcoming in *Journal of Finance*.

George, T. J., and C.-Y. Hwang (2010): "A resolution of the distress risk and leverage puzzles in the cross section of stock returns," *Journal of Financial Economics*, 96, 56–79.

Geweke, J., and G. Amisano (2010): "Optimal prediction pools," Working paper, Center for Quantitative Economic Research.

Geweke, J., and G. Zhou (1996): "Measuring the pricing error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557–587.

Ghysels, E. (1998): "On stable factor structures in the pricing of risk: Do time-varying betas help or hurt?," *Journal of Finance*, 53, 549–573.

Gomes, J., L. Kogan, and L. Zhang (2003): "Equilibrium cross section of returns," *Journal of Political Economy*, 111, 693–732.

Good, I. (1952): "Rational decisions," *Journal of the Royal Statistical Society Series B*, 14, 107–114.

Griffin, J. M., and M. L. Lemmon (2002): “Book-to-market equity, distress risk, and stock returns,” *Journal of Finance*, 57, 2317–2335.

Guidolin, M., and A. Timmermann (2009): “Forecasts of U.S. short-term interest rates: A flexible forecast combination approach,” *Journal of Econometrics*, 150, 297–311.

Güntay, L., and D. Hackbarth (2010): “Corporate bond credit spreads and forecast dispersion,” Working paper, University of Illinois.

Hall, S., and J. Mitchell (2007): “Combining density forecasts,” *International Journal of Forecasting*, 23, 1–13.

Harvey, C. R. (1989): “Time-varying conditional covariances in tests of asset pricing models,” *Journal of Financial Economics*, 25, 289–317.

Harvey, C. R., and G. Zhou (1990): “Bayesian inference in asset pricing tests,” *Journal of Financial Economics*, 26, 221–254.

Haugen, R. A., and N. L. Baker (1996): “Commonality in the determinants of expected stock returns,” *Journal of Financial Economics*, 41, 401–439.

Hendry, D., and M. Clements (2002): “Pooling of forecasts,” *Econometrics Journal*, 5, 1–26.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1998): “Bayesian model averaging: A tutorial,” *Statistical Science*, 14, 382–417.

Ikenberry, D., J. Lakonishok, and T. Vermaelen (1995): “Market underreaction to open market share repurchases,” *Journal of Financial Economics*, 39, 181–208.

Jagannathan, R., and Z. Wang (1996): “The conditional CAPM and the cross-section of expected returns,” *Journal of Finance*, 51, 3–53.

Jegadeesh, N. (1990): “Evidence of predictable behavior of security returns,” *Journal of Finance*, 45, 881–898.

Jegadeesh, N., and S. Titman (1993): “Returns to buying winners and selling losers: Implications for stock market efficiency,” *Journal of Finance*, 48, 65–91.

——— (2001): “Profitability of momentum strategies: An evaluation of alternative explanations,” *Journal of Finance*, 56, 699–720.

Jensen, M. C. (1978): “Some anomalous evidence regarding market efficiency,” *Journal of Financial Economics*, 6, 95–102.

Johnson, T., T. Chebonenko, I. Cunha, F. D’Almeida, and X. Spencer (2010): “Endogenous leverage and expected stock returns: A comment,” Working paper, University of Illinois.

Johnson, T. C. (2004): "Forecast dispersion and the cross section of expected returns," *Journal of Finance*, 59, 1957–1978.

Kan, R. (2004): "On the explanatory power of asset pricing models across and within portfolios," Working paper, University of Toronto.

Kandel, S., R. McCulloch, and R. F. Stambaugh (1995): "Bayesian inference and portfolio efficiency," *Review of Financial Studies*, 8, 1–53.

Kandel, S., and R. F. Stambaugh (1995): "Portfolio inefficiency and the cross section of expected returns," *Journal of Finance*, 50, 157–184.

Karolyi, G. A. (1992): "Predicting risk: Some new generalizations," *Management Science*, 38, 57–74.

Korteweg, A., and N. Polson (2009): "Corporate credit spreads under parameter uncertainty," Working paper, Stanford University.

Leland, H. E. (1994): "Corporate debt value, bond covenants, and optimal capital structure," *Journal of Finance*, 49, 1213–1252.

Lettau, M., and S. Ludvigson (2001): "Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying," *Journal of Political Economy*, 109, 1238–1287.

Lewellen, J., and S. Nagel (2006): "The conditional CAPM does not explain asset-pricing anomalies," *Journal of Financial Economics*, 82, 289–314.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset-pricing tests," *Journal of Financial Economics*, 96, 175–194.

Li, Y., and L. Yang (2008): "Underconditioning and overconditioning: Testing the conditional CAPM and the conditional Fama–French three-factor model," Working paper, Cornell University.

Lintner, J. (1965): "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, 47, 13–37.

Litzenberger, R. H., and K. Ramaswamy (1979): "The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence," *Journal of Financial Economics*, 7, 163–195.

Lo, A. W., and A. C. MacKinlay (1990): "Data-snooping biases in tests of financial asset pricing models," *Review of Financial Studies*, 3, 431–468.

Loughran, T. (1997): "Book-to-market across firm size, exchange, and seasonality," *Journal of Financial and Quantitative Analysis*, 32, 249–268.

Loughran, T., and J. R. Ritter (1995): “The new issues puzzle,” *Journal of Finance*, 50, 23–51.

Mamaysky, H., M. Spiegel, and H. Zhang (2007): “Improved forecasting of mutual fund alphas and betas,” *Review of Finance*, 11, 359–400.

McCulloch, R., and P. E. Rossi (1991): “A Bayesian approach to testing the Arbitrage Pricing Theory,” *Journal of Econometrics*, 49, 141–168.

Merton, R. C. (1974): “On the pricing of corporate debt: The risk structure of interest rates,” *Journal of Finance*, 29, 449–470.

Newbold, P., and D. Harvey (2001): “Forecast combination and encompassing,” in *A Companion to Economic Forecasting*, ed. by M. Clements, and D. Hendry. Blackwell, Oxford.

Ohlson, J. A. (1980): “Financial ratios and the probabilistic prediction of bankruptcy,” *Journal of Accounting Research*, 18, 109–131.

Pástor, Ľ., and R. Stambaugh (2003): “Liquidity risk and rxpected stock returns,” *Journal of Political Economy*, 111, 642–685.

Pástor, Ľ., and P. Veronesi (2003): “Stock valuation and learning about profitability,” *Journal of Finance*, 58, 1749–1789.

Petkova, R., and L. Zhang (2005): “Is value riskier than growth?,” *Journal of Financial Economics*, 78, 187–202.

Pontiff, J., and A. Woodgate (2008): “Share issuance and cross-sectional returns,” *Journal of Finance*, 63, 921–945.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010): “Out-of-sample equity premium prediction: Combination forecasts and links to the real economy,” *Review of Financial Studies*, 23, 821–862.

Roll, R. (1977): “A critique of the asset pricing theory’s tests – Part I: On past and potential testability of the theory,” *Journal of Financial Economics*, 4, 129–176.

Rosenberg, B., K. Reid, and R. Lanstein (1985): “Persuasive evidence of market inefficiency,” *Journal of Portfolio Management*, 11, 9–17.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2005): *Bayesian Statistics and Marketing*. John Wiley & Sons, West Sussex, England.

Schwert, G. W. (2003): “Anomalies and market efficiency,” in *Handbook of the Economics of Finance*, ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. North Holland.

Shanken, J. (1987): “A Bayesian approach to testing portfolio efficiency,” *Journal of Financial Economics*, 19, 195–215.

Shanken, J., and G. Zhou (2007): "Estimating and testing beta pricing models: Alternative methods and their performance in simulations," *Journal of Financial Economics*, 84, 40–86.

Shanken, J. A. (1990): "Intertemporal asset pricing: An empirical investigation," *Journal of Econometrics*, 45, 99–120.

Sharpe, W. F. (1964): "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, 19, 425–442.

Shumway, T. (1997): "The delisting bias in CRSP data," *Journal of Finance*, 52, 327–340.

——— (2001): "Forecasting bankruptcy more accurately: A simple hazard model," *Journal of Business*, 74, 101–124.

Sloan, R. G. (1996): "Do stock prices fully reflect information in accruals and cash flows about future earnings?," *The Accounting Review*, 71, 289–315.

Stock, J., and M. Watson (2003): "Forecating output and inflation: The role of asset prices," *Journal of Economic Literature*, 41, 788–829.

——— (2004): "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23, 405–430.

Timmermann, A. (2006): "Forecast combination," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, pp. 135–196. North–Holland, Amsterdam.

Titman, S., K. C. J. Wei, and F. Xie (2004): "Capital investments and stock returns," *Journal of Financial and Quantitative Analysis*, 39, 677–700.

Vasicek, O. A. (1973): "A note on using cross-sectional information in Bayesian estimation of security betas," *Journal of Finance*, 28, 1233–1239.

Vassalou, M., and Y. Xing (2004): "Default risk in equity returns," *Journal of Finance*, 59, 831–868.

Wallis, K. (2005): "Combining density and interval forecasts: A modest proposal," *Oxford Bulletin of Economics and Statistics*, 67, 983–994.

Welch, I., and A. Goyal (2008): "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 21, 1455–1508.

Xie, H. (2001): "The mispricing of abnormal accruals," *The Accounting Review*, 76, 357–373.

Zellner, A. (1971): *An introduction to Bayesian inference in econometrics*. Wiley, Hoboken, NJ.

Zhang, L. (2005): "The value premium," *Journal of Finance*, 60, 67–103.

Zmijewski, M. E. (1984): "Methodological issues related to the estimation of financial distress prediction models," *Journal of Accounting Research*, 22, 59–82.