ESSAYS IN EMPIRICAL ASSET PRICING

A Dissertation

Presented to

The Faculty of the C.T. Bauer College of Business

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

by

Yang Luo

November, 2018

Acknowledgments

I want to express my sincere gratitude to my advisor, James (Seung Min) Yae, and the co-chair, Raul Susmel, for their invaluable time and guidance through the ups and downs of the journey. James has been extremely patient while listening to my random thoughts about research and life and continuously helping me through the whole process. Raul always shares his passion and knowledge. His insight and enthusiasm always inspire me. I am deeply indebted to my advisors, James and Raul. Without their support, I would not be able to complete the dissertation.

I am grateful to Hitesh Doshi for his encouragement and invaluable comments. He taught me how to conduct research. Kris Jacobs encouraged me to keep an eye on the broader view in everything. I have greatly benefited from his ability to focus on the big picture especially for research. I thank my dissertation committee member, Tong Lu, for his valuable time and advice.

I am also thankful to Thomas George, Praveen Kumar, Stuart Turnbull, Nisan Langberg, Vijay Yerramilli, Kevin Roshak, Sang Byung Seo, Zack Liu, Paola Pederzoli and Rajkamal Vasu. They taught me fundamental knowledge of finance and empirical skills and provided supportive advice whenever I approach them. Mary Gould, Anesia Lamont, and Ieshia Deal helped me deal with the bureaucracy. Thanks to Shunlin Liang, Xin Gao, Guanglian Hu, Rui Liu, David Buchuk, Steven Chen, Sangwon Lee, Yu Li, Eshwar Venugopal, and other Ph.D. students for all the spirited discussions, encouragement, and companion.

I am eternally grateful to my parents for their understanding and support which has afforded me the freedom to live and work far away from them. I thank my son, Eli, for bringing me happiness and joy. Finally, I cannot thank my wife, Fei, enough for being there for me every day with love and support. She is a lifelong partner who makes my life worthwhile.

ESSAYS IN EMPIRICAL ASSET PRICING

Abstract

Yang Luo

November, 2018

This dissertation consists of three essays in empirical asset pricing. In the first essay, I propose a machine learning based approach to monitor the relative forecasting performance between two forecasts and select the conditionally better forecast. When I apply this approach to the combination forecast and the historical average benchmark forecast, the resulting new return predictor leads to statistically and economically significant out-of-sample gains consistently over time. Such improvements come from expecting a conditionally poor performance of the combination forecast and switching to the historical average benchmark. This approach also works well for other return forecasts using individual economic predictors and can be applied to combine individual forecasts more efficiently. Interestingly, the weight on the combination forecast produced by the machine learning based monitoring approach is high during NBER recessions and periods with high macro uncertainty, which captures the well-known fact that return predictability is concentrated in bad times.

In the second essay, we compute implied dividend yields using equity options and show that they are negatively related to the subsequent stock returns. This finding is in contrast with the theory and evidence at the market level where dividend yield is positively related to the future market return. The panel data analysis reveals that the normal relation between the dividend yield and individual stock returns recovers in longer horizon. I further find that the mixed evidence regarding option implied skewness and stock returns could also be reconciled with varying forecast horizon. The opposite to theory relation between option implied measures and stock returns is stronger when analyst forecast dispersion is at a higher level.

In the third essay, we investigate the relationship between systematic risk and credit default swap (CDS) returns and discovers that cross-sectional dispersion in future CDS returns can be rationalized by differences in firm's sensitivities to the market return. Further analysis shows that investors in the CDS market demand higher compensation to provide default protection to firms with higher sensitivities to downside market movements. The reward for bearing downside risk is not simply the compensation for systematic risk nor is it explained by other firm characteristics. The relation between downside risk and CDS returns is stronger for longer maturity CDS contracts.

Table of Contents

List of Tables

List of Figures

Chapter 1

Monitoring Forecasts with Machine Learning

1.1 Introduction

Stock return forecasts are of great interest to both academics and practitioners in finance as they affect areas ranging from asset pricing to portfolio allocation and risk management. The quest for accurate and reliable return forecasts led to numerous economic variables discovered to have significant in-sample predictability (Campbell and Shiller, 1988; Fama and French, 1988, 1989; Ferson and Harvey, 1993; Lettau and Ludvigson, 2001; Pontiff and Schall, 1998). However, as noted by Goyal and Welch (2008), these potential predictors are unable to deliver consistent out-of-sample forecasts of the market return when compared to a simple benchmark: the historical average forecast. The relative forecasting performance of economic predictors are found to be unstable and changes over time. Rapach, Strauss, and Zhou (2010) find that combining individual predictive regression models can improve out-of-sample performance in both statistical and economic terms. However, the out-of-sample performance of this approach still deteriorates from 1980 onward. I find that the deteriorated performance of combining strategy is because several economics variables could not reliably predict market return after around 1980. Thus, model uncertainty and instability not only seriously impair the forecasting ability of individual predictive regression models, but can also spill over and negatively affect the combination forecast. Removing poor forecasts from the combination pool could potentially boost out-of-sample performance, but this approach requires econometricians to identify in real time which forecast will perform worse (or better) in future.

Motivated by these empirical observations, I answer the following two questions: (1) can we monitor the forecasting performance ex-ante; (2) if so, can we improve forecast accuracy using the information gathered from monitoring. I propose a machine learning based approach to monitor relative performance between a candidate forecast and the historical average benchmark forecast (theoretically they could be any two forecasts but the current setting uses the historical average as a benchmark as this is the most commonly used baseline in evaluating any market return forecast). This approach accommodates model instability and utilizes a large number of potential conditioning variables to monitor forecasting model's predictive power. I then investigate whether the information provided by machine learning based monitoring can help us make more accurate forecasts in a pseudo-real-time exercise. I find that machine learning based monitoring approaches can provide consistent out-of-sample gains to both combination forecast and individual forecasts statistically and economically.

The approach of monitoring forecasting performance via machine learning has several essential features. First, this method naturally takes into account the structural instability of forecasting models. Paye and Timmermann (2006), Goyal and Welch (2008), Pettenuzzo and Timmermann (2011), Henkel et al. (2011) document substantial variation in predictive power of regression models when forecasting stock returns. Numerous factors including business cycles, macroeconomic uncertainty, policy shocks, and advances in information technology combine to produce a highly complex and constantly evolving data-generating process for market return. To capture the time-varying predictability of forecasting models, monitoring the performance of these models in "real time" is crucial. If evidence indicates that the forecasting performance is deteriorating or extremely poor during certain periods, then investors should adjust the weights put on this forecast or just switch to the alternative forecast. The intuition is similar to the case when an asset manager adjusts the portfolio weight by evaluating portfolio risk constantly.

Second, I apply a monitoring approach to facilitate forward-looking optimal selection of candidate forecasts. The standard literature of forecasting with multiple models usually use past forecast performance to gauge future forecast accuracy. This idea of choosing optimal weight based on past performance, dating back to Bates and Granger (1969), works in theory but not in practice especially when compared to simply putting equal weights on models (Granger (1989), Stock and Watson (2004), Timmermann (2006)). A key contributor to the poor performance of optimal forecast combination is that these methods do not take into account expected future forecasting performance. The models that perform well in the past (unconditionally) may not necessarily continue to work well in the future (conditionally). Timmermann and Zhu (2017) and Gibbs and Vasnev (2017) show that the forward-looking approach to model selection or model combination using conditionally optimal weights leads to better performance. Provided that a model's superior performance can be ex-ante identified with conditional information, this model could be the preferred one for certain time periods even if it performs poor on average. I implement this idea of identifying conditional forecasting performance with a broader set of conditioning variables compared with the empirical study conducted in Timmermann and Zhu (2017) and Gibbs and Vasnev (2017).

Third, the use of machine learning methods in the process of monitoring accommodates a far more broad list of potential conditioning variables (monitoring instruments) and richer specifications of the functional form. The flexibility comes from the high-dimensional nature of machine learning methods and allows me to monitor the forecasting performance in real time which otherwise is hard to implement by using traditional econometric techniques. It is thus possible to approximate the unknown and complex data-generating process underlying market return and forecasting performance of candidate models. There is a concern of a higher propensity with overfitting using a high-dimensional method like machine learning. I address this by implementing ensemble based machine learning algorithms with regularization to classify whether a forecasting prediction model will perform better than the historical average benchmark forecast in the next period. I use sequential ensemble methods (boosting trees) and parallel ensemble methods (random forest and extremely randomized trees) to convert small, weak learners into a strong learner. Besides, time-series cross-validation is applied in real time to tune hyperparameters of each machine learning model to further control in-sample overfitting. Above considerations describe advantages of machine learning methods to learn complex dynamics of relative forecasting performance and emphasize considerations to guard against overfitting the data.

My empirical findings suggest that the relative forecasting performance of the combination forecast can be monitored with machine learning methods by conditioning on the time-series features of past relative performance. Hybrid forecasts based on selecting the conditionally better one between the combination forecast and the historical average benchmark forecast provide gains in predicting market returns. The gains obtained through monitoring are statistically and economically significant, especially during recent decades when model instability is a severe concern for both individual predictive models and combination forecast. Besides, the monitoring approach can also improve individual forecasts and be used to generate new combination forecasts that consistently outperform the historical average benchmark. I further show that the monitoring approach has links that are related to the macro economy. The weights on the combination forecast provided by the monitoring approach can be seen as the level of predictability of the market return, and they align well with NBER recession indicators and macro uncertainty measures which is consistent with previous literature.

The contribution of my paper to the literature is twofold. First, I document the evidence that model instability, a common concern in individual prediction models, also negatively affects the combination forecast, especially in recent decades. Second, I propose a novel machine learning based monitoring approach, evaluate its usefulness for model selection and model combination, and examine the out-of-sample gains in forecasting the market return.

My work extends the empirical literature on stock return prediction. Goyal and Welch (2008) note that it remains difficult to find models that can improve on even the most naive benchmarks in out-of-sample. Rapach, Strauss, and Zhou (2010) find that combination method could reduce model uncertainty. Campbell and Thompson (2008) show that restricting the return forecasts to be non-negative and forcing the regression coefficients to have the theoretically expected signs lead to clear gains in out-of-sample predictability. Along the same lines, Pettenuzzo, Timmermann, and Valkanov (2014) propose a Bayesian approach to impose non-negative equity premia and bounds on the conditional Sharpe ratio of univariate predictive regressions and find that their approach leads to more accurate forecasts. I build on the work of Rapach, Strauss, and Zhou (2010), document the evidence that forecast combination methods also undergo model instability problem, and propose a novel approach using machine learning methods to monitor the relative forecasting performance between the combination forecast and the historical average benchmark forecast.

Machine learning methods have appeared much more in the asset pricing literature recently. Rapach et al. (2013) apply LASSO to predict global equity market returns using lagged returns of all countries. Giglio and Xiu (2016) and Kelly et al. (2017) use dimension reduction methods to estimate and test factor pricing models. Moritz and Zimmermann (2016) apply tree-based models to portfolio sorting. Kozak et al. (2017) and Freyberger et al. (2017) use shrinkage and selection methods to, respectively, approximate a stochastic discount factor and a nonlinear function for expected returns. The focus of my paper is to explore the usefulness of machine learning methods in monitoring the forecasting performance between two forecasts and study whether monitoring forecasts via machine learning could improve the predictability of the market return.

There are a few papers in the literature that estimate conditional forecasting performance for model selection or model averaging. Aiolfi and Timmermann (2006) exploit the persistence of the past forecast errors in an optimal way for constructing weights for model averaging. Timmermann and Zhu (2017), Gibbs and Vasnev (2017), and Granziera and Sekhposyan

(2018) consider optimal weighting strategy conditional on expected future performance of the models given their past performance. Their approach results in sizable improvements in the accuracy of industrial production and inflation forecasts. Kim and Swanson (2016), on the other hand, use a hybrid modeling strategy, where they use a threshold, controlling for the severity of the business cycle, to switch between a naive benchmark and sophisticated index driven models for forecasting. They find that this strategy delivers sizable improvements in the accuracy of the GDP growth forecasts. My paper focuses on forecasting the market return and propose a machine learning based monitoring approach to accommodate an extensive list of monitoring instruments to maximize the information one can extract from past forecasting performance.

The remainder of this chapter is organized as follows. Section 1.2 outlines the machine learning methodology I rely on to monitor the forecasting performance and how to select conditionally better forecast. Data descriptions are presented in Section 1.3. Section 1.4 presents the main results of monitoring on the combination forecast and individual forecasts. Section 1.5 examines the link with the real economy. Section 1.6 concludes.

1.2 Methodology

In this section, I introduce conditionaly predictive ability, the machine learning algorithms, how to identify conditionally better forecast from past forecasting performance, and finally the criteria I use to evaluate the out-of-sample forecasts.

1.2.1 Conditional Predictive Ability

Conventional forecast combination methods use past forecasting performance to identify the more precise forecast. It is usually assumed that past forecasting performance is a good indicator of future forecasting performance. However, it is not clear how to select the proper holdout window of historical forecasting performance to use as a guidance and more critically past forecasting performance may not be the best indicator for predicting future forecasting performance. It is possible that one forecast is expected to be more accurate than another forecast conditional on some state variables (monitoring instruments) taking certain values, while the ranking of the two forecasts' expected performance is reversed for other values of the state variables. The state variables could be the mean, the maximum, the trend of the relative forecasting error or other features that are not obviously straightforward. I introduce more formally how to idenity which forecast will be more accurate for the next period by learning from a large number of monitoring instruments with machine learning algorithms.

Pairwise comparisons of predictive accuracy are routinely carried out in macroeconomic and financial studies. In this study, I will focus on comparing two forecasts with one of them being the historical average benchmark forecast. This setting has a theoretical background as investors would assume no predictability of the market return if the historical average benchmark forecast dominates other candidate forecasts. If the market return is constant, then the historical average benchmark forecast should be the best estimate for next period stock return. On the other hand, if investors believe the market return is time-varying, then an alternative forecast should be better than the historical average in predicting stock return. The historical average of past market return is the most common used benchmark for evaluating out-of-sample predictability from any forecast.

Let $\hat{r}_{1,t+1|t}$ and $\hat{r}_{2,t+1|t}$ be two one-step-ahead forecasts of r_{t+1}: the return on a stock market index in excess of the risk-free interest rate. The second forecast $\hat{r}_{2,t+1|t}$ is the historical average forecast, $\hat{r}_{2,t+1|t} = \sum_{\tau=1}^{t} r_\tau$, serving as a benchmark forecasting model corresponding to a constant expected market return. Both forecasts are generated using information known at time t. I calculate the square forecast error

$$L(\hat{r}_{t+1|t}, r_{t+1}) = (r_{t+1} - \hat{r}_{t+1|t})^2. \tag{1.1}$$

Under square forecast error loss, the loss differential between the two forecasts, $\Delta L_{t+1} \equiv L(\hat{r}_{1,t+1|t}, r_{t+1}) - L(\hat{r}_{2,t+1|t}, r_{t+1})$, takes the form

$$\Delta L_{t+1} = e^2_{1,t+1} - e^2_{2,t+1}, \tag{1.2}$$

where $e_{k,t+1} = r_{t+1} - \hat{r}_{j,t+1|t}$ for $k = 1, 2$ are the individual forecast errors. Negative values of ΔL_{t+1} indicates that the first forecast produced a smaller square forecast error than the second forecast in period $t + 1$.

If the market return is constant, then $\mathbb{E}(\Delta L_t)$ should be positive for any t, meaning that the historical average forecast will be a better model for predicting the market return. On the other hand, if the market return is predictable by any forecast, then $\mathbb{E}(\Delta L_t)$ should be netative for any t. However, as reported from earlier literature, most models relative predictive power change signs over time. Both individual predictive models or combination methods undergo up and downs in performance when compared to the historical average benchmark forecast. Even if one forecast is worse on average than the historical average benchmark forecast, it might perform better in certain states of the world. This suggests that using conditioning variables (monitoring instruments) when evaluating competing forecasts' relative accuracy could help to identify the more precise forecast at each time.

Suppose the conditioning variables could be written as a feature vector x_t and label the sign of outcome ΔL_t as y_t where

$$y_t = \begin{cases} 0, & \text{if } \Delta L_t > 0 \\ 1, & \text{if } \Delta L_t <= 0 \end{cases} \tag{1.3}$$

Thus, identifying the more accurate forecast between $\hat{r}_{1,t+1|t}$ and $\hat{r}_{2,t+1|t}$ could be converted to a supervised classification problem with a set of t training examples of the form $\{(x_1, y_1), \ldots (x_t, y_t)\}$ where a learning algorithm seeks a function $g : X \to Y$.

1.2.2 Machine Learning Algorithms

After establishing the supervised classification learning problem, it is still needed to introduce the specific functions g that will be used.

I begin model description with the least complex method, the logistic regression model. Logistic regression model is a statistical model that is usually taken to apply to binary dependent variable. The log-odds of the probability of an event is modeled as a linear function of independent variables. This model imposes a simple regression specification and does not allow for interactions between predictors.

To account for interactions among predictors, one way that is popular in machine learning literature is to use decision trees. Unlike logistic regression model, decision trees are fully nonparametric and imply a logic that departs markedly from traditional logistic regression. Decision trees are like flow-chart-like structure, where it grows in a sequence of steps. At each step, a new branch sorts the data remained from the previous step into bins based on one of the predictor variables. This sequential branching slices the space of predictors into rectangular partitions, and approximates the unknown function g with the majority value of the binary outcome within each partition.

The advantages of a tree model include the following: it is invariant to monotonic transformations of predictors, it can accommodate interactions between predictors, and it can approximate severe nonlinearities in the functional form. However, the flexibility also limits its usefulness as decision trees are prone to overfit the training data and thus some sort of regularization should be applied. In my analysis, I consider three types of ensemble tree regularizers that combine forecasts from many small trees into a single forecast.

The first regularization method is random forest classifier. It is an ensemble method that combines forecasts from many small decision trees. This algorithm introduces extra randomness when growing trees. Instead of trying to fit the whole sample, each small decision tree will only use a sub-sample of the dataset. Averaging the results from small decision trees can improve the predictive accuracy and control over-fitting. This brings greater diversity, which trades a higher bias for a lower variance and yields an overall better model.

The second regularization method is a variant of random forest and is called extremely randomized trees classifier. Instead of using sub-sample of the dataset and searching for the best possible threshold for each feature when splitting a node, this algorithm uses the full sample and random thresholds for each feature. This trades more biaes for a lower variance and usually will be faster to train than regular random forest since finding the best possible threshold for each feature at every node is very time-consuming.

The last ensemble tree method I include is gradient boosting classifier. This algorithm works by sequentially adding a new decesion tree to an ensemble of previous trees with each new one trying to correct the forecasting errors from its predecessor. It fit the new predictor to the residual errors made by the previous predictor. Shallow trees on their own are "weak learners" with weak predictive power. The theory behind boosting suggests that many weak learners may, as an ensemble, comprise a single "strong learner" with greater stability than a single complex tree.

For all three machine learning algorithms, depth of the trees and number of small estimators will be the tuning parameters determined via cross validation using past sample. All of the above learning algorithms including logistic regression will result in a probability measure, labeled as p_{t+1} representing the probability of outcome of y_{t+1} in Equation (1.3) to be 1. I apply another layer of ensemble to average all three machine learning algorithms outputs and end up with the averaged probability measure from three machine learning algorithms. This extra layer of ensemble is implemented to reduce the power of any particular algorithm and reduce overfitting in a further step.

Given the probability measure p_{t+1}, I apply a model selection rule and select forecast 1, $\hat{r}_{1,t+1|t}$, if $p_{t+1} > 0.5$ and select forecast 2, $\hat{r}_{2,t+1|t}$, if $p_{t+1} < 0.5$.

1.2.3 Conditioning Variables

After describing the learning algorithms, I will now introduce the conditioning variables (monitoring instruments) used as explanatory variables in the logistic regression and machine learning algorithms.

It is natural to use historical information to compute the combination forecast such as the Bates and Granger (1969) optimal weighting scheme and the discounted mean squared forecast error (DMSFE) in Stock and Watson (2004) and Rapach, Strauss, and Zhou (2010). These methods assume that forecasts performed well in the recent past will continue to work well in future. Thus, the first set of conditioning variables are past k month relative square forecast errors between forecast 1 and forecast 2. DMSFE is the simplest case which averages the past k month forecast errors from a hold-out sample. Here, I list each of past k month relative square forecast error as one independent variable since the best indicator of future performance may not usually be the mean. Choosing the correct number of k and figuring out the best indicator for future forecasting performance ex-ante is not possible without knowing the actual data. Thus, it is purly an empirical question of how to learn from the past forecasting performance. With the machine leaning algorithms, it is possible to learn from a wide range of lag forecasting performance and decide which of these conditioning variables are good indicators for future performance.

However, it may not be the lag forecasting performance that directly determines the future performance but rather some of the features of the past history of performance could be more useful. I follow Christ, Braun, Neuffer, and Kempa-Liehr (2018) and calculate hundreds of time-series features from the history of past performance using the TSFRESH package in Python. It automatically extracts more than 300 features from the time-series describing basic characteristics of the time-series such as the number of peaks, the average, maximal value, autocorrelation, and linear trend. The combination of machine learning algorithms and the large number of time-series features will maximize the information one can extract from past forecasting performance and help to identify the more precise forecast.

1.2.4 Forecast Evaluation

I use the out-of-sample R^2 statistic, suggested by Campbell and Thompson (2008) to evaluate the forecasting performance of $\hat{r}_i$, where $\hat{r}_i$ is either an individual forecast based on univariate predictive regression model, a combination forecast or a monitored forecast. The out-of-sample R^2 is given by

$$R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{i,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}. \tag{1.4}$$

where $\hat{r}_{HA}$ is the historical average benchmark forecast ($\hat{r}_{HA,t+1|t} = \sum_{\tau=1}^{t} r_\tau$). The out-of-sample R^2 statistic measures the reduction in mean square prediction error for a forecast relative to the historical average benchmark forecast. When $R^2_{OS} > 0$, the forecast $\hat{r}_i$ outperforms the historical average benchmark forecast according to the mean square prediction error metric.

Even if there is evidence that R^2_{OS} is greater than zero, its values are typically small for predicting the market return. This raises the issue of economic significance. Campbell and Thompson (2008) argue that even a small out-of-sample R^2 , such as 0.5% for the monthly data, can signal economically meaningful degree of return predictability in terms of increased annual portfolio returns for a mean-variance investor. This provides a simple assessment of forecastability in practice.

A limitation to the out-of-sample R^2 measure is that it does not explicitly account for the risk taken by an investor during the out-of-sample period. To address this, I also measure the economic value of market return forecasts for a risk-averse investor. Following Campbell and Thompson (2008) and Ferreira and Santa-Clara (2011), among others, I compute the certainty equivalent return (CER) for an investor with mean-variance preferences who monthly allocates across equities and risk-free bills using various market return forecasts.

At the end of month t, the investor optimally allocates the following share of the portfolio to equities during month t+1:

$$w_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right), \tag{1.5}$$

where $\hat{r}_{t+1}$ is a forecast of the market return and $\hat{\sigma}^2_{t+1}$ is a forecast of its variance. The share $1 - w_t$ is allocated to risk-free bills. The portfolio return of month $t+1$ is:

$$R_{p,t+1} = w_t r_{t+1} + R_{f,t+1}. \tag{1.6}$$

I assume that investor estimates the variance of the market return from five-year moving window of past monthly returns, set w_t to lie between 0 and 1.5, assume the relative risk coefficient to be 5 following Campbell and Thompson (2008).

The CER of the portfolio is

$$CER = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p, \tag{1.7}$$

where $\hat{\mu}_p$ is the mean of the investor's portfolio return over the forecast evaluation period and $\hat{\sigma}^2_p$ is the variance of portfolio return. The CER can be interpreted as the risk-free rate of return that an investor is willing to accept instead of choosing the given risky portfolio. The CER gain is the difference between the CER for the investor who uses a candidate forecast of the market return and the CER for an investor who uses the historical average benchmark forecast. I multiply the difference by 1200 so that it represents the annual percentage portfolio management fee that an investor would be willing to accept to have access to the predictive model instead of the historical average benchmark forecast.

1.3 Data

The monthly data are downloaded from Amit Goyal's website where detailed descriptions of the data are provided. This analysis include 14 predictor variables originally included in

Goyal and Welch (2008) and the sample period starts from January 1927 to December 2017. Below, I provide a list of the predictors I relied on to form the mean combination forecast:

1. Dividend-price ratio (dp): log of a 12-month moving sum of dividends paid on the S&P 500 Index minus the log of stock prices (S&P 500 Index)

2. Dividend yield (dy): log of a 12-month moving sum of dividends minus the log of lagged stock prices.

3. Earning-price ratio (ep): log of a 12-month moving sum of earnings on the S&P 500 Index minus the log of stock prices.

4. Dividend–payout ratio (de): log of a 12-month moving sum of dividends minus the log of a 12-month moving sum of earnings.

5. Equity risk premium volatility (rvol): based on a 12-month moving standard deviation estimator (Mele 2007)

6. Book-to-market ratio (bm): book-to-market value ratio for the Dow Jones Industrial Average

7. Net equity expansion (ntis): ratio of a 12-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of New York Stock Exchange (NYSE) stocks.

8. Treasury bill rate (tbl): interest rate on a three-month Treasury bill (secondary market).

9. Long-term yield (lty): long-term government bond yield.

10. Long-term return (ltr): return on long-term government bonds.

11. Term spread (tms): long-term yield minus the Treasury bill rate.

12. Default yield spread (dfy): difference between Moody's BAA- and AAA-rated corporate bond yields.

13. Default return spread (dfr): long-term corporate bond return minus the long-term government bond return.

14. Inflation (infl): calculated from the CPI for all urban consumers.

Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends, and the Treasury bill rate is used to compute the equity premium. As in Rapach, Stratuss, and Zhou (2010), I use standard predictive regression model for the equity premium:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \epsilon_{t+1} \tag{1.8}$$

where r_{t+1} is the return on a stock market index in excess of the risk-free rate, $x_{i,t}$ is a variable listed above. As in Goyal and Welch (2008), I generate out-of-sample forecasts of the equity premium using an expanding estimation window. The first out-of-sample forecast starts at January 1947, consistent with the starting point of out-of-sample evaluation in Goyal and Welch (2008).

Then I calculate the equal-weighted combination forecast of r_{t+1} made at time t as the simple average of 14 individual forecasts estimated from Equation (1.8):

$$\hat{r}_{CF,t+1} = \frac{1}{14}\sum_{i=1}^{14} \hat{r}_{i,t+1}. \tag{1.9}$$

Relative square forecast error is calculated using the combination forecast $\hat{r}_{CF,t+1}$ and the historical average benchmark forecast $\bar{r}_{t+1}$ as in Equation (1.1) and (1.2). In the main analysis, I use past 60 month lag relative square forecast error (60 explanatory variables) and time-series features of past 60 month lag relative square forecast error (hundrends of explanatory variables) as two sets of conditioning variables to classify whether $\hat{r}_{CF,t+1}$ will have lower forecast error compared with $\bar{r}_{t+1}$. For either logistic regression and machine learning algorithms, I use 120 past sets of feature and outcome pair as training data to learn what features are important in determining whether the mean combination forecast is better than the historical average forecast.

1.4 Empirical Results

1.4.1 Model Instability of Combination Forecast

Goyal and Welch (2008) comes to a pessimistic conclusion about return predictability and attributes the lack of consistent out-of-sample predictability of the equity premium to model uncertainty and model instability of individual forecasts using economic predictors. Following the application of forecast combination in literatures predicting macro economic variables (Hendry and Clements (2004), Clements and Hendry (2006), and Timmermann (2006)), Rapach, Strauss, and Zhou (2010) reports that various versions of combinations of forecasts from individual predictive regression models generate consistent and significant out-of-sample gains relative to the historical average. The intuition is that combining across individual forecasts provides a convenient and straightforward way to reduce the instability risk associated with relying on any single model. The results reported in Rapach, Strauss, and Zhou (2010) support the hypothesis that by combining forecast, and usually putting equal weights on individual forecasts can significantly improve out-of-sample predictability.

I revisit this exercise with monthly data and extend the sample period to December, 2017, which is the longest data period used in the literature. In Figure 1.1, I report time-series plot of the difference of square forecast error between the historical average benchmark forecast and the combination forecast. The difference of square forecast error is computed as $e^2_{HA,t} - e^2_{CF,t}$, where $e_{HA,t}$ is the forecast error of the historical average benchmark forecast and $e_{CF,t}$ is the forecast error of the combination forecast. The combination forecast is the equal-weighted average of fourteen individual forecasts calculated following Goyal and Welch (2008). This plot provides a visual impression of the relative forecasting performance of combination forecast over time. When the line in Figure 1.1 is above the horizontal dotted line at $y = 0$, the combination forecast outperforms the historical average benchmark forecast, while the opposite holds when the line is below the horizontal dotted line. At each month, I compute and plot the rolling mean of difference in square forecast error from past 60 month.

The plot in Figure 1.1 have positive values for most of the time from 1950 to 1990, indicating that the combination forecast delivers resonable out-of-sample predictability compared with the historical average benchmark forecast. The plot is highly positive before 1960 and during 1970 to 1990, highlighting that forecast combination was an effective strategy for market return prediction. However, after 1995, there are two sharp downward sloping periods where combination forecast performs much worse than the historical average. The first decrease of performance is from 1995 to 2000 before the Dot-com bubble and the second from 2007 to 2010 which covers the financial crisis period. The difference of square forecast error changes from positive to negative around 1996 and around 2007 suggesting the combination forecast is not a good candidate for predicting market return since the historical average benchmark forecast have smaller square forecast errors.

Actually, the predictability of the market return has decreased gradually over time. In Figure 1.3, I plot out-of-sample R^2 with moving evaluation starting date from January 1947 to January 2012. The evaluation ending date is always December 2017. The out-of-sample R^2 is computed as the percentage reduction in mean square forecast error of the model of interest relative to the historical average benchmark forecast: $R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{1,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}$. where $\hat{r}_{1,t|t-1}$ is the forecast of interest and $\hat{r}_{HA,t|t-1}$ is the historical average benchmark forecast. The dotted line in Figure 1.3 depicts the time series trend of out-of-sample forecasting performance of the combination forecast. The figure shows that combination forecast has positive out-of-sample R^2 of around 0.50% when the evaluation is for the whole period from 1947 to 2017. When we gradually move the evaluation starting date month by month, the out-of-sample R^2 will decrease slowly, indicating the predictive power of the combination forecast has diminished over time. If the evaluation starting date is set around 1984, the mean combination forecast will have negative out-of-sample R^2 suggesting it is worse than the historical average benchmark forecast. The latest evaluation starting date is January 2014 since the later the evaluation starting date, the less oberservations will be available. Thus the line is natually more volatile near the end of the x axis. However, we can still observe the clear decreasing trend of market return predictability using combinaiton forecast.

It is well known that model uncertainty and model uncertainty are serious concerns for individual predictive regression models. Therefore, it is difficult to identify individual forecasts capable of generating reliable equity premium forecast over time. Forecast combination methods are able to alleviate the concern of model instability to some extent. As shown in Figure 1.1 and Figure 1.3, the combination forecast works well from 1950 to 1990 but performs much worse since 1990. It suggests that model instability not only negatively affect predictability of individual models but can also affect the combination forecast. Table 1.1 reports the out-of-sample R^2 with various evaluation starting date for 14 individual predictive models in Panel A and for several versions of combination forecasts in Panel B. All columns are for evaluation periods with moving starting date from January 1947 to January 2007 and with fixed evaluation ending date at December 2017. As shown in Panel B, the out-of-sample R^2 is decreasing from 1947 to 2017. In row 1 of Panel B in Table 1.1, it is reported that the combination forecast, as shown in Figure 1.3, has out-of-sample R^2 of around 0.50% for the whole sample period. When we look at the evaluation period starting from 1987, 1997 and 2007, the out-of-sample R^2 decreases to negative values (-0.09%, -0.10% and -0.24%). We observe similar trend in most individual forecasts in Panel A. For example, for earnings-price ratio (ep), the out-of-sample R^2 decreases from -1.50% to -3.59% when we transit from whole sample period to the most recent decade, and for net equity expansion (ntis), it changes from -0.52% to -4.54%. Similar decreasing pattern can be found for 8 out of 14 economic predictors. Combining individual forecasts may be able to address the problem of model instability when the concern is not too severe. However, when the individual forecasts perform much worse than the benchmark forecast (the historical average of market return), combination strategy is not able to deliver consistent gains.

In Panel B of Table 1.1, I also report out-of-sample R^2 for other combining methods other than equal-weighted mean. The row labeled as 'Median' represents using the median of individual forecasts and it has some improvement in predictability during 1987 to 2017 compared with equal-weighted combination forecast. It indicates that ignoring outliers in individual forecasts may be an effective way to tackle the problem of model instability which spills over from individual forecasts to combination forecast. However, median combination forecast gets slightly worse results for the whole evaluation period from 1947 to 2017. Besides, from 2007 to 2017, the out-of-sample R^2 is about 0.04% suggesting that the forecasting performance no better than that of the historical average benchmark forecast.

The following rows starting with 'DMSFE' is a combining method based on Stock and Watson (2004), where the combining weights formed at time t are functions of the historical forecasting performance of the individual forecasts over the holdout out-of-sample period. I consider holdout periods of 60, 24, 12 and 1 month and discount factor of 1 and 0.5. There are several interesting findings. First, setting discount rate to 0.5 seems to have improved the predictability for all evaluation periods compared with discount rate of 1. This suggests that we may need to put more weights on forecasts with better recent outperformance. However, the extreme case of using performance of last month as guidance for the next period weight (labeled as 'DMSFE 1 1.0') leads to much better results for the evaluation period from 1947 to 2017 (out-of-sample R^2 of 1.17%) and much worse result for evaluation period starting from 2007 to 2017 (out-of-sample R^2 of -1.26%). These contrasting results imply that it is hard to identify the best combination weights that work throughout the whole sample period based on ad hoc choice of holdout period and discount factor. It is possible to try more combinations of holdout period and discount factor to boost out-of-sample R^2 . However, any combination with superior results could be purely random and not meaningful for future reference as theory suggests no prior belief about which combination should work better than others.

The results presented above can be summarized as follows:

- Combination forecast may be effective when model instability in individual forecasts is not strong. When certain individual models perform much worse than the benchmark, equal-weighted mean of individual forecasts is not able to beat the historical average benchmark consistently.

- Various modifications of combination forecast show some potential of improvement of performance. However, it is hard to know the right form of combination ex-ante and the best combination method may not work well consistently throughtout the whole period.

1.4.2 Monitoring the Combination Forecast

Based on the previous results that the combination forecast does not outperform the historical average benchmark forecast consistently over time, there could be improvement in forecasting the market return if the conditionally better forecast could be identified and selected for each period. I next provide the results of monitoring the relative forecasting performance between the combination forecast and the historical average benchmark forecast.

The results are presented in Table 1.2. It report out-of-sample R^2 for moving evaluation starting date ranging from January 1947 to January 2007. All evaluation period ends at December 2017. Panel A reports baseline results of the combination forecast and a simple equal-weighted average of the combination forecast and the historical average benchmark forecast. The first row in Panel A of Table 1.2 reports the out-of-sample R^2 of the combination forecast. It shows that for the most recent several decades, the combination forecast does not perform better than the historical average benchmark forecast as the out-of-sample

R^2 is negative especially after 2007. It is natural to consider taking the average of the combination forecast and the historical average as a way to improve the forecasting performance. The second row of Panel A (labeled as '(CF + HA)/2') indicates that there are some small improvement from 1987 to 2017 but a large loss of predictability for earlier evaluation periods as the out-of-sample R^2 of the full evaluation period reduced from 0.50% to 0.29% with this simple adjustment.

Panel B of Table 1.2 reports the results of a dummy monitoring method by comparing discounted mean squared forecast error (DMSFE) with a holdout window of 60 months and discount factor of 0.5 and 1. It selects the forecast with better recent out-of-sample performance. DMSFE is computed for the two forecasts: the combination forecast and the historical average benchmark forecast. Both rows in Panel B report improvement over the combination forecast after 1967. The second row of Panel B using a discount factor of 0.5 puts more weight on recent performance and results in larger improvement compared with using a discount factor of 1. But for the full sample period from 1947 to 2017, both DMSFE method perform worse compared to the combination forecast with out-of-sample R^2 lower than 0.50%. Besides, there are numerous ways to choose holdout window and discount factor while no theory suggests an ideal combiantion that would be the best for any data and any condition. Furthermore, the mean of past square forecast error may not be the best indicator for future square forecast error given the time-varying feature of relative forecasting performance shown in Figure 1.1. The empirical evidence also suggests that the dummy monitoring using DMSFE is not improving the overall results.

Thus, I use a monitoring approach to identify the conditionally better forecast through using past 120 month observations as training sample. Every month, it is observed that one of the two forecasts will perform better. This will output the left-hand side binary response variable. Right-hand side conditioning variables will be the past 60 month lag relative square forecast error. If the mean of lag relative square forecast error is indeed the best indicator for comparing future performance, then all 60 right-hand variable (lag relative square forecast error) will naturally have equal weights. Thus, dummy monitoring using DMSFE is nested in the general monitoring approach.

Results of monitoring through using past 60 month lag relative square forecast error as conditioning variables are reported in Panel C of Table 1.2. I implement logistic regression and an ensemble of three machine learning algorithms to classify which forecast will perform better conditional on past relative performance. Both logistic regression and machine learning algorithms will output a probability measure p_t between 0 and 1 to indicate the probability that the combination forecast will have lower square forecast error for next period. Then I select the combination forecast (the historical average benchmark forecast) when p_t is larger than 0.5 (less than 0.5). The first row in Panel C present out-of-sample R^2 of monitoring with lag relative square forecast error through logistic regression. For all evaluation periods, monitoring through logistic regression results in worse performance compared to the combination forecast. It suggests logistic regression is not able to identify the conditionally better forecast through learning from lag relative square forecast error. On the other hand, as shown in the second row of Panel C, using the machine learning algorithms have higher out-of-sample R^2 compared with using the logistic regression. However, the performance of monitoring with machine learning algorithms is similar to the dummy monitoring approach 'DMSFE 60 0.5' reported in Panel B. This suggests that if we restrict ourselves to using the past 60 month lag relative forecast error as conditioning variables, discounted mean of the past performance with discount factor of 0.5 is one of the best indicators as machine learning algorithms are not able to beat this with different weights. Results in Panel C indicates that machine learning algorithms are better than logistic regression in identifying useful indicator for future performance, but the out-of-sample R^2 is still lower than the combination forecast itself without monitoring.

In Panel D of Table 1.2, I report the results of monitoring through using more than 300 time-series features of past 60 month lag relative square forecast error as conditioning variables. The time-series features include basic statistics like mean and standard deviation and also other features like autocorrelation coefficients that could potentially describe the past relative performance better. For example, the cyclical changes in business cycles and seasonal patterns in forecasting performance can not be detected by looking at the simple mean. First row of Panel D present out-of-sample R^2 of monitoring with logistic regression. The predicting performance of logistic regression based monitoring is worse than any other previous cases. With hundreds of possibly correlated features, this is expected as logistic regression is not able to handle the situation when the number of conditioning variables are larger than the number of observations.

The second row of Panel D report the out-of-sample R^2 of monitoring using an ensemble of three machine learning algorithms. Ensemble is a term used in machine learning literature to indicate averaging several machine learning algorithms' output. This is a common practice in machine learning classification to reduce over-fitting. Second row of Panel D reports that for all evaluation periods, the forecasting performance has increased compared with no monitoring, dummy monitoring with DMSFE, and monitoring with lag relative square forecast error. For example, for evaluation from 1977 to 2017, the out-of-sample R^2 of is 0.34%, higher than that of the mean combination forecast (0.14%) and other monitoring methods mentioned above. For the period from 1947 to 2017, the out-of-sample R^2 is 0.57% while it is 0.50% for the combination forecast without monitoring. The difference between monitoring with machine learning algorithms and no monitoring gets bigger as the evaluation period moves closer to the most recent date. The combination forecast has negative out-of-sample R^2 starting from 1987 meaning that generally spreaking, the mean combination forecast is worse than the historical average benchmark forecast in predicting the market return. But when I use machine learning algorithms to monitor and select the conditionally better forecast among the combination forecast and the historical average benchmark forecast, all the out-of-sample R^2 are positive. The out-of-sample R^2 is 0.18% from 2007 to 2017, 0.32% from 1997 to 2017, and 0.35% from 1987 to 2017. It shows that there are certain periods from

1987 to 2017 when combiantion forecast delivers superior forecasts and the machine learning algorithms are able to capture these periods when combination forecast is usefull.

The results in Table 1.2 show that using machine leaning algorithms to monitor and select conditionally better forecast by learning from time-series features of past relative performance is able to improve the forecasting performance consistently over time. The improvements can not be achieved with other traditional monitoring method like comparing DMSFE or using logistic regression.

Figure 1.2 displays the time-series plot of the difference of square forecast error between the historical average benchmark forecast and the monitored combination forecast using machine learning algorithms and time-series features of past relative performance. The monitored combination forecast consistently outperforms the historical average benchmark forecast as the square forecast errors of the historical average are almost always larger than that of the combination forecast resulting in the line above the dotted horizontal line at $y = 0$. Compared with Figure 1.1 for the combination forecast without monitoring, the monitored combination forecast performs pretty well when the combination forecast is worse than the historical average benchmark. It is not surprising as the switching rule from monitoring would have no room to improve when the combination forecast is dominating the historical average benchmark forecast. Though there are also periods when monitoring reduce the forecasting performance from incorrectly choosing the better forecast, the overall outperformance of monitored combinaiton forecast over the historical average benchmark forecast makes the monitoring approach a good method to generate useful forecast consistently over time.

Figure 1.3 displays the comparison of out-of-sample R^2 of the combination forecast and the monitored combination forecast with moving evaluation starting point from 1947 to 2012 (all ending at 2017) as a complement of Table 1.2. The soid line represents the out-of-sample R^2 of the monitored combination forecast while the dotted line reprents the out-of-sample R^2 of the combination forecast without monitoring. It is clear from the graph that the monitored combination forecast performs better than combination forecast for almost all evaluation periods ending at 2017. Since there are less observations when the evaluation starts later, both lines start to have more up and downs after 1999 but it can still be observed that monitoring improves the performance.

I also measure the economic value of forecasting market return through monitoring forecasting performance for a risk-averse investor. Numbers in Table 1.3 denote the annualized certainty equivalent return (CER) gains based on different evaluation starting date. I evaluate the quality of the forecast using certainty equivalent return (CER) gains calculated as following. Each period, investors allocates the following share of the portfolio to equities: $w_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right)$, where $\hat{r}_{t+1}$ is a forecast of the equity premium and $\hat{\sigma}^2_{t+1}$ is a forecast of stock variance. The share $1 - w_t$ is allocated to risk-free bills. I assume investor estimates the variance of the equity premium from five-year moving window of past monthly returns, set w_t to lie between -0.5 and 1.5, assume the relative risk coefficient γ to be 5 following Campbell and Thompson (2008). The CER of the portfolio is $CER = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p$, where $\hat{\mu}_p$ is the mean of the portfolio returnand $\hat{\sigma}^2_p$ is the variance of the portfolio return. The CER can be seen as the risk-free rate of return that an investor is willing to accept instead of choosing the risky portfolio. The CER gain is the difference between the CER for the investor who uses a candidate forecast of the equity risk premium and the CER for an investor who uses the historical average forecast.

The first row in Panel A represents the CER gains of the combination forecast. It provides CER gains of 0.90% over the full evaluation period from 1947 to 2017 but the gains decrease as the evaluation starting dates moves from 1947 to 2007. Similarly as the out-of-sample R^2 displayed in Table 1.2, the average of combination forecast and historical average, dummy monitoring using DMSFE, and monitoring with lag relative square forecast errors all produce CER gains lower than the combination forecast without any monitoring for most of the evaluation periods.

In the second two row of Panel D, it displays the CER gains of monitoring through machine learning algorithms with time-series features of past 60 month relative square forecast error. It is shown that machine learning based monitoring increases the CER gains for all evaluation periods. For example, the CER gain is 1.05% from 1947 to 2017 while the combination forecast has CER gain of 0.90%. The improvement gets larger when we evaluate more recent samples.

Overall, these results provide statistical and economic evidence in support of the proposed monitoring approach with machine learning algorithms and using time-series features of relative forecasting performance as conditioning variables, particularly so for the recent decades.

1.4.3 Monitoring Individual Forecasts

Previous results show that monitoring the combination forecast could improve the forecasting performance by dynamically switching between the combinaiton forecast and the historical average benchmark forecast. Similarly, the monitoring approach could be applied to individual forecasts too as model instability also affects individual forecasts. If an individual forecast of market return does not dominate the historical average benchmark forecast and it is not dominated by the historical average benchmark forecast, identifing the conditionally better forecast in real time and switching to the better one should also increase the predictability of market return.

I report the results from monitoring individual forecasts in Table 1.4. The results are for the full evaluation period from 1947 to 2017. Because of the good performance in monitoring the combination forecast, the table only reports results utilizing an ensemble of three machine learning algorithms as monitoring method and time-series features of past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and each candidate individual forecast as conditioning variables. The first column of Table 1.4 lists the 14 individual economic predictors used for generating individual forecasts. The second column reports the out-of-sample R^2 of the individual forecasts without monitoring while the third column reports the out-of-sample R^2 after monitoring. For 13 out of 14 forecasts, the monitoring approach increase the out-of-sample R^2 . For example, the out-of-sample R^2 increases from -1.54% to 0.74% for bm and from 0.09% to 0.60% for tms. The average of improvement in out-of-sample R^2 is 0.57%. The magnitude of improvement across each individual forecast differs slightly as different forecasts have different forecasting performance compared to the historical average benchmark forecast. How much the monitoring approach could improve depends on how often each individual forecast outperforms the benchmark and whether the algorithm can correctly identify these periods. Last two columns of Table 1.4 reports the CER gains before and after monitoring. The increase in economic gains is observed for 11 out of 14 individual forecasts. The universal improvement in out-of-sample R^2 and CER gains indicates that the monitoring approach is indeed useful.

The previous table reports results from monitoring individual forecasts. It is a natural extension to consider if there are any benefits to combine the monitored individual forecasts. Table 1.5 displays the out-of-sample R^2 of this approach (monitoring first and combining later). Panel A of Table 1.5 shows the benchmark results from previous section including the out-of-sample R^2 of combination forecast and monitored combination forecast (combining first and monitoring later).

By monitoring each individual forecast, the algorithm will output a probability measure $p_{i,t}$ indicating the weight to put on the i-th individual forecast. The basic monitoring approach selects the individual forecast (the historical average forecast) when $p_{i,t}$ is larger than 0.5 (less than 0.5). First, consider the equal-weighted average of monitored individual forecasts which is reported in first row of Panel B in Table 1.5. It is observed that this approach of monitoring first and combining later has similar out-of-sample R^2 for all evaluation periods as the monitored combination forecast (combining first and monitoring later). Both approach increase the out-of-sample R^2 especially for later evaluation periods when model instability is a severe concern.

Instead, the cross-section of $p_{i,t}$ could be used as weights to apply for each individual forecast when generating a new combination forecast. The second row of Panel B in Table 1.5, labeled as 'Weighted Average', reports the results of using individual monitoring ouput as combination weights. The new combination forecast improves the out-of-sample R^2 for all evaluation periods compared with equal-weigthed combination of individual forecasts. For the most recent evaulation period from 2007 to 2017, the out-of-sample R^2 increases from -0.24% to 0.24% and for the full evaluation period from 1947 to 2017, it increases from 0.50% to 0.63%.

On the other hand, given the cross-section of $p_{i,t}$ at each time period, it is possible to consider the extreme case of selecting the individual forecast with the highest $p_{i,t}$. The last row of Panel B in Table 1.5, labeled as 'Weighted Selection', shows the out-of-sample R^2 from this extreme approach. During the evaluation period from 2007 to 2017, the out-of-sample R^2 reaches a magnitude of 3.64% indicating a significant improvement by selecting the conditionally best individual forecast from monitoring. Similar large improvement are observed from 1977 to 2017 though the out-of-sample R^2 is not much different as other approaches during the full evaluation period.

In summary, this section confirms the usefullness of the monitoring approach by showing universal improvement on individual forecasts and the benefits to consider new combination methods using the output of monitoring on individual forecasts. The improvements are observed for the full evaluation period and even more strong during past 30 years when model instability is a severe concern for individual forecasts.

1.5 Links to the Real Economy

1.5.1 Predictability and NBER Recessions

Mounting evidence shows that return predictability fluctuates over the business cycles. Particularly, various papers find that return predictability concentrates in bad times. Rapach, Strauss, and Zhou (2010), Henkel, Martin, and Nardari (2011), and Dangl and Halling (2012) find that macro variables or combination of macro variables' forecast have better predictive power in recessions.

The selection between the combination forecast and the historical average benchmark forecast boils down to the question of whether market return is predictable or not. If the market return is constant and unpredictable, the historical average forecast should be the best estimate for the next period. If the market return is time-varying and predictable, the combination forecast should capture the business-cycle fluctuations in the market return and thus be a better forecast. Fama and French (1989) and Cochrane (1999, 2007) argue that heightened risk aversion during economic downturns demands a higher risk premium, thereby generating market return predictability. Therefore, whether the combination forecast provides a good prediction of the market return is linked to business-cycles movements.

The monitoring approach output a probability measure p_t indicating the weight to put on the combination forecast. It can be seen as the level of return predictability. This output of probability measure should have some connection with business-cycles according to previous literature. Figure 1.4 plots the rolling mean of weights on the combination forecast from past 24 month, along with gray shaded vertical lines indicating NBER recessions. There are upward spikes of the weight on the combination forecast at or shortly after most NBER recessions. For example, for recent recession between December 2007 to June 2009 and between March 2001 to November 2001, there are distinct increasing patterns of the weights on the combination forecast. It shows that the monitoring approach indeed correctly identifies the increase of predictability in market return during economic recessions. However, there are other periods when return predictability is high and not captured by the NBER recessions. For example, the two recent spikes after 2009 is not associated with any recession periods. Next, I explore other possible origins of time-varying return predictability.

1.5.2 Predictability and Macro Uncertainty

Cujean and Hasler (2017) provide theoretical model to explain why stock return predictability concentrates in bad times. They argue that the key feature is that investors assess uncertainty differently as they use different forecasting models. As economic conditions deteriorate, uncertainty rises and investors' opinions polarize. Disagreement thus spikes in bad times, causing returns to react to past economic news. Thus whether return is predictable by economic predictors or combinaiton forecast of individual economic predictors is related to the level of uncertainty in investors.

Figure 1.5 plots the estimated weight on the combination forecast and macro uncertainty. Rolling mean of the estimated weights from past 24 month are reported in the figure. Macro uncertainty measure is calculated following Jurado, Ludvigson, and Ng (2015) which is available from 1960 to 2017. Figure 1.5 shows that the weight on the combination forecast and macro uncertainty comoves throught the sample. Whenever the macro uncertainty increases, the estimated weight on the combination forecast also increases. The results confirm the theoretical model in Cujean and Hasler (2017) indicating higher return predictability from economic predictors during high uncertainty periods (bad times).

To gauge the effect of recession and uncertainty on the return predictability, Table 1.6 reports the regression results. On the left hand side of the regression, I put the weight of each forecast from monitoring the relative forecasting performance between the candidate forecast and the historical average benchmark forecast. The explanatory variables include NBER recession indicator and macro uncertainty measure. The estimation is implemented for each forecast seperately. First row of Table 1.6 shows that about 10% of the estimated weight on the combination forecast is explained by recession and uncertainty. Both recession and uncertainty have positive coefficients confiring previous empirical findings and theoretical models. The coefficient on recession indicator has a t-statistic of 1.22 while the uncertainty measure has a t-statistic of 4.03. This suggests that macro uncertainty plays the essential role of driving the time-varying return predictability. The following rows of Table 1.6 reports the results for other individual forecasts. Unlike the case for combination forecast, the weight on other individual forecasts are usually less explained by the macro conditions with R^2 less than 5% except for dp and dy. The estimations for dp and dy also result in significantly positive coefficient for macro uncertainty. This similarity suggests that the forecasts generated from dividend related predictors follow similar pattern of the combination forecast. It highlights that the dividend related ratios are important measures driving the time-varying market return.

1.6 Conclusion

I conduct a comprehensive evaluation of the out-of-sample performance of the forecast combination methods in predicting the market return. I find that forecast combination methods fail to deliver consistent out-of-sample gains especially during the recent decades as model instability of individual models can be large enough to impair the effectiveness of combination forecasts. I then investigate whether the relative performance between the combination forecast and the historical average benchmark forecast can be monitored through tracking lag relative forecasting performance. I propose a novel approach to monitor the relative forecasting performance between the two forecasts using machine learning algorithms as the estimation method and using time-series features from the lag relative performance as conditioning variables. I document that selecting the conditionally better forecast based on monitoring improves the out-of-sample predictability of the combination forecast. The improvements are also observed for individual forecasts and several novel methods to combine individual forecasts using the output from the monitoring approach. My results suggest that considering conditional predictive ability and examining a large panel of monitoring instruments via machine learning methods could indeed be useful for model selection and model combination strategies. Further analysis shows that monitoring approach provides information with links to the macro economy. The results indicate that the market return is more predictable during economic downturns and during the periods when macro uncertainty is high, in agreement with previous literature.

Table 1.1 Out-of-sample R^2 of Traditional Forecasts

This table reports the forecasting performance of 14 individual predicting varialbes used in Goyal and Welch (2008), as well as several combination forecasts constructed from the 14 individual forecasts used in Rapach, Strauss, and Zhou (2010). The combination methods include the equal-weighted average (Mean), the median of individual forecasts (Median), combination using discounted mean square forecast error (DMSFE) with holdout window of 1, 12, 24 and 60 months and discount factor of 0.5 and 1. I evaluate the quality of the forecast using the out-of-sample R^2 defined as the percentage reduction in mean square forecast error of the forecast of interest relative to the historical average benchmark: $R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{i,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}$. where $\hat{r}_{i,t|t-1}$ is the forecast of interest and $\hat{r}_{HA,t|t-1}$ is the historical average benchmark forecast of the market return. I multiply the R^2_{OS} by 100 to denote percentage values. A positive R^2_{OS} indicates that the forecast of interest generates more accurate predictions than the historical average benchmark. I report R^2_{OS} for different evaluation starting date ranging from January 1947 to January 2007. All evaluation period ends at December 2017.

	Evaluation starts at (ends at December 2017)						
	1947	1957	1967	1977	1987	1997	2007
Panel A: Individual forecasts							
dp	-0.12	-0.33	-0.29	-0.92	-1.39	-1.04	-0.80
dy	-0.45	-0.72	-0.56	-1.41	-1.99	-1.58	-0.87
ep	-1.50	-1.93	-1.68	-2.13	-1.48	-1.73	-3.59
de	-1.44	-1.87	-1.57	-0.87	-0.56	-1.10	-1.70
rvol	-0.05	-0.19	-0.08	-0.01	-0.01	0.01	0.03
bm	-1.54	-1.96	-2.26	-2.86	-2.33	-1.37	-1.08
ntis	-0.52	-0.64	-0.80	-0.97	-1.90	-2.84	-4.54
tbl	0.08	0.12	-0.00	-0.62	-0.14	0.15	0.46
lty	-0.68	-1.00	-0.85	-0.80	0.03	0.31	0.56
ltr	-0.78	-0.11	-0.01	0.02	0.15	-0.30	0.39
tms	0.09	0.08	0.14	-0.25	-0.96	-0.68	-0.40
dfy	-0.18	-0.06	-0.04	-0.10	-0.21	-0.28	-0.44
dfr	-0.24	0.00	0.03	0.23	-0.09	-0.10	0.23
infl	-0.06	-0.04	-0.07	-0.33	-0.33	-0.59	-0.90
Panel B: Traditional combination forecasts							
Mean	0.50	0.37	0.36	0.14	-0.09	-0.10	-0.24
Median	0.40	0.37	0.38	0.21	0.09	0.08	0.04
DMSFE 60 1.0	0.50	0.37	0.37	0.15	-0.08	-0.09	-0.24
DMSFE 24 1.0	0.49	0.36	0.37	0.14	-0.04	-0.03	-0.19
DMSFE 12 1.0	0.56	0.43	0.42	0.18	-0.03	-0.00	-0.14
DMSFE 1 1.0	1.17	1.09	1.18	1.13	-0.31	-0.34	-1.26
DMSFE 60 0.5	0.57	0.45	0.43	0.14	-0.08	-0.01	-0.08
DMSFE 24 0.5	0.57	0.45	0.43	0.14	-0.08	-0.01	-0.08
DMSFE 12 0.5	0.57	0.45	0.43	0.14	-0.08	-0.01	-0.08

Table 1.2 Out-of-sample R^2 of Monitored Combination Forecast

This table reports the forecasting performance of the (equal-weighted) combination forecast, as well as several variations of monitored combination forecast by monitoring the relative forecasting performance beteen the combination forecast and the historical average benchmark forecast. Panel A includes the combination forecast and the simple average of the combination forecast and the historical average benchmark forecast (labeled as '(CF+HA)/2'). Panel B selects one of the two forecasts by comparing past discounted mean squared forecast error (DMSFE) with a holdout window of 60 months and discount factor of 0.5 and 1. Panel C and D report results from selecting the conditionally better of the two forecasts using logistic regression or ensemble of machine learning algorithms as monitoring method. Both logistic regression or ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the combination forecast. I use past 120 month outcome as training sample and use time-series cross-validation to select hyper-parameters if necessary. The approach selects the combination forecast (the historical average forecast) when p_t is larger than 0.5 (less than 0.5). Panel C reports the results from monitoring through past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and the combination forecast as conditioning variables. Panel D utilizes the time-series features of the past 60 month lagged difference of squared forecast error (DSFE) as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. I evaluate the quality of each forecast using the out-of-sample R^2 defined as the percentage reduction in mean square forecast error of the forecast of interest relative to the historical average benchmark: $R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{i,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}$. where $\hat{r}_{i,t|t-1}$ is the forecast of interest and $\hat{r}_{HA,t|t-1}$ is the historical average benchmark forecast of the market return. I multiply the R^2_{OS} by 100 to denote percentage values. A positive R^2_{OS} indicates that the forecast of interest generates more accurate predictions than the historical average benchmark. I report R^2_{OS} for different evaluation starting date ranging from January 1947 to January 2007. All evaluation period ends at December 2017.

	Evaluation starts at (ends at December 2017)						
	1947	1957	1967	1977	1987	1997	2007
Panel A: Baseline							
Combination Forecast	0.50	0.37	0.36	0.14	-0.09	-0.10	-0.24
(CF + HA)/2	0.29	0.21	0.21	0.09	-0.02	-0.03	-0.12
Panel B: Select best by comparing past DMSFE							
DMSFE 60 1.0	0.40	0.30	0.39	0.24	0.04	0.01	-0.10
DMSFE 60 0.5	0.45	0.42	0.38	0.19	0.10	0.14	0.09
Panel C: Monitoring with past DSFE as conditioning variables							
Logistic Regression	0.37	0.23	0.25	0.15	-0.08	-0.30	-0.60
Machine Learning	0.46	0.33	0.23	0.12	0.09	0.14	0.04
Panel D: Monitoring with time-series features of past DSFE							
Logistic Regression	0.21	0.11	0.06	-0.05	-0.36	-0.48	-0.46
Machine Learning	0.57	0.55	0.52	0.34	0.35	0.32	0.18

Table 1.3 Certainty Equivalent Return Gains of Monitored Combination Forecast

This table reports the certainty equivalent return gains of the equal-weighted combination forecast as well as several variations of the monitored combination forecast by monitoring the relative forecasting performance beteen combination forecast and historical average forecast. Panel A includes the combination forecast and the simple average of the combination forecast and the historical average forecast (labeled as (CF+HA)/2). Panel B selects the better of the two forecasts by comparing discounted mean squared forecast error (DMSFE) with a holdout window of 60 months and discount factor of 0.5 and 1. Panel C and D report results from selecting the conditionally better of the two forecasts using logistic regression or ensemble of ensemble of machine learning algorithms as monitoring method. Both logistic regression or ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the combination forecast. The approach selects the combination forecast (the historical average forecast) when p_t is larger than 0.5 (less than 0.5). Panel C reports the results from monitoring through past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and the combination forecast as conditioning variables. Panel D utilizes the time-series features of the past 60 month lagged difference of squared forecast error (DSFE) as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. Each period, investors allocates the following share of the portfolio to equities: $w_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right)$, where $\hat{r}_{t+1}$ is a forecast of the equity premium and $\hat{\sigma}^2_{t+1}$ is a forecast of stock variance. The share $1 - w_t$ is allocated to risk-free bills. I assume investor estimates the variance of the equity premium from five-year moving window of past monthly returns, set w_t to lie between -0.5 and 1.5, assume the relative risk coefficient γ to be 5 following Campbell and Thompson (2008). The CER of the portfolio is $CER = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p$, where $\hat{\mu}_p$ is the mean of the portfolio returnand $\hat{\sigma}^2_p$ is the variance of the portfolio return. CER can be seen as the risk-free rate of return that an investor is willing to accept instead of choosing the risky portfolio. CER gain is the difference between the CER for the investor who uses a candidate forecast of the equity risk premium and the CER for an investor who uses the historical average forecast. I multiply the difference by 1200 so that it represents the annual percentage portfolio management fee that an investor would be willing to accept to have access to the predictive model instead of the historical average forecast. I report CER gains for different evaluation starting date ranging from January 1947 to January 2007. All evaluation period ends at December 2017.

	Evaluation starts at (ends at December 2017)						
	1947	1957	1967	1977	1987	1997	2007
Panel A: Baseline							
Combination Forecast	0.90	0.77	0.81	0.33	0.10	0.36	0.40
(CF + HA)/2	0.51	0.44	0.46	0.21	0.10	0.23	0.18
Panel B: Select best by comparing past DMSFE							
DMSFE 60 1.0	0.69	0.59	0.76	0.45	0.26	0.30	0.54
DMSFE 60 0.5	0.91	0.94	0.90	0.51	0.48	0.63	0.67
Panel C: Monitoring with past DSFE as conditioning variables							
Logistic Regression	0.54	0.35	0.44	0.30	0.06	-0.33	-0.61
Machine Learning	0.85	0.71	0.59	0.35	0.38	0.45	0.82
Panel D: Monitoring with time-series features of past DSFE							
Logistic Regression	0.46	0.36	0.30	0.13	-0.17	0.01	0.14
Machine Learning	1.05	1.08	1.13	0.81	0.92	1.22	0.84

Table 1.4 Improvement From Monitoring Individual Forecasts

This table reports the forecasting performance of 14 individual forecasts examined in Goyal and Welch (2008) with or without monitoring. The monitoring approach uses ensemble of machine learning algorithms to select the conditionally better forecast between each individual forecast and the historical average forecast. Time-series features of past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and each candidate individual forecast are used as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. The ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the individual forecast. The approach selects the individual forecast (the historical average forecast) when p_t is larger than 0.5 (less than 0.5). The second and the third column report the out-of-sample R^2 defined as: $R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{i,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}$. where $\hat{r}_{i,t|t-1}$ is the forecast of interest and $\hat{r}_{HA,t|t-1}$ is the historical average benchmark forecast of the market return. I multiply the R^2_{OS} by 100 to denote percentage values. A positive R^2_{OS} indicates that the forecast of interest generates more accurate predictions than the historical average benchmark. The last two columns report certainty equivalent return (CER) gains calculated as following. Each period, investors allocates the following share of the portfolio to equities: $w_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right)$, where $\hat{r}_{t+1}$ is a forecast of the equity premium and $\hat{\sigma}^2_{t+1}$ is a forecast of stock variance. The share $1 - w_t$ is allocated to risk-free bills. I assume investor estimates the variance of the equity premium from five-year moving window of past monthly returns, set w_t to lie between -0.5 and 1.5, assume the relative risk coefficient to be 5 following Campbell and Thompson (2008). The CER of the portfolio is $CER = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p$, where $\hat{\mu}_p$ is the mean of the portfolio returnand $\hat{\sigma}^2_p$ is the variance of the portfolio return. The CER can be seen as the risk-free rate of return that an investor is willing to accept instead of choosing the risky portfolio. The CER gain is the difference between the CER for the investor who uses a candidate forecast of the equity risk premium and the CER for an investor who uses the historical average forecast. I multiply the difference by 1200 so that it represents the annual percentage portfolio management fee that an investor would be willing to accept to have access to the predictive model instead of the historical average forecast. Both out-of-sample R^2 and CER gains are evaluated from January 1947 to December 2017.

	OOS R^2		CER Gains	
	Individual	Monitored	Individual	Monitored
dp	-0.12	0.00	-0.05	0.51
dy	-0.45	0.05	-0.13	0.49
ep	-1.50	0.65	-0.09	2.12
de	-1.44	-1.19	-0.75	-0.60
rvol	-0.05	0.00	-0.11	-0.01
bm	-1.54	0.74	-1.59	1.42
ntis	-0.52	-0.14	-0.01	-0.22
tbl	0.08	0.17	0.57	0.56
lty	-0.68	-0.05	0.26	0.49
ltr	-0.78	-0.29	-0.86	-0.07
tms	0.09	0.60	0.00	0.92
dfy	-0.18	-0.09	-0.42	-0.14
dfr	-0.24	0.34	0.09	0.72
infl	-0.06	-0.16	-0.07	-0.19

Table 1.5 Out-of-sample $\mathbf{R^2}$ of Combining Individual Forecasts Through Monitoring

This table reports the forecasting performance of combining individual forecasts throught monitoring. Panel A includes baseline results of combination forecast (combining individual forecasts with equal weights) and monitored combination forecast. Panel B includes results from several different ways to combine individual forecasts from monitoring each one of them. Fist row of Panel B (labeled as 'Equal-weighted Average') reports results of monitoring each individual forecast, selecting the conditionally better one of the individual forecast and the historical average benchmark forecast, and then computing equal weighted average of monitored individual forecasts. Second row of Panel B (labeled as 'Weighted Average') reports results of combining individual forecasts using weights resulted from an ensemble of machine learning algorithms. The ensemble of machine learning algorithms monitor the relative forecasting performance between each individual forecast and historical average forecast and output a probability measure $p_{i,t}$ indicating the weight to put on the i-th individual forecast. Then individual forecasts are combined using normalized value of weights. Last row of Panel B (labeled as 'Weighted Selection') reports results of selecting the conditionally best individual forecast by comparing $p_{i,t}$ reported by machine learning algorithms. I evaluate the quality of each forecast using the out-of-sample R^2 defined as the percentage reduction in mean square forecast error of the forecast of interest relative to the historical average benchmark: $R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{i,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}$. where $\hat{r}_{i,t|t-1}$ is the forecast of interest and $\hat{r}_{HA,t|t-1}$ is the historical average benchmark forecast of the market return. I multiply the R^2_{OS} by 100 to denote percentage values. A positive R^2_{OS} indicates that the forecast of interest generates more accurate predictions than the historical average benchmark. I report R^2_{OS} for different evaluation starting date ranging from January 1947 to January 2007. All evaluation period ends at December 2017.

	Evaluation starts at (ends at December 2017)						
	1947	1957	1967	1977	1987	1997	2007
Panel A: Baseline: Equal-weighted combination forecast (monitored)							
Combination Forecast	0.50	0.37	0.36	0.14	-0.09	-0.10	-0.24
Monitored Combination Forecast	0.57	0.55	0.52	0.34	0.35	0.32	0.18
Panel B: Monitor each individual forecast first, then combine							
Equal-weighted Average	0.50	0.41	0.45	0.46	0.36	0.25	0.36
Weighted Average	0.63	0.48	0.50	0.42	0.18	0.18	0.24
Weighted Selection	0.54	0.32	0.30	0.95	1.02	1.74	3.64

Table 1.6 Weight on Forecasts and the Real Economy

This table reports estimation results of regressing weight put on each forecast on macro conditions. The ensemble of machine learning algorithms are used to monitor the relative forecasting performance between each forecast and historical average benchmark forecast and output a probability measure $p_{i,t}$ indicating the weight to put on the candidate forecast. Then the weight of each candidate forecast is regressed on NBER recession indicator and macro uncertainty measure. Macro uncertainty measure is calculated following Jurado, Ludvigson, and Ng (2015). Sample period is from August 1960 to December 2017 due to the data avalability of macro uncertainty. Point estimates of macro variables, t statistics, and R^2 of regression are reported.

Weight	Recession		Uncertainty		
	β	t-stat	β	t-stat	R^2
CF	0.04	1.22	0.50	4.03	0.08
dp	0.01	0.26	0.64	4.56	0.09
dy	0.01	0.40	0.79	5.44	0.12
ep	-0.00	-0.03	0.42	2.80	0.04
de	-0.09	-2.47	-0.13	-0.73	0.04
rvol	-0.01	-0.13	0.00	0.01	0.00
bm	0.02	0.47	0.41	3.41	0.04
ntis	-0.09	-2.32	0.02	0.13	0.03
tbl	0.09	1.99	-0.15	-0.76	0.02
lty	0.09	2.63	-0.08	-0.47	0.02
ltr	-0.01	-0.35	0.02	0.11	0.00
tms	0.01	0.19	0.16	0.94	0.01
dfy	-0.00	-0.07	0.39	2.28	0.03
dfr	0.07	2.46	-0.00	-0.01	0.01
infl	0.05	1.34	-0.27	-1.72	0.01

Figure 1.1 Difference of Square Forecast Error Between the Historical Average Benchmark Forecast and the Combination Forecast

This figure plots the difference of square forecast error betweeen the historical average benchmark forecast and the equal-weighted combination forecast. The combination forecast is based on fourteen individual forecasts calculated following Goyal and Welch (2008). Each forecast is estimated using expanding window from January 1927 to current period. The difference of square forecast error is computed as $e^2_{HA,t} - e^2_{CF,t}$, where $e_{HA,t}$ is the forecast error of the historical average benchmark forecast and $e_{CF,t}$ is the forecast error of the combination forecast. The differenec of square forecast error is then multiplied by 10^4 and the past 60 month rolling mean is computed and reported in this figure.

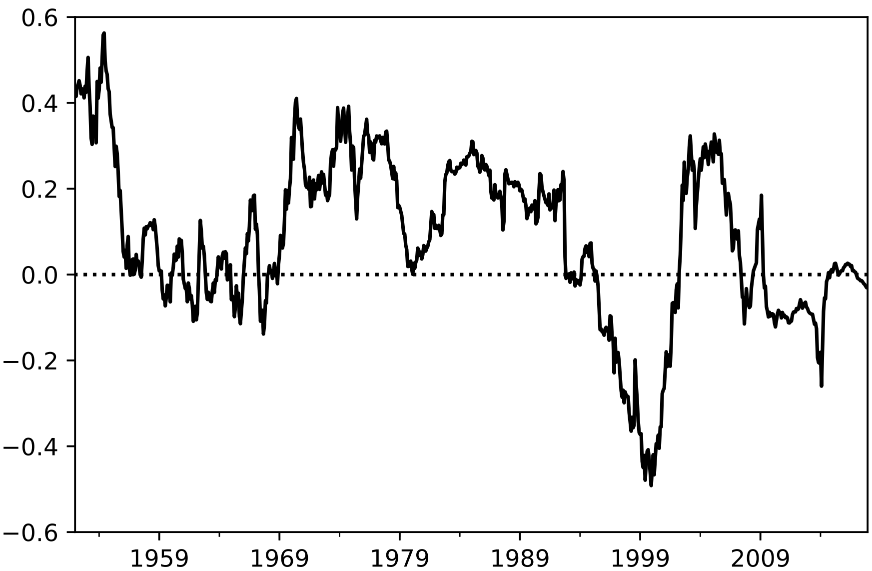


Figure 1.2 Difference of Square Forecast Error Between the Historical Average Benchmark Forecast and the Monitored Combination Forecast

This figure plots the difference of square forecast error betweeen the historical average benchmark forecast and the monitored combination forecast. The difference of square forecast error is computed as $e^2_{HA,t} - e^2_{MCF,t}$, where $e_{HA,t}$ is the forecast error of the historical average benchmark forecast and $e_{MCF,t}$ is the forecast error of the monitored combination forecast. The monitoring is conducted using ensemble of machine learning algorithms and extracting more than 300 time-series features extracted from past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and the combination forecast88 as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. The ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the combination forecast. The approach selects the combination forecast (the historical average forecast) when p_t is larger than 0.5 (less than 0.5). The differenec of square forecast error is then multiplied by 10^4 and the past 60 month rolling mean is computed and reported in this figure.

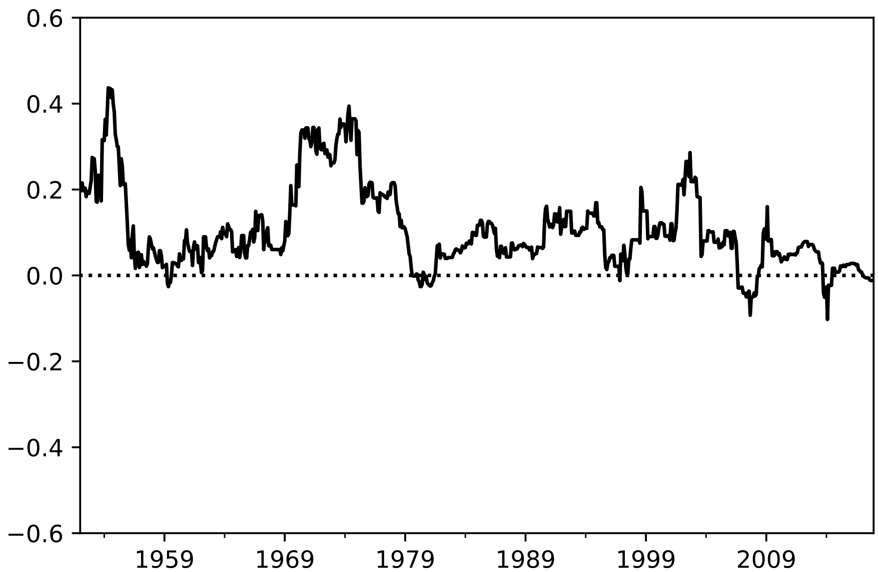


Figure 1.3 Out-of-sample R^2 (%) of Combination Forecast and Monitored Combination Forecast with Moving Evaluation Starting Date

This figure plots out-of-sample R^2 of the equal-weighted combination forecast and the monitored combination forecast. The monitoring is conducted using ensemble of machine learning algorithms and extracting more than 300 time-series features extracted from past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and the combination forecast as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. The ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the combination forecast. The approach selects the combination forecast (the historical average forecast) when p_t is larger than 0.5 (less than 0.5). The evaluation starting date starts from January 1947 to December 2012. The evaluation always ends at December 2017. The out-of-sample R^2 is computed as the percentage reduction in mean square forecast error of the model of interest relative to the historical average benchmark: $R^2_{OS} = 1 - \frac{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{i,t|t-1})^2}{\sum_{t=t_0}^{t_1}(r_t - \hat{r}_{HA,t|t-1})^2}$. where $\hat{r}_{i,t|t-1}$ is the forecast of interest and $\hat{r}_{HA,t|t-1}$ is the historical average benchmark forecast. The out-of-sample R^2 is multiplied by 100, to denote percentage values. A positive R^2_{OS} implies that the model of interest generates more accurate forecasts than the benchmark. Each point in the figure represents a evaluation period with starting date in the x-axis and ending date at December 2017.

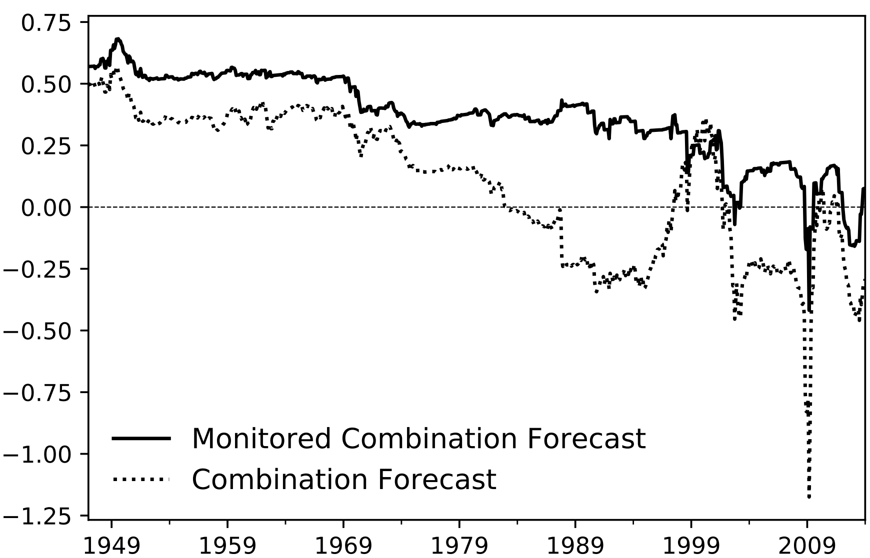


Figure 1.4 Weight on the Combination Forecast From Monitoring and NBER Recession

This figure plots the estimated weight on the combination forecast and NBER recessions. The weight on the combination forecast is calculated from monitoring via ensemble of machine learning algorithms. The monitoring approach utilizes more than 300 time-series features extracted from past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and the combination forecast as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. The ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the combination forecast. Rolling mean of the estimated weights from past 24 month are reported in the figure. The gray shaded areas represent periods labeled by NBER as recessions.

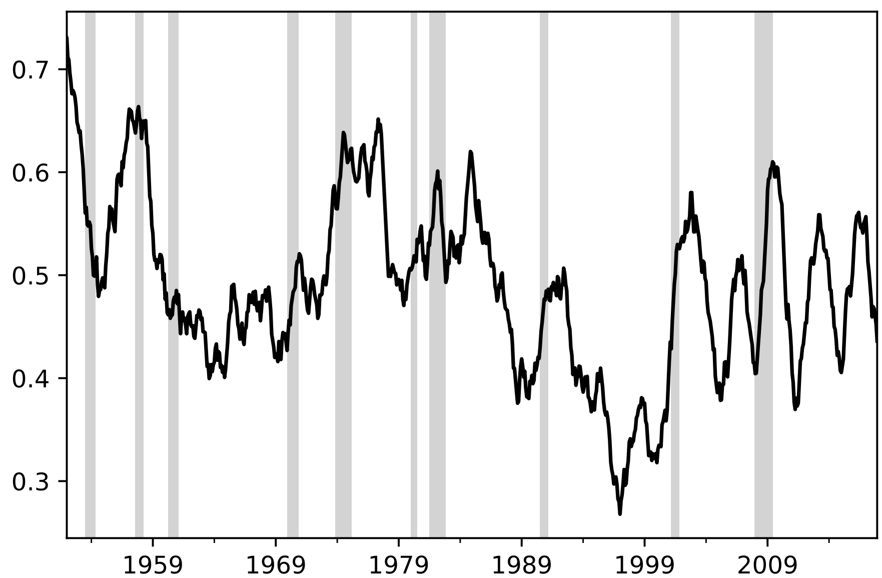
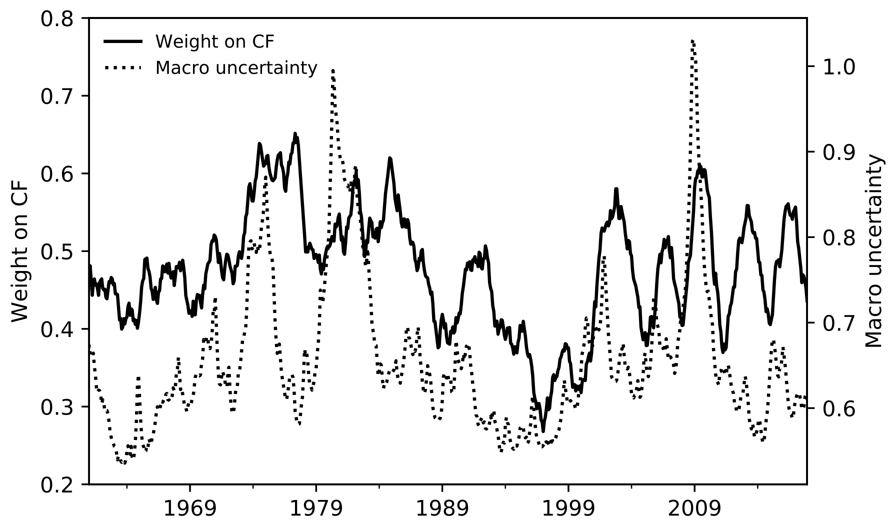


Figure 1.5 Weight on the Combination Forecast From Monitoring and Macro Uncertainty

This figure plots the estimated weight on the combination forecast and macro uncertainty. The weight on the combination forecast is calculated from monitoring via ensemble of machine learning algorithms. The monitoring approach utilizes more than 300 time-series features extracted from past 60 month lagged difference of squared forecast error (DSFE) between the historical averge benchmark forecast and the combination forecast as conditioning variables. The time-series features are calculated following Christ, Braun, Neuffer, and Kempa-Liehr (2018) using the python package TSFRESH. The ensemble of machine learning algorithms output a probability measure p_t indicating the weight to put on the combination forecast. Rolling mean of the estimated weights from past 24 month are reported in the figure. Macro uncertainty measure is calculated following Jurado, Ludvigson, and Ng (2015) which is available from 1960 to 2017.



Bibliography

Aharony, J. and I. Swary (1980). Quarterly Dividend and Earnings Announcements and Stockholders' Returns: An Empirical Analysis. *Journal of Finance 35*(1), 1–12.

Aiolfi, M. and A. Timmermann (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of econometrics 135*(1-2), 31–53.

Ball, R. and P. Brown (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research 6*(2), 159.

Barber, B. M. and T. Odean (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies 21*(2), 785–818.

Barber, B. M., T. Odean, and N. Zhu (2009). Systematic noise. *Journal of Financial Markets 12*(4), 547–569.

Bates, J. M. and C. W. J. Granger (2017). The Combination of Forecasts. *Journal of the Operational Research Society 20*(4), 451–468.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*.

Campbell, J. Y. and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies 21*(4), 1509–1531.

Christ, M., N. Braun, J. Neuffer, and A. W. Kempa-Liehr (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing 307*, 72–77.

Clements, M. P. and D. F. Hendry (2006). Chapter 12 Forecasting with Breaks. *Handbook of Economic Forecasting 1*, 605–657.

Cochrane, J. H. (1999). New Facts in Finance. *Economic Perspectives 23*, 36–58.

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. *The Review of Financial Studies 21*(4), 1533–1575.

Cujean, J. and M. Hasler (2017). Why Does Return Predictability Concentrate in Bad Times? *Journal of Finance 72*(6), 2717–2758.

Da, Z., J. Engelberg, and P. Gao (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *The Review of Financial Studies 28*(1), 1–32.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157–181.

Dellavigna, S. and J. M. Pollet (2009). Investor Inattention and Friday Earnings Announcements. *Journal of Finance 64*(2), 709–749.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(1), 3–25.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

Ferreira, M. A. and P. Santa-Clara (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics 100*(3), 514–537.

Ferson, W. E. and C. R. Harvey (1993). The Risk and Predictability of International Equity Returns. *The Review of Financial Studies 6*(3), 527–566.

Freyberger, J., A. Neuhierl, and M. Weber (2017). Dissecting Characteristics Nonparametrically. *University of Wisconsin-Madison*.

Gibbs, C. G. (2017). Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts. *SSRN Electronic Journal*.

Giglio, S. and D. Xiu (2017). Inference on Risk Premia in the Presence of Omitted Factors. *University of Chicago*.

Granger, C. W. J. (1989). Invited review combining forecasts—twenty years later. *Journal of Forecasting 8*(3), 167–173.

Granziera, E. and T. Sekhposyan (2017). How to Predict Your Next Forecasting Model: Conditional Predictive Ability Approach.

Gu, S., B. T. Kelly, and D. Xiu (2018). Empirical Asset Pricing Via Machine Learning. *SSRN Electronic Journal*.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *The Econometrics Journal 7*(1), 1–31.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*(3), 560–580.

Jurado, K., S. Ludvigson, and S. Ng (2015). Measuring Uncertainty. *American Economic Review 105*(3), 1177–1216.

Kelly, B. T., S. Pruitt, and Y. Su (2017). Some Characteristics Are Risk Exposures, and the Rest Are Irrelevant. *SSRN Electronic Journal*.

Kim, K. and N. R. Swanson (2016). Mixing Mixed Frequency and Diffusion Indices in Good Times and in Bad. *SSRN Electronic Journal*.

Kozak, S., S. Nagel, and S. Santosh (2017). Shrinking the Cross Section. *University of Michigan*.

Lettau, M. and S. Ludvigson (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance 56*(3), 815–849.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics 86*(2), 446–478.

Moritz, B. and T. Zimmermann (2016). Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns. *SSRN Electronic Journal*.

Pan, Z., D. Pettenuzzo, and Y. Wang (2017). Forecasting Stock Returns: A Predictor-Constrained Approach. *SSRN Electronic Journal*.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*(3), 274–315.

Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of econometrics 164*(1), 60–78.

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics 114*(3), 517–553.

Pontiff, J. and L. Schall (1998). Book-to-market ratios as predictors of market returns1. *Journal of Financial Economics 49*(2), 141–160.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies 23*(2), 821–862.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International Stock Return Predictability: What Is the Role of the United States? *Journal of Finance 68*(4), 1633–1662.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

Timmermann, A. (2006). Forecast Combinations. *Handbook of Economic Forecasting 1*, 135–196.

Welch, I. and A. Goyal (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies 21*(4), 1455–1508.

Zhu, Y. and A. Timmermann (2017). Monitoring forecasting performance. *UCSD working paper*.

Chapter 2

A Comprehensive Look at the Option-Implied Predictors of Stock Returns

2.1 Introduction

Stock return predictability is not necessarily clear evidence against efficient markets, as opposed to conventional wisdom. At the aggregate level, market return predictability can be viewed as a result of the time-varying equity premium. At the individual firm or portfolio level, the cross-sectional variations of expected stock returns can be attributed to different exposures to risk factors. However, evidence on return predictability is subject to statistical biases, poor out-of-sample return predictability, and difficulty in identifying risk factors.[1] As a result, literature has provided mixed evidence on the nature of stock return predictability, in particular, at the individual firm or portfolio level.

First, we examine whether ex-ante measures of firm-level dividend yields can forecast future stock returns. Theory suggests that the ex-ante dividend yield is positively related to future stock returns. This is due to the time-varying equity premium and the present value relation as described in Fama and French (1988) and Campbell and Shiller (1988). Consistent with this theoretical argument, numerous papers including Cochrane (2011) find a positive relationship between the dividend yield and future market returns. Golez (2014) refines the well-known return predictor dividend-price ratio by extracting the expected dividend growth

[1] See Stambaugh (1999) and Ferson et al. (2003) for statistical biases and see Welch and Goyal (2008) for out-of-sample return predictability.

from options data and finds stronger predictability of market returns. His results provide further support to the view that the market return predictability is due to time-varying equity premium. At the firm-level, we compute ex-ante measures of dividend yields using equity options data. When we sort individual stocks based on their implied dividend yields every month, the subsequent monthly return difference between the top and bottom quintile is -0.58% per month with t-stat of -2.34. This cross-sectional variation in the expected return is puzzling in that the firm-level implied dividend yield is negatively related to the subsequent stock return as opposed to the positive relation theoretically expected.

We argue that the main driver of this puzzling empirical fact is contamination of firm-level option-implied dividend yields. For example, information asymmetry between traders in options and equity markets can severely distort measurements of option-implied dividend yields based on the no-arbitrage principle. The intuition is simple. When the traders in the options market have superior information than the traders in the equity market, the underlying stock price temporarily deviates from what option prices in the put-call parity imply. Such deviation is captured by the expected dividend when we extract the expected dividends from the put-call parity. Therefore, the measure of the expected dividend is contaminated, and so are the measures of implied dividend yield and the corrected dividend-price ratio. It is expected that all options will suffer from information asymmetry, probably with less effects for longer maturity options. Besides, short-term options will unlikely be able to contain predictive information about future dividend payments since short-term future dividend payments are rather fixed. Thus, we control for the information asymmetry by looking at the term structure of implied dividend yield and isolate the information about implied dividend yield. We show that the positive relationship between implied dividend yield and expected returns will recover after information asymmetry is controlled.

We also develop a hypothesis that the effect of information asymmetry will disappear within a few months, and thereafter the normal relationship between the implied dividend yield and the expected return will recover. We test this hypothesis by both cross-sectional and panel data analyses. In the cross-sectional analysis, when we sort individual stocks based on their implied dividend yields every month, the negative future return difference between the top and bottom quintile persists up to 3 months and become insignificant for longer holding periods. In panel analysis with firm fixed effects, we find the return predictability pattern consistent with our hypothesis. Our results are robust to the use of an alternative valuation ratio, the corrected dividend-price ratio by Golez (2014).

The reason why the cross-sectional analysis fails to find the same return-predictability pattern as the panel analysis is as follows. First, the dividend yield represents the expected return based on the present-value relation which is originally defined in the time-series environment in Campbell and Shiller (1988). In the cross-sectional setting, excessive variations in the expected dividend growth across firms completely dominate the effects of the dividend yields. Second, the cross-sectional analysis ignores the effect of the common time-variation in firm-level implied dividend yields. As a result, how the expected return varies with the dividend yield within a given firm is ignored in estimation. Third, a cross-sectional analysis is not suitable for the data with a short sample period, that is, when conditional information matters, in general. Fourth, most importantly, cross-sectional variations in expected returns can be driven by unidentified risk factors. We include the firm fixed effects to circumvent the issue of unidentified risk factors.

Our findings shed light on the existing mixed evidence on the relationship between risk-neutral skewness and expected return. Xing et al. (2010) show that high (low) slope of the volatility smirk predicts low (high) stock returns, implying the traders in the options market have superior information than the equity traders. Since the high slope of the volatility smirk is often assumed to be associated with negative risk-neutral skewness of stock returns, their finding implies the positive relationship between risk-neutral skewness and the expected return, as opposed to Conrad et al. (2013) and the asset pricing models with idiosyncratic skewness developed in Brunnermeier et al. (2007) and Barberis and Huang (2008).

We examine if information asymmetry can explain the existing mixed evidence on the relationship between risk-neutral skewness and the expected return. We find that option-implied measures such as the slope of the volatility smirk (Xing et al., 2010), and the model-free risk-neutral skewness (Conrad et al., 2013) also suffer from the same issue of contamination.[2] With such option-implied measures, we confirm that the effect of information asymmetry disappears within a few months, and thereafter the normal negative relationship between risk-neutral skewness and the expected return recovers. Further, we look at situations where information asymmetry is more prominent with the use of proxy extracted from analyst forecast data. We argue that the higher the forecast dispersion about future earnings, the easier for traders with information advantage to hide and exploit this private news in the options market, which will contaminate the option-implied measures further. We show that when analyst forecast dispersion is at a higher level, the effect of information asymmetry on option-implied measures will be stronger within a few months. To summarize, the issues in option-implied measures and cross-sectional analysis cause such mixed evidence on the relationship between option-implied skewness and the expected return, and we reconcile the mixed evidence using a panel analysis with varying forecasting horizons.[3]

The remainder of the paper is organized as follows. Section 2.2 introduces different option-implied measures and explains why they are potentially contaminated by nature. Section 2.3 describes the data sources and data treatments. Section 2.4 provides empirical results to support our hypothesis. Section 2.5 concludes.

[2]The contamination issue exists also in other option-implied measures used in Ofek et al. (2004), Yan (2011), Cremers and Weinbaum (2010), and Kalay et al. (2014).

[3] Yan (2011) shows that the relationship between risk-neutral skewness and the expected return can be either positive or negative depending on the parameter values in a model with jumps, and he further argues that it is purely an empirical question and so introducing information asymmetry is not necessary. However, we explain why it is difficult to reject the existence of information asymmetry in Section 2.2.5 .

2.2 Option-Implied Predictors for Stock Returns

2.2.1 Implied Dividend Yield

There are two ways to extract implied dividend yields from index derivatives. The first one is to use put-call parity of options introduced in Binsbergen and Koijen (2010). The second method is to use both options and futures data described in Golez (2014). Apart form estimating implied dividend yields, the second method also simultaneously estimates interest rates from futures prices. However, the estimates of interest rates could be different from the interest rates inferred from bond markets data. Besides, the estimated interest rates could contain measurement errors. Therefore, in this paper, we only use options data to derive dynamics of implied dividend yields.

To compute implied dividends from options data, we require only the absence of arbitrage opportunities. Under this condition, put-call parity for European options holds:

$$c_{t,T} + Ke^{-r_{t,T}(T-t)} = p_{t,T} + S_t - D_{t,T}$$

where $c_{t,T}$ and $p_{t,T}$ are prices of call and put options at time t, with maturity T and strike price K.[4] $r_{t,T}$ is the interest rate between time t and T. $D_{t,T}$ is the expected dividends paid between time t and T under the risk-neutral probability defined as

$$D_{t,T} = \sum_{i=1}^{T} E_t(M_{t:t+i} d_{t+i})$$

Here, $M_{t:t+i}$ is a stochastic discount factor to discount future dividends d_{t+i}. Equity price is given by the sum of discounted dividend values:

$$S_t = \sum_{i=1}^{\infty} E_t(M_{t:t+i} d_{t+i})$$

[4]Since individual firm options are American options, the put-call parity becomes a band with inequality. Therefore, we acknowledge that the firm-level option-implied measures based on strict put-call parity are still noisy proxies at best even without information asymmetry.

Thus, the implied dividend yield is $IDY = D_{t,T}/S_t$. We use pairs of call option and put option with same strike price and same time to maturity to estimate implied dividend yield. We will interpolate the term structure of interest rate to get an appropriate discount rate.

The implied dividend yield can be used as a predictor of future stock returns because of the present value relation explained in Fama and French (1988) and Campbell and Shiller (1988). From a simple dividend growth model of stock prices, we have

$$S_{t-1} = \frac{D_t}{r - g}$$

Therefore, the historical dividend yield $D_t/S_{t-1} = r - g$ includes information about the discount rate which is the expected return. The option-implied dividend yield $IDY = D_{t,T}/S_t$ we construct is an ex-ante version of this the historical dividend yield.

2.2.2 Corrected Dividend-Price Ratio

The corrected dividend price ratio measure comes from the model in Campbell and Shiller (1988) and Golez (2014) where time-series property of expected returns are assumed. We first define log return r_{t+1}, log dividend growth rate Δd_{t+1}, and log dividend-price ratio dp_t as follows:

$$r_{t+1} = log[\frac{P_{t+1} + D_{t+1}}{P_t}], \Delta d_{t+1} = log[\frac{D_{t+1}}{D_t}], dp_t = log[\frac{D_t}{P_t}]$$

where P_t is the price at time t and D_t is the dividend paid from $t-1$ to t. Then, use taylor expansion around the average of dividend-price radio $\bar{dp}$,

$$r_{t+1} = \kappa + dp_t + \Delta d_{t+1} - \rho dp_{t+1}$$

where $\kappa = log[1 + exp(-\bar{dp})] + \rho\bar{dp}$ and $\rho = \frac{exp(-\bar{dp})}{1+exp(-\bar{dp})}$. After iterations, we obtain the Campbell and Shiller (1988) present value identity

$$dp_t = -\frac{\kappa}{1-\rho} + E_t\sum_{j=0}^{\infty}\rho^j(r_{t+1+j}) - E_t\sum_{j=0}^{\infty}\rho^j(\Delta d_{t+1+j})$$

Denote $\mu_t = E_t(r_{t+1})$. Assume μ_t follows AR(1) processes:

$$\mu_{t+1} \quad = \quad \delta_0 + \delta_1\mu_t + \epsilon^{\mu}_{t+1}$$

By plugging in the AR(1) assumption of expected return, we find the dividend price ratio as follows,

$$dp_t = \phi + (\frac{1}{1-\rho\delta_1})\mu_t - E_t\sum_{j=0}^{\infty}\rho^j(\Delta d_{t+1+j})$$

Finally, we derive a return forecasting equation:

$$\mu_t = E(r_{t+1}) = \psi + (1-\rho\delta_1)dp_t + (1-\rho\delta_1)E_t\sum_{j=0}^{\infty}\rho^j(\Delta d_{t+1+j}) + v^r_{t+1} \tag{2.1}$$

where $\Delta d_{t+1+j} = log[\frac{D_{t+1+j}}{D_{t+j}}] = log(D_{t+1+j}) - log(D_{t+1})$. We can also write the return forecasting equation with a single factor as

$$E(r_{t+1}) = \psi + (1-\rho\delta_1)dp^{CorrTS}_t + v^r_{t+1} \tag{2.2}$$

where $dp^{CorrTS}_t = dp_t + E_t\sum_{j=0}^{\infty}\rho^j(\Delta d_{t+1+j})$ is the dividend-price ratio corrected for term structure of implied dividend growth rates.

According to Equation (2.1), the future expected return is a function of historical dividend-price ratio and the expected forward dividend growth rates. In Equation (2.2), expected return can be more accurately measured by the dividend-price ratio after subtracting the term structure of implied dividend growth rates. We will use options with different maturities to estimate implied forward dividends. Golez (2014) assume AR(1) process for implied dividend growth and derive the dividend-price ratio corrected for single implied dividend growth rate estimated from option pairs with six months to maturity. The dividend-price ratio derived in Golez (2014) is

$$dp_t^{Corr} = dp_t + g_t(\frac{1}{1 - \rho\gamma_1})$$

where g_t is conditional expected dividend growth rate and γ_1 is AR(1) coefficient of the process of expected dividend growth rate.

2.2.3 Model-Free Risk-Neutral Skewness

We calculate individual firms' risk-neutral skewness following the results in Bakshi and Madan (2000) and Bakshi et al. (2003). They show that the payoff to any security can be replicated and priced using a set of options with different strike prices on that security. Bakshi and Madan (2000) define quadratic contract, cubic contract, and quadratic contracts as having payoffs

$$H[S] = \begin{cases} R(t,\tau)^2 & \text{volatility contract} \\ R(t,\tau)^3 & \text{cubic contract} \\ R(t,\tau)^4 & \text{quartic contract} \end{cases}$$

where $R(t,\tau) \equiv \ln[S(t+\tau)] - \ln[S(t)]$ is the log-return of the stock. Using the prices of these contracts, model-free risk-neutral moments may be computed as

$$\mathrm{VAR}(t,\tau) = e^{r\tau}V(t,\tau) - \mu(t,\tau)^2$$

$$\mathrm{SKEW}(t,\tau) = \frac{e^{r\tau}W(t,\tau) - 3\mu(t,\tau)e^{r\tau}V(t,\tau) + 2\mu(t,\tau)^3}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2)]^{2/3}}$$

$$\mathrm{KURT}(t,\tau) = \frac{e^{r\tau}X(t,\tau) - 4\mu(t,\tau)e^{r\tau}W(t,\tau) + 6e^{r\tau}\mu(t,\tau)^2V(t,\tau) - 3\mu(t,\tau)^4}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2)]^2}$$

where V, W and X represent the fair values of the volatility, cubic and quadratic contract, respectively. These prices are computed by integrating over a set of strike prices, as

$$\begin{aligned} V(t,\tau) &= \int_{S(t)}^{\infty} \frac{2(1-\ln[K/S(t)])}{K^2} C(t,\tau;K)dK \\ &\quad + \int_0^{S(t)} \frac{2(1+\ln[S(t)/K])}{K^2} P(t,\tau;K)dK \end{aligned}$$

$$\begin{aligned} W(t,\tau) &= \int_{S(t)}^{\infty} \frac{6\ln[K/S(t)] - 3(\ln[K/S(t)])^2}{K^2} C(t,\tau;K)dK \\ &\quad - \int_0^{S(t)} \frac{6\ln[S(t)/K] + 3(\ln[S(t)/K])^2}{K^2} P(t,\tau;K)dK \end{aligned}$$

$$\begin{aligned} X(t,\tau) &= \int_{S(t)}^{\infty} \frac{12(\ln[K/S(t)])^2 - 4(\ln[K/S(t)])^3}{K^2} C(t,\tau;K)dK \\ &\quad + \int_0^{S(t)} \frac{12(\ln[S(t)/K])^2 + 4(\ln[S(t)/K])^3}{K^2} P(t,\tau;K)dK \end{aligned}$$

In the above equations, $C(t,\tau;K)$ and $P(t,\tau;K)$ are the prices of European calls and puts written on the underlying stock with strike price K and expiration τ periods from time t. As shown in the equation, we use a weighted sum of out of the money options across different strike prices to construct the ex-ante risk-neutral skewness of stock returns. Following Conrad et al. (2013), we set apart four maturity buckets. Each time to maturity is assigned to one of 1-month, 3-month, 6-month, and 12-month maturity. We calculate the risk-neutral skewness from options with time to maturity closest to 3 month.

2.2.4 Volatility Skew

Xing et al. (2010) define volatility skew (the slope of volatility smirk) as the difference between the implied volatilities of out-of-the-money puts and at-the-money calls.

$$Vol\_skew_{i,t} = VOL_{i,t}^{OTM,P} - VOL_{i,t}^{ATM,C}$$

where $VOL_{i,t}^{OTM,P}$ is the implied volatility of an out-of-the-money put option with the ratio of the strike price to the stock price lower than 0.95 (but higher than 0.80), and $VOL_{i,t}^{ATM,C}$ is the implied volatility of an at-the-money call option with the ratio of the strike price to the stock price between 0.95 and 1.05. We follow Xing et al. (2010) to restrict our attention to options with time to maturity between 10 and 60 days. When there are more than one pair of out-of-the-money put option and at-the-money call option, we weight all available options with positive volume equally.

2.2.5 Contamination of Option-Implied Measures

Option-implied measures are supposed to be contaminated by information asymmetry between traders in options and equity markets. The intuition is simple. When the implied dividend yield is constructed from options data, the main assumption is the absence of arbitrage opportunities. Under this condition, the present (expected) value of dividend before maturity is extracted from the put-call parity:

$$c_{t,T} + Ke^{-r_{t,T}(T-t)} = p_{t,T} + S_t - D_{t,T}$$

where $c_{t,T}$ and $p_{t,T}$ are prices of call and put options at time t, with maturity T and strike price K. $r_{t,T}$ is the interest rate between time t and T. $D_{t,T}$ is the expected dividends paid between time t and T.

When the traders in options market have superior information than the traders in equity market as argued in Xing et al. (2010) and Cremers and Weinbaum (2010), the underlying stock price S_t temporarily deviates from what option prices in the put-call parity imply. Such deviation is captured by the expected dividend when we extract the expected dividends from the put-call parity. Therefore, the measure of expected dividend is contaminated, and so are the measures of implied dividend yield and the corrected dividend-price ratio.

When the volatility skew is computed following Xing et al. (2010), deviations of underlying stock price S_t from the fair price will affect implied-volatility calculations. For example, if some negative news is available only in options market, then the observed stock price will be higher than the full-information price. The out-of-the-money put option price will be seen too high to equity traders, and the calculated implied volatility will be higher than the true implied volatility. Therefore, the measure of the volatility skew will be contaminated by information asymmetry. In fact, Xing et al. (2010) acknowledge the existence of information asymmetry and interpret the contaminated volatility skew as a proxy for information asymmetry. However, if we view the volatility skew as a proxy for negative ex-ante risk-neutral skewness, their findings contradict Conrad et al. (2013) and the asset pricing models with skewness developed in Brunnermeier et al. (2007) and Barberis and Huang (2008). Our view is that the measured volatility skew captures both information asymmetry and negative skewness.

When the model-free risk-neutral skewness is computed following Bakshi et al. (2003), deviations of underlying stock price S_t from the fair prices also play a role since the stock price S_t is used in calculation. Similar to the case of volatility skew, the measure of model-free risk-neutral skewness will be contaminated by information asymmetry and will represent both information asymmetry and risk-neutral skewness.[5]

Based on potential contamination of option-implied measures, we develop a hypothesis that the effect of information asymmetry will disappear within a few months, and thereafter the normal relationship between the option-implied measures and the expected return will recover. Here, the normal relationship means a positive association with expected returns in case of the implied dividend yield, the corrected dividend price ratio, and volatility skew while negative in case of the risk-neutral skewness. Contamination of option-implied measures by information asymmetry can even revert the sign of the relationship if contamination is severe. However, if traders in equity market resolve information asymmetry in a few months, the normal relationship can appear thereafter. We empirically test this hypothesis in the rest of the paper in order to reconcile the well-known mixed evidence on the relationship between option-implied skewness and expected returns: Xing et al. (2010) and Yan (2011) vs. Conrad et al. (2013).

In fact, the positive relationship between risk-neutral skewness and expected return is not necessarily abnormal. Yan (2011) shows that such relationship can be either positive or negative depending on the parameter values in a model with jumps and argues that it is purely an empirical question and introducing information asymmetry is not necessary. However, we do not rule out existence of information asymmetry for three reasons. First, the model cannot explain why the relationship between risk-neutral skewness and expected return changes from "positive" to "negative" over forecasting horizons. Second, it is very difficult to find parameter values in the model that can explain the negative relationship at the market level while positive at the firm-level even though we focus on only short-term expected returns. Third, the evidence about information asymmetry in literature is quite

[5]Although we do not study in this paper, measures for historical (realized) skewness such as Amaya et al. (2013) are potentially contaminated as well since the realized return due to information asymmetry will be included when realized skewness is calculated.

strong. Xing et al. (2010) show information asymmetry is linked to the subsequent earning surprises, and Cremers and Weinbaum (2010) find the short sale constraint cannot explain the put-call parity deviation and its return predictability as opposed to Ofek et al. (2004).

2.3 Data

Our sample period is from January 1996 to December 2013. Options data are from Optionmetrics (provided through Wharton Research Data Services). Closing prices are calculated as the average of closing bid and ask prices. Data on stock returns are obtained from Center for Research in Security Prices. We use monthly returns from 1996 to 2013 for all individual securities with positive common shares outstanding. Balance sheet data for the computation of book-to-market ratios and leverage ratios are from Compustat. Interest rates are obtained from a collection of continuously compounded zero-coupon interest rates at different maturities from OptionMetrics.

To calculate risk-neutral skewness, we follow the procedure as in Bakshi, Kapadia, and Madan (2003) using out-of-the-money puts and calls. We employ options with time to maturity close to 3 month and with positive open interest. For each day, we require at least two OTM puts and two OTM calls to calculate risk-neutral skewness. If there are more puts than calls, then we use the puts that have the most similar strike to price ratio as the calls, vice versa if there are more calls than puts. For each month, we average the risk-neutral skewness for each day in this month to get a monthly measure of risk-neutral skewness. The sample consists of 70,095 firm-month observations of risk-neutral skewness with mean of -0.48 and standard deviation 0.33 over the time period January 1996 to December 2013.

When calculating the volatility skew, we apply the following filters to daily options data as in Xing, Zhang, and Zhao (2010): including options with positive volume for underlying stock, implied volatility between 3% and 200%, price larger than $0.125, positive open interest and nonmissing volume, and maturity between 10 to 60 days. We use ATM call options with moneyness between 0.95 and 1.05 and OTM put options with moneyness between 0.80

and 0.95. We first calculate daily volatility skew by using the differences between implied volatilities between OTM puts and ATM calls. Then we average the daily volatility skew to get monthly measures for each firm. We end up with 151,771 firm-month observations of volatility skew with mean 4.08 and standard deviation 4.91.

In estimating the implied dividend yield from options, we follow the procedure described in Binsbergen, Brandt, and Koijen (2012). Options with positive volume or open interest greater than 200 contracts are considered. Each day, we find paris of call and put with same strike price and same time to maturity and calculate implied dividend yields for this pair from put-call parity. For each day, we compute the mean of implied dividend yield to get the daily measure. For each month, we aggregate all daily estimates of implied dividend yields and calculate the mean of daily estimates. The sample has 392,725 firm-month observations of implied dividend yields with mean of 0.02 and standard deviation of 0.05. We also calculated dividend price ratio corrected for implied dividend growth for firms paying dividends. We have 38,656 firm-month measures of log corrected dividend price ratio with mean -3.24 and standard deviation of 1.07.

Table 2.1 provides the summary statistics of monthly firm-level option-implied measures and control variables used in the paper. IDY is the ex-ante option-implied dividend yield from put-call parity. log DP^c is the corrected dividend-price ratio from Golez (2014). RNSKEW is the model-free risk-neutral skewness from Bakshi et al. (2003). HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and log LEV are firm market capitalization, the book-to-market ratio, and the leverage in a log scale, respectively.

2.4 Empirical Results

2.4.1 Option-implied Measures

Figure 2.1 shows how the distribution of firm-level implied dividend yield evolves over time. The cross-sectional mean and the median of implied dividend yield change over time significantly, implying there are substantial common time-variations of firm-level implied dividend yield.

Table 2.2 shows the correlation matrix of monthly firm-level option-implied measures and control variables. The ex-ante option-implied dividend yield (IDY) and the corrected dividend-price ratio (DP^c) from Golez (2014) are positively correlated and the correlation coefficient is 0.746. They are, in fact, very similar measures. The only difference is that the implied dividend yield (IDY) captures both the expected return and expected dividend growth by construction while the corrected dividend-price ratio (DP^c) represents only the expected return by subtracting the expected dividend growth term from the historical dividend price ratio. Thus, the implied dividend yield is still a noisy proxy for the expected return. However, we consider the implied dividend yield, hoping that it can be cleaner measure than the historical dividend yield in that the historical dividend yield includes the realized dividend which is noisy by nature. On the other hand, the corrected dividend-price ratio (DP^c) is still a noisy proxy for the expected return because measuring the dividend growth term is subject to estimation errors and model misspecification embedded in the procedure by Golez (2014). Therefore, we consider both the implied dividend yield (IDY) and the corrected dividend-price ratio (DP^c) in our analysis.

Volatility skew (VOLSKEW) from Xing et al. (2010), which is the slope of volatility smirk, is positively correlated with the implied dividend yield (IDY) and the corrected dividend-price ratio (DP^c). This positive correlation represents two different aspects of these measures. First, high implied dividend yield or high corrected dividend-price ratio implies high expected return. High volatility skew generally translates into more negative skewness which means high expected returns as shown in Brunnermeier et al. (2007) and Barberis and Huang (2008). Therefore, they are supposed to be positively correlated. Second, as explained in Section 2.2.5, these option-implied measures are potentially contaminated by information asymmetry between options and equity markets to the same direction. Therefore, these option-implied measures are supposed to be positively correlated for this reason as well. Note the model-free risk-neutral skewness (RNSKEW) from Bakshi et al. (2003) is negatively correlated with the implied dividend yield (IDY), the corrected dividend-price ratio (DP^c), and the volatility skew (VOLSKEW) since high risk-neutral skewness generally translates into low volatility skew (the slope of volatility smirk).

Figure 2.2 shows the time-series of the cross-sectional medians of four option-implied measures: IDY, log DP^c, RNSKEW, and VOLSKEW. All variables are standardized by their sample means and standard deviations for better visualization. Note we draw −RNSKEW instead of RNSKEW so that all four option-implied measures in the plot are associated with expected returns in the same way. In time-series, the implied dividend yield (IDY) and the corrected dividend-price ratio (DP^c) are highly correlated as expected from their definition. The negative model-free risk-neutral skewness (−RNSKEW) and the volatility skew (VOLSKEW) are also highly correlated, confirming that they both measure negative skewness. All four measures are highly correlated after 2006 suggesting skewness becomes a dominant factor in equity valuations.

Table 2.3 reports the average firm-characteristics of each decile portfolio sorted by the implied dividend yield every month. Although the correlations between the implied dividend yield and firm characteristics in Table 2.2 are not very high, the average firm-characteristics except Sharpe ratio have monotonic relationship with the implied dividend yield. Since the implied dividend yield is a valuation ratio, it is supposed to be related to firm-characteristics. By definition, the implied dividend yield does not compete with firm-characteristics. Rather, it summarizes all information contained in firm-characteristics related to firm valuation.

Roughly speaking, low implied dividend yield firms are small, growth, low-leveraged, more volatile, more positively skewed firms with higher kurtosis.

2.4.2 Cross-sectional Analysis

Table 2.4 shows raw and risk-adjusted average returns of five equal-weighted portfolios sorted by the ex-ante option-implied dividend yield (IDY). We find that high implied dividend yield is associated with lower subsequent returns. Panel A shows that the top implied-dividend-yield portfolio has the subsequent monthly return significantly lower than the bottom implied-dividend-yield portfolio. The subsequent monthly return difference between top and bottom quintile is -0.58% per month with t-stat -2.34 computed using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. This difference remains robust after we adjust risk by CAPM, Fama-French 3-factor model, and Carhart 4-factor model. Therefore, this sorting exercise implies that high implied dividend yield means low expected return as opposed to the well-known market level evidence and what the present-value relation in Fama and French (1988) and Campbell and Shiller (1988) implies. We repeat this analysis with different holding periods up to twelve months and find the similar pattern up to three-month holding periods but not longer.

We examine the effects of information asymmetry on IDY by differentiating between short-term implied dividend yield and long-term implied dividend yield. Short-term implied dividend yield is the IDY measure computed using only options with maturity less than 60 days while long-term implied dividend yield is IDY calculated with options maturing no earlier than 6 months. If there are possible information advantage to option traders, it would affect both short-term and, to a less extent, long-term put-call parity deviation. And the short-term put and call options will unlikely to have any significant superior information about expected dividend payments since the short maturity and smooth dividend payments in short period. Thus, if we take the difference between long-term IDY and short-term IDY and cancel out the effect of information asymmetry, the remaining part will reflect solely information on expected dividend yield. Table 2.5 reports the raw and risk-adjusted average returns of five equal-weighted quintile portfolios sorted by four measures related to option-implied dividend yield. Panel A reports the results with respect to IDY calculated from all available options. Panel B and C shows quintile returns of portfolios sorted by short-term IDY (from options with maturity less than 60 days) and long-term IDY (from options maturing 6 months later). Portfolio returns sorted by the term structure of IDY (the difference between long-term IDY and short-term IDY) are presented in Panel D. We observe from first three panels that the negative relationship between IDY and returns are persistent with respect to different maturities of options used to compute this measure. When we sort all firms by the term structure of IDY, monotonically increasing portfolio average returns and risk-adjusted returns are reported in Panel D. It suggests that the positive relationship between dividend yield and expected returns recovers once we control for the information asymmetry by taking the difference between long-term IDY and short-term IDY.

We perform a conventional cross-sectional analysis on stock returns with option-implied measures. Table 2.6 shows Fama-Macbeth regressions with monthly stock returns. For each month t, we run the following cross-sectional regression:

$$r_{i,t+1} = \alpha_t + \beta_t X_{i,t} + \gamma_t^\top Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+1}$ is the monthly stock return (%), $X_{i,t}$ is either the implied dividend yield $IDY_{i,t}$ or the corrected dividend-price ratio $\log DP^c_{i,t}$, and $Z_{i,t}$ is control variables. Then we compute the time-average of β_t and γ_t to find the point estimates and report their t-statistics in parentheses using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The coefficient on the implied dividend yield is negative and its t-statistic is about 5 with or without control variables including size, book-to-market, leverage, historical volatility, and historical skewness. Again, this result is counter intuitive since the present vale relation which is a simple accounting identity implies a positive coefficient. One potential explanation is that implied dividend yield $IDY_{i,t}$ and the corrected dividend-price ratio $\log DP^c_{i,t}$ are dominated by the effect of put-call parity deviations caused by information asymmetry between traders in options and equity markets. In that case, negative coefficients should be observed as Table 2.6.

In fact, if the implied dividend yield and the corrected dividend-price ratio are not contaminated by put-call parity deviations, the control variables should be excluded in the regression since the dividend yield and the dividend-price ratio are valuation ratios and thus already include information embedded in control variables related to the expected returns. However, we include the control variables because what we observe in the regression is mostly from the effect of put-call parity deviations caused by information asymmetry.

Table 2.7 repeats Table 2.4 with other option-implied measures such as model-free risk-neutral skewness (RNSKEW) and volatility skew (VOLSKEW) which is the slope of volatility smirk. Table 2.7 shows exactly the same pattern as Table 2.4. The coefficient on volatility skew is negative and significant with or without control variables. In case of risk-neutral skewness, the opposite sign on the coefficient is actually the same pattern as Table 2.4 because of its definition. The coefficient on risk-neutral skewness is positive and significant with or without control variables. Therefore, we suspect that we have counter-intuitive results here since all four option-implied measures are potentially contaminated by deviations from put-call parity caused by information asymmetry between traders in options and equity markets. The next section further investigates this hypothesis with panel data analysis.

2.4.3 Panel Data Analysis

Previous section shows that high implied dividend yield is associated with lower subsequent returns in the cross-section as opposed to exiting theories and market level evidence. In case of the subsequent monthly return, it is consistent with contamination of the option-implied measures by deviations from put-call parity caused by information asymmetry between traders in options and equity markets.

However, more puzzling fact is that such pattern persists for twelve months since information asymmetry is difficult to explain why such pattern persists for such a long period. Here, we argue that the conventional cross-sectional analysis is not suitable to reveal the true relationship between the implied dividend yield and the expected returns. First, the dividend yield represents the expected return based on the present-value relation which is originally defined in the time-series environment in Campbell and Shiller (1988). In the cross-sectional setting, the dividend yield will be difficult to capture variations in the expected return due to excessive variations in the expected dividend growth across firms. Second, the cross-sectional analysis ignores the effect of the common time-variation in firm-level implied dividend yields as shown in Figure 2.1. As a result, how the expected return varies with the dividend yield within a given firm is not considered at all in cross-sectional estimation. Third, a cross-sectional analysis is not suitable for the data with short sample period, that is, when conditional information matters, in general. Fourth, most importantly, cross-sectional variations in expected returns might be due to unidentified risk factors. We resolve these issue by performing a panel data analysis. In particular, we circumvent the issue of unidentified risk factors by including firm fixed effects. After all, we focus on how the expected return varies with the dividend yield within a given firm.

Table 2.8 shows the time-series predictability of monthly S&P500 index returns. We compare five different option-implied measures: the implied dividend yield (IDY), the corrected dividend-price ratio (log DP^c), the model-free risk-neutral skewness (RNSKEW), the volatility skew (VOLSKEW), and the corrected dividend-price ratio using information in the term structure of the expected dividend growth from Bilson et al. (2015). All these predictor variables are computed using the index options. Due to the short sample period from January 1996 to December 2013, evidence on predictability is somewhat weak, yet the signs of coefficients are all consistent with theories. High dividend yield and low skewness are associated with high expected return.

Table 2.9 shows the pooled time-series predictability of individual stock returns. We run the following panel regression:

$$r_{i,t+1} = \alpha_i + \beta \; IDY_{i,t} + \gamma^\top Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+1}$ is the monthly stock return (%), α_i is the firm fixed effect, IDY is the ex-ante option-implied dividend yield from put-call parity, and $Z_{i,t}$ is control variables. The table shows that the coefficient on the implied dividend yield is negative and highly significant as t-stat is around 10, consistent with the cross-sectional result. One interesting result here is that control variables become relatively less significant when the firm fixed effects are included whilst the implied dividend yield does not. If firm fixed effects capture all risk premium associated with identified and unidentified risk factors, remaining significant coefficients represent time-variation of risk premium. Since the coefficients on the implied dividend yield remain the same even after the firm fixed effects are included, we conclude the time-series relationship between the implied dividend yield and the expected return is more important or more dominant in the data than the cross-sectional relationship.

Table 2.10 extends Table 2.9 using different forecasting horizons up to twelve months. We run the following panel regression with each of forecasting horizon $h = 1, 2, 3, ..., 12$:

$$r_{i,t+h} = \alpha_i + \beta \; X_{i,t} + e_{i,t+1}$$

where $r_{i,t+h}$ is the h-months ahead non-overlapping monthly stock return (%), α_i is the firm fixed effect, and $X_{i,t}$ is either implied dividend yield $IDY_{i,t}$ in Panel A or the corrected dividend-price ratio $\log DP^c_{i,t}$ in Panel B. Here, including firm fixed effects is important since it eliminates all cross-sectional variations in expected returns due to risk factors. In Panel A, the coefficient on the implied dividend yield starts from a strongly negative value but gradually increases and becomes significantly positive after three month. We interpret this coefficient pattern as the information asymmetry disappears within a few months and the normal relationship between the implied dividend yield and the expected return recovers. The same pattern is discovered when the corrected dividend price ratio is used as a predictor variable since two measures are constructed in a very similar way. Here, the control variables should be excluded in the regression since the dividend yield and the dividend-price ratio are a valuation ratio and so already include information in control variables related to the expected returns.

Table 2.11 repeats Table 2.10 with the model-free risk-neutral skewness (RNSKEW) and the volatility skew (VOLSKEW), which is the slope of volatility smirk, as a predictor variable. We find exactly the same pattern as Table 2.10. In the first few months, information asymmetry dominates and so the coefficient is the opposite of what theory implies, yet the normal relationship between the skewness measures and the expected return recovers thereafter. Note the signs of the coefficients on the model-free risk-neutral skewness (RNSKEW) should be the opposite of the volatility skew (VOLSKEW) because of their definitions. In case of the model-free risk-neutral skewness and the volatility skew, control variables should be included because skewness is just one determinant of the expected return. Accordingly, we have stronger results when control variables are included as expected. Table 2.12 repeats Table 2.10 and 2.11 with overlapping holding period returns. After six month, the normal relationship with all option-implied measures and the expected return dominates the overall return predictability.

To confirm the validity of the option-implied measures, we test whether the option-implied dividend actually include information about future dividend. This is an important fundamental question since the option-implied dividend from put-call parity is a main building block of the implied dividend yield and the corrected dividend-price ratio. The first two rows of Table 2.13 show that the option-implied dividend actually include information about future dividend, and so two option-implied measures are not pure noises. The remaining part of Table 2.13 shows how the implied dividend yield predicts the dividend growth. The result is consistent with information asymmetry in the short term and existing literature in the long term.

2.4.4 Option-implied Measures and Information Asymmetry

We argue that the information asymmetry between traders in options and equity markets creates the puzzling evidence of opposing relationship between option-implied measures and future stock returns over different horizons. When the traders in options market have informational advantage of some news (positive or negative) about a specific stock, all option-implied measures will include this information asymmetry and its effect will dominate other underlying predicators (dividend yield or skewness). In this section, we further look into this hypothesis and measure the influence of this information asymmetry using a proxy calculated from analyst forecast data.

When some traders receive private information about specific stocks, options market is naturally a good place to exploit the private information since it provides higher leverage and it is easier to hide their private information. Further, if the analyst forecast dispersion is high, which means public information is noisy about the firm's future earnings, it will be easier for these traders to hide their intension and profit from the informational advantage. Thus, we develop a hypothesis that the higher the analyst forecast dispersion, the stronger the relationship between information asymmetry and future stock returns.

For the analyst forecast dispersion measure, we use the standard deviation of analyst forecasts of quarterly earnings scaled by stock price to eliminate firm-level differences. Table 2.14 reports the result of interaction term between option-implied measures and analyst forecast dispersion. We run the following panel regression with each of forecasting horizon $h = 1, 2, 3, ..., 12$:

$$r_{i,t+h} = \alpha_i + \beta X_{i,t} + \gamma DISP_{i,t} + \delta\ X_{i,t} \cdot DISP_{i,t} + e_{i,t+1}$$

where $r_{i,t+h}$ is the h-months ahead non-overlapping monthly stock return (%), α_i is the firm fixed effect, and $X_{i,t}$ is either implied dividend yield IDY, volatility skew VOLSKEW (the slope of volatility smirk) or model-free risk-neutral skewness RNSKEW. The interaction term of option-implied measure and analyst forecast dispersion (DISP) is included in each regression. The table shows that the coefficient on the interaction term between IDY and $DISP$ is significantly negative for first three month ahead non-overlapping returns with t-statistics ranging from -7 to -2. This is a break down of Table 2.10 where we report the coefficient of IDY changes from negatively significant (opposing theory) to positive after a few months. From Table 2.14, it is observed that for higher analyst forecast dispersion stocks, this negative relationship between IDY and future stock returns within three months is stronger. Similar patterns are shown for skewness measures volatility skew (VOLSKEW) and model-free risk-neutral skewness (RNSKEW) as a predictor variable. The coefficient of interaction term between RNSKEW and DISP are positively significant for first few months, which drives the overall coefficient of RNSKEW to be positive (opposing theory) for future returns within first few months reported in Table 2.11. The effect of analyst forecast dispersion on VOLSKEW is, however, not significant especially for first month return. But for 2 to 4 month ahead returns, the sign is consistent with the hypothesis that higher dispersion stocks tend to have a stronger and opposite to theory relationship between skewness and future stock returns.

2.5 Conclusion

As Fama and French (1988) and Campbell and Shiller (1988) explain, the dividend yield should be positively related to the future stock returns. We construct the option-implied measure of dividend yield from equity options at firm-level and find that it is negatively related to the subsequent monthly stock returns, as opposed to the market level evidence and what theory suggests. This puzzling empirical finding is mainly driven by the deviations from put-call parity caused by information asymmetry between traders in options and equity markets. Our panel data analysis reveals that the normal relationship between the dividend yield and the future returns recovers after information asymmetry dissipates within a few months. The biggest problem with the cross-sectional analysis is that the cross-sectional variations in expected returns can be driven by unidentified risk factors. We include the firm fixed effects to circumvent the issue of unidentified risk factors. We further investigate whether the existence of information asymmetry contaminates ex-ante option-implied skewness measures as well, which explains why existing literature finds both positive and negative relationship between option-implied skewness and expected returns in the cross-section. Finally, we reconcile such mixed evidence by showing that the normal negative relationship between option-implied skewness and expected returns appears regardless of choice of option-implied skewness measures after the false positive relationship due to information asymmetry vanishes in the panel data analysis. To conclude, the issues in option-implied measures and cross-sectional analysis cause such mixed evidence on the relationship between option-implied skewness and the expected return, and we reconcile the mixed evidence using a panel analysis with varying forecasting horizons.

Table 2.1 Summary Statistics

The table presents the summary statistics of monthly firm-level option-implied measures and control variables used in the paper. IDY is the ex-ante option-implied dividend yield from put-call parity. log DP^c is the corrected dividend-price ratio from Golez (2014). RNSKEW is the model-free risk-neutral skewness from Bakshi et al. (2003). VOLSKEW is volatility skew (VOLSKEW), which is the slope of volatility smirk, from Xing et al. (2010). HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and LEV are firm market capitalization, the book-to-market ratio, and the leverage, respectively. The sample period is from January 1996 to December 2013.

Variable	Mean	Percentile				
		5%	25%	50%	75%	95%
IDY	0.022	-0.020	0.004	0.014	0.031	0.089
log DP^c	-3.247	-4.971	-3.799	-3.226	-2.613	-1.611
RNSKEW	-0.428	-0.909	-0.568	-0.398	-0.253	-0.060
VOLSKEW	4.084	-0.181	2.046	3.217	4.985	10.872
HSKEW	0.171	-1.291	-0.303	0.155	0.640	1.707
HVOL	0.099	0.032	0.055	0.081	0.122	0.225
log Size	0.099	0.032	0.055	0.081	0.122	0.225
log BM	-1.077	-2.871	-1.672	-1.054	-0.483	0.561
LEV	0.320	0.020	0.100	0.251	0.480	0.870

Table 2.2 Correlation Matrix

The table presents the correlation matrix of monthly firm-level option-implied measures and control variables used in the paper. IDY is the ex-ante option-implied dividend yield from put-call parity. log DP^c is the corrected dividend-price ratio from Golez (2014). RNSKEW is the model-free risk-neutral skewness from Bakshi et al. (2003). VOLSKEW is volatility skew (VOLSKEW), which is the slope of volatility smirk, from Xing et al. (2010). HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and LEV are firm market capitalization, the book-to-market ratio, and the leverage, respectively. The sample period is from January 1996 to December 2013.

	IDY	DP^c	RN−SKEW	VOL−SKEW	HSKEW	HVOL	log Size	log BM	LEV
IDY	1								
log DP^c	0.726	1							
RNSKEW	-0.175	-0.134	1						
VOLSKEW	0.544	0.384	-0.517	1					
HSKEW	0.010	0.010	0.006	0.010	1				
HVOL	0.249	0.197	0.021	0.438	0.031	1			
log Size	-0.231	-0.136	-0.145	-0.238	0.001	-0.353	1		
log BM	0.096	0.054	-0.049	0.113	-0.002	0.136	-0.242	1	
LEV	0.142	0.139	-0.135	0.147	-0.003	0.114	-0.097	0.665	1

Table 2.3 Implied Dividend Yield and Firm Characteristics

The table presents the average firm-characteristics of each decile portfolio sorted by the implied dividend yield every month. IDY is the ex-ante option-implied dividend yield from put-call parity. log DP^c is the corrected dividend-price ratio from Golez (2014). RNSKEW is the model-free risk-neutral skewness from Bakshi et al. (2003). HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and LEV are firm market capitalization, the book-to-market ratio, and the leverage, respectively. HKURT is the monthly historical return kurtosis calculated using daily returns. Sharpe Ratio is calculated using monthly return and volatility. The sample period is from January 1996 to December 2013.

	Low 1	2	3	4	5	6	7	8	9	High 10	High-Low (t-stat.)
log Size	6.45	7.22	7.71	7.78	7.82	7.89	7.90	7.88	7.61	7.04	0.60
											(26.87)
log BM	-1.04	-1.15	-1.28	-1.23	-1.16	-1.10	-1.03	-0.96	-0.89	-0.93	0.11
											(4.16)
LEV	0.26	0.26	0.25	0.27	0.30	0.32	0.35	0.37	0.40	0.38	0.12
											(22.86)
HVOL	3.53	3.18	3.03	2.97	2.85	2.73	2.65	2.58	2.65	2.99	-0.54
											(-11.69)
HSKEW	0.27	0.21	0.19	0.18	0.17	0.17	0.16	0.15	0.16	0.15	-0.12
											(-11.57)
HKURT	1.63	1.42	1.29	1.23	1.16	1.14	1.15	1.13	1.13	1.28	-0.34
											(-10.22)
Sharpe Ratio	0.72	0.57	0.52	0.56	0.57	0.61	0.62	0.63	0.68	0.62	-0.10
											(-1.63)

Table 2.4 Average Returns of Decile Portfolios Sorted by the Implied Dividend Yield

The table presents raw and risk-adjusted average returns of five equal-weighted quintile portfolios sorted by the ex-ante option-implied dividend yield. Panel A shows one-month holding period returns and risk-adjusted alphas. Avg. return is the raw sample average of returns. CAPM α is the CAPM alpha of each portfolio return. FF3 α is the Fama-French three-factor model alpha of each portfolio return. Carhart4 α is the Carhart four-factor model alpha of each portfolio return. The last two columns show the difference between the top and the bottom quintile portfolio returns and their t-statistics computed using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. Panel B shows longer holding period average raw returns. The sample period is from January 1996 to December 2013.

Panel A: 1-month Holding Period Return (%)

	Low 1	2	3	4	High 5	H-L	t(H-L)
Avg. return	0.97	0.89	0.86	0.69	0.39	-0.58	-2.34
CAPM α	-0.15	-0.12	-0.12	-0.27	-0.59	-0.44	-1.72
FF3 α	0.07	0.01	-0.04	-0.22	-0.53	-0.59	-2.72
Carhart4 α	-0.08	-0.07	-0.12	-0.30	-0.68	-0.60	-2.78

Panel B: Longer Holding Period Raw Returns (%)

	Low 1	2	3	4	High 5	H-L	t(H-L)
3-month	2.40	2.58	2.65	2.44	1.47	-0.93	-1.91
6-month	4.12	4.82	5.36	4.96	3.57	-0.54	-0.71
12-month	9.95	10.64	10.74	10.62	9.88	-0.07	-0.06

Table 2.5 Average Returns of Deciles Portfolios Sorted by Term Structure of the Implied Dividend Yield

The table presents raw and risk-adjusted average returns of equal-weighted quintile portfolios. Panel A shows results sorted by implied dividend yield (IDY) calculated from opitons with all maturities. Panel B presents results sorted by short-term IDY, calculated from options with less than 60 days maturity. Panel C includes results sorted by long-term IDY computed from options with longer than 6 month maturity. Panel D reports results sorted by the difference between long-term IDY and short-term IDY. Avg. return is the raw sample average of returns. CAPM α is the CAPM alpha. FF3 α is the Fama-French three-factor model alpha. Carhart4 α is the Carhart four-factor model alpha. The last two rows in each panel show the difference between the top and the bottom quintile portfolio returns and their t-statistics computed using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Measure		Rank	Avg. return	CAPM α	FF3 α	Carhart4 α
Panel A: Decile Returns Sorted by IDY (%)						
IDY	Low	1	0.97	-0.15	0.07	-0.08
		2	0.89	-0.12	0.01	-0.07
		3	0.86	-0.12	-0.04	-0.12
		4	0.69	-0.27	-0.22	-0.30
	High	5	0.39	-0.59	-0.53	-0.68
		H-L	-0.58	-0.44	-0.59	-0.60
		T	(-2.34)	(-1.72)	(-2.72)	(-2.78)
Panel B: Decile Returns Sorted by short-term IDY (%)						
IDY_{short}	Low	1	0.92	-0.18	0.07	-0.13
		2	0.92	-0.06	0.03	-0.02
		3	0.88	-0.10	-0.05	-0.08
		4	0.67	-0.31	-0.23	-0.33
	High	5	0.41	-0.59	-0.54	-0.70
		H-L	-0.51	-0.42	-0.61	-0.57
		T	(-2.56)	(-1.98)	(-3.18)	(-3.01)
Panel C: Decile Returns Sorted by long-term IDY (%)						
IDY_{long}	Low	1	1.07	-0.05	0.10	-0.02
		2	0.79	-0.28	-0.13	-0.22
		3	0.81	-0.21	-0.08	-0.20
		4	0.70	-0.22	-0.18	-0.26
	High	5	0.43	-0.48	-0.42	-0.56
		H-L	-0.64	-0.43	-0.52	-0.55
		T	(-2.19)	(-1.55)	(-2.28)	(-2.44)
Panel D: Decile Returns Sorted by term structure of IDY (%)						
$IDY_{long} - IDY_{short}$	Low	1	0.52	-0.52	-0.45	-0.60
		2	0.79	-0.23	-0.14	-0.23
		3	0.84	-0.17	-0.10	-0.12
		4	0.80	-0.17	-0.05	-0.14
	High	5	0.85	-0.16	0.03	-0.16
		H-L	0.33	0.35	0.48	0.44
		T	(2.03)	(2.05)	(2.89)	(2.68)

Table 2.6 Fama Macbeth Regression on the Implied Divided Yield and the Corrected Dividend-Price Ratio

The table presents Fama-Macbeth regressions with monthly stock returns. For each month t, we run the following cross-sectional regression:

$$r_{i,t+1} = \alpha_t + \beta_t X_{i,t} + \gamma_t^\top Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+1}$ is the monthly stock return (%), $X_{i,t}$ is either the implied dividend yield $IDY_{i,t}$ or the corrected dividend-price ratio $\log DP^c_{i,t}$, and $Z_{i,t}$ is control variables. HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and LEV are firm market capitalization, the book-to-market ratio, and the leverage, respectively. Then we compute the time-average of β_t and γ_t to find the point estimates and report their t-statistics in parentheses using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The last column shows the time-average of adjusted R^2 of each cross-sectional regression. The sample period is from January 1996 to December 2013.

Model	IDY	log DP^c	log Size	log BM	LEV	HVOL	HSKEW	Adj. R^2 (%)
I	-6.060							0.3
	(-4.73)							
II	-5.894		-0.384	0.135	0.322	2.812	-0.011	5.8
	(-5.18)		(-4.89)	(1.50)	(0.69)	(1.01)	(-0.22)	
III		-0.035						0.5
		(-0.48)						
IV		-0.122	-0.140	-0.019	0.229	-1.195	-0.089	6.5
		(-1.59)	(-1.92)	(-0.19)	(1.87)	(-0.35)	(-1.11)	

Table 2.7 Fama Macbeth Regression on Risk-Neutral Skewness and the Slope of Volatility Smirk.

The table presents Fama-Macbeth regressions with monthly stock returns. For each month t, we run the following cross-sectional regression:

$$r_{i,t+1} = \alpha_t + \beta_t X_{i,t} + \gamma_t^{\top} Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+1}$ is the monthly stock return (%), $X_{i,t}$ is either model-free risk-neutral skewness $RNSKEW$ or volatility skew $VOLSKEW$ (the slope of volatility smirk), and $Z_{i,t}$ is control variables. HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and LEV are firm market capitalization, the book-to-market ratio, and the leverage, respectively. Then we compute the time-average of β_t and γ_t to find the point estimates and report their t-statistics in parentheses using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The last column shows the time-average of adjusted R^2 of each cross-sectional regression. The sample period is from January 1996 to December 2013.

Model	RN−SKEW	VOL−SKEW	log Size	log BM	LEV	HVOL	HSKEW	Adj. R^2 (%)
I	0.915							0.7
	(1.77)							
II	1.360		0.016	-0.036	0.083	-2.232	0.067	7.8
	(2.99)		(0.17)	(-0.31)	(0.75)	(-0.65)	(0.81)	
III		-0.060						0.8
		(-2.66)						
IV		-0.062	-0.038	-0.006	0.037	-1.245	0.004	7.2
		(-2.87)	(-0.49)	(-0.07)	(0.34)	(-0.42)	(0.06)	

Table 2.8 Time-series Predictability of Market Returns

The table presents time-series predictability of market returns. Each column shows the result of the time-series predictive regression:

$$r_{m,t+1} = \alpha + \beta X_t + e_{t+1}$$

where $r_{m,t+1}$ is the monthly S&P500 index return (%) and X_t is a single predictor variable. IDY is the ex-ante option-implied dividend yield from put-call parity. log DP^c is the corrected dividend-price ratio from Golez (2014). log $\mathrm{DP}^{c,ts}$ is the corrected dividend-price ratio using information in the term structure of the expected dividend growth from Bilson et al. (2015). RNSKEW is the model-free risk-neutral skewness from Bakshi et al. (2003). VOLSKEW is volatility skew, which is the slope of volatility smirk, from Xing et al. (2010). All these predictor variables are computed using the index options. Numbers in parentheses are their t-statistics using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Single Predictor	IDY	log DP^c	log $\mathrm{DP}^{c,ts}$	RNSKEW	VOLSKEW
Coefficient β	0.0057	0.0043	0.0080	-0.0001	0.7295
(t-statistics)	(1.56)	(1.24)	(2.31)	(-0.29)	(1.67)
In-sample R^2 (%)	1.56	0.89	2.60	0.01	1.59
Pseudo Out-of-sample R^2 (%)	0.17	-0.35	2.46	-2.80	0.19

Table 2.9 Panel Regression of Monthly Stock Returns on the Lagged Implied Dividend Yield

The table presents the pooled time-series predictability of individual stock returns. We run the following panel regression:

$$r_{i,t+1} = \alpha_i + \beta\ IDY_{i,t} + \gamma^\top Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+1}$ is the monthly stock return (%), α_i is the firm fixed effect, IDY is the ex-ante option-implied dividend yield from put-call parity, and $Z_{i,t}$ is control variables. HSKEW and HVOL are the monthly historical return skewness and volatility calculated using daily returns, respectively. log Size, log BM, and LEV are firm market capitalization, the book-to-market ratio, and the leverage, respectively. All explanatory variables are standardized by their sample means and standard deviations. Numbers in parentheses are t-statistics calculated using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Model	IDY	log Size	log BM	LEV	HVOL	HSKEW	Fixed Effect	Adj R^2 (%)
I	-0.32						Yes	0.03
	(-8.55)							
II	-0.44	-5.56	0.44				Yes	2.06
	(-11.02)	(-56.56)	(7.04)					
III	-0.41	-5.69	0.28	0.41	-0.41	0.04	Yes	2.12
	(-10.22)	(-54.41)	(3.91)	(3.69)	(-10.02)	(1.36)		
IV	-0.33						No	0.04
	(-10.05)							
V	-0.33	-0.44	0.34				No	0.19
	(-9.38)	(-13.36)	(10.35)					
VI	-0.34	-0.52	0.21	0.17	-0.12	0.05	No	0.21
	(-9.53)	(-14.20)	(4.73)	(4.03)	(-3.46)	(1.56)		

Table 2.10 Panel Regression of Monthly Stock Returns with Varying Forecasting Horizon

The table presents the pooled time-series predictability of individual stock returns with different forecasting horizons up to twelve months. We run the following panel regression with each of forecasting horizon $h = 1, 2, 3, ..., 12$:

$$r_{i,t+h} = \alpha_i + \beta \, X_{i,t} + e_{i,t+1}$$

where $r_{i,t+h}$ is the h-months ahead non-overlapping monthly stock return (%), α_i is the firm fixed effect, and $X_{i,t}$ is either implied dividend yield $IDY_{i,t}$ or the corrected dividend-price ratio log $DP^c_{i,t}$. All explanatory variables are standardized by their sample means and standard deviations. Under the coefficient estimates, we report t-statistics calculated using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Panel A: Regression on IDY

h^{th} month	1	2	3	4	5	6	7	8	9	10	11	12
1. Without Firm Fixed Effects												
β	-0.33	-0.18	-0.06	0.03	0.05	0.11	0.03	0.02	0.04	0.01	0.07	0.13
t-stat.	-10.05	-5.09	-1.77	0.83	1.45	2.82	1.38	0.53	1.17	0.26	1.67	3.09
2. With Firm Fixed Effects												
β	-0.32	-0.13	-0.02	0.13	0.18	0.26	0.12	0.13	0.16	0.10	0.15	0.25
t-stat.	-8.55	-3.24	-0.48	3.09	4.10	5.84	2.64	2.84	3.62	2.15	3.26	5.07

Panel B: Regression on log DP^c

h^{th} month	1	2	3	4	5	6	7	8	9	10	11	12
1. With Firm Fixed Effects												
β	0.05	-0.13	0.21	0.24	0.20	0.23	0.24	0.31	0.29	-0.04	0.17	0.30
t-stat.	0.76	-1.73	2.81	3.05	2.56	3.12	3.02	3.82	3.80	-0.50	1.92	3.75

Table 2.11 Panel Regression of Monthly Stock Returns on the Lagged Option-implied Skewness Measures with Varying Forecasting Horizon

The table presents the pooled time-series predictability of individual stock returns with different forecasting horizons up to twelve months. We run the following panel regression with each of forecasting horizon $h = 1, 2, 3, ..., 12$:

$$r_{i,t+h} = \alpha_i + \beta\, X_{i,t} + \gamma^\top Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+h}$ is the h-months ahead non-overlapping monthly stock return (%), α_i is the firm fixed effect, $X_{i,t}$ is either model-free risk-neutral skewness $RNSKEW$ or volatility skew $VOLSKEW$ (the slope of volatility smirk), and $Z_{i,t}$ is control variables: HSKEW, HVOL, log Size, log BM, and LEV as defined in Table 2.1. All explanatory variables are standardized by their sample means and standard deviations. Under the coefficient estimates, we report t-statistics calculated using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Panel A: Regression on VOLSKEW												
h^{th} month	1	2	3	4	5	6	7	8	9	10	11	12
1. With Firm Fixed Effects												
β	-0.39	-0.11	0.02	0.21	0.36	0.55	0.37	0.28	0.19	0.33	0.53	0.57
t-stat.	-7.39	-1.99	0.43	3.57	5.99	8.92	6.08	4.43	2.93	5.16	8.07	8.61
2. With Firm Fixed Effects, log Size, and log BM												
β	-0.62	-0.33	-0.21	0.01	0.17	0.39	0.21	0.08	-0.11	0.14	0.39	0.37
t-stat.	-10.34	-5.52	-3.47	0.08	2.75	5.96	3.26	1.33	-0.17	2.08	5.67	5.21
3. With Firm Fixed Effects and All Control Variables												
β	-0.42	-0.25	-0.23	0.13	0.07	0.2	0.21	0.08	-0.11	0.13	0.37	0.23
t-stat.	-6.97	-5.47	-3.58	-0.23	1.04	2.94	3.03	1.18	-1.51	1.87	5.14	3.04
Panel B: Regression on RNSKEW												
h^{th} month	1	2	3	4	5	6	7	8	9	10	11	12
1. With Firm Fixed Effects												
β	0.55	0.23	0.09	0.10	0.06	0.07	-0.03	0.01	0.29	0.07	-0.07	0.08
t-stat.	7.62	3.23	1.23	1.23	1.37	0.87	-0.45	0.11	3.78	0.93	-0.94	1.05
2. With Firm Fixed Effects, log Size, and log BM												
β	0.19	-0.09	-0.28	-0.28	-0.26	-0.34	-0.40	-0.24	0.04	-0.23	-0.34	-0.25
t-stat.	2.38	-1.23	-3.48	-3.39	-3.21	-4.07	-4.75	-2.86	0.42	-2.65	-3.97	-2.92
3. With Firm Fixed Effects and All Control Variables												
β	0.16	-0.07	-0.23	-0.22	-0.15	-0.22	-0.33	-0.22	0.07	-0.21	-0.33	-0.24
t-stat.	3.04	-0.85	-2.84	-2.68	-1.89	-2.69	-3.92	-2.60	0.82	-2.44	-3.84	-2.73

Table 2.12 Panel Regression of Stock Returns with Varying Holding Period

The table presents the pooled time-series predictability of individual stock returns with different holding periods up to twelve months. We run the following panel regression with each of holding periods n = 1, 3, 6, and 12:

$$r_{i,t+n} = \alpha_i + \beta\, X_{i,t} + \gamma^\top Z_{i,t} + e_{i,t+1}$$

where $r_{i,t+n}$ is the overlapping n-month holding period stock return (%), α_i is the firm fixed effect, $X_{i,t}$ is implied dividend yield $IDY_{i,t}$, the corrected dividend-price ratio $\log DP^c_{i,t}$, model-free risk-neutral skewness $RNSKEW$, or volatility skew $VOLSKEW$ (the slope of volatility smirk), and $Z_{i,t}$ is control variables: HSKEW, HVOL, log Size, log BM, and LEV as defined in Table 2.1. All explanatory variables are standardized by their sample means and standard deviations. Under the coefficient estimates, we report t-statistics calculated using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Predictor	Control Variables		Holding Period (month)			
			1	3	6	12
IDY	Firm Fixed Effects	β	-0.32	-0.30	0.77	3.05
		t-stat.	-8.55	-3.40	4.81	10.67
log DP^c	Firm Fixed Effects	β	0.05	0.17	0.84	2.30
		t-stat.	0.76	1.21	3.82	6.83
VOLSKEW	Firm Fixed Effects	β	-0.39	-0.34	1.26	4.98
		t-stat.	-7.39	-2.98	6.50	15.30
	Firm Fixed Effects	β	-0.62	-1.05	-0.10	2.28
	+ log Size + log BM	t-stat.	-10.34	-8.50	-0.49	7.16
	Firm Fixed Effects	β	-0.42	-0.88	-0.26	1.73
	+ All Control Variables	t-stat.	-6.97	-7.06	-1.29	5.46
RNSKEW	Firm Fixed Effects	β	0.55	0.86	0.99	0.77
		t-stat.	7.62	5.51	3.92	1.92
	Firm Fixed Effects	β	0.19	-0.26	-1.42	-3.79
	+ log Size + log BM	t-stat.	2.38	-1.52	-5.26	-9.54
	Firm Fixed Effects	β	0.16	-0.21	-1.09	-3.28
	+ All Control Variables	t-stat.	3.04	-1.21	-3.99	-8.23

Table 2.13 Panel Regression of Dividend Growth with Varying Horizon

The table presents the pooled time-series predictability of individual dividend growth with different dividend horizons up to twelve months. We run the following panel regression with each of dividend horizons n = 1, 3, 6, and 12:

$$\log D_{i,t+n}^{(12)} - \log D_{i,t}^{(12)} = \alpha_i + \beta X_{i,t} + e_{i,t+1}$$

where $D_{i,t}$ is the overlapping twelve-month trailing sum of dividends, α_i is the firm fixed effect, and $X_{i,t}$ is either option-implied dividend growth $IDG_{i,t}$ or option-implied dividend yield $IDY_{i,t}$. Under the coefficient estimates, we report t-statistics calculated using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

Predictor	Firm Fixed Effects		Dividend Growth Horizon (month)			
			1	3	6	12
IDG	Yes	β	0.696	0.108	0.198	0.379
		t-stat.	2.55	3.57	6.93	7.60
	No	β	0.078	0.104	0.165	0.351
		t-stat.	3.89	4.63	7.28	7.93
IDY	Yes	β	0.037	0.035	0.238	-0.114
		t-stat.	2.93	1.23	0.49	-1.57
	No	β	0.085	-0.028	0.093	-0.272
		t-stat.	0.84	-1.08	-1.94	-3.29

Table 2.14 Panel Regression of Monthly Stock Returns with Interaction of Analyst Forecast Dispersion

The table presents the pooled time-series predictability of individual stock returns with option-implied measure and its interaction with analyst forecast dispersion. We run the following panel regression with each of forecasting horizon $h = 1, 2, 3, ..., 12$:

$$r_{i,t+h} = \alpha_i + \beta X_{i,t} + \gamma DISP_{i,t} + \delta \; X_{i,t} \cdot DISP_{i,t} + e_{i,t+1}$$

where $r_{i,t+h}$ is the h-months ahead non-overlapping monthly stock return (%), α_i is the firm fixed effect, and $X_{i,t}$ is either implied dividend yield IDY, volatility skew $VOLSKEW$ (the slope of volatility smirk) or model-free risk-neutral skewness $RNSKEW$. The interaction term of option-implied measure and analyst forecast dispersion $DISP$ is included in each regression. All explanatory variables are standardized by their sample means and standard deviations. Under the coefficient estimates, we report t-statistics calculated using Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. The sample period is from January 1996 to December 2013.

h^{th}-mon	1	2	3	4	5	$\cdots$	10	11	12
1. $X_{i,t} = IDY$ with Firm Fixed Effects									
δ	-0.09	-0.11	-0.06	0.07	0.02	$\cdots$	-0.99	-0.28	0.03
t-stat.	-7.27	-7.52	-2.84	2.16	0.51	$\cdots$	-0.99	-0.62	0.59
2. $X_{i,t} = VOLSKEW$ with Firm Fixed Effects									
δ	0.01	-0.03	-0.02	-0.07	-0.04	$\cdots$	-0.06	-0.07	-0.17
t-stat.	0.53	-2.70	-2.17	-2.91	-1.51	$\cdots$	-1.18	-1.29	-3.56
3. $X_{i,t} = RNSKEW$ With Firm Fixed Effects									
δ	0.33	0.35	0.65	0.35	-0.07	$\cdots$	0.32	-0.37	0.65
t-stat.	3.42	3.32	5.81	3.08	-0.67	$\cdots$	2.13	-2.48	4.44

Figure 2.1 Evolution of the Cross-sectional Distribution of Implied Dividend Yield

The plot shows the cross-sectional distribution of monthly option-implied dividend yield. Each line represents the cross-sectional mean, median, 10^{th} percentile, and 90^{th} percentile calculated every month, respectively.

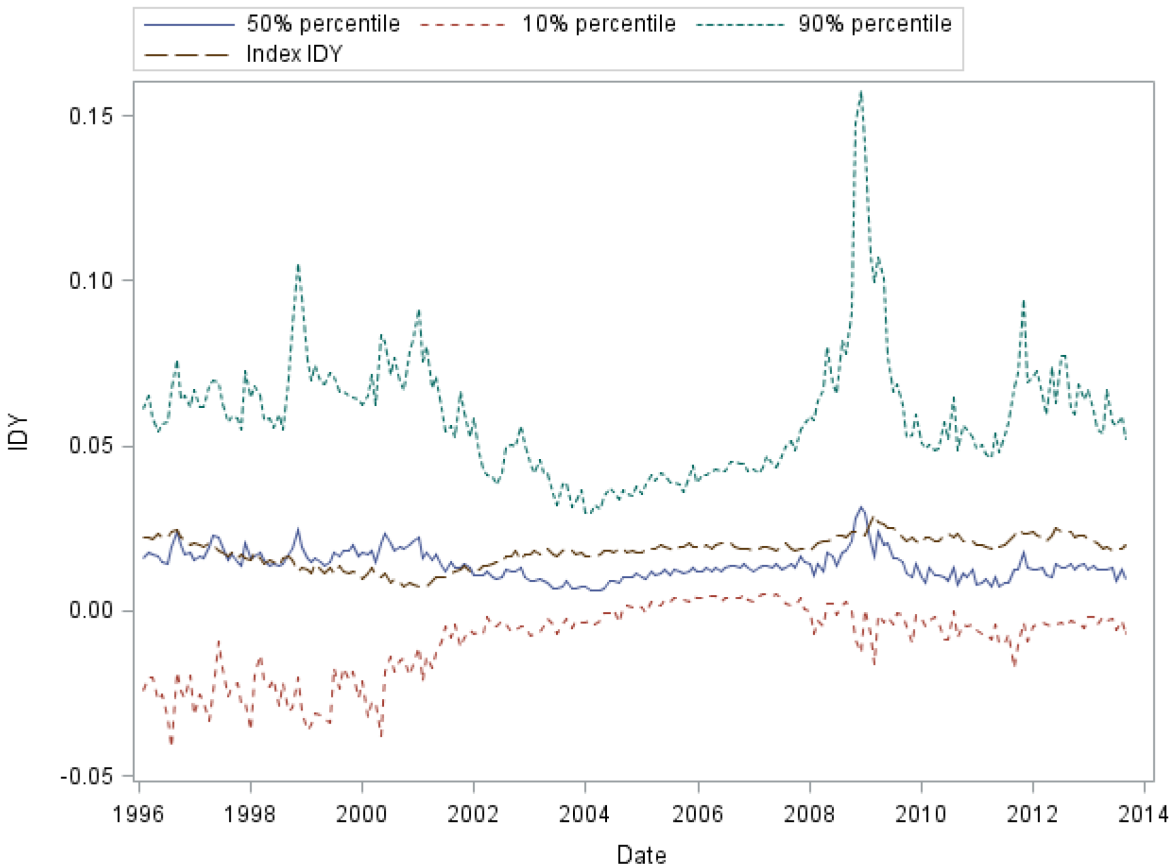
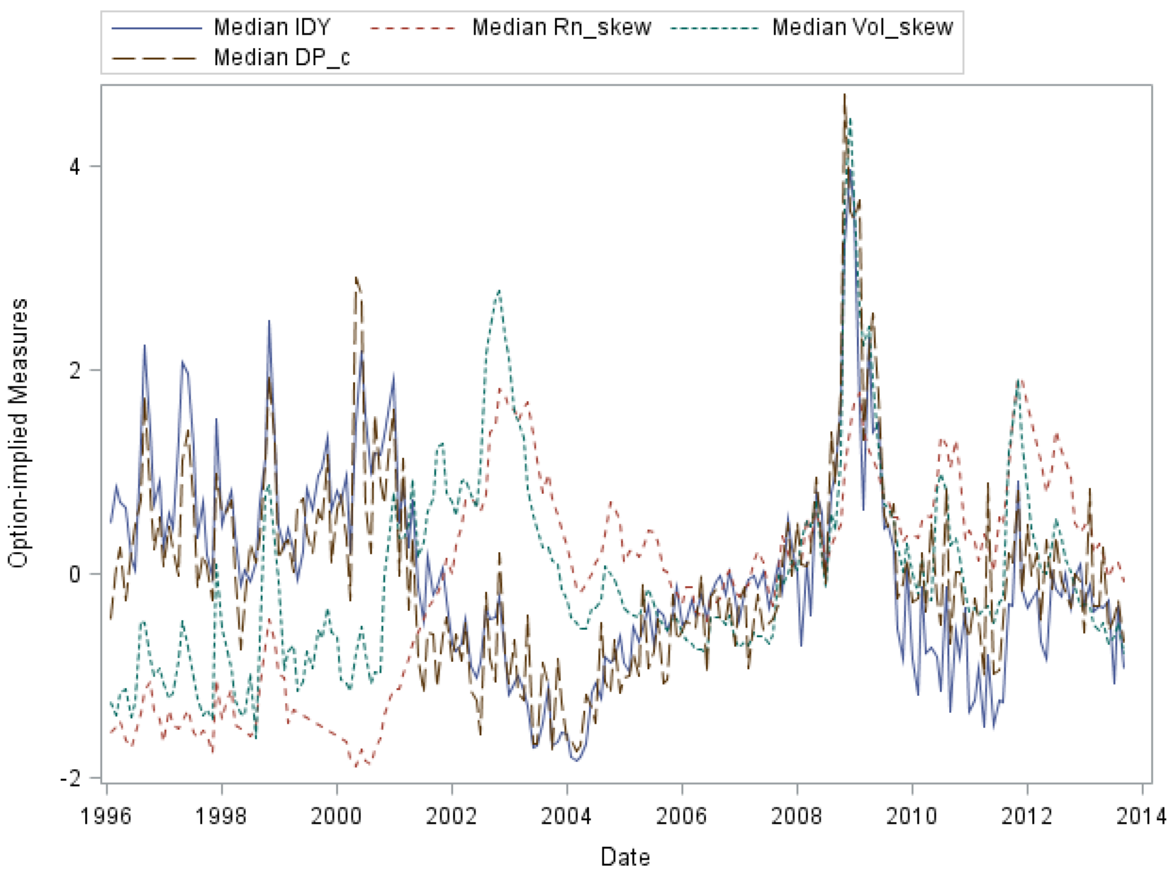


Figure 2.2 Cross-sectional Medians of Option-implied Predictors for Stock Returns

The plot shows the time-series of the cross-sectional median of four monthly option-implied predictors for stock returns: IDY, log DP^c, RNSKEW, and VOLSKEW where IDY is the implied dividend yield, log DP^c is the corrected dividend-price ratio, RNSKEW is the model-free risk-neutral skewness, and VOLSKEW is the volatility skew (the slope of volatility smirk). All variables are standardized by their sample means and standard deviations for better visualization. We draw −RNSKEW instead of RNSKEW so that all four option-implied measures in the plot are associated with expected returns in the same way. The sample period is from January 1996 to December 2013.



Bibliography

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez (2013). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics, Forthcoming*.

Bakshi, G., N. Kapadia, and D. Madan (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies 16*(1), 101–143.

Bakshi, G. and D. Madan (2000). Spanning and derivative-security valuation. *Journal of Financial Economics 55*(2), 205–238.

Barberis, N. and M. Huang (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review 98*(5), 2066–2100.

Bilson, J. F., S. B. Kang, and H. Luo (2015). The term structure of implied dividend yields and expected returns. *Economics Letters*.

Binsbergen, J. H. v. and R. S. J. Koijen (2010). Predictive Regressions: A Present-Value Approach. *Journal of Finance 65*(4), 1439–1471.

Brunnermeier, M. K., C. Gollier, and J. A. Parker (2007). Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review 97*(2), 159–165.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*.

Cochrane, J. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Conrad, J., R. F. Dittmar, and E. Ghysels (2013). Ex Ante Skewness and Expected Stock Returns. *Journal of Finance 68*(1), 85–124.

Cremers, M. and D. Weinbaum (2010). Deviations from put-call parity and stock return predictability. *Journal of Financial and Quantitative Analysis 45*(02), 335–367.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*.

Ferson, W., S. Sarkissian, and T. Simin (2003). Is stock return predictability spurious. *Journal of Investment Management 1*(3), 1–10.

Golez, B. (2014). Expected Returns and Dividend Growth Rates Implied by Derivative Markets. *The Review of Financial Studies 27*(3), 790–822.

Kalay, A., O. Karakaş, and S. Pant (2014). The market value of corporate votes: Theory and evidence from option prices. *The Journal of Finance 69*(3), 1235–1271.

Ofek, E., M. Richardson, and R. F. Whitelaw (2004). Limited arbitrage and short sales restrictions: Evidence from the options markets. *Journal of Financial Economics 74*(2), 305–342.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Welch, I. and A. Goyal (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies 21*(4), 1455–1508.

Xing, Y., X. Zhang, and R. Zhao (2010). What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? *Journal of Financial and Quantitative Analysis 45*(03), 641–662.

Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics 99*(1), 216–233.

Chapter 3

Systematic Risk and the Cross-Section of Credit Default Swap Returns

3.1 Introduction

The basic insight from the standard unconditional capital asset pricing model (CAPM) (Sharpe (1964), Lintner (1965)) states that the cross-sectional asset returns should be explained by sensitivities of asset returns to the return of the aggregate market. There are a large number of papers exploring this relationship between systematic risk and return in the equity market. However, it has been shown (see Fama and French 1992, 1993) that, market beta, the systematic risk measure described in CAPM model, has no ability to explain the cross-section of stock returns. Rather, individual firm characteristics such as size, book to market ratios, and momentum are found to be important and more robust predictors of the cross-sectional variation in stock returns. These findings create some controversy over whether firm characteristics or factor loadings are more relevant.

A natural extension of the CAPM model is to consider the asymmetric treatment of risk by investors. Ang, Chen, and Xing (2006) define upside and downside risks by computing market beta conditioning on market returns being larger or smaller than average market returns. They find that firms with high downside beta have higher contemporaneous stock returns while firms with high relative upside beta (relative to CAPM beta) have lower contemporaneous stock returns. The underlying assumption is that investors have asymmetric preferences for losses versus gains, as suggested in Markowitz (1952) and Bawa and Lindenberg (1977). When the market goes down, loss-averse investors have a larger increase in marginal utility because of their asymmetric utility function (the disappointment utility function presented in Gul (1991) satisfies this assumption and could be used to motivate the role of downside beta in asset prices). Assets that are more sensitive to market downturns than market uptrends are undesirable for loss-averse investors as these assets will provide low payoffs exactly when the wealth of the investors are decreasing. Thus, stocks with higher covariation with the market during downside market movements should earn higher rewards for investors to hold in equilibrium.

There exist several studies in the literature examining the relationship between conditional market risk and expected returns in different asset classes or different markets (see, for example, Lettau, Maggiori, and Weber (2014) and Atilgan, Bali, Demirtas, and Gunaydin (2015)). However, no earlier papers study the relationship between (un)conditional market risk and the cross-section of credit default swap (CDS) returns. In the current paper, we first test whether the systematic market risk is priced in the cross-section of CDS returns and then examine the relation between downside (upside) market risk and CDS returns. We contribute to a growing literature on identifying determinants of the cross-section of returns for bond (see, e.g., Kwan (1996), Gebhardt, Hvidkjaer, and Swaminathan (2005), Bessembinder et al. (2009), Lin, Wang, and Wu (2011), Jostova, Nikolova, Philipov, and Stahel (2013), Chordia et al. (2017), Choi and Kim (2017) and Bai, Bali and Wen (2017)) and firm-level CDS contracts (see, e.g., Berndt and Obreja (2010), Bongaerts, de Jong, and Driessen (2011), Junge and Trolle (2015) and Lee, Naranjo, and Sirmans (2014)). We contribute to the literature by studying the systematic risk as determinants of the cross-section of CDS returns and rely on specific features of CDS contracts and decomposition of systematic risk to show which part of the systematic beta risk is priced in the cross-section while controlling for firm characteristics.

Credit default swaps are single-name over-the-counter derivatives that provide default insurance. The payoff to the protection buyer covers losses up to the notional value in the event of a default by the reference entity. Default events are triggered by either bankruptcy, failure to pay, or a debt-restructuring event. The buyer of the protection pays a quarterly premium, quoted as an annualized percentage of the notional value, and in return receives the payoff from the seller of protection if a credit event occurs. Structural models following Merton (1974) generate a simple insight for the relation between CDS returns and equity returns: risk premia on the equity market and CDS (debt) market are related as all claims on the same underlying assets should earn the same amount of compensation per unit of risk. Empirical studies along this line successfully confirm the model-suggested relationship between stock returns and bond returns (Schaefer and Strebulaev (2008)) and the relationship between stock returns and credit-implied risk premia (Friewald, Wagner, and Zechner (2014)). Though most earlier studies (e.g., Huang and Huang (2003)) find that structural models are not sufficient to generate a high level of credit spreads as observed in data, simplest structural model is capable of describing the association between corporate bond (CDS) returns and equity returns. This paper contributes to the literature by constructing theoretic relationship (implied by Merton (1974) structural model and CAPM model) between CDS returns and firm's market risk (CAPM beta) estimated from stock returns and empirically confirm the positive relation between systematic risk and CDS returns.

The literature has recognized the role of downside risk in asset pricing since Roy (1952) and Markowitz (1952). Ang, Chen, and Xing (2006) build on the idea that investors dislike assets that are more sensitive to market downturns than market uptrends and motivate a downside beta measure relying on the disappointment utility function of Gul (1991). They find that firms that have higher downside beta have higher contemporaneous stock returns. However, they did not find a significant association between ex-ante downside beta and future stock returns unless they exclude high volatility stocks. Lettau, Maggiori, and Weber

(2014) further report that downside-risk-CAPM can jointly explain the cross-sectional returns of currencies, commodities, sovereign bonds, and U.S. stocks. However, Atilgan, Bali, Demirtas, and Gunaydin (2015) find that contrary to the findings in the U.S. equity market, downside beta does not explain the cross-sectional differences in future and contemporaneous returns in an international setting for individual stocks. We contribute to the literature by examining the relation between downside beta and CDS returns as CDS contracts are designed to be more sensitive to downside risk than equity contracts. Although CDS contracts reflect the same firm fundamentals as stocks, they are more sensitive to downside risk because of their specific structure as firms are more likely to default in bad market states and CDS protection sellers are more likely to be responsible for the loss of value during market downside movements. Thus, CDS protection sellers are expected to require a premium for bearing the risk of negative externalities spilling over from other sectors in the market. In this paper, we test and confirm the hypothesis that CDS protection sellers do get higher compensation for the downside risk embedded in the firm.

We focus on returns of CDS contract for two other reasons. First, Merton (1974) implies a clear structural link between stock and CDS returns through the firm's capital structure. Second, Friewald, Wagner, and Zechner (2014) confirms that CDS spreads are consistent in their assessment of the underlying credit risk with stock prices without being affected by other liquidity-related measures. In contrast, earlier literature studying credit risk through corporate bond returns document a "distress risk puzzle" in which expected stock returns are negatively related to credit risk.

Our strategy for finding a premium for bearing unconditional and conditional systematic risk in the cross-section of CDS contracts is as follows. First, we show that CDS protection sellers' earn higher average returns for firms with higher CAPM beta. Second, we show that firms with higher downside beta risk or upside beta risk have larger returns for CDS protection sellers. This result is seemingly inconsistent with protection sellers willing to accept a discount for firms with high upside potential. However, it is shown that CAPM beta, downside beta and upside beta are all correlated by construction, thus, sorting on each of the conditional beta risk measures alone will not rule the impact of CAPM beta on CDS returns. This observation leads us to next step where we explore two methods to disentangle the effects of conditional beta on CDS returns that are unrelated to unconditional beta. We use relative downside (upside) beta measures as in Ang, Chen, and Xing (2006) and also extract the residual part of downside (upside) beta that is orthogonal to CAPM beta. The decomposition of impact from conditional beta risk provides more evidence that is consistent with theory. It is confirmed that CDS protection sellers demand higher returns for firms with larger relative (residual) downside beta and accept a discount for firms with larger relative (residual) upside beta. Finally, we identify the reward for unconditional systematic risk and conditional systematic risk while controlling for other known characteristics that may affect CDS returns in theory. Fama-Macbeth regression results show that higher CDS returns are required by CDS sellers to provide protection for firms with higher CAPM beta, downside beta, and relative downside beta.

The remainder of the paper is organized as follows. Section 3.2 introduces the relation between CDS returns, stock returns and market beta implied by Merton structural model and CAPM model. Section 3.3 describes the data sources and data treatments to obtain CDS returns and various risk measures. Section 3.4 provides empirical results. Section 3.5 concludes.

3.2 Structural Models and CDS Returns

3.2.1 Merton Model

In this section, we use a simple Merton framework to illustrate that returns to a firm's credit instrument (CDS) should be related to returns to its equity. CDS and equity contracts are two claims on the same underlying firm. Thus, in structural models, their returns are intrinsically related where firm value, V, is the driving state variable.

First, consider the equity claim which could be viewed as a European call option on the firm's assets with strike price equal to face value of debt D and maturity T. Using Ito's lemma on value of equity, we obtain that the expected excess return on equity $\mu_E - r$ as a function of excess asset return which is denoted as $\mu - r$:

$$\mu_E - r = \frac{V}{E} E_V(\mu - r), \tag{3.1}$$

where E_V denotes the partial derivative of E with respect to V which is the call option delta with $E_V > 0$.

To derive characteristics of CDS returns, we follow the insight in Berndt and Obreja (2010). Note that a portfolio that combines a long position in a T-year par defaultable bond issued by firm i and a short position in a T-year par riskless bond generates same cash-flows, to a close approximation, as from selling protection on the firm through a T-year CDS contract. Therefore, we have

$$\Delta V_{CDS} = \Delta P_D - \Delta P_{RF}, \tag{3.2}$$

where ΔP_D and ΔP_{RF} denote changes in the value of the risky and risk-free bond. After we divided both side by par, the excess return on defaultable bond is given as

$$\mu_D - r = \Delta V_{CDS}. \tag{3.3}$$

We refer to ΔV_{CDS} as excess CDS return for protection seller, or simply as CDS return and will describe how to calculate it later.

Then, similar to the equity return, CDS return could also be written as a function of asset return

$$\Delta V_{CDS} = \mu_D - r = \frac{V}{D} D_V(\mu - r), \tag{3.4}$$

where D_V denotes the partial derivative of D with respect to V.

Dividing Equation (4) by Equation (1) and considering the fact that $V = D + E$, we have

$$\Delta V_{CDS} = h_E(\mu_E - r), \tag{3.5}$$

where h_E denotes the sensitivity of the return on CDS to the return on equity and $h_E = \frac{E}{D}(\frac{1}{E_V} - 1)$.

3.2.2 CDS Returns and CAPM

Now we suppose the CAPM model holds. The expected excess equity return will be determined by the following equation

$$\mu_E - r = \beta_E(\mu_M - r), \tag{3.6}$$

where μ_M is expected return on market portfolio and β_E is the CAPM beta of the equity. Substitute Equation (3.6) into Equation (3.5), we have

$$\Delta V_{CDS} = \beta_{CDS}(\mu_M - r), \tag{3.7}$$

where CDS beta is given as $h_E \beta_E$.

CDS beta depends on firm characteristics and equity beta, we will use equity beta as a first step in empirical analysis. And we also use Merton model implied elasticity to generate heterogeneous CDS beta to analyze direct relation between CDS returns and CDS beta.

3.3 Data and Empirical Methodologies

3.3.1 Data

Our sample period is from January 2000 to December 2015. All U.S. firms with actively traded stock and single-name CDS contract are included. Total number of firms in the sample is 820. Equity data are obtained from CRSP. CDS data are acquired from Markit. We consider CDS contracts denominated in U.S. dollars with one, three, five, seven and ten years maturity. Markit constructs a composite CDS spread using input from a variety of market makers and requires each daily observation passes a rigorous cleaning test to ensure accuracy and reliability.

Individual stocks data are used to calculate CAPM beta, downside beta and upside beta. We use twelve month of daily data to calculate each risk measure following Ang, Chen, and Xing (2006). It is a choice of balancing between long enough time period to have enough data for downside/upside variation and short period to account for time-varying risk exposures.

3.3.2 Calculating CDS Returns

To construct excess CDS returns for protection sellers, we follow the procedure in Bongaerts, de Jong, and Driessen (2011). Consider an investor at time $t-\Delta t$ who sells protection to the k-th firm using a CDS contract and receive spread at $CDS_{k,t-\Delta t}$ each quarter until default or maturity. Assume this protection seller purchases an offsetting contract at time t with spread $CDS_{k,t}$. This investor will receive $-\frac{1}{4}\Delta CDS_{k,t}$ each quarter until default or maturity. The value of this stream plus adjustment for initial accrued spread during holding period will give the excess holding return

$$\Delta V_{CDS,k,t} = \frac{\Delta t}{4} CDS_{k,t-\Delta t} - \frac{\Delta CDS_{k,t}}{4} \sum_{j=1}^{(T-t)} B_t(t+j) Q_{k,t}^{SV}(t+j), \tag{3.8}$$

where $Q_{k,t}^{SV}(t+j)$ is the survival probability up to time $t+j$ under risk neutral measure and $B_t(t+j)$ is the price of a riskless zero coupon bond with maturity at time $t+j$. The time frequency for CDS returns we considered here is month by month. Since the investor has zero cost when initiating the contract, the excess return is equal to the value of the stream.

To facilitate the calculation of risk neutral survival probability, it is assumed for simplicity that CDS spread only reflects default risk, the default intensity is constant over the maturity period and there is a constant recovery rate $\rho = 40\%$. Then, we follow Duffie and Singleton (2003) to solve for risk neutral survival probability:

$$CDS_{k,t} = 4\frac{(1-\rho)\sum_{j=1}^{(T-t)} Q_{k,t}^{def|SV}(t+j)B_t(t+j)}{\sum_{j=1}^{(T-t)} Q_{k,t}^{SV}(t+j)B(t,t+j)}, \tag{3.9}$$

$$Q_{k,t}^{SV}(t+j) = exp(-\lambda_{k,t}(t+j)), \tag{3.10}$$

where $Q_{k,t}^{def|SV}(t+j)$ represents the risk neutral probability of a default occurring in period $t+j$ conditional on survival up to time $t+j-1$.

3.3.3 Systematic Risk and Downside Risk Measures

There is sufficient evidence supporting the view that general investors are loss averse following the seminal work of Kahneman and Tversky (1979). Asymmetric preferences were already used in the early finance literature to provide alternatives to the standard CAPM, which is based on the variance as a symmetric risk concept. Markowitz (1952), for example, introduces the notion of semi-variance as a measure of risk. The notion is exploited and extended in asset pricing theory by Hogan and Warren (1974), Bawa and Lindenberg (1977), and Harlow and Rao (1989).

Regular CAPM beta is calculated as $\beta = \frac{Cov(r_i, r_m)}{Var(r_m)}$. For downside beta and upside beta, we follow Ang, Chen, and Xing (2006) to calculate downside beta as conditional CAPM beta when market excess return is below average market excess return. Downside beta is denoted as β^-,

$$\beta^- = \frac{Cov(r_i, r_m | r_m < \overline{r_m})}{Var(r_m | r_m < \overline{r_m})}, \tag{3.11}$$

where r_i and r_m are individual stock excess return and market excess return, $\overline{r_m}$ is average market excess return. Similarly, upside beta is defined as conditional CAPM beta when market excess return is above average market excess return,

$$\beta^{+} = \frac{Cov(r_i, r_m | r_m > \overline{r_m})}{Var(r_m | r_m > \overline{r_m})}. \tag{3.12}$$

Everything else being equal, firms with high downside potential are not as desirable as firms with low downside potential, since former firms tend to correlate with market return with higher magnitude during market downturns when investors' wealth is low. On the other hand, consider two stocks with same downside beta, but with different payoff potential in up markets. The stock that covaries more with the market when market is going up has larger payoff. Investors will not require a high expected return to hold this asset. Thus, everything else being equal, stock with high upside beta will have a discount and earn less reward. However, regular beta, downside beta, and upside beta are related by construction. It is important to differentiate effects from different risk, we introduce three additional relative measures to control for other risk measures.

The relative downside beta (denoted by $\beta^{-}-\beta$) is downside beta relative to regular CAPM beta. Similarly, upside beta relative to regular beta is the relative upside beta computed as $\beta^{+} - \beta$. Another relative risk measure is asymmetric beta which is the difference between downside beta and upside beta, $(\beta^{-} - \beta^{+})$. In a simple model with disappointment utility where investors care more about downside risk, Ang, Chen, and Xing (2006) show that stocks with either high regular beta, high downside beta, or high upside beta are rewarded with higher expected returns since these three beta measures are correlated. Further, stocks with high relative downside beta or high asymmetric beta are remunerated by higher expected returns, while stocks with high relative upside beta have lower expected returns in the model.

3.4 Empirical Results

3.4.1 Systematic Risk

In this section, we investigate whether firms with higher systematic risk have higher average returns for CDS protection sellers. Systematic risk is measured by CAPM beta which describes how strong a firm's return covary with the market return. If there is a cross-sectional relation between systematic risk and returns, then there should exist a pattern between average realized returns and the factor loadings associated with exposure to systematic risk. The CAPM model implies that stocks that covary strongly with the market have contemporaneously high average returns. According to the Merton model implications described in Section 3.2.2, we would also expect a positive relationship between equity CAPM beta and CDS sellers' returns. Though this relationship may depend on firm characteristics that explains the CDS/equity elasticity, the sign of the relationship should maintain positive if we focus on cross-sectional relationship.

Every month, we use daily stock returns from past twelve months to compute a stock's CAPM beta. At the beginning of each month at time t, we sort single-name CDS contracts into five quintiles based on systematic risk estimated using realized equity returns over the past twelve months. We concentrate on presenting the results of equal-weighted CDS portfolios and equal-weighted Fama-Macbeth (1973) regressions.

In the columns under label "CDS returns (%)" of Table 3.2, it reports the average realized excess CDS sellers' returns from t to $t+1$ in each equally weighted quintile portfolio. CDS contracts with maturity from one year (labeled as "1 yr") to ten years (labeled as "10 yr") are included for analysis. For five year maturity CDS contracts (labeled as "5 yr"), quintile 1 has an average excess return of 0.06% per month for protection sellers while quintile 5 has an average excess return of 0.42% per month. The spread in average excess returns between quintile portfolios 5 and 1 is 0.36% per month which is significantly positive with t statistic of 2.64. The results are consistent across each CDS contract maturity with slightly increasing return spread between portfolio 5 and 1 as we move from one year maturity to ten years maturity. The results are also consistent with the notion that investors are rewarded for bearing systematic risk. It does not necessarily mean CAPM model holds as there might be other variable that also explain expected returns. However, it implies that CDS protection sellers bearing high market risk are rewarded with high average returns.

Table 3.2 also reports unconditional and conditional systematic risk measures for each quintile portfolio under label "Systematic risk". Conditional market risk measures: negative beta and positive beta (β^- and β^+) estimated from past twelve month data (same as β) are reported. Negative beta is estimated conditional on market return being less than average market return while positive beta is estimated conditional on market return greater than average market return. By construction, unconditional and conditional beta measures are positively correlated. It means that higher β^- or higher β^+ must also imply higher unconditional beta. Note that, from last three columns in Table 3.2, for these portfolios sorted by β, the spread in β^- and β^+ are similar to the spread in β. Thus, in next subsection, to analyze the rewards for downside and upside risk, we apply several different methods to distinguish rewards from conditional market risk with rewards from unconditional market risk while taking into account the fact that the risk measures are highly correlated.

3.4.2 Downside Beta and Upside beta

In this subsection, we explore the relationship between conditional market risk measures and cross-sectional CDS sellers' returns. We report returns of equal-weighted CDS portfolios sorted by downside beta and upside beta in Table 3.3.

Panel A of Table 3.3 displays the CDS portfolio returns sorted by downside beta β^-. It shows that for higher β^- firms, CDS protection sellers earn higher excess returns. Selling five year maturity CDS contracts in the quintile with the lowest (highest) β^- earns an average excess return of 0.08% (0.43%) per month. The average difference between quintile portfolio 5 and 1 is 0.35% per month which is statistically significant with t statistic of 2.89. The return spread is uniformly positive for CDS contracts with maturity from one year to ten years and the return spread is increasing as we move from short maturity to long maturity. These results are consistent with the argument that CDS protection sellers dislike downside risk and request higher reward for selling protection to firms that covary strongly when the market goes down. Firms with high β^- should carry a premium in order for agents to provide CDS protection.

There is an alternative explanation to explain the results: agents simply have no specific emphasis on downside risk versus upside potential. And the positive relation between β^- and CDS sellers' returns could be a mechanical result caused by construction of β^- and β. Indeed, under label "Systematic risk" in Panel A of Table 3.3, it is shown that high β^- portfolios do have higher β and β^+. However, the average β spread between quintile portfolio 5 and portfolio 1 is 1.13 (from 0.57 to 1.69), smaller than the average β^- spread which is 1.37 (from 0.48 to 1.85). This indicates that the variation in β is not as disperse as the variation in β^- across these quintile portfolios, suggesting downside risk β^- may have independent risk-award relationship that is not subsumed by β.

Panel B of Table 3.3 displays the CDS portfolio returns sorted on upside beta β^+. It reports a positive relationship between β^+ and CDS sellers' returns. For CDS contracts with five years maturity, the return spread between quintile portfolio 5 and 1 is 0.22% per month. The return spread is significantly positive with a t statistics of 2.17. However, the magnitude of the return spread for five year CDS contracts sorted on β^+ (0.22%) is smaller than when sorted on β^- (0.35% as shown in Panel A of Table 3.3) and when sorted on β (0.36% as shown in Table 3.2).

Theory suggests that firms with rising stock prices when market goes up should be more attractive to investors and thus earn low returns for protection sellers. However, we do not observe a discount for firms that have high β^+. Rather, firms with high β^+ earn higher returns for CDS protection sellers. This pattern seems to be inconsistent with agents having a preference for upside potential. To explain this seemingly contradicting result, note that sorting on the upside beta measure $\beta+$ alone does not control for the effects of unconditional β or of β^-. Thus, the positive relationship between β^+ and CDS returns shown in Panel B of Table 3.3 may be a result of increasing β or β^- from quintile portfolio 1 to 5. It is reported in last three columns in Panel B of Table 3.3 that the spread in β is 1.11 and the spread in β^- is 0.94. It is confirmed that higher CAPM beta and downside beta both imply higher CDS sellers' returns. Hence, we study the conditional risk measures while controlling for the confounding effects in the next subsection.

3.4.3 Pure Downside Beta and Upside beta

To control for unconditional systematic risk, we examine the reward for downside (upside) risk in CDS markets by first sorting firms by relative downside (upside) beta ($\beta^-(\beta^+) - \beta$) and asymmetric beta ($\beta^- - \beta^+$). Relative downside (upside) beta is computed to focus on the incremental effect of downside (upside) beta over the regular CAPM beta. Asymmetric beta describes how asymmetric a firm's return sensitivity to market return will be depending on whether the market is performing bad or good. These measures will alleviate the concern of the potential impact of unconditional risk on conditional risk.

Panel A of Table 3.4 shows that the relation between relative downside beta and CDS sellers' returns is not significant for one year to seven years CDS contracts. For ten years CDS contract, there is a positive return spread of 0.17% significant at 10% level. It is also shown that when moving from quintile portfolio 1 to quintile portfolio 5, there is a U-shape pattern for average CDS excess returns for each maturity group. This U-shape pattern for CDS returns is aligned well with the pattern for β across these portfolios sorted on relative downside beta. Thus, by sorting on relative downside beta, we have not completely ruled out the impact of β on CDS returns.

Interestingly, quintile portfolio 1 (with lowest relative downside beta) has average β of 1.31 which is larger than the average β (1.11) for quintile portfolio 5 (with highest relative downside beta). If β is the only risk that explains the cross-section of CDS returns, we should expect higher average returns for quintile portfolio 1. Yet, though insignificantly, the average CDS returns of quintile portfolio 1 is slightly smaller than those of quintile portfolio 5. It indicates that higher downside beta has an impact on CDS returns that is not subsumed by CAPM beta. We will apply another method to fully disentangle downside beta from regular CAPM beta later in this subsection.

In panel B of Table 3.4, we examine the effects of β^+ while controlling for regular beta by sorting firms according to relative upside beta ($\beta^+ - \beta$). We find that higher relative upside beta firms have lower CDS returns. The return spread between quintile portfolio 5 and 1 are ranging from -0.27% to -0.15% for five different maturities which are all significantly negative. This pattern of low CDS returns to high relative upside beta firms is consistent with agents willing to accept a discount for selling protection for firms with high upside potential. Though we have not fully rule out the impact of regular beta, the difference of average β between quintile portfolio 1 and 5 would actually work against us. Note that quintile portfolio 5, with highest relative upside beta, also has higher average regular beta (1.24) than quintile portfolio 1 (1.16). Higher regular beta should be associated with higher CDS returns for quintile portfolio 5, compared with portfolio 1. However, since quintile portfolio 5 also has higher average upside beta, lower average CDS returns for this portfolio are discovered to imply that protection sellers accept a discount of expected return because of the higher upside potential of these firms.

We also sort firms by asymmetric beta ($\beta^- - \beta^+$) in panel C of Table 3.4. This measure is computed to provide an examination of the impact of downside beta relative to upside beta. In panel C of Table 3.4, we observe an increasing pattern in average CDS sellers' returns with increasing asymmetric beta. The return spread between quintile portfolio 5 and 1 is significantly positive at 1% level and ranges from 0.12% for one year maturity CDS to 0.30% for ten years maturity CDS. Average β of quintile portfolio 5 is smaller than that of quintile portfolio 1, thus higher average CDS returns of portfolio 5 indicates that firms covary more with market during bad times than good times compensate CDS protection sellers' with a larger reward for bearing undesirable risk.

Table 3.4 demonstrates that CDS sellers providing protection for firms that covary strongly with the market conditional on positive moves accept discounts in rewards. It also shows some weak evidence that CDS sellers require higher returns to sell protection for firms with higher relative downside beta. However, sorting on relative downside (upside) beta does not fully eliminate the impact of regular CAPM beta, which we now provide further analysis.

To disentangle downside (upside) beta from regular CAPM beta, we extract out the projection of downside (upside) beta on CAPM beta and call it fitted downside (upside) beta (component of conditional market beta that is explained by unconditional market risk). Then the residual downside (upside) beta that is not explained by unconditional market beta will now be orthogonal to CAPM beta. This residual downside (upside) beta measure will provide information about conditional risk that is completely unrelated to unconditional market risk.

We first run time series regression of β^- on β for each firm individually,

$$\beta^- = c_0 + c_1\beta + e. \tag{3.13}$$

Then we obtain fitted downside beta $\hat{\beta}^- = \hat{c}_0 + \hat{c}_1\beta$ and residual downside beta $\hat{e}(\beta^-) = \beta^- - \hat{\beta}^-$ from downside beta, beta and estimation results from regression. Fitted upside beta and residual upside beta will be calculated similarly.

Table 3.5 reports the average CDS returns for portfolios sorted on fitted downside beta and residual downside beta. Panel A of Table 3.5 confirms the positive relationship between CAPM beta and CDS returns. Quintile portfolio 5 have highest fitted downside beta (project of β^- on β) and significantly higher CDS returns than quintile portfolio 1. Panel B of Table 3.5 shows the impact of pure downside beta by sorting firms on residual downside beta.

We can confirm that the risk-return pattern will not be influenced by regular beta since quintile portfolio 1 and 5 have very close average beta (1.19 and 1.18). A significant positive relationship between residual downside beta and CDS returns is observed for CDS contracts with maturity from three years to ten years. The return spread between quintile portfolio 5 and 1 increases from 0.04% for CDS with one year maturity to 0.15% for CDS with ten years maturity. Hence, it is confirmed that exposure to pure downside beta risk implies high returns for CDS protection sellers, consistent with agents requiring higher rewards for bearing higher downside risk. Besides, the positive relationship between downside beta and CDS returns are stronger for longer maturity CDS contracts.

The results sorted on fitted upside beta and residual upside beta are reported in Table 3.6. Panel A of Table 3.6 presents the average CDS returns of portfolios sorted on fitted upside beta. It explains the seemingly puzzling evidence of positive relationship between upside beta and CDS returns previously shown in panel B of Table 3.3. $\beta+$ and β are highly correlated and sorting on the projection of β^+ on β reflects the CAPM suggested pattern between systematic risk and return. Thus, sorting on β^+ alone will not rule out the effects of β. To illustrate the impact of pure upside beta on CDS returns, we turn to panel B of Table 3.6 where firms are sorted on residual upside beta. The return spreads between quintile portfolio 5 and 1 are significantly negative across all contract maturities, ranging from -0.11% to -0.22% as we move from one year maturity to ten years maturity. It is consistent with protection sellers willing to accept a discount of return for firms with strong upside potential. The evidence shown above is also consistent with the results displayed in panel B of Table 3.4. However, note that β does not show much variation across each portfolio sorted on residual upside beta. By examining this pure upside beta, we identify a clear pattern between upside beta and CDS returns that is consistent with theory.

In summary, this subsection demonstrates that downside risk is rewarded in the cross-section of CDS returns. This effect is not mechanically driven by CAPM beta since pure downside risk is priced too. On the other hand, CDS protection sellers accept a discount to sell CDS for firms that covary strongly with the market conditional on positive moves of the market. The pattern between downside (upside) beta and CDS returns is robust while controlling for variation in CAPM beta. However, these sorting results does not control for other firm characteristics that may potentially affect CDS returns. We consider these firm control variables in the next subsection.

3.4.4 Fama-MacBeth Regressions

Section 3.2.2 describes the relation between CDS returns and CAPM beta and shows that the sensitivity of CDS returns to the stock returns h_E is also an important factor determining the CDS returns apart from beta. In Merton model, h_E is determined by leverage, maturity and asset volatility. Besides, size and book-to-market have been found to be critical in explaining equity returns. Thus, in this subsection, we run Fama-Macbeth regressions of excess returns on (un)conditional betas and firm characteristics including leverage, asset volatility, size, btm and credit ratings.

Table 3.7 reports the results of Fama-Macbeth regression of one year maturity CDS sellers' returns over t to $t+1$ on various sets of (un)conditional risk measures and firm characteristics that are available to investors at time t. Newey-West t-statistics are reported in brackets to determine the statistical significance of the average intercept and slope coefficients. Each independent variable is standardized to have straightforward interpretation of the estimates.

Consistent with single-sorting results reported in Section 3.4.1, Model 1 in Table 3.7 confirms that CAPM beta carries a significantly positive coefficient. Model 2 and 3 confirm that both downside beta (β^-) and upside beta (β^+) also have significant positive coefficients when they are included separately as the single regressor. Upside beta has a seemingly contradicting sign of coefficient compared with theory. As mentioned earlier, the reason is that beta, downside beta, and upside beta are all positively correlated and the relation between upside beta and CDS returns may be biased if not controlling for other risk measures.

In model 4 where both β^- and β^+ are included as regressors, we find that only downside risk is priced with a positive sign. The coefficient on downside beta is 0.176, highly significant (t-statistics of 2.96), consistent with portfolio sorts on downside beta in Section 3.4.2. The coefficient on upside beta is negative (-0.015) with insignificant t-statistic (-0.64). These results support the hypothesis that CDS protection sellers demand positive premium on β^-, but the evidence for sellers accepting discount on β^+ is weak. Model 5 includes both β and pure downside beta $\hat{e}(\beta^-)$. β still has a significantly positive coefficient while pure downside beta has a positive coefficient (0.018) which is insignificant. Model 6 includes β and pure upside beta $\hat{e}(\beta^+)$. In this model, pure upside beta is found to carry a negative risk premium with t statistic of -2.35. Model 7 to model 8 confirms model 5 and model 6 by controlling for firm characteristics that might be related to CDS returns. Higher numerical rating number means worse credit rating in the data sample. Thus, it is natural that this measure has a positive coefficient (about 0.077) meaning CDS sellers providing protection for worse credit rating firms should be rewarded with higher returns.

Table 3.8 and table 3.9 display similar Fama-Macbeth regression results for CDS contracts with five years maturity and ten years maturity respectively. β^- is priced in the cross-section while β^+ has insignificant negative coefficient when both of them are included as regressors in model 4. Note that pure downside risk measure $\hat{e}(\beta^-)$ has significant positive coefficient in model 5 and model 7 for both CDS contracts with maturity of five years and ten years.

The consistent results from the regressions in Table 3.7 to Table 3.9 suggest that rewards in CDS markets for (pure) downside beta is always positive and statistically significant. CDS sellers providing protection to high (pure) downside firms are compensated by higher average returns, and this result is robust to controlling for other firm characteristics and conditional risk characteristics. The pattern between pure downside beta and CDS returns are stronger for longer maturity CDS contracts. On the other hand, the reward or discount for upside beta is found to be insignificant in the cross-section once we control for other conditional risk measures and firm characteristics. We expect the coefficient of β^+ to be negative once controlling for β^-. However, in empirical results, it sometimes flips sign and is always insignificantly negative when we control other risk attributes. Overall, aversion to market risk and downside risk is priced in CDS markets more strongly, and more robustly, in the cross-section than investors' attraction to upside potential. This is in accordance with the fact that CDS contracts are more valuable during market downturns and CDS investors mainly care about downside market risk.

3.5 Conclusion

Theory suggests that systematic risk should be priced in the cross-section of asset returns. Besides, previous literature suggests a role for downside risk based on rational preferences. If investors treat a downside loss differently from an upside gain of equal magnitude, there should be a significantly positive link between downside risk and expected CDS returns required by protection sellers. This pattern is expected to be discovered in the CDS market as these contracts provide valuable protection especially when the market is going down. There have been various earlier studies testing this relation in the equity markets. However, the evidence is mixed and only a contemporaneous relationship between downside risk and stock returns is discovered in literature. Our paper contributes to the literature by analyzing the systematic (downside) risk-expected return relation at the credit market level.

The univariate sorting analysis based on CAPM beta and downside beta find a significant positive return spread of CDS contracts between high-risk firms and low-risk firms. A closer look of the results sorted on conditional downside beta calls for more analysis to rule out the impact of CAPM beta since unconditional and conditional beta measures are correlated. We use two methods to isolate the impact of downside beta on CDS returns and confirm that firms with high downside risk are associated with a positive premium required by CDS protection sellers. The pattern between downside risk and CDS returns are stronger for longer maturity CDS contracts. We test the robustness of our results by using firm-level Fama-Macbeth regressions and controlling for other firm characteristics. These regression results confirm that higher risk premium is required by investors to provide protection for firms with higher systematic risk and downside risk.

Table 3.1 Sample Statistics

This table presents summary statistics of CDS returns and spreads. The sample period is from January 2001 to December 2015 and observations are at monthly frequency. CDS data are provided by Markit.

	Mean	Stdev	Min	Max
CDS Return (%, 1 yr)	0.1049	1.8174	-52.6563	109.1511
CDS Return (%, 3 yr)	0.1391	2.5470	-55.8476	156.2189
CDS Return (%, 5 yr)	0.1639	3.0161	-51.2839	253.3413
CDS Return (%, 7 yr)	0.1725	3.4418	-51.1963	348.1378
CDS Return (%, 10 yr)	0.1789	3.8171	-65.1980	387.4469
CDS Spread (1 yr)	0.0150	0.0799	0.0001	7.5031
CDS Spread (3 yr)	0.0178	0.0645	0.0002	6.5620
CDS Spread (5 yr)	0.0204	0.0553	0.0001	3.0763
CDS Spread (7 yr)	0.0213	0.0512	0.0003	3.0455
CDS Spread (10 yr)	0.0219	0.0477	0.0005	2.6723

Table 3.2 CDS Sellers' Returns Sorted by CAPM Beta

This table lists the monthly equal-weighted average CDS sellers' returns and risk characteristics of the firms sorted by CAPM beta. For each month, we calculate CAPM beta (β) using daily continuously compounded returns over past 12 months. CDS contracts are ranked into quintiles ($1-5$) by β and form equal-weighted portfolios each month. The columns under label "CDS returns (%)" reports the average percentage excess returns for protection sellers over one month. CDS contracts with maturity of 1 year, 3 years, 5 years, 7 years, and 10 years are included in table. The columns under label "Systematic risk" reports the CAPM beta (β), downside beta (β^-) and upside beta (β^+) which are estimated from past 12 months. The row labeled "High-Low" presents the difference between the CDS returns or systematic risk between portfolio 5 and portfolio 1. The entry labeled "t-stat" in square brackets is the t-statistic computed using Newey-West heteroskedastic-robust standard errors with 12 lags for the High-Low difference. The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

	CDS returns (%)					Systematic risk		
	1 yr	3 yr	5 yr	7 yr	10 yr	β	β^-	β^+
1 Low β	0.03	0.04	0.06	0.06	0.05	0.50	0.55	0.46
2	0.06	0.08	0.09	0.10	0.11	0.79	0.83	0.75
3	0.07	0.10	0.13	0.13	0.14	1.01	1.03	0.99
4	0.12	0.16	0.19	0.21	0.21	1.26	1.26	1.25
5 High β	0.27	0.35	0.42	0.45	0.48	1.78	1.76	1.77
High-Low	0.25***	0.31**	0.36***	0.39***	0.42***	1.27***	1.21***	1.30***
t-stat	[3.06]	[2.56]	[2.64]	[2.65]	[2.75]	[48.11]	[50.19]	[29.09]

Table 3.3 CDS Sellers' Returns Sorted by Downside/Upside Beta

This table lists the monthly equal-weighted average CDS sellers' returns and risk characteristics of the firms sorted by downside beta and upside beta. For each month, we calculate downside beta (β^-) and upside beta (β^+) using daily continuously compounded returns over past 12 months. In Panel A, CDS contracts are ranked into quintiles ($1-5$) by downside beta: β^- and form equal-weighted portfolios each month. In Panel B, CDS contracts are sorted by upside beta: β^+. The columns under label "CDS returns (%)" reports the average percentage excess returns for protection sellers over one month. CDS contracts with maturity of 1 year, 3 years, 5 years, 7 years, and 10 years are included in table. The columns under label "Systematic risk" reports the CAPM beta (β), downside beta (β^-) and upside beta (β^+) which are estimated from past 12 months. The row labeled "High-Low" presents the difference between the CDS returns or systematic risk between portfolio 5 and portfolio 1. The entry labeled "t-stat" in square brackets is the t-statistic computed using Newey-West heteroskedastic-robust standard errors with 12 lags for the High-Low difference. The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

	CDS returns (%)					Systematic risk		
	1 yr	3 yr	5 yr	7 yr	10 yr	β	β^-	β^+
Panel A: Sorted by downside beta: $\boldsymbol{\beta^-}$								
1 Low β^-	0.05	0.05	0.08	0.08	0.05	0.57	0.48	0.57
2	0.05	0.06	0.07	0.07	0.06	0.82	0.79	0.81
3	0.07	0.10	0.13	0.14	0.17	1.02	1.02	1.01
4	0.10	0.14	0.18	0.18	0.20	1.25	1.29	1.22
5 High β^-	0.28	0.38	0.43	0.48	0.50	1.69	1.85	1.62
High-Low	0.23***	0.33***	0.35***	0.40***	0.46***	1.13***	1.37***	1.05***
t-stat	(3.48)	(3.07)	(2.89)	(2.95)	(3.23)	(37.65)	(52.39)	(23.33)
Panel B: Sorted by upside beta: $\boldsymbol{\beta^+}$								
1 Low β^+	0.08	0.11	0.12	0.14	0.16	0.59	0.69	0.37
2	0.06	0.09	0.11	0.11	0.11	0.81	0.87	0.72
3	0.07	0.12	0.15	0.17	0.15	1.02	1.04	0.99
4	0.11	0.14	0.17	0.18	0.19	1.24	1.23	1.28
5 High β^+	0.22	0.28	0.34	0.37	0.39	1.70	1.63	1.87
High-Low	0.14**	0.17**	0.22**	0.22**	0.23*	1.11***	0.94***	1.51***
t-stat	(2.60)	(2.08)	(2.17)	(2.14)	(1.97)	(31.89)	(29.42)	(40.94)

Table 3.4 CDS Sellers' Returns Sorted by Relative Downside/Upside Beta

This table lists the monthly equal-weighted average CDS sellers' returns and risk characteristics of the firms sorted by relative betas. For each month, we calculate CAPM beta (β), downside beta (β^-) and upside beta (β^+) using daily continuously compounded returns over past 12 months. In Panel A, CDS contracts are ranked into quintiles ($1-5$) by relative downside beta: $\beta^- - \beta$ and form equal-weighted portfolios each month. In Panel B, CDS contracts are sorted by relative upside beta: $\beta^+ - \beta$ and in Panel C, they are sorted by asymmetric beta: $\beta^- - \beta^+$. The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

	CDS returns (%)					Systematic risk		
	1 yr	3 yr	5 yr	7 yr	10 yr	β	β^-	β^+
Panel A: Sorted by relative downside beta: $\boldsymbol{\beta^- - \beta}$								
1 Low $\beta^- - \beta$	0.17	0.20	0.26	0.27	0.23	1.31	1.00	1.40
2	0.07	0.10	0.10	0.09	0.10	1.05	0.96	1.07
3	0.06	0.09	0.10	0.11	0.11	0.97	0.98	0.95
4	0.08	0.10	0.12	0.14	0.16	0.95	1.07	0.89
5 High $\beta^- - \beta$	0.18	0.27	0.31	0.37	0.41	1.11	1.47	0.95
High-Low	0.01	0.08	0.05	0.10	0.17*	-0.21***	0.47***	-0.45***
t-stat	(0.11)	(1.11)	(0.59)	(1.18)	(1.95)	(-5.40)	(8.66)	(-10.24)
Panel B: Sorted by relative upside beta: $\boldsymbol{\beta^+ - \beta}$								
1 Low $\beta^+ - \beta$	0.25	0.34	0.39	0.44	0.42	1.16	1.29	0.75
2	0.10	0.13	0.15	0.15	0.20	0.99	1.05	0.86
3	0.05	0.08	0.10	0.09	0.08	0.97	0.99	0.96
4	0.06	0.08	0.10	0.12	0.13	1.01	1.00	1.12
5 High $\beta^+ - \beta$	0.10	0.12	0.16	0.18	0.18	1.24	1.14	1.57
High-Low	-0.15***	-0.23***	-0.23***	-0.27***	-0.24**	0.08*	-0.15***	0.81***
t-stat	(-2.81)	(-3.10)	(-3.06)	(-3.02)	(-2.49)	(1.67)	(-3.78)	(14.85)
Panel C: Sorted by asymmetric beta: $\boldsymbol{\beta^- - \beta^+}$								
1 Low $\beta^- - \beta^+$	0.11	0.11	0.16	0.15	0.13	1.28	1.03	1.55
2	0.07	0.10	0.12	0.12	0.11	1.04	0.97	1.12
3	0.07	0.11	0.11	0.15	0.16	0.96	0.98	0.95
4	0.08	0.13	0.16	0.17	0.19	0.96	1.07	0.85
5 High $\beta^- - \beta^+$	0.23	0.32	0.35	0.41	0.43	1.14	1.43	0.79
High-Low	0.12***	0.20***	0.19***	0.25***	0.30***	-0.14***	0.40***	-0.76***
t-stat	(3.75)	(3.93)	(3.22)	(3.45)	(3.93)	(-3.43)	(8.37)	(-16.41)

Table 3.5 CDS Sellers' Returns Sorted by Fitted/Residual Downside Beta

This table lists the monthly equal-weighted average CDS sellers' returns and risk characteristics of the firms sorted by fitted and residual downside beta. For each month, we calculate CAPM beta (β), downside beta (β^-) and upside beta (β^+) with respect to the market of all stocks using daily continuously compounded returns over past 12 months. Then we run time series regression of β^- on β for each firm to obtain fitted downside beta $\hat{\beta}^-$ (projection of β^- on β) and residual downside beta $\hat{e}(\beta^-)$ (part of β^- orthogonal to β). In Panel A, CDS contracts are ranked into quintiles $(1-5)$ by fitted downside beta and form equal-weighted portfolios each month. In Panel B, CDS contracts are sorted by residual downside beta. The columns under label "CDS returns (%)" reports the average percentage excess returns for protection sellers over next one month. CDS contracts with maturity of 1 year, 3 years, 5 years, 7 years and 10 years are included in table. The columns under label "Systematic risk" reports the CAPM beta (β), downside beta (β^-) and upside beta (β^+) which are estimated from past 12 months. The row labeled "High-Low" presents the difference between the CDS returns or systematic risk between portfolio 5 and portfolio 1. The entry labeled "t-stat" in square brackets is the t-statistic computed using Newey-West heteroskedastic-robust standard errors with 12 lags for the High-Low difference. The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

	CDS returns (%)					Systematic risk		
	1 yr	3 yr	5 yr	7 yr	10 yr	β	β^-	β^+
Panel A: Sorted by fitted downside beta (projection on β): $\hat{\boldsymbol{\beta}}^- = \hat{\boldsymbol{c}}_0 + \hat{\boldsymbol{c}}_1\boldsymbol{\beta}$								
1 Low $\hat{\beta}^-$	0.03	0.07	0.08	0.09	0.07	0.53	0.54	0.50
2	0.05	0.07	0.09	0.10	0.08	0.80	0.83	0.78
3	0.07	0.11	0.15	0.15	0.19	1.02	1.04	1.01
4	0.11	0.15	0.19	0.20	0.21	1.26	1.27	1.25
5 High $\hat{\beta}^-$	0.27	0.38	0.43	0.45	0.48	1.75	1.78	1.73
High-Low	0.24***	0.32***	0.35***	0.36**	0.41***	1.23***	1.23***	1.23***
t-stat	(3.31)	(2.78)	(2.65)	(2.51)	(2.72)	(60.03)	(64.87)	(36.5)
Panel B: Sorted by residual downside beta: $\hat{\boldsymbol{e}}(\boldsymbol{\beta}^-) = \boldsymbol{\beta}^- - \hat{\boldsymbol{\beta}}^-$								
1 Low $\hat{e}(\beta^-)$	0.15	0.21	0.24	0.24	0.24	1.19	0.93	1.24
2	0.07	0.10	0.13	0.11	0.12	1.01	0.93	1.03
3	0.07	0.10	0.13	0.15	0.14	0.98	0.99	0.97
4	0.07	0.11	0.14	0.15	0.17	1.00	1.11	0.96
5 High $\hat{e}(\beta^-)$	0.19	0.27	0.32	0.34	0.39	1.18	1.49	1.05
High-Low	0.04	0.07*	0.08*	0.10**	0.15***	-0.01	0.56***	-0.20***
t-stat	(1.38)	(1.77)	(1.67)	(2.11)	(2.76)	(-0.14)	(17.48)	(-7.22)

Table 3.6 CDS Sellers' Returns Sorted by Fitted/Residual Upside Beta

This table lists the monthly equal-weighted average CDS sellers' returns and risk characteristics of the firms sorted by fitted and residual upside beta. For each month, we calculate CAPM beta (β), downside beta (β^-) and upside beta (β^+) with respect to the market of all stocks using daily continuously compounded returns over past 12 months. Then we run time series regression of β^+ on β for each firm to obtain fitted upside beta $\hat{beta}^+$ (projection of β^+ on β) and residual upside beta $\hat{e}(\beta^+)$ (part of β^+ orthogonal to β). In Panel A, CDS contracts are ranked into quintiles ($1-5$) by fitted upside beta and form equal-weighted portfolios each month. In Panel B, CDS contracts are sorted by residual upside beta. The columns under label "CDS returns (%)" reports the average percentage excess returns for protection sellers over next one month. CDS contracts with maturity of 1 year, 3 years, 5 years, 7 years and 10 years are included in table. The columns under label "Systematic risk" reports the CAPM beta (β), downside beta (β^-) and upside beta (β^+) which are estimated from past 12 months. The row labeled "High-Low" presents the difference between the CDS returns or systematic risk between portfolio 5 and portfolio 1. The entry labeled "t-stat" in square brackets is the t-statistic computed using Newey-West heteroskedastic-robust standard errors with 12 lags for the High-Low difference. The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

	CDS returns (%)					Systematic risk		
	1 yr	3 yr	5 yr	7 yr	10 yr	β	β^-	β^+
Panel A: Sorted by fitted upside beta (projection on β): $\hat{\beta}^+ = \hat{c}_0 + \hat{c}_1\beta$								
1 Low $\hat{\beta}^+$	0.05	0.07	0.09	0.09	0.11	0.53	0.59	0.44
2	0.05	0.09	0.12	0.14	0.12	0.80	0.86	0.75
3	0.07	0.10	0.13	0.13	0.12	1.02	1.03	1.00
4	0.13	0.19	0.23	0.24	0.26	1.26	1.26	1.27
5 High $\hat{\beta}^+$	0.23	0.33	0.38	0.39	0.43	1.75	1.71	1.80
High-Low	0.19***	0.26***	0.29***	0.29***	0.32**	1.22***	1.12***	1.35***
t-stat	(3.36)	(3.05)	(2.85)	(2.65)	(2.54)	(58.04)	(55.52)	(41.87)
Panel B: Sorted by residual upside beta: $\hat{e}(\beta^+) = \beta^+ - \hat{\beta}^+$								
1 Low $\hat{e}(\beta^+)$	0.22	0.31	0.36	0.40	0.41	1.18	1.29	0.83
2	0.09	0.13	0.15	0.13	0.17	1.01	1.05	0.90
3	0.06	0.09	0.11	0.12	0.14	0.97	0.99	0.96
4	0.07	0.10	0.12	0.13	0.14	1.01	1.00	1.09
5 High $\hat{e}(\beta^+)$	0.11	0.17	0.20	0.22	0.20	1.20	1.13	1.47
High-Low	-0.11**	-0.14**	-0.17***	-0.18**	-0.22**	0.02	-0.16***	0.64***
t-stat	(-2.58)	(-2.58)	(-2.76)	(-2.45)	(-2.57)	(0.75)	(-6.67)	(17.07)

Table 3.7 Fama-Macbeth Regressions of Seller's Returns on One Year CDS Contract

This table displays the results of Fama-Macbeth regression of CDS sellers' return of one year maturity contract on systematic risk and firm characteristics. For each month, we calculate CAPM beta (β), downside beta (β^-) and upside beta (β^+) with respect to the market of all stocks using daily continuously compounded returns over past 12 months. Then we run time series regression of β^- on β for each firm to obtain fitted downside beta (projection of β^+ on β) and residual downside beta $\hat{e}(\beta^+)$ (part of β^+ orthogonal to β). Other control variables in the regression include market equity (size), book-to-market ratio (btm), leverage (lev), asset volatility (avol), and credit rating (Rating). The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

Model	1	2	3	4	5	6	7	8
Intercept	-0.107	-0.100	-0.036	-0.097	-0.110	-0.101	-0.264	-0.254
	[-2.69]	[-2.53]	[-1.43]	[-2.49]	[-2.64]	[-2.72]	[-2.81]	[-2.88]
β	0.174				0.173	0.171	0.113	0.105
	[2.98]				[2.91]	[3.05]	[1.95]	[1.78]
β^-		0.166		0.176				
		[2.97]		[2.96]				
β^+			0.111	-0.015				
			[2.89]	[-0.64]				
$\hat{e}(\beta^-)$					0.099		0.076	
					[1.40]		[1.27]	
$\hat{e}(\beta^+)$						-0.113		-0.117
						[-2.35]		[-2.31]
lev							-0.013	-0.017
							[-0.66]	[-0.82]
avol							-3.441	-3.563
							[-1.88]	[-1.81]
size							-0.005	-0.005
							[-0.93]	[-0.95]
btm							0.002	0.006
							[0.18]	[0.47]
Rating							0.077	0.078
							[3.76]	[3.87]
$Adj R^2$	0.040	0.045	0.026	0.057	0.053	0.053	0.122	0.124
	[8.17]	[6.19]	[8.01]	[7.56]	[8.92]	[10.25]	[13.93]	[15.05]

Table 3.8 Fama-Macbeth Regressions of Seller's Returns on Five Year CDS Contract

This table displays the results of Fama-Macbeth regression of CDS sellers' return of five year maturity contract on systematic risk and firm characteristics. For each month, we calculate CAPM beta (β), downside beta (β^-) and upside beta (β^+) with respect to the market of all stocks using daily continuously compounded returns over past 12 months. Then we run time series regression of β^- on β for each firm to obtain fitted downside beta (projection of β^+ on β) and residual downside beta $\hat{e}(\beta^+)$ (part of β^+ orthogonal to β). Other control variables in the regression include market equity (size), book-to-market ratio (btm), leverage (lev), asset volatility (avol), and credit rating (Rating). The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

Model	1	2	3	4	5	6	7	8
Intercept	-0.132	-0.138	-0.028	-0.125	-0.135	-0.125	-0.362	-0.364
	[-2.39]	[-2.24]	[-0.59]	[-2.09]	[-2.34]	[-2.39]	[-1.71]	[-1.75]
β	0.246				0.242	0.242	0.193	0.189
	[2.57]				[2.47]	[2.63]	[2.28]	[2.29]
β^-		0.249		0.284				
		[2.62]		[2.75]				
β^+			0.156	-0.051				
			[2.46]	[-1.26]				
$\hat{e}(\beta^-)$					0.197		0.177	
					[1.99]		[2.04]	
$\hat{e}(\beta^+)$						-0.174		-0.158
						[-2.87]		[-2.80]
lev							0.001	-0.004
							[0.03]	[-0.13]
avol							-8.468	-8.988
							[-3.09]	[-3.18]
size							-0.012	-0.011
							[-0.89]	[-0.83]
btm							-0.039	-0.036
							[-2.20]	[-1.94]
Rating							0.122	0.124
							[3.15]	[3.19]
$Adj R^2$	0.038	0.039	0.024	0.050	0.046	0.047	0.117	0.119
	[11.17]	[7.61]	[8.02]	[9.31]	[10.99]	[13.14]	[18.45]	[20.00]

Table 3.9 Fama-Macbeth Regressions of Seller's Returns on Ten Year CDS Contract

This table displays the results of Fama-Macbeth regression of CDS sellers' return of ten years maturity contract on systematic risk and firm characteristics. For each month, we calculate CAPM beta (β), downside beta (β^-) and upside beta (β^+) with respect to the market of all stocks using daily continuously compounded returns over past 12 months. Then we run time series regression of β^- on β for each firm to obtain fitted downside beta (projection of β^+ on β) and residual downside beta $\hat{e}(\beta^+)$ (part of β^+ orthogonal to β). Other control variables in the regression include market equity (size), book-to-market ratio (btm), leverage (lev), asset volatility (avol), and credit rating (Rating). The sample period is from January 2001 to December 2015 and observations are at monthly frequency.

Model	1	2	3	4	5	6	7	8
Intercept	-0.181	-0.195	-0.048	-0.176	-0.183	-0.173	-0.554	-0.536
	[-2.40]	[-2.43]	[-0.62]	[-2.22]	[-2.37]	[-2.40]	[-2.05]	[-1.99]
β	0.311				0.305	0.306	0.229	0.223
	[2.64]				[2.56]	[2.70]	[2.26]	[2.26]
β^-		0.321		0.388				
		[2.83]		[3.26]				
β^+			0.194	-0.088				
			[2.38]	[-1.67]				
$\hat{e}(\beta^-)$					0.286		0.265	
					[2.30]		[2.33]	
$\hat{e}(\beta^+)$						-0.248		-0.220
						[-2.78]		[-2.67]
lev							0.053	0.041
							[1.06]	[0.80]
avol							-7.556	-8.483
							[-1.95]	[-2.16]
size							-0.002	-0.003
							[-0.11]	[-0.15]
btm							-0.065	-0.060
							[-2.56]	[-2.34]
Rating							0.146	0.146
							[3.06]	[3.01]
$AdjR^2$	0.035	0.035	0.021	0.044	0.041	0.044	0.104	0.106
	[8.64]	[6.76]	[6.85]	[8.23]	[8.96]	[10.54]	[16.67]	[17.30]

Bibliography

Ang, A., J. Chen, and Y. Xing (2006). Downside Risk. *The Review of Financial Studies 19*(4), 1191–1239.

Atilgan, Y., T. G. Bali, K. O. Demirtas, and A. D. Gunaydin (2015). Downside Beta and Equity Returns around the World. *SSRN Electronic Journal*.

Bai, J., T. G. Bali, and Q. Wen (2017). Common Risk Factors in the Cross-Section of Corporate Bond Returns. *SSRN Electronic Journal*.

Bawa, V. S. and E. B. Lindenberg (1977). Capital Market Equilibrium in a Mean-Lower Partial Moment Framework. *Journal of Financial Economics 5*(2), 189–200.

Berndt, A. and I. Obreja (2010). Decomposing European CDS Returns. *Review of Finance 14*(2), 189–233.

Bessembinder, H., K. M. Kahle, W. F. Maxwell, and D. Xu (2009). Measuring Abnormal Bond Performance. *The Review of Financial Studies 22*(10), 4219–4258.

Bongaerts, D., F. de Jong, and J. Driessen (2011). Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market. *Journal of Finance 66*(1), 203–240.

Choi, J. and Y. Kim (2017). Anomalies and Market (Dis)Integration. *SSRN Electronic Journal*.

Chordia, T., A. Goyal, Y. Nozawa, A. Subrahmanyam, and Q. Tong (2017). Are Capital Market Anomalies Common to Equity and Corporate Bond Markets? An Empirical Investigation. *Journal of Financial and Quantitative Analysis 52*(4), 1301–1342.

Duffie, D. and K. J. Singleton (2003). *Credit Risk: Pricing, Management, and Measurement.* Princeton, NJ: Princeton University Press.

Fama, E. F. and K. R. French (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics 33*(1), 3–56.

Friewald, N., C. Wagner, and J. Zechner (2014). The Cross-Section of Credit Risk Premia and Equity Returns. *Journal of Finance 69*(6), 2419–2469.

Gebhardt, W. R., S. Hvidkjaer, and B. Swaminathan (2005). The Cross-Section of Expected Corporate Bond Returns: Betas or Characteristics? *Journal of Financial Economics 75*(1), 85–114.

Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica 59*(3), 667.

Harlow, W. V. and R. K. S. Rao (1989). Asset Pricing in a Generalized Mean-Lower Partial Moment Framework: Theory and Evidence. *Journal of Financial and Quantitative Analysis 24*(3), 285–311.

Hogan, W. W. and J. M. Warren (1974). Toward the Development of an Equilibrium Capital-Market Model Based on Semivariance. *Journal of Financial and Quantitative Analysis 9*(1), 1–11.

Huang, J.-Z. and M. Huang (2012). How Much of the Corporate-Treasury Yield Spread Is Due to Credit Risk? *Review of Asset Pricing Studies 2*(2), 153–202.

Jostova, G., S. Nikolova, A. Philipov, and C. W. Stahel (2013). Momentum in Corporate Bond Returns. *The Review of Financial Studies 26*(7), 1649–1693.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*.

Kwan, S. H. (1996). Firm-Specific Information and the Correlation Between Individual Stocks and Bonds. *Journal of Financial Economics 40*(1), 63–80.

Lee, J., A. Naranjo, and S. Sirmans (2014). CDS Momentum: Slow Moving Credit Ratings and Cross-Market Spillovers. *SSRN Electronic Journal*.

Lettau, M., M. Maggiori, and M. Weber (2014). Conditional Risk Premia in Currency Markets and Other Asset Classes. *Journal of Financial Economics 114*(2), 197–225.

Lin, H., J. Wang, and C. Wu (2011). Liquidity Risk and Expected Corporate Bond Returns. *Journal of Financial Economics 99*(3), 628–650.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains From Diversification. *Journal of Finance 20*(4), 587–615.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance 7*(1), 77–91.

Merton, R. C. (1974). On the Pricing of Corporate Debt: the Risk Structure of Interest Rates. *Journal of Finance 29*(2), 449–470.

Roy, A. D. (1952). Safety First and the Holding of Assets. *Econometrica 20*(3), 431.

Schaefer, S. M. and I. A. Strebulaev (2008). Structural Models of Credit Risk Are Useful: Evidence From Hedge Ratios on Corporate Bonds. *Journal of Financial Economics 90*(1), 1–19.

Sharpe, W. F. (1964). Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance 19*(3), 425–442.