Three Essays on Empirical Cross-Sectional Asset Pricing

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2020

Shuwen Yang

Alliance Manchester Business School
Accounting and Finance Division

Table of Contents

This thesis contains 47,196 words, including words in the main text, headers, captions, and footnotes.

Abstract

The University of Manchester
Shuwen Yang
Doctor of Philosophy (PhD)
Three Essays on Empirical Cross-Sectional
Asset Pricing
March 2020

This thesis broadly covers three different topics in empirical cross-sectional asset pricing and consists of three papers. The first paper prices the cross-sectional delta-hedged option and straddle returns in a consumption-based asset pricing model. Delta-hedged options are particularly sensitive to the underlying asset's volatility, which is in turn determined by the fundamental consumption volatility. The strong connection between delta-hedged options and consumption volatility provides us with powerful test assets to identify the consumption volatility premium and hence the preference of the representative agent. As indicated by our results, exposures to consumption growth, expected consumption growth, and consumption volatility are all significantly priced in the cross-section of delta-hedged option and straddle returns. Consumption growth and expected consumption growth command positive risk premiums, whereas consumption volatility commands a negative risk premium, suggesting that investors prefer early resolution of uncertainty. Our results further suggest that consumption risk exposures provide rational foundations for well-known relations between option moneyness or idiosyncratic underlying-stock volatility and the cross-section of delta-hedged option or straddle returns.

The second paper relies on a hazard-model prediction of failure as proxy for firm-level distress risk. The paper discovers a significantly negative relation between firm-level distress risk and the cross-section of corporate bond returns, which is analogous to the often negative relation between distress risk and stock returns found in prior studies ("distress anomaly"). Our finding casts doubts on theories arguing that the distress anomaly arises due to shareholders shifting financial risk onto debtholders in distress. In accordance, proxy variables suggested by such theories do not condition the distress risk-bond return relation. Theories suggesting that distressed firms own valuable disinvestment options and thus have a low levered asset risk are more promising to explain the anomaly, with some of the proxy variables suggested by these theories conditioning the former relation.

The third paper evaluates the prediction performance of machine learning methods in predicting the cross-sectional bond returns out-of-sample. Recent studies show that machine learning methods, especially neural networks, perform well in predicting the cross-sectional stock returns when the number of predictors is large. Prior research indicate that bond returns can be predicted by not only macroeconomic factors, bond market factors, and bond-level characteristics, but also stock market factors and stock-level characteristics. Therefore, the number of predictors in the bond market is even larger than that in the stock market, and the advantage of machine learning will be more pronounced in forecasting bond returns. In this work, I show that machine learning methods perform much better than the simple linear model in predicting bond returns out-of-sample.

Declaration

I, Shuwen Yang, declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

At the very beginning, I would like to express my great thanks to my supervisors, Prof. Hening Liu and Prof. Kevin Aretz. They provide generous help to me in my research, and teach me the way to conduct research and polish papers. Without their guidance, I could not be able to complete this thesis. In my first research paper (Chapter 2), *Early Resolution of Uncertainty: Evidence from Equity Options*, Prof. Hening Liu teaches me how to estimate a Markov-switching process and solve the consumption-based asset pricing model adopted in the paper. He also simulates how consumption risks influence delta-hedged option returns inside the model. Prof. Kevin Aretz teaches me the way to use Stata to filter options data, calculate raw option returns and conduct empirical analysis. In my second research paper (Chapter 3), *Switching Perspective: How Does Firm-Level Distress Risk Price the Cross-Section of Corporate Bond Returns?*, Prof. Kevin Aretz estimates firm-level distress risk by employing the Campbell et al.'s (2008) hazard model and simulates the relations between distress risk and expected asset, stock, and debt returns inside the shareholder advantage model and the real options model.

Secondly, I am thankful for conferences and seminars participants who give me useful comments on my first and second chapters in the following conferences and seminars. My first research paper has been presented in *the European Meeting of the Econometric Society*, *the China Finance Review International Conference*, *the International Accounting and Finance Doctoral Symposium*, *a Seminar in the Shanghai University of Finance and Economics*, *the Econometric Conference on Big Data and AI in Honor of Ronald Gallant*, *the Guanghua International Symposium on Finance*, *the Lingnan Macro Workshop "Financial Risks and Macroprudencial Policy"*, *the NWSSDTP Job Market and Employability Skills Workshop for Accounting and Finance*, *the China Meeting of the Econometric Society*, *the "Frontiers of Factor Investing" Conference*. My second research paper has been presented and discussed in *the Young Finance Scholars' Conference*, *the International Symposium in Finance*, and *the Paris December Finance Meeting*.

Thirdly, I want to thank my colleagues and friends, Wei Liu, Yue Ye, Xinyu

Cui, Xiangshang Cai, Adnan Gazi, Nestor Romero Navarro, who offer me kind help, and useful suggestions and information during my four years of study.

Finally, I am particularly grateful for my family members, who give me unconditional support and encouragement during my study years in United Kingdom.

Shuwen Yang

Manchester, United Kingdom

Chapter 1

Introduction

The core of empirical cross-sectional asset pricing is to investigate whether assets with different risk exposures and characteristics are priced differently and thus earn different rates of return. The most famous research in this field are Fama and French (1992) and Fama and French (1993), which claim that stocks with larger exposure to the market risk, higher ratio of book-to-market equity, and smaller equity size will have higher returns. This thesis covers three different topics in the field of empirical cross-sectional asset pricing. The first paper (Chapter 2) identifies three consumption risks which can price cross-sectional delta-hedged option and straddle returns. Empirical results in Chapter 2 provide strong evidence to support that the representative agent prefers early resolution of uncertainty, which is a long-lasting debate in the consumption-based asset pricing literature. Moreover, different loadings on consumption risks of delta-hedged options can explain the well-known anomalies in the option market, which are the negative relation between idiosyncratic underlying-stock volatilities and delta-hedged option returns and the positive relation between option moneyness and delta-hedged option returns. The second paper (Chapter 3) investigates how firm-level distress risk is priced in the cross-sectional corporate bond returns. We discover a novel negative relation between distress risk and the cross-section of corporate bond returns which can not be explained by the shareholder advantage theory. The shareholder advantage theory is well-accepted in the literature to explain the hump-shaped or negative relation between distress risk and stock returns. Moreover, we

find that valuable disinvestment options of firms in distress are able to explain the negative relations between distress risk and both stock and bond returns. The third paper (Chapter 4) uses a collection of machine learning models to predict the cross-sectional corporate bond returns out-of-sample. I show that when the number of predictors is large, machine learning models perform much better than the simple linear model. The following sections will discuss each chapter of this thesis with more details.

1.1 Consumption-Based Asset Pricing Models and Delta-Hedged Options

In the consumption-based asset pricing framework, only consumption risks can affect asset prices. In the standard consumption-based asset pricing model (CCAPM) pioneered by Breeden (1979), asset betas are measured with respect to changes in the aggregate real consumption rate. In the following long-run risk models with recursive utilities and richer specifications of the consumption growth process, expected consumption growth and consumption volatility can also influence asset prices. Moreover, whether investors prefer early resolution of uncertainty decides the signs of risk premiums on expected consumption growth and consumption volatility betas. The long-run risk literature usually assumes a high value for the elasticity of intertemporal substitution and a preference for early resolution of uncertainty, which results in a positive risk premium for expected consumption growth beta and a negative risk premium for consumption volatility beta. (See Bansal and Yaron (2004), Lettau, Ludvigson and Wachter (2008), and Bansal, Kiku, Shaliastovich and Yaron (2014).)

The majority of previous research in the consumption-based asset pricing literature mainly focus on how consumption risks affect stock returns. However, in its essence, the consumption-based asset pricing framework can be applied to all traded assets. We consider delta-hedged options and straddles as alternative test assets because we discover in our simulation work that delta-hedged options returns

are more sensitive to consumption volatility than stocks returns. Therefore, using delta-hedged options and straddles may be able to better identify the consumption volatility risk premium.

In Chapter 2, we price the cross-section of delta-hedged option and straddle returns in a consumption-based asset pricing model, where the representative agent prefers early resolution of uncertainty and the consumption growth process follows a four-state Markov-switching process. By linearizing the pricing kernel, we identify three consumption risks, consumption growth risk, expected consumption growth risk, and consumption volatility risk, to price the cross-sectional delta-hedged option and straddle returns. We find that the consumption growth beta and the expected consumption growth beta command significantly positive risk premiums, while the consumption volatility beta commands a significantly negative risk premium.

Since we adopt the model established by Lettau et al. (2008), we do not contribute to the theoretical models, and our contributions mainly focus on the empirical analysis. We contribute to the literature in three aspects. First, our finding of a large, positive and significant risk premiun for the consumption growth beta provides strong support for the CCAPM. Second, our empirical results of a positive risk premium for the expected consumption growth beta and a negative risk premium for the consumption volatility beta further support the long-run risk models. Third and most important, we provide macroeconomic explanations for the well-known anomalies in the option market, namely, the negative relation between idiosyncratic underlying-stock volatilities and delta-hedged option returns pointed out by Cao and Han (2013), and the positive relation between option moneyness and delta-hedged option returns discovered by Bakshi and Kapadia (2003).

1.2 Distress Risk and Corporate Bond Returns

Previous studies find a flat, hump-shaped, or negative relation between the probablity that a firm fails to fulfill its obligations ("distress risk") and the cross-

section of stock returns (see e.g. Griffin and Lemmon (2002), Campbell, Hilscher and Szilagyi (2008), Garlappi, Shu and Yan (2008), and George and Hwang (2010)), which is the so-called "distress anomaly". The most convincing explanation for the non-positive relation between distress risk and stock returns is Garlappi et al.'s (2008) and Garlappi and Yan's (2011) shareholder advantage theory, which claims that shareholders can extract economic rents from bondholders when a firm is in distress, and thus lower the risks and returns of distressed stocks. Their theory is further supported by evidence from Favara, Schroth and Valta (2012), Hackbarth, Haselmann and Schoenherr (2015), and Aretz, Florackis and Kostakis (2018). Favara et al. (2012) show that in countries whose institutions favor stockholders over debtholders, the stock betas and volatilities are lower, and Aretz et al. (2018) discover that the distress risk-stock return relation is more negative in the same countries. Moreover, Hackbarth et al. (2015) find that an exogenous increase in shareholder advantage in the United States in 1978 lowered stock betas and returns for all but most strongly distressed firms. Those previous research mainly investigate the distress anomaly from the perspective of stockholders. However, we try to re-examine that anomaly from the perspective of bondholders.

In the second paper (Chapter 3), we use Campbell et al.'s (2008) hazard model to capture the probability of failure of firms and discover a significantly negative relation between distress risk and the cross-section of corporate bond returns. We extend the simulation results of Garlappi et al. (2008) to bonds, and find a positive relation between distress risk and bond returns in the shareholder advantage framework. Moreover, we find that shareholder advantage proxies are not able to condition the distress risk-bond return relation empirically. Those findings confirm that shareholder advantage cannot explain our bond pricing evidence. Considering the limited success of shareholder advantage theory, we turn to find other explanations for the distress anomaly in both stocks and bonds. We study a modified version of the standard real options model of Aretz and Pope (2018), which allows for the gradual disinvestment of productive capacity, and our simulation results show that the real options model can produce negative

relations between distress risk and both expected stock and debt returns. Besides, our empirical results show that the disinvestment options proxies can condition the relations between distress risk and both stock and debt returns with correct signs.

Here are our main contributions of Chapter 3. First, we document a negative relation between distress risk and the cross-section of corporate bond returns, which is analogous to the non-positive and often negative relation between distress risk and stock returns. Second, we show that the shareholder advantage theory fails both empirically and theoretically to explain the negative distress risk-bond return relation. Finally, we find that valuable real options of distressed firms can empirically and theoretically explain the distress anomaly from the perspectives of both stockholders and bondholders.

1.3 Machine Learning Models and Corporate Bond Returns

With the development of machine learning techniques, applying machine learning models in finance research becomes a new trend, especially using machine learning models to predict asset returns. Gu, Kelly and Xiu (2020) show that machine learning models offer an improved description of expected stock returns compared with traditional forecasting methods when the number of predictors is huge, and the best performing method, neural networks, improves its predictive power through allowing for nonlinear predictor interactions. Similarly, Bianchi, Büchner and Tamoni (2019) find that neural networks are useful to detect predictable variations in Treasury bond excess returns, and macroeconomics information has substantial out-of-sample forecasting power for Treasury bond excess returns across maturities.

Meanwhile, there is a growing number of papers which try to discover predictors of corporate bond returns. Extant literature show that the cross-section of corporate bond returns can be predicted by not only bond market factors and bond-level

characteristics (see e.g. Gebhardt, Hvidkjaer and Swaminathan (2005), Bao, Pan and Wang (2011), Lin, Wang and Wu (2011), Jostova, Nikolova, Philipov and Stahel (2013), and Bai, Bali and Wen (2019)), but also stock market factors and stock-level characteristics (see e.g. Acharya, Amihud and Bharath (2013) , Bongaerts, de Jong and Driessen (2017), Chordia, Goyal, Nozawa, Subrahmanyam and Tong (2017), and Choi and Kim (2018)) and macroeconomic factors (see e.g. Bali, Subrahmanyam and Wen (2019)). Therefore, the number of predictors of corporate bond returns is even larger than that of stock returns, and the advantage of machine learning models will be more pronounced in forecasting bond returns.

In the third paper (Chapter 4), I try to answer three research questions: (1) whether machine learning methods also possess strong predictive power in predicting corporate bond returns out-of-sample; (2) which predictors play the main roles in predicting bond returns; (3) which predictors can explain the cross-section of both stock and bond returns at the same time. The current version of Chapter 4 is still preliminary, which only uses a small set of bond and stock market predictors to predict the cross-sectional corporate bond returns from July 2006 to June 2017. The machine learning models I use include, the OLS with Huber loss (OLS+H), the partial least squares (PLS), the principal component regression (PCR), the Lasso, the Ridge, the elastic net (Enet), the Group Lasso, the gradient boosted regression trees and the random forests. There are 4 bond market factors, 11 bond-level characteristics, and 6 stock-level characteristics. Considering the interaction terms between bond market factors and bond-level and stock-level characteristics, the number of predictors is 85 in total. Among all those machine learning models, the random forests perform the best with the highest out-of-sample predictive R^2, and all the machine learning models perform better than the simple linear model. Therefore, I can answer the research question (1) and conclude that machine learning methods do possess stronger predictive power in predicting corporate bond returns out-of-sample than the simple linear model.

The contribution of Chapter 4 based on current results is that machine learning

models can better predict corporate bond returns out-of-sample than the traditional simple linear model.

1.4 Thesis Structure

This thesis has three self-contained journal-format essays, Chapter 2, Chapter 3, and Chapter 4. Each chapter has its own introduction, model and methodology, data and empirical analysis, conclusion, references, and appendix parts. The equations, tables, figures, and footnotes are listed in sequential orders throughout the thesis.

The thesis is organized as follows. Chapter 2 prices the cross-sectional delta-hedged option and straddle returns in a consumption-based asset pricing model where the representative agent has recursive utilities and the consumption growth follows a four-state Markov-switching process. Chapter 3 investigates the relation between distress risk and the cross-section of corporate bond returns, and provide possible explanations for that relation. Chapter 4 applies machine learning models in predicting the cross-sectional corporate bond returns out-of-sample. Chapter 5 concludes major findings of the thesis.

Chapter 2 is co-authored with Prof. Hening Liu and Prof. Kevin Aretz at the Alliance Manchester Business School and Dr. Yuzhao Zhang at the Rutgers Business School, and Chapter 3 is co-authored with Prof. Kevin Aretz at the Alliance Manchester Business School. Chapter 4 is finished completely by myself.

References

Acharya, V.V., Amihud, Y., Bharath, S.T., 2013. Liquidity risk of corporate bond returns: conditional approach. Journal of Financial Economics 110, 358–386.

Aretz, K., Florackis, C., Kostakis, A., 2018. Do stock returns really decrease with default risk? new international evidence. Management Science 64, 3821–3842.

Aretz, K., Pope, P.F., 2018. Real options models of the firm, capacity overhang, and the cross section of stock returns. Journal of Finance 73, 1363–1415.

Bai, J., Bali, T.G., Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. Journal of Financial Economics 131, 619–642.

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. Review of Financial Studies 16, 527–566.

Bali, T.G., Subrahmanyam, A., Wen, Q., 2019. Economic uncertainty premium in the corporate bond market. Working Paper.

Bansal, R., Kiku, D., Shaliastovich, I., Yaron, A., 2014. Volatility, the macroeconomy, and asset prices. Journal of Finance 69, 2471–2511.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. Journal of Finance 66, 911–946.

Bianchi, D., Büchner, M., Tamoni, A., 2019. Bond risk premia via machine learning. Working Paper.

Bongaerts, D., de Jong, F., Driessen, J., 2017. An asset pricing approach to liquidity effects in corporate bond markets. Review of Financial Studies 30, 1229–1269.

Breeden, D.T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 265–296.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. Journal of Finance 63, 2899–2939.

Cao, J., Han, B., 2013. Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics 108, 231–249.

Choi, J., Kim, Y., 2018. Anomalies and market (dis) integration. Journal of Monetary Economics 100, 16–34.

Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A., Tong, Q., 2017. Are capital market anomalies common to equity and corporate bond markets? an empirical investigation. Journal of Financial and Quantitative Analysis 52, 1301–1342.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Favara, G., Schroth, E., Valta, P., 2012. Strategic default and equity risk across countries. Journal of Finance 67, 2051–2095.

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

Garlappi, L., Yan, H., 2011. Financial distress and the cross-section of equity returns. Journal of Finance 66, 789–822.

Gebhardt, W.R., Hvidkjaer, S., Swaminathan, B., 2005. Stock and bond market interaction: Does momentum spill over? Journal of Financial Economics 75, 651–690.

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

Griffin, J.M., Lemmon, M.L., 2002. Book-to-market equity, distress risk, and stock returns. Journal of Finance 57, 2317–2336.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Hackbarth, D., Haselmann, R., Schoenherr, D., 2015. Financial distress, stock returns, and the 1978 bankruptcy reform act. Review of Financial Studies 28, 1810–1847.

Jostova, G., Nikolova, S., Philipov, A., Stahel, C., 2013. Momentum in corporate bond returns. Review of Financial Studies 20, 1649–1693.

Lettau, M., Ludvigson, S.C., Wachter, J.A., 2008. The declining equity premium: What role does macroeconomic risk play? Review of Financial Studies 21, 1653–1687.

Lin, H., Wang, J., Wu, C., 2011. Liquidity risk and expected corporate bond returns. Journal of Financial Economics 99, 628–650.

Chapter 2

Early Resolution of Uncertainty: Evidence from Equity Options

We offer evidence that exposures to consumption growth, expected consumption growth, and consumption volatility are significantly priced in the cross-section of delta-hedged option and straddle returns. Consumption growth and expected consumption growth command a positive risk premium, whereas consumption volatility commands a negative risk premium. In the context of a representative-agent economy with Epstein-Zin (1989) recursive preferences, our results suggest that investors prefer early resolution of uncertainty. Our results further suggest that consumption risk exposures provide rational foundations for well-known relations between option moneyness or idiosyncratic underlying-stock volatility and the cross-section of delta-hedged option or straddle returns.

KEYWORDS: Consumption growth, option returns, recursive utility, volatility risk.

2.1 Introduction

It is of central importance to understand how consumption risks influence asset prices. In the standard consumption-based asset pricing model (CCAPM) pioneered by Breeden (1979), the risk premium on an asset is a multiple of its exposure to consumption risk, the covariance of the asset return with contemporaneous aggregate consumption growth. In long-run risk models with recursive preferences and richer dynamics of consumption growth (e.g., Bansal and Yaron (2004)), the expected consumption growth and consumption volatility are also priced. Moreover, the signs of the risk premiums on the mean and volatility of consumption growth crucially depend on whether investors prefer early resolution of uncertainty. The long-run risk literature commonly postulates that the elasticity of intertemporal substitution (EIS) for a representative agent is sufficiently high such that the agent prefers early resolution of uncertainty. The resulting premiums on the expected consumption growth and consumption volatility are then, respectively, positive and negative (see Bansal and Yaron (2004), Lettau, Ludvigson, and Wachter (2008), and Bansal, Kiku, Shaliastovich, and Yaron (2014)).

Most of the early studies in the consumption-based asset pricing literature focus on the impact of the first moment of consumption growth on stocks (e.g., Lettau and Ludvigson (2001), Parker and Julliard (2005), and Yogo (2006)). More recent studies also stress the importance of consumption volatility for stocks (e.g., Bansal, Kiku, and Yaron (2012), Boguth and Kuehn (2013), and Bansal et al. (2014)). Yet, while the prior literature has so far only studied the ability of consumption risks to price stocks, the consumption-based framework is, in theory, applicable to all traded assets, not only stocks. In our paper, we thus evaluate the ability of consumption risks to price alternative assets, namely delta-hedged call and put options as well as straddles. As shown in our impulse response analysis described in Section 2.3, delta-hedged options and straddles returns positively respond to changes in consumption volatility and negatively to shocks in expected consumption growth while stock returns do in contrast ways, making them interesting alternative assets to consider. Furthermore, when there are only shocks to consumption volatility,

delta-hedged options returns are more sensitive to consumption volatility than stocks returns. Therefore, using delta-hedged options and straddles may be able to better identify the consumption volatility risk premium.

Using our alternative test assets, we identify highly significant consumption growth, expected consumption growth, and consumption volatility risk premiums that support the standard CCAPM with constant relative risk aversion and the model with Epstein and Zin's (1989) recursive utility featuring a preference for early resolution of uncertainty. More importantly, we show that consumption risks offer rational explanations for well-known anomalies in delta-hedged options, as, for example, the moneyness and idiosyncratic underlying-stock volatility anomalies.

To better understand how consumption risks affect options, we study a delta-hedged call option in the representative-agent model of Lettau, Ludvigson, and Wachter (2008). In that model, consumption growth follows a Markov-switching process in which the mean growth rate and the volatility of innovation shocks are characterized by two independent Markov chains, each with two unobservable states. The representative agent has Epstein and Zin's (1989) recursive preferences, allowing his risk attitude and preferences over intertemporal substitution to be disentangled. The log-linearized pricing kernel is an affine function of consumption growth and the change in the wealth-consumption ratio. In turn, the change in the wealth-consumption ratio is approximately affine in the changes in the conditional mean and volatility of consumption growth. Thus, the agent's estimates of the conditional mean and volatility of consumption growth are priced, with the signs of their risk premiums depending on the parameters in the agent's utility function.

When the coefficient of relative risk aversion (RRA) exceeds the reciprocal of the EIS, the agent prefers early resolution of the intertemporal risk arising from unobservable states and thus demands a positive (negative) risk premium for shocks to the conditional mean (volatility).[1] Because of Bayesian learning, a negative shock to consumption growth leads the agent to lower his estimate of mean consumption growth, but to raise his estimate of consumption volatility. Using an impulse response

[1]If the agent prefers late resolution of uncertainty, the risk premiums for the conditional mean and volatility of consumption growth switch signs.

analysis, we show that this shock decreases the stock price, increases stock volatility, and lowers the price of a call option written on the stock despite raising the option's implied volatility. The call option price, however, decreases less than the gain from shorting delta stocks because the call option price is convex in the underlying stock's price and the implied volatility increases. Thus, the value of the delta-hedged option (which is long the call option and short delta stocks) increases, implying that the delta-hedged option is negatively exposed to consumption growth and mean consumption growth but that it is positively exposed to consumption volatility.

Motivated by our theoretical analysis, we estimate the Markov-switching model and obtain estimates of the conditional mean and volatility of consumption growth. We then test whether consumption risks are priced in option returns. To match quarterly consumption data with options data, we choose a cross-section of at-the-money (ATM) options with times-to-maturity between about three to six months at the end of each quarter. Our time-to-maturity choice guarantees that the options expire after the end of the coming quarter and confines the times-to-maturity to be within a reasonable range. We then follow Cao and Han (2013) and compute the quarterly return of a portfolio that buys one call (or put) option and delta-hedges it with the underlying stock. Delta-hedging the option neutralizes the effect of movements in the underlying stock's price on option returns, ensuring that our results are not simply reflecting stock returns. We finally form ten equally-weighted delta-hedged call option portfolios ranked on the underlying stock's idiosyncratic volatility (IVOL). Using idiosyncratic stock volatility as sorting variable is motivated by Cao and Han's (2013) result that this variable is significantly negatively related to delta-hedged option returns.

In line with existing studies, our evidence suggests that the mean returns of all delta-hedged option portfolios are negative. Why? We offer a macroeconomic-based explanation by looking at the portfolios' consumption exposures as well as their exposures with respect to the estimated mean and volatility of consumption growth. Most of the portfolios have negative exposures toward both consumption growth and its expectation, enabling investors to use the portfolios to hedge against unfavorable

macroeconomic conditions. According to the standard CCAPM, investors accept a lower or even negative return on assets offering protection against consumption risk by paying out more in adverse macroeconomic conditions, explaining the negative mean portfolio returns. Moreover, options are particularly sensitive to the underlying stock's volatility. Coval and Shumway (2001) show that volatility risk is priced in options, whereas Cao and Han (2013) and Hu and Jacobs (2020) find that the cross-section of option returns is strongly affected by that volatility. Since stock volatility depends on consumption volatility in many theoretical models (e.g., Bansal and Yaron (2004)), it seems likely that consumption volatility risk also matters for options. Indeed, we find that all delta-hedged option portfolios have positive consumption volatility exposures, further suggesting that options are countercyclical assets accommodating investors' hedging needs.

The mean spread return between the highest and lowest IVOL delta-hedged call option portfolio is –2.97% per quarter (t=–6.91). While, as already said, most portfolios have negative consumption growth exposures, the exposures become more negative with underlying idiosyncratic volatility, implying that options written on higher idiosyncratic volatility stocks offer better protection against adverse conditions. Similarly, the usually positive consumption volatility exposures become more positive with idiosyncratic volatility, suggesting that options written on higher idiosyncratic volatility stocks also protect against high consumption-volatility states. We next show that the three consumption exposures are all priced in Fama-MacBeth (Fama and MacBeth (1973), henceforth FM) regressions. The consumption growth risk premium is positive and significant, and the product of average consumption growth exposure over the portfolios (–0.019) and that premium (0.57) is –0.62% per quarter. The significant consumption risk premium is consistent with Jagannathan and Wang (2007). The risk premium of expected consumption growth, which is considered to be the long-run risk component in Bansal and Yaron (2004), is also positive and significant, and the product of average exposure and premium is –0.10% per quarter in its case. Finally, consistent with Boguth and Kuehn (2013), the consumption volatility risk premium is negative and significant, and the product

of average exposure and premium is –0.68% per quarter in its case. The three sources of consumption risk all contribute to the negative mean delta-hedged option returns and explain an average of over 50% of the cross-sectional variations in these returns. Also, the intercepts from the FM regressions are all only insignificantly different from zero, supporting the consumption-based model.

In our model, the stochastic discount factor (SDF) is approximately affine in consumption growth, the change in its expectation, and the change in its volatility. We use Hansen's (1982) generalized method of moments (GMM) to test this Euler-equation implication. Using the delta-hedged call option portfolios as test assets, we find that consumption growth and its volatility are significantly priced, with risk premiums that are quantitatively similar to those obtained from the FM regressions. The overidentifying restrictions test fails to reject the model, and the cross-sectional R^2 is above 85%. In accordance, the mean absolute pricing error and root mean squared error are a modest 0.27% and 0.32% per quarter, respectively. Most crucially, the risk premiums of consumption growth and its expectation continue to be positive, while the consumption volatility premium continues to be negative, again suggesting the EIS exceeds the inverse of the RRA. Given the EIS-RRA relation is crucial for a large literature on long-run risks, our evidence critically informs that literature.

Bakshi and Kapadia (2003) show that delta-hedged option returns measure the variance risk premium, defined as the difference between a stock's realized variance (RV) and the implied variance paid for an option written on that stock at the start of the realized variance period (IV). While options written on high IVOL stocks are indeed more expensive in terms of implied variance than those on low IVOL stocks, realized variance increases, on average, less steeply with idiosyncratic volatility, implying a more negative variance risk premium for options on higher IVOL stocks. In turn, the options on higher IVOL stocks earn lower delta-hedged returns. Interestingly, the variance risk premium is time-varying and covaries with the state of the economy. In particular, the premium is negatively correlated with consumption growth, with the correlation being most negative for high

IVOL stocks. Thus, options written on the highest IVOL stocks perform best in adverse economic conditions, explaining why they offer the best hedge against such conditions.

While we use call options in our main tests, we find similar results when switching to put options or straddles. We also find similar results when studying the monthly delta-hedged returns of shorter maturity options, showing that our results are robust to the choice of option maturity and return interval. Motivated by Bakshi and Kapadia's (2003) conclusions, we further corroborate our evidence by showing that the risk premiums estimated from delta-hedged option portfolios based on moneyness are close to those estimated from the IVOL portfolios.

Our work adds to the literature relating consumption to asset prices. Jagannathan and Wang (2007) point out that the CCAPM explains stock returns when consumption exposures are computed using December-to-December consumption growth. Bansal et al. (2012) evaluate the ability of the long-run risks model to explain asset returns, highlighting the importance of low-frequency movements and time-varying uncertainty in economic growth for understanding risk-return tradeoffs. Bansal, Khatchatrian, and Yaron (2005) find that consumption volatility predicts and is predicted by valuation ratios at long horizons. Lettau et al. (2008) study a consumption-based model with Markov-switching consumption growth, showing that learning about volatility regimes can explain the decrease in the equity risk premium during the 1990s. Eraker and Shaliastovich (2008) examine equilibrium models with Epstein-Zin preferences in a framework in which exogenous state variables follow affine jump diffusion processes. Calvet and Fisher (2007) introduce a parsimonious equilibrium model with regime shifts of heterogeneous durations in fundamentals. Romeo (2015) shows that changes in consumption volatility are the key driver for explaining major asset pricing anomalies across risk horizons. Boguth and Kuehn (2013) show that consumption volatility is negatively priced in stock portfolios. We contribute to this literature by more precisely estimating the risk premiums of consumption growth and its first two moments using delta-hedged options and straddles.

Some recent papers extend the long-run risk framework to investigate the pricing implications of time-varying uncertainty for the variance risk premium. Drechsler and Yaron (2011) report that time-varying economic uncertainty and a preference for early resolution of uncertainty are required to generate a positive time-varying variance risk premium predicting excess stock market returns. Bansal and Shaliastovich (2011) model investors' optimal decisions to identify the unobserved state. Their model predicts that income volatility (but not income growth) predicts future jump periods. Bollerslev, Tauchen, and Zhou (2009) study the volatility of volatility and the variance risk premium.

There is also a large literature on the cross-section of option returns. Cao and Han (2013) show that delta-hedged option returns decrease with the underlying stock's idiosyncratic volatility. They argue that options written on high-idiosyncratic-volatility stocks are more difficult to hedge, inducing dealers to charge a higher premium in the presence of limits to arbitrage. The question remains why investors are willing to pay that higher premium. We complement their explanation by showing that the options written on high-idiosyncratic-volatility stocks provide a better hedge against adverse macroeconomic conditions, making investors willing to accept low or negative returns. Hu and Jacobs (2020) show that returns on call (put) stock-option portfolios decrease (increase) with underlying stock volatility.[2] We contribute to the literature by studying the cross-section of option returns under the long-run risk framework, identifying links between option returns and covariance risk with respect to consumption growth, mean consumption growth, and consumption volatility.

Our paper is organized as follows. Section 2.2 introduces the theoretical framework motivating our empirical work. In Section 2.3, we conduct a numerical analysis of that framework to understand how consumption risks affect delta-hedged option returns within it. In Section 2.4, we test whether loadings on consumption growth and changes in its conditional moments forecast the cross-section of delta-hedged option and straddle returns. Section 2.5 summarizes and concludes. The Appendix

[2] Aretz, Lin, and Poon (2019) use an SDF model to illustrate that expected European option returns are not unambiguously related to their underlying asset's volatility, with the sign of the relation depending on the option's moneyness.

contains additional derivations and empirical results.

2.2 The Model

In this section, we introduce the consumption-based asset pricing model of Lettau et al. (2008). In that model, the growth rate of consumption follows a Markov-switching process, and the representative agent has the recursive preferences of Epstein and Zin (1989). We next follow Boguth and Kuehn (2013) in linearizing the pricing kernel to derive an equation for expected returns. We do not make any changes to the model, and just use the model to guide our empirical analysis.

2.2.1 Consumption Dynamics

Following McConnell and Perez-Quiros (2000) and Lettau et al. (2008), we assume that the log consumption growth rate follows a Markov-switching process in which the conditional mean and volatility states follow two independent Markov chains. More specifically, we assume that the log consumption growth rate, Δc_{t+1}, follows:

$$\Delta c_{t+1} \equiv \ln\left(\frac{C_{t+1}}{C_t}\right) = \mu_t + \sigma_t \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0,1), \tag{2.1}$$

where C_t is consumption at time t, and μ_t the conditional mean and σ_t the conditional volatility of the log consumption growth rate. We assume two states for the mean growth rate, $\mu_t \in \{\mu_l, \mu_h\}$, and two states for the volatility of the growth rate, $\sigma_t \in \{\sigma_l, \sigma_h\}$. The transition matrices for the mean and volatility states are $\boldsymbol{P}^{\mu}$ and $\boldsymbol{P}^{\sigma}$, respectively. The two matrices are given by:

$$\boldsymbol{P}^{\mu} = \begin{bmatrix} p^{\mu}_{ll} & 1 - p^{\mu}_{hh} \\ 1 - p^{\mu}_{ll} & p^{\mu}_{hh} \end{bmatrix}, \quad \boldsymbol{P}^{\sigma} = \begin{bmatrix} p^{\sigma}_{ll} & 1 - p^{\sigma}_{hh} \\ 1 - p^{\sigma}_{ll} & p^{\sigma}_{hh} \end{bmatrix}. \tag{2.2}$$

Since the mean and volatility states switch independently, the joint transition matrix is the product of the marginal transition probabilities and can be fully characterized by p^{μ}_{ll}, p^{μ}_{hh}, p^{σ}_{ll}, and p^{σ}_{hh}.

We assume that the agent does not observe the state of the economy and must

infer it from available consumption growth data. The posterior belief that the economy is in specific states at date $t+1$ conditional on observations available until date t is denoted by the vector $\boldsymbol{\xi}_{t+1|t}$. Bayesian inference implies that the belief vector evolves according to:

$$\boldsymbol{\xi}_{t+1|t} = \boldsymbol{P}\frac{\left(\boldsymbol{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}{\mathbf{1}'\left(\boldsymbol{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}, \tag{2.3}$$

where $\boldsymbol{\eta}_t$ is a vector of conditional Gaussian densities, $\odot$ represents element-by-element multiplication, $\boldsymbol{P} = \boldsymbol{P}^{\mu} \otimes \boldsymbol{P}^{\sigma}$ is the joint transition matrix, and $\otimes$ is the Kronecker product. Despite the mean and volatility states switching independently, Lettau et al. (2008) and Boguth and Kuehn (2013) stress that Bayesian learning implies that the agent's beliefs about those states are dependent.

2.2.2 The Pricing Kernel

The agent's preferences obey Epstein-Zin's (1989) recursive utility function, given by:

$$U_t = \left[(1-\beta)\, C_t^{1-\frac{1}{\psi}} + \beta \left[E_t\left(U_{t+1}^{1-\gamma}\right)\right]^{\frac{1-\frac{1}{\psi}}{1-\gamma}}\right]^{\frac{1}{1-\frac{1}{\psi}}}, \tag{2.4}$$

where β is the time discount factor, γ the relative risk aversion parameter, ψ the elasticity of intertemporal substitution, and U_{t+1} the continuation value at time $t+1$. The γ and ψ parameters are required to satisfy $\gamma > 0$, $\psi > 0$, and $\psi \neq 1$. For $\gamma = \frac{1}{\psi}$, the representative agent has standard constant relative risk aversion (CRRA) preferences. As demonstrated in the long-run risk literature (e.g., Bansal and Yaron (2004)), $\gamma > \frac{1}{\psi}$ signals that the agent prefers early resolution of uncertainty, whereas $\gamma < \frac{1}{\psi}$ signals that the agent prefers late resolution of uncertainty.

The general asset pricing equation pricing any asset is given by:

$$E_t\left[M_{t+1} R_{i,t+1}\right] = 1, \tag{2.5}$$

where $R_{i,t+1}$ is the gross return on asset i. The equation for an excess return $R^e_{i,t+1}$

is:

$$E_t\left[M_{t+1}R^e_{i,t+1}\right] = 0. \tag{2.6}$$

The SDF under recursive utility, M_{t+1}, is:

$$M_{t+1} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\psi}}\left(\frac{U_{t+1}}{\left(E_t\left[U_{t+1}^{1-\gamma}\right]\right)^{\frac{1}{1-\gamma}}}\right)^{\frac{1}{\psi}-\gamma}. \tag{2.7}$$

Epstein and Zin (1989) show that the wealth-consumption ratio, $Z_t \equiv W_t/C_t$, satisfies:

$$\frac{W_t}{C_t} = \frac{1}{1-\beta}\left(\frac{U_t}{C_t}\right)^{1-\frac{1}{\psi}}, \tag{2.8}$$

so that the SDF, M_{t+1}, can alternatively be expressed as:

$$M_{t+1} = \beta^{\frac{1-\gamma}{1-\frac{1}{\psi}}}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\left(\frac{Z_{t+1}}{Z_t-1}\right)^{\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}}. \tag{2.9}$$

A log-linear approximation of the SDF in Equation (2.9), m_{t+1}, is:

$$m_{t+1} \approx \left(\frac{1-\gamma}{1-\frac{1}{\psi}}\right)\ln\beta - \gamma\Delta c_{t+1} + \left(\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}\right)\Delta z_{t+1}, \tag{2.10}$$

where Δz_{t+1} is the change in log wealth-consumption ratio from time t to $t+1$. The Euler equation defined for wealth implies the following functional equation for $z_t = \log(Z_t)$:

$$E_t\left[\exp\left(\theta\left(\log\beta + \left(1-\frac{1}{\psi}\right)\Delta c_{t+1} + z_{t+1} - \log\left(e^{z_t}-1\right)\right)\right)\right] = 1, \tag{2.11}$$

where $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}$. In an endowment economy in which consumption growth is driven by independent and identically distributed (i.i.d.) shocks, the wealth-consumption ratio is thus constant. When the mean and volatility of consumption growth are time varying, however, the agent's posterior beliefs about the state of the economy become state variables of the model. Consequently, the wealth-consumption ratio becomes a function of the agent's beliefs, $Z_t = Z\left(\boldsymbol{\xi}_{t+1|t}\right)$.

2.2.3 Approximating the Change in the Wealth-Consumption Ratio

We define the posterior belief that the mean or volatility of consumption growth is in the high state at time $t+1$ conditional on the current information set $\mathcal{F}_t$ by:

$$b_{\mu,t} = \Pr\left(\mu_{t+1} = \mu_h | \mathcal{F}_t\right) \quad b_{\sigma,t} = \Pr\left(\sigma_{t+1} = \sigma_h | \mathcal{F}_t\right), \tag{2.12}$$

allowing us to write the perceived mean and volatility as belief-weighted averages:

$$\hat{\mu}_t = b_{\mu,t}\mu_h + (1 - b_{\mu,t})\,\mu_l \quad \hat{\sigma}_t = b_{\sigma,t}\sigma_h + (1 - b_{\sigma,t})\,\sigma_l. \tag{2.13}$$

We define the changes in the perceived moments by:

$$\Delta\hat{\mu}_t = \hat{\mu}_t - \hat{\mu}_{t-1} \quad \Delta\hat{\sigma}_t = \hat{\sigma}_t - \hat{\sigma}_{t-1}. \tag{2.14}$$

In our empirical analysis, we assume that changes in the log wealth-consumption ratio are approximately linear in changes in the perceived moments:

$$\Delta z_t \approx \kappa + A\Delta\hat{\mu}_t + B\Delta\hat{\sigma}_t. \tag{2.15}$$

In Table A1, we show that the approximation is highly accurate for various combinations of parameter values, confirming the numerical results of Boguth and Kuehn (2013).[3] When $\psi > 1$, the slope coefficients of $\Delta\hat{\mu}$ and $\Delta\hat{\sigma}$ are positive and negative, respectively (and vice versa).

2.2.4 Asset Pricing Implications

We next illustrate that, in our model, the first two conditional moments of consumption growth price assets. To do so, we plug approximation (2.15) into Equation (2.10), then plug Equation (2.10) into Equation (2.6), and finally use the

[3] We approximate the wealth-consumption ratio with Chebyshev polynomials and rely on function iterations to find the fixed-point of the wealth-consumption ratio implied by the Euler equation. Using linear interpolation produces very similar results.

definition of covariance. It is then obvious that asset i's expected excess return, $E_t\left[R^e_{i,t+1}\right]$, is (approximately) linear in the asset's consumption growth exposure, $\beta^i_{\Delta c,t}$, mean consumption growth exposure, $\beta^i_{\Delta\mu,t}$, and consumption volatility exposure, $\beta^i_{\Delta\sigma,t}$:

$$E_t\left[R^e_{i,t+1}\right] \approx \beta^i_{\Delta c,t}\lambda_{\Delta c,t} + \beta^i_{\Delta\mu,t}\lambda_{\Delta\mu,t} + \beta^i_{\Delta\sigma,t}\lambda_{\Delta\sigma,t}, \tag{2.16}$$

where $\lambda_{\Delta c,t}$, $\lambda_{\Delta\mu,t}$, and $\lambda_{\Delta\sigma,t}$ are risk premiums of the three consumption exposures.[4]

Due to risk aversion, the model always predicts a positive risk premium for consumption growth exposure. When the EIS exceeds the inverse of the RRA, the agent prefers early resolution of uncertainty and thus the risk premiums on the conditional mean and volatility of consumption growth are positive and negative, respectively.[5]

2.3 Impulse Response Analysis

In this section, we use an impulse response analysis to find out how consumption risks affect delta-hedged option returns in the representative-agent model in Section 2.2.[6] We first explain how we conduct the impulse response analysis. We next discuss the results from that analysis.

2.3.1 Computing Model-Implied Option Returns

To compute the delta-hedged return, we begin by considering a stock paying dividends that are positively correlated with aggregate consumption. Specifically, we follow Abel (1999) and Bansal and Yaron (2004) and assume that the dividend growth process, Δd_t, is given by:

$$\Delta d_t \equiv \ln\left(\frac{D_t}{D_{t-1}}\right) = \Phi\Delta c_t + g_d + \sigma_d\epsilon_{d,t} \tag{2.17}$$

[4] See Appendix Section 2.6.1 for the derivation and expressions of the risk premiums.

[5] See Table A1 in Appendix Section 2.6.1 for related results.

[6] Developing a full-fledged general equilibrium model with heterogeneous firms is beyond the scope of this paper. We neither aim to match the magnitude of the delta-hedged returns observed in the data.

where D_t is the dividend at time t, Φ is the leverage parameter, Δc_t is the log consumption growth rate, g_d determines mean dividend growth conditional on a zero consumption growth, σ_d is the dividend growth volatility, and $\epsilon_{d,t}$ is a standard normal i.i.d. shock independent of other shocks in the model. In equilibrium, the price-dividend ratio $\frac{S_t}{D_t} \equiv \varphi\left(\boldsymbol{\xi}_{t+1|t}\right)$ satisfies the Euler equation:

$$\frac{S_t}{D_t} = E_t\left[M_{t+1}\left(\frac{S_{t+1}}{D_{t+1}} + 1\right)\frac{D_{t+1}}{D_t}\right]$$

or equivalently,

$$\varphi\left(\boldsymbol{\xi}_{t+1|t}\right) = E_t\left[M_{t+1}\left(\varphi\left(\boldsymbol{\xi}_{t+2|t+1}\right) + 1\right)\exp\left(\Delta d_{t+1}\right)\right]. \tag{2.18}$$

We solve the fixed point of the price-dividend ratio as determined by Equation (2.18) using the linear interpolation method.[7] The model-implied risk-free rate is $r_t^f \equiv \ln\left(R_t^f\right)$, with $R_t^f = 1/E_t\left[M_{t+1}\right]$.

We assume that the option price is equal to the option value implied by the equilibrium model. For instance, the current value of a call option expiring in n periods, $C_t^{(n)}$, is given by the expectation of the option's future cash flow multiplied by the multi-period pricing kernel:

$$C_t^{(n)} = E_t\left[M_{t,t+n}\max\left(0, S_{t+n} - K\right)\right], \tag{2.19}$$

where S_{t+n} is the price of the underlying asset at time $t+n$, K is the option's strike price, and $M_{t,t+n}$ is the multi-period pricing kernel, $M_{t,t+n} = M_{t,t+1}M_{t+1,t+2}\cdots M_{t+n-1,t+n}$ in which $M_{t,t+1}$ is the one-period pricing kernel.

We set the leverage parameter $\Phi = 3$, in line with previous studies such as Bansal and Yaron (2004), Lettau et al. (2008) and Boguth and Kuehn (2013). The parameters g_d and σ_d are set to match the unconditional mean and standard deviation of dividend growth in the post-war data, which yields quarterly values $g_d = -0.009$ and $\sigma_d = 0.028$. Our calibration analysis suggests that the model

[7] We choose 50 grid points on each dimension of the state variables and use function iterations to find the fixed points for both wealth-consumption and price-dividend ratios.

with ($\gamma = 60$, $\psi = 1.5$) and our empirical estimates of the Markov-switching model parameters in Table 2.1 can generate a sizable equity premium of about 4% per year when the model is calibrated at the quarterly frequency. Thus, we choose ($\gamma = 60$, $\psi = 1.5$) as our benchmark case. Since ($\gamma = 60$, $\psi = 1.5$) implies that investors prefer early resolution of uncertainty, we also run simulations with ($\gamma = 60$, $\psi = 0.01$) to study the case in which investors prefer late resolution of uncertainty.

We use Monte Carlo simulations to compute the values of an ATM European call options with three months (twelve weeks) to maturity according to Equation (2.19).[8] The computation involves simulating 40,000 sample paths of stock prices and the multi-period SDF using the parameters that match the aggregate consumption process, as shown in Table 2.1. Along a sample path, we track the contract and compute the option prices given time-t state belief $\boldsymbol{\xi}_{t+1|t}$, the current dividend D_t, and the pre-specified strike price K. We compute the Black-Scholes model implied volatility and delta for each option that is still alive at time t. Because the asset underlying the option is a dividend-paying stock, we make appropriate adjustments to the equilibrium price S_t and use the ex-dividend price in computing the implied volatility and delta of the options.

We calculate the model-implied delta-hedged gain of a call option over its lifetime as:

$$\begin{aligned}\Pi\left(t, t-11\right) &= C_t^{(1)} - C_{t-11}^{(12)} - \sum_{n=0}^{10} \Delta_{c,t-11+n}\left(S_{t-10+n} - S_{t-11+n}\right) \\ &\quad - \sum_{n=0}^{10} r_{t-11+n}^{f}\left(C_{t-11+n}^{(12-n)} - \Delta_{c,t-11+n} S_{t-11+n}\right),\end{aligned}$$

where $C_{t-11}^{(12)}$ is the value of the option when issued, $C_t^{(1)}$ the option value one period (week) before expiration, and $\Delta_{c,t-11+n}$, $C_{t-11+n}^{(12-n)}$, S_{t-11+n}, and r_{t-11+n}^{f} ($n = 0, 1, ..., 10$) are, respectively, the option delta, option value, stock price, and the

[8] Before running Monte Carlo simulations to compute delta-hedged returns, we solve the model (the SDF and price-dividend ratio) numerically at the weekly frequency by appropriately scaling relevant parameters in the model. An alternative approach is to develop a continuous-time asset pricing model with recursive utility and a hidden Markov model. However, (semi)closed-form solutions are not available for such a model. Moreover, because volatility is instantaneously observable in the continuous-time setting, it would be infeasible to analyze the impact of learning about the volatility state on equilibrium prices.

risk-free rate within the horizon of the delta-hedged gain. The value of the delta-hedged option at the start of the horizon is $\Delta_{c,t-11}S_{t-11} - C^{(12)}_{t-11}$. The delta-hedged gain divided by the absolute value of the delta-hedged option portfolio yields the delta-hedged return.

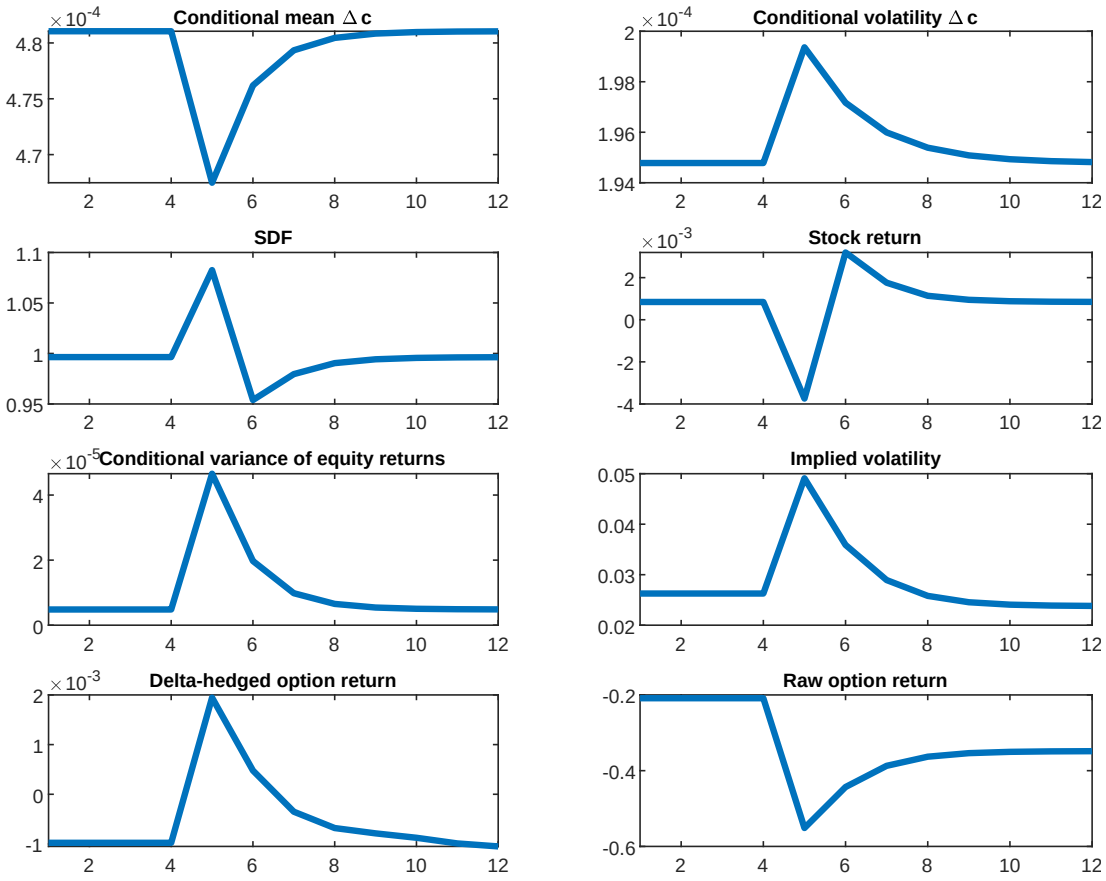
The model-implied delta-hedged return resembles its empirical counterpart. First, the delta-hedged return is an excess return derived from a self-financing strategy. Second, the computation of the delta-hedged gain requires multiple intermediate delta-hedging opportunities within the horizon. Third, the model-implied delta-hedged return crucially depends on stock and option prices determined in equilibrium. As such, we can investigate the mechanism of the model by examining the impact of consumption risks on the pricing kernel and equilibrium asset prices.

2.3.2 Simulation Results

We next discuss the results from our simulations. We perform impulse response analysis to study the impacts of changing beliefs about the consumption growth regimes on the SDF, the stock return, the call option return, and the delta-hedged call option return. First, we assume that consumption growth stays at its long-run mean implied by the estimated Markov-switching model. Due to Bayesian learning, beliefs of consumption growth regimes converge to the stationary level. We then suppose that a negative shock to the expectation of consumption growth occurs in the fifth period. The agent updates beliefs according to Bayes' rule, leading to a decline in the posterior probability of the high mean growth regime and an increase in the posterior probability of the high volatility regime. The agent's expectation of mean consumption growth falls, whereas his expectation of consumption volatility rises. The top two panels of Figure 2.1 display these results.

Assuming ($\gamma = 60$, $\psi = 1.5$), the other panels in Figure 2.1 present the impulse responses of the SDF, the stock return, the conditional variance of the stock return, the implied volatility, the call option return, and the delta-hedged call option return in response to the negative shock. In case of recursive preferences, the continuation value falls as a result of the lower conditional mean and the

Figure 2.1: Impulse Responses: Conditional Mean and Volatility, $\gamma = 60, \psi = 1.5$



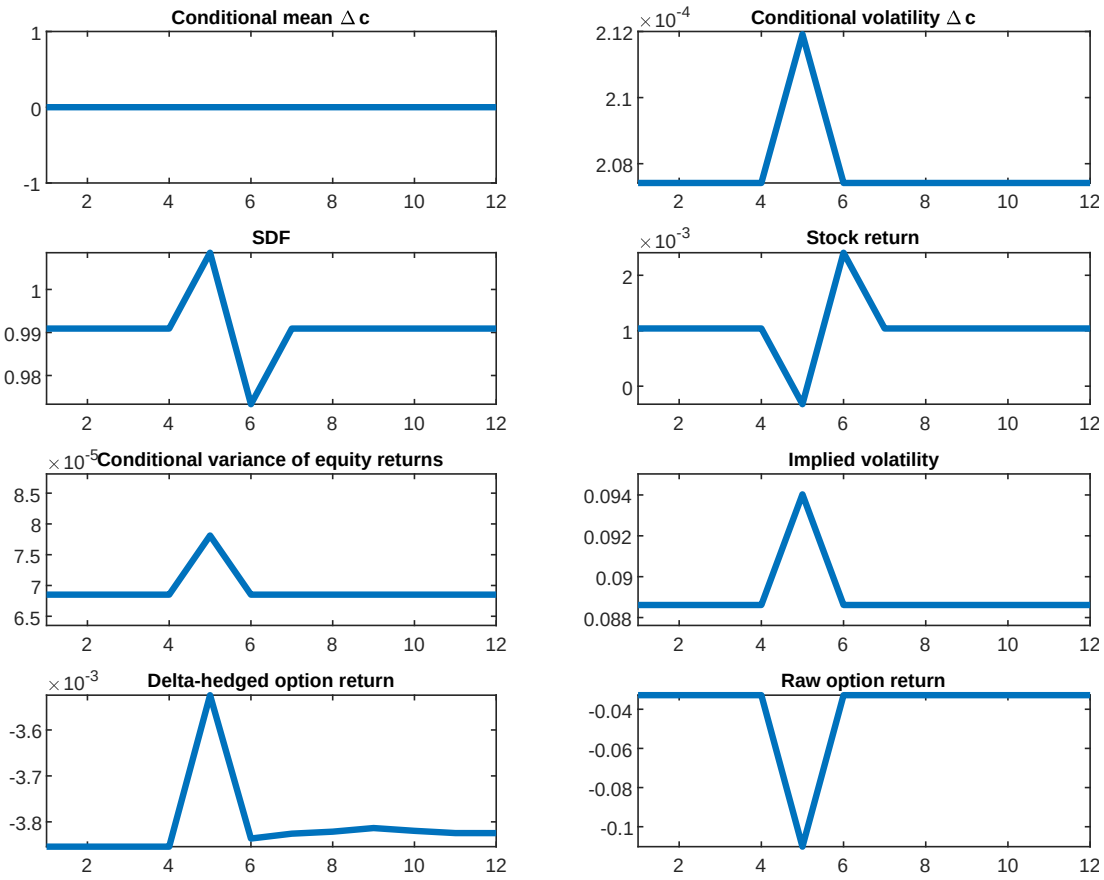
This figure plots the impulse response functions when the growth rate of consumption shifts from its long-run mean to the low mean growth rate. The agent's belief vector $\boldsymbol{\xi}_{t+1|t}$ is updated according to Bayes' rule. The risk aversion parameter is set at $\gamma = 60$, and the EIS parameter at $\psi = 1.5$.

higher conditional volatility of consumption growth. Because the values of γ and ψ imply a preference for early resolution of uncertainty, the SDF rises significantly in response to the shock. Moreover, the stock return drops as the equity value depreciates, while the conditional variance of the stock return rises due to an enhanced pessimism about the state of the economy. The co-movement of the SDF and the conditional stock variance suggests that stock return variance carries a risk premium. This is also evident from the observation that the implied volatility of the call option increases substantially. The lowest panel in the figure shows that the call option return falls because the effect of lower equity value dominates that of higher implied volatility. On the contrary, the delta-hedged call option gain (return) rises due to the elimination of the impact of the underlying

stock price movement on the call option price.

The co-movement of the SDF and the delta-hedged call option return in Figure 2.1 implies that the delta-hedged option enables investors to hedge against systematic risk. Because consumption growth and its conditional mean are negatively related to the SDF, both factors have positive risk premiums. In contrast, since conditional consumption volatility is positively related to the SDF, it has a negative risk premium. These results are consistent with our empirical findings in Section 2.4.3 below. Meanwhile, we can observe that delta-hedged options returns positively respond to shocks to the conditional mean of consumption growth and negatively to shocks to the conditional volatility of consumption growth while stocks returns do in contrast ways.

Figure 2.2: Impulse Responses: Conditional volatility, $\gamma = 60, \psi = 1.5$

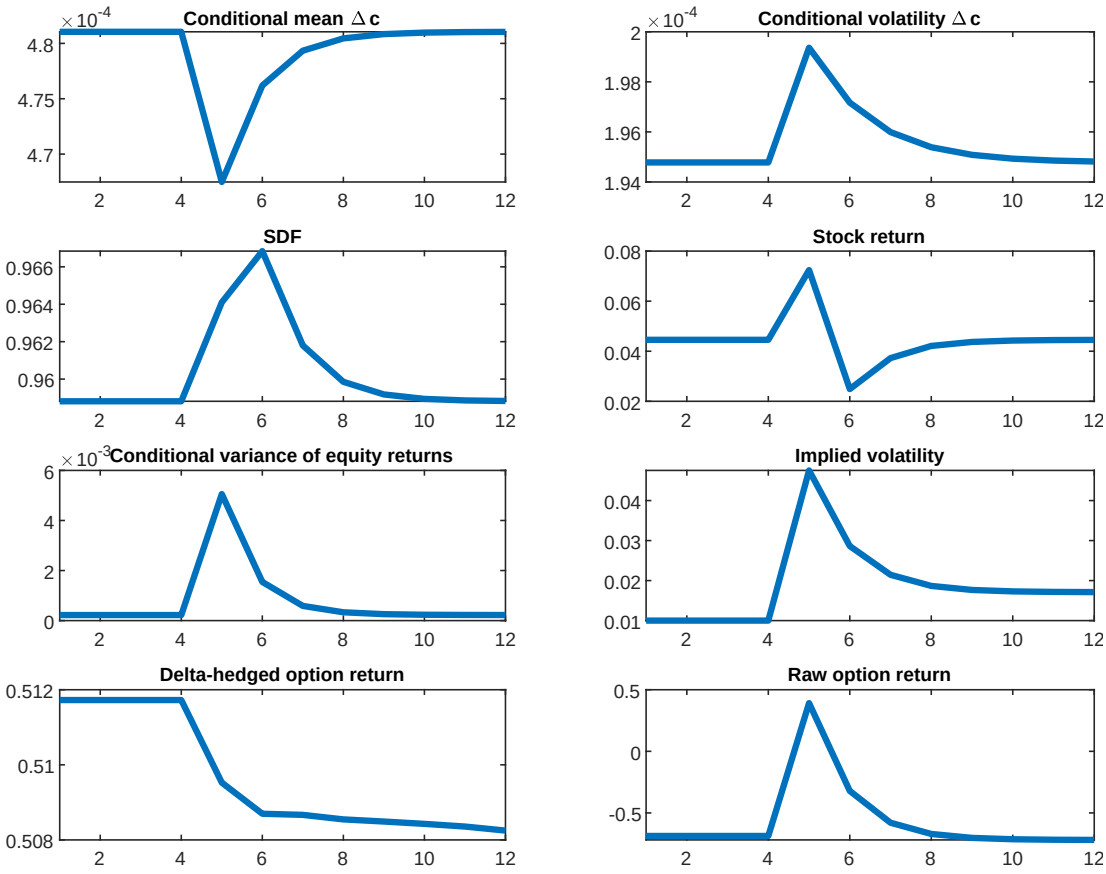


This figure plots the impulse response functions when the conditional volatility of consumption growth rises. The risk aversion parameter is set at $\gamma = 60$, and the EIS parameter at $\psi = 1.5$.

Boguth and Kuehn (2013) find that consumption volatility is important to

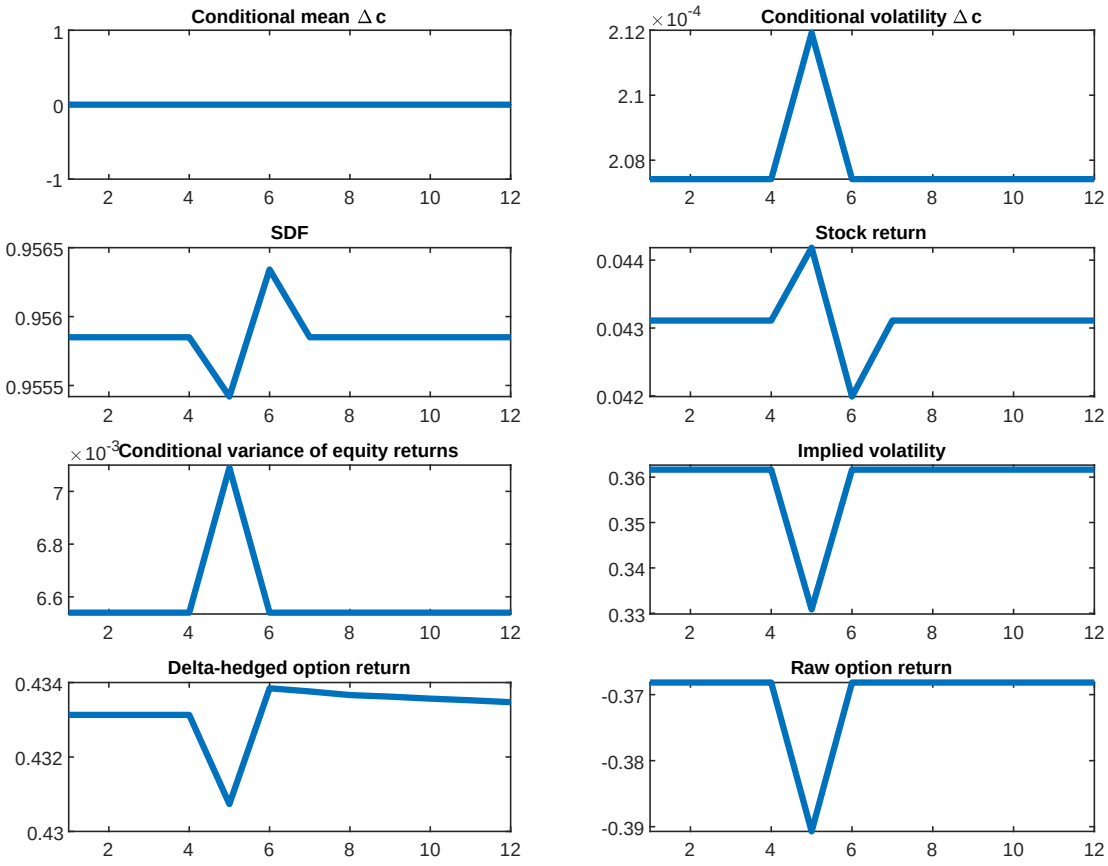
price stocks. To focus on volatility risk on its own, we run a second impulse response analysis by assuming that consumption growth and its expectation remain unchanged as consumption volatility rises. Figure 2.2 shows that the responses of our analysis variables are largely similar to those in the previous case. The rise in conditional volatility alone leads to an increase in the SDF. Also similar to before, both the conditional stock variance and the implied volatility rise in response to the shock. The option return falls, whereas the delta-hedged return rises. Thus, the delta-hedged option represents a hedging opportunity against consumption volatility, and consumption volatility carries a negative risk premium. Moreover, as displayed in Figure 2.2, delta-hedged options returns are more sensitive to shocks to consumption volatility than stocks returns.

Figure 2.3: Impulse Responses: Conditional Mean and Volatility, $\gamma = 60, \psi = 0.01$



This figure plots the impulse response functions when the growth rate of consumption shifts from its long-run mean to the low mean growth rate. The agent's belief vector $\boldsymbol{\xi}_{t+1|t}$ is updated according to Bayes' rule. The risk aversion parameter is set at $\gamma = 60$, and the EIS parameter at $\psi = 0.01$.

Figure 2.4: Impulse Responses: Conditional Volatility, $\gamma = 60, \psi = 0.01$



This figure plots the impulse response functions when the conditional volatility of consumption growth rises. The risk aversion parameter is set at $\gamma = 60$, and the EIS parameter at $\psi = 0.01$.

Figures 2.3 and 2.4 plot simulation results for the ($\gamma = 60$, $\psi = 0.01$) case, in which the agent prefers late resolution of uncertainty. Figure 2.3 shows that the delta-hedged return and SDF move in the opposite directions in response to a negative mean consumption growth shock despite both conditional stock variance and implied volatility rising on impact. Contrary to the ($\gamma = 60$, $\psi = 1.5$) case and our empirical evidence, the delta-hedged call option has a positive exposure to systematic risk in that the return on the delta-hedged option performs poorly when the SDF is high. On the other hand, the stock return increases and co-moves with the SDF. Thus, the implied risk premiums on the conditional mean and volatility of consumption, respectively, have opposite signs compared to the ($\gamma = 60$, $\psi = 1.5$) case. As shown in Figure 2.4, when the conditional volatility of consumption growth rises on its own, implied volatility falls, resulting in a decline in the delta-hedged option return. Because of the agent's preference for late resolution of uncertainty

and the absence of shocks to the level of consumption, the SDF drops in response to the volatility shock. Although the delta-hedged option return and the SDF co-move in the same direction, the implied risk premium on consumption volatility is positive, opposite to that in the ($\gamma = 60$, $\psi = 1.5$) case and in our empirical evidence. In Section 2.4.3, our GMM estimation identifies a negative relation between the SDF and the conditional mean of consumption growth while a positive relation between the SDF and consumption volatility.

2.4 Empirical Tests

In this section, we use options data to estimate the risk premiums of consumption growth, mean consumption growth, and consumption growth volatility risk. We first fit a Markov-switching model to obtain estimates of the conditional mean and volatility of consumption growth. We next sort single-name options into portfolios. We finally use the option portfolios in conjunction with the Markov-switching model estimates to study the pricing of consumption exposures.

2.4.1 Estimating Consumption Dynamics

Defining total consumption as the sum of non-durable goods consumption expenditures and service consumption expenditures, we obtain quarterly per capita real expenditures data on the two total consumption components from the Bureau of Economic Analysis (BEA). Following Yogo (2006), Lettau et al. (2008), and Boguth and Kuehn (2013), we start our consumption data sample in the first quarter of 1952. Conversely, we end the sample in the first quarter of 2018.

We estimate a four-state Markov-switching model on our consumption data. While maintaining the assumption that the agent has preferences over total consumption, we follow Boguth and Kuehn (2013) in separately using non-durable and service consumption expenditures in our estimation to improve state identification and to reduce standard errors. In particular, we assume that both log non-durable goods consumption growth and the log change in the share of non-durable to total consumption follow Markov chains. Doing so, the difference between the two con-

sumption components, log total consumption growth, also follows a Markov chain. More precisely, we express total consumption C_t as non-durable goods consumption N_t divided by the non-durable consumption share V_t. That is, $C_t = N_t/V_t$. Thus, log total consumption growth is log non-durable consumption growth, Δn_t, minus the log change in the non-durable consumption share, Δv_t:

$$\Delta c_{t+1} = \Delta n_{t+1} - \Delta v_{t+1}. \tag{2.20}$$

Given Equation (2.20), we assume that both Δn_{t+1} and Δv_{t+1} follow Markov chains:

$$\Delta n_{t+1} = \mu_t^n + \sigma_t^n \epsilon_{t+1}^n \quad \Delta v_{t+1} = \mu_t^v + \sigma_t^v \epsilon_{t+1}^v, \tag{2.21}$$

where μ_t^k and σ_t^k, with $k \in \{n, v\}$, are, respectively, the conditional expectation and standard deviation of log non-durable consumption growth ($k = n$) or the log change in the non-durable consumption share ($k = v$). Next, ϵ_{t+1}^k is a standard normal residual, with $\text{Cov}_t\left(\epsilon_{t+1}^n, \epsilon_{t+1}^v\right) = \rho_{nv}$. The dynamics specified in Equations (2.20) and (2.21) together with the fact that the information set $\mathcal{F}_t$ contains Δn_t and Δv_t plus their histories also imply a Markov process for total consumption growth, with dynamics specified in Equation (2.1) and $\mu_t = \mu_t^n - \mu_t^v$ and $\sigma_t^2 = (\sigma_t^n)^2 + (\sigma_t^v)^2 - 2\rho_{nv}\sigma_t^n\sigma_t^v$. Thus, the estimates of μ_t^n, μ_t^v, σ_t^n, σ_t^v, and ρ_{nv} allow us to recover the dynamics of log total consumption growth, which we use to solve the consumption-based model in Section 2.2.

Table 2.1 presents the estimates of the Markov chains for log non-durable consumption growth and the log change in the non-durable consumption share. Panel A shows that expected non-durable consumption growth is positive in the high state ($\mu_h^n = 0.58\%$) and negative in the low state ($\mu_l^n = -0.03\%$). State-conditional non-durable consumption volatilities are $\sigma_l^n = 0.40\%$ and $\sigma_h^n = 0.83\%$. The estimated parameters for the non-durable consumption share (shown in Panel B) are $\mu_l^v = -0.16\%$ and $\mu_h^v = 0.00\%$ for the expected drift and $\sigma_l^v = 0.34\%$ and $\sigma_h^v = 0.58\%$ for the conditional volatilities. The correlation between log changes in the two

Table 2.1: Markov Model of Consumption Growth

This table reports parameter estimates for the Markov models fitting log non-durable goods consumption growth, Δn_{t+1}, and changes in the log non-durable consumption share, Δv_t,

$$\Delta n_{t+1} = \mu_t^n + \sigma_t^n \epsilon_{t+1}^n \quad \Delta v_{t+1} = \mu_t^v + \sigma_t^v \epsilon_{t+1}^v,$$

where for $i \in \{n, v\}$, μ_t^i denotes the conditional expectation, σ_t^i denotes the conditional standard deviation, and ϵ_{t+1}^i is standard normal with $\text{Cov}_t\left(\epsilon_{t+1}^n, \epsilon_{t+1}^v\right) = \rho_{nv}$. The conditional first and second moments of both processes switch jointly with transition matrices $\boldsymbol{P}^\mu$ and $\boldsymbol{P}^\sigma$, respectively, given by

$$\boldsymbol{P}^\mu = \begin{bmatrix} p_\mu^{ll} & 1 - p_\mu^{hh} \\ 1 - p_\mu^{ll} & p_\mu^{hh} \end{bmatrix} \quad \boldsymbol{P}^\sigma = \begin{bmatrix} p_\sigma^{ll} & 1 - p_\sigma^{hh} \\ 1 - p_\sigma^{ll} & p_\sigma^{hh} \end{bmatrix}.$$

The estimation procedure follows Hamilton (1994). We use quarterly per capita real consumption expenditures for non-durable goods and services for the years 1952.Q1 to 2018.Q1. t-statistics are reported in parentheses.

Panel A: Non-durable Consumption (%)			
μ_l^n	μ_h^n	σ_l^n	σ_h^n
-0.0269	0.5835	0.4002	0.8255
(-0.57)	(21.51)	(17.67)	(45.73)
Panel B: Non-durable Consumption Share (%)			
μ_l^v	μ_h^v	σ_l^v	σ_h^v
-0.1576	0.0000	0.3422	0.5835
(-4.46)	(0.00)	(17.63)	(22.01)
Panel C: Marginal Transition Probabilities			
p_μ^{ll}	p_μ^{hh}	p_σ^{ll}	p_σ^{hh}
0.87	0.95	0.91	0.91
(18.96)	(47.59)	(28.87)	(24.46)
Panel D: Correlation			
ρ_{nv}			
0.8256			
(45.73)			

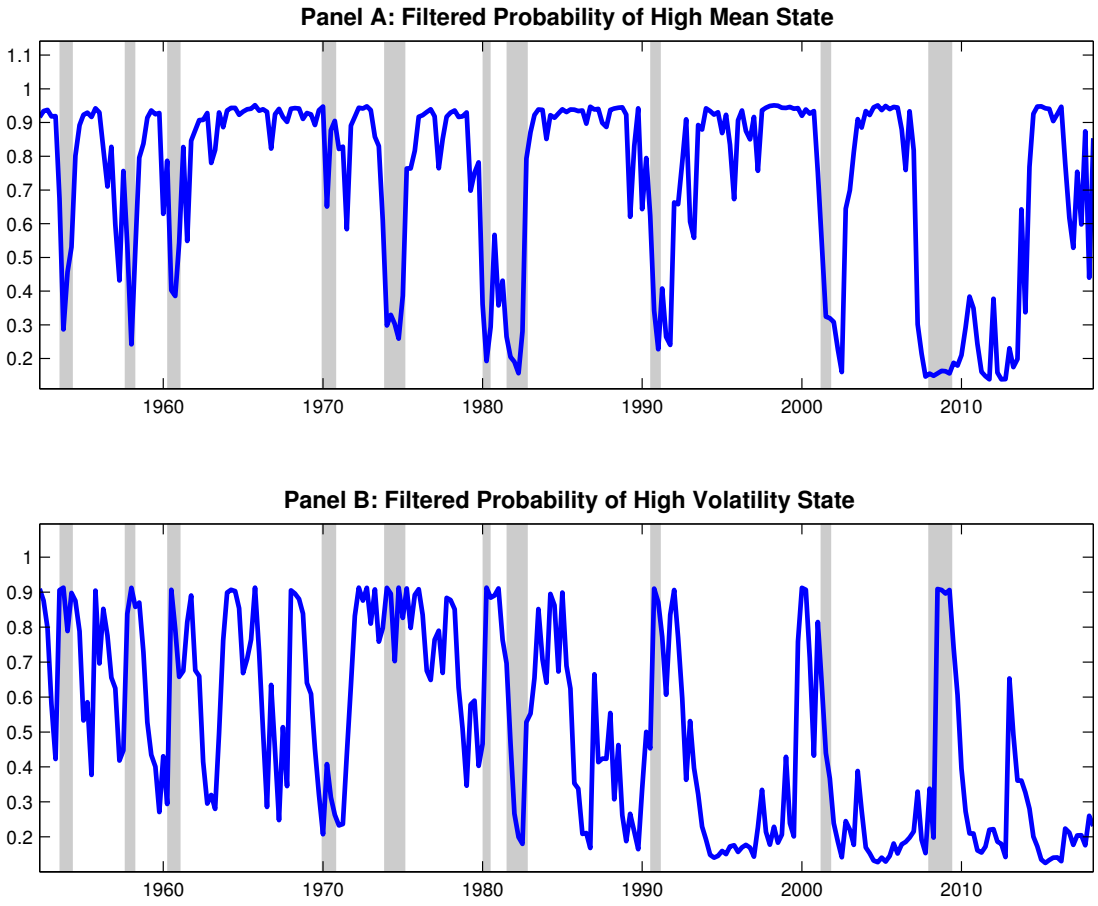
variables is $\rho_{nv} = 0.83$. Turning to total consumption growth, its expected

growth is $\mu_l = 0.13\%$ in the low state and $\mu_h = 0.58\%$ in the high state, while its volatility is $\sigma_l = 0.23\%$ in the low state and $\sigma_h = 0.50\%$ in the high state.

The two transition probabilities for the mean growth regimes, p_μ^{ll} and p_μ^{hh}, are 0.87 and 0.95 respectively. Consistent with Lettau et al. (2008), the high mean state is thus markedly more persistent than the low mean state. Both consumption growth volatility states are persistent, with transition probabilities being about 0.91 for both volatility regimes. Interestingly, these estimates differ a bit from those in Lettau et al. (2008), who find the volatility states to be even more persistent. Given differences in the sample and the consumption measure across the two papers, uncertainty after the financial crisis greatly reduces the persistence of volatility regimes in our analysis.

Figure 2.5 presents the filtered beliefs for the regimes. The upper panel depicts the belief dynamics for the high mean consumption growth regime, $b_{\Delta\mu,t}$, and the lower panel for the high standard deviation regime, $b_{\Delta\sigma,t}$. The grey bars in the graph indicate economic recession periods defined by the National Bureau of Economic Research. The figure further suggests that the low volatility regime becomes more prevalent from 1990 on, as also observed by Kim and Nelson (1999) and Boguth and Kuehn (2013). Despite that, consumption volatility appears to have returned to the high regime during both the 2000-2001 dot-com crash and the 2008-2009 global financial crisis in the post-1990 period. When the economy is in a recession, the probability of being in the high mean state is low, while the probability of being in the high volatility state tends to be high in certain periods. However, their correlation is far from being perfect due to the assumption of independent switching between mean regimes and volatility regimes. Overall, the Markov-switching model captures most recessions during our sample period, with the exceptions of the mild 1969-1970 recession and the 1981-1982 recession caused by contractionary monetary policy.

Figure 2.5: Bayesian Beliefs about the Mean and Volatility State



This figure displays the estimated Bayesian belief processes for being in the high expected growth rate state (top figure) and high volatility state (bottom figure). The estimation procedure follows Hamilton (1994). We use quarterly per capita real consumption expenditure for non-durable goods and services for the years from 1952.Q1 to 2018.Q1. The Markov model is estimated using both components of consumption. The gray bars indicate economic recession periods provided by the National Bureau of Economic Research.

2.4.2 Calculation of Delta-Hedged Returns

We obtain call and put options data over the period from January 1996 to December 2017 from Optionmetrics. The data include the daily closing bid and ask quotes, the trading volume, the strike price, and the maturity date of each option. The data further include each option's delta, calculated by Optionmetrics using standard market conventions, the closing price of and the dividends paid out by the stocks underlying the options, and the risk-free rate of return.

We apply standard filters to the options data (see Goyal and Saretto (2009) and Cao and Han (2013)). First, we exclude an option if the stock underlying the

option pays out dividends over the option's remaining time-to-maturity. Second, we exclude option observations violating well-known arbitrage bounds. More specifically, we exclude an option observation if the option's price does not fulfill $S \geq C \geq \max(0, S - Ke^{-rT})$ (or $K \geq P \geq \max(0, K - S)$) where C (or P) is the call (or put) option's price, S the underlying stock's price, K the strike price, T the option's time-to-maturity, and r the risk-free rate of return. Third, we only retain option observations with positive trading volume, a positive bid quote, a bid price strictly smaller than the ask price, and a bid-ask midpoint of at least 1/8. Finally, we only keep option observations whose last trade date matches the record date and whose option price date matches the underlying stock's price date.

We use quarterly delta-hedged option returns in our main tests, and monthly delta-hedged option returns in robustness tests. In either case, we calculate the return from the start of a calendar quarter or month to its end. In line with Bakshi and Kapadia (2003) and Cao and Han (2013), we define the delta-hedged option return as the delta-hedged option gain over the period scaled by the absolute value of the delta-hedged option at the start of the period, where the delta-hedged option is a self-financing portfolio consisting of a long (call or put) option, a hedging position in the underlying stock, and a money market investment. The value of a perfectly delta-hedged option would be insensitive to changes in the value of the underlying stock. Assuming that the delta-hedge is re-balanced at the end of every trading day, we calculate the delta-hedged call option gain over the quarter or month starting at time $t-1$ and ending at time t, $\Pi(t-1, t)$, as:

$$\begin{aligned}\Pi(t-1,t) = C_t - C_{t-1} - \sum_{n=0}^{N-1} \Delta_{c,t_n} \left[S(t_{n+1}) - S(t_n)\right] \\ - \sum_{n=0}^{N-1} \frac{a_n r_{t_n}}{365} \left[C(t_n) - \Delta_{c,t_n} S(t_n)\right], \end{aligned} \tag{2.22}$$

where C_t is the call option price at time t, Δ_{c,t_n} the option delta, r_{t_n} the annualized risk-free rate of return, $S(t_n)$ the underlying stock price at the end of trading day t_n, where $t_n \in \{t_0, t_1, \ldots, t_{N-1}\}$ are the N trading days within the period from

time $t-1$ to t, and a_n is the number of calendar days between t_n and t_{n+1}. We use an analogous equation for the delta-hedged put option gain, with, however, put option price and delta replacing call option price and delta, respectively. We finally calculate the value of the delta-hedged call option at time $t-1$ as the absolute value of $C_{t-1} - \Delta_{c,t-1} S_{t-1}$ and the value of a delta-hedged put option as the absolute value of $P_{t-1} - \Delta_{p,t-1} S_{t-1}$, where P_{t-1} is the value of the put option and $\Delta_{p,t-1}$ the put option delta.

2.4.3 The Pricing of Consumption Risks

In this section, we study how consumption growth, expected consumption growth, and consumption volatility exposures price the cross-section of option returns using various test assets. We start our investigation by looking at the cross-section of quarterly returns on idiosyncratic volatility sorted portfolios of delta-hedged call or put options or straddles. We also consider the realized and implied variance of these portfolios over the same return horizon. As robustness tests, we next analyze the monthly returns of the idiosyncratic volatility sorted portfolios as well as the quarterly returns of moneyness sorted portfolios of delta-hedged call or put options or straddles. [9]

A. The Quarterly Returns of Idiosyncratic-Volatility-Sorted Option Portfolios

We first test whether consumption exposures explain the negative relation between delta-hedged stock-option returns and idiosyncratic stock volatility discovered in Cao and Han (2013). Idiosyncratic volatility is the standard deviation of the residual with respect to the Fama-French three-factor model estimated using daily stock returns over the previous quarter. To mitigate the influence of time-to-maturity, we keep options with a time-to-maturity between 106 to 176 days so that the option expires after the end of the following quarter and the time-to-maturity range is not too wide. We next only keep options with a moneyness (defined as stock price

[9] The empirical results hardly change before or after the financial crisis. To maintain the brevity of this paper, we do not report the results before or after the financial crisis periods.

divided by strike price) from 0.8 to 1.2 and choose the call (or put) option that is closest to ATM. If two or more options have the same moneyness, we choose the option with the shortest time-to-maturity. The selection results in 24,126 quarterly observations for call options and 12,355 quarterly observations for put options. Table 2.2 shows summary statistics for the quarterly option sample. Panel A suggests that the average delta-hedged call option return is –1.82% per quarter, with a variation of 9.80%. The average call option has a time-to-maturity of 137 days and a daily idiosyncratic volatility of 2.79%. Panel B suggests that the average delta-hedged put option return is –1.38% per quarter, with a variation of 5.22%. The average put option has a time-to-maturity of 133 days and a daily idiosyncratic volatility of 2.59%.

In addition to delta-hedged call and put option returns, we also calculate straddle returns. To do so, we match call and put options written on the same underlying stock and with the same strike price and time-to-maturity, applying the same selection criteria as for call or put options. We then compute the straddle return as the arithmetic average of the delta-hedged call and put option return. The straddle sample includes 6,613 quarterly observations and is thus smaller than both the call and put option sample simply because we require a matching call-and-put pair to form a straddle. Panel C of Table 2.2 presents summary statistics for the straddle sample, which align with the summary statistics for the call and put option samples reported in Panels A and B, respectively.

Table 2.2: Summary Statistics for the Options Sample

This table reports descriptive statistics on the delta-hedged option returns, days-to-maturity, and idiosyncratic underlying-stock volatility of option contracts sampled at a quarterly frequency. We exclude the following option observations: the stock underlying the option pays out cash over the options' remaining time-to-maturity; option price violates arbitrage bounds; reported trading volume is 0; option bid quote is 0 or midpoint of bid and ask quotes is less than 1/8. For each optionable stock, we keep that call or put closest to being at-the-money and having the shortest time-to-maturity among others. Then we only keep calls and puts with moneyness within the range from 0.8 to 1.2 and days-to-maturity within the range from 106 to 176 days. Delta-hedged returns are calculated through option delta-hedged gains (given by equation (2.22)) scaled by $\Delta S - C$ for calls and $P - \Delta S$ for puts, where Δ is the Black-Scholes option delta, S is the underlying stock price, and C (P) is the price of call (put) option at the beginning of a quarter. Straddle returns are computed as the average returns of calls and puts which are written on the same stock and have the same strike price and time-to-maturity. We select straddles with moneyness closest to 1 and within the range from 0.8 to 1.2. Then we only keep straddles with days-to-maturity between 106 and 176 days. Days-to-maturity is the number of calendar days until option expiration. Idiosyncratic volatility is the standard deviation of the residuals of Fama-French 3-factors model estimated using the daily stock returns over the previous quarter. Calls, puts, and straddles are reported in Panels A, B, and C, respectively. The option sample period is from January 1996 to December 2017.

	N	Mean	Median	SD	5th	10th	25th	75th	90th	95th
Panel A: Call Options										
Delta-hedged returns (%)	24,126	-1.82	-2.00	9.80	-15.66	-11.36	-6.09	2.06	7.64	12.33
Days-to-maturity	24,126	137	140	26	108	108	112	169	172	173
Idiosyncratic volatility (%)	24,126	2.79	2.41	1.83	1.01	1.24	1.70	3.43	4.73	5.76
Panel B: Put Options										
Delta-hedged returns (%)	12,355	-1.38	-1.79	5.22	-9.41	-7.47	-4.56	1.19	5.36	8.39
Days-to-maturity	12,355	133	114	26	107	108	110	168	172	173

Table 2.2 continued

Idiosyncratic volatility (%)	12,355	2.59	2.27	1.47	0.98	1.18	1.61	3.18	4.33	5.26
Panel C: Straddles										
Delta-hedged returns (%)	6,613	-1.73	-2.10	6.80	-10.85	-8.15	-4.90	1.05	5.22	18.08
Days-to-maturity	6,613	131	113	25	107	108	110	144	171	173
Idiosyncratic volatility (%)	6,613	2.63	2.28	1.55	0.99	1.19	1.62	3.22	4.45	5.36

Table 2.3: Idiosyncratic Risk Sorted Option Portfolios

This table reports characteristics of equally-weighted option portfolios sorted on idiosyncratic underlying-stock volatility (IVOL). Portfolios are rebalanced every quarter. The average return, standard deviation (SD), skewness, and kurtosis of each portfolio are reported. Column "High–Low" shows the average return of the long-short strategy which buys the highest IVOL portfolio and sells the lowest IVOL portfolio. Full sample loadings on consumption growth (Δc_t), the change in the perceived conditional mean of consumption growth ($\Delta\hat{\mu}_t$), and the change in the perceived consumption growth volatility ($\Delta\hat{\sigma}_t$) are reported for each call and put option portfolio. Panels A, B, and C present portfolios formed from call options, put options, and straddles, respectively. Newey-West (Newey and West (1987)) adjusted t-statistics are reported in parentheses. The sample period is from January 1996 to December 2017.

Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High–Low
Panel A: Call Options											
Mean Return (%)	-0.80	-0.75	-0.96	-0.78	-1.46	-1.82	-1.20	-1.97	-2.05	-3.77	-2.97
	(-2.10)	(-1.95)	(-2.38)	(-1.59)	(-3.31)	(-4.21)	(-2.70)	(-3.32)	(-3.58)	(-6.21)	(-6.91)
SD (%)	4.61	5.50	6.18	6.80	7.41	7.62	9.41	9.22	9.98	13.75	
Skewness	0.46	0.45	0.41	0.51	0.33	0.27	0.43	0.03	0.12	0.10	
Kurtosis	6.08	5.21	5.11	5.45	5.54	4.86	5.63	4.89	4.44	5.63	
$\beta_{\Delta c}$	0.000	-0.006	-0.003	0.004	-0.011	-0.010	-0.017	-0.016	-0.021	-0.029	
$\beta_{\Delta\hat{\mu}}$	-0.032	-0.008	-0.041	-0.052	0.023	-0.059	0.004	-0.049	-0.014	-0.009	
$\beta_{\Delta\hat{\sigma}}$	0.164	0.138	0.126	0.100	0.139	0.205	0.100	0.187	0.200	0.244	
Panel B: Put Options											

Table 2.3 continued

Mean Return (%)	-0.44	-0.61	-1.05	-0.90	-0.86	-1.06	-1.29	-1.21	-1.70	-2.38	-1.95
	(-1.19)	(-1.92)	(-3.47)	(-2.86)	(-2.61)	(-2.77)	(-3.87)	(-3.06)	(-4.37)	(-6.39)	(-7.32)
SD (%)	3.26	3.44	3.74	3.90	4.12	4.43	4.58	4.83	5.06	5.82	
Skewness	0.15	0.14	0.21	0.07	0.13	0.15	0.20	0.14	0.25	0.29	
Kurtosis	3.44	3.19	3.25	2.88	3.01	2.80	2.91	2.68	2.81	2.68	
$\beta_{\Delta c}$	0.000	-0.003	-0.011	-0.008	-0.007	0.000	-0.014	-0.005	-0.020	-0.010	
$\beta_{\Delta \hat{\mu}}$	-0.004	0.018	0.031	0.011	0.033	-0.049	0.008	-0.027	-0.039	-0.036	
$\beta_{\Delta \hat{\sigma}}$	0.110	0.023	0.047	0.017	0.043	0.057	-0.034	0.089	0.057	0.119	
Panel C: Straddles											
Mean Return (%)	-0.51	-0.70	-1.49	-0.67	-0.78	-1.89	-1.77	-1.30	-2.01	-3.17	-2.66
	(-1.41)	(-1.81)	(-5.23)	(-1.63)	(-1.61)	(-5.50)	(-4.25)	(-1.93)	(-3.58)	(-4.81)	(-5.74)
SD (%)	2.80	3.45	3.00	4.32	4.47	4.63	4.62	5.64	5.95	9.52	
Skewness	0.17	0.22	0.16	0.28	0.21	0.03	0.05	0.01	0.04	0.02	
Kurtosis	2.28	2.63	2.34	2.53	2.40	2.57	2.59	2.52	2.49	2.70	

At the end of each quarter $t-1$ in our sample period, we sort the delta-hedged call options (alternatively: delta-hedged put options or straddles) into ten decile portfolios according to the idiosyncratic volatility of the underlying stock measured until the end of that quarter. Portfolio 1 contains options or straddles written on the lowest idiosyncratic volatility stocks, while portfolio 10 contains those written on the highest. We construct equally-weighted portfolios and hold them over quarter t. We also create a spread portfolio long on portfolio 10 and short on portfolio 1 ("H–L"). Table 2.3 reports the average option and straddle returns for each portfolio. Supporting Cao and Han (2013), the delta-hedged option returns are all negative and become monotonically more negative with an increase in idiosyncratic volatility. In accordance, the mean return of the H–L IVOL portfolio is –2.97% per quarter for call options (t=–6.91), –1.95% per quarter for put options (t=–7.32), and –2.66% for straddles (t=–5.74). Interestingly, the standard deviation of portfolio returns also increases in idiosyncratic volatility. The skewness of all portfolio returns is slightly above zero, while their kurtosis is close to three. Thus, the portfolio returns are close to normally distributed, alleviating the concern that non-normality in them could distort our statistical inferences.[10]

Motivated by the long-run risks model in Section 2.2 (see Equation (2.16)), we next run time-series regressions of option portfolio returns on consumption growth, Δc_t, the change in the conditional mean of consumption growth, $\Delta\hat{\mu}_t$, and the change in consumption volatility, $\Delta\hat{\sigma}_t$:

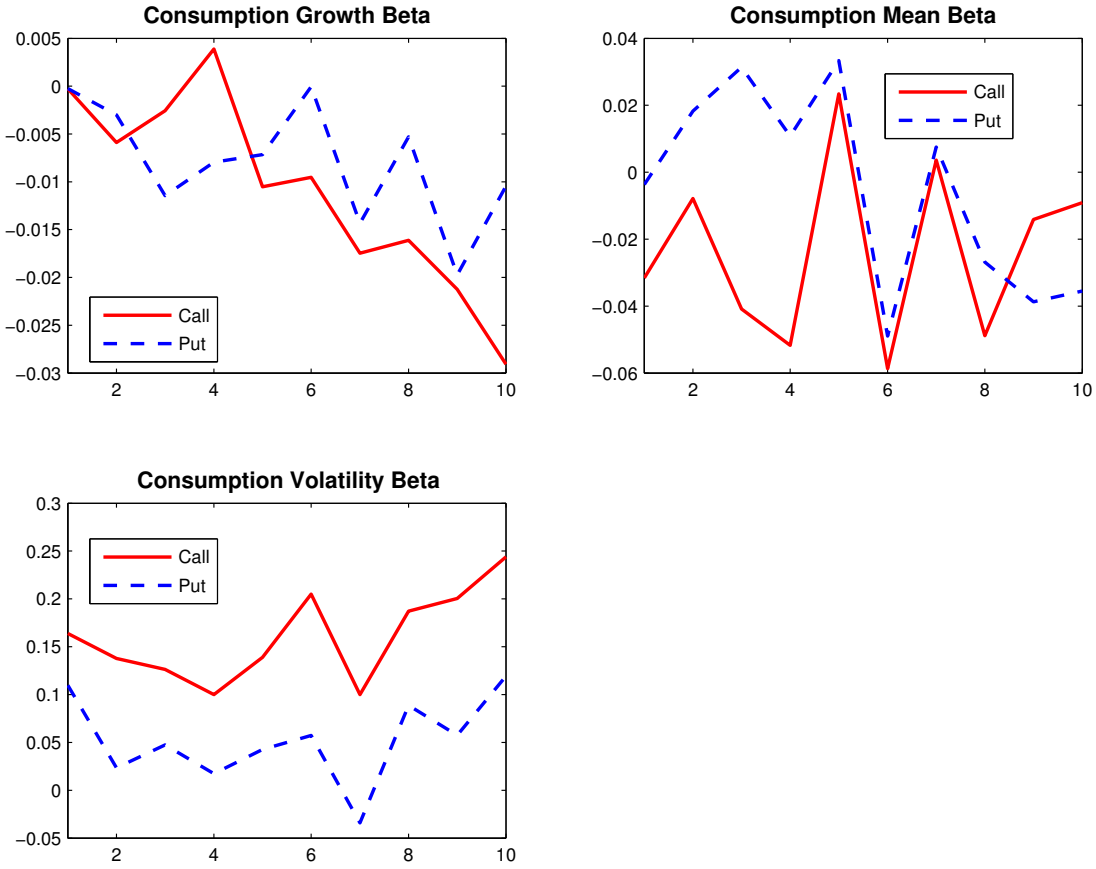
$$R_t^i = \alpha_t^i + \beta_{\Delta c,t}^i \Delta c_t + \beta_{\Delta \mu,t}^i \Delta\hat{\mu}_t + \beta_{\Delta\sigma,t}^i \Delta\hat{\sigma}_t + \epsilon_t^i, \tag{2.23}$$

where R_t^i is the delta-hedged quarterly return of option portfolio i over period t, α_t^i is a constant, $\beta_{\Delta c,t}^i$, $\beta_{\Delta\mu,t}^i$, and $\beta_{\Delta\sigma,t}^i$ are consumption exposures, and ϵ_t^i is the residual. We obtain the conditional mean and volatility of consumption growth ($\hat{\mu}_t$ and $\hat{\sigma}_t$, respectively) from the estimates of the Markov switching model in

[10]Consistent with the mild skewness and excess kurtosis of the portfolio returns, we have found that bootstrap inference levels used in either our portfolio sorts or FM regressions are similar to asymptotic inference levels. For the sake of brevity, we do not report the bootstrap inference levels in the paper.

Section 2.4.1. We estimate regression model (2.23) over rolling windows spanning ten years of quarterly data, expanding the rolling windows on a quarterly basis. The first rolling window stretches from the second quarter of 1996 to the first quarter of 2006.

Figure 2.6: Betas of Idiosyncratic Risk Sorted Option Portfolios



This graph displays consumption beta loadings estimated using the full sample for each option portfolio sorted on the IVOL of the underlying stock. Call option betas are plotted with the red line and put option betas are plotted with the blue dashed line. The upper left graph shows consumption growth beta, $\hat{\beta}_c^i$, for each portfolio. Consumption mean beta ($\hat{\beta}_\mu^i$) and volatility beta ($\hat{\beta}_\sigma^i$) for each portfolio are shown in the upper right and lower left graphs, respectively. The sample period is from January 1996 to December 2017.

Figure 2.6 and Table 2.3 present the full-sample exposures of the call and put option portfolios. Both types of portfolios produce negative consumption growth exposures, with the exposures becoming more negative over the IVOL portfolios. Conversely, the same portfolios produce positive consumption volatility exposures, with those exposures becoming more positive over the IVOL portfolios. There is no clear trend in the mean consumption exposures over the portfolios. That the

high IVOL portfolios are more negatively exposed to consumption growth and more positively to consumption volatility suggests that they are better suited to hedge consumption risks than the low IVOL portfolios.

To test whether the three consumption exposures are priced, we next run FM regressions of the quarterly returns of the option portfolios on subsets of the exposures. In our most comprehensive specification, we regress the quarterly return of option portfolio i over quarter $t+1$, R^i_{t+1}, on the consumption growth exposure, $\hat{\beta}^i_{\Delta c,t}$, the mean consumption growth exposure, $\hat{\beta}^i_{\Delta \mu,t}$, and the consumption volatility exposure, $\hat{\beta}^i_{\Delta \sigma,t}$, of the portfolio:

$$R^i_{t+1} = \varphi_{0,t+1} + \lambda_{\Delta c,t+1}\hat{\beta}^i_{\Delta c,t} + \lambda_{\Delta \mu,t+1}\hat{\beta}^i_{\Delta \mu,t} + \lambda_{\Delta \sigma,t+1}\hat{\beta}^i_{\Delta \sigma,t} + \eta^i_{t+1}, \tag{2.24}$$

where $\varphi_{0,t+1}$ is a constant and $\lambda_{\Delta c,t+1}$, $\lambda_{\Delta \mu,t+1}$, and $\lambda_{\Delta \sigma,t+1}$ are the risk premiums. The exposure estimates, $\hat{\beta}^i_{\Delta c,t}$, $\hat{\beta}^i_{\Delta \mu,t}$, and $\hat{\beta}^i_{\Delta \sigma,t}$, are obtained from the rolling-window time-series regressions.

Table 2.4 presents the results from the FM regressions, with Panels A, B, and C focusing on the results from call options, put options, and straddles, respectively. Only explicitly discussing the call option results in Panel A, model specification I tests the standard CCAPM. The risk premium estimate, $\lambda_{\Delta c}$, is significant at 0.57 (t=4.46) and the average cross-sectional R^2 is around 30%. Prior studies (see, e.g., Mankiw and Shapiro (1986), Lettau and Ludvigson (2001) and Boguth and Kuehn (2013)) show that quarterly contemporaneous consumption growth exposures do not explain stock returns. Different from prior studies using stock returns, we find that consumption growth exposures significantly explain delta-hedged option returns. Specification II adds the mean consumption growth exposure to the consumption growth exposure. Doing so, the $\lambda_{\Delta c}$ estimate is hardly affected, the mean consumption growth estimate, $\lambda_{\Delta \mu}$, is 0.04 and significant (t=2.51), and the average R^2 increases to around 36%. Conversely, Specification III adds the consumption volatility exposure to the consumption growth exposure. Doing so, the $\lambda_{\Delta c}$ estimate is again hardly affected, the $\lambda_{\Delta \sigma}$ estimate is –0.08 and significant (t=–3.34), and the average R^2 increases to around 44%. The negative $\lambda_{\Delta \sigma}$ estimate,

which is similar to the estimate in Boguth and Kuehn (2013), identifies a channel for macroeconomic volatility to be priced in options.[11]

Table 2.4: Fama-MacBeth Regressions with Idiosyncratic Risk Sorted Option Portfolios

This table reports quarterly risk premium estimates for the consumption betas. The consumption betas are estimated from ten-year rolling window time-series regressions (see equation (2.23)). In the cross-section, we regress quarterly future option returns over quarter $t+1$ on estimated beta loadings (see equation (2.24)). We report quarterly Fama-MacBeth regression results in Panel A, B, and C for call options, put options, and straddles, respectively. Newey-West (Newey and West (1987)) adjusted t-statistics using four lags are shown in parentheses. The sample period is from January 1996 to December 2017.

	Intercept	$\beta^i_{\Delta c,t}$	$\beta^i_{\Delta \mu,t}$	$\beta^i_{\Delta \sigma,t}$	Avg.R^2
			Panel A: Call Options		
I	-0.0018	0.5699			0.3011
	(-0.34)	(4.46)			
II	-0.0012	0.5936	0.0400		0.3633
	(-0.21)	(4.47)	(2.51)		
III	0.0037	0.5480		-0.0829	0.4427
	(0.80)	(4.37)		(-3.34)	
IV	0.0035	0.5658	0.0406	-0.0427	0.5085
	(0.70)	(4.38)	(2.19)	(-3.09)	
			Panel B: Put Options		
I	-0.0079	0.3164			0.1933
	(-2.39)	(3.89)			
II	-0.0095	0.2625	0.0166		0.2742
	(-2.45)	(3.07)	(1.05)		
III	-0.0101	0.2881		-0.0853	0.3608
	(-2.41)	(2.80)		(-4.96)	
IV	-0.0112	0.2423	-0.0025	-0.0806	0.4354
	(-2.77)	(2.63)	(-0.13)	(-4.63)	

[11] In a production-based asset pricing model, Liu and Zhang (2018) also highlight the connection between macroeconomic risk and the variance risk premium.

Table 2.4 continued.

Panel C: Straddles					
I	0.0011	0.6573			0.2270
	(0.24)	(4.67)			
II	-0.0004	0.6632	0.0792		0.3533
	(-0.11)	(4.36)	(2.54)		
III	0.0009	0.6429		-0.0418	0.3385
	(0.19)	(4.22)		(-3.10)	
IV	-0.0030	0.6068	0.0847	-0.0395	0.4629
	(-0.68)	(3.75)	(2.63)	(-2.74)	

Specification IV presents the full three-factor model. The jointly estimated risk premiums $\lambda_{\Delta c}$, $\lambda_{\Delta \mu}$, and $\lambda_{\Delta \sigma}$ are, respectively, 0.57 (t=4.38), 0.04 (t=2.19), and –0.04 (t=–3.09), indicating that all consumption exposures have independent significant explanatory power. The signs of the risk premiums are consistent with the agent preferring early resolution of uncertainty. Table 2.3 reveals substantial cross-sectional variation in the exposures, generating quarterly risk premium spreads of 1.86%, 0.33%, and 0.62% for consumption growth, mean consumption growth, and consumption volatility, respectively. The three spreads add up to 2.81%, close to the H–L spread return of 2.97% shown in Table 2.3. The average exposures over the portfolios imply average risk premiums of –0.62%, –0.10%, and –0.68% per quarter, respectively. Adding these up, the combined risk premium is –1.40% per quarter, almost identical to the pooled mean portfolio return of –1.56% per quarter.

Since delta-hedged option portfolios are zero-cost portfolios, we further test whether the intercepts are zero. Interestingly, we cannot reject that hypothesis, neither for the comprehensive model in specification IV nor for the less comprehensive models in specifications I, II, and III.

Panels B and C suggest that the results obtained from put options and straddles are similar to those obtained from call options, with the following two exceptions. First, the put option portfolios do not produce a significant risk premium for

expected consumption growth, regardless of the presence of consumption volatility (see Panel B). Second, the same portfolios also fail to produce insignificant intercepts, with their t-statistics ranging from –2.77 to –2.39 (again see Panel B).

Equations (2.10) and (2.15) in Section 2.2 suggest that the log-linearized SDF is approximately affine in consumption growth and the changes in its first two moments in our long-run risks model. As an alternative to running FM regressions, we thus now use two-stage GMM[12] to explicitly test the Euler equation of our model. The second stage uses the optimal weighting matrix. The full-model moment condition can be written as:

$$E[(1 - b_{\Delta c}\Delta c_{t+1} - b_{\Delta\mu}\Delta\hat{\mu}_{t+1} - b_{\Delta\sigma}\Delta\hat{\sigma}_{t+1})R^i_{t+1}] = 0, \tag{2.25}$$

where $b_{\Delta c}$, $b_{\Delta\mu}$, and $b_{\Delta\sigma}$ are the SDF loadings. The GMM estimation of the Euler equation can generate useful results for elucidating the relation between the SDF and the consumption risk factors.

Table 2.5 presents the model estimates (both SDF loadings and implied risk premiums λ) and test statistics, with Panels A, B, and C focusing on those from the call options, put options, and straddles, respectively. Standard errors are Newey-West (1987) adjusted with four lags. Again only explicitly discussing the call option results in Panel A, we find that, consistent with the FM regression results, the risk premiums for consumption growth and consumption volatility are significantly positive and negative, respectively. More specifically, the full model produces risk premium estimates for consumption growth, mean consumption growth, and consumption volatility of 0.64, 0.03, and –0.05, respectively, which are all close to the FM regression estimates in Panel A of Table 2.4. It further produces a mean absolute error (MAE) of 0.27% per quarter and an R^2 of over 86%, and the J-test of the over-identifying restrictions never rejects it (p=0.50). Turning to the put option and straddle results in Panels B and C, respectively, we find them to align with those from the call options, with the important exception that in their cases the risk premium of mean consumption growth tends to be significantly positive.

[12]Iterative GMM yields results virtually identical to those reported.

Table 2.5: Asset Pricing Tests with Idiosyncratic Risk Sorted Option Portfolios

This table reports GMM estimates of the moment conditions in equation (2.25), showing both the b estimates as well as the implied risk premia (λ). MAE and RMSE refer to the mean absolute pricing error and the root mean squared error, respectively. Panel A presents the results using ten quarterly call option portfolios, Panel B using ten quarterly put option portfolios, and Panel C using ten quarterly straddle portfolios. Newey-West (Newey and West (1987)) adjusted t-statistics using four lags are reported in parentheses and p-values for J-statistics are shown in parentheses below the associated J-statistics. The sample period is from January 1996 to December 2017.

	Δc	$\Delta \mu$	$\Delta \sigma$	MAE	RMSE	J	R^2
			Panel A: Call Options				
b	8.3689			0.0040	0.0045	4.5204	0.7338
	(2.97)					(0.87)	
λ	1.0801						
	(2.97)						
b	6.7396	16.9193		0.0034	0.0041	4.9864	0.7811
	(1.93)	(0.98)				(0.76)	
λ	1.0252	0.1372					
	(2.49)	(1.96)					
b	4.9862		-27.3480	0.0027	0.0032	6.3970	0.8632
	(2.22)		(-2.51)			(0.60)	
λ	0.6599		-0.0484				
	(2.26)		(-2.62)				
b	5.0306	-2.5268	-28.9577	0.0027	0.0032	6.3069	0.8638
	(2.14)	(-0.38)	(-2.58)			(0.50)	
λ	0.6434	0.0346	-0.0511				
	(2.05)	(0.91)	(-2.69)				

Table 2.5 continued.

Panel B: Put Options							
b	8.6057			0.0041	0.0049	7.1406	0.1396
	(2.76)					(0.62)	
λ	1.1107						
	(2.76)						
b	7.7899	8.3335		0.0042	0.0048	7.6869	0.1650
	(2.16)	(0.58)				(0.46)	
λ	1.0819	0.1086					
	(2.62)	(2.00)					
b	6.4881		-39.9611	0.0029	0.0034	8.0227	0.6012
	(3.63)		(-2.49)			(0.43)	
λ	0.8613		-0.0702				
	(3.62)		(-2.56)				
b	6.5837	-1.2344	-40.4442	0.0029	0.0033	7.9825	0.6017
	(3.32)	(-0.18)	(-2.50)			(0.33)	
λ	0.8626	0.0544	-0.0711				
	(3.53)	(1.95)	(-2.57)				

Table 2.5 continued.

Panel C: Straddles							
b	7.5161			0.0066	0.0084	5.7981	-0.1539
	(2.80)					(0.76)	
λ	0.9701						
	(2.80)						
b	4.1908	32.3586		0.0033	0.0045	2.9889	0.6676
	(2.11)	(1.86)				(0.94)	
λ	0.8381	0.1825					
	(2.90)	(2.34)					
b	4.9834		-45.9753	0.0045	0.0055	11.8537	0.4941
	(3.98)		(-2.30)			(0.16)	
λ	0.6707		-0.0793				
	(3.93)		(-2.35)				
b	3.7641	23.6245	-24.0243	0.0032	0.0036	3.6863	0.7847
	(3.23)	(1.70)	(-1.42)			(0.82)	
λ	0.7172	0.1394	-0.0418				
	(3.81)	(2.27)	(-1.47)				

Overall, our evidence in this section suggests that consumption risks can explain the negative relation between idiosyncratic stock volatility and the cross-section of delta-hedged option returns discovered by Cao and Han (2013). While these authors attribute the relation to market makers charging a premium for options that are difficult to delta hedge, showing that the relation is weakened after controlling for such difficulties, the question remains why investors are content to pay the high premiums. We show that investors are content to do so because options written on high IVOL stocks pay out more in adverse economic conditions, as evidenced by large negative consumption growth exposures and large positive consumption volatility exposures. Thus, these options are better hedging tools.

B. The Variance Risk Premiums of Idiosyncratic Volatility Sorted Portfolios

Section A. suggests that the mean returns of delta-hedged call option, put option, and straddle portfolios decrease with idiosyncratic stock volatility partially because options written on high IVOL stocks have higher (i.e., more positive) consumption volatility exposures than options written on low IVOL stocks. In accordance, Bakshi and Kapadia (2003) show that mean delta-hedged returns directly measure the underlying stock's variance risk premium. An alternative to estimate the variance risk premium is to calculate the difference between a stock's realized variance over quarter t (RV) and its implied variance paid for the options at the end of the previous quarter $t-1$ (IV). To see whether that alternative variance risk premium estimate also supports our conclusions, Table 2.6 presents the time-series average of the simple cross-sectional average of the difference between RV and IV for the idiosyncratic volatility sorted portfolios formed from call options (Panel A) and put options (Panel B).

Focusing on the call option results in Panel A, the table suggests that the implied variances of options are indeed increasing in the IVOL of the underlying stocks, from 0.10 for portfolio 1 to 0.74 for portfolio 10. The spread across the portfolios, 0.64, is statistically significant. The literature usually consider an option's implied variance as a standardized measure for the option's price, and it is natural that options on higher-volatility underlying stocks are more expensive. Higher IVOL stocks, however, do not only have a higher IV, but they also continue to be more volatile over the next quarter, as shown by their mean RV. To be precise, the difference in mean RV between portfolio 10 and portfolio 1 is a significant 0.60. Panel B shows that the results for put options are similar.

Table 2.6: Relation between Variance and Consumption

This table presents the average realized volatility (RV), average implied volatility (IV), and correlations between the difference in RV and IV (RVIV) with consumption growth (CG), the change in mean consumption growth (CM), and the change in consumption volatility (CV). Panel A presents call option results and Panel B put option results. *P*-values associated with the correlations are shown in parentheses. The sample period is from January 1996 to December 2017.

Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High–Low
						Panel A: Call Options					
$\overline{IV}$	0.1005	0.1464	0.1870	0.2267	0.2778	0.3239	0.3814	0.4417	0.5296	0.7392	0.6387
											(0.00)
$\overline{RV}$	0.1068	0.1576	0.1950	0.2466	0.2876	0.3183	0.3947	0.4522	0.5378	0.7051	0.5983
											(0.00)
$\rho_{RVIV,CG}$	-0.14	-0.14	-0.09	-0.03	-0.14	-0.20	-0.24	-0.17	-0.23	-0.28	-0.29
	(0.18)	(0.19)	(0.43)	(0.77)	(0.19)	(0.06)	(0.02)	(0.12)	(0.03)	(0.01)	(0.01)
$\rho_{RVIV,CM}$	0.01	-0.01	-0.06	-0.05	-0.03	-0.12	-0.12	-0.10	-0.09	-0.06	-0.07
	(0.89)	(0.96)	(0.60)	(0.63)	(0.75)	(0.27)	(0.25)	(0.36)	(0.42)	(0.59)	(0.51)
$\rho_{RVIV,CV}$	0.09	0.09	0.10	0.01	0.09	0.12	0.04	0.07	0.07	0.08	0.07
	(0.38)	(0.38)	(0.36)	(0.92)	(0.43)	(0.28)	(0.69)	(0.50)	(0.51)	(0.46)	(0.53)

Table 2.6 continued

Panel B: Put Options											
$\overline{IV}$	0.1076	0.1489	0.1918	0.2340	0.2693	0.3148	0.3642	0.4356	0.5089	0.7345	0.6269
											(0.00)
$\overline{RV}$	0.1056	0.1513	0.1829	0.2320	0.2738	0.3184	0.3537	0.4356	0.4839	0.6734	0.5678
											(0.00)
$\rho_{RVIV,CG}$	-0.15	-0.10	-0.22	-0.13	-0.14	-0.12	-0.19	-0.19	-0.28	-0.20	-0.19
	(0.17)	(0.34)	(0.04)	(0.24)	(0.18)	(0.29)	(0.09)	(0.08)	(0.01)	(0.07)	(0.08)
$\rho_{RVIV,CM}$	-0.03	-0.03	-0.06	-0.08	-0.05	-0.10	-0.09	-0.11	-0.13	-0.16	-0.18
	(0.78)	(0.76)	(0.58)	(0.44)	(0.64)	(0.38)	(0.42)	(0.30)	(0.23)	(0.15)	(0.09)
$\rho_{RVIV,CV}$	0.13	0.05	0.07	0.01	0.06	0.07	-0.10	0.06	-0.04	0.08	0.06
	(0.23)	(0.64)	(0.51)	(0.93)	(0.57)	(0.50)	(0.34)	(0.60)	(0.73)	(0.45)	(0.59)

We next calculate the time-series correlation between the difference in RV and IV and contemporaneous consumption growth at the portfolio level. We find those correlations to be consistently negative. Interpreting the difference in RV and IV as the payoff from a volatility hedging strategy, we can thus conclude that, as consumption growth decreases, volatility hedging strategies based on all portfolios observe positive payoffs. Interestingly, however, we find a more negative correlation for higher IVOL stocks, with the H–L spread portfolio having a correlation of –0.29 (p=0.01). Even after accounting for the higher IV paid for higher IVOL options, hedging strategies based on high IVOL stocks thus pay off more than those based on low IVOL stocks in low consumption growth states, explaining why options on high IVOL stocks offer a better hedge against such states than those on low IVOL stocks. As before, Panel B reports similar results for put options.

C. Robustness Test: Monthly Returns of Idiosyncratic-Volatility Sorted Option Portfolios

We next establish that our main conclusions are not sensitive to our choice of option maturity and return frequency. We do so as follows. At the end of every month $t-1$ in our sample period, we restrict our attention to options with a time-to-maturity between 43 to 55 days and choose the option that is closest to ATM. We then delta-hedge the options. We sort the delta-hedged call options, put options, and straddles into ten decile portfolios according to the idiosyncratic volatility of the underlying stock. We equally-weight the ten portfolios and hold them over month t.[13]

Table 2.7 shows the results from using the monthly option portfolios in the FM regressions. The estimation identifies the risk premiums of consumption growth, mean consumption growth, and consumption volatility (see Equation (2.24)). Panels A, B, and C focus on the results obtained from call options, put options, and straddles, respectively. We follow Boguth and Kuehn (2013) and obtain the results from estimating time-series regression (2.23) on quarterly data over the prior ten

[13]We present summary statistics for the monthly-rebalanced portfolios in Tables A2 and A3 in the Appendix.

years, using the exposures from the quarterly regression run over the window coming closest to but not exceeding the FM regression month. Focusing on the call option results in Panel A, the estimated monthly risk premiums $\lambda_{\Delta c}$, $\lambda_{\Delta \mu}$, and $\lambda_{\Delta \sigma}$ are, respectively, 0.25 (t=4.63), 0.01 (t=2.37), and –0.02 (t=–5.81), indicating that all consumption exposures have independent significant explanatory power. The three monthly risk premiums translate into quarterly risk premiums of 0.75, 0.03, and –0.06, in line with our quarterly estimates shown earlier. Overall, Table 2.7 suggests that the estimates, statistical significance, and the model fit of the monthly portfolios are all similar to their quarterly counterparts in Table 2.4. In particular, with only one exception, the risk premiums for consumption growth and mean consumption growth are again positive and significant, while those for consumption volatility are again negative and significant.

Table 2.7: Fama-MacBeth Regressions with Monthly Idiosyncratic Risk Sorted Option Portfolios

This table reports monthly risk premium estimates for the consumption betas. The consumption betas are estimated from ten-year rolling window time-series regressions (see equation (2.23)). In the cross-section, we regress monthly future option returns over month $t+1$ on estimated beta loadings (see equation (2.24)). We report monthly Fama-MacBeth regression results in Panel A, B, and C for call options, put options, and straddles, respectively. Newey-West (Newey and West (1987)) adjusted t-statistics using twelve lags are shown in parentheses. The sample period is from January 1996 to December 2017.

	Intercept	$\beta^i_{\Delta c,t}$	$\beta^i_{\Delta \mu,t}$	$\beta^i_{\Delta \sigma,t}$	Avg.R^2
			Panel A: Call Options		
I	0.0008	0.3068			0.3458
	(-0.76)	(6.27)			
II	0.0007	0.2885	0.0130		0.4090
	(0.38)	(6.58)	(2.83)		
III	0.0036	0.2694		-0.0200	0.4832
	(2.39)	(4.59)		(-4.92)	
IV	0.0033	0.2543	0.0110	-0.0216	0.5479
	(2.24)	(4.63)	(2.37)	(-5.81)	

Table 2.7 continued.

Panel B: Put Options					
I	-0.0031	0.2634			0.2048
	(-1.37)	(3.39)			
II	-0.0045	0.1948	0.0144		0.3092
	(-1.52)	(2.14)	(1.39)		
III	-0.0043	0.2721		-0.0758	0.4481
	(-2.02)	(6.33)		(-6.30)	
IV	-0.0053	0.2116	-0.0016	-0.0728	0.5446
	(-2.51)	(4.16)	(-0.14)	(-5.98)	
Panel C: Straddles					
I	0.0029	0.3991			0.2572
	(1.36)	(5.34)			
II	0.0026	0.4075	0.0497		0.3827
	(1.29)	(5.27)	(5.98)		
III	0.0046	0.4169		-0.0201	0.3464
	(2.18)	(5.35)		(-6.86)	
IV	0.0039	0.4159	0.0525	-0.0195	0.4685
	(1.93)	(5.06)	(6.28)	(-7.52)	

Table 2.8: Asset Pricing Tests with Monthly Idiosyncratic Risk Sorted Option Portfolios

This table reports GMM estimates of the moment conditions in equation (2.25), showing both the b estimates as well as the implied risk premia (λ). MAE and RMSE refer to the mean absolute pricing error and the root mean squared error, respectively. Panel A presents the results using ten monthly call option portfolios, Panel B using ten monthly put option portfolios, and Panel C using ten monthly straddle portfolios. Newey-West (Newey and West (1987)) adjusted t-statistics using twelve lags are reported in parentheses and p-values for J-statistics are shown in parentheses below the associated J-statistics. The sample period is from January 1996 to December 2017.

	Δc	$\Delta \mu$	$\Delta \sigma$	MAE	RMSE	J	R^2
			Panel A: Call Options				
b	8.9548			0.0068	0.0079	6.9811	0.7211
	(3.45)					(0.64)	
λ	1.1557						
	(3.45)						
b	9.6201	-7.8005		0.0067	0.0079	5.9017	0.7238
	(2.20)	(-0.64)				(0.66)	
λ	1.1700	0.0537					
	(2.44)	(1.64)					
b	6.7593		-28.0413	0.0048	0.0062	7.3419	0.8265
	(2.52)		(-1.74)			(0.50)	
λ	0.8892		-0.0506				
	(2.53)		(-1.84)				
b	5.2038	15.1031	-31.4505	0.0047	0.0061	9.4077	0.8348
	(2.00)	(1.26)	(-1.79)			(0.22)	
λ	0.8292	0.1146	-0.0551				
	(2.70)	(2.41)	(-1.86)				

Table 2.8 continued.

Panel B: Put Options							
b	11.6732			0.0088	0.0108	11.0693	0.3186
	(3.34)					(0.27)	
λ	1.5066						
	(3.34)						
b	13.6556	-15.2714		0.0091	0.0106	11.3080	0.3354
	(3.29)	(-0.78)				(0.18)	
λ	1.6222	0.0575					
	(3.27)	(0.71)					
b	7.8617		-44.9914	0.0076	0.0092	8.3814	0.4995
	(3.06)		(-1.72)			(0.40)	
λ	1.0416		-0.0794				
	(3.05)		(-1.79)				
b	8.3582	-3.3472	-44.2580	0.0076	0.0092	8.2683	0.5002
	(2.96)	(-0.25)	(-1.63)			(0.31)	
λ	1.0745	0.0613	-0.0785				
	(3.02)	(1.08)	(-1.70)				

Table 2.8 continued.

Panel C: Straddles							
b	9.4370			0.0104	0.0143	10.6698	-0.0335
	(2.84)					(0.30)	
λ	1.2180						
	(2.84)						
b	10.6360	-31.4740		0.0099	0.0138	8.6748	0.0320
	(2.48)	(-1.41)				(0.37)	
λ	1.0836	-0.0424					
	(2.44)	(-0.54)					
b	4.2948		-100.2685	0.0058	0.0071	6.5055	0.7433
	(2.36)		(-2.79)			(0.59)	
λ	0.6143		-0.1690				
	(2.47)		(-2.81)				
b	3.2201	18.3734	-107.5526	0.0051	0.0069	8.1810	0.7615
	(1.63)	(1.64)	(-2.61)			(0.32)	
λ	0.6487	0.1098	-0.1802				
	(2.63)	(2.43)	(-2.62)				

In Table 2.8, we present the results from using the monthly option portfolios in GMM tests (see Equation (2.25)). To match the monthly option and the quarterly consumption data, we compound monthly returns to quarterly returns. As before, Panels A, B, and C focus on the results obtained from call options, put options, and straddles, respectively. Focusing on the call option results in Panel A, the estimated quarterly risk premiums $\lambda_{\Delta c}$, $\lambda_{\Delta \mu}$, and $\lambda_{\Delta \sigma}$ are, respectively, 0.83 (t=2.80), 0.11 (t=2.41), and –0.06 (t=–1.86). The J-test fails to reject the overidentifying restriction with the p-value equal to 0.22. The estimated risk premiums are qualitatively similar with those estimated using quarterly options in Table 2.5. Table 2.8 suggests that the most comprehensive model explains the

monthly option portfolio returns well, with it always passing the J-test of the overidentifying restrictions and it producing high R^2s and small pricing errors. As in the FM regressions, consumption growth and mean consumption growth are (with one exception) significantly positively priced, while consumption volatility is significantly negatively priced.

D. Alternative Test Assets: Moneyness Sorted Option Portfolio Returns

We finally switch to an alternative set of test assets to study the pricing of consumption risks in delta-hedged options: moneyness-sorted portfolios. Bakshi and Kapadia (2003) show that delta-hedged call option returns are negative and monotonically decrease with option moneyness, defined as the ratio of the underlying stock's price to the option's strike price.[14] Boyer and Vorkink (2014) offer evidence that a call option's ex-ante skewness is negatively related to its return and monotonically decreases with option moneyness. Taken together, these two findings imply that lower-moneyness (OTM) call options have higher ex-ante skewness and earn more negative returns than higher-moneyness (ITM) call options do. Our aim in this section is to find out whether consumption risks can help us understand why delta-hedged call option returns are more negative for OTM than ITM options.

We conduct the moneyness-portfolio based tests in this section on the quarterly returns of delta-hedged options with a time-to-maturity between 106 to 176 days.[15] The option selection criteria used in these tests are identical to those used before, except for the following three differences. First, we only consider American call options written on non-dividend stocks in these tests since American put options contain an early exercise risk premium correlating with moneyness. Second, we allow for more than one call option on each stock since we require a spectrum of

[14]Note that Bakshi and Kapadia (2003) also present evidence that the dollar gains (losses) are smaller for OTM than ITM options. This is to be expected because the dollar gains are not normalized by price and OTM options are much cheaper than ITM options. Once the gains are normalized, OTM options have more negative returns than ITM options.

[15]We also study the monthly returns of delta-hedged call options with a time-to-maturity of 44 to 86 days. As shown in the Appendix Tables A4, A5, and A6, the results obtained from the monthly returns are similar to those obtained using quarterly options.

options with different moneyness levels.[16] Third, we only choose actively traded options, defining an option as actively traded if its number of zero-trading-volume days within a quarter is less than ten. Panel A of Table 2.9 reports summary statistics for the resulting delta-hedged call option sample. The sample consists of 34,437 quarterly observations. The mean and median delta-hedged returns are both negative. The average time-to-maturity is 124 days, comparable to that of the IVOL sorted option portfolios. While mean moneyness is about 99% and thus close to ATM, the standard deviation is an impressive 18%.

We form the delta-hedged call option portfolios sorted on moneyness as follows. At the end of each quarter $t-1$ in our sample period, we sort the options into ten decile portfolios according to their moneyness. Portfolio 1 contains options with the lowest moneyness (most OTM), while portfolio 10 contains options with the highest moneyness (most ITM). We equally-weight the portfolios and hold them over quarter t. We also create a spread portfolio long on portfolio 10 and short on portfolio 1 ("H–L"). Panel B of Table 2.9 presents the mean returns of the portfolios. In line with Bakshi and Kapadia (2003), all portfolios produce negative mean returns, with the mean returns becoming more negative the lower the moneyness of the options in a portfolio (i.e., the more the options are OTM). In particular, the mean delta-hedged return ranges from –4.28% per quarter (t=–5.25) for the lowest moneyness portfolio to –0.85% for the highest portfolio (t=–2.42), with the spread being a significant 3.42% (t=5.70).

[16]We also construct the sample using only one option per stock-quarter. As shown in Appendix Tables A7 and A8, the results obtained from that strategy are qualitatively similar to those reported here.

Table 2.9: Moneyness Sorted Call Option Portfolios

This table reports summary statistics and characteristics of moneyness-sorted call-option portfolios. Panel A reports descriptive statistics on delta-hedged option returns, days-to-maturity, and moneyness of the options. We exclude the following option observations: the stock underlying the option pays out cash over the options' remaining time-to-maturity; option price violates arbitrage bounds; reported trading volume is 0; option bid quote is 0 or midpoint of bid and ask quotes is less than 1/8; options have more than ten recorded zero-trading-volume days within a quarter. We keep options with days-to-maturity between 106 to 176 days. Delta-hedged returns are calculated through option delta-hedged gains (given by equation (2.22)) scaled by $\Delta S - C$, where Δ is the Black-Scholes option delta, S is the underlying stock price, and C is the price of call option. Days-to-maturity is the number of calendar days until option expiration. Moneyness is the ratio of stock price over option strike price. Panel B reports characteristics of equally-weighted call option portfolios sorted by options moneyness. Portfolios are rebalanced every quarter. Average return, standard deviation (SD), skewness, and kurtosis of each portfolio are reported. The column "High–Low" shows the average return of the long-short strategy which buys the highest moneyness portfolio and sells the lowest moneyness portfolio. Newey-West (Newey and West (1987)) adjusted t-statistics are reported in parentheses. The option sample period is from January 1996 to December 2017.

Panel A: Summary Statistics

	N	Mean	Median	SD	5th	10th	25th	75th	90th	95th
Delta-hedged returns (%)	34,437	-2.50	-2.11	7.25	-14.53	-10.64	-5.84	0.99	5.18	8.54
Days-to-maturity	34,437	124	113	18	107	108	110	141	144	168
Moneyness=S/K (%)	34,437	98.70	96.79	17.76	74.58	80.00	88.25	106.33	119.49	129.76

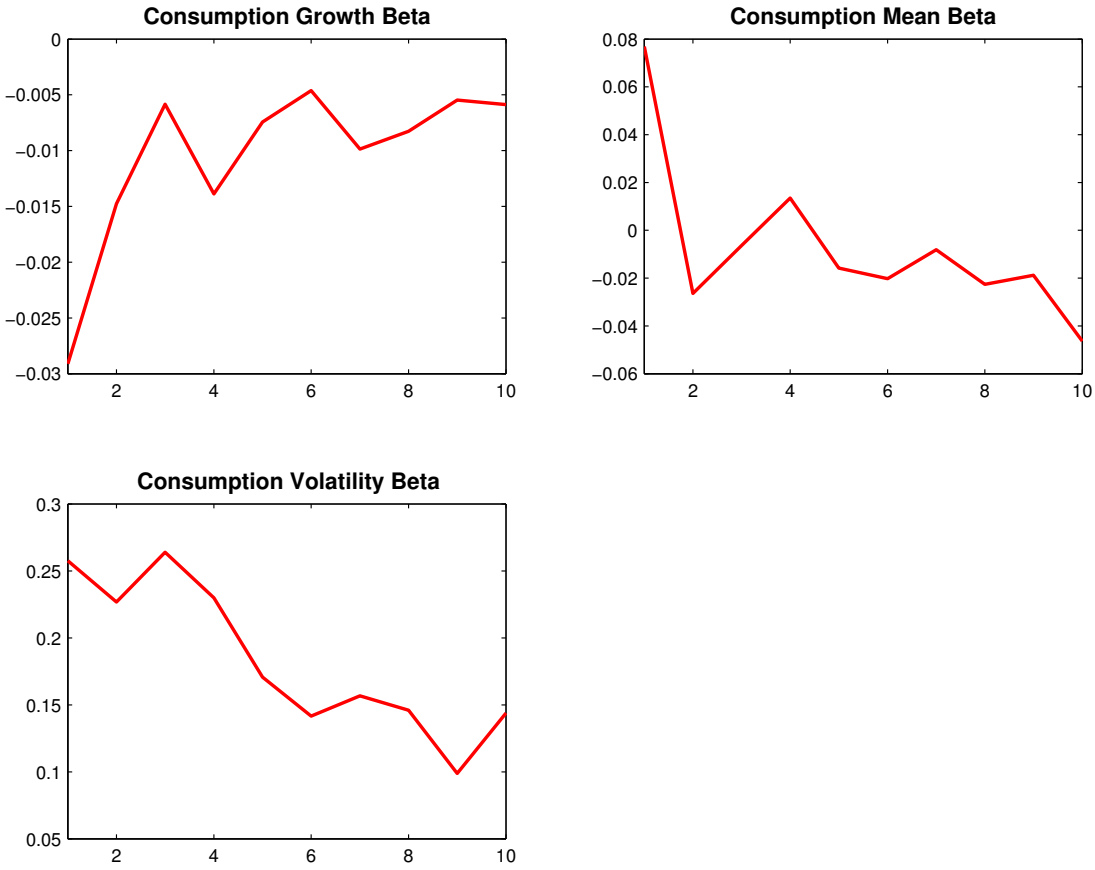
Table 2.9 continued

Panel B: Moneyness Sorted Portfolios											
Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High–Low
Mean Return (%)	-4.28	-3.46	-2.79	-2.51	-1.93	-1.57	-1.56	-1.3	-1.14	-0.85	3.42
	(-5.25)	(-5.08)	(-5.16)	(-5.94)	(-4.39)	(-4.02)	(-4.80)	(-3.54)	(-3.53)	(-2.42)	(5.70)
SD (%)	8.66	7.31	6.28	5.84	5.04	4.76	4.63	4.30	4.18	4.23	
Skewness	0.29	0.17	0.17	0.15	0.09	0.01	-0.05	-0.15	-0.18	-0.27	
Kurtosis	2.38	2.72	3.04	3.38	3.62	3.86	4.64	4.21	4.50	4.69	

Panel B of Table 2.9 also presents the standard deviation, skewness, and kurtosis of the portfolio returns. While the low moneyness options are more volatile than the high moneyness options, skewness monotonically declines with moneyness, with OTM options being positively skewed and ITM options negatively skewed. Thus, sorting options into portfolios according to moneyness is akin to sorting them into portfolios according to Boyer and Vorkink's (2014) ex-ante skewness proxy derived from the Black-Scholes model, even though we do not explicitly use skewness in our sorts. We also differ with Boyer and Vorkink (2014) in that they focus on raw option returns while we examine delta-hedged returns. The return spread between portfolios with high and low skewness further suggests that investors pay more for the tail probability (lottery feature) in OTM options even after controlling for the directional movement of the underlying asset. Finally, the kurtosis of the portfolios' returns is close to three, which is the kurtosis of a normal random variable.

We next estimate time-series regression (2.23) on the delta-hedged call option portfolios sorted on moneyness, plotting the full-sample estimates of $\hat{\beta}^i_{\Delta c,t}$, $\hat{\beta}^i_{\Delta \mu,t}$, and $\hat{\beta}^i_{\Delta \sigma,t}$ in Figure 2.7. The figure reveals that the consumption growth exposure, $\hat{\beta}^i_{\Delta c,t}$, is negative for all portfolios, suggesting that delta-hedged call options are counter-cyclical assets. More crucially, the low-moneyness (OTM) portfolios have larger negative consumption growth exposures, indicating that they are better suited to hedge against adverse conditions than the high-moneyness (ITM) portfolios. Further adding to their hedging ability, the same portfolios also have more positive consumption volatility exposures. The only concerning aspect of the low-moneyness portfolios is that they can have a higher mean consumption growth exposure than the other portfolios, which lowers their ability to hedge against adverse conditions.

Betas of Moneyness Sorted Call Option Portfolios



This graph displays consumption beta loadings estimated using the full sample for each option portfolio sorted by call option moneyness. The sample constitutes only actively traded call options. The upper left graph shows consumption growth beta, $\hat{\beta}_c^i$, for each portfolio. Consumption mean beta ($\hat{\beta}_\mu^i$) and volatility beta ($\hat{\beta}_\sigma^i$) for each portfolio are shown in the upper right and lower left graphs, respectively. The sample period is from January 1996 to December 2017.

Table 2.10 reports the results from running FM regressions of the quarterly returns of the moneyness-sorted portfolios on the rolling-window consumption exposures. Model specification I tests the standard CCAPM results. The risk premium estimate, $\lambda_{\Delta c}$, is a significant 0.65 (t=3.73) and the average cross-sectional R^2 is around 51%. The magnitude of the risk premium is close to that estimated using the IVOL sorted option portfolios. Specification II adds the expected mean consumption growth exposure to the standard CCAPM. Doing so, the estimate of $\lambda_{\Delta c}$ remains highly significant, while the estimate for expected consumption growth, $\lambda_{\Delta \mu}$, is 0.05 but statistically insignificant. The average R^2 increases to around 62%. Specification III adds the consumption volatility exposure to the standard CCAPM. Doing so, the estimate for $\lambda_{\Delta c}$ is again similar to before, while

the estimate for $\lambda_{\Delta\sigma}$ is a negative and significant –0.10 (t=–5.18). The average R^2 further increases to around 67%.

Table 2.10: Fama-MacBeth Regressions with Moneyness Sorted Call Option Portfolios

This table reports quarterly risk premium estimates for the consumption betas. The sample features only actively traded call options. The consumption betas are estimated from ten-year rolling window time-series regressions using equation (2.23). In the cross-section, we regress quarterly future call option returns over quarter $t+1$ on estimated beta loadings (see equation (2.24)). Newey-West (Newey and West (1987)) adjusted t-statistics with a lag of four are displayed in parentheses. The sample period is from January 1996 to December 2017.

	Intercept	$\beta^i_{\Delta c,t}$	$\beta^i_{\Delta\mu,t}$	$\beta^i_{\Delta\sigma,t}$	Avg.R^2
I	-0.0046	0.6524			0.5118
	(-1.15)	(3.73)			
II	-0.0036	0.7860	0.0488		0.6230
	(-1.07)	(4.29)	(1.30)		
III	-0.0011	0.6006		-0.1048	0.6726
	(-0.29)	(3.45)		(-5.18)	
IV	-0.0019	0.7601	0.0631	-0.1071	0.7310
	(-0.50)	(4.05)	(1.99)	(-5.41)	

Specification IV presents the results from the full three-factor model. The risk premium estimates $\lambda_{\Delta c}$, $\lambda_{\Delta\mu}$, and $\lambda_{\Delta\sigma}$ are 0.76 (t=4.05), 0.06 (t=1.99), and –0.11 (t=–5.41), respectively. Thus, all three consumption exposures have significant explanatory power for option returns. The average R^2 reaches 73%. The estimated risk premiums are consistent with those estimated using the IVOL portfolios, and their signs again suggest that agents prefer early resolution of uncertainty, lending support to the long-run risks model of Bansal and Yaron (2004). Because delta-hedged returns measure the returns of zero-cost portfolios, we also again test whether the intercepts are equal to zero. As before, we fail to reject that hypothesis in all four specifications.

Table 2.11: Asset Pricing Tests with Moneyness Sorted Call Option Portfolios

This table reports GMM estimates of the moment conditions in equation (2.25), showing both the b estimates as well as the implied risk premia (λ). The sample features only actively traded call options. MAE and RMSE refer to the mean absolute pricing error and the root mean squared error, respectively. The table presents the results using ten quarterly call option portfolios. Newey-West (Newey and West (1987)) adjusted t-statistics using four lags are reported in parentheses and p-values for J-statistics are shown in parentheses below the associated J-statistics. The sample period is from January 1996 to December 2017.

	Δc	$\Delta\mu$	$\Delta\sigma$	MAE	RMSE	J	R^2
b	13.8248			0.0046	0.0061	4.7015	0.6544
	(2.27)					(0.86)	
λ	1.7843						
	(2.27)						
b	14.2182	-6.7146		0.0047	0.0061	4.6723	0.6604
	(2.00)	(-0.36)				(0.79)	
λ	1.7734	0.1007					
	(2.04)	(1.17)					
b	6.1819		-41.8062	0.0036	0.0045	6.9142	0.8183
	(2.62)		(-3.71)			(0.55)	
λ	0.8229		-0.0731				
	(2.66)		(-3.71)				
b	5.5416	-24.1990	-53.0705	0.0028	0.0036	6.3358	0.8836
	(2.48)	(-2.55)	(-2.62)			(0.50)	
λ	0.5247	-0.0575	-0.0917				
	(1.76)	(-1.29)	(-2.67)				

Table 2.11 presents the estimates and model statistics from using GMM to test our model's Euler equation on the moneyness-sorted portfolios. In line with the FM regression results, the risk premium for consumption growth is consistently positive and significant, while the risk premium for consumption volatility is consistently negative and significant. In the full three-factor model, the risk premium estimates

for consumption growth and consumption volatility are 0.52 and –0.09, respectively, similar to those obtained in the FM regressions. Conversely, the risk premium for expected consumption growth is insignificant. The full model prices the ten portfolios with small pricing errors. The J-test fails to reject the over-identifying restrictions (p=0.50) and the full model's R^2 is over 88%.

Overall, our evidence in this section suggests that consumption risks can explain the negative relation between delta-hedged call options and moneyness. Boyer and Vorkink (2014) argue that the negative relation arises because OTM options have more right-skewed payoffs than other options, which can be attractive to investors with non-standard preferences. Our evidence supports their finding that option moneyness is negatively related to option skewness. More importantly, we show that the low returns of low-moneyness and/or high-skewness options can be explained by these options being better instruments to hedge against adverse economic conditions. Here, adverse economic conditions not only include shocks to consumption growth but also variations in expected consumption growth and consumption volatility that can lead to a rise in the SDF under recursive utility.

2.5 Conclusion

We study the impact of consumption risks on the cross-section of delta-hedged option and straddle returns using portfolios sorted on idiosyncratic underlying-stock volatility and moneyness. Our consumption-based pricing factors consist of consumption growth, an estimate of its conditional expectation, and an estimate of its conditional volatility. The three factors explain the cross-section of delta-hedged option returns well and support a risk-based explanation for option returns. Consumption growth and expected consumption growth command a positive risk premium, whereas consumption volatility commands a negative risk premium. Our evidence suggests that, in a representative-agent economy with Epstein-Zin's (1989) recursive preferences, the agent prefers early resolution of uncertainty. They further suggest that options written on high IVOL stocks and options with a low moneyness earn more negative returns than other options because they provide a better hedge

against bad macroeconomic conditions. Overall, our empirical findings are robust to the choice of test assets, option maturities, and the testing horizon. Taken together, our evidence provides a strong foundation for consumption risks explaining the cross-section of delta-hedged option and straddle returns.

References

Abel, A. B., 1999. Risk premia and term premia in general equilibrium. Journal of Monetary Economics 43 (1), 3–33.

Aretz, K., Lin, M.-T., Poon, S.-H., 2019. Moneyness, underlying asset volatility, and the cross-section of option returns, working paper.

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. Review of Financial Studies 16 (2), 527–566.

Bansal, R., Khatchatrian, V., Yaron, A., 2005. Interpretable asset markets? European Economic Review 49 (3), 531–560.

Bansal, R., Kiku, D., Shaliastovich, I., Yaron, A., 2014. Volatility, the macroeconomy, and asset prices. Journal of Finance 69 (6), 2471–2511.

Bansal, R., Kiku, D., Yaron, A., 2012. An empirical evaluation of the long-run risks model for asset prices. Critical Finance Review 1 (1), 183–221.

Bansal, R., Shaliastovich, I., 2011. Learning and asset-price jumps. Review of Financial Studies 24 (8), 2738–2780.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59 (4), 1481–1509.

Boguth, O., Kuehn, L.-A., 2013. Consumption volatility risk. Journal of Finance 68 (6), 2589–2615.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Review of Financial Studies 11 (1), 4463–4492.

Boyer, B. H., Vorkink, K., 2014. Stock options as lotteries. Journal of Finance 69 (4), 1485–1527.

Breeden, D. T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7 (3), 265–296.

Calvet, L., Fisher, A., 2007. Multifrequency news and stock returns. Journal of Financial Economics 86 (1), 178–212.

Cao, J., Han, B., 2013. Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics 108 (1), 231–249.

Coval, J. D., Shumway, T., 2001. Expected option returns. Journal of Finance 56 (3), 983–1009.

Drechsler, I., Yaron, A., 2011. What's vol got to do with it. Review of Financial Studies 24 (1), 1–45.

Epstein, L. G., Zin, S. E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica 57 (4), 937–969.

Eraker, B., Shaliastovich, I., 2008. An equilibrium guide to designing affine pricing models. Mathematical Finance 18 (4), 519–543.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81 (3), 607–636.

Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. Journal of Financial Economics 94 (2), 310–326.

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Hu, G., Jacobs, K., 2020. Volatility and expected option returns. Journal of Financial and Quantitative Analysis 55 (3), 1025–1060.

Jagannathan, R., Wang, Y., 2007. Lazy investors, discretionary consumption, and the cross-section of stock returns. Journal of Finance 62 (4), 1623–1661.

Kim, C.-J., Nelson, C. R., 1999. Has the U.S. economy become more stable? A Bayesian approach based on a markov-switching model of the business cycle. Review of Economics and Statistics 81 (4), 608–616.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy 109 (6), 1238–1287.

Lettau, M., Ludvigson, S. C., Wachter, J. A., 2008. The declining equity premium: What role does macroeconomic risk play? Review of Financial Studies 21 (4), 1653–1687.

Liu, H., Zhang, Y., 2018. Financial uncertainty with ambiguity and learning, working paper.

Mankiw, N. G., Shapiro, M. D., 1986. Risk and return: Consumption beta versus market beta. Review of Economics and Statistics 68 (3), 452–459.

McConnell, M. M., Perez-Quiros, G., 2000. Output fluctuations in the United States: What has changed since the early 1980's? American Economic Review 90 (5), 1464–1476.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Parker, J. A., Julliard, C., 2005. Consumption risk and the cross section of expected returns. Journal of Political Economy 113 (1), 185–222.

Romeo, T., 2015. Consumption volatility and the cross-section of stock returns. Review of Finance 19, 367–405.

Yogo, M., 2006. A consumption-based explanation of expected stock returns. Journal of Finance 61 (2), 539–580.

2.6 Appendix

2.6.1 Derivation of Expected Asset Return

The derivation below closely follows Boguth and Kuehn (2013). To test the model in the cross-section of returns, with the approximation of changes in the log wealth-consumption ratio, it is convenient to restate the Euler equation in terms of betas,

$$\begin{aligned}
\mathrm{E}_t\left[R^e_{i,t+1}\right] &\approx -\mathrm{Cov}_t\left(R_{i,t+1}, m_{t+1}\right) \\
&= \gamma \mathrm{Cov}_t\left(R_{i,t+1}, \Delta c_{t+1}\right) - \left(\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}\right)\mathrm{Cov}_t\left(R_{i,t+1}, \Delta z_{t+1}\right) \\
&= \gamma \mathrm{Cov}_t\left(R_{i,t+1}, \Delta c_{t+1}\right) - \left(\frac{\frac{1}{\psi}-\gamma}{1-\frac{1}{\psi}}\right)\left[A\mathrm{Cov}_t\left(R_{i,t+1}, \Delta\hat{\mu}_{t+1}\right)\right. \\
&\left. + B\mathrm{Cov}_t\left(R_{i,t+1}, \Delta\hat{\sigma}_{t+1}\right)\right] \\
&= \beta^i_{\Delta c,t}\lambda_{\Delta c,t} + \beta^i_{\Delta\mu,t}\lambda_{\Delta\mu,t} + \beta^i_{\Delta\sigma,t}\lambda_{\Delta\sigma,t} \qquad \text{(A1)}
\end{aligned}$$

with

$$\begin{aligned}
\beta^i_{\Delta c,t} &= \frac{\mathrm{Cov}_t\left(R_{i,t+1}, \Delta c_{t+1}\right)}{\mathrm{Var}_t\left(\Delta c_{t+1}\right)} \quad \beta^i_{\Delta\mu,t} = \frac{\mathrm{Cov}_t\left(R_{i,t+1}, \Delta\hat{\mu}_{t+1}\right)}{\mathrm{Var}_t\left(\Delta\hat{\mu}_{t+1}\right)} \\
\beta^i_{\Delta\sigma,t} &= \frac{\mathrm{Cov}_t\left(R_{i,t+1}, \Delta\hat{\sigma}_{t+1}\right)}{\mathrm{Var}_t\left(\Delta\hat{\sigma}_{t+1}\right)},
\end{aligned}$$

and

$$\begin{aligned}
\lambda_{\Delta c,t} &= \gamma \mathrm{Var}_t\left(\Delta c_{t+1}\right) \quad \lambda_{\Delta\mu,t} = A\left(\frac{\gamma-\frac{1}{\psi}}{1-\frac{1}{\psi}}\right)\mathrm{Var}_t\left(\Delta\hat{\mu}_{t+1}\right) \\
\lambda_{\Delta\sigma,t} &= B\left(\frac{\gamma-\frac{1}{\psi}}{1-\frac{1}{\psi}}\right)\mathrm{Var}_t\left(\Delta\hat{\sigma}_{t+1}\right),
\end{aligned}$$

$\beta^i_{\Delta c,t}$, $\beta^i_{\Delta\mu,t}$, $\beta^i_{\Delta\sigma,t}$ denote risk loadings of asset i at date t with respect to consumption growth and changes in the perceived first and second moments of consumption growth, and $\lambda_{\Delta c,t}$, $\lambda_{\Delta\mu,t}$, $\lambda_{\Delta\sigma,t}$ are market prices of those betas.

Table A1: Change of Wealth-Consumption Ratio

We simulate 500 economies for 150 years at the quarterly frequency. Panel A and Panel B display regression results when $\gamma > \frac{1}{\psi}$ and $\psi > 1$ and $\psi < 1$ respectively. Panel C and Panel D display results when $\gamma < \frac{1}{\psi}$ and $\psi > 1$ and $\psi < 1$ respectively. In all the panels, the representative agent has a rate of time preference of 0.995. We report the average regression coefficients and average R^2s.

	Const.	$\Delta\hat{\mu}$	$\Delta\hat{\sigma}$	R^2
Panel A: $\gamma > \frac{1}{\psi}$, $\psi > 1$ ($\psi = 1.5$)				
$\gamma = 10$	-3.8e-9	2.0196	-0.1684	0.9981
$\gamma = 20$	-9.8e-9	2.1444	-0.3410	0.9915
$\gamma = 30$	-1.8e-8	2.2057	-0.5249	0.9800
$\gamma = 35$	-2.3e-8	2.2100	-0.6158	0.9729
Panel B: $\gamma > \frac{1}{\psi}$, $\psi < 1$ ($\psi = 0.5$)				
$\gamma = 10$	1.1e-8	-5.8516	0.4784	0.9982
$\gamma = 20$	2.8e-8	-6.2206	0.9738	0.9919
$\gamma = 30$	5.1e-8	-6.4232	1.5056	0.9809
$\gamma = 35$	6.4e-8	-6.4522	1.7710	0.9739
Panel C: $\gamma < \frac{1}{\psi}$, $\psi > 1$ ($\psi = 1.5$)				
$\gamma = 0.3$	-8.2e-11	1.8637	-0.0206	0.9999
$\gamma = 0.4$	-1.1e-10	1.8654	-0.0220	0.9999
$\gamma = 0.5$	-1.4e-10	1.8671	-0.0234	0.9999
$\gamma = 0.6$	-1.7e-10	1.8688	-0.0248	0.9999
Panel D: $\gamma < \frac{1}{\psi}$, $\psi < 1$ ($\psi = 0.5$)				
$\gamma = 1.6$	1.4e-9	-5.4679	0.1081	0.9999
$\gamma = 1.7$	1.5e-9	-5.4727	0.1122	0.9999
$\gamma = 1.8$	1.6e-9	-5.4774	0.1164	0.9999
$\gamma = 1.9$	1.7e-9	-5.4822	0.1205	0.9999

2.6.2 Robustness Test Results

In this section, we present additional empirical results for robustness.

Table A2: Summary Statistics of Options – Monthly

This table reports the descriptive statistics of the days-to-maturity, delta-hedged option returns, and idiosyncratic volatility of the underlying stock of selected contracts in **monthly** frequency. We exclude the following option observations: the stock underlying the option pays out cash over the options' remaining time-to-maturity; option price violates arbitrage bounds; reported trading volume is 0; option bid quote is 0 or midpoint of bid and ask quotes is less than 1/8. For each optionable stock, we keep one call and one put which are closest to being at-the-money and have the shortest time-to-maturity among others. Then we only keep calls and puts with moneyness within the range from 0.8 to 1.2 and days-to-maturity within the range from 31 to 53 days. Delta-hedged returns are calculated through option delta-hedged gains (given by equation (2.22)) scaled by $\Delta S - C$ for calls and $P - \Delta S$ for puts, where Δ is the Black-Scholes option delta, S is the underlying stock price, and C (P) is the price of call (put) option at the beginning of a month. Straddle returns are computed as the average returns of calls and puts which are written on the same stock and have the same strike price and time-to-maturity. We select straddles with moneyness closest to 1 and within the range from 0.8 to 1.2. Then we only keep straddles with days-to-maturity between 31 and 53 days. Days-to-maturity is the number of calendar days until option expiration. Idiosyncratic volatility is the standard deviation of the residuals of Fama-French 3-factors model estimated using the daily stock returns over the previous month. Calls, puts and straddles are reported in Panel A, Panel B and Panel C respectively. The option sample period is from January 1996 to December 2017.

	N	Mean	Median	SD	5th	10th	25th	75th	90th	95th
Panel A: Call Options										
Delta-hedged returns (%)	150,390	-0.97	-1.04	6.27	-9.20	-6.49	-3.24	0.87	4.06	7.34
Days to maturity	150,390	49	50	4	44	46	49	51	52	53
Idiosyncratic volatility (%)	150,390	2.24	1.80	1.77	0.58	0.75	1.14	2.83	4.23	5.34
Panel B: Put Options										
Delta-hedged returns(%)	101,496	-0.76	-0.99	4.94	-6.93	-5.09	-2.73	0.55	3.23	6.15
Days to maturity	101,496	49	50	4	39	46	47	51	52	53

Table A2 continued

Idiosyncratic volatility (%)	101,496	2.10	1.68	1.78	0.55	0.71	1.07	2.63	3.92	4.95
Panel C: Straddles										
Delta-hedged returns (%)	65,538	-0.79	-0.91	3.99	-6.07	-4.44	-2.45	0.58	2.82	4.99
Days to maturity	65,538	49	50	3	45	46	49	51	52	53
Idiosyncratic volatility(%)	65,538	2.13	1.71	1.78	0.55	0.71	1.09	2.67	3.99	5.04

Table A3: Idiosyncratic Risk Sorted Option Portfolios – Monthly

This table reports characteristics of equally-weighted option portfolios sorted by idiosyncratic volatility (IVOL) of the underlying stocks. Portfolios are rebalanced every **month**. Average return, standard deviation (SD), skewness and kurtosis of each portfolio are reported. Column "High–Low" shows the average return of the long-short strategy which buys in the highest IVOL portfolio and sells the lowest IVOL portfolio. Panel A presents portfolios formed with call option returns. Panel B presents put option portfolios and Panel C presents straddle portfolios. Newey-West (Newey and West (1987)) adjusted t-statistics are reported in parentheses. The sample period is from January 1996 to December 2017.

Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High–Low
Panel A: Call Options											
Mean Return (%)	-0.37	-0.46	-0.58	-0.65	-0.70	-0.90	-0.96	-1.12	-1.41	-2.19	-1.83
	(-3.37)	(-4.29)	(-4.93)	(-5.67)	(-5.23)	(-6.98)	(-5.78)	(-6.20)	(-7.44)	(-11.49)	(-9.69)
SD (%)	2.90	3.55	3.97	4.36	4.84	5.17	5.85	6.32	6.92	8.41	
Skewness	0.81	0.97	0.92	0.98	0.98	0.71	0.78	0.55	0.56	0.52	
Kurtosis	9.47	10.83	9.90	9.79	9.05	9.02	9.36	9.11	8.35	8.29	
Panel B: Put Options											
Mean Return (%)	-0.45	-0.38	-0.38	-0.36	-0.43	-0.47	-0.49	-0.70	-0.93	-1.90	-1.45
	(-4.11)	(-2.78)	(-2.37)	(-2.39)	(-2.37)	(-2.41)	(-2.49)	(-3.51)	(-4.27)	(-9.29)	(-8.69)
SD (%)	2.46	2.87	3.33	3.59	3.70	4.08	4.30	4.73	5.04	5.75	

Table A3 continued

Skewness	0.24	0.59	0.66	0.79	0.63	0.61	0.77	0.59	0.45	0.37	
Kurtosis	7.15	7.75	7.38	7.37	6.51	6.79	6.99	6.93	6.06	6.03	
					Panel C: Straddles						
Mean Return (%)	-0.38	-0.35	-0.39	-0.47	-0.45	-0.55	-0.54	-0.85	-1.13	-1.97	-1.60
	(-4.15)	(-3.68)	(-3.50)	(-3.81)	(-3.08)	(-3.29)	(-2.84)	(-4.48)	(-5.82)	(-11.24)	(-11.19)
SD (%)	1.53	1.97	2.30	2.58	2.79	3.18	3.44	3.89	4.69	5.58	
Skewness	0.35	0.46	0.53	0.55	0.44	0.32	0.35	0.23	0.25	0.11	
Kurtosis	4.74	4.75	4.77	5.24	4.69	4.85	4.48	4.85	5.23	4.79	

Table A4: Actively Traded Call Options – Monthly

This table reports summary statistics and characteristics of moneyness sorted portfolios of actively traded call options in **monthly** frequency. Panel A reports the descriptive statistics of days-to-maturity, delta-hedged option returns, and moneyness of actively traded call options. We exclude the following option observations: the stock underlying the option pays out cash over the options' remaining time-to-maturity; option price violates arbitrage bounds; reported trading volume is 0; option bid quote is 0 or midpoint of bid and ask quotes is less than 1/8; options have more than 5 recorded zero-trading-volume days within a month. We keep options with days-to-maturity between 44 to 86 days. Delta-hedged returns are calculated through option delta-hedged gains (given by equation (2.22)) scaled by $\Delta S - C$, where Δ is the Black-Scholes option delta, S is the underlying stock price, and C is the price of call option. Days-to-maturity is the number of calendar days until option expiration. Moneyness is the ratio of stock price over option strike price. Panel B reports characteristics of equally-weighted call option portfolios sorted by options moneyness. Portfolios are rebalanced every month. Average return, standard deviation (SD), skewness and kurtosis of each portfolio are reported. Column "High–Low" shows the average return of the long-short strategy which buys in the highest moneyness portfolio and sells the lowest moneyness portfolio. Newey-West (Newey and West (1987)) adjusted t-statistics with a lag of 12 are reported in parentheses. The option sample period is from January 1996 to December 2017.

Panel A: Summary Statistics

	N	Mean	Median	SD	5th	10th	25th	75th	90th	95th
Delta-hedged returns (%)	273,450	-0.82	-0.72	3.98	-6.97	-5.01	-2.51	0.75	3.21	5.37
Days-to-maturity	273,450	63	52	15	46	47	50	79	82	84
Moneyness=S/K (%)	273,450	99.34	98.21	14.36	79.40	84.50	91.60	105.36	114.95	123.04

Table A4 continued

Panel B: Moneyness Sorted Portfolios											
Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High–Low
Mean Return (%)	-1.73	-1.37	-1.12	-0.85	-0.68	-0.56	-0.48	-0.34	-0.23	-0.14	1.60
	(-8.87)	(-7.99)	(-7.85)	(-6.74)	(-5.67)	(-5.10)	(-4.56)	(-2.83)	(-1.91)	(-1.06)	(10.99)
SD (%)	4.84	4.18	3.70	3.29	2.99	2.73	2.57	2.44	2.37	2.41	
Skewness	0.32	0.27	0.25	0.25	0.22	0.16	0.08	0.05	0.04	0.01	
Kurtosis	2.46	2.85	3.27	3.77	4.31	4.75	5.18	5.25	5.97	5.84	

Table A5: Fama-MacBeth Regressions with Moneyness Sorted Call Option Portfolios – Monthly

This table reports estimated **monthly** risk premia on consumption betas. The sample constitutes only actively traded call options. Beta loadings are estimated from 10-year rolling window time-series regressions using equation (2.23). In the cross-section, we regress monthly future call option returns (in time $t+1$) on estimated beta loadings (see equation (2.24)). Newey-West (Newey and West (1987)) adjusted t-statistics with a lag of 12 are in parentheses. The sample period is from January 1996 to December 2017.

	Intercept	$\beta^i_{\Delta c,t}$	$\beta^i_{\Delta \mu,t}$	$\beta^i_{\Delta \sigma,t}$	Avg.R^2
I	-0.0013	0.1888			0.5412
	(-0.76)	(2.92)			
II	-0.0007	0.2300	0.0000		0.6664
	(-0.56)	(3.54)	(0.00)		
III	0.0008	0.1895		-0.0440	0.6955
	(0.88)	(3.48)		(-5.12)	
IV	0.0002	0.1970	0.0038	-0.0387	0.7635
	(0.18)	(3.56)	(0.46)	(-4.50)	

Table A6: Asset Pricing Tests with Moneyness Sorted Call Option Portfolios – Monthly

This table reports GMM estimates of the moment conditions in equation (2.25), the b estimates as well as the implied risk premia (λ). The sample constitutes only actively traded call options. MAE and RMSE refer to the mean absolute pricing error and the root mean squared error, respectively. The table presents the results using 10 **monthly** call option portfolios. Newey-West (Newey and West (1987)) adjusted t-statistics by using 12 lags are reported in the parentheses and p-values for J-statistics are shown in parentheses below the associated J-statistics. The sample period is from January 1996 to December 2017.

	Δc	$\Delta \mu$	$\Delta \sigma$	MAE	RMSE	J	R^2
b	8.0742			0.0032	0.0037	7.6629	0.9330
	(3.26)					(0.57)	
λ	1.0421						
	(3.26)						
b	8.5434	-6.7659		0.0032	0.0036	7.8067	0.9360
	(3.22)	(-0.46)				(0.45)	
λ	1.0405	0.0484					
	(2.81)	(0.68)					
b	5.2176		-21.6667	0.0028	0.0034	8.9654	0.9441
	(3.75)		(-2.67)			(0.35)	
λ	0.6864		-0.0391				
	(3.80)		(-2.85)				
b	4.1548	-20.1964	-40.3464	0.0026	0.0028	9.1786	0.9627
	(3.10)	(-2.83)	(-2.72)			(0.24)	
λ	0.3749	-0.0523	-0.0697				
	(2.10)	(-1.63)	(-2.79)				

Table A7: Fama-MacBeth Regressions with Moneyness Sorted Call Option Portfolios (A Smaller Sample)

This table reports estimated quarterly and monthly risk premia on consumption betas. The sample constitutes only one call option for each stock quarter (month). We select options with the shortest time-to-maturity for each stock per period. Beta loadings are estimated from 10-year rolling window time-series regressions using equation (2.23). In the cross-section, we regress quarterly (monthly) future call option returns (in time $t+1$) on estimated beta loadings (see equation (2.24)). We report quarterly Fama-MacBeth regression results in Panel A and monthly results in Panel B. Newey-West (Newey and West (1987)) adjusted t-statistics by using 4 lags for quarterly frequency and using 12 lags for monthly frequency are in parentheses. The sample period is from January 1996 to December 2017.

	Intercept	$\beta^i_{\Delta c,t}$	$\beta^i_{\Delta \mu,t}$	$\beta^i_{\Delta \sigma,t}$	Avg.R^2
			Panel A: Quarterly Risk Premium		
I	-0.0080	0.4117			0.3694
	(-0.84)	(1.60)			
II	-0.0004	0.5037	-0.0799		0.5719
	(-0.07)	(3.86)	(-2.69)		
III	0.0005	0.4905		-0.0613	0.6294
	(0.10)	(2.74)		(-3.39)	
IV	-0.0005	0.5404	0.0122	-0.0537	0.6861
	(-0.11)	(3.37)	(0.36)	(-2.98)	
			Panel B: Monthly Risk Premium		
I	-0.0013	0.2277			0.3524
	(-0.41)	(2.45)			
II	-0.0001	0.1962	-0.0402		0.5466
	(-0.06)	(3.47)	(-4.74)		
III	0.0002	0.2318		-0.0333	0.6249
	(0.09)	(3.08)		(-4.24)	
IV	-0.0004	0.2503	0.0136	-0.0315	0.6807
	(-0.20)	(3.27)	(1.13)	(-3.42)	

Table A8: Asset Pricing Tests with Moneyness Sorted Call Option Portfolios (A Smaller Sample)

This table reports GMM estimates of the moment conditions in equation (2.25), the b estimates as well as the implied risk premia (λ). The sample constitutes only one call option for each stock quarter (month). MAE and RMSE refer to the mean absolute pricing error and the root mean squared error, respectively. Panel A presents the results using 10 quarterly call option portfolios. Panel B presents the results using 10 monthly call option portfolios. Newey-West (Newey and West (1987)) adjusted t-statistics by using four lags for quarterly frequency and by using twelve lags for monthly frequency are reported in the parentheses and p-values for J-statistics are shown in parentheses below the associated J-statistics. The sample period is from January 1996 to December 2017.

	Δc	$\Delta\mu$	$\Delta\sigma$	MAE	RMSE	J	R^2
			Panel A: Quarterly Option Results				
b	12.4141			0.0106	0.0158	10.8390	-0.1825
	(3.84)					(0.29)	
λ	1.6022						
	(3.84)						
b	17.6847	-69.8072		0.0078	0.0090	9.0396	0.6134
	(3.49)	(-2.68)				(0.34)	
λ	1.6412	-0.1482					
	(3.26)	(-1.71)					
b	1.8282		-55.5122	0.0052	0.0056	7.5185	0.8503
	(2.02)		(-3.66)			(0.48)	
λ	0.2692		-0.0932				
	(2.25)		(-3.67)				
b	6.3846	-26.7550	-42.2080	0.0037	0.0044	9.6581	0.9078
	(4.58)	(-1.74)	(-1.81)			(0.21)	
λ	0.6035	-0.0610	-0.0743				
	(2.40)	(-0.86)	(-1.91)				

Table A8 continued

Panel B: Monthly Option Results							
b	10.0192			0.0050	0.0057	6.6398	0.8905
	(3.00)					(0.67)	
λ	1.2931						
	(3.00)						
b	10.6921	-10.0854		0.0048	0.0056	6.1379	0.8941
	(2.69)	(-0.77)				(0.63)	
λ	1.2873	0.0533					
	(2.76)	(1.03)					
b	4.7514		-49.5998	0.0028	0.0036	5.8072	0.9559
	(3.27)		(-3.82)			(0.67)	
λ	0.6429		-0.0852				
	(3.40)		(-3.92)				
b	5.5355	-14.3181	-51.2130	0.0029	0.0033	6.3147	0.9632
	(3.14)	(-1.29)	(-3.01)			(0.50)	
λ	0.6135	-0.0136	-0.0885				
	(2.71)	(-0.29)	(-3.11)				

Chapter 3

Switching Perspective: How Does Firm-Level Distress Risk Price the Cross-Section of Corporate Bond Returns?

We document a significantly negative relation between firm-level distress risk and the cross-section of corporate bond returns, analogous to the often negative relation between distress risk and stock returns in the prior literature ("distress anomaly"). Our evidence casts doubts on theories attributing the distress anomaly to shareholders exploiting debtholders in distress ("shareholder advantage"). In accordance, shareholder advantage proxies do not condition the distress risk-bond return relation. Conversely, we show that real options theories with disinvestment also have the potential to explain the anomaly, with disinvestment proxies conditioning the relation between distress risk and both stock and bond returns.
KEYWORDS: Distress risk, corporate bonds, shareholder advantage, disinvestment options.

3.1 Introduction

Recent empirical work finds a flat, hump-shaped, or negative relation between the probability that a firm fails to honor its fixed obligations ("distress risk") and the cross-section of stock returns.[1] The most convincing explanation for that finding is Garlappi, Shu and Yan's (2008) and Garlappi and Yan's (2011) shareholder advantage theory, which argues that shareholders' ability to extract economic rents from debtholders in default lowers stock risk and thus the returns of distressed stocks. Further evidence supporting that theory comes from Favara, Schroth and Valta (2012), who show that stock betas and volatilities are lower in countries whose institutions favor shareholders over debtholders, and Aretz, Florackis and Kostakis (2018), who show that the distress risk-stock return relation is more negative in the same countries. Also, Hackbarth, Haselmann and Schoenherr (2015) find that an exogenous increase in shareholder advantage in the United States in 1978 lowered stock betas and returns for all but most strongly distressed firms.

In our paper, we document that, analogous to the non-positive and often negative relation between firm-level distress risk and the cross-section of stock returns in the prior literature, there is also a negative relation between firm-level distress risk and the cross-section of corporate bond returns. In particular, using Campbell et al.'s (2008) hazard model to capture the probability of failure (defined as a default, bankruptcy filing, or performance-related delisting),[2] we find a monthly distress premium in bonds of –30 to –50 basis points in both portfolio sorts and Fama-MacBeth (FM; 1973) regressions. Akin to stocks, the premium is, however, only statistically significant when we control for popular stock and bond pricing factors, such as the bond market beta and bond-price momentum (see Bai, Bali and Wen (2019), Bali, Subrahmanyam and Wen (2019a) and Bali, Subrahmanyam and Wen (2019b)). Finally, the negative premium is attributable to inter-firm variations in distress risk. Keeping firm-level distress risk constant, intra-firm variations in distress risk due to variations in bond

[1] See, for example, Dichev (1998), Campbell, Hilscher and Szilagyi (2008), and Da and Gao (2010).

[2] A large literature suggests that hazard-model predictions of failure in general — and Campbell et al.'s (2008) prediction in particular — are vastly superior to, for example, discriminant-analysis or structural model-based predictions (see Shumway (2001), Chava and Jarrow (2004), Bharath and Shumway (2008), Campbell et al. (2008), and Aretz et al. (2018)).

indentures (as, e.g., in seniority, coupons, or collateral) are typically positively priced.

Our evidence that distress risk is negatively priced in corporate bonds comprises a serious blow to the shareholder advantage theory. The shareholder advantage theory starts from the premise that debtholders are entitled to a perpetual stream of coupon payments, but that they have awarded shareholders the option to cease payments in return for a to-be-negotiated fraction of firm value. Given that the option issued by debtholders is a perpetual American put option, its systematic risk — if held short — increases with the option exercise probability, which is equivalent to distress risk. Thus, the shareholder advantage theory predicts that the expected debt return increases with distress risk. To put that intuition on a more formal footing, we extend the simulation evidence of Garlappi et al. (2008), who employ Fan and Sundaresan's (2000) shareholder advantage model to create an artificial cross-section of expected stock returns and distress risk under realistic model inputs. Doing so, they find that high shareholder advantage can turn the expected stock return-distress risk relation negative. Picking up where they left off, we, however, show that the expected debt return-distress risk relation is consistently positive in their simulations, confirming that shareholder advantage cannot explain our bond pricing evidence.

To further show that shareholder advantage is not behind our bond evidence, we next condition the bond distress premium on popular shareholder advantage proxies, such as research and development (R&D) expenses, the Herfindahl sales index, and asset tangibility (see Garlappi et al. (2008) and Favara et al. (2012)). Since lower R&D expenses predict fewer cash-flow-related debt covenants, while a higher Herfindahl index and a lower asset tangibility predict greater fire-sale discounts in distress, low R&D expenses, a high Herfindahl index, and a low asset tangibility indicate high shareholder advantage. Double-sorted portfolios and FM regressions with interaction terms suggest that, while the shareholder advantage proxies usually continue to condition the distress risk-stock return relation (even within our smaller data sample), they are completely powerless to condition the distress risk-bond return relation. To make matters worse, the proxies tend to condition the bond distress premium with the wrong signs.

Given the limited success of the shareholder advantage theory to yield a unified explanation for the pricing of distress in stocks and bonds, we next take a fresh look at what could lie behind the distress anomaly. As first pointed out by Guthrie (2011), the relation between the expected return on a claim of a firm and the firm's condition is jointly determined by asset and financial risk in neoclassical finance models. Focusing on asset risk, Hackbarth and Johnson (2015), Aretz and Pope (2018), and Gu, Hackbarth and Johnson (2018) show that real-asset disinvestment options can lower the expected asset return of economically distressed firms because the disinvestment options can be interpreted as American put options with a negative systematic risk. Interestingly therefore, Table 2 in Campbell et al. (2008) suggests that firms classified by them as distressed are, on average, not only more financially levered but also less profitable than other firms.[3] It is thus entirely possible that a low asset risk, spurred by highly valuable negative systematic risk disinvestment options, lies behind the distress anomaly in both stocks and bonds.

We use numerical methods to value an equity claim and a zero-coupon debt claim on a firm owning production assets with embedded disinvestment options to study whether disinvestment risk can explain the distress anomaly. Assuming disinvestment proceeds go to shareholders unless they fall into a "suspect period" shortly before a debt default, in which case they go to debtholders, the model can produce a hump-shaped relation between distress risk and both stock and bond returns, which is more consistent with the empirical evidence than the shareholder advantage theory. To offer some more support for asset risk driving the distress anomaly, we then condition the stock and bond distress premia on Novy-Marx's (2013) gross profitability and Aretz and Pope's (2018) capacity overhang, defined as the difference between a firm's installed production capacity and its ex-ante optimal capacity.[4] While the first proxy measures economic profitability, the second measures how close a firm is to exercising

[3]Given that Campbell et al.'s (2008) profitability variable, NIMTA, contains financial expenses, it is not a pure proxy for *economic* profitability. Using operating profitability, defined as the difference between sales and costs of good sold scaled by total assets, we, however, find that operating profitability also strongly declines over their distress risk portfolios.

[4]Aretz and Pope (2018) define the ex-ante optimal capacity as that capacity level equalizing the marginal benefit of assets-in-place with the marginal cost of exercising growth options. See their paper for more technical details. An updated version of the capacity overhang proxy can be downloaded from: <https://www.kevin-aretz.com>.

its real-asset disinvestment options and thus also the value of these options. Double-sorted portfolios and FM regressions with interaction terms suggest that, with one exception, both gross profitability and capacity overhang significantly condition the relations between distress risk and both stock and bond returns with the correct signs.

Our work adds to studies on the distress premium in stocks. Dichev (1998), Griffin and Lemmon (2002), and George and Hwang (2010) show that Altman's (1968) Z-Score and Ohlson's (1980) O-Score, two accounting distress risk proxies, are flat in or decrease with stock returns. Extracting a distress risk proxy from Merton's (1974) model, Vassalou and Xing (2004) find a positive premium. Da and Gao (2010), however, question that premium, arguing it is attributable to illiquid stocks. Using the alternative structural distress risk proxy of Moody's KMV Corporation, Garlappi et al. (2008) and Garlappi and Yan (2011) find a hump-shaped relation between distress risk and stock returns. Anginer and Yıldızhan (2018) report that corporate credit spreads, which increase with risk-neutral distress risk, do not price stocks. Avramov, Chordia, Jostova and Philipov (2009) show that stock returns increase with credit ratings, implying a negative distress risk-stock return relation. Using an efficient hazard model proxy, Campbell et al. (2008) report a negative distress premium in stocks. We contribute to those studies by showing that, analogous to the often negative stock distress premium, the corporate bond distress premium can also be negative.

We also add to the literature by coming up with a new rationale for why both stock and bond returns decrease with distress risk. Prior studies often argue that financial risk lies behind the negative stock distress premium. As we already said, Garlappi et al.'s (2008) and Garlappi and Yan's (2011) shareholder advantage theory is the best-known example in that literature. Other examples include George and Hwang (2010), who reason that firms with high systematic risk induced through high financial distress costs endogenously choose low financial leverage ratios, and O'Doherty (2012), who speculates that high asset-value uncertainty drives down the systematic risk of distressed stocks. One caveat about these theories is that they often implicitly predict opposite effects of distress risk on stock and bond returns, inconsistent with our main empirical evidence. In contrast, we propose a real-asset based explanation for the

distress anomaly suggesting that both stock and bond returns decline with distress risk.

We proceed as follows. Section 3.2 describes our analysis variables and data sources. In Section 3.3, we study the relations between distress risk and the cross-sections of corporate bond, stock, and asset returns. In Section 3.4, we investigate whether the shareholder advantage theory explains our empirical findings. Section 3.5 explores whether real-asset disinvestment risk explains them. Section 3.6 gives the results from several robustness tests. Section 3.7 summarizes and concludes our paper.

3.2 Methodology and Data

In this section, we describe our methodology and data. We first outline the hazard model and credit ratings used to measure distress risk at the firm- and the bond-level, respectively. We next explain how we calculate the returns on bonds and other assets. We finally discuss our data sources.

3.2.1 Calculating Firm- and Bond-Level Distress Risk

We follow Campbell et al.'s (2008) hazard model methodology to measure twelve-month-ahead firm-level distress risk. In particular, we estimate a logit model of a dummy variable equal to one if a firm defaults on its debt obligations, files for bankruptcy, or delists for performance reasons over the next twelve months and else zero, *Failure*, on distress risk predictors measured at the start of the twelve-month period.[5] We can compactly write the logit model as:

$$\text{Prob}(Failure_{i,t} = 1|\mathbf{X}_{i,t-12}) = \frac{1}{1 + \exp(-\alpha - \beta\mathbf{X}_{i,t-12})}, \tag{3.1}$$

where α is a free parameter, β a vector of free parameters, and $\mathbf{X}_{i,t-12}$ a vector containing the distress risk predictors. Campbell et al. (2008) estimate logit model (3.1) recursively, using data from January 1963 to December of calendar year t, with t ranging from 1980 to 2003 in unit increments. They next combine the logit model

[5] Shumway (2001) shows that a logit model is a special form of hazard model.

estimates obtained from the estimation window extending to December of calendar year t with the distress risk predictor values over calendar year $t+1$. Doing so, they ensure that the logit model prediction could have been computed by investors in real-time.

The distress risk predictors in $\mathbf{X}$ contain *NIMTA*, the ratio of net income to the sum of the market value of equity and the book value of total liabilities ("market-value-adjusted total assets"); *TLMTA*, the ratio of the book value of total liabilities to market-value-adjusted total assets; *CASHMTA*, the sum of cash and short-term assets to market-value-adjusted total assets; and *MB*, the market-to-book ratio. To mitigate the effects of outliers on *MB*, Campbell et al. (2008) add 10% of the difference between the market value and the book value of equity to the book value of equity, setting book values of equity that continue to be negative to \$1. The vector $\mathbf{X}$ further contains *EXRET*, the monthly log stock return minus the monthly log S&P 500 return; *SIGMA*, a stock's volatility obtained from daily data over the prior three months;[6] *SIZE*, the log ratio of a stock's market capitalization to the S&P 500's total market capitalization; and *PRICE*, the log stock price truncated at \$15.

To enhance the distress risk predictors' timeliness, Campbell et al. (2008) use quarterly accounting data in their calculations, assuming that the accounting variable values become publicly available with a two-month reporting gap (i.e., two months after the end of the fiscal quarter). To guard against outlier effects, they winsorize the distress risk predictors at the $5th$ and $95th$ percentiles.[7]

We rely on corporate bond ratings issued by Moody's and S&P's to measure intra-firm variations in distress risk induced through the characteristics of a bond

[6] More specifically, they calculate volatility as the square root of 252 times the average of the squared daily stock return over the prior three months, assuming that the expected daily stock return is equal to zero. In case of stocks with fewer than five non-zero returns over the three-month period, they replace the volatility estimate with the cross-sectional mean of the volatility estimates of stocks with more than five non-zero returns over the same period.

[7] Since we do not have access to the failure data used by Campbell et al. (2008), we are unable to estimate logit model (3.1) ourselves. Fortunately, however, Jens Hilscher sent us the output from recursively estimating that model as described in the text. We use the logit model output obtained from the longest estimation window (1980-2008) to calculate our firm-level distress risk proxy for the post-2010 sample period. Doing so is unlikely to cause problems since the recursive estimates sent to us show strong signs of converging over the sample period extending to 2008. We thank Jens Hilscher and his co-authors for sharing the estimation output from their recursive logit model estimations with us.

issue — as opposed to the inter-firm variations captured by the firm-level distress risk proxy. To that end, we follow Bai et al. (2019) and assign a number to different ratings. In particular, we assign a value of one to AAA ratings, a value of two to AA+ ratings, and so on, until ultimately assigning a value of 21 to C ratings. As a result, investment-grade bonds have a value between one (AAA) and ten (BBB–), while non-investment-grade bonds have a value above ten. We finally compute *Rating* as the value associated with the most recent rating if only one agency issues ratings or the average of the most recent values if both agencies issue ratings.

3.2.2 Calculating the Returns on Corporate Bonds and Other Assets

In line with Bessembinder, Kahle, Maxwell and Xu (2009), Bao, Pan and Wang (2011), and Jostova, Nikolova, Philipov and Stahel (2013), we calculate the net return of corporate bond i over month t, $r_{i,t}$, using:

$$r_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1, \tag{3.2}$$

where P is the bond price, AI the accrued interest, and C the coupon payment. The price P is calculated as follows. To minimize confounding effects arising from bid-ask spreads, we start by calculating a bond's daily price as the trading-volume-weighted average of intra-day transaction prices over that day, as also done by Bessembinder et al. (2009). In line with Bai et al. (2019), we next calculate two types of bond returns, namely: (i) the return from the start of month t to the end of month t; and (ii) the return from the start of month t to the start of month $t+1$, where we define the start (end) of a month as the first (last) five trading days within that month. If we have more than one non-missing daily bond price within either the start- or end-of-month window, we choose the daily price closest to the first/last trading day of a month in our calculations. Finally, if we are able to calculate both types of returns, we use the start-of-month to start-of-month (type (ii)) return in our empirics.

To calculate the accrued interest AI, we first compute the daily coupon rate. The daily coupon rate is the coupon rate divided by 360 if a bond's day-count basis is "30/360" or "ACT/360," and it is the coupon rate divided by the actual number of calendar days per year if the day-count basis is "ACT/ACT." We next count the calendar days between the current month-end t and the previous coupon payment date, assuming that a month has 30 calendar days if the day-count basis is "30/360" and the actual number of days per month when it is "ACT/360" or "ACT/ACT." Also, we use the date of the first coupon payment and the coupon payment frequency to infer on which days the coupons are paid. We finally calculate the accrued interest AI as the daily coupon rate multiplied by the number of days between the current month-end t and the previous coupon payment date.

As is standard in the literature, we impose the following filters on our bond return data. First, we remove bonds not traded or listed in U.S. public markets. Second, we exclude bonds that are structured notes, are mortgage-, asset-, or agency-backed, or are equity-linked. Third, we remove convertible bonds. Fourth, we remove bonds with a price below \$5 or above \$1,000. Fifth, we keep only fixed and zero coupon bonds. Sixth, we remove bonds with less than one year to maturity. Seventh, we eliminate bond transactions that are labeled as when-issued or lock-in or have special sales conditions. Eighth, we remove transaction records that are canceled, subsequently corrected, or reversed. Finally, we only keep transactions with a trading volume that is larger than \$10,000.

In addition to bond returns, we also investigate the stock returns of the subsample of firms with bonds outstanding over our bond sample period (July 2002 to June 2017). While we directly obtain the stock returns from CRSP, we replace a stock's return over its delisting month with its delisting return if the delisting return is non-missing. If a stock's return over its delisting month is missing, we replace the return with –30% for NYSE and AMEX stocks and –55% for NASDAQ stocks, as advocated by Shumway (1997) and Shumway and Warther (1999). We do not exclude stocks with low prices from the stock subsample associated with our bond sample since only large well-capitalized firms issue bonds, rendering that restriction unnecessary. When we

later, however, shift our focus to a more comprehensive cross-section of stocks, we exclude stocks with a one-month-lagged price below \$1.

We finally also take a look at a firm's asset return, defined as the return to both its shareholders and debtholders. Since we are unable to observe the return on private debt, we approximate the asset return using a value-weighted average of the returns on a firm's stock and its outstanding bonds, using either the book or market leverage ratio to derive the weights. We assume that firms have only common stock outstanding (i.e., we ignore preferred stock), and we calculate the return on outstanding bonds as the value-weighted average of the returns on all of the firm's outstanding bond issues. In line with Fama and French (1992), we define the book leverage ratio as the ratio of the book value of assets to the book value of equity, while we define the market leverage ratio as the ratio of the book value of assets to the market value of equity, using the sum of common equity plus balance-sheet deferred taxes as book value of equity. We use the ratios from the fiscal year ending in calendar year $t-1$ to calculate weights from July of calendar year t to June of calendar year $t+1$.

3.2.3 Calculating Risk Factors and Control Variables

We use portfolio sorts and FM regressions to investigate the pricing of distress risk. In the portfolio sorts, we adjust for risk by regressing a portfolio's return on risk factors and reporting the intercept from that regression ("alpha"). As risk factors, we choose either the Fama and French (1993) five-factor model factors or the Bai et al. (2019) nine-factor model factors. The five Fama-French (1993) factors are the excess stock market return ($\text{MKT}^{\text{Stock}}$), the returns of stock spread portfolios formed on size (SMB) and the book-to-market ratio (HML), as well as the returns of bond spread portfolios formed on the term structure (TERM) and default risk (DEF). The term structure spread portfolio is long on long-term government bonds and short on one-month Treasury bills. Conversely, the default risk spread portfolio is long on long-term corporate bonds and short on long-term government bonds. The nine Bai et al. (2019) factors add to the former five factors the return on a stock spread

portfolio on momentum (MOM^{Stock}), Pastor and Stambaugh's (2003) stock liquidity risk factor (LIQ), the excess bond market return (MKT^{Bond}), and the return on a bond momentum spread portfolio (MOM^{Bond}). The excess bond market return is the return on a value-weighted portfolio of our sample bonds minus the one-month Treasury-bill rate. The bond momentum spread portfolio is long an equally-weighted portfolio of bonds with a past return over months $t-6$ to $t-1$ in the top decile and short an equally-weighted portfolio of bonds with that past return in the bottom decile (see Jostova et al. (2013)).[8]

In the FM regressions, we control for risk by including both stock and bond factor exposures and characteristics as control variables in our estimations. In particular, the bond-return regressions include a bond's exposures to the excess stock (MKT^{Stock}) and bond (MKT^{Bond}) market returns and to the SMB, HML, MOM^{Stock}, MOM^{Bond}, TERM, DEF, and LIQ spread portfolio returns. They further include a bond's years-to-maturity, log bond amount outstanding, most recent credit rating, and lagged one-month excess return. Conversely, the stock regressions include a stock's exposures to the excess stock and bond market returns and to the MOM^{Bond}, TERM, DEF, and LIQ spread portfolio returns, while directly adding the stock's one-month-lagged log market value of equity, log book-to-market ratio, and past-eleven-month compounded return.[9] In case of both the stock and bond portfolios, we estimate the exposures using rolling window regressions over the past 36 months of monthly data, winsorizing the estimated exposures at the $1st$ and $99th$ percentiles per month to mitigate outlier effects.

[8]To avoid losing the first seven months of our sample period, we use the bond momentum spread portfolio return from Gergana Jostova's website over the July 2002-January 2003 period in our empirical work.

[9]Following Fama and French (1992), we calculate the log book-to-market ratio as the log of the ratio of the book value of equity to the market value of equity, where the book value of equity is total assets minus total liabilities plus deferred taxes minus preferred stock from the fiscal year-end in calendar year $t-1$ and the market value of equity is the stock price times shares outstanding at the end of calendar year $t-1$. We use the computed value from July of calendar year t to June of calendar year $t+1$. Following Carhart (1997), we calculate the past-eleven-month momentum return as the compounded return over months $t-12$ to $t-2$, leaving a one-month gap between the compounding period and the current month t to avoid that the momentum return also captures short-term reversal effects.

3.2.4 Data Sources

We obtain stock data from CRSP and accounting data from Compustat. We collect bond data, including intraday transaction prices, trading volumes, and buy and sell indicators, from the enhanced version of the Trade Reporting and Compliance Engine (TRACE). In contrast to the Lehman Brothers Fixed Income Database, Datastream, and Bloomberg, which are quote-based databases, TRACE is a trade-based database, offering higher market transparency (see Bessembinder, Maxwell and Venkataraman (2006)) and covering about 99% of all public bond-market transactions since February 2005 (see Bao et al. (2011)). We rely on the Mergent Fixed Income Securities Database (FISD) to obtain bond characteristics, including offering-amount and -date, maturity date, coupon-rate, -type, and -payout frequency, bond-type, -rating, and -option features, and issuer information. We obtain MKT^{Stock}, SMB, HML, and MOM^{Stock} from Ken French's website, while we obtain LIQ from Lubos Pastor's website. We retrieve the corporate and government bond portfolio returns underlying the bond risk factors TERM and DEF from DataStream.

Our main bond sample period, determined by the availability of TRACE data, is July 2002 to June 2017. In our stock tests, we, however, sometimes rely on the longer sample period from January 1981 to December 2017, which is determined by our firm-level distress risk proxy.

3.3 The Pricing of Distress Risk in Corporate Bonds

In this section, we study the relation between firm-level distress risk and the cross-section of corporate bond returns. We start with offering summary statistics on our analysis variables. We next provide the mean excess returns and alphas of bond portfolios and their associated stock portfolios univariately sorted based on our firm-level distress risk proxy. We finally report the same statistics for bond portfolios double-sorted on both the firm-level distress risk proxy and intra-firm distress risk as captured by bond ratings as well as asset portfolios univariately sorted on firm-level

distress risk.

3.3.1 Summary Statistics

Table 3.1 reports summary statistics on our analysis variables, with Panels A, B, and C focusing on bond, stock, and firm characteristics, respectively. The summary statistics include the number of observations, the mean, standard deviation, and the $1st$, $5th$, $25th$, $50th$, $75th$, $95th$, and $99th$ percentiles. The table shows that our bond sample contains 556,965 bond-month observations over the sample period from July 2002 to June 2017. While the number of observations in our sample appears low compared to the number of observations used in other studies, we note that we lose many observations in the process of merging with the stock and firm characteristics data.[10] The average bond in our sample has a monthly return of 0.62%, a rating of 7.49 (BBB+), a market size of 0.57 billion dollars, and a time-to-maturity of 9.63 years. Conversely, the average stock has a monthly return of 0.96% and a market size of 58.6 billion dollars. The average twelve-month-ahead distress risk of the firms in our sample is only 0.09%, which is much lower than the average reported in Campbell et al. (2008). The reason is that bonds are almost exclusively issued by large firms, which tend to have a low distress risk.

[10]More specifically, our initial bond return sample contains 826,845 bond-month return observations, so that 269,880 observations are lost in the process of merging.

Table 3.1: Descriptive Statistics

In this table, we present descriptive statistics for our analysis variables. Panel A reports the number of bond-month observations, the cross-sectional mean, median, standard deviation, and selected percentiles of the monthly corporate bond return, and bond characteristics including the most recent credit rating, the years-until-maturity, and the market size (in billions). The credit rating is an integer between one and 21, with one referring to a triple A rating and 21 to a C rating. Panel B reports the number of stock-month observations, the cross-sectional mean, median, standard deviation and selected percentiles of the monthly stock return, and stock characteristics including market size (in billions). Panel C reports firm characteristics. Distress risk is the probability that the firm fails over the coming twelve months, calculated using the methodology of Campbell et al. (2008). Book leverage is the ratio of book assets to book equity, and market leverage is the ratio of book assets to market equity. Asset size is the book value of the firm's total assets, measured in billions. The sample period is from July 2002 to June 2017.

			Standard	Percentiles						
	Obs	Mean	Deviation	1	5	25	50	75	95	99
Panel A: Bond Characteristics										
Return (%)	556,965	0.62	4.48	–9.51	–3.68	–0.51	0.44	1.69	5.13	11.39
Credit Rating	556,965	7.49	3.49	1.00	1.00	5.00	7.00	9.50	14.50	16.00
Time-to-Maturity (years)	556,965	9.63	8.82	1.17	1.62	3.72	6.55	11.13	27.97	29.93
Market Size (in billions)	556,965	0.57	0.61	0.00	0.01	0.25	0.40	0.75	1.75	3.00
Panel B: Stock Characteristics										
Return (%)	556,965	0.96	10.21	–27.05	–13.15	–3.49	0.96	5.31	14.33	28.45
Market Size (in billions)	556,965	58.60	82.70	0.23	1.12	7.34	21.60	70.20	240.00	364.00

Table 3.1 continued

Panel C: Firm Characteristics										
Distress Risk (%)	556,965	0.09	0.32	0.01	0.01	0.03	0.04	0.06	0.21	1.14
Book Leverage	556,965	6.71	26.11	1.41	1.66	2.26	3.19	6.92	16.12	32.24
Market Leverage	556,965	4.23	6.51	0.28	0.49	1.11	1.94	4.00	17.07	32.25
Asset Size (in billions)	556,965	255.22	523.35	0.70	2.41	12.94	39.12	178.35	1787.63	2265.79

3.3.2 Portfolios Univariately Sorted on Firm-Level Distress Risk

We next analyze the relation between firm-level distress risk and the cross-section of corporate bond returns. To do so, we sort our bond sample into portfolios according to the decile breakpoints of the firm-level distress risk proxy distribution at the end of month $t-1$. We value- or equally-weight the portfolios, using the notional bond value outstanding at the end of month $t-1$ to calculate the value weights, and hold the portfolios over month t. We follow an analogous procedure to also sort the subsample of stocks associated with the bonds into value- or equally-weighted portfolios, using the market value of equity at the end of month $t-1$ to calculate the value weights. For each set of portfolios (i.e., the value or equally-weighted stock or bond portfolios), we create a spread portfolio long the highest distress risk portfolio and short the lowest portfolio. To adjust for systematic risk, we regress each portfolio's return on the five Fama and French (1993) factors or the nine Bai et al. (2019) factors introduced in Section 3.2.3 and report the alphas from these regressions.

Table 3.2: Bond and Stock Portfolios Univariately Sorted on Firm-Level Distress Risk

In this table, we present the mean excess returns and alphas of bond and stock portfolios univariately sorted on firm-level distress risk. We form the portfolios by sorting either bonds or stocks into portfolios according to the decile breakpoints of our firm-level distress risk proxy at the end of month $t-1$. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. We either value- (Panel A) or equally-weight the portfolios (Panel B) and hold them over month t. We calculate the bond weights using notional bond values outstanding and the stock weights using market equity values. We also form a spread portfolio long the highest distress risk decile and short the lowest ("High–Low"). The table reports the time-series average of the cross-sectional averages of distress risk, the average numbers of bonds/stocks per portfolio, and the average excess bond/stock returns, Fama-French five-factor alphas and Bai et al. (2019) nine-factor alphas for each bond/stock portfolio. Average distress risk, the average excess returns, and the alphas are in monthly percentage terms. We obtain the alphas from regressing a portfolio's return on the relevant factors and reporting the intercept from that regression. The five-factor model factors are the excess stock market return ($\text{MKT}^{\text{Stock}}$), the size factor (SMB), the value factor (HML), the term factor (TERM) and the default factor (DEF). The nine-factor model adds to these the stock momentum factor ($\text{MOM}^{\text{Stock}}$), the stock liquidity risk factor (LIQ), the bond market factor (MKT^{Bond}) and the bond momentum factor (MOM^{Bond}). Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are given in parentheses.

	Bonds					Stocks				
Decile	Mean Dist. Risk	Mean # Bonds	Mean Return	FF5 Alpha	B9 Alpha	Mean Dist. Risk	Mean # Stocks	Mean Return	FF5 Alpha	B9 Alpha
Panel A: Value-Weighted Distress Risk Portfolios										
1 (L)	0.01	377	0.56	0.49	0.21	0.01	71	0.45	–0.10	–0.12
2	0.02	384	0.50	0.46	0.14	0.02	71	0.58	0.06	0.06
3	0.03	382	0.60	0.54	0.20	0.03	71	0.56	–0.03	–0.08
4	0.03	382	0.50	0.42	0.08	0.03	71	0.65	–0.01	–0.08

Table 3.2 continued

5	0.04	383	0.46	0.37	0.04	0.04	71	0.81	0.12	0.10
6	0.05	382	0.45	0.31	–0.06	0.05	71	0.92	0.20	0.13
7	0.07	387	0.48	0.35	0.02	0.06	71	0.87	0.08	0.02
8	0.09	382	0.42	0.26	–0.10	0.09	71	0.65	–0.22	–0.24
9	0.14	388	0.45	0.23	–0.15	0.15	71	0.72	–0.15	–0.17
10 (H)	0.51	408	0.47	0.07	–0.34	0.76	72	0.33	–1.07	–1.12
H–L			–0.09	–0.41	–0.55			–0.13	–0.97	–1.01
t-stat.			[–0.34]	[–2.32]	[–2.69]			[–0.15]	[–1.92]	[–2.18]
Panel B: Equally-Weighted Distress Risk Portfolios										
1 (L)	0.01	377	0.58	0.51	0.26	0.01	71	0.73	0.11	0.05
2	0.02	384	0.53	0.48	0.19	0.02	71	0.75	0.13	0.07
3	0.03	382	0.64	0.58	0.27	0.03	71	0.80	0.12	0.05
4	0.03	382	0.57	0.50	0.17	0.03	71	0.89	0.16	0.05
5	0.04	383	0.48	0.40	0.12	0.04	71	1.10	0.31	0.19

Table 3.2 continued

6	0.05	382	0.52	0.40	0.01	0.05	71	1.17	0.35	0.26
7	0.07	387	0.52	0.39	0.09	0.06	71	1.00	0.11	–0.02
8	0.09	382	0.49	0.30	–0.12	0.09	71	1.07	0.03	–0.03
9	0.14	388	0.52	0.30	–0.11	0.15	71	1.08	–0.03	–0.20
10 (H)	0.51	408	0.52	0.10	–0.35	0.76	72	1.01	–0.57	–0.72
H–L			–0.05	–0.41	–0.62			0.28	–0.68	–0.77
t-stat.			[–0.17]	[–2.07]	[–2.72]			[0.30]	[–1.09]	[–1.81]

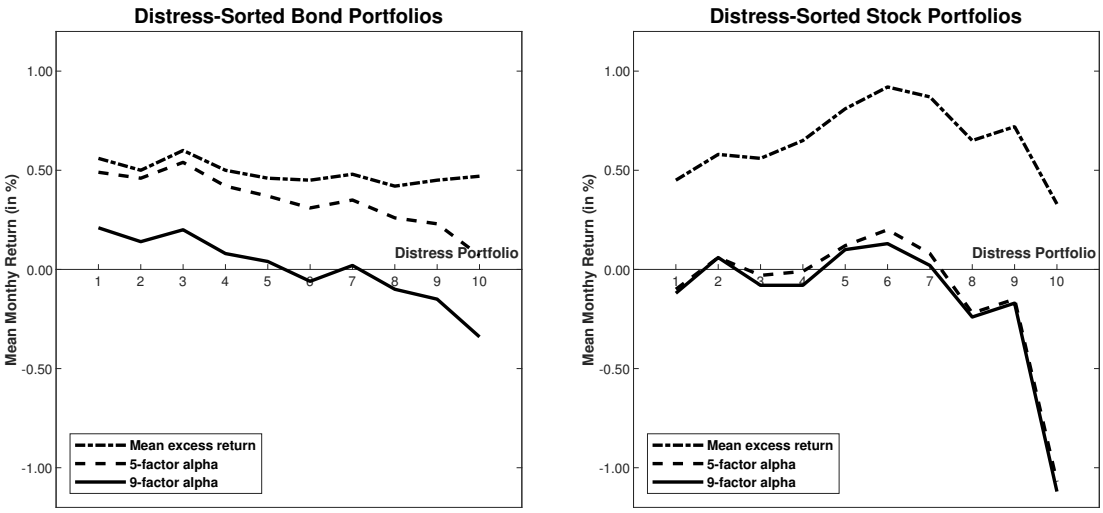
Table 3.2 [11] presents the mean excess returns, alphas, and other characteristics of the stock and bond distress portfolios, with Panel A focusing on the value-weighted and Panel B on the equally-weighted portfolios. Plain numbers are estimates, whereas the numbers in square parentheses are t-statistics calculated using Newey and West (1987) standard errors with a lag length of twelve months. The other characteristics are the time-series averages of the cross-sectional distress risk average and the number of assets (bonds or stocks) per portfolio. Consistent with Campbell et al.'s (2008) evidence on the stock pricing of distress risk, Table 3.2 suggests that the mean excess returns and alphas of the value- and equally-weighted bond portfolios decrease with distress risk. [12] Also consistent with Campbell et al. (2008), only the decreases in the alphas but not those in the mean excess returns are, however, statistically significant. [13] For example, Panel A suggests that, while the bond spread portfolio long the top and short the bottom value-weighted distress portfolio has an insignificant mean excess return of –0.09% per month (t-statistic: –0.34), its five-factor alpha is a significant –0.41% (t-statistic: –2.32) and its nine-factor alpha a significant –0.55% (t-statistic: –2.69). The left panel of Figure 3.1 graphically shows the relations between the mean excess bond returns and alphas and the distress portfolios.

[11] The empirical results are hardly affected by the financial crisis. To maintain the brevity of this paper, we do not report the empirical results before or after the financial crisis periods.

[12] We have verified that the bond offering yield increases with distress risk, and the emprical results are shown in Appendix A..

[13] We have conducted the monotonic relation test on whether the relations between distress risk and mean excess returns, FF5 alphas, and B9 alphas of the ten value-weighted bond portfolios are monotonic through applying the method developed by Patton and Timmermann (2010). Relevant results and analysis are discussed in Appendix B..

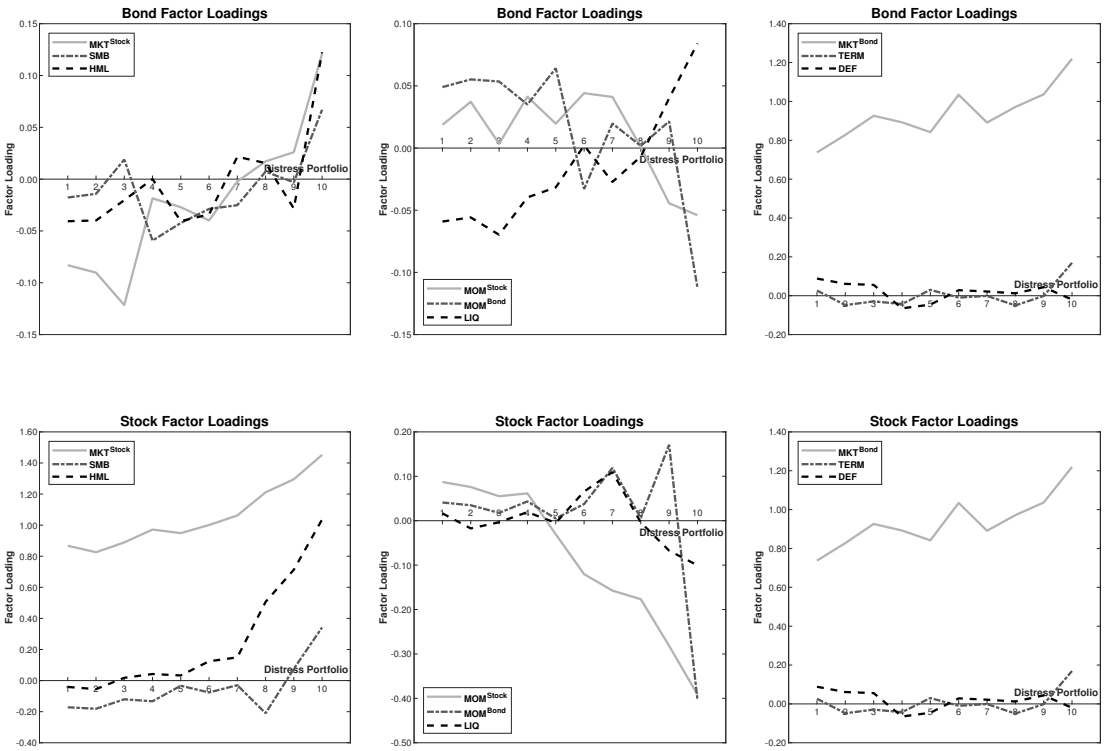
Figure 3.1: Mean Excess Returns and Alphas of Distress-Sorted Portfolios



This figure plots the mean excess returns and Fama-French five-factor and Bai et al. nine-factor model alphas of the value-weighted distress-sorted bond (left panel) and stock portfolios (right panel) over our sample period.

More directly corroborating Campbell et al.'s (2008) evidence, Table 3.2 further shows that the mean excess returns and alphas of the stock portfolios formed using only stocks associated with the bonds also decrease with distress risk. As before, however, only the decreases in the alphas but not those in the mean excess returns are significant. For example, Panel A suggests that, while the stock spread portfolio long the top and short the bottom value-weighted distress portfolio has an insignificant mean excess return of –0.13% (t-statistic: –0.15), its five- and nine-factor alphas are a significant –0.97% (t-statistic: –1.92) and –1.01% (t-statistic: –2.18), respectively. The right panel of Figure 3.1 graphically shows the relations between the mean excess stock returns and alphas and the distress portfolios.

Figure 3.2: Factor Exposures of Distress-Sorted Bond and Stock Portfolios



This figure plots the factor exposures of the value-weighted distress-sorted bond (upper panels) and stock portfolios (lower panels) over our sample period. We use the nine-factor model to estimate the factor exposures for each portfolio. We show the factor exposures on the factors, $\text{MKT}^{\text{Stock}}$, SMB and HML, in the left panel, the factors ($\text{MOM}^{\text{Stock}}$, MOM^{Bond} and LIQ) in the middle panel and the factors (MKT^{Bond}, TERM and DEF) in the right panel.

Figure 3.2 plots the exposures of the stock and bond distress portfolios on Bai et al.'s (2019) nine risk factors, shedding light on why the mean excess returns of the portfolios are so different from their alphas. The figure shows striking trends in the exposures over the portfolios. Starting with the bond portfolios, we see that, of the stock risk factors, the stock market and liquidity exposures increase almost monotonically over the distress portfolios, while the SMB, HML, and stock MOM exposures produce no discernable patterns. Conversely, of the bond market factors, only the bond market exposure but not the bond MOM, DEF, or TERM exposures increase over the distress portfolios. Given that both the stock and the bond market as well as the LIQ factor produce, on average, positive excess returns over our sample period,[14] it is no surprise that the alphas of the bond distress spread portfolios are

[14]To be more specific, the average monthly returns of the stock market portfolio, the bond market

significantly lower than their mean excess returns. Turning to the stock portfolios, the stock market, HML, and bond market exposures increase almost monotonically over the distress portfolios. Given that the HML factor also produces a positive average return over our sample period, it is also not surprising that the alphas of the stock distress spread portfolios are significantly lower than their mean excess returns.

3.3.3 Portfolios Double-Sorted on Firm-Level and Intra-Firm Distress Risk

A potential explanation for the negative relation between firm-level distress risk and corporate bond returns obtained in the previous subsection could be that distressed firms issue higher quality bonds than safer firms. Distressed firms may, for example, grant a higher priority to their bondholders and may issue more secured bonds. To refute that explanation, we now measure the quality of a bond issue using its most recently available credit rating and sort our bond sample into double-sorted portfolios according to their firm- and bond-level distress risk at the end of month $t-1$. As before, we either value- or equally-weight the portfolios and hold them over month t. We adjust for risk by regressing a portfolio's return on Bai et al.'s (2019) nine risk factors and reporting the intercept.

portfolio, and the liquidity spread portfolio are 0.70%, 0.53%, and 0.24% per month, respectively.

Table 3.3: Bond Portfolios Double-Sorted on Firm- and Bond-Distress Risk

In this table, we present the nine-factor model alphas of bond portfolios double-sorted on firm- and bond-level distress risk. In Panel A, we form the portfolios by sorting bonds into portfolios according to their most recent credit rating at the end of month $t-1$. Within each credit rating portfolio, we then sort into portfolios according to the quintile breakpoints of our firm-level distress risk proxy at the same time. In Panel B, we reverse the exercise, first sorting into quintile firm-level distress risk portfolios and then into credit-rating portfolios. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. The credit rating is an integer between one and 21, with one referring to a triple A rating and 21 to a C rating. Investment grade bonds have numbers from 1 to 10, speculative bonds from 11 to 13, highly speculative bonds from 14 to 16, and junk bonds from 17 to 21. We either value- (Panels A.1 and B.1) or equally-weight the portfolios (Panels A.2 and B.2) and hold them over month t. We calculate the bond weights using notional bond values outstanding and the stock weights using market equity values. Within each first-sorting-variable portfolio, we form a spread portfolio long the highest second-sorting-variable portfolio and short the lowest ("High–Low"). The table shows the average number of bonds per portfolio and the Bai et al. (2019) nine-factor alpha, in monthly percentage terms. See the caption of Table 3.2 for details on how we calculate the nine-factor model alpha. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses.

Panel A: First Sorting Variable: Credit Rating; Second: Distress Risk

	Credit Rating							
	Investment Grade		Speculative		Highly Speculative		Junk	
Dist. Risk	Obs	Alpha	Obs	Alpha	Obs	Alpha	Obs	Alpha
Panel A.1: Value-Weighted Distress Risk Portfolios								
1 (L)	605	0.15	99	0.31	39	0.34	33	0.55
2	600	0.13	97	0.17	38	0.14	33	0.36
3	602	–0.02	97	0.18	38	0.13	32	0.36
4	595	–0.13	97	0.10	38	0.02	33	0.01
5 (H)	565	–0.29	95	–0.16	37	–0.36	31	–0.73
H–L		–0.43		–0.47		–0.70		–1.28
t-stat.		[–2.58]		[–2.74]		[–2.73]		[–3.05]
Panel A.2: Equally-Weighted Distress Risk Portfolios								
1 (L)	605	0.20	99	0.32	39	0.41	33	0.61
2	600	0.21	97	0.22	38	0.28	33	0.38

Table 3.3 continued

3	602	0.05	97	0.24	38	0.24	32	0.38
4	595	–0.13	97	0.16	38	0.14	33	0.02
5 (H)	565	–0.30	95	–0.15	37	–0.29	31	–0.74
H–L		–0.50		–0.47		–0.70		–1.34
t-stat.		[–2.47]		[–2.81]		[–2.72]		[–3.10]

Panel B: First Sorting Variable: Distress Risk; Second: Credit Rating

	Distress Risk							
	Q1		Q2		Q3		Q4	
Cred. Rat.	Obs	Alpha	Obs	Alpha	Obs	Alpha	Obs	Alpha
Panel B.1: Value-Weighted Distress Risk Portfolios								
Invest. Gra.	756	0.15	747	0.08	744	–0.12	708	–0.27
Spec.	122	0.23	123	0.10	125	0.15	138	–0.12
High. Spec.	44	0.26	44	0.09	45	–0.04	45	–0.22
Junk	42	0.57	42	0.30	42	0.32	36	–0.14
H–L		0.42		0.22		0.44		0.13
t-stat.		[4.52]		[1.87]		[3.39]		[0.51]
Panel B.2: Equally-Weighted Distress Risk Portfolios								
Invest. Gra.	756	0.20	747	0.17	744	–0.09	708	–0.29
Spec.	122	0.27	123	0.18	125	0.18	138	–0.06
High. Spec.	44	0.33	44	0.18	45	0.09	45	–0.20
Junk	42	0.62	42	0.44	42	0.37	36	–0.08
H–L		0.42		0.26		0.46		0.21
t-stat.		[4.69]		[2.79]		[2.89]		[0.88]

Table 3.3 presents the results from the double-sorted portfolio formation exercise. In Panel A, we start with sorting our bond sample into four credit rating classes:

investment-grade (*Rating*: 1-10), speculative (11-13), highly speculative (14-16), and junk bonds (17-21). Within each class, we sort bonds into quintile portfolios according to their firm-level distress risk. Looking at value- and equally-weighted portfolios in Panels A.1 and A.2, respectively, Panel A suggests that the nine-factor alphas significantly decrease over the distress portfolios within each class. In fact, controlling for a bond's credit rating, the negative distress risk-bond alpha relations become more pronounced, with the alphas of the high-minus-low distress spread portfolios now never attracting a t-statistic above –2.50. Panel B reverses the exercise, first sorting into firm-level distress quintiles and then into the four credit rating classes. Looking at value- and equally-weighted portfolios in Panels B.1 and B.2, respectively, Panel B suggests that, except for the top distress quintile, the nine-factor alphas significantly increase over the credit rating classes within each distress quintile. Most pronouncedly, within the bottom distress quintile, the alphas increase by 0.42% as we move from the value- or equally-weighted investment-grade portfolio to the corresponding junk-bond portfolio (t-statistics about 4.60; see Panels B.1 and B.2).

3.3.4 Asset Portfolios Univariately Sorted on Firm-Level Distress Risk

We finally take a look at the relation between firm-level distress risk and asset returns. Table 3.4 shows the nine-factor alphas of value- or equally-weighted asset portfolios sorted on firm-level distress risk, with the portfolios being formed using the same procedures as before. The table suggests that the nine-factor alphas of the asset portfolios decrease with distress risk, which is perhaps unsurprising given that both stock and bond returns do so, too. Interestingly, however, the magnitudes of the decreases are slightly smaller than for the stock and bond portfolios, with, for example, the alphas of the high-minus-low distress spread portfolio now only being between –0.22% and –0.38% per month. In accordance, the t-statistics of the spread portfolio alphas are now slightly less significant, with them being around –1.90 except for the book-leverage value-weighted asset return.

Table 3.4: Asset Portfolios Univariately Sorted on Firm-Level Distress Risk

In this table, we present the nine-factor model alphas of portfolios of firms' assets univariately sorted on firm-level distress risk. We form the portfolios by sorting assets into portfolios according to the decile breakpoints of our firm-level distress risk proxy at the end of month $t-1$. We calculate a firm's asset return as a value-weighted average of its common stock return and its aggregate bond return. We either use the book values of equity and total liabilities ("book leverage asset return") or the market value of equity and the book value of total liabilities to compute the weights ("market leverage asset return"). The aggregate bond return is a value-weighted average of the returns on all of the firm's outstanding bond issues, using notional amounts to calculate the weights. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. We either value- or equally-weight the portfolios and hold them over month t. We also form a spread portfolio long the highest distress risk decile and short the lowest ("High–Low"). The table reports the time-series average of the cross-sectional averages of distress risk, the average numbers of assets per portfolio, and the Bai et al. (2019) nine-factor alphas per portfolio. Average distress risk and the alphas are in monthly percentage terms. See the caption of Table 3.2 for details on how we calculate the nine-factor model alpha. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses.

			Value-Weighted Portfolios		Equally-Weighted Portfolios	
			Book Lev. Asset Return	Market Lev. Asset Return	Book Lev. Asset Return	Market Lev. Asset Return
Decile	Mean Dist. Risk	Mean # Firms	9 Factor Alpha	9 Factor Alpha	9 Factor Alpha	9 Factor Alpha
1 (L)	0.01	54	0.31	0.50	0.26	0.28
2	0.02	53	0.30	0.36	0.26	0.31
3	0.02	54	0.25	0.34	0.22	0.29
4	0.03	53	0.23	0.27	0.22	0.25
5	0.04	53	0.28	0.31	0.29	0.35

Table 3.4 continued

6	0.04	54	0.24	0.29	0.22	0.32
7	0.06	54	0.24	0.32	0.22	0.24
8	0.08	54	0.14	0.09	0.18	0.19
9	0.12	54	0.36	0.35	0.21	0.17
10 (H)	0.54	53	0.09	0.12	0.01	-0.01
High–Low			–0.22	–0.38	–0.24	–0.28
t-statistic			[–1.24]	[–1.90]	[–1.98]	[–1.88]

Overall, we conclude from this section that there is a robust negative relation between firm-level distress risk and the cross-section of corporate bond returns, which becomes statistically significant once we control for popular risk factors. We are able to draw the same conclusions for the subsample of stocks associated with our bond sample. Variations in bond quality across differentially-distressed firms are not responsible for the negative distress risk-bond return relation, but work against it. Controlling for such variations, the negative relation becomes more pronounced and significant. Given the effects of firm-level distress risk on bond and stock returns, we also find a negative relation between firm-level distress risk and asset returns, which are the value-weighted average of stock and bond returns.

3.4 Does Financial Risk Explain the Bond Distress Premium?

In this section, we study whether financial risk can explain why stock and corporate bond returns decrease with distress risk. Garlappi et al.'s (2008) and Garlappi and Yan's (2011) shareholder advantage theory, for example, suggests that shareholders' ability to extract economic rents from debtholders in distress explains the negative stock distress premium. To see whether that theory can also explain a negative bond distress premium, we first repeat Garlappi et al.'s (2008) simulation exercise to identify the sign of the effect of shareholder advantage on the bond distress premium. We next rerun our asset pricing tests allowing the bond distress premium to depend on popular shareholder advantage proxies.

3.4.1 Shareholder Advantage and the Pricing of Distressed Debt

A. A Shareholder Advantage Model of the Firm

In line with Garlappi et al. (2008), we now study whether the shareholder advantage model of Fan and Sundaresan (2000) can explain the distress anomaly in stocks and

corporate bonds. Fan and Sundaresan (2000) look at a debt and equity-financed firm operating in continuous time indexed by t. The firm is exposed to a flat corporate tax rate of τ and loses a fraction of firm value α in bankruptcy ("deadweight costs of bankruptcy"). The value of the firm's unlevered assets, V_t, obeys:

$$dV_t = (\mu - \delta)V_t dt + \sigma V_t dB_t, \tag{3.3}$$

where μ is the expected return on the unlevered assets, $\delta < \mu$ the dividend yield, σ the volatility of the unlevered assets, and dB_t is the increment of a standard Brownian motion.

Turning to the financing side of the model, Fan and Sundaresan (2000) assume that the firm's entire debt takes the form of a single perpetuity with a coupon payment of c per time unit. Since the coupon payment is tax-deductable, it creates a tax shield. Shareholders are able to strategically default on the coupon payment. They use that possibility when the unlevered asset value V_t drops below the threshold level $\tilde{V}_S$ endogenously chosen by them. In default, shareholders and debtholders negotiate about the residual levered firm value, with shareholders ultimately receiving the fraction $\tilde{\theta}$ of residual value and debtholders the fraction $1-\tilde{\theta}$. The fractions are determined by maximizing the joint benefit to shareholders and debtholders in a Nash bargaining game:

$$\begin{aligned}\tilde{\theta}^* &= \operatorname{argmax}\left[\tilde{\theta}\upsilon(V) - 0\right]^{\eta}\left[(1-\tilde{\theta})\upsilon(V) - (1-\alpha)V\right]^{(1-\eta)} \\ &= \eta\left(1 - \frac{(1-\alpha)V}{\upsilon(V)}\right),\end{aligned} \tag{3.4}$$

where $\upsilon(V)$ is the levered asset value, and η shareholders' bargaining power. Equation (3.4) shows that the fraction of firm value allocated to shareholders in default, $\tilde{\theta}$, increases with shareholders' bargaining power, η, and the fraction of firm value lost in bankruptcy, α.

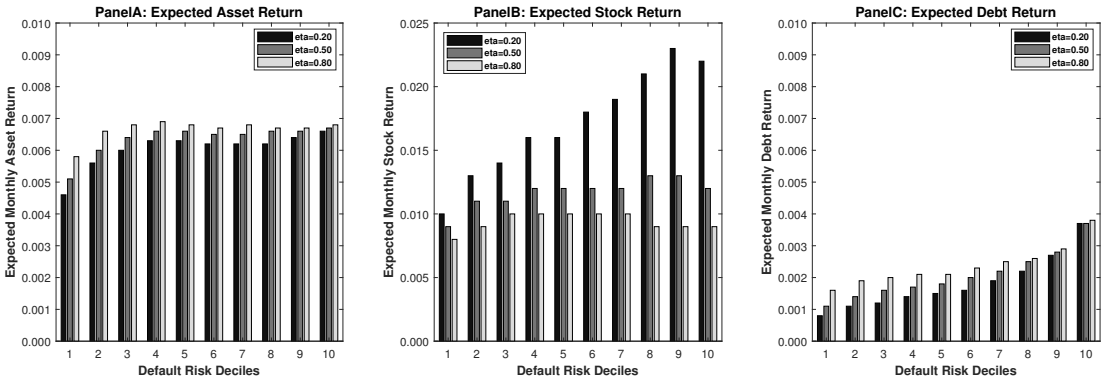
Using standard real options techniques outlined in, for example, Dixit and Pindyck (1994), Fan and Sundaresan (2000) derive closed-form solutions for the levered firm value, $\upsilon(V)$, the equity value, $\tilde{E}(V)$, the debt value, $\tilde{D}(V)$, and

the default threshold, $\tilde{V}_S$. Building up on Fan and Sundaresan's (2000) results, Garlappi et al. (2008) derive closed-form solutions for the time-0 expectation of the equity value at time t, $\mathbb{E}_0(\tilde{E}(V_t))$, and the probability that the unlevered asset value V_t hits the threshold $\tilde{V}_S$ over the period from time 0 to T, $\text{Prob}_{(0,T]}(V_0)$ ("strategic default probability"). Using the equity value expectation, they calculate the expected equity return, defined as the ratio of the expected equity value to its current value. We show the closed-form solutions derived by Fan and Sundaresan (2000) and Garlappi et al. (2008) in Appendix A. In the same appendix, we also derive the time-0 expectation of the debt value at time t, $\mathbb{E}_0(\tilde{D}(V_t))$, which, since $\tilde{D}(V) = \upsilon(V) - \tilde{E}(V)$, only requires us to derive the time-0 expectation of the levered asset value, $\mathbb{E}_0(\upsilon(V_t))$. Using the expected levered asset value and the expected debt value, we calculate the expected levered asset return and the expected debt return over and above the expected equity return.

B. Simulation Results

We use the closed-form solutions derived in Section A. to extend the simulation exercise of Garlappi et al. (2008). In line with them, we calculate expected returns over a one-month horizon and default risk over a one-year horizon, and we set the risk-free rate, r, to 0.04, the payout rate, δ, to 0.04, the tax rate, τ, to 0.35 and the bankruptcy costs, α, to 0.50. Also in line with them, we draw the coupon rate, c, the expected unlevered asset return, μ, and the initial unlevered asset value, V_0, from uniform distributions with support $[0.05, 0.10]$, $[\delta + \frac{1}{2}\sigma, 3(\delta + \frac{1}{2}\sigma)]$, and $[V_S, V_S + 1.25]$, respectively. Since we do not have access to the asset volatility estimates from Moody's KMV Corporation, we also draw these from a uniform distribution with support $[0.10, 0.30]$. Relying on a shareholder bargaining power η of 0.20, 0.50, or 0.80, we simulate 100,000 firms and calculate their expected levered asset returns, expected equity returns, expected debt returns, and default risk using the formulas in Appendix 3.8.2. We next sort the firms into ten decile portfolios according to their distress risk. We finally compute the equally-weighted expected levered asset-, equity-, and debt-returns of the portfolios.

Figure 3.3: Shareholder Advantage and Expected Asset, Stock, and Debt Returns



The figure shows the monthly expected levered asset returns (Panel A), stock returns (Panel B), and debt returns (Panel C) of decile portfolios sorted according to strategic default risk in the shareholder advantage model of Fan and Sundaresan (2000) with a shareholder bargaining power (eta) of either 0.20, 0.50, or 0.80. We describe the simulations and analytical formulas used to create the figure in Section 3.4.1 and Appendix 3.8.2, respectively.

Figure 3.3 plots the results from the simulations, with Panels A, B, and C focusing on the expected levered asset, equity, and debt return, respectively. Panel B corroborates Garlappi et al.'s (2008) result that a higher shareholder bargaining power, η, can turn the default risk-expected equity return relation from being almost monotonically positive to hump-shaped, with high default risk firms having a (marginally) lower expected equity return than low default risk firms. Conversely, consistent with the intuition that debtholders hold a zero-risk long perpetuity and a high-risk short put option entitling shareholders to default on their debt payments, and that the short option's risk increases with default risk, Panel C shows that the default risk-expected debt return relation is consistently positive in all our simulations. Perhaps surprisingly, the panel, however, also suggests that the relation does not become more but less positive with a higher shareholder bargaining power. Panel A hints at the reason, with it suggesting that the negative effect of shareholder bargaining power on the default risk-expected debt return relation stems from a similarly negative effect of shareholder bargaining power on the default risk-expected levered asset return relation. Notwithstanding, the most important takeaway is that, under realistic model input parameters, the Fan and Sundaresan (2000) shareholder advantage model produces a consistently positive default risk-expected debt return relation.

3.4.2 Conditioning the Bond Distress Premium on Shareholder Advantage Proxies

Section 3.4.1 presents theoretical evidence that shareholder advantage theories are unable to explain a negative distress premium in corporate debt including bonds. To further substantiate that evidence, we next condition the bond distress premium estimate obtained in Section 3.3 on popular shareholder advantage proxies, including a firm's R&D intensity, its industry concentration, and its asset tangibility. Opler and Titman (1994) show that highly levered firms with a high R&D intensity often encounter cash flow problems in recessions, triggering their cash-flow-related covenants and preventing them from renegotiating their debt. Conversely, Shleifer and Vishny (1992) and Acharya, Sundaram and John (2011) show that firms operating in concentrated industries and mostly owning intangible assets are often forced to sell their assets at fire-sale discounts in distress, making debtholders more willing to compromise to avoid a liquidation. Thus, the literature usually interprets a lower R&D intensity, a higher industry concentration, and a lower asset tangibility as signalling greater shareholder advantage.

We calculate a firm's R&D intensity as the ratio of its R&D expenses to its total assets. In accordance with Garlappi et al. (2008), we employ the sales-based Herfindahl index to measure an industry's concentration. We calculate that Herfindahl index for industry j as:

$$Herfindahl_j = \sum_{i=1}^{I_j} s_{i,j}^2,$$

where $s_{i,j}$ is the fraction of firm i's sales over the total sales of FF49 industry j, and I_j is the number of firms belonging to that industry. We calculate a firm's asset tangibility as the ratio of its gross property, plant and equipment (PPE) to its total assets. We take all accounting variables required to calculate the shareholder advantage proxies from the fiscal-year end in calendar year $t-1$. We use the proxies from July of calendar year t to June of calendar year $t+1$.

We start with using portfolio sorts to gauge the effect of the shareholder advantage

proxies on the distress premium in corporate bonds. To do so, we sort our bond sample (alternatively, the associated stock sample) into portfolios according to the tercile breakpoints of each shareholder advantage proxy at the end of month $t-1$. Within each shareholder advantage portfolio, we next sort the same assets into portfolios according to the quintile breakpoints of firm-level distress risk at the end of month $t-1$, giving us portfolios double-sorted on each shareholder advantage proxy and distress risk.[15] We either value- or equally-weight the double-sorted portfolios and hold them over month t, adjusting for risk by regressing each portfolio's return on the nine Bai et al. (2019) risk factors.

Table 3.5 presents the nine-factor alphas of the bond and stock portfolios double-sorted on shareholder advantage and distress risk. In Panels A to C, we use R&D intensity, the Herfindahl index, and asset tangibility to proxy for shareholder advantage, respectively. In each panel, the column titled "Strong (Weak) Shareholder Power" shows the alphas of those portfolios containing the 33% of firms with the highest (lowest) shareholder advantage according to the proxy used in the panel. Remarkably, the table suggests that, despite them being almost always significant, the declines in the bond alphas over the distress portfolios are virtually unrelated to shareholder advantage. Using asset tangibility to measure shareholder advantage, Panel C, for example, suggests that the decline in the value-weighted bond alpha is 0.31% (t-statistic: –2.48) for strong shareholder advantage firms and 0.48% (t-statistic: –2.38) for weak shareholder advantage firms. Looking at either value- or equally-weighted portfolios, the two other proxies, R&D intensity and the Herfindahl index, yield similarly narrow differences in the bond alpha declines over the set of distress portfolios (see Panels A and B).

[15] We only sort into two median portfolios in case of R&D intensity. We do so since, when following other studies and setting missing R&D expenditures equal to zero, more than half of all firms have zero R&D expenditures, making it impossible to sort into more granular (e.g., decile or quintile) portfolios.

Table 3.5: Bond and Stock Portfolios Double-Sorted on Firm-Level Distress Risk and Shareholder Advantage

In this table, we present the nine-factor model alphas of bond and stock portfolios double-sorted on firm-level distress risk and shareholder advantage. We form the portfolios by first sorting bonds or stocks into portfolios according to the tercile breakpoints of one of the shareholder advantage proxies at the end of month $t-1$. Within each shareholder advantage portfolio, we then sort them into portfolios according to the quartile breakpoints of our firm-level distress risk proxy at the same time. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. The shareholder advantage proxy is R&D intensity (Panel A), the sales-based Herfindahl index (Panel B), and asset tangibility (Panel C), with a low R&D intensity, a high Herfindahl index, and a low asset tangibility indicating strong shareholder advantage. R&D intensity is R&D expenses scaled by total assets. The Herfindahl index is the sum over firms' squared sales proportions within an industry. Asset tangibility is gross PP&E scaled by total assets. In case of R&D intensity, we are only able to sort into two (median-based) shareholder advantage portfolios since most firms have a zero R&D intensity. We either value- or equally-weight the portfolios and hold them over month t. We calculate the bond weights using notional bond values outstanding and the stock weights using market equity values. Within each shareholder advantage portfolio, we form a spread portfolio long the highest distress risk portfolio and short the lowest ("High–Low"). The table shows the Bai et al. (2019) nine-factor alphas for those double-sorted portfolios within the highest or lowest shareholder advantage portfolio. The alphas are in monthly percentage terms. See the caption of Table 3.2 for details on how we calculate the nine-factor model alpha. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses.

	Value-Weighted Portfolios				Equally-Weighted Portfolios			
	Bonds		Stocks		Bonds		Stocks	
	Shareholder Power		Shareholder Power		Shareholder Power		Shareholder Power	
Portfolio	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
	Panel A: Shareholder Power Proxy = R&D Expenses							
1 (L)	0.20	0.13	–0.09	0.02	0.26	0.19	–0.04	0.10
2	0.03	0.15	–0.10	0.04	0.15	0.22	0.07	0.18

Table 3.5 continued

3	–0.04	0.05	–0.18	0.21	–0.02	0.12	0.06	0.20
4 (H)	–0.18	–0.39	–0.45	–0.18	–0.24	–0.38	–0.36	0.09
High–Low	–0.38	–0.52	–0.35	–0.20	–0.50	–0.57	–0.33	–0.01
t-statistic	[–2.76]	[–2.53]	[–1.08]	[–0.63]	[–3.06]	[–2.45]	[–1.18]	[–0.03]
Panel B: Shareholder Power Proxy = Herfindahl Index								
1 (L)	0.17	0.18	0.01	–0.31	0.24	0.20	0.03	–0.12
2	0.18	0.09	0.06	–0.10	0.25	0.20	0.06	0.03
3	–0.09	–0.05	0.02	–0.04	0.01	–0.11	0.04	0.15
4 (H)	–0.25	–0.16	–0.62	–0.39	–0.22	–0.22	–0.31	–0.10
High–Low	–0.42	–0.34	–0.62	–0.08	–0.46	–0.43	–0.34	0.02
t-statistic	[–2.34]	[–2.44]	[–2.05]	[–0.26]	[–2.44]	[–2.43]	[–1.34]	[0.08]

Table 3.5 continued

Panel C: Shareholder Power Proxy = Asset Tangibility								
1 (L)	0.17	0.18	–0.01	0.03	0.23	0.28	0.07	0.08
2	0.10	0.05	0.06	–0.06	0.15	0.18	0.14	0.03
3	0.06	–0.03	0.15	–0.08	0.14	–0.04	0.16	0.05
4 (H)	–0.14	–0.30	–0.48	–0.30	–0.10	–0.27	–0.13	–0.34
High–Low	–0.31	–0.48	–0.47	–0.34	–0.33	–0.55	–0.20	–0.42
t-statistic	[–2.48]	[–2.38]	[–1.10]	[–0.95]	[–2.26]	[–2.76]	[–0.85]	[–1.01]

Turning to the stock portfolios, the situation changes dramatically. Supporting Garlappi et al. (2008), we find that the stock alpha declines over the distress portfolios are far more pronounced for strong than weak shareholder advantage firms. Using the Herfindahl index to measure shareholder advantage, Panel B, for example, suggests that the decline in the value-weighted stock alpha is 0.62% (t-statistic: –2.05) for strong and 0.08% (t-statistic: –0.26) for weak shareholder advantage firms. Using the other two shareholder advantage proxies, we find similarly large differences between the two types of firms (see Panels A and C). Notwithstanding, presumably due to the fact that we study a relatively narrow cross-section of stocks, the stock alpha declines are often insignificant.

We next also run FM regressions of bond (alternatively, stock) returns over month t on combinations of firm-level distress risk, the shareholder advantage proxies, interactions between firm-level distress risk and the shareholder advantage proxies, and controls measured at the end of month $t-1$. To mitigate that the firm-level distress risk proxy is heavily right-skewed, we take its natural log before entering it into the regressions. Also, instead of directly including the shareholder advantage proxies in the regressions, we rely on dummy variables signalling that shareholder advantage is high according to either shareholder advantage proxy. *LowR&D* is a dummy variable equal to one if a firm's R&D intensity is below the third quartile in a month, else zero; *HighHSI* is a dummy variable equal to one if a firm operates in an industry with a Herfindahl index value above the median, else zero; and *LowTangibibility* is a dummy variable equal to one if a firm's asset tangibility is below the median in a month, else zero. Table 3.6 [16] reports the results from the regressions, with Panel A focusing on the bond return regressions and Panel B on the stock return regressions. Plain numbers are monthly risk premium estimates (in percent), while the numbers in square parentheses are t-statistics calculated from Newey and West (1987) standard errors with a lag length of twelve months.

[16] In the 10^{th} comment, we are required to control for bond liquidity, rating, and time and industry fixed effects in Table 3.6. However, in our current table, we have already included bond liquidity beta and credit rating as control variables. Besides, it is impossible to add time fixed effects in FM regressions since FM regressions automatically include such a time fixed effect by allowing the intercept estimate to vary across the cross-sectional regressions. We try to add industry fixed effects in the FM regressions through applying the 49 Fama-French industry classification, but the results are similar to the current ones. Therefore, we decide to keep our current results.

Table 3.6: Regressions on Distress Risk and Shareholder Advantage

This table shows the results from Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead excess bond returns (Panel A) and excess stock returns (Panel B) on our firm-level distress risk proxy, the shareholder advantage proxies, interactions between the distress risk and the shareholder advantage proxies, and control variables. The firm-level distress risk proxy is the natural log of Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. The shareholder advantage proxies are based on R&D intensity, the sales-based Herfindahl index, and asset tangibility. LowR&D is a dummy variable equal to one if R&D expenses scaled by total assets is below the third quartile per month, else zero. HighHSI is a dummy variable equal to one if the Herfindahl index, the sum over firms' squared sales proportions within an industry, is above its median per month, also zero. LowTangibility is a dummy variable if gross PP&E scaled by total assets is below its median per month, else zero. In case of the bond return regressions, the control variables are $\beta^{MKT^{Stock}}$, β^{SMB}, β^{HML}, β^{TERM}, β^{DEF}, $\beta^{MOM^{Stock}}$, β^{LIQ}, $\beta^{MKT^{Bond}}$ and $\beta^{MOM^{Bond}}$, years-to-maturity, the natural log of bond amount outstanding, the most recent credit rating, and the lagged excess bond return. In case of the stock return regressions, the control variables are $\beta^{MKT^{Stock}}$, β^{TERM}, β^{DEF}, β^{LIQ}, $\beta^{MKT^{Bond}}$, $\beta^{MOM^{Bond}}$, the natural log of market equity, the natural log of book-to-market ratio, and the past eleven-month return. Betas are estimated using two-year rolling windows and are winsorized at the first and 99th percentiles. To keep the table concise, we do not report the estimates on the control variables. Plain numbers are estimates, in monthly percentage terms. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses. The final row of each panel further shows the average adjusted R^2 obtained from each Fama and MacBeth (1973) regression.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Bond Return Regressions				
Distress	–0.26	–0.23	–0.23	–0.27	–0.20	–0.23	–0.28	–0.24
	[–3.23]	[–2.91]	[–2.05]	[–3.15]	[–2.41]	[–2.92]	[–2.89]	[–1.54]
LowR&D		0.10	0.07					–0.02
		[1.94]	[0.09]					[–0.03]
Distress × *LowR&D*			0.01					0.00
			[0.07]					[–0.01]

Table 3.6 continued

HighHSI				–0.06	–0.80			–0.22
				[–1.77]	[–1.21]			[–0.38]
Distress × *HighHSI*					–0.10			–0.02
					[–1.19]			[–0.29]
LowTangibility						–0.03	0.80	0.45
						[–0.81]	[1.15]	[0.74]
Distress × *LowTangibility*							0.11	0.06
							[1.25]	[0.89]
Constant	–1.40	–1.26	–1.26	–1.43	–0.87	–1.14	–1.46	–1.13
	[–1.60]	[–1.32]	[–1.06]	[–1.61]	[–0.88]	[–1.21]	[–1.41]	[–0.78]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R^2	0.35	0.36	0.37	0.36	0.36	0.37	0.37	0.38

Table 3.6 continued

Panel B: Stock Return Regressions								
Distress	–0.06	–0.05	0.19	–0.06	0.08	–0.06	–0.05	0.37
	[–0.48]	[–0.39]	[1.48]	[–0.47]	[0.54]	[–0.46]	[–0.39]	[2.51]
LowR&D		–0.22	–2.65					–2.69
		[–1.45]	[–2.38]					[–2.36]
Distress × *LowR&D*			–0.30					–0.31
			[–2.36]					[–2.30]
HighHSI				–0.18	–2.18			–2.22
				[–1.20]	[–2.10]			[–2.24]
Distress × *HighHSI*					–0.25			–0.26
					[–2.11]			[–2.25]
LowTangibility						0.05	–0.01	–0.36
						[0.40]	[–0.01]	[–0.38]
Distress × *LowTangibility*							0.00	–0.04
							[–0.02]	[–0.38]

Table 3.6 continued

Constant	2.18	2.43	4.35	2.26	3.36	2.14	2.23	5.92
	[1.15]	[1.23]	[2.29]	[1.20]	[1.62]	[1.16]	[1.23]	[3.09]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R^2	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11

Starting with the bond regressions in Panel A, model (1) suggests that, using only distress risk and the controls as exogenous variables, distress risk earns a significantly negative premium of –26 basis points per month (t-statistic: –3.23). Allowing the shareholder advantage proxies to independently or jointly condition the negative distress premium, models (2) to (9) suggest that neither does so, with no interaction term attracting an absolute t-statistic larger than 1.25. Turning to the stock regressions in Panel B, model (1) suggests that using only distress risk and the controls as exogenous variables produces a negative albeit insignificant relation between distress risk and stock returns (t-statistic: –0.48). More importantly, models (2) to (9) show that two shareholder advantage proxies suggest that high shareholder advantage produces a significantly more negative distress risk-stock return relation, consistent with Garlappi et al. (2008). In particular, model (3) shows that a low R&D intensity leads the stock distress premium to decline by 30 basis points (t-statistic: –2.36), while model (5) shows that operating in a high Herfindahl index industry lowers it by 25 basis points (t-statistic: –2.11). In contrast, model (7) shows that a low asset tangibility does not affect the stock distress premium.

Overall, this section offers evidence that popular shareholder advantage proxies do not condition the bond distress premium obtained in Section 3.3, despite them continuing to condition the same premium in stocks even in our narrow cross-section and short sample period. Thus, shareholder advantage does not offer a consistent explanation for the distress premia in stocks and corporate bonds.

3.5 Does Asset Risk Explain the Bond Distress Premium?

In this section, we ask whether real options models of the firm are more successful in explaining why stock and corporate bond returns decrease with distress risk. Assuming investments are only partially reversible, the models suggest that *economically* unprofitable firms are close to exercising their disinvestment options, lowering their expected asset returns. Yet, if economic and financial distress change in tandem, disinvestment

options may also lead expected stock and debt returns to decline with financial distress. We first study that possibility within a real options model in which the firm can gradually disinvest capacity. The model is standard except for allowing the firm to be equity- *and* debt-financed. We next rerun our asset pricing tests allowing the stock and bond distress premiums to depend on disinvestment proxies.

3.5.1 The Pricing of Distressed Debt Under Disinvestment

A. A Real Options Model of the Firm Allowing for Disinvestment

We study a modified version of the standard real options model of Aretz and Pope (2018), who extend Pindyck's (1988) model to allow for the gradual disinvestment of productive capacity. In the model, a monopolistic firm operating in continuous time indexed by t optimally makes capacity and production decisions to maximize profits from producing and instantaneously selling some quantity of a homogenous output good. The firm has an initial productive capacity of $\bar{K}$. Each capacity unit allows the firm to produce and sell one unit of output per time unit, so that quantity, Q, is within $\{0; \bar{K}\}$. Each output unit is sold at a stochastic price, θ, evolving according to the differential equation:

$$d\theta = (\mu - \delta)\theta dt + \sigma\theta dW, \tag{3.5}$$

where μ is the total expected return, δ the dividend yield, and σ the volatility of the return of a traded asset replicating the variations in price, and W is a Brownian motion. The variable costs of producing Q units of output, $C(Q)$, are: $c_1 Q + \frac{1}{2}c_2 Q^2$, while the fixed costs, $F(\bar{K})$ are: $f\bar{K}$, where $c_1 \geq 0$, $c_2 \geq 0$, and $f \geq 0$ are parameters. The firm's total profits per time unit, $\pi(Q)$, are then:

$$\pi(Q) = \theta Q - c_1 Q - \frac{1}{2}c_2 Q^2 - f\bar{K}, \tag{3.6}$$

implying that the firm maximizes profits by choosing $Q = \min(\frac{\theta - c_1}{c_2}; \bar{K})$ in each instant. Finally, the firm is able to sell off productive capacity for a unit price equal to $s \geq 0$. In comparison to Aretz and Pope (2018), the only differences between our model and theirs is that (i) we do not allow the firm to expand its productive capacity, and (ii) we include fixed production costs $F(\bar{K})$. In Appendix 3.8.3, we show how to derive the firm's optimal disinvestment policy and how to value the firm.

Turning to the financing side of the model, we assume that the firm's entire debt takes the form of one single zero-coupon bond with a contractual payment of C and a maturity time of T. If the value available to debtholders exceeds C at time T, shareholders pay off debtholders, and the firm continues to exist. If it does not, the firm defaults and is liquidated. To satisfy their claims, debtholders have full recourse to the firm's productive assets at time T, but not past profits, which the firm instantaneously distributes to shareholders as dividends. In addition, although the firm also instantaneously distributes disinvestment proceeds to shareholders, debtholders are able to reclaim these proceeds in default if they fall within a "suspect period" (a legally-defined period preceding the default time; see Wood (2007, Chapter 17)). Using risk-neutral pricing, the value of debt, $D(\theta, C)$, is then:

$$D(\theta, C) = E^{\mathbb{Q}}\left[e^{-r(T-t)} \min(C, V(\theta(T), \bar{K}(T)) + S)\right], \tag{3.7}$$

where $E^{\mathbb{Q}}$ is the risk-neutral expectation, r the risk-free rate, $V(\theta(T), \bar{K}(T))$ the value of the remaining productive capacity at time T, and S the compounded-up value of the disinvestment proceeds that fell within the suspect period. Conversely, Cox and Rubinstein (1985) show that the instantaneous expected debt return is $\frac{\partial D(\theta,C)}{\partial \theta} \times \frac{\theta}{D(\theta,C)}$ multiplied by the expected excess return of the asset replicating variations in the price θ. Given that there is no closed-form solution for the expectation in Equation (3.7), we use Monte Carlo simulations to find the value and expected return of the debt claim.

We also calculate the equity value using a discounted risk-neutral expectation. Having done so, we again use Cox and Rubinstein's (1985) formula to derive the

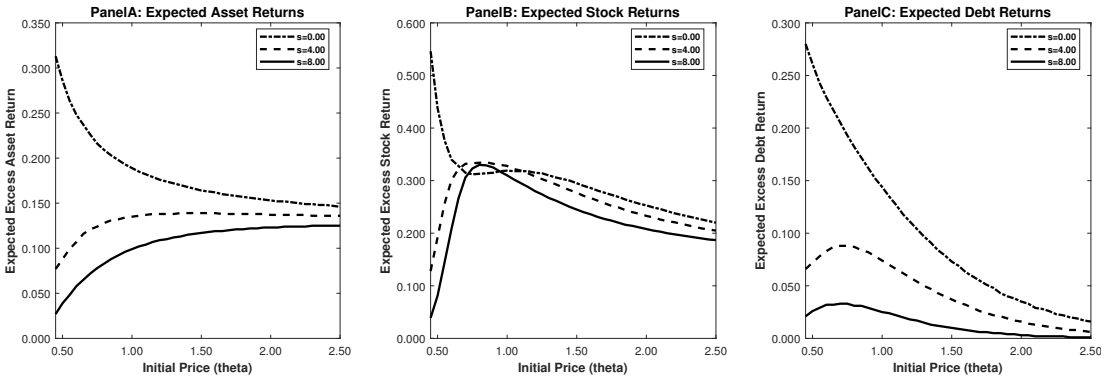
expected equity return.

B. Monte Carlo Simulation Results

We use Monte Carlo simulations to find out whether the real options model can produce negative relations between distress risk and both expected stock and debt returns, using 100,000 iterations for each set of parameters. We assume that the firm starts with an initial capacity, $\bar{K}$, of 1.00. Also, we set the total expected return, μ, the dividend yield, δ, and the volatility, σ, of the price replication asset equal to 16%, 4%, and 30%, respectively, while we set the production cost parameters, c_1, c_2, and f, equal to 0.00, 0.30, and 0.20, respectively. The zero-coupon bond has a contractual payment, C, of 10 and a maturity time, T, of 2.00. To vary the attractiveness of disinvestment, we set the disinvestment price, s, to 0.00, 4.00, or 8.00. We assume that shareholders capture disinvestment proceeds over the first year, but debtholders over the second. To vary the firm's economic (and also financial) health, we choose an initial price, θ, between 0.45 and 2.50, where 0.45 is slightly above the level below which the firm would instantaneously start disinvesting when the disinvestment price is at its highest value ($s = 8.00$).

Figure 3.4 plots the results from the simulation exercise, with Panels A, B, and C displaying the expected asset, equity, and debt return, respectively. Panel A suggests that the expected asset return increases (decreases) with the firm's economic health when the disinvestment price is low (high), in line with Aretz and Pope's (2018) main conclusions. The intuition is that when firms have disinvestment options and are close to exercising them, disinvestment options reduce asset risks and produce lower expected asset returns. Panels B and C reveal that disinvestment options also reduce the expected stock and debt returns of economically distressed firms since the benefits from disinvestment can accrue to both shareholders and debtholders. Taken together, the real options model thus confirms that disinvestment options can explain why both stock and bond returns decrease with distress risk.

Figure 3.4: Disinvestment Option and Expected Asset, Stock, and Debt Returns



The figure shows the expected asset returns (Panel A), stock returns (Panel B), and debt returns (Panel C) over different economic health conditions of firms in a modified version of the standard real options model of Aretz and Pope (2018) with a disinvestment attractiveness (s) of either 0.00, 4.00, or 8.00. We describe the simulations and analytical formulas used to create the figure in Section 3.5.1 and Appendix 3.8.3, respectively.

3.5.2 Conditioning the Bond Distress Premium on Disinvestment Options

Section 3.5.1 offers theoretical evidence that disinvestment options may be behind the negative relations between distress risk and both stock and corporate bond returns. To empirically test that possibility, we condition the stock and bond distress premia obtained in Section 3.3 on two disinvestment option value proxies, gross profitability and the extent to which a firm's installed capacity exceeds its optimal capacity ("capacity overhang"). A low gross profitability and high capacity overhang signal that a firm is economically unprofitable, suggesting that its disinvestment options are deep in-the-money. In line with Novy-Marx (2013), we calculate a firm's gross profitability as the ratio of the difference between its sales and costs of good sold ("gross profits") to its total assets. We take the values of the accounting variables from the fiscal year ending in calendar year $t-1$ and use the calculated ratio from July of calendar year t to June of calendar year $t+1$. In line with Aretz and Pope (2018), we use a recursively estimated stochastic frontier model to measure capacity overhang. The stochastic frontier model decomposes a firm's installed capacity into the sum of an optimal capacity estimate and a positively-signed capacity overhang residual. Installed capacity is proxied for using the log of the sum of PPE and

long-term intangibles. Conversely, optimal capacity is a linear function of optimal capacity determinants (as, e.g., log sales, log costs of goods sold, and stock volatility) and a normally-distributed mean-zero error term. Finally, the capacity overhang residual is a normally-distributed error term truncated from below at zero. Crucially, the expectation of that residual is a linear function of capacity overhang determinants (as, e.g., the percent decline in sales over some past period if positive, else zero). We provide more details about the capacity overhang variable in Appendix 3.8.4.

We again start with portfolio sorts to estimate the conditioning effect of disinvestment option value on the bond and stock distress premia. To do so, we sort our bond sample (alternatively, the associated stock sample) into portfolios according to the tercile breakpoints of each disinvestment value proxy at the end of month $t-1$. Within each disinvestment value portfolio, we next sort the same assets into portfolios according to the quintile breakpoints of firm-level distress risk at the end of month $t-1$, giving us portfolios double-sorted on disinvestment value and distress risk. As before, we either value- or equally-weight the portfolios and hold them over month t. We form a high-minus-low distress spread portfolio within each disinvestment value portfolio, once again adjusting the spread portfolio for risk by regressing its return on the nine Bai et al. (2019) factors.

Table 3.7 presents the nine-factor alphas of the bond and stock portfolios double-sorted on disinvestment value and distress risk, with Panel A using gross profitability and Panel B capacity overhang to measure disinvestment value. In each panel, the column titled "High (Low) Disinvestment Value" shows the alphas of those portfolios containing the 33% of firms with the highest (lowest) disinvestment values according to the proxy used in the panel. Starting with the bond portfolios, we see that a higher disinvestment value predicts a more negative distress risk-bond return relation. For example, Panel A suggests that, while the value-weighted bond distress spread portfolio has a mean monthly return of –0.24% (t-statistic: –2.65) for high gross-profitability (i.e., low disinvestment value) stocks, the same portfolio attracts a more than double mean return of –0.57% (t-statistic: –2.55) for low gross-profitability (i.e., high disinvestment value) stocks. In the same vein, Panel B suggests that the mean

return of that spread portfolio is –0.25% (t-statistic: –3.04) for low capacity overhang (i.e., low disinvestment value) stocks, but a much higher –0.42% (t-statistic: –2.38) for high capacity overhang (i.e., high disinvestment value) stocks.

Table 3.7: Stock and Bond Portfolios Double-Sorted on Firm-Level Distress Risk and Disinvestment Option Value

In this table, we present the nine-factor model alphas of bond and stock portfolios double-sorted on firm-level distress risk and disinvestment option value. We form the portfolios by first sorting bonds or stocks into portfolios according to the tercile breakpoints of one of the disinvestment option value proxies at the end of month $t-1$. Within each disinvestment option value portfolio, we then sort them into portfolios according to the quartile breakpoints of our firm-level distress risk proxy at the same time. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. The disinvestment option value proxies are operating profitability (Panel A) and capacity overhang (Panel B), with a lower operating profitability and a higher capacity overhang signalling more valuable disinvestment options. Operating profitability is gross profits scaled by total assets, while capacity overhang is an estimate of the difference between a firm's installed productive capacity and its optimal capacity derived using a stochastic frontier model. We either value- or equally-weight the portfolios and hold them over month t. We calculate the bond weights using notional bond values outstanding and the stock weights using market equity values. Within each disinvestment option value portfolio, we form a spread portfolio long the highest distress risk portfolio and short the lowest ("High–Low"). The table shows the Bai et al. (2019) nine-factor alphas for those double-sorted portfolios within the highest or lowest disinvestment option value portfolio. The alphas are in monthly percentage terms. See the caption of Table 3.2 for details on how we calculate the nine-factor model alpha. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses.

	Value-Weighted Portfolios				Equally-Weighted Portfolios			
	Bonds		Stocks		Bonds		Stocks	
	Divest. Option		Divest. Option		Divest. Option		Divest. Option	
Port.	Low	High	Low	High	Low	High	Low	High
Panel A: Divestment Option Proxy = Firm Profitability								
1 (L)	0.17	0.22	0.07	–0.40	0.23	0.30	0.15	–0.10
2	0.13	–0.03	0.08	0.02	0.20	0.04	0.21	0.19
3	0.08	–0.12	0.15	–0.23	0.16	–0.09	0.24	–0.04
4 (H)	–0.06	–0.35	–0.06	–0.63	–0.03	–0.38	0.32	–0.20
H–L	–0.24	–0.57	–0.13	–0.23	–0.26	–0.67	0.17	–0.10
t-stat.	[–2.65]	[–2.55]	[–0.49]	[–0.72]	[–2.54]	[–2.53]	[0.78]	[–0.33]

Table 3.7 continued

Panel B: Divestment Option Proxy = Capacity Overhang								
1 (L)	0.18	0.18	0.08	–0.04	0.22	0.27	0.16	0.14
2	0.10	0.18	0.03	0.12	0.18	0.24	0.27	0.21
3	0.02	–0.02	0.12	0.25	0.12	0.06	0.28	0.28
4 (H)	–0.06	–0.24	–0.30	–0.38	0.00	–0.28	0.08	–0.15
H–L	–0.25	–0.42	–0.38	–0.34	–0.22	–0.54	–0.07	–0.28
t-stat.	[–3.04]	[–2.38]	[–1.10]	[–1.05]	[–2.96]	[–2.78]	[–0.31]	[–0.81]

Importantly, the table shows that the conditioning effect of disinvestment value on the distress risk-stock return relation is also negative in three out of four cases. For example, Panel A suggests that, while the equally-weighted stock distress spread portfolio has a mean monthly return of 0.17% (*t*-statistic: 0.78) for high gross-profitability (i.e., low disinvestment value) stocks, the same portfolio attracts a much lower mean return of –0.10% (*t*-statistic: –0.33) for low gross-profitability (i.e., high disinvestment value) stocks. However, presumably again due to us studying a narrow cross-section of stocks over a short sample period, the stock spread portfolios never attract a significant alpha for either the high or low disinvestment value stocks (most negative *t*-statistic: –1.10).

Table 3.8 presents the results from FM regressions of bond (Panel A) or stock (Panel B) returns over month t on combinations of distress risk, the disinvestment value proxies, interactions between these variables, and control variables at the end of month $t-1$. Running the FM regressions, we are able to test for the significance of the conditioning effect of disinvestment value on the bond or stock distress premium. To alleviate skewness and kurtosis effects, we take the log of distress risk and the disinvestment value proxies before entering them into the regressions. As before, instead of directly including the disinvestment value proxies in the regressions, we again rely on dummy variables signalling that a firm is close to exercising its disinvestment options according to either proxy. *LowGrossProfits* is a dummy variable equal to one if a firm's gross profitability is below the median, else zero;

and *HighOverhang* is a dummy variable equal to one if a firm's capacity overhang is above the median, else zero. The control variables are exactly the same as those also used in Table 3.6.

Starting with the bond regressions, Panel A suggests that a higher disinvestment value significantly decreases the distress risk-bond return relation. Using either disinvestment value proxy, columns (3) and (5), for example, show that a lower gross profitability (signalling a higher disinvestment value) decreases the bond distress premium by 0.27% per month (t-statistic: –2.08), while a higher capacity overhang (also signalling a higher value) decreases that premium by 0.17% (t-statistic: –1.98). Jointly using the disinvestment value proxies, column (6) suggests that only gross profitability, but not capacity overhang, significantly conditions the distress premium in bonds.

Turning to the stock regressions, Panel B suggests that gross profitability, but not capacity overhang, also significantly conditions the distress risk-stock return relation with the anticipated sign in the model featuring both disinvestment value proxies (see column (6)). However, likely as a result of the limited sample size in these tests, the other conditioning effects fail to attract significance.

Overall, our empirical findings in this section suggest that low gross profitability or high capacity overhang, both signalling valuable disinvestment options, can go some way toward explaining the negative distress risk-bond return relation, as suggested by real options models of the firm. Alas, the same variables lack power to explain the distress risk-stock return relation in the subsample of stocks associated with our bond sample. In the next section, we thus study the variables' ability to explain that relation in a more comprehensive cross-section of stocks over a longer sample period.

Table 3.8: Regressions On Distress Risk and Disinvestment Option Value

This table shows the results from Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead excess bond returns (Panel A) and excess stock returns (Panel B) on our firm-level distress risk proxy, disinvestment option value proxies, interactions between the distress risk and the disinvestment option value proxies, and control variables. The firm-level distress risk proxy is the natural log of Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. The disinvestment option value proxies are based on operating profitability and capacity overhang, with a lower operating profitability and a higher capacity overhang signalling more valuable disinvestment options. LowGrossProfits is a dummy variable equal to one if gross profits scaled by total assets is below its median per month, else zero. HighOverhang is a dummy variable equal to one if capacity overhang, an estimate of the difference between a firm's installed productive capacity and its optimal capacity derived using a stochastic frontier model, is above its median per month, else zero. In case of the bond return regressions, the control variables are $\beta^{MKT^{Stock}}$, β^{SMB}, β^{HML}, β^{TERM}, β^{DEF}, $\beta^{MOM^{Stock}}$, β^{LIQ}, $\beta^{MKT^{Bond}}$ and $\beta^{MOM^{Bond}}$, years-to-maturity, the natural log of bond amount outstanding, the most recent credit rating, and the lagged excess bond return. In case of the stock return regressions, the control variables are $\beta^{MKT^{Stock}}$, β^{TERM}, β^{DEF}, β^{LIQ}, $\beta^{MKT^{Bond}}$, $\beta^{MOM^{Bond}}$, the natural log of market equity, the natural log of book-to-market ratio, and the past eleven-month return. Betas are estimated using two-year rolling windows and are winsorized at the first and 99th percentiles. To keep the table concise, we do not report the estimates on the control variables. Plain numbers are estimates, in monthly percentage terms. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses. The final row of each panel further shows the average adjusted R^2 obtained from each Fama and MacBeth (1973) regression.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Bond Return Regressions						
Dist.	–0.28	–0.28	–0.21	–0.28	–0.17	–0.10
	[–3.47]	[–3.36]	[–3.16]	[–3.52]	[–3.66]	[–2.22]
LowGrossProfits		0.05	–1.94			–1.82
		[0.77]	[–1.92]			[–1.82]
Dist. × *LowGrossProfits*			–0.27			–0.25
			[–2.08]			[–1.96]
HighOverhang				–0.01	–1.92	–1.17
				[–0.18]	[–1.95]	[–2.01]
Dist. × *HighOverhang*					–0.17	–0.15
					[–1.98]	[–2.04]
Constant	–1.57	–1.53	–1.09	–1.57	–0.79	–0.08
	[–1.85]	[–1.82]	[–1.44]	[–1.88]	[–1.19]	[–0.12]

Table 3.8 continued

Controls	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R2	0.41	0.41	0.43	0.41	0.42	0.43
Panel B: Stock Return Regressions						
Dist.	–0.04	–0.03	0.10	–0.03	0.01	0.08
	[–0.28]	[–0.17]	[0.80]	[–0.22]	[0.05]	[0.80]
LowGrossProfits		–0.22	–2.35			–2.33
		[–1.68]	[–2.11]			[–2.22]
Dist. × *LowGrossProfits*			–0.27			–0.27
			[–1.90]			[–2.01]
HighOverhang				–0.07	–0.56	0.09
				[–0.81]	[–0.54]	[0.10]
Dist. × *HighOverhang*					–0.06	0.02
					[–0.46]	[0.19]
Constant	2.47	2.63	3.52	2.57	2.89	3.43
	[1.21]	[1.29]	[1.99]	[1.27]	[1.62]	[2.06]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R2	0.10	0.11	0.11	0.11	0.11	0.11

3.6 Robustness Test

Since our idea to explain the negative relations between distress risk and the cross-sections of stock and corporate bond returns using disinvestment options is new to the literature, it is somewhat unnatural to immediately test that idea on the subsample of stocks of firms that also have bonds outstanding. To remedy that problem, we next estimate the conditioning effect of disinvestment options on the distress risk-stock return relation using a more comprehensive cross-section of stocks over a longer sample period. In particular, we now consider the entire cross-section of CRSP common stocks

traded on the NYSE, AMEX, and Nasdaq over the 1981 to 2017 sample period.[17] In line with other studies, we exclude financial (SIC code: 6000-6999) and utility stocks (4900-4949). To alleviate market microstructure biases, we further exclude stocks with a one-month-lagged market capitalization in the bottom quartile from both the equally-weighted portfolios and the FM regressions (see Hou, Xue and Zhang (2016)).

Table 3.9 presents the mean returns, alphas, and characteristics of value-weighted (Panel A) and equally-weighted distress risk portfolios (Panel B) formed using the same conventions as in Table 3.2, but featuring the more comprehensive cross-section of stocks over the longer sample period. Corroborating the evidence of Campbell et al. (2008), the main message of the table is that the more comprehensive data also produces a distress anomaly, which typically becomes significant when we control for the Fama and French (1993) three factors or the five factors which add to the former the stock momentum factor and the stock liquidity risk factor. More importantly, Table 3.10 shows the results from repeating the FM regressions conditioning the distress risk-stock return relation on the disinvestment proxies (gross profitability and capacity overhang) in Table 3.8 using the more comprehensive data. The table shows that a lower gross profitability (signalling a higher disinvestment value) yields a significantly more negative stock distress premium (see columns (3) and (6)), while it does not suggest that capacity overhang significantly conditions that premium.

In sum, the stock pricing tests conducted on the more comprehensive sample over the longer sample period thus offer some more evidence that operating profitability significantly conditions the distress risk-stock return relation, but, unfortunately, not that capacity overhang does the same.

[17]The starting point of the longer sample period is dictated by the availability of the distress risk proxy.

Table 3.9: Stock Portfolios Univariate Sorted on Firm-Level Distress Risk (1981–2017)

In this table, we present the mean excess returns and alphas of stock portfolios univariately sorted on firm-level distress risk over the extended sample period from 1981 to 2017. We form the portfolios by sorting stocks into portfolios according to the decile breakpoints of our firm-level distress risk proxy at the end of month $t-1$. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. We either value- (Panel A) or equally-weight the portfolios (Panel B) and hold them over month t. In case of the equally-weighted portfolios, we exclude stocks with a market size below the first quartile at the end of month $t-1$. We calculate the stock weights using market equity values. We also form a spread portfolio long the highest distress risk decile and short the lowest ("High–Low"). The table reports the time-series average of the cross-sectional averages of distress risk, the average numbers of stocks per portfolio, and the average excess stock returns, Fama-French three-factor alphas, and five-factor model alphas for each portfolio. Average distress risk, the average excess returns, and the alphas are in monthly percentage terms. We obtain the alphas from regressing a portfolio's return on the relevant factors and reporting the intercept from that regression. The three factors are the excess stock market return ($\text{MKT}^{\text{Stock}}$), the size factor (SMB), and the value factor (HML). The five factors add to the former the stock momentum factor ($\text{MOM}^{\text{Stock}}$) and the stock liquidity risk factor (LIQ). Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are given in parentheses.

Decile	Mean Distress Risk	Mean Number Stocks	Mean Excess Return	FF3 Alpha	5-Factor Alpha
Panel A: Value-Weighted Distress Risk Portfolios					
1 (L)	0.006	406	0.55	0.03	0.01
2	0.012	406	0.66	0.12	0.02
3	0.018	406	0.49	–0.01	–0.06
4	0.024	406	0.66	0.09	0.05
5	0.033	406	0.74	0.13	0.10
6	0.045	406	0.79	0.15	0.16
7	0.065	406	0.82	0.05	0.09
8	0.104	406	0.83	–0.01	0.12
9	0.199	406	0.34	–0.61	–0.42
10 (H)	0.999	407	0.10	–1.05	–0.62
H–L			–0.45	–1.07	–0.64
t-stat.			[–1.33]	[–3.72]	[–2.62]

Table 3.9 continued

Panel B: Equally-Weighted Distress Risk Portfolios					
1 (L)	0.006	304	0.60	–0.05	–0.02
2	0.011	305	0.64	–0.03	–0.04
3	0.016	304	0.71	0.04	0.05
4	0.022	305	0.83	0.12	0.14
5	0.028	305	0.81	0.06	0.11
6	0.037	304	0.80	0.02	0.09
7	0.052	304	0.81	0.00	0.11
8	0.076	305	0.80	–0.07	0.07
9	0.132	304	0.67	–0.30	–0.09
10 (H)	0.545	305	0.37	–0.76	–0.34
H–L			–0.24	–0.71	–0.32
t-stat.			[–0.91]	[–3.03]	[–1.56]

Table 3.10: Regressions On Distress Risk and Disinvestment Option Value (1981–2017)

This table shows the results from Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead excess stock returns on our firm-level distress risk proxy, disinvestment option value proxies, interactions between the distress risk and the disinvestment option value proxies, and control variables, estimated over the extended sample period from 1981 to 2017. The regressions exclude stocks with a market size below the first quartile at the end of month $t-1$. The firm-level distress risk proxy is the natural log of Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. The disinvestment option value proxies are based on operating profitability and capacity overhang, with a lower operating profitability and a higher capacity overhang signalling more valuable disinvestment options. LowGrossProfits is a dummy variable equal to one if gross profits scaled by total assets is below its median per month, else zero. HighOverhang is a dummy variable equal to one if capacity overhang, an estimate of the difference between a firm's installed productive capacity and its optimal capacity derived using a stochastic frontier model, is above its median per month, else zero. The control variables are $\beta^{MKT^{Stock}}$, β^{LIQ}, the natural logarithm of the market equity value, the natural logarithm of book-to-market ratio and the past 11-month average monthly returns as control variables. Betas are estimated using two-year rolling windows and are winsorized at the first and 99th percentiles. To keep the table concise, we do not report the estimates on the control variables. Plain numbers are estimates, in monthly percentage terms. Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are in parentheses. The final row of each panel further shows the average adjusted R^2 obtained from each Fama and MacBeth (1973) regression.

	(1)	(2)	(3)	(4)	(5)	(6)
Distress	0.09	0.12	0.18	0.13	0.16	0.22
	[1.43]	[1.78]	[2.75]	[1.89]	[2.45]	[3.29]
LowGrossProfits		–0.36	–1.28			–1.20
		[–4.04]	[–3.36]			[–3.12]
Distress × *LowGrossProfits*			–0.11			–0.11
			[–2.43]			[–2.36]
HighOverhang				–0.33	–0.84	–0.52
				[–3.40]	[–2.46]	[–1.50]
Distress × *HighOverhang*					–0.06	–0.03
					[–1.62]	[–0.80]
Constant	1.30	1.58	2.08	1.60	1.93	2.44
	[1.42]	[1.72]	[2.33]	[1.72]	[2.07]	[2.67]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R2	0.05	0.05	0.06	0.05	0.05	0.06

3.7 Conclusion

We offer empirical evidence suggesting a negative relation between firm-level distress risk and the cross-section of corporate bond returns, similar to the often negative relation between distress risk and stock returns obtained in prior studies. The negative distress risk-bond return relation becomes economically larger and statistically significant when controlling for popular stock and bond pricing factors, shows up in both value- and equally-weighted portfolio sorts and FM regressions, and is not attributable to distressed firms issuing higher-quality bonds than safer firms. Combining stock and bond returns to calculate a proxy for the asset return, we further offer evidence that distress risk is also negatively, albeit less significantly so, related to the cross-section of asset returns.

Our findings have important implications for the literature. In particular, they are first in casting some doubt on shareholder advantage explaining the distress anomaly, in particular, and shareholder advantage theories, in general. They do so since, as we show, shareholder advantage theories are unable to produce a negative relation between distress risk and debt returns. Consistent with that observation, popular shareholder advantage proxies fail to condition the bond distress premium estimate in our empirical work. We finally show that real options asset pricing models are more promising to explain why both stock and bond returns decrease with distress risk. These models predict that disinvestment options can lead asset returns to decrease with distress risk, with the low asset returns likely dragging down stock and bond returns, too. Supporting these models, disinvestment value proxies have some ability to condition the relations between distress risk and both stock and bond returns.

References

Acharya, V., Sundaram, R.K., John, K., 2011. Cross-country variations in capital-structures: The role of bankruptcy codes. Journal of Financial Intermediation 20, 25–54.

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

Anginer, D., Yıldızhan, Ç., 2018. Is there a distress risk anomaly? pricing of systematic default risk in the cross-section of equity returns. Review of Finance 22, 633–660.

Aretz, K., Florackis, C., Kostakis, A., 2018. Do stock returns really decrease with default risk? new international evidence. Management Science 64, 3821–3842.

Aretz, K., Pope, P.F., 2018. Real options models of the firm, capacity overhang, and the cross section of stock returns. Journal of Finance 73, 1363–1415.

Avramov, D., Chordia, T., Jostova, G., Philipov, A., 2009. Credit ratings and the cross-section of stock returns. Journal of Financial Markets 12, 469–499.

Bai, J., Bali, T.G., Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. Journal of Financial Economics 131, 619–642.

Bali, T.G., Subrahmanyam, A., Wen, Q., 2019a. Economic uncertainty premium in the corporate bond market. Working Paper.

Bali, T.G., Subrahmanyam, A., Wen, Q., 2019b. Long-term reversals in the corporate bond market. Journal of Financial Economics, forthcoming.

Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. Journal of Finance 66, 911–946.

Bessembinder, H., Kahle, K.M., Maxwell, W.F., Xu, D., 2009. Measuring abnormal bond performance. Review of Financial Studies 22, 4219–4258.

Bessembinder, H., Maxwell, W., Venkataraman, K., 2006. Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. Journal of Financial Economics 82, 251–288.

Bharath, S.T., Shumway, T., 2008. Forecasting default with the merton distance to default model. Review of Financial Studies 21, 1339–1369.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. Journal of Finance 63, 2899–2939.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Carlson, M., Fisher, A., Giammarino, R., 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. Journal of Finance 59, 2577–2603.

Chava, S., Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. Review of Finance 8, 537–569.

Cox, J.C., Rubinstein, M., 1985. Options markets. Prentice Hall.

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

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

Dixit, A.K., Pindyck, R.S., 1994. Investment under Uncertainty. Princeton University Press, Princeton, N.J.

Fama, E., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–36.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 42, 427–465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fan, H., Sundaresan, S.M., 2000. Debt valuation, renegotiation, and optimal dividend policy. Review of Financial Studies 13, 1057–1099.

Favara, G., Schroth, E., Valta, P., 2012. Strategic default and equity risk across countries. Journal of Finance 67, 2051–2095.

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

Garlappi, L., Yan, H., 2011. Financial distress and the cross-section of equity returns. Journal of Finance 66, 789–822.

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

Griffin, J.M., Lemmon, M.L., 2002. Book-to-market equity, distress risk, and stock returns. Journal of Finance 57, 2317–2336.

Gu, L., Hackbarth, D., Johnson, T., 2018. Inflexibility and stock returns. Review of Financial Studies 31, 278–321.

Guthrie, G., 2011. A note on operating leverage and expected rates of return. Finance Research Letters 8, 88–100.

Hackbarth, D., Haselmann, R., Schoenherr, D., 2015. Financial distress, stock returns, and the 1978 bankruptcy reform act. Review of Financial Studies 28, 1810–1847.

Hackbarth, D., Johnson, T., 2015. Real options and risk dynamics. Review of Economic Studies 82, 1449–1482.

Hou, K., Xue, C., Zhang, L., 2016. A comparison of new factor models. Working Paper.

Jostova, G., Nikolova, S., Philipov, A., Stahel, C., 2013. Momentum in corporate bond returns. Review of Financial Studies 20, 1649–1693.

Lewellen, J., Nagel, S., 2006. The conditional capm does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.

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

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, 1–28.

O'Doherty, M.S., 2012. On the conditional risk and performance of financially distressed stocks. Management Science 58, 1502–1520.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 18, 109–131.

Opler, T.C., Titman, S., 1994. Financial distress and corporate performance. Journal of Finance 49, 1015–1040.

Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Patton, A.J., Timmermann, A., 2010. Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. Journal of Financial Economics 98, 605–625.

Pindyck, R., 1988. Irreversible investment, capacity choice, and the value of the firm. American Economic Review 78, 969–85.

Shleifer, A., Vishny, R.W., 1992. Liquidation values and debt capacity: A market equilibrium approach. Journal of Finance 47, 1343–1366.

Shumway, T., 1997. The delisting bias in crsp data. Journal of Finance 52, 327–340.

Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. Journal of Business 74, 101–124.

Shumway, T., Warther, V.A., 1999. The delisting bias in crsp's nasdaq data and its implications for the size effect. Journal of Finance 54, 2361–2379.

Vassalou, M., Xing, Y., 2004. Default risk in equity returns. Journal of Finance 59, 831–868.

Wood, P.R., 2007. International loans, bonds, guarantees, legal opinions. volume 3. Sweet & Maxwell.

3.8 Appendix

3.8.1 Supplementary Empirical Tests

A. The Relation between Bond Yields and Distress Risk at Issuance

In Table A1, we show the relation between bond offering yields and distress risk through both portfolio sorts and regressions [18]. We find that, different from bond returns, bond yields at issuance increase as distress risk increases.

[18] We report the OLS regressions results of the pooled panel data here, and the FM regressions produce similar results.

Table A1: Bond Offering Yields and Firm-Level Distress Risk – Portfolio Sorts and Regressions

In this table, we present the mean offering yields and alphas of bond portfolios univariately sorted on firm-level distress risk (Panel A) and report the coefficients of regressing bond offering yields on distress risk and some control variables (Panel B). In Panel A, we form the portfolios by sorting bonds into portfolios according to the quantile breakpoints of our firm-level distress risk proxy at the end of month $t-1$. The firm-level distress risk proxy is Campbell et al.'s (2008) hazard-model probability that a firm fails over the coming twelve months. We either equally- or value-weight the portfolios and hold them over month t. We calculate the bond weights using notional bond values outstanding. We also form a spread portfolio long the highest distress risk quantile and short the lowest ("H–L"). The table reports the time-series average of the cross-sectional averages of distress risk, the average numbers of bonds per portfolio, and the average bond offering yields, Fama-French five-factor alphas and Bai et al. (2019) nine-factor alphas for each portfolio. Average distress risk, the average bond yields, and the alphas are in monthly percentage terms. We obtain the alphas from regressing a portfolio's return on the relevant factors and reporting the intercept from that regression. The five-factor model factors are the excess stock market return ($\text{MKT}^{\text{Stock}}$), the size factor (SMB), the value factor (HML), the term factor (TERM) and the default factor (DEF). The nine-factor model adds to these the stock momentum factor ($\text{MOM}^{\text{Stock}}$), the stock liquidity risk factor (LIQ), the bond market factor (MKT^{Bond}) and the bond momentum factor (MOM^{Bond}). Newey and West (1987)-adjusted t-statistics calculated using a twelve-month lag-length are given in parentheses. In Panel B, we regress bond offering yields on distress risk and some control variables. The control variables include bonds' years-to-maturity (*Maturity*), the most recent credit rating (*Rating*), the natural log of bond amount outstanding (*BondSize*), the natural log of market equity (*StockSize*), the natural log of book-to-market ratio (*BMRatio*) and the past eleven-month return (*StockMom*). The sample period is from July 2002 to June 2017.

Panel A: Portfolio Sorts

			Equally-Weighted			Value-Weighted		
Quantile	Mean Dist. Risk	Mean # Bonds	Mean Yield	FF5 Alpha	B9 Alpha	Mean Yield	FF5 Alpha	B9 Alpha
1 (L)	0.02	15	0.39	0.39	0.39	0.39	0.39	0.39
2	0.03	16	0.40	0.40	0.39	0.39	0.39	0.39
3	0.05	16	0.39	0.39	0.39	0.38	0.38	0.38

Table A1 continued

4 (H)	0.12	17	0.41	0.41	0.40	0.41	0.41	0.40
H–L			0.01	0.01	0.01	0.02	0.02	0.02
t-stat.			[2.46]	[2.20]	[2.41]	[2.11]	[1.94]	[2.16]

Panel B: Regressions

	(1)	(2)
Distress	0.37	0.26
	[17.67]	[14.14]
Maturity		0.003
		[21.89]
Rating		0.005
		[8.63]
BondSize		-0.01
		[-15.97]
StockSize		-0.04
		[-32.55]

Table A1 continued

BMRatio		0.002
		[1.10]
StockMom		-0.26
		[-4.36]
Constant	0.42	1.18
	[209.08]	[53.35]
Adj. R^2	0.03	0.27

B. Monotonicity Test

To investigate whether the relations between mean excess returns, FF5 alphas, and B9 alphas of bond portfolios and distress risk are monotonic, we apply the monotonic relation (MR) test method developed by Patton and Timmermann (2010), and report the results in Table A2. As shown in Table A2, none of the p-values of the MR test statistics reaches the critical value, which indicates that the MR test fails to find evidence in favor of a monotonic relation between distress risk and mean excess returns, FF5 alphas, or B9 alphas of bond portfolios. The hypothesis of MR test is very strict, and requires that the relation between two variables is strictly monotonic. However, we did not expect a strictly monotonicaly negative relation between bond returns and distress risk. In contrast, we expect a first flat and then decreasing relation between distress risk and bond returns. Therefore, failure to prove that the distress risk-bond returns relation is monotonically negative does not affect our conclusion that the distress risk and bond returns is generally negatively related.

Table A2: Monotonic Relation Test

This table reports the High–Low spread of mean excess returns, FF5 alphas, and B9 alphas of the ten value-weighted bond portfolios displayed in Table 3.2 and their corresponding t-statistics and t-statistic p-values. The last column displays the p-values of monotonic relation (MR) test statistics, which intend to test whether the relations between mean excess returns, FF5 alphas, and B9 alphas of bond portfolios and distress risk are monotonic. The MR test statistic p-values are obtained through appying the monotonic relation (MR) test method developed by Patton and Timmermann (2010).

	Mean Excess Return	FF5 Alpha	B9 Alpha
High–Low Spread	-0.09	-0.41	-0.55
t-statistic	-0.34	-2.32	-2.69
t-statistic p-value	0.737	0.021	0.008
MR p-value	0.841	0.756	0.805

3.8.2 The Fan and Sundaresan (2000) Model

A. Valuing the Equity and Debt Claims

Using contingent claims analysis, Fan and Sundaresan (2000) show that the value of a firm's levered assets in their shareholder advantage model, $\upsilon(V)$, is equal to:

$$\upsilon(V) = \begin{cases} V + \frac{\tau c}{r} - \frac{\lambda_2}{\lambda_2 - \lambda_1}\frac{\tau c}{r}\left(\frac{V}{\tilde{V}_S}\right)^{\lambda_1} & \text{if } V > \tilde{V}_S, \\ V + \frac{-\lambda_1}{\lambda_2 - \lambda_1}\frac{\tau c}{r}\left(\frac{V}{\tilde{V}_S}\right)^{\lambda_2} & \text{if } V \leq \tilde{V}_S, \end{cases} \tag{A1}$$

where the optimal (endogenous) strategic default threshold $\tilde{V}_S$ is given by:

$$\tilde{V}_S = \frac{c(1-\tau+\eta\tau)}{r}\frac{-\lambda_1}{1-\lambda_1}\frac{1}{1-\eta\alpha}, \tag{A2}$$

and:

$$\lambda_1 = \left(\frac{1}{2} - \frac{r-\delta}{\sigma^2}\right) - \sqrt{\left(\frac{1}{2} - \frac{r-\delta}{\sigma^2}\right)^2 + \frac{2r}{\sigma^2}} < 0, \tag{A3}$$

$$\lambda_2 = \left(\frac{1}{2} - \frac{r-\delta}{\sigma^2}\right) + \sqrt{\left(\frac{1}{2} - \frac{r-\delta}{\sigma^2}\right)^2 + \frac{2r}{\sigma^2}} > 1. \tag{A4}$$

Conversely, they show that the value of equity, $\tilde{E}(V)$, is equal to:

$$\tilde{E}(V) = \begin{cases} V - \frac{c(1-\tau)}{r} + \left[\frac{c(1-\tau)}{(1-\lambda_1)r} - \frac{\lambda_1(1-\lambda_2)\eta}{(\lambda_2-\lambda_1)(1-\lambda_1)}\frac{\tau c}{r}\right]\left(\frac{V}{\tilde{V}_S}\right)^{\lambda_1} & \text{if } V > \tilde{V}_S, \\ \theta^* \upsilon(V) & \text{if } V \leq \tilde{V}_S, \end{cases} \tag{A5}$$

where θ^* is given in Equation (3.4) in the main text. Finally, the value of debt, $\tilde{D}(V)$, is the value of the levered assets minus the value of equity, $\upsilon(V) - \tilde{E}(V)$.

B. Deriving the Expected Equity Value

Garlappi et al. (2008) show that the time-0 expectation of the equity value at time t, $\mathbb{E}_0(\tilde{E}(V_t))$, is:

$$\begin{aligned}\mathbb{E}_0(\tilde{E}(V_t)) = {} & \eta\alpha V_0 e^{(\mu-\delta)t} N\left(h(t)-\sigma\sqrt{t}\right) \\ & - \eta\frac{\lambda_1}{\lambda_2-\lambda_1}\frac{\tau c}{r}\left(\frac{V_0}{\tilde{V}_S}\right)^{\lambda_2} e^{\lambda_2(\gamma-\lambda_2)t} N\left(h(t)-\lambda_2\sigma\sqrt{t}\right) \\ & + V_0 e^{(\mu-\delta)t} N\left(-h(t)+\sigma\sqrt{t}\right) - \frac{c(1-\tau)}{r} N\left(-h(t)\right) \\ & + \left[\frac{c(1-\tau)}{(1-\lambda_1)r} - \frac{\lambda_1(1-\lambda_2)\eta}{(\lambda_2-\lambda_1)(1-\lambda_1)}\frac{\tau c}{r}\right]\left(\frac{V}{\tilde{V}_S}\right)^{\lambda_1} \\ & \times e^{\lambda_1(\gamma-\lambda_1)t} N\left(-h(t)+\lambda_1\sigma\sqrt{t}\right), \end{aligned} \tag{A6}$$

where $\gamma = \mu - \delta - \frac{1}{2}\sigma^2$, $h(t) = \frac{\ln(\tilde{V}_S/V_0)-\gamma t}{\sigma\sqrt{t}}$, and $N(.)$ is the cumulative standard normal distribution.

They further show that the probability of the unlevered asset value V hitting the strategic default threshold $\tilde{V}_S$ over the period from $t = 0$ to T ("strategic default risk"), $\Pr_{(0,T]}$ is:

$$\begin{aligned}\Pr_{(0,T]} = {} & N\left(\frac{\ln(\tilde{V}_S) - \ln(V_0) - \gamma T}{\sigma\sqrt{T}}\right) \\ & + e^{\frac{2\gamma(\ln(\tilde{V}_S)-\ln(V_0))}{\sigma^2}} N\left(\frac{\ln(\tilde{V}_S) - \ln(V_0) + \gamma T}{\sigma\sqrt{T}}\right). \end{aligned} \tag{A7}$$

C. Deriving the Expected Debt Value

In this section, we apply the methods used in Garlappi et al. (2008) to derive the time 0 expectation of the debt value at time t, $\mathbb{E}_0(\tilde{D}(V_t))$. Given that $\tilde{D}(V) = \upsilon(V) - \tilde{E}(V)$, we can easily achieve that goal by deriving the time 0 expectation of the levered asset value at time t, $\mathbb{E}_0(\upsilon(V_t))$. Under the assumptions in Section 3.4.1, the unlevered asset value at time t can be written as:

$$V_t = V_0 e^{(\mu-\delta-\frac{1}{2}\sigma^2)t+\sigma(B_t-B_0)}, \tag{A8}$$

which is log-normally distributed. Again defining $\gamma = \mu - \delta - \frac{1}{2}\sigma^2$, the location and scale parameters of the natural log of V_t are $E[\ln V_t] = \ln V_0 + \gamma t$ and $\text{Var}[\ln V_t] = \sigma^2 t$, respectively.

Consider the integral $\int_0^a V_t^b p(V_t) dV_t$, where a and b are constants and $p(V_t)$ is the probability density function of the log-normal variable V_t. Plugging in for $p(V_t)$, we obtain:

$$\int_0^a V_t^b p(V_t) dV_t = \int_0^a V_t^b \frac{1}{\sqrt{2\pi\sigma^2 t} V_t} e^{-\frac{1}{2}\left(\frac{\ln V_t - (\ln V_0 + \gamma t)}{\sigma\sqrt{t}}\right)^2} dV_t. \tag{A9}$$

Using the change of variable $X_t = \frac{\ln V_t - \ln V_0 - \gamma t}{\sigma\sqrt{t}}$, we can rewrite the right-hand side as:

$$\begin{aligned}
& \int_{-\infty}^{\frac{\ln(a/V_0)-\gamma t}{\sigma\sqrt{t}}} e^{b(\ln V_0 + \gamma t + \sigma\sqrt{t} X_t)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}X_t^2} dX_t & \text{(A10)} \\
= \; & V_0^b e^{b\gamma t} \int_{-\infty}^{\frac{\ln(a/V_0)-\gamma t}{\sigma\sqrt{t}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}X_t^2 + b\sigma\sqrt{t}X_t - \frac{1}{2}b^2\sigma^2 t + \frac{1}{2}b^2\sigma^2 t} dX_t & \text{(A11)} \\
= \; & V_0^b e^{b\gamma t + \frac{1}{2}b^2\sigma^2 t} \int_{-\infty}^{\frac{\ln(a/V_0)-\gamma t}{\sigma\sqrt{t}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_t^2 - 2b\sigma\sqrt{t}X_t + b^2\sigma^2 t)} dX_t & \text{(A12)} \\
= \; & V_0^b e^{b\gamma t + \frac{1}{2}b^2\sigma^2 t} \int_{-\infty}^{\frac{\ln(a/V_0)-\gamma t}{\sigma\sqrt{t}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_t - b\sigma\sqrt{t})^2} dX_t. & \text{(A13)}
\end{aligned}$$

Using the change of variable $Y_t = X_t - b\sigma\sqrt{t}$, we can write:

$$\begin{aligned}
& V_0^b e^{b\gamma t + \frac{1}{2}b^2\sigma^2 t} \int_{-\infty}^{\frac{\ln(a/V_0)-\gamma t}{\sigma\sqrt{t}} - b\sigma\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Y_t^2} dY_t & \text{(A14)} \\
= \; & V_0^b e^{b\gamma t + \frac{1}{2}b^2\sigma^2 t} N\left(\frac{\ln(a/V_0) - \gamma t}{\sigma\sqrt{t}} - b\sigma\sqrt{t}\right). & \text{(A15)}
\end{aligned}$$

Following the same steps, we can, conversely, also show that:

$$\int_a^\infty V_t^b p(V_t) dV_t = V_0^b e^{b\gamma t + \frac{1}{2}b^2\sigma^2 t} N\left(-\frac{\ln(a/V_0) - \gamma t}{\sigma\sqrt{t}} + b\sigma\sqrt{t}\right). \tag{A16}$$

Using Equation (A1), we can write the expected levered asset value, $\mathbb{E}_0(\upsilon(V_t))$,

as:

$$\begin{aligned}
\mathbb{E}_0(v(V_t)) &= \int_0^{\infty} V_t p(V_t) dV_t + \int_{\tilde{V}_S}^{\infty} \frac{\tau c}{r} p(V_t) dV_t \\
&- \int_{\tilde{V}_S}^{\infty} \frac{\lambda_2}{\lambda_2 - \lambda_1} \frac{\tau c}{r} \left(\frac{V_t}{\tilde{V}_S} \right)^{\lambda_1} p(V_t) dV_t \\
&+ \int_0^{\tilde{V}_S} \frac{-\lambda_1}{\lambda_2 - \lambda_1} \frac{\tau c}{r} \left(\frac{V_t}{\tilde{V}_S} \right)^{\lambda_2} p(V_t) dV_t \\
&= \int_0^{\infty} V_t p(V_t) dV_t + \frac{\tau c}{r} \int_{\tilde{V}_S}^{\infty} p(V_t) dV_t \\
&- \frac{\lambda_2}{\lambda_2 - \lambda_1} \frac{\tau c}{r} \left(\frac{1}{\tilde{V}_S} \right)^{\lambda_1} \int_{\tilde{V}_S}^{\infty} V_t^{\lambda_1} p(V_t) dV_t \\
&+ \frac{-\lambda_1}{\lambda_2 - \lambda_1} \frac{\tau c}{r} \left(\frac{1}{\tilde{V}_S} \right)^{\lambda_2} \int_0^{\tilde{V}_S} V_t^{\lambda_2} p(V_t) dV_t. \qquad \text{(A17)}
\end{aligned}$$

Using Equations (A15) and (A16), we finally have:

$$\begin{aligned}
\mathbb{E}_0(v(V_t)) &= V_0 e^{(\mu-\delta)t} + \frac{\tau c}{r} N\left(-h(t)\right) \\
&- \frac{\lambda_2}{\lambda_2 - \lambda_1} \frac{\tau c}{r} \left(\frac{V_0}{\tilde{V}_S} \right)^{\lambda_1} e^{\lambda_1(\gamma + \frac{1}{2}\lambda_1\sigma^2)t} N\left(-h(t) + \lambda_1 \sigma\sqrt{t}\right) \\
&+ \frac{-\lambda_1}{\lambda_2 - \lambda_1} \frac{\tau c}{r} \left(\frac{V_0}{\tilde{V}_S} \right)^{\lambda_2} e^{\lambda_2(\gamma + \frac{1}{2}\lambda_2\sigma^2)t} N\left(h(t) - \lambda_2 \sigma\sqrt{t}\right), \qquad \text{(A18)}
\end{aligned}$$

where we again use $h(t) = \frac{\ln(\tilde{V}_S/V_0) - \gamma t}{\sigma\sqrt{t}}$ to simplify the notation.

3.8.3 A Real Options Model with Disinvestment

A. Valuing the Operating Assets of the Firm

We use contingent claims analysis to value the *incremental* production options owned by the firm described in Section A.. Using $K \in \{0; \bar{K}\}$ to number the incremental options, incremental option K produces a cash flow of $\theta - c_1 - c_2 K - f$ per time unit when switched on to produce output and a payoff of $-f$ per time unit when switched off. Denoting the value of incremental option K by $\Delta V(\theta; K)$ and assuming that there is a traded asset whose value perfectly replicates variations in

the price θ, it is well known that the value of incremental option K has to satisfy:

$$\frac{1}{2}\sigma^2\theta^2\frac{\partial^2\Delta V(\theta;K)}{\partial\theta^2}+(r-\delta)\theta\frac{\partial\Delta V(\theta;K)}{\partial\theta}-r\Delta V(\theta;K)+\pi(\theta,K)=0, \quad \text{(A19)}$$

where $\pi(\theta,K)$ is the cash flow produced by the option.

In the θ-region in which the firm uses the incremental option to produce output (i.e., in which $\pi(\theta,K)=\theta-c_1-c_2K-f$), the value of the option takes on the general form:

$$\Delta V(\theta,K)=A_O\theta^{\beta_1}+B_O\theta^{\beta_2}+\frac{\theta}{\delta}-\frac{c_1+c_2K+f}{r}, \quad \text{(A20)}$$

where A_O and B_O are free parameters, and:

$$\beta_1 = \frac{1}{2}-(r-\delta)/\sigma^2+\sqrt{\left[(r-\delta)/\sigma^2-\frac{1}{2}\right]^2+2r/\sigma^2}>1, \quad \text{(A21)}$$

$$\beta_2 = \frac{1}{2}-(r-\delta)/\sigma^2-\sqrt{\left[(r-\delta)/\sigma^2-\frac{1}{2}\right]^2+2r/\sigma^2}<0. \quad \text{(A22)}$$

Given that $\lim_{\theta\to+\infty}\Delta V(\theta,K)$ needs to be $\frac{\theta}{\delta}-\frac{c_1+c_2K+f}{r}$, it is obvious that $A_O=0$. Conversely, in the region in which the firm does not use the incremental option to produce output (i.e., in which $\pi(\theta,K)=-f$), the value of the option takes on the general form:

$$\Delta V(\theta,K)=A_I\theta^{\beta_1}+B_I\theta^{\beta_2}+-\frac{f}{r}, \quad \text{(A23)}$$

where A_I and B_I are free parameters. Finally, in the region in which the firm instantaneously sells the option (i.e., when θ drops below the disinvestment threshold θ^D, which is another free parameter), the value of the incremental option is equal to the disinvestment price s.

To find the values of the free parameters B_O, A_I, B_I, and θ^D, we ensure that the three regions value-match and smooth-paste into one another. In particular,

we ensure that:

$$B_O(\theta^P)^{\beta_2} + \frac{(\theta^P)}{\delta} - \frac{c_1 + c_2K + f}{r} = A_I(\theta^P)^{\beta_1} + B_I(\theta^P)^{\beta_2} - \frac{f}{r}, \quad \text{(A24)}$$

$$B_O\beta_2(\theta^P)^{\beta_2-1} + \frac{1}{\delta} = A_I\beta_1(\theta^P)^{\beta_1-1} + B_I\beta_2(\theta^P)^{\beta_2-1}, \quad \text{(A25)}$$

$$A_I(\theta^D)^{\beta_1} + B_I(\theta^D)^{\beta_2} - \frac{f}{r} = s, \quad \text{(A26)}$$

$$A_I\beta_1(\theta^D)^{\beta_1-1} + B_I\beta_2(\theta^D)^{\beta_2-1} = 0, \quad \text{(A27)}$$

where $\theta^D = c_1 + c_2K$ is the price θ at which the firm switches on the incremental option. Equation (A24) ensures that at θ^P the value of the used option is identical to the value of the idle option, while Equation (A25) ensures that, at that price, the two option values do so with identical partial derivatives (i.e., smoothly). Conversely, Equation (A26) ensures that at the price at which the firm disinvests off the incremental option, θ^D, the value of the option is identical to the disinvestment price, while Equation (A27) ensures that, at that price, the option value has a zero partial derivative.

Solving for B_O, A_I, B_I, and θ^D, we obtain:

$$p^D = \left(\frac{r\delta(\beta_1 - \beta_2)(s + \frac{f}{r})(c_1 + c_2K)^{\beta_1 - 1}}{(r - \beta_2(r - \delta))\left(1 - \frac{\beta_1}{\beta_2}\right)} \right)^{\frac{1}{\beta_1}}, \quad \text{(A28)}$$

$$A_I = \frac{s + \frac{f}{r}}{(p^D)^{\beta_1}\left(1 - \frac{\beta_1}{\beta_2}\right)}, \quad \text{(A29)}$$

$$B_I = \frac{s + \frac{f}{r}}{(p^D)^{\beta_2}\left(1 - \frac{\beta_2}{\beta_1}\right)}, \quad \text{(A30)}$$

$$B_O = A_I(c_1 + c_2K)^{\beta_1 - \beta_2} - \left(\frac{r - \delta}{r\delta}\right)(c_1 + c_2K)^{1 - \beta_2} + B_I. \quad \text{(A31)}$$

Having valued the incremental options, total firm value, $V(\theta, \bar{K})$, is now:

$$V(\theta, \bar{K}) = \int_0^{\bar{K}} \Delta V(\theta, K) dK, \quad \text{(A32)}$$

and the expected instantaneous excess asset return of the firm, $E[R_A] - r$, is:

$$E[R_A] - r = \frac{\partial V(\theta, \bar{K})}{\partial \theta} \times \frac{\theta}{V(\theta, \bar{K})} \times (\mu - r), \tag{A33}$$

as shown in, for example, Cox and Rubinstein (1985) or Carlson, Fisher and Giammarino (2004).

3.8.4 Measuring Capacity Overhang

Aretz and Pope (2018) use a stochastic frontier model to estimate the difference between a firm's installed production capacity and the capacity level setting the marginal benefit of additional capacity equal to its marginal cost ("optimal capacity"). As they show, real options models often imply that installed capacity cannot fall below optimal capacity, implying that the difference between the two capacity levels is truncated from below at zero. Given that, stochastic frontier models are an appealing method to estimate the difference. Intuitively speaking, such models decompose a variable (in this case: installed capacity) into a component capturing the minimum value the variable can take on (optimal capacity) and a positively-signed residual component ("capacity overhang"). More specifically, we can write a stochastic frontier model decomposing a firm's installed capacity as:

$$\ln(\bar{K}_{i,t}) = \alpha_k + \boldsymbol{\beta}'\mathbf{X}_{i,t} + v_{i,t} + u_{i,t} = \alpha_k + \boldsymbol{\beta}'\mathbf{X}_{i,t} + \epsilon_{i,t}, \tag{A34}$$

where $\bar{K}_{i,t}$ is installed capacity, $\alpha_k + \boldsymbol{\beta}'\mathbf{X}_{i,t} + v_{i,t}$ optimal capacity, $u_{i,t}$ capacity overhang, and $\epsilon_{i,t} \equiv v_{i,t} + u_{i,t}$. Optimal capacity is modeled as a linear function of industry fixed effects, α_k, optimal capacity determinants contained in the vector $\mathbf{X}_{i,t}$, and a normally distributed residual $v_{i,t}$, with mean zero and variance σ_v^2. Conversely, capacity overhang, $u_{i,t}$, is a normally-distributed residual truncated from below at zero. The mean of the normally-distributed variable, $\boldsymbol{\gamma}'\boldsymbol{Z}_{i,t}$, is modeled as a linear function of capacity overhang determinants contained in the vector $\boldsymbol{Z}_{i,t}$, while its variance is σ_u^2. Finally, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are parameter vectors. The parameter vectors and variance parameters are estimated recursively using maximum likelihood techniques.

The first estimation window is July 1963 to December 1971, and the end of the estimation window is rolled forward on an annual basis until December 2017.

Having estimated the parameters, the estimates obtained from the window ending with year $t-1$ are combined with the values of the optimal capacity determinants and capacity overhang determinants for year t. We then define $\mu^*_{i,t} = \frac{\epsilon_{i,t}\sigma_u^2 + \boldsymbol{\gamma}'\boldsymbol{Z}_{i,t}\sigma_v^2}{\sigma_u^2+\sigma_v^2}$ and $\sigma^*_{i,t} = \sigma_u\sigma_v/\sqrt{\sigma_u^2+\sigma_v^2}$. We finally calculate firm i's capacity overhang at time t as the conditional expectation of the capacity overhang residual:

$$\hat{u}_{i,t} = E[u_{i,t}|\epsilon_{i,t}, \boldsymbol{Z}_{i,t}] = \mu^*_{i,t} + \sigma^*_{i,t}\left(\frac{n(-\mu^*_{i,t}/\sigma^*_{i,t})}{N(\mu^*_{i,t}/\sigma^*_{i,t})}\right), \tag{A35}$$

where $n(.)$ and $N(.)$ are the standard normal-density and -cumulative density, respectively.

Aretz and Pope (2018) use the log of the sum of gross property, plant, and equipment and intangible assets (intan or intanq) to measure installed capacity.[19] As optimal capacity determinants, they use:

- Sales: Log of sales over the prior four fiscal quarters (sale or saleq).
- COGS: Log of COGS over the prior four fiscal quarters (cogs or cogsq).
- SG&A: Log of SGA costs over the prior four quarters (xsga or xsgaq).
- Volatility: Log of the volatility of daily returns (ret) over the prior twelve months.
- Market beta: Sum of slope coefficients from a stock-level regression of excess stock returns (ret) on current, one-day lagged, and the sum of two-, three-, and four-day lagged excess market returns, where the regression is run using daily data over the prior twelve months (see Lewellen and Nagel (2006) for more details about the market beta estimation methodology).
- Risk-free rate: Three-month Treasury bill rate (see Kenneth French's website).

As capacity overhang determinants, they use:

[19] The terms in parentheses are the database (CRSP or Compustat) mnemonics.

- Recent sales decline: Percentage decrease in sales (sale or saleq) over the most recent four fiscal quarters; the variable is set to zero if the decrease is negative.

- More distant sales decline: Percentage decrease in sales (sale or saleq) from a stock's historical maximum of sales, measured twelve months ago, to its sales twelve months ago; the variable is set to zero if the decrease is negative.

- Loss dummy: Dummy set equal to one if a firm ran a loss (negative ni or niq) over the prior four fiscal quarters; otherwise, the variable is set to zero.

To improve timeliness, Aretz and Pope (2018) use the most recent quarterly estimate of installed capacity whenever quarterly accounting data are available. Else they use the most recent estimate from annual accounting data. With the same objective, they use four-quarter trailing sums of accounting flow variables (e.g., sales, COGS, and SG&E) whenever quarterly accounting data are available. Else they use annual accounting data. In line with Campbell et al. (2008), they assume that quarterly accounting data are released with a two-month reporting gap, while annual accounting data are released with a three-month reporting gap. They use stock market data from CRSP, accounting data from Compustat, and data on the market return and risk-free rate from Kenneth French's website.

Chapter 4

Corporate Bond Return Prediction via Machine Learning

I perform a comparative analysis of machine learning models in predicting corporate bond returns. I show that all the machine learning models have better predictive performance than the traditional simple linear model, for which the monthly out-of-sample R^2 (R^2_{oos}) is -43.00%. Among those linear machine learning models, the penalized linear models (Lasso, Ridge, and Elastic Net) perform the best, with the R^2_{oos} ranging from 0.76% to 1.09%. The generalized linear model (Group Lasso) works the second best, with an R^2_{oos} of 0.90%. The correlations among predictors are quite low, and thus the advantages of dimension reduction methods are not fulfilled. Therefore, dimension reduction models have the worst predictive power, with an R^2_{oos} of -13.95% for the partial least squares and an R^2_{oos} of -0.66% for the principal component regression. Through allowing for more flexibility in the functional form of the fitted model and interactions among predictors, non-linear machine learning models greatly improve their out-of-sample prediction performance, with an R^2_{oos} of 1.80% for gradient boosted regression trees and of 1.91% for random forests.

KEYWORDS: Machine Learning, Return Prediction, Corporate Bond Returns.

4.1 Introduction

Predicting stock returns is a long-standing research objective in the asset pricing literature.[1] Green, Hand and Zhang (2013) analyze the population of predictive signals for returns and show that there are more than 330 signals discovered over the 40-year period from 1970 to 2010. With such a large number of predictors, the predictive performance of traditional simple linear models will deteriorate seriously because of potential in-sample overfit. As machine learning techniques become more popular in finance, Gu, Kelly and Xiu (2020) point out that non-linear machine learning models perform the best in describing stock returns out-of-sample when the number of predictors is enormous. They summarize three advantages of machine learning models in predicting returns, (1) machine learning methods are highly specialized for prediction tasks, (2) machine learning models which emphasize on variable selection and dimension reduction are powerful in solving prediction problems when the number of predictors approaches the number of observations and predictors are highly correlated, (3) diversity of machine learning models sets few restrictions on the functional form of predictive regressions.

Recently, a growing literature has examined bond return predictability by bond characteristics and bond market factors.[2] Choi and Kim (2018) find evidence that equity and corporate bond markets are integrated. Besides, Chordia, Goyal, Nozawa, Subrahmanyam and Tong (2017) find that past equity returns are positively related to current bond returns, indicating that the equity market leads the bond market. Therefore, bond returns can be predicted not only by bond characteristics and bond market factors, but also by stock characteristics, stock market factors and macroeconomic factors. That is, the advantage of applying machine learning models in predicting bond returns may be more pronounced than that in the stock market, since there is an even larger collection of predictors in the bond market.

[1] Ang and Bekaert (2006) show that dividend yield predicts excess stock returns only at short horizons, but not at long horizons. Welch and Goyal (2007) argue that predicting excess stock returns using historical average return performs better than using predictive regressions. However, Campbell and Thompson (2007) find that many predictive regressions beat the historical average return in predicting returns. Cochrane (2007) concludes that stock returns are predictable using the price-dividend ratio.

[2] See, for example, Bao, Pan and Wang (2011), Jostova, Nikolova, Philipov and Stahel (2013), and Bai, Bali and Wen (2019).

This research aims to answer three questions, (1) whether machine learning models outperform the traditional simple linear model in predicting corporate bond returns out-of-sample, (2) which predictors are important in predicting corporate bond returns, (3) which predictors are important in predicting cross-sectional stock and bond returns simultaneously. In the empirical analysis, I investigate 24,945 individual bonds over the period from 2006 to 2017. I construct 11 bond-level and 6 stock-level characteristics and 4 bond market factors, and the number of baseline predictive covariates is 85 (=(11+6)×(4+1)) including interaction terms between bond-level and stock-level characteristics and bond market factors. I adopt 7 different linear and 2 non-linear machine learning models in predicting bond returns, and compare the monthly out-of-sample R^2s (R^2_{oos}s) of machine learning models to that of the simple linear model. All of the machine learning models perform better than the simple linear model, with R^2_{oos} ranging from -39.51% to 1.91%, while the R^2_{oos} of the simple linear model is -43.00%. The simple linear model with Huber loss, which controls the influence of outliers, improves the R^2_{oos} to -39.51%. The Lasso, Ridge, and elastic net models, which add penalization to the OLS loss function and reduce in-sample overfit to ensure stable out-of-sample prediction, work very well and have R^2_{oos}s of 1.00%, 1.09%, and 0.76% respectively. Compared with penalized linear models, dimension reduction models that shrink the predictor set by combining highly correlated predictors, work less well. The R^2_{oos}s of principal component regression (PCR) and partial least squares (PLS) are -13.95% and -0.66% respectively. The empirical fact that few predictors are redundant or highly correlated may cause the inferior performance of dimension reduction models. The generalized linear model that adopts a spline expansion on predictors (group lasso), works relatively well, with R^2_{oos} equal to 0.90%, indicating that univariate expansion of predictors provides incremental information and performance to the simple linear model. Most importantly, through allowing for more flexibility in the functional form of the fitted model and interactions among predictors, non-linear machine learning models greatly improve their out-of-sample prediction performance, with an R^2_{oos} of 1.80% for gradient boosted regression trees

and of 1.91% for random forests. These findings suggest that machine learning models outperform the traditional simple linear model in predicting corporate bond returns, even when a small bond sample and a relatively small set of predictors are used in the analysis. Future work could attempt to find more robust evidence to address question (1) and examine question (2) and (3). The future plan is described in Section 4.4 in details.

This paper contributes to three strands of literature. First, this paper adds to the literature that applies machine learning models in predicting returns. Gu et al. (2020) find that non-linear machine learning models perform the best in predicting stock returns out-of-sample. Bianchi, Büchner and Tamoni (2019) also find that non-linear machine learning models can be highly useful for the out-of-sample prediction of government bond excess returns. Second, this paper connects to a growing literature that predicts corporate bond returns. Lin, Wang and Wu (2011) find that liquidity risk is an important determinant of expected corporate bond returns. Bao et al. (2011) find that bond-level illiquidity can explain individual bond yields spread significantly. Jostova et al. (2013) document significant momentum effects in bond returns. Lin, Wu and Zhou (2018) indicate that corporate bond returns are highly predictable through an iterated combination model and using 27 macroeconomic, stock, and bond predictors. Bai et al. (2019) show that the downside risk is the strongest predictor of future bond returns. Finally, this paper contributes to the literature that aims to find the common predictors of stock and bond returns. Bongaerts, de Jong and Driessen (2017) find that the liquidity level and exposure to equity market liquidity risk affect expected bond returns. Chordia et al. (2017) show that profitability and asset growth are negatively related to corporate bond returns, while past equity returns are positively related to bond returns. Gebhardt, Hvidkjaer and Swaminathan (2005) discover that firms earning high (low) equity returns over the previous year earn high (low) bond returns in the following year.

The rest of the paper is organized as follows. Section 4.2 briefly discusses the machine learning models used in empirical tests. Section 4.3 calculates corporate

bond returns and predictors, and shows the predictive performance of each machine learning model. Section 4.4 describes the future research plan.

4.2 Methodology

In this section, I describe the collection of machine learning methods to be used in the following empirical analysis. For each model, I provide a detailed description of the statistical model and an objective function for estimating the model parameters. All of the models use mean squared predictions error (MSE) as the objective function. To avoid overfitting problems and improve models' out-of-sample predictive performance, I add variant regularization on the estimation objective function. Finally, I briefly discuss the computational methods for estimating each model and describe the specific implementation choices of variant algorithms in Appendix 4.5.1.

I express an asset's excess return in its most general form, an additive prediction error model:

$$r_{i,t+1} = \mathrm{E}_t(r_{i,t+1}) + \epsilon_{i,t+1}, \tag{4.1}$$

where

$$\mathrm{E}_t(r_{i,t+1}) = g^\star(\boldsymbol{z_{i,t}}). \tag{4.2}$$

Corporate bonds are indexed as $i = 1, \ldots, N$ and months by $t = 1, \ldots, T$. For simplicity of presentation, I assume a balanced panel of corporate bonds and discuss the method to deal with missing data in Section 4.3.3. $\boldsymbol{z_{i,t}}$ is a P-dimensional vector and denotes predictor variables. I assume that the conditional expected return, $g^\star(\cdot)$, is a flexible function of the predictors. My purpose is to estimate $\mathrm{E}_t(r_{i,t+1})$ as a function of the predictors which can maximize the out-of-sample explanatory power for realized $r_{i,t+1}$.

Although the form of $g^\star(\cdot)$ is flexible, there are some important restrictions on $g^\star(\cdot)$. The $g^\star(\cdot)$ function depends neither on i nor t. The restriction indicates that the form of $g^\star(\cdot)$ is the same across different bonds and over time. Therefore, the model can use information from the entire panel in estimation and brings

stability to estimates of risk premia for any individual bond. It is in contrast to standard asset pricing approaches that re-estimate a cross-sectional model each time period, or that independently estimate time-series models for each asset. Furthermore, $g^\star(\cdot)$ depends on $\boldsymbol{z}$ only through $\boldsymbol{z_{i,t}}$. It means that the prediction only uses information on the time period t and the i^{th} individual bond.

4.2.1 Sample Splitting

Before discussing specific models and regularization approaches, I introduce how I design disjoint sub-samples for estimation, validation, and testing. In particular, I divide the sample into three disjoint time periods that maintain the temporal ordering of the data.

The first or "training" subsample is used to estimate the model subject to a specific set of tuning parameter values. The second or "validation" sample is used for tuning the hyperparameters. Hyperparameters are parameters used to control the complexity of machine learning models, and different machine learning models have different hyperparameters. I first specify a suitably wide range of values for hyperparameters. [3] I then construct forecasts for data points in the validation sample using the estimated model with a specific set of values for hyperparameters from the training sample. Next, I calculate the performance evaluation measure (i.e., the out-of-sample R^2) based on forecast errors from the validation sample. Each set of values for hyperparameters has its corresponding performance evaluation value from the validation sample. I iteratively search for hyperparameters that optimize the evaluation measure. In its essence, the validation is to simulate an out-of-sample test of the model. Hyperparameters tuning intends to search for a degree of model complexity that tends to produce reliable out-of-sample performance. The validation sample is not truly out-of-sample because it is used to tune the hyperparameters, and thus an input to the estimation. Therefore, the third or "testing" subsample is truly out-of-sample and is used to evaluate a model's predictive performance.

[3] A summary of ranges of values for hyperparameters for each model is provided in Appendix 4.5.3.

Following Gu et al. (2020), I adopt a sample splitting scheme by recursively increasing the training sample size, periodically re-estimating the model once per year, and making out-of-sample predictions over the subsequent year. [4] Each time I re-estimate a model, I increase the training sample size by one year. Meanwhile, I maintain the same size of the validation sample, but roll it forward to include the most recent year's data. I choose not to cross-validate in order to maintain the temporal ordering of the data for prediction, and calculate the final out-of-sample R^2 as the average value of the R^2s from different testing samples.

4.2.2 Simple Linear Model with Huber Loss

I describe the models according to their complexity and begin with the least complex model, the simple linear predictive regression with Huber loss. The simple linear regression has been the most widely used model in return predictions over the past decades and remains one of the most important tools in the empirical asset pricing literature. [5] Although the simple linear model is expected to perform poorly with a large number of predictors, I treat it as a benchmark model to analyze distinctive features of advanced machine learning models.

Model: The simple linear model assumes that the conditional expected return, $g^\star(z_{i,t})$, can be approximated by a linear function of all the predictors,

$$g(\boldsymbol{z_{i,t}}; \boldsymbol{\theta}) = \boldsymbol{z'_{i,t}\theta}, \tag{4.3}$$

where $\boldsymbol{\theta}$ is the parameter vector. This model imposes a linear relation between predictors and expected returns and does not allow for non-linear effects or interactions among predictors.

Objective Function and Computational Algorithm: Financial returns and predictors are well-known for their non-normal distributions and heavy tails. Estimation through ordinary least squares (OLS) puts extreme weights on outliers, therefore, outliers can weaken the stability of OLS-based predictions. To solve the

[4] A detailed explanation for the sample splitting scheme is provided in Appendix 4.5.2.

[5] See Ang and Bekaert (2006), Campbell and Thompson (2007), Cochrane (2007), and Welch and Goyal (2007).

problem, the statistics literature has developed modified least squares objective function that is able to produce more stable predictions in presence of extreme observations. [6] In the machine learning literature, a common choice to mitigate the harmful effect of extreme observations is the Huber robust objective function,

$$\mathcal{L}_H(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} H(r_{i,t+1} - g(\boldsymbol{z_{i,t}};\boldsymbol{\theta}), \xi), \tag{4.4}$$

where

$$H(x,\xi) = \begin{cases} x^2, & \text{if } |x| \leq \xi; \\ 2\xi|x| - \xi^2, & \text{if } |x| > \xi. \end{cases}$$

The Huber loss, $H(\cdot)$, is a hybrid of squared loss for relatively small errors and absolute loss for relatively large errors, and the degree of combination is governed by a tuning parameter, ξ, which can be optimized adaptively from the validation sample. When ξ approaches ∞, the Huber objective function becomes a standard least squares objective function, which yields the pooled OLS estimator.

4.2.3 Penalized Linear Model (Ridge, Lasso, and Elastic Net)

When the number of predictors P approaches the number of observations T, estimates of the simple linear model become inefficient and inconsistent. [7] It is well known as the curse of dimensionality (see, e.g., Stein (1956)). As pointed out in Gu et al. (2020), the curse of dimensionality leads to overfitting noise rather than extracting signal and can be particularly problematic in forecasting asset returns where the signal-to-noise ratio is low.

Reducing the number of estimated parameters is crucial for avoiding overfitting. A popular strategy is to impose parameter sparsity through appending a penalty term to the objective function. The idea behind the strategy is that it can reduce in-sample overfitting and improve out-of-sample performance of the linear model by

[6] See Box (1953), Tukey (1960), and Huber (1964).

[7] Like stock returns, corporate bond returns strongly depend on each other cross-sectionally. The incremental information from new cross-sectional observations is limited. Therefore, I intentionally compare the number of predictors P and the sample time horizon T here.

keeping the predictive variables with the highest predictive power and discarding the least relevant ones. This strategy works when the penalization manages to reduce the model's fit of noise while preserve its fit of signals.

Objective Function and Computational Algorithm: The functional form of the penalized linear model is the same as that of the simple linear model in Equation (4.3). Penalized regressions add a penalty term in the original loss function,

$$\mathcal{L}(\boldsymbol{\theta};\cdot) = \mathcal{L}(\boldsymbol{\theta}) + \phi(\boldsymbol{\theta};\cdot), \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{i,t+1} - g(\boldsymbol{z_{i,t}};\boldsymbol{\theta}))^2 \tag{4.5}$$

where $\mathcal{L}(\boldsymbol{\theta})$ is the loss function of the simple linear model and $\phi(\boldsymbol{\theta};\cdot)$ is the penalty term. More specifically, the penalization term can take the following form (see Chapter 3 of Friedman, Hastie and Tibshirani (2001)),

$$\phi(\boldsymbol{\theta};\cdot) = \begin{cases} \frac{1}{2}\lambda\sum_{j=1}^{P}\boldsymbol{\theta_j^2} & \text{Ridge;} \quad (4.6a) \\ \lambda\sum_{j=1}^{P}|\boldsymbol{\theta_j}| & \text{Lasso;} \quad (4.6b) \\ \lambda(1-\rho)\sum_{j=1}^{P}|\boldsymbol{\theta_j}| + \frac{1}{2}\lambda\rho\sum_{j=1}^{P}\boldsymbol{\theta_j^2} & \text{Elastic Net.} \quad (4.6c) \end{cases}$$

Ridge regression shrinks the regression coefficients by imposing a penalty on their size. In equation (4.6a), $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage: the larger the value of λ, the greater the amount of shrinkage. The coefficients are shrunk toward zero. By imposing a size constraint on coefficients, the ridge regression can alleviate the problem that very large coefficients with opposite signs could arise when there exist many correlated variables in a linear regression model. Unlike ridge, the lasso method shrinks those sufficiently small coefficients to be exactly 0.[8] In this sense, the lasso imposes parsimony on the model specification and thus can be seen as a variable selection method.

[8] In the case of orthonormal inputs, the ridge estimates are just a scaled version of the least squares estimates, that is, $\hat{\theta}_j^{Ridge} = \hat{\theta}_j^{OLS}/(1+\frac{1}{2}\lambda)$, while the lasso translates the least squares estimates by a constant factor λ, truncating at zero, that is, $\hat{\theta}_j^{Lasso} = \text{sign}(\hat{\theta}_j^{OLS})(|\hat{\theta}_j^{OLS}| - \lambda)_+$.

Introduced by Hastie and Zou (2005), the elastic net is a combination of the ridge and the lasso. The elastic net selects variables like the lasso and shrinks the coefficients of the highly correlated predictors like the ridge. I adaptively optimize the tuning parameters, λ and ρ, through the validation sample. I use the accelerated proximal gradient algorithm to implement the penalized linear regressions (see Appendix A. for more details).

4.2.4 Dimension Reduction: PCR and PLS

Penalized linear methods use shrinkage and variable selection to relieve the curse of dimensionality by forcing the coefficients of redundant predictors near or exactly to zero. When predictors are highly correlated, penalized linear models can produce suboptimal predictions. In a simple situation where all the predictors are the forecast target plus an iid noise term, using a simple average of the predictors in a univariate regression performs better than the penalized linear methods do.

The essence of dimension reduction is predictors averaging, in contrast to predictors selection. Forming linear combinations of predictors helps reduce noise and better isolate signals from predictors, and meanwhile de-correlate highly dependent predictors. Two classic techniques of dimension reduction are principal components regression (PCR) and partial least squares (PLS).

I implement PCR in two steps. First, I use the principal components analysis (PCA) to form linear combinations of predictors that best preserve the covariance structure among predictors. Second, a few leading components (i.e., components with leading sample variances amongst all normalized linear combinations of predictors) are used in the predictive regression. PCA condenses regressors into components only based on the covariation among predictors. Therefore, a drawback of PCR is that it fails to incorporate our forecasting objective in the dimension reduction procedure.

Stone and Brooks (1990) and Frank and Friedman (1993) point out that, unlike PCR, which seeks directions that only maximize the variances of components, the PLS maximizes both the variances of components and their correlations with the

forecast target. That is, PLS performs dimension reduction by directly exploiting covariation of predictors with the forecast target. [9] PLS proceeds iteratively in a two-step procedure. In the first step, I estimate the univariate regression coefficient for each predictor. Next, I form a weighted linear combination of all the predictors into a single component by using weights proportional to their univariate regression coefficients, placing the highest weight on the strongest univariate predictor and the least weight on the weakest one. Through this process, PLS reduces dimension while still considers the ultimate forecasting objective. In the second step, the forecast target and all the predictors are orthogonalized with previous constructed components and the procedure of the first step is repeated on the orthogonalized dataset. This two-step procedure is iterated until the desired number of PLS components is reached.

Model: To better present the implementation of PCR and PLS, I rewrite the linear regression $r_{i,t+1} = \boldsymbol{z}_{\boldsymbol{i,t}}'\boldsymbol{\theta} + \epsilon_{i,t+1}$ as,

$$\boldsymbol{R} = \boldsymbol{Z\theta} + \boldsymbol{E}, \tag{4.7}$$

where $\boldsymbol{R}$ is the $NT \times 1$ vector of $r_{i,t+1}$, $\boldsymbol{Z}$ is the $NT \times P$ matrix of stacked predictors $\boldsymbol{z_{i,t}}$ and $\boldsymbol{E}$ is a $NT \times 1$ vector of residuals $\epsilon_{i,t+1}$.

Both PCR and PLS use the same principal to reduce the dimensionality. Both methods transform P predictors into K (smaller than P) linear combinations of predictors. Therefore, the statistical forecasting model for both methods can be written as,

$$\boldsymbol{R} = (\boldsymbol{Z\Omega_K})\boldsymbol{\theta_K} + \tilde{\boldsymbol{E}}. \tag{4.8}$$

$\boldsymbol{\Omega_K}$ is a $P \times K$ matrix with columns $\boldsymbol{w_1}$, $\boldsymbol{w_2}$, ..., $\boldsymbol{w_K}$. Vector $\boldsymbol{w_j}$ represents the set of linear combination weights used to construct the j^{th} component, and $\boldsymbol{Z\Omega_K}$ is the dimension-reduced version of the original predictors. Accordingly, the coefficient vector $\boldsymbol{\theta_K}$ has dimension $K \times 1$ instead of $P \times 1$.

Objective Function and Computational Algorithm: PCR chooses the

[9] Kelly and Pruitt (2013) and Kelly and Pruitt (2015) analyze the asymptotic characteristics of PLS and apply PLS in forecasting risk premia in financial markets.

combination weights, $\boldsymbol{w_j}$, by solving [10]

$$\boldsymbol{w_j} = \underset{\boldsymbol{w_j}}{\operatorname{argmax}} \operatorname{Var}(\boldsymbol{Z}\boldsymbol{w_j}), \text{ s.t. } \boldsymbol{w_j'}\boldsymbol{w_j} = 1, \ \operatorname{Cov}(\boldsymbol{Z}\boldsymbol{w_j}, \ \boldsymbol{Z}\boldsymbol{w_l}) = 0, \ l = 1, 2, \ldots, j-1. \tag{4.9}$$

Intuitively, PCR seeks K linear combinations of $\boldsymbol{Z}$ that retain the most possible common variation within the predictor set. A well-known solution to Problem (4.9) computes $\boldsymbol{\Omega_K}$ through eigen decomposition of $\boldsymbol{Z'Z}$, and thus the PCR algorithm is very computationally efficient.

In contrast to PCR, the PLS objective function searches K linear combinations of $\boldsymbol{Z}$ that have maximal predictive association with the forecast target. The j^{th} PLS component is constructed by using weights that solve

$$\boldsymbol{w_j} = \underset{\boldsymbol{w_j}}{\operatorname{argmax}} \operatorname{Cov}^2(\boldsymbol{R}, \boldsymbol{Z}\boldsymbol{w_j}), \text{ s.t. } \boldsymbol{w_j'}\boldsymbol{w_j} = 1, \ \operatorname{Cov}(\boldsymbol{Z}\boldsymbol{w_j}, \ \boldsymbol{Z}\boldsymbol{w_l}) = 0, \ l = 1, 2, \ldots, j-1. \tag{4.10}$$

The main difference between PCR and PLS can be observed from the objective function. PLS sacrifices the accuracy of $\boldsymbol{Z\Omega_K}$ approximating $\boldsymbol{Z}$ in exchange for more return predictability. We use the SIMPLS algorithm of de Jong (1993) to solve Problem (4.10).

For both PCR and PLS, $\boldsymbol{\theta_K}$ is estimated through OLS by regressing $\boldsymbol{R}$ on $\boldsymbol{Z\Omega_K}$ after I have a solution for $\boldsymbol{\Omega_K}$.[11] For both models, K is a hyperparameter that can be determined adaptively from the validation sample.

4.2.5 Generalized Linear Model (Group Lasso)

Linear models are very popular in practice, because they are simple, easy to implement, and can be considered as a first-order approximation to the data generating process. However, when the "true" model is not linear, using a linear form to describe the model will introduce approximation error due to model misspecification. Following Gu et al. (2020), I highlight the importance of specifying a proper functional form of the model as follows. A model's forecast error can be

[10] For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, I denote $\operatorname{Cov}(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a} - \bar{a})'(\boldsymbol{b} - \bar{b})$, where $\bar{a}$ is the average value of vector $\boldsymbol{a}$. Following the same rule, I define $\operatorname{Var}(\boldsymbol{a}) = \operatorname{Cov}(\boldsymbol{a}, \boldsymbol{a})$.

[11] Detailed algorithms to solve $\boldsymbol{\Omega_K}$ for both methods are provided in Appendix B..

decomposed as,

$$r_{i,t+1} - \hat{r}_{i,t+1} = \underbrace{g^{\star}(\boldsymbol{z_{i,t}}) - g(\boldsymbol{z_{i,t}}; \boldsymbol{\theta})}_{\text{approximation error}} + \underbrace{g(\boldsymbol{z_{i,t}}; \boldsymbol{\theta}) - g(\boldsymbol{z_{i,t}}; \hat{\boldsymbol{\theta}})}_{\text{estimation error}} + \underbrace{\epsilon_{i,t+1}}_{\text{intrinsic error}}, \quad (4.11)$$

where $g^{\star}(\boldsymbol{z_{i,t}})$ denotes the true model, $g(\boldsymbol{z_{i,t}}; \boldsymbol{\theta})$ is the specified functional form of the model, and $\hat{r}_{i,t+1}$ is the forecasted return from the fitted model $g(\boldsymbol{z_{i,t}}; \hat{\boldsymbol{\theta}})$. Intrinsic error, which is irreducible, is the unpredictable part of returns induced by new information and unknown randomness in financial markets. Estimation error is determined by the data, and can be reduced by having more observations. Approximation error can be reduced by incorporating flexible specifications of the model and improving the model's ability to approximate the true model. However, one noteworthy fact is that increasing flexibility raises the risk of overfitting and makes the model instable out-of-sample. In this and following subsections, I introduce more flexible nonparametric models of $g(\cdot)$ and their regularization methods to mitigate overfit.

Model: I first consider the generalized linear model, which is the closest to the linear models. The generalized linear model introduces nonlinear transformations of the original predictors as new predictors in a linear model. The model I study replaces the predictors with their K-term spline series expansion in the simple linear model,

$$g(\boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{p}(\cdot)) = \sum_{j=1}^{P} \boldsymbol{p}(\boldsymbol{z_j})' \boldsymbol{\theta_j}, \quad (4.12)$$

where $\boldsymbol{p}(\cdot) = (\boldsymbol{p_1}(\cdot), \boldsymbol{p_2}(\cdot), \ldots, \boldsymbol{p_K}(\cdot))'$ is a vector of basis functions, and the parameters matrix becomes a $K \times P$ matrix $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \ldots, \boldsymbol{\theta_P})$. There are many choices for spline functions, and I adopt a spline series of order two, $(1, \boldsymbol{z}, (\boldsymbol{z} - c_1)^2, (\boldsymbol{z} - c_2)^2, \ldots, (\boldsymbol{z} - c_{K-2})^2)$, where $c_1, c_2, \ldots, c_{K-2}$ are knots.

Objective Function and Computational Algorithm: Although higher order terms of predictors enter into the model, the functional form of the model is still linear. Therefore, I can use the same estimation tools of linear models. In particular, I use a least squares objective function with a penalty term as the objective function for this model (see equation (4.5)). The penalty term is used

to control the number of model parameters which grow rapidly because of series expansion. The penalization function is specialized for the spline expansion setting and is known as group lasso. The form of the penalty function is

$$\phi(\boldsymbol{\theta}; \lambda, K) = \lambda \sum_{j=1}^{P} \left(\sum_{k=1}^{K} \theta_{j,k}^2 \right)^{\frac{1}{2}}. \tag{4.13}$$

The group lasso selects either all K spline terms of a predictor or none of them. I also use the accelerated proximal gradient descent to estimate the group lasso (see Appendix A. for details), and the two tuning parameters, λ and K, are chosen adaptively from the validation sample.

4.2.6 Gradient Boosted Regression Trees and Random Forests

The generalized linear model (group lasso) described in Section 4.2.5 captures how the nonlinear transformations of individual predictors influence expected returns, however, it does not include the interaction terms among predictors. One possible way is to include multivariate functions of predictors in the generalized linear model. But the difficulty lies in that multi-way interactions will increase the parameterization combinatorially, and the generalized linear model will become computationally infeasible without knowing which interactions to include.

Different from traditional regressions, regression trees are nonparametric. The logic of regression trees makes them popular alternative machine learning approaches to incorporate multi-way predictor interactions. The fundamental logic of regression trees is to find groups of observations that behave similarly to each other. A tree "grows" in sequential steps. At each step, a new "branch" sorts the data from the previous step into different bins based on one of the predictors. This sequential branching process slices the space of predictors into rectangular partitions, and approximates the unknown function $g^\star(\cdot)$ with the average value of the outcome variable (the excess bond returns) within each partition.

Model: The prediction of a tree, $\mathcal{T}$, with K "leaves" or terminal nodes, and

depth L, can be formally written as

$$g(z_{i,t};\theta,K,L)=\sum_{k=1}^{K}\theta_k\mathbf{1}_{\{z_{i,t}\in C_k(L)\}}, \tag{4.14}$$

where $C_k(L)$ is one of the K partitions of the data. Each partition is a product of up to L indicator functions of the predictors. θ_k, the constant associated with partition k, is defined to be the simple average of outcomes within the partition.

Objective Function and Computational Algorithm: The core of regression trees lies in finding bins that best distinguish among the potential outcomes. The specific predictor on which a branch is based and the specific value where the branch is split, are chosen to minimize forecast errors. However, the large quantity of potential tree structures precludes exact optimization. To quickly converge on approximately optimal trees, I follow the algorithm of Breiman, Friedman, Stone and Olshen (1984), which is discussed in Appendix C. in detail. The basic logic of the algorithm is to optimize the forecast error at the start of and within the branch. At each branch, a predictor and the corresponding split value are chosen to maximize the discrepancy among avarage outcomes of bins without considering other branches. The loss associated with the forecast error for a branch C is called "impurity", and captures how similarly observations behave within each bin. We use the most popular l_2 loss function:

$$H(\theta,C)=\frac{1}{|C|}\sum_{z_{i,t}\in C}(r_{i,t+1}-\theta)^2, \tag{4.15}$$

where $|C|$ is the number of observations in set C, and $\theta=\frac{1}{|C|}\sum_{z_{i,t}\in C}r_{i,t+1}$. This process tries to search for the branch which locally minimizes the loss function and will stop when the number of leaves "K" or the depth of the tree "L" reaches a pre-specified value which can be chosen adaptively from the validation sample.

The outstanding advantages of a tree model are: (1) it is the same with monotonic transformations of predictors, (2) it can accommodate categorical and numerical data at the same time, (3) it can incorporate potentially severe nonlinearities and a tree of depth L can capture $(L-1)$-way interactions. However,

trees suffer greatly from overfit due to their flexibility, and thus must be heavily regularized. In line with Gu et al. (2020), I apply two "ensemble" tree regularizers that combine forecasts from many trees into a single forecast.

Boosting (Gradient Boosted Regression Trees): Although shallow trees in their own are "weak learners" with minor predictive power, however, many weak learners may, as an ensemble, comprise a single "strong learner" with greater stability than a single complex tree. Therefore, the "boosting" regularization method tries to improve the predictive power through recursively combining forecasts from many over-simplified trees.

The specific boosting procedure I use is called gradient boosted regression trees (GBRT). It begins with fitting a shallow tree, for example, a tree with depth $L = 1$. Next, a second simple tree with the same depth L is used to fit the prediction residuals from the first tree. At each new step b, a shallow tree is fitted to the prediction residuals from the tree in step $b - 1$, and its forecast is added to the total prediction with a shrinkage weight of v ($v \in (0, 1)$ is applied to prevent from overfitting the prediction residuals.). This procedure is iterated until there are a total of B trees in the ensemble. The final result is an additive model of shallow trees with three tuning parameters (L, v, B) which are adaptively chosen from the validation sample. The detailed algorithm to implement GBRT is displayed in Appendix C..

Random Forest: Similar to boosting, a random forest is also a kind of ensemble methods that combine forecasts from many different trees. It is a variation of a more general procedure known as bootstrap aggregation, or "bagging" (Breiman (2001)). The basic tree bagging procedure randomly draws B different bootstrap samples from the data, fits a separate regression tree to each sample, and averages their forecasts. It is likely for trees of individual bootstrap samples to be overfit, thus making their individual predictions inefficient and vary a lot. Averaging over multiple predictions can smooth the prediction, and stabilize the trees' predictive performance.

There is a drawback of bagging. If, for example, bond credit rating is the

dominant predictor in the data, then most of the bagged trees will have low-branch splits on bond credit rating, which will result in substantial correlation among their ultimate predictions. The random forest method considers a randomly drawn subset of predictors for splitting at each branch. Through that way, early branches for some trees will split on predictors other than bond credit rating in the example. It further improves the variance reduction relative to standard bagging by reducing the correlation among predictions. Depth L of trees and number of bootstrap samples B are tuning parameters optimized through validation. Details of the algorithm to implement random forests are shown in Appendix C..

4.2.7 Performance Evaluation

To evaluate the predictive performance of each machine learning model, I calculate the out-of-sample R^2 as

$$R^2_{oos} = 1 - \frac{\sum_{(i,t)\in\tau_3}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t)\in\tau_3}(r_{i,t+1} - \bar{r})^2}, \tag{4.16}$$

where τ_3 indicates data which is only available in the testing sample and never enters into model estimation or hyperparameter tuning. $\bar{r}$ is the historical mean of excess bond returns and obtained by averaging excess returns in the training sample. The R^2_{oos} pools prediction errors across bonds and over time into a tremendous panel-level assessment of each model.

As the following extortion, one can use the Diebold and Mariano (2002) test to make pairwise comparisons and examine the significance of predictive performance improvement among models. The data structure decides that the time-series dependence in returns is weak, while the cross-sectional dependence is strong. It is unlikely that the conditions of weak error dependence underlying the Diebold-Mariano test apply to the bond-level analysis. Therefore, one can adapt Diebold-Mariano test to the data by comparing the cross-sectional average of prediction errors from each model, instead of doing it from the individual return perspective. To be specific, the test statistic used to test the predictive performance of model (1) against model (2) is defined as,

$$DM_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}_{12}}, \quad d_{12,t+1} = \frac{1}{n_{3,t+1}} \sum_{i=1}^{n_{3,t+1}} \left((\hat{e}^{(1)}_{i,t+1})^2 - (\hat{e}^{(2)}_{i,t+1})^2 \right), \tag{4.17}$$

where $\hat{e}^{(1)}_{i,t+1}$ and $\hat{e}^{(2)}_{i,t+1}$ are prediction errors for bond i at time t of model (1) and model (2), and $n_{3,t+1}$ is the number of bonds in the testing sample at year $t+1$. $\bar{d}_{12}$ and $\hat{\sigma}_{\bar{d}_12}$ denote the mean and Newey-West standard error of $d_{12,t}$ over the testing sample. This modified Diebold-Mariano test statistic now bases on a single time-series of $d_{12,t+1}$ with small autocorrelation. Therefore, it is likely to satisfy the conditions for asymptotic normality and provide proper p-values for the model comparison tests.

4.3 Empirical Study

4.3.1 Bond Data

I collect bond data, including intraday transaction prices, transaction dates, trading volumes, sale conditions, when-issued indicators, locked-in indicators, trade status, and commission indicators from the enhanced version of the Trade Reporting and Compliance Engine (TRACE). The sample period spands from July 2002 to June 2017. In contrast to the Lehman Brothers Fixed Income Database, Datastream, and Bloomberg, which are quote-based databases, TRACE is a trade-based database, offering higher market transparency (see Bessembinder, Maxwell and Venkataraman (2006)) and covering about 99% of all public bond-market transactions since February 2005 (see Bao et al. (2011)). I collect bond characteristics from the Mergent Fixed Income Securities Database (FISD), including offering-amount and -date, maturity date, coupon-rate, -type, and -payout frequency, bond-type, -rating, and -option features, and issuer information. The monthly risk-free rate (one-month Treasury Bill rate) is obtained from Ibbotson and Associates.

I impose the following standard filters in the literature on the bond data. 1) I remove bonds not traded or listed in U.S. public markets. 2) I exclude bonds that

are structured notes, are mortgage-, asset-, or agency-backed, or are equity-linked. 3) I remove convertible bonds. 4) I keep only fixed and zero coupon bonds. 5) I remove bonds with less than one year to maturity. 6) I eliminate bond transactions that are labeled as when-issued or locked-in or have special sales conditions. 7) I remove canceled, corrected, or commission trades.

4.3.2 Calculating Corporate Bond Returns

Following Bessembinder, Kahle, Maxwell and Xu (2009), Bao et al. (2011), and Jostova et al. (2013), I calculate the net return of corporate bond i over month t, $r_{i,t}$, using:

$$r_{i,t} = \frac{B_{i,t} + AI_{i,t} + C_{i,t}}{B_{i,t-1} + AI_{i,t-1}} - 1, \tag{4.18}$$

where B is the bond price, AI the accrued interest, and C the coupon payment. The price B is calculated as follows. Consistent with Bessembinder et al. (2009), I calculate a bond's daily price as the trading-volume-weighted average of intraday transaction prices over that day to minimize microstructure effects caused by bid-ask spreads. In line with Bai et al. (2019), I next calculate three types of bond returns, namely: (i) the return from the end of month $t-1$ to the end of month t; (ii) the return from the start of month t to the end of month t; and (iii) the return from the start of month t to the start of month $t+1$, where I define the start (end) of a month as the first (last) five trading days within that month. If I have more than one non-missing daily bond price within either the start- or end-of-month window, I choose the daily price closest to the first/last trading day of a month in my calculations. Finally, if it is able to calculate more than one type of returns, the selection order is type (i), type (ii), and type (iii).

To calculate the accrued interest AI, I first compute the daily coupon rate. The daily coupon rate is the coupon rate divided by 360 if a bond's day-count basis is "30/360" or "ACT/360," and it is the coupon rate divided by the actual number of calendar days per year if the day-count basis is "ACT/ACT." I next count the calendar days between the current month-end t and the previous coupon

payment date, assuming that a month has 30 calendar days if the day-count basis is "30/360" and the actual number of days per month when it is "ACT/360" or "ACT/ACT." Also, I use the date of the first coupon payment and the coupon payment frequency to infer on which days the coupons are paid. I finally calculate the accrued interest AI as the daily coupon rate multiplied by the number of days between the current month-end t and the previous coupon payment date.

4.3.3 Calculating Predictors

I construct a collection of bond-level predictive characteristics and predictive bond market factors based on the literature of cross-sectional bond returns. There are 11 bond-level characteristics and 4 bond market factors in total. Besides, I build 6 stock-level characteristics according to predictions of cross-sectional stock returns literature. All of the predictors are updated monthly. Brief descriptions and references of all the predictors are provided in Table 4.1. [12] Appendix 4.5.4 provides detailed construction processes of the bond-level characteristics and bond market factors.

Table 4.2 shows the descriptive statistics of excess corporate bond returns and bond-level characteristics. The final sample includes 24,945 unique bonds and a total of 659,299 bond-month return observations during the sample period from July 2006 to June 2017. There are on average 4,995 bonds per month in the whole sample. The average monthly excess bond return is 0.55%.

[12] To construct predictive bond-level characteristics and bond market factors, I lose the first 4 years' data. The sample used in prediction goes from July 2006 to June 2017. To avoid the forward-looking bias, the monthly bond-level and stock-level characteristics are delayed by 1 month. Furthermore, I replace missing characteristics with the cross-sectional median of each month for each bond.

Table 4.1: Details of the Predictors

This table lists the bond-level characteristics, bond market factors, and stock-level characteristics I use in predicting bond returns. All of the predictors are updated monthly. Except for def (default spread) and term (term spread), the data used to construct bond-level characteristics and bond market factors are from TRACE and FISD. Default spread and term spread are constructed with the data from DataStream. Stock-level characteristics are built with the data from CRSP.

No.	Symbol	Description	Paper's Author(s)	Year, Journal
		Bond-level Characteristics		
1	dsrisk	Downside risk	Bai, Bali & Wen	2019, JFE
2	illiq	Illiquidity	Bao, Pan & Wang	2011, JF
3	ltr	Long-term reversal	Bali, Subrahmanyam & Wen	2019, WP
4	mat	Years-to-maturity	Bai, Bali & Wen	2019, JFE
5	mom	Momentum	Jostova, Nikolova, Philipov & Stahel	2013, RFS
6	rating	Credit rating	Jostova, Nikolova, Philipov & Stahel	2013, RFS
7	size	Amount outstanding	Bai, Bali & Wen	2019, JFE
8	skew	Return skewness	Bai, Bali & Wen	2016, WP
9	str	Short-term reversal	Bali, Subrahmanyam & Wen	2019, WP
10	vol	Return volatility	Bai, Bali & Wen	2016, WP

Table 4.1 continued

11	uncbeta	Economic uncertainty beta	Bali, Subrahmanyam & Wen	2019, WP
		Bond Market Factors		
1	bmkt	Bond market excess return	Bai, Bali & Wen	2019, JFE
2	def	Default spread	Fama & French	1993, JFE
3	liq	Market liquidity risk	Lin, Wang & Wu	2011, JFE
4	term	Term spread	Fama & French	1993, JFE
		Stock-level Characteristics		
1	beta	Beta loading on the stock market excess return	Fama & MacBeth	1973, JPE
2	betasq	Beta squared	Fama & MacBeth	1973, JPE
3	ill	Stock illiquidity ratio	Amihud	2002, JFM
4	maxret	Maximum daily return	Bali, Cakici & Whitelaw	2011, JFE
5	mom1m	1-month momentum	Jegadeesh	1990, JF
6	mom12m	12-month momentum	Jegadeesh	1990, JF

Table 4.2: Descriptive Statistics

This table shows descriptive statistics of excess bond returns, and bond-level and stock-level characteristics. ret, dsrisk, illiq, ltr, mat, mom, rating, size, skew, str, vol, uncbeta, beta, betasq, ill, maxret, mom1m, mom12m are excess bond return, downside risk, illiquidity, long-term reversal, years-to-maturity, momentum, credit rating, amount outstanding, return skewness, short-term reversal, return volatility, economic uncertainty beta, beta loading on the excess stock market return, beta squared, stock illiquidity ratio, maximum daily return, 1-month momentum and 12-month momentum respectively. Details of construction processes of those bond-level characteristics are provided in Appendix 4.5.4. The sample period is from July 2006 to June 2017.

			Standard	Percentiles						
	Obs	Mean	Deviation	1	5	25	50	75	95	99
Excess Bond Return										
ret (%)	659,299	0.55	4.63	-11.70	-3.83	-0.49	0.40	1.58	5.15	13.33
Bond-level Characteristics										
dsrisk (%)	659,299	4.20	3.61	0.65	1.15	2.58	2.99	5.40	8.80	20.92
illiq	659,299	1.44	3.95	-1.12	-0.04	0.10	0.39	1.21	6.24	20.15
ltr	659,299	0.19	0.18	-0.08	-0.00	0.06	0.16	0.25	0.47	0.58
mat	659,299	14.23	10.13	2.99	4.93	7.70	10.03	20.03	30.07	33.74
mom	659,299	0.03	0.12	-0.19	-0.06	0.00	0.03	0.05	0.14	0.32
rating	659,299	8.47	3.85	1.00	3.00	6.00	8.00	10.00	16.00	22.00

Table 4.2 continued

size	659,299	12.36	1.72	7.76	8.87	12.07	12.90	13.46	14.38	14.91
skew	659,299	0.09	0.76	-2.19	-0.94	-0.10	0.07	0.29	1.34	2.70
str (%)	659,299	0.61	4.04	-9.82	-3.43	-0.32	0.47	1.51	4.76	11.90
vol (%)	659,299	0.15	0.31	0.01	0.02	0.04	0.07	0.16	0.43	1.63
uncbeta	659,299	-0.01	0.37	-1.19	-0.44	-0.09	0.01	0.07	0.42	1.00
Stock-level Characteristics										
beta	659,299	0.02	0.56	-0.21	-0.04	0.00	0.01	0.02	0.07	0.31
betasq	659,299	0.3121	13.1874	0.0000	0.0000	0.0001	0.0002	0.0007	0.0126	0.7584
ill	659,299	-0.0069	4.3781	0.0000	0.0000	0.0000	0.0001	0.0002	0.0023	0.0308
maxret (%)	659,299	4.23	5.21	0.86	1.21	2.01	2.95	4.59	10.97	25.00
mom1m (%)	659,299	0.80	11.20	-30.82	-14.21	-3.76	0.92	5.36	14.78	30.13
mom12m (%)	659,299	0.73	11.11	-30.47	-14.43	-3.90	0.88	5.35	14.67	29.63

Table 4.3: Correlations among Bond-Level and Stock-Level Characteristics

This table shows pairwise correlations among bond-level and stock-level characteristics. dsrisk, illiq, ltr, mat, mom, rating, size, skew, str, vol, uncbeta, beta, betasq, ill, maxret, mom1m, mom12m are excess bond return, downside risk, illiquidity, long-term reversal, years-to-maturity, momentum, credit rating, amount outstanding, return skewness, short-term reversal, return volatility, economic uncertainty beta, beta loading on the excess stock market return, beta squared, stock illiquidity ratio, maximum daily return, 1-month momentum and 12-month momentum respectively. Details of construction processes of those bond-level characteristics are provided in Appendix 4.5.4. The sample period is from July 2006 to June 2017.

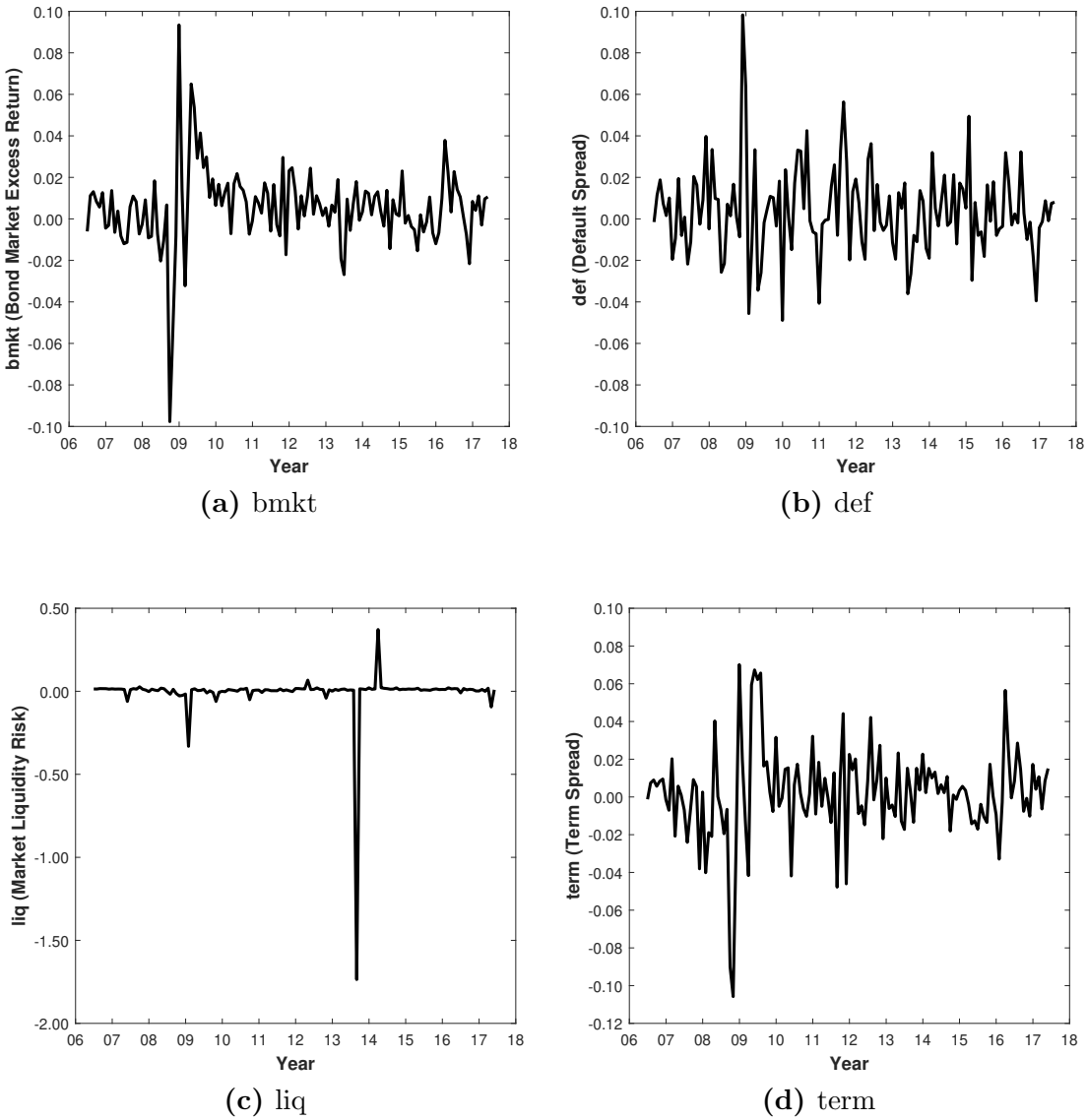
	drisk	illiq	ltr	mat	mom	rating	size	skew	str	vol	uncbeta	beta	betasq	ill	maxret	mom1m	mom12m
drisk	1.00																
illiq	0.20	1.00															
ltr	-0.20	-0.07	1.00														
mat	0.14	0.13	0.04	1.00													
mom	0.19	-0.01	-0.10	0.03	1.00												
rating	0.18	0.08	0.03	-0.12	0.02	1.00											
size	-0.07	-0.12	0.05	-0.07	-0.02	0.07	1.00										
skew	0.03	-0.04	0.07	0.03	0.13	-0.02	0.03	1.00									
str	0.09	0.01	-0.03	0.02	0.25	0.00	-0.02	0.12	1.00								
vol	**0.67**	0.13	0.06	0.11	-0.25	0.17	-0.05	0.13	0.13	1.00							
uncbeta	-0.00	-0.00	-0.08	0.04	-0.05	-0.10	-0.01	-0.05	-0.01	-0.09	1.00						

Table 4.3 continued

beta	-0.01	0.00	-0.00	-0.00	-0.01	-0.01	-0.02	-0.00	0.00	-0.01	0.02	1.00					
betasq	0.00	-0.00	-0.02	-0.00	0.01	0.01	0.01	0.00	-0.01	0.01	-0.01	-0.30	1.00				
ill	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	1.00			
maxret	0.24	0.19	-0.10	-0.05	-0.08	0.21	-0.07	-0.09	0.01	0.15	-0.01	0.01	0.00	-0.00	1.00		
mom1m	0.02	-0.04	0.02	0.00	0.12	-0.03	0.02	0.07	0.29	0.04	-0.00	0.05	-0.02	-0.00	0.08	1.00	
mom12m	-0.02	-0.03	0.05	0.00	-0.08	-0.01	0.02	0.02	-0.03	-0.00	0.00	0.00	-0.00	0.00	-0.07	0.02	1.00

Table 4.3 displays pairwise correlations among bond-level and stock-level characteristics. Except for the high correlation between return volatility (vol) and downside risk (dsrisk) which is 0.67, the correlations among other bond-level and stock-level characteristics are low.

Figure 4.1: Bond Market Factors



(a) bmkt **(b)** def

(c) liq **(d)** term

This figure shows 4 bond market factors, with bmkt (bond market excess return) in Panel (a), def (default spread) in Panel (b), liq (market liquidity risk) in Panel (c), and term (term spread) in Panel (d).

Figure 4.1 shows the time series of 4 bond market factors. The bond market excess return and term spread reach their lowest levels during the 2008-2009 financial crisis. By contrast, the default spread reaches its highest level since

investors become more risk-averse. Besides, the market liquidity risk increases in the subprime financial crisis period, indicating low liquidity in the market. Moreover, there is another large decline in market liquidity during 2013-2014 debt ceiling crisis, which limits the amount of national debt and reduces liquidity in the bond market.

4.3.4 Performance of Machine Learning Models

The baseline set of predictive covariates can be defined as,

$$\boldsymbol{z_{i,t}} = \boldsymbol{c_{i,t}} \otimes \boldsymbol{x_t}, \tag{4.19}$$

where $\boldsymbol{c_{i,t}}$ is a $P_c \times 1$ ($P_c = 17 = 11 + 6$) vector of characteristics for each bond i, and $\boldsymbol{x_t}$ is a $P_x \times 1$ ($P_x = 5 = 1 + 4$) vector of bond market factors, which include a constant and are common to all bonds. Therefore, $\boldsymbol{z_{i,t}}$ is a $P \times 1$ ($P = P_c \times P_x = 85$) vector of covariates, which are used in predicting individual bond returns and include interaction terms of bond-level and stock-level characteristics and bond market factors.

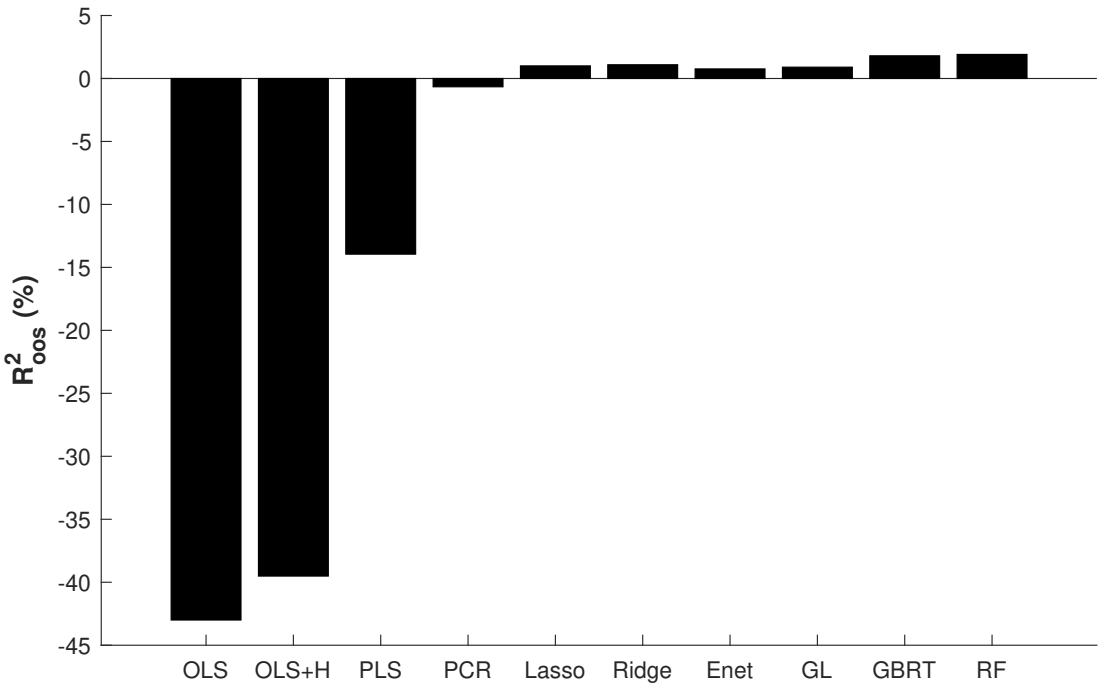
Table 4.4 compares the out-of-sample predictive R^2_{oos}s of different machine learning models. R^2_{oos}s for the entire pooled sample of 10 different models are reported, including OLS, OLS with Huber loss (OLS+H), PLS, PCR, Lasso, elastic net (Enet), Group Lasso (GL), gradient boosted regressions (GBRT), and random forests (RF). The OLS model generates an R^2_{oos} of -43.00%. Using Huber loss to alleviate the outlier problem slightly improves the R^2_{oos} to -39.51%. The bad performance of the simple linear model is unsurprising, because the lack of regularization greatly improves the possibility of in-sample overfit and reduces the model's out-of-sample prediction power. Moreover, adding penalization to the OLS loss function substantially improves the performance of the simple linear model, which can be seen by the positive R^2_{oos}s of the Lasso (1.00%), Ridge (1.09%), and Enet (0.76%) models. However, regularization through dimension reduction does not work very well, though the performances of PLS ($R^2_{oos} = -13.95\%$) and PCR ($R^2_{oos} = -0.66\%$) are much better than that of the OLS model. It is because

correlations among bond characteristics are low and few covariates are redundant, and thus the advantages of dimension reduction methods are not fulfilled. PCR keeps a few high-variance directions and discards the rest. Meanwhile, it has been shown that PLS also tends to shrink the low-variance directions, but can actually inflate some of the higher variance directions. This property makes PLS a little unstable and causes the inferior performance of PLS compared with PCR. The generalized linear model with group lasso penalty performs relatively well with an R^2_{oos} of 0.90%. It suggests that univariate expansion of covariates does provide incremental information and performance to the simple linear model. Most importantly, the non-linear machine learning models, GBRT and RF, perform the best among all the machine learning models with R^2_{oos}s of 1.80% and 1.91% respectively. Allowing for more flexibility of the functional form and interactions among predictors greatly improves the predictive power of non-linear models with respect to linear models.

Table 4.4: Monthly Out-of-sample Bond-level Prediction Performance

This table reports monthly R^2_{oos}s for the entire panel of bonds using OLS, OLS with Huber loss (OLS+H), PLS, PCR, Lasso, Ridge, elastic net (Enet), Group Lasso (GL), gradient boosted regression trees (GBRT), and random forests (RF). R^2_{oos}s are displayed in percentage. The lower figure provides a visual comparison of R^2_{oos} measures displayed in the table.

OLS	OLS+H	PLS	PCR	Lasso	Ridge	Enet	GL	GBRT	RF
-43.00	-39.51	-13.95	-0.66	1.00	1.09	0.76	0.90	1.80	1.91



4.4 Future Plan

This paper aims to answer three main questions: (1) whether machine learning models outperform the traditional simple linear model in predicting corporate bond returns out-of-sample, (2) which predictors are important in predicting corporate bond returns, (3) which predictors are important in predicting the cross-sectional stock and bond returns simultaneously.

My current empirical results indicate that machine learning models indeed perform much better than the simple linear model, even when a small bond sample and a relatively small set of bond and stock market predictors are used in the

analysis.[13] To answer question (1) with stronger evidence, I plan to improve the empirical results from three aspects. First, I will expand the current bond sample to a large one which ranges from January 1973 to March 2019 and includes bond transaction data from TRACE, National Association of Insurance Commission (NAIC), Lehman Brothers fixed income database, DataStream, and Bloomberg. Second, I will add more stock market predictors in predicting bond returns. In total, there are 190 predictors, including 94 stock-level characteristics, 11 bond-level characteristics, 74 firm industry dummies, and 11 market and macroeconomic factors. Therefore, the total number of covariates will be $(11+94)\times(11+1)+74 = 1334$, when considering the interactions between macroeconomic variables and bond- and stock-level characteristics. Third, I will also apply more advanced non-linear machine learning models, i.e., neural networks, in predicting bond returns.

To answer question (2), I plan to identify predictors that have an important impact on the cross-sectional bond returns while controlling for other predictors. To accomplish that goal, I rank influential predictors according to their importance which is measured by two different methods. The first method calculates the reduction in R^2_{oos} when setting the value of examined predictor to 0, and holding the other covariates parameter estimates untouched. The second method measures the sensitivity of the fit of model to changes in the examined predictor, and calculates the sum of squared partial derivatives of the model with respect to the examined variable.[14]

The plan for answering question (3) is to apply the same set of predictors used in predicting bond returns to predict stock returns, and identify predictors which are influential in predicting both bond and stock returns.

[13] The bond pricing literature documents that corporate bond returns can be predicted not only by bond market predictors, but also by stock market predictors. (See, e.g., Choi and Kim (2018).)

[14] The first and second methods have been used in Kelly, Pruitt and Su (2019) and Dimopoulos, Bourret and Lek (1995) respectively.

References

Ang, A., Bekaert, G., 2006. Stock return predictability: Is it there? Review of Financial Studies 20, 651–707.

Bai, J., Bali, T.G., Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. Journal of Financial Economics 131, 619–642.

Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. Journal of Finance 66, 911–946.

Bessembinder, H., Kahle, K.M., Maxwell, W.F., Xu, D., 2009. Measuring abnormal bond performance. Review of Financial Studies 22, 4219–4258.

Bessembinder, H., Maxwell, W., Venkataraman, K., 2006. Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. Journal of Financial Economics 82, 251–288.

Bianchi, D., Büchner, M., Tamoni, A., 2019. Bond risk premia with machine learning. Working Paper.

Bongaerts, D., de Jong, F., Driessen, J., 2017. An asset pricing approach to liquidity effects in corporate bond markets. Review of Financial Studies 30, 1229–1269.

Box, G.E.P., 1953. Non-normality and tests on variances. Biometrika 40, 318–335.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. CRC press.

Campbell, J.Y., Thompson, S.B., 2007. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Choi, J., Kim, Y., 2018. Anomalies and market (dis) integration. Journal of Monetary Economics 100, 16–34.

Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A., Tong, Q., 2017. Are capital market anomalies common to equity and corporate bond markets? an empirical investigation. Journal of Financial and Quantitative Analysis 52, 1301–1342.

Cochrane, J.H., 2007. The dog that did not bark: A defense of return predictability. Review of Financial Studies 21, 1533–1575.

Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Dimopoulos, Y., Bourret, P., Lek, S., 1995. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Processing Letters 2, 1–4.

Frank, I.E., Friedman, J.H., 1993. A statistical view of some chemometrics regression tools. Technometrics 35, 109–135.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer Series in Statistics, Springer New York Inc., New York, NY, USA.

Gebhardt, W.R., Hvidkjaer, S., Swaminathan, B., 2005. Stock and bond market interaction: Does momentum spill over? Journal of Financial Economics 75, 651–690.

Green, J., Hand, J.R., Zhang, X.F., 2013. The supraview of return predictive signals. Review of Accounting Studies 18, 692–730.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Hastie, T., Zou, H., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320.

Huber, P.J., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73–101.

de Jong, S., 1993. Simpls: An alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems 18, 251–263.

Jostova, G., Nikolova, S., Philipov, A., Stahel, C., 2013. Momentum in corporate bond returns. Review of Financial Studies 20, 1649–1693.

Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. Journal of Finance 68, 1721–1756.

Kelly, B., Pruitt, S., 2015. The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics 186, 294–316.

Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Lin, H., Wang, J., Wu, C., 2011. Liquidity risk and expected corporate bond returns. Journal of Financial Economics 99, 628–650.

Lin, H., Wu, C., Zhou, G., 2018. Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. Management Science 64, 4218–4238.

Nesterov, Y.E., 1983. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. Soviet Mathematics Doklady 27, 372–376.

Parikh, N., Boyd, S., 2014. Proximal algorithms. Foundations and Trends in Optimization 1, 127–239.

Polson, N.G., Scott, J.G., Willard, B.T., 2015. Proximal algorithms in statistics and machine learning. Statistical Science 30, 559–581.

Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press, Berkeley, Calif.. pp. 197–206.

Stone, M., Brooks, R.J., 1990. Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. Journal of the Royal Statistical Society. Series B (Methodological) 52, 237–269.

Tukey, J.W., 1960. A survey of sampling from contaminated distributions, in: Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., Mann, H.B. (Eds.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press, Stanford, CA, USA, pp. 448–485.

Welch, I., Goyal, A., 2007. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

4.5 Appendix

4.5.1 Algorithms in Details

A. Ridge, Lasso, Elastic Net, and Group Lasso

In this section, I present the procedure of using the accelerated proximal algorithm to estimate coefficients of the Lasso, Ridge, elastic net and Group Lasso models. [15] Their regularized objective functions can be written as

$$\mathcal{L}(\boldsymbol{\theta};\cdot) = \mathcal{L}(\boldsymbol{\theta}) + \phi(\boldsymbol{\theta};\cdot), \tag{A1}$$

and

$$\phi(\boldsymbol{\theta};\cdot) = \begin{cases} \frac{1}{2}\lambda \sum_{j=1}^{P} \boldsymbol{\theta_j^2} & \text{Ridge;} \\ \lambda \sum_{j=1}^{P} |\boldsymbol{\theta_j}| & \text{Lasso;} \\ \lambda(1-\rho) \sum_{j=1}^{P} |\boldsymbol{\theta_j}| + \frac{1}{2}\lambda\rho \sum_{j=1}^{P} \boldsymbol{\theta_j^2} & \text{Elastic Net;} \\ \lambda \sum_{j=1}^{P} ||\boldsymbol{\theta_j}|| & \text{Group Lasso.} \end{cases}.$$

$\boldsymbol{\theta^\star}$ minimizes the objective function (A1) if and only if

$$\boldsymbol{\theta^\star} = \mathbf{prox}_{\gamma\phi}(\boldsymbol{\theta^\star} - \gamma\nabla\mathcal{L}(\boldsymbol{\theta^\star})), \tag{A2}$$

where γ is the learning rate. The proximal operators are

$$\mathbf{prox}_{\gamma\phi}(\boldsymbol{\theta}) = \begin{cases} \frac{\boldsymbol{\theta}}{1+\lambda\gamma} & \text{Ridge;} \\ S(\boldsymbol{\theta}, \lambda\gamma) & \text{Lasso;} \\ \frac{1}{1+\lambda\gamma\rho} S(\boldsymbol{\theta}, (1-\rho)\lambda\gamma) & \text{Elastic Net;} \\ (\tilde{S}(\boldsymbol{\theta_1}, \lambda\gamma)^\intercal, \tilde{S}(\boldsymbol{\theta_2}, \lambda\gamma)^\intercal, \ldots, \tilde{S}(\boldsymbol{\theta_P}, \lambda\gamma)^\intercal)^\intercal & \text{Group Lasso.} \end{cases},$$

[15] Parikh and Boyd (2014) and Polson, Scott and Willard (2015) provide more detailed descriptions of the accelerated proximal algorithm.

where $S(x,\mu)$ and $\tilde{S}(x,\mu)$ are vector-valued functions, whose i^{th} component is

$$(S(x,\mu))_i = \begin{cases} x_i - \mu, & \text{if } x_i > 0 \text{ and } \mu < |x_i|; \\ x_i + \mu, & \text{if } x_i < 0 \text{ and } \mu < |x_i|; \\ 0, & \text{if } \mu \geq |x_i|. \end{cases},$$

and

$$(\tilde{S}(x,\mu))_i = \begin{cases} x_i - \mu\frac{x_i}{||x_i||}, & \text{if } ||x_i|| > \mu; \\ 0, & \text{if } ||x_i|| \leq \mu. \end{cases}.$$

The detailed algorithm is described in Algorithm 1.

B. PCR and PLS

I obtain the principal components of the variables in $\boldsymbol{Z}$ through the eigen decomposition of $\boldsymbol{Z'Z}$,

$$\boldsymbol{Z'Z} = \boldsymbol{WD^2W'}. \tag{A3}$$

$\boldsymbol{D}$ is a $P \times P$ diagonal matrix, with diagonal entries $d_1 \geq d_2 \geq \ldots \geq d_P \geq 0$ called the singular values of $\boldsymbol{Z}$. The eigenvectors $\boldsymbol{w_j}$ (columns of $\boldsymbol{W}$) are the principal components directions of $\boldsymbol{Z}$. The first principal component direction $\boldsymbol{w_1}$ has the property that $\boldsymbol{Zw_1}$ has the largest sample variance amongst all normalized linear combinations of the columns of $\boldsymbol{Z}$.

Principal component regression forms the derived input columns $\boldsymbol{x_j} = \boldsymbol{Zw_j}$, and then regress $\boldsymbol{R}$ on $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_K}$ for some $K \leq P$. The number of K is optimized through the validation sample.

I use the SIMPLS algorithm developed by de Jong (1993) to estimate the coefficients of partial least squares regression. The detailed algorithm is displayed in Algorithm 2.

Algorithm 1 Accelerated Proximal Gradient Method

Standardize each predictor vector $\boldsymbol{Z_j}$ to have mean 0 and variance 1. Center the excess return vector $\boldsymbol{R}$.
Initialize $\boldsymbol{\theta_0} = 0$, $k = 0$, γ.
while $\boldsymbol{\theta_k}$ not converged **do**
  $\bar{\boldsymbol{\theta}}_{\boldsymbol{k+1}} \leftarrow \boldsymbol{\theta_k} - \gamma \nabla \mathcal{L}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta_k}}$
  $\tilde{\boldsymbol{\theta}}_{\boldsymbol{k+1}} \leftarrow \mathbf{prox}_{\gamma\phi}(\bar{\boldsymbol{\theta}}_{\boldsymbol{k+1}})$ {The first two iteration steps implement equation (A2).}
  $\boldsymbol{\theta}_{\boldsymbol{k+1}} \leftarrow \tilde{\boldsymbol{\theta}}_{\boldsymbol{k+1}} + \frac{k}{k+3}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{k+1}} - \tilde{\boldsymbol{\theta}}_{\boldsymbol{k}})$ {This step is a Nesterov momentum (Nesterov (1983)) adjustment that accelerates convergence.}
  $k \leftarrow k + 1$
end while
return $\boldsymbol{\theta_k}$

Algorithm 2 SIMPLS

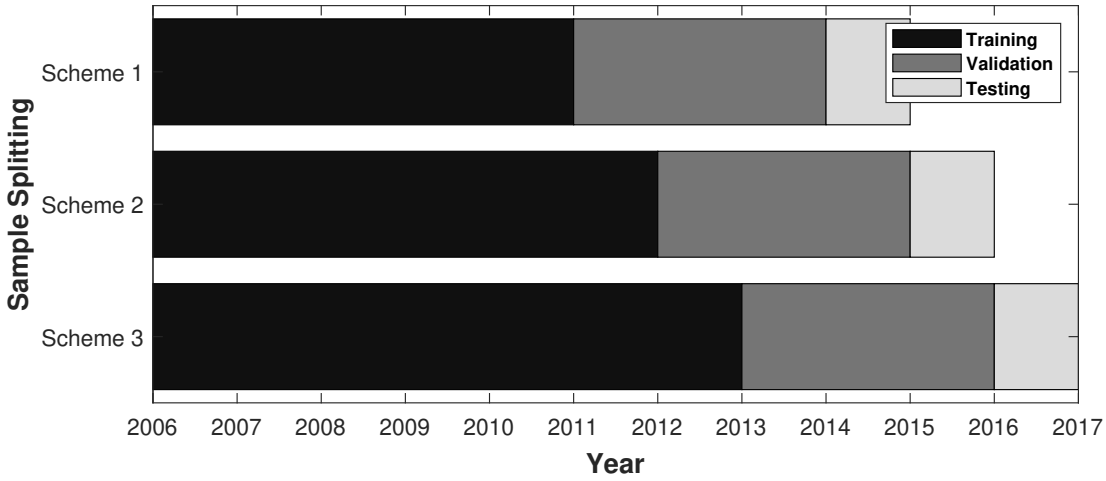
Input: $NT \times P$ matrix $\boldsymbol{X}$ {the standardized version of the predictor matrix $\boldsymbol{Z}$}
  $NT \times 1$ vector $\boldsymbol{Y}$ {the excess return vector $\boldsymbol{R}$}
  the number of factors A {P in the main text}
Output: $\boldsymbol{B}$ {regression coefficients}
  $\boldsymbol{Y_0} = \boldsymbol{Y} - MEAN(\boldsymbol{Y})$ {center $\boldsymbol{Y}$}
  $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{Y_0}$
  for $a = 1, \ldots, A$ **do**
    $\boldsymbol{q}$ = dominant eigenvector of $\boldsymbol{S}'\boldsymbol{S}$
    $\boldsymbol{r} = \boldsymbol{S}\boldsymbol{q}$
    $\boldsymbol{t} = \boldsymbol{X}\boldsymbol{r}$
    $\boldsymbol{t} = \boldsymbol{t} - MEAN(\boldsymbol{t})$
    $normt = SQRT(\boldsymbol{t}'\boldsymbol{t})$
    $\boldsymbol{t} = \boldsymbol{t}/normt$
    $\boldsymbol{r} = \boldsymbol{r}/normt$
    $\boldsymbol{p} = \boldsymbol{X}'\boldsymbol{t}$
    $\boldsymbol{q} = \boldsymbol{Y_0}'\boldsymbol{t}$
    $\boldsymbol{u} = \boldsymbol{Y_0}\boldsymbol{q}$
    $\boldsymbol{v} = \boldsymbol{p}$
    if $a > 1$ **then**
      $\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{V}(\boldsymbol{V}'\boldsymbol{p})$
      $\boldsymbol{u} = \boldsymbol{u} - \boldsymbol{T}(\boldsymbol{T}'\boldsymbol{u})$
    end if
    $\boldsymbol{v} = \boldsymbol{v}/SQRT(\boldsymbol{v}'\boldsymbol{v})$
    $\boldsymbol{S} = \boldsymbol{S} - \boldsymbol{v}(\boldsymbol{v}'\boldsymbol{S})$
    Store $\boldsymbol{r}$, $\boldsymbol{t}$, $\boldsymbol{p}$, $\boldsymbol{q}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ into $\boldsymbol{R}$, $\boldsymbol{T}$, $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{U}$, and $\boldsymbol{V}$, respectively.
  end for
  $\boldsymbol{B} = \boldsymbol{R}\boldsymbol{Q}'$

C. Binary Regression Trees, Gradient Boosted Regression Trees, and Random Forests

Algorithm 3 grows a complete binary regression tree. Algorithm 4 implements the gradient boosted regression tree, and Algorithm 5 delivers the random forest.

4.5.2 Sample Splitting

Figure A1: Sample Splitting Scheme



This figure displays the sample splitting scheme used for tuning the hyperparameters and evaluating the out-of-sample performance of machine learning models. The dark, gray, light gray bars indicate training, validation, and testing samples respectively.

There is 11 years' data in total (from 2006 to 2017) in the bond sample. The training sample is started with the first 5 years' data and the validation sample is the following 3 years' data. I refit all the models once every year, and each time I refit, I increase the training sample by one year. I maintain the same size of the validation sample, but roll it forward to include the most recent one year data each time I refit the models. The test sample uses the data from the subsequent year. Figure A1 vividly displays the sample splitting scheme.

4.5.3 Hyperparameter Tuning

Table A1 describes the set of hyperparameters and their value ranges used for tuning each machine learning model.

Algorithm 3 Binary Regression Trees

Initialize the stump. $C_1(0)$ denotes the range of all covariates, and $C_l(d)$ denotes the l-th node of depth d.

for d from 1 to L **do**

for i in $\{C_l(d-1),\, l = 1, \ldots, 2^{d-1}\}$ **do**

i) For each feature $j = 1, 2, \ldots, P$, and each threshold level α, define a split as $s = (j, \alpha)$, which divides $C_l(d-1)$ into C_{left} and C_{right}:

$$C_{left}(s) = \{z_j \leq \alpha\} \cap C_l(d-1); \quad C_{right}(s) = \{z_j > \alpha\} \cap C_l(d-1),$$

where z_j denotes the jth covariate.

ii) Define the impurity function:

$$\mathcal{L}(C, C_{left}, C_{right}) = \frac{|C_{left}|}{|C|} H(C_{left}) + \frac{|C_{right}|}{|C|} H(C_{right}),$$

where $H(C) = \frac{1}{|C|} \sum_{z_{i,t} \in C} (r_{i,t+1} - \theta)^2$, $\theta = \frac{1}{|C|} \sum_{z_{i,t} \in C} r_{i,t+1}$, and $|C|$ denotes the number of observations in set C.

iii) Select the optimal split:

$$s^* \leftarrow \underset{s}{\operatorname{argmin}}\, \mathcal{L}(C(s), C_{left}(s), C_{right}(s)).$$

iv) Update the nodes:

$$C_{2l-1}(d) \leftarrow C_{left}(s^*), \quad C_{2l}(d) \leftarrow C_{right}(s^*).$$

end for

end for

Result: The output of a binary regression tree is given by:

$$g(z_{i,t}; \theta, L) = \sum_{k=1}^{2^L} \theta_k \mathbf{1}\{z_{i,t} \in C_k(L)\},$$

where $\theta_k = \frac{1}{|C_k(L)|} \sum_{z_{i,t} \in C_k(L)} r_{i,t+1}$. For a single binary complete regression tree $\mathcal{T}$ of depth L, the VIP for the covariate z_j is

$$\text{VIP}(z_j, \mathcal{T}) = \sum_{d=1}^{L-1} \sum_{i=1}^{2^{d-1}} \Delta im(C_i(d-1), C_{2i-1}(d), C_{2i}(d)) \mathbf{1}\{z_j \in \mathcal{T}(i, d)\},$$

where $\mathcal{T}(i, d)$ represents the covariate on the i-th (internal) node of depth d, which splits $C_i(d-1)$ into two sub-regions $\{C_{2i-1}(d), C_{2i}(d)\}$, and $\Delta im(\cdot, \cdot, \cdot)$ is defined by:

$$\Delta im(C, C_{left}, C_{right}) = H(C) - \mathcal{L}(C, C_{left}, C_{right}).$$

Algorithm 4 Gradient Boosted Regression Trees

Initialize the predictor as $\hat{g}_0(\cdot) = 0$;

for b from 1 to B **do**

Compute for each $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, the negative gradient of the loss function $l(\cdot, \cdot)$:

$$\varepsilon_{i,t+1} \leftarrow -\frac{\partial l(r_{i,t+1}, g)}{\partial g}|_{g=\hat{g}_{b-1}(z_{i,t})}.$$

Grow a (shallow) regression tree of depth L with dataset $\{(z_{i,t}, \varepsilon_{i,t+1}) : \forall i, \forall t\}$

$$\hat{f}_b(\cdot) \leftarrow g(z_{i,t}; \theta, L).$$

Update the model by

$$\hat{g}_b(\cdot) \leftarrow \hat{g}_{b-1}(\cdot) + v\hat{f}_b(\cdot),$$

where $v \in (0, 1]$ is a tuning parameter that controls the step length.

end for

Result: The final model output is

$$\hat{g}_B(z_{i,t}; B, v, L) = \sum_{b=1}^{B} v\hat{f}_b(\cdot).$$

Algorithm 5 Random Forests

for b from 1 to B **do**

Generate Bootstrap samples $\{(z_{i,t}, r_{i,t+1}), (i,t) \in \mathit{Bootstrap}(b)\}$ from the original dataset, for which a tree is grown using Algorithm 3. At each step of splitting, use only a random subsample of all predictors. The resulting bth tree is:

$$\hat{g}_b(z_{i,t}; \hat{\theta}_b, L) = \sum_{k=1}^{2^L} \theta_b^{(k)} \mathbf{1}\{z_{i,t} \in C_k(L)\}.$$

end for

Result: The final random forests output is given by the average of the outputs of all B trees.

$$\hat{g}(z_{i,t}; L, B) = \frac{1}{B}\sum_{b=1}^{B} \hat{g}_b(z_{i,t}; \hat{\theta}_b, L).$$

Table A1: Value Ranges of Hyperparameters for Machine Learning Models

This table shows the value ranges of hyperparameters I use to tune the corresponding machine learning model. ξ represents quantiles of residuals from the simple OLS model. ρ equals 0 for the Lasso method and 1 for the Ridge method. K in the Group Lasso method indicates the number of knots. P under PCR and PLS is the number of predictors. GBRT and RF denote gradient boosted regression trees and random forests respectively. L indicates the depth of trees. B under GBRT is the number of trees in the ensemble, and under RF is the number of bootstrap samples. v is the shrinkage weight and F is the number of features in each split.

Linear Machine Learning Models

Model	OLS+H	Lasso	Ridge	Elastic Net	Group Lasso	PCR	PLS
Hyperparameters	ξ	λ	λ	λ, ρ	λ, K	K	K
Value Range	ξ =85% to 99%	$\lambda \in (10^{-4}, 10^{10})$	$\lambda \in (10^{-4}, 10^{10})$	$\lambda \in (10^{-4}, 10^{10})$ $\rho \in (0.1, 0.9)$	$\lambda \in (10^{-4}, 10^{10})$ $K \in (4, 10)$	$K \in (1, P-1)$	$K \in (1, P-1)$

Non-Linear Machine Learning Models

Model	GBRT	RF
Hyperparameters	L, B, v	L, B, F
Value Range	$L = 1$, $B = 1 \sim 50$ $v \in (10^{-1}, 1)$	$L = 1 \sim 6$, $B = 300$ $F \in \{3, 5, 10, 20, 30, 50\}$

4.5.4 Construction of Predictors

Bond-level characteristics:

1. **dsrisk** (downside risk): The proxy for downside risk, 5% VaR, is based on the lower tail of the empirical return distribution, that is, the second lowest monthly return observation over the past 36 months. The original measure is multiplied by -1 for convenience of interpretation. A bond is included in VaR calculation if it has at least 24 monthly return observations in the 36-month rolling window before the test month.

2. **illiq** (illiquidity): $illiq = -\text{Cov}_t(\Delta b_{itd}, \Delta b_{itd+1})$, where $\Delta b_{itd} = ln(B_{itd}) - ln(B_{itd-1})$ is the log price change for bond i on day d of month t and B_{itd} is the clean end-of-day bond price. Price changes may be between prices over multiple days if a bond does not trade in some days and the maximum difference in days is limited to one week.

3. **ltr** (long-term reversal): The long-term reversal is quantified as the past 36-month cumulative returns from month $t-48$ to $t-13$, skipping the 12-month momentum and the short-term reversal months.

4. **mat** (years-to-maturity): Years-to-maturity is calculated as (bond maturity date-bond offering date)/365.

5. **mom** (momentum): Momentum is quantified as the past 6-month cumulative returns from month $t-7$ to $t-2$, skipping the short-term reversal month.

6. **rating** (credit rating): The S&P ratings are converted into numerical ratings to facilitate the analysis, for example, 1 = AAA, 2 = AA+, 3 = AA, ..., 20 = CC, 21 = C, and 22 =D. The most recent rating is merged with the relevant bond-month observation.

7. **size** (amount outstanding): It is the natural logarithm of bond amount outstanding.

8. **skew** (skewness) and **vol** (volatility): A 60-month rolling window is used to generate the monthly time-series measures of skewness and volatility of

excess bond returns. A bond is included in the sample if it has at least 24 monthly return observations in the 60-month rolling window before the test month.

9. **str** (short-term reversal): The short-term reversal of a bond for month t is measured by its previous month return.

10. **uncbeta** (economic uncertainty beta): The economic uncertainty beta is estimated from the monthly rolling regressions of excess bond returns on the change in the economic uncertainty index over a 36-month fixed window while controlling for the bond market portfolio:

$$r_{i,t} = \alpha_{i,t} + \beta_{i,t}^{UNC} \Delta UNC_t + \beta_{i,t}^{MKT} MKT_t + \epsilon_{i,t}.$$

$r_{i,t}$ is the excess return of bond i in month t. ΔUNC_t is the change in the economic uncertainty index over month t (One-month-ahead economic uncertainty index is obtained from Sydney Ludvigson's website.). MKT_t is the excess return on the bond market portfolio in month t, proxied by the value-weighted average returns of all corporate bonds in the sample. $\beta_{i,t}^{UNC}$ is the uncertainty beta of bond i in month t. A bond is included in economic uncertainty beta calculation if it has at least 24 monthly return observations in the 36-month rolling window before the test month.

Bond market factors:

1. **bmkt** (bond market excess return): The bond market excess return is calculated as the value-weighted average returns of all corporate bonds in the sample minus the one-month Treasury bill rate.

2. **def** (default spread): The default spread is the difference between the return on a market portfolio of long-term corporate bonds and the long-term government bond return.

3. **liq** (market liquidity risk): First, estimate the liquidity level for an individual

bond ($\pi_{i,t}$) in month t using the regression,

$$r^e_{i,j+1,t} = \rho_0 + \rho_1 r_{i,j,t} + \pi_{i,t} sign(r^e_{i,j,t}) Vol_{i,j,t} + \epsilon_{i,j+1,t}.$$

$r_{i,j,t}$ is the return of bond i on day j in month t, $r^e_{i,j,t}$ is the bond's return in excess of the bond market return, $sign(r^e_{i,j,t})$ is an indicator function whose value is equal to 1 if $r^e_{i,j,t}$ is positive and -1 if it is negative, and $Vol_{i,j,t}$ is the dollar volume of bond i on day j in month t. A bond is included in the sample only if there are at least 10 daily return observations with which to estimate the regression in a month. Second, aggregate individual liquidity measures month by month to generate the marketwide liquidity series: $\pi_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \pi_{i,t}$, where N_t is the number of corporate bonds in each month. Third, scale the difference in monthly aggregate liquidity measures by the ratio of capitalization of the bonds in the sample to account for the effects of changes in the growth in size of the bond market and the sample size, i.e., $\Delta\pi_t = (M_t/M_1)(\pi_t - \pi_{t-1})$, where M_t is the total dollar value at the end of month $t-1$ of bonds included in month t. Fourth, obtain liquidity innovations ($\hat{e}_t$) from the following regression, $\Delta\pi_t = a_0 + a_1 \Delta\pi_{t-1} + a_2(\frac{M_{t-1}}{M_1})\pi_{t-1} + e_t$. Finally, the market liquidity factor is calculated as residual estimates $\hat{e}_t$ multiplied by 100, $L_t = 100\hat{e}_t$, because the residual term is typically small.

4. **term** (term spread): The term spread is the difference between the monthly long-term government bond return and the one-month Treasury bill rate measured at the end of the previous month.

Chapter 5

Conclusions

This thesis consists of three papers that cover three different topics in empirical cross-sectional asset pricing. In the first paper (Chapter 2), we identify three consumption risks from a consumption-based asset pricing model where the representative agent has recursive utilities and the consumption growth follows a four-state Markov-switching process through linearizing the pricing kernel. We find that all the three consumption risks are significantly priced in the cross-section of options and straddles returns. The consumption growth risk and expected consumption growth risk command positive risk premiums, while the consumption volatility risk commands a negative risk premium. Positive consumption growth and expected consumption growth risk premiums and negative consumption volatility risk premium indicate that investors prefer early resolution of uncertainty. Moreover, our empirical results are able to explain the negative relation between underlying stock idiosyncratic volatilities and delta-hedged options returns, and the positive relation between option moneyness and delta-hedged options returns. Options with higher underlying stock idiosyncratic volatilities tend to have larger negative consumption growth beta and larger positive consumption volatility beta, and thus have lower returns. Similarly, options with higher moneyness tend to have smaller negative consumption growth beta and smaller positive consumption volatility beta, and thus have higher returns.

In the second paper (Chapter 3), we use a harzard model to capture a firm's probability of failure in the future twelve months which is also called distress

risk, and discover a significantly negative relation between distress risk and the cross-sectional corporate bond returns. The negative relation between distress risk and bond returns is analogous to the often negative relation between distress risk and stock returns, which is so-called "distress anomaly". Our finding of a negative relation between distress risk and bond returns casts serious doubts to shareholder advantage theory, which predicts a positive relation between distress risk and bond returns. Besides, shareholder advantage proxies are not able to condition the negative relation between distress risk and bond returns empirically. Therefore, we try to investigate whether asset risks can explain the negative relations between distress risk and both stock and bond returns instead of financial risks. In a real options model which allows for gradual disinvestments of productive capacity, we are able to simulate negative relations between distress risk and both expected stock and debt returns. We conclude that valuable disinvestment options of distressed firms can explain the distress anomaly in both stock and bond returns consistently. Moreover, we find that disinvestment options proxies can condition the relations between distress risk and stock and bond returns with correct signs empirically.

In the third paper (Chapter 4), I use 7 linear and 2 non-linear machine learning models to predict the cross-sectional corporate bond returns out-of-sample. Predictors include 11 bond-level and 6 stock-level characteristics, and their interaction terms with 4 bond market factors. Therefore, there are 85 predictors in total. I show that when the number of predictors is large, the simple linear model does not possess stable out-of-sample predictive power and generates a huge negative out-of-sample R^2. The simple linear model with Huber loss which can mitigate the influence of outliers improves the performance of the simple linear model, but still generates a relatively large negative out-of-sample R^2. Among all the linear machine learning models, the penalized linear methods, which include the Lasso, the Ridge, and the elastic net models, perform the best due to their ability to reduce influences of noises in in-sample fit and thus generate stable out-of-sample predictions. The Group Lasso performs the second best since it considers the

non-linear transformations of predictors and those non-linear transformations do add extra information to the regression. The dimension reduction methods, which include the partial least squares (PLS) and the principal component regression (PCR) models, perform the least best because the correlations among predictors are low. Compared with linear machine learning models, the non-linear models which include gradient boosted regression trees and random forests, posses stronger predictive power. The outstanding performance of non-linear machine learning models are due to their ability to allow for flexible functional form of the regression equation and multi-way interactions among predictors.