Essays in Empirical Asset Pricing with Machine Learning

by

Matthias Büchner

Thesis

Submitted to The University of Warwick

for the degree of

Doctor of Philosophy

in Finance & Econometrics

Warwick Business School

September 2020

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

"I think that when we know that we actually do live in uncertainty, then we ought to admit it; it is of great value to realize that we do not know the answers to different questions. This attitude of mind - this attitude of uncertainty - is vital to the scientist, and it is this attitude of mind which the student must first acquire."

—Richard P. Feynman; The Relation of Science and Religion (1956)

I wish to thank my supervisors during my time at Warwick, Prof. Daniele Bianchi, Prof. Roman Kozhan and Prof. Philippe Müller, for their invaluable guidance and unwavering support through all stages of my degree. I am very grateful for the many inspiring discussions, comments, and advices that have substantially contributed to the success of this thesis and my development as a researcher.

Also, I would like to thank greatly my co-authors, Prof. Bryan Kelly at the Yale School of Management and Prof. Andrea Tamoni at Rutgers Business School. I thank Bryan for hosting me during a research visit to Yale that has been instrumental in my academic development beyond my expectations. I thank Andrea for his unending willingness to provide feedback on my research ideas and for always motivating me to keep pressing on. In addition, I would like to thank the members of the Finance Group at Warwick Business School for their support, including stimulating feedback during my Brown Bag Seminar presentations, funding for many conference attendances, and for giving me the opportunity to teach on Warwick's masters courses. Furthermore, I would like to thank Prof. Arie Gozluklu for his comments and suggestions during my upgrade and annual review panels over the past four years.

My PhD journey would not have been half as happy without the many friends I have made during this time. At Warwick I would like to thank Kolja Johannsen, Mehrshad Motahari, Alexander Dickerson, my desk neighbour Francesco Saverio Gaudio, Andrea De Polis, Giulia Mantoan, and Vashisht Bhatt. At Yale, I would like to thank Natee Amornsiripanitch, Thomas Bonczek, and Kaushik Vausdevan.

The accomplishment of this thesis would not have been possible without the unconditional support of my family, that is my parents Agnes and Ulrich-Michael, as well as my brother, Martin, for which I will be forever grateful. Finally, and most importantly my beloved Theresa cannot be credited enough with motivating me through these four years, enduring all the ups and downs on my journey, and pushing me to realise my ambitions. She is who I dedicate this thesis to.

Declarations

This thesis is submitted to The University of Warwick in support of the requirements for the degree of Doctor of Philosophy. I confirm that I have not submitted the thesis or any work therein for a degree at another university.

In addition, I declare that the first chapter is the product of joint work with Prof. Roman Kozhan and Prof. Daniele Bianchi, the second chapter is the product of joint work with Prof. Daniele Bianchi and Prof. Andrea Tamoni and parts of the chapter are accepted for publication in the *Review of Financial Studies*, the third chapter is the product of joint work with Prof. Bryan Kelly. Finally, the fourth chapter is my sole work.

Abstract

This thesis consists of four papers on topics in empirical asset pricing with a particular focus on applications of machine learning. The first paper investigates the interplay of predictable trading behaviour and asset prices. We show that predictable order flow is associated with improved liquidity and market efficiency. In addition, we find evidence for a priced factor constructed from order flow predictability, contributing to the literature that connects market microstructure features and asset prices. The second paper evaluates the efficacy of machine learning based forecasts of bond excess returns and contributes to a better understanding of the formation of bond risk premia. We show that machine learning techniques outperform the principal components benchmarks used in extant literature and deliver substantial economic gains to investors. The third paper investigates the risk-reward trade-off in index options through the lens of a factor modelling approach. We show that a factor model with five factors and time-varying loadings instrumented with option characteristics, explains the vast majority of variation in delta-hedged option returns. The recovered factors point to jump, volatility and term structure spread risks. Finally, the fourth paper studies the systematic drivers of asset holdings in a novel factor modelling approach. I document the existence of a factor structure in holdings changes that points to distinct, well-understood economic channels as drivers of asset holdings. Using investor-specific factor loadings I find evidence for pro-cyclical trading of banks and mutual funds as well as counter-cyclical trading of investment advisors and pension funds. Furthermore, I document that changes to institutional investor holdings driven by systematic factors are negatively associated with future returns, suggesting a price pressure channel as a driver for return reversals.

Introduction

Over the past few decades we have witnessed a unique marriage of two hitherto distinct scientific fields, statistics & computer science, culminating in one powerful, new paradigm: machine learning. In a multitude of applications ranging from image recognition in medical diagnostics to natural language processing performed in smart phones, machine learning has proven to be of essence to model complex statistical relations. However, in the context of financial economics in general and asset pricing in particular, our understanding of the benefits and shortcomings of machine learning is relatively less well developed.[1] This thesis explores various applications of machine learning in the context of empirical asset pricing.

In this thesis, machine learning features both in its supervised and unsupervised form.[2] In supervised machine learning the objective is to find the mapping between one (or more) target variables, e.g. a stock's return, and a potentially large number of predictive signals, e.g. a stock's characteristics.[3] In this regime, commonly used models include neural networks, regression trees, or penalised regressions. In contrast, unsupervised machine learning aims to uncover systematic patterns in data without reference to a target variable - a typical example of such a technique is principal components analysis.[4] In (financial) economics, the choice of supervised versus unsupervised learning is often motivated by economic principles and prior knowledge. For example, in the context of the measurement of risk premia, asset pricing theory posits the existence of a stochastic discount factor that links conditioning information and expected returns, hence giving rise to a supervised

[1]Weigand (2019) performs a review of the recent, emerging literature on machine learning in empirical asset pricing.

[2]For a comprehensive overview and textbook treatment of machine learning techniques in supervised and unsupervised form refer to, e.g., Friedman et al. (2001); Bishop (2006); Murphy (2012).

[3]This is essentially the setting faced in the context of measuring risk premia (cf., Gu et al., 2020).

[4]In the empirical asset pricing literature examples of unsupervised learning settings include Kelly et al. (2019); Gu et al. (2019).

learning problem. In this context, supervised learning is a tool to recover, with potentially little constraint on functional form, the mapping between conditioning information and expected returns. In contrast, the study of asset returns in a statistical factor modelling approach is a situation that allows the use of unsupervised machine learning. In this context, the researcher forgoes the imposition of economic theory and instead follows a data-driven approach to recover relevant patterns from the observed data.

This thesis consists of four chapters that, although distinct in their focus on different asset classes and market aspects, are connected by their statistical methodology, i.e. applications of both supervised (Chapters 1 and 2) and unsupervised machine learning (Chapters 3 and 4). Chapter 1 investigates the predictability of order imbalances and the consequences of such predictability for market quality (i.e. liquidity and efficiency) and asset prices. Methodologically, the chapter employs a dynamic linear model in a big data context with millions of observations. The empirical results point towards higher predictability of order flow being associated with improved market quality and further suggest the existence of a risk factor associated with order imbalance predictability that is priced in the cross-section of stock returns. Chapter 2 investigates the use of machine learning techniques in the context of regression-based measurement of bond risk premia and finds notable incremental predictive power of machine learning models vis-à-vis benchmark approaches that also manifests in significant economic gains to investors. Chapter 3 studies option returns through the lens of an advanced statistical factor model that circumvents the common shortcomings of simpler factor models such as principal components analysis by modelling an option's time-varying factor loadings as functions of asset characteristics. The proposed model is able to capture the vast majority of variation in delta-hedged option returns and delivers an economically meaningful factor structure. Finally, Chapter 4 extends the factor modelling technique used in Chapter 3 to study the systematic drivers of institutional equity portfolio holdings, thereby contributing to a better understanding of the main economic channels that determine portfolio choice as well as investor-specific behaviour.

The chapters are presented in the form of self-contained papers and are summarised in more detail below. Chapter 1 is co-authored with Prof. Daniele Bianchi and Prof. Roman Kozhan, Chapter 2 is co-authored with Prof. Daniele Bianchi and Prof. Andrea Tamoni, Chapter 3 is co-authored with Prof. Bryan Kelly, and Chapter 4 is my sole work.

Chapter 1 - Predictability of Order Imbalance, Market Quality and Equity Cost of Capital

Order imbalances measure the difference between the amount of buyer- versus seller-initiated trades and have been recognised as important drivers of asset prices in many models of the market microstructure literature. This literature often assumes market makers that decompose order imbalances into an informed and uninformed component to efficiently set prices and provide liquidity. Anecdotal evidence suggests that in practice market makers do not only try to assert the informativeness of order imbalances, but also try to *predict* future order imbalances. The predictability of order imbalances could affect market quality and expected returns in two contrasting ways. On the one hand, high predictability of the uninformed component of order imbalance could allow market makers to mitigate adverse selection and inventory risks, therefore improving market quality. On the other hand, a high predictability of the uninformed component of order imbalances could be detrimental for market quality due to predatory trading or front-running by informed traders. While the academic literature has studied the impact of order imbalances on market quality and expected returns, little attention has been paid to the impact of *predictability of order imbalances*. The aim of this paper is to address this aspect.

In order to investigate the impact of predictable order imbalances we take the stance of a sophisticated market participant (e.g., a market maker) and predict day-ahead order imbalances for a large sample of U.S. equities using a dynamic linear model. From the forecasts we obtain a measure of realised forecast accuracy as a proxy of order imbalance predictability - the predictive likelihood. To determine the origins of predictability in order imbalances we use two different samples, i.e., the total order imbalance across all market participants and the order imbalance of active pension funds. We document both cross-sectional as well as time series variation in the degree of predictability in both types of order imbalances. Importantly, we find evidence that the degree of predictability of total order flow is related to the predictability of order imbalances from pension fund trading.

Since the direction of the effect of predictability of order imbalances on market quality is not clear ex ante, we investigate the impact of predictability on measures of liquidity (spread measures) and proxies of market efficiency (return variance ratios / autocorrelations). We find evidence of a negative and statistically significant relation between order imbalance predictability and illiquidity, while controlling for variables typically considered as important determinants of illiquidity. The effect is economically large: a one standard deviation increase in predictability leads to a 3.00 bps change in effective spreads (a 3.6% change relative to the sample average). Similarly, we find that price efficiency improves when order imbalances are more predictable. These results suggest order imbalance predictability as an additional channel in determining market quality.

We complement the results on market quality by exploring consequences of order imbalance predictability for asset pricing. If the degree of predictability of order imbalances captures a cost of asymmetric information it could pose a systematic risk to investors (Easley and O'Hara, 2004). In a portfolio sorting exercise we document that the stocks with the highest degree of order imbalance predictability earn substantially lower returns than those with low order imbalance predictability. Following this evidence, we construct a low-minus-high factor, POI, that shorts stocks with high predictability and longs stocks with low predictability. We find that this factor earns an annualised Sharpe ratio of 1.00. In a Fama-MacBeth regression exercise we find that the POI factor is priced in the cross-section of stock returns after controlling for conventional risk factors (size, value, momentum) as well as a host of factors considered as proxies of asymmetric information (liquidity, volatility of order imbalances, trade-to-quote ratios).

Chapter 2 - Bond Risk Premia with Machine Learning

Interest rates, and in particular the interest rates on government-issued bonds such as Treasury bonds, are of relevance to many market participants. A central question in this regard is whether and to which extent changes in interest rates are predictable and by means of which conditioning information. Our paper contributes to this literature by investigating the use of machine learning methodologies in the context of two traditional forecasting applications; one that uses only information from the current yield curve as in Cochrane and Piazzesi (2005); and one that complements yield curve information with a large data set of macroeconomic indicators as in

Ludvigson and Ng (2009).

We compare a number of machine learning approaches such as regression trees (including random forests, extreme trees, gradient boosted trees) and neural networks in terms of their ability to predict the one-year holding period bond excess returns with the de-facto academic benchmark, i.e. principal components regression. Across both applications (yields-only and yield plus macroeconomic information), we show that machine learning methods, and especially neural networks, are able to detect predictable variation in bond excess returns above and beyond what is explained by data compression techniques, i.e., principal components regressions. A portfolio allocation exercise confirms that gains in statistical predictability translate into economic gains for investors.

The finding that macroeconomic information is relevant above and beyond yield curve information suggests significant deviations from the expectations hypothesis which states that the time t term structure should contain all information relevant for future interest rates. Relatedly, we document heterogeneity in the type of macroeconomic variables that are relevant for forecasts of bond returns at different maturities: stock and labour market variables are more important for short-term maturities, whereas order and inventory variables are more relevant for long-term maturities.

In addition to shedding light on the statistical aspects of predictability, we investigate the economic properties of the machine learning based forecasts. We find that the forecasts of bond excess returns are counter-cyclical and relate to proxies of macroeconomic uncertainty and time-varying risk aversion. Hence, our findings are consistent with theory models that feature both aforementioned channels as in, e.g., Bekaert et al. (2009) and Creal and Wu (2018).

Overall, the paper demonstrates that machine learning can be useful to detect predictable variation in a "small data" setting, suggesting that the benefits of the machine learning toolbox are not only relegated to big data contexts. However, it is important to note that heavy model regularisation and careful model choice are important for the approach to be effective. In addition, the paper makes a contribution to the discussion around unspanned macro factors in bond returns, suggesting that macroeconomic information contains useful predictive power for bond excess returns beyond what is captured in the yield curve.

Chapter 3 - A Factor Model for Option Returns

Factor structures present a common empirical approach to study risk-reward trade-offs in asset pricing. While this approach has been employed successfully in asset classes including equities and fixed income, less is known about the efficacy of factor modelling approaches in the context of option returns. This is due to two challenges unique to the dynamics of option contracts. Firstly, options often have short lifespans, making it challenging to estimate factor loadings using, e.g., rolling window regressions. Secondly, the risk attributes that determine the "identity" of an option contract change more rapidly than those of the underlying asset. Instead, the options literature has focussed on parametric, no-arbitrage based models to describe options - an approach that benefits from enforcing economic restrictions such as no-arbitrage and mathematical elegance, but that is likely too simplistic to describe empirically observed patterns in option returns.

In this paper, we investigate the use of a latent factor model with time-varying loadings, namely, Instrumented Principal Components Analysis (IPCA) (Kelly et al., 2019), for describing option returns. In particular, IPCA is based on employing asset-specific conditioning information, i.e. characteristics, to instrument time-varying loadings. In the context of index options on the S&P 500 that we study in this paper, a natural choice of characteristics are the basic option attributes such as option strike price and time-to-maturity. In addition, a number of characteristics implied by option Greeks from the Black and Scholes (1973) model, present themselves as characteristics of primary relevance to determine an option's risk attributes.

We estimate IPCA models for the cross-section of S&P 500 options and assess their performance using variations of measures of explained variance (R^2). We find that a single latent factor is sufficient to explain more than two thirds of the return variation in monthly, delta-hedged option returns. For comparison, none of the observable option factor models (Coval and Shumway, 2001; Frazzini and Pedersen, 2012) proposed in earlier literature achieves such performance, even if instrumented with the same set of characteristics. Model performance increases become small after four to five factors, at which point more than 90% of the variation in delta-hedged returns is captured in the model.

To interpret the factors, we study the correlation structure of the factors with portfolios of option contracts sorted on moneyness (strike) and time-to-maturity as well as observable factors proposed in earlier literature. We find evidence for a tail risk factor, a maturity risk factor and a volatility risk factor (variance risk), suggesting that the factors recovered by IPCA are not only relevant in a statistical, but also economic sense. In further exercises, we assess the importance of the characteristics used to instrument time-varying loadings and find that, overall, moneyness and implied volatility contribute most strongly to the model fit performance. Finally, we complement the monthly horizon result with results at the daily frequency. At daily frequency, a traders' rule-of-thumb approach, implied by the Black-Merton-Scholes model (formalised in Carr and Wu (2020)), is used to benchmark our results.

Overall, the paper aims to contribute to a more systematic understanding of the major drivers of index option returns, similar in spirit to the developments in the literature focussed on factor models for the underlying asset. IPCA can provide a rather universal benchmark approach that allows to distinguish alphas from factor risk, hence contributing to the understanding of the risk-return trade-off investors face when trading in options.

Chapter 4 - What Drives Asset Holdings? Commonality in Investor Demand

A common assumption made in the asset pricing literature is to assume atomistic investors, that is, trading behaviour across investors is assumed to be uncorrelated. However, the fundamental mechanism of market clearing demands that - akin to a zero-sum game - for every quantity of an asset bought an equal quantity is sold. This induces an almost mechanical form of correlated trading across investors. The aim of this paper is to model the correlated trading behaviour of investors in a novel factor modelling approach to uncover the systematic drivers of asset holdings.

To track trading behaviour of investors, this paper uses institutional holdings data. The use of holdings data in asset pricing research is still rather recent with most of the literature relying on price and consumption data. However, holdings data is an interesting complement to more commonly used data sets as it allows the researcher to trace investor demands across assets and over time. In addition, the study of holdings data allows a closer look into the processes through which information gets incorporated into prices. More precisely, I use changes in holdings as a proxy for investor demands and investigate the existence of a factor structure in holdings changes to account for correlated investor behaviour. This empirical approach reflects an economy with aggregate shocks in the presence of which heterogeneous investors choose from a pool of differentiated assets (cf. Campbell et al., 2003).

Methodologically, I extend a factor modelling approach known as *Instrumented Principal Components Analysis* (IPCA) to incorporate investor-specific factor loadings. IPCA is distinguished by its use of conditioning information, i.e. asset characteristics, to instrument time-varying factor loadings. Upon estimation of the extended IPCA model, dubbed "IPCA3D" for distinction, I firstly recover a set of factors in holdings changes common to all investors and secondly investor-specific time-varying loadings on the aforementioned factors. Furthermore, the model incorporates an intercept term that allows for asset characteristics to explain changes in holdings without the interaction of factors. This latter component is motivated by a recent strand of literature (cf. Koijen and Yogo, 2019a,b) that has proposed a particular way to think about holdings data, i.e. characteristics-based demand.

Using the data set of 13F U.S. investor holdings aggregated to (investor) sector level, I estimate the IPCA3D model and find that holdings changes exhibit a low-dimensional factor structure. Three factors are sufficient to capture the bulk of the systematic variation in holdings changes. In a comprehensive correlation analysis, I find that the factors are related, firstly, to the state of the economy / business cycle, and secondly, to financial conditions / constraints of investors. The sector-specific factor loadings on the business cycle factor provide an interesting result on the cyclicality of trading behaviour for different investor types: while banks and mutual funds exhibit pro-cyclical trading, hedge funds and to some extent pension funds exhibit counter-cyclical trading. Analysing the importance of different asset characteristics used in the model, I find that contrary to the typical characteristics used in factor models for returns such as the Fama and French (2015) five factors, past returns and liquidity related characteristics dominate the importance ranking. Finally, in an asset pricing analysis I use the decomposition of changes in holdings implied by IPCA3D to reveal two interesting pricing effects. Firstly, systematic demand from banks, mutual funds and pension funds is associated with negative future returns on average. This finding is consistent with an institutional demand pressure induced return reversal effect. Secondly, I find that idiosyncratic demand of hedge funds is a positive predictor of future returns up to four quarters ahead.

Overall, the paper contributes to a recent and growing literature that focusses on using institutional holdings data to further our understanding of asset price formation and expected returns. My paper makes an empirical contribution to this literature by studying investor demand through the lens of a factor modelling approach that allows for correlated investor behaviour. The findings of the paper can be a useful guide for theory literature as to the economic channels that should be considered as key drivers of asset demand.

Chapter 1

Predictability of Order Imbalance, Market Quality and Equity Cost of Capital[1]

> *"The math works. Over the course of a season, there's some predictability to baseball. When you play 162 games, you eliminate a lot of random outcomes. There's so much data that you can predict individual players' performances and also the odds that certain strategies will pay off."*
>
> – Billy Beane; General Manager Oakland A's

1.1 Introduction

Most market micro-structure models recognise order imbalance as the main driver of the dynamics of asset prices. On the one hand, market makers use the information contained in the observed order flow to efficiently set up prices and provide liquidity. Therefore, the ability to accurately predict the uninformed part of order imbalance is essential for both liquidity provision and market efficiency as it allows market makers to mitigate both adverse selection and inventory risks. On the other hand, a high degree of order imbalance predictability can be detrimental for market quality as it facilitates predatory trading and front-running by informed traders. While academic

[1]This chapter is based on a research paper jointly authored with Daniele Bianchi and Roman Kozhan. The paper has been presented at the Lancaster-Warwick PhD Workshop 2018, London Empirical Asset Pricing Workshop (2018), Goethe University Finance Seminar 2020, and Warwick University Finance Brown Bag Seminar.

literature studies the effect of uninformed order imbalance on trading costs and price efficiency, little attention has been devoted to the effect of the predictability of order imbalance on market quality and expected returns. In this paper we fill this gap and empirically address this question.

We estimate a dynamic linear regression model to predict the day-ahead order imbalance for each stock in the sample and use a forecasting accuracy measure – the predictive likelihood – to measure the degree of order imbalance predictability. In order to contrast the effect of predictability of total versus uninformed order imbalance we use two distinct samples: total order imbalance of a large cross-section of U.S. equities for the period from 1995 to 2013 and order imbalance from pension funds on a subset of U.S. stocks with active trading for the period from 2006 to 2010. We document substantial time series and cross-sectional variation in the degree of predictability of the order imbalance for both types of order imbalance. In addition, we show that order imbalance predictability is cross-sectionally correlated with several stock characteristics, such as size, bid-ask spreads, share turnover, trade-to-quote ratios, and the volatility of order imbalance.

We provide evidence of a negative and statistically significant relation between the predictability of order imbalance and illiquidity, while controlling for variables which have been widely shown to affect liquidity in addition to fixed effects. Given the predictive nature of these regressions, the results strongly suggest that the predictability of order imbalance can be thought of as an additional channel in determining market liquidity. The effect is economically large: a one standard deviation change in predictive likelihoods leads to a 3.00 bps change in effective spreads; this corresponds to about 3.6% change relative to the sample average. The main empirical results also show that the predictability of order imbalance has a significant impact on price efficiency. By using the variance ratios and the autocorrelations of daily stock returns as proxies for price efficiency, we show that an increase in the predictability of order imbalance leads to increasing market efficiency; that is, a higher predictive likelihood corresponds to a decrease in both variance ratios and absolute autocorrelation compared to their sample averages. We obtain qualitatively similar effects when the uninformed component of the order imbalance is proxied by the order imbalance from active pension funds. Furthermore, the results remain strong when splitting our sample into quintiles based on several firm characteristics, such as market capitalisation, trade-to-quote ratios, idiosyncratic volatility, passive institutional ownership. Finally, we show that both effects for liquidity and price efficiency are more pronounced during periods of high market uncertainty, as proxied by the VIX index.

If the degree of order imbalance predictability captures the costs of information asymmetry, it poses a systematic risk for investors (see, e.g., Easley and O'Hara, 2004). To test this intuition we examine if the predictability of order imbalance proxies for a risk factor that is priced in the cross-section of stock returns. We provide evidence that the forecasting accuracy of order imbalance correlates with monthly future stock returns: the stocks with the highest predictive likelihoods underperform stocks with lowest predictive likelihoods by 17.16%, on an annual basis. Based on this evidence, we construct a low-minus-high Predictability of Order Imbalance portfolio, denoted *POI*, and show that such strategy yields an annualised Sharpe ratio of 1.00. For small stocks, the portfolio with highest predictive likelihood under-performs stocks with lowest predictive likelihood by 23.96% annually. Such economic significance is robust to a double-sorting procedure in which we control for other proxies of adverse selection, such as bid-ask spread, share turnover, trade-to-quote ratio, and both the level and the volatility of the order imbalance.

Delving further into the asset pricing implications of order imbalance predictability, we run a battery of regression-based asset pricing tests. First, we document the existence of a positive relation between future monthly return and the exposure to our *POI* factor. That is, for a given market cap quintile, stocks with lower predictability of order imbalance exhibit higher beta to the *POI* factor. By implementing a conventional Fama-MacBeth two-step regression we show that the risk proxied by our *POI* factor is priced in the cross-section of stock returns. More precisely, the price of risk associated with our *POI* factor is positive and statistically significant after controlling for Fama and French (1993a) factors, the Jegadeesh and Titman (1993) momentum factor, and a number of factors that proxy asymmetric information, such as the Pastor and Stambaugh (2003) liquidity factor, returns on stocks sorted on both the level and the volatility of order imbalance (see, e.g., Chordia et al., 2018), as well as on the trade-to-quote ratio. These results suggest that the main mechanism behind the effect of order imbalance predictability on market quality is via changes in adverse selection rather than reduction of inventory risk.

Our paper contributes to an existing debate on the role of order imbalance predictability in price discovery and the consequential effect on market quality. A number of papers document an improvement of market quality during events where order imbalance of uninformed parties can be predicted by market participants.[2] The main argument in favour of this intuition is that an announced, or anticipated, uninformed trade can attract additional liquidity suppliers and hence might improve liquidity and market resiliency.[3] On the contrary, there are a number of theoretical and empirical results advocating for a negative effect of order imbalance predictability on liquidity and efficiency due to predatory trading and front-running, especially around events with predictable order flow.[4] By estimating a dynamic regression model with time-varying parameters we show, both statistically and economically, that the positive effect of order imbalance predictability on market quality significantly prevails in the data.

This paper also connects to a substantial literature that investigates the asset pricing implications of asymmetric information and inventory risk. Several theory papers argue that adverse selection costs generate a non-diversifiable risk for which investors require compensation (see Admati, 1985, 1993; Dow and Gorton, 1953; Easley and O'Hara, 2004). Madhavan and Smidt (1993), Brunnermeier and Pedersen (2005), Duffie (2010), Hendershott and Menkveld (2014) demonstrate theoretically how inventories of intermediaries generate substantial pricing errors in asset returns. There is also a large body of empirical research testing for the pricing effects of adverse selection and inventory risks, although the evidence is mixed. For instance, a number of papers using different proxies of adverse selection show significant pricing effects.[5] On the opposite, existing research claims that some of these effects are not

[2]See Admati and Pfleiderer (1991); Degryse et al. (2013); Bessembinder et al. (2015); Skjeltorp et al. (2017) among others.

[3]An emerging literature on endogenous liquidity provision also makes an implicit assumption that predictability of order imbalance is a main factor that attracts high-frequency liquidity provision (see Raman et al., 2016; Aït-Sahalia and Saglam, 2017; Saglam, 2020; Brogaard et al., 2018).

[4]See Brunnermeier and Pedersen (2005); Chen et al. (2006); Carlin et al. (2007); Coval and Stafford (2007); Cheng and Madhavan (2009); Petajisto (2011); Mou (2011); Tuzan (2013) among others.

[5]Papers include, but are not limited to, Easley et al. (2002) who estimate the probability of informed trading (PIN), Pastor and Stambaugh (2003) who estimate the price impact, Akbas et al. (2011) who use volatility of liquidity, Hwang and Qian (2011) who estimate an information risk measure based on the price discovery of large trades, Johnson and So (2018) who use the option-to-stock volume ratio as a measure of information asymmetry, Choi et al. (2016) who consider institutional ownership volatility, Yang et al. (2020) who use abnormal idiosyncratic volatility as a measure of information asymmetry and Chordia et al. (2018) who consider volatility of order imbalance.

robust, possibly due to the fact that adverse selection costs are hard to measure empirically (see, e.g. Mohanram and Rajgopal, 2009). Hendershott and Menkveld (2014) show that specialist's inventory contributes to pricing errors in stock prices and Comerton-Forde et al. (2010) show that inventories affect market liquidity. We contribute to this literature by documenting the asset pricing effects of predictability of order imbalance, which is related to both adverse selection costs and inventory risks.

Finally, this paper also speaks to a recent literature that investigates the role of order imbalance in price discovery. Chordia et al. (2002) demonstrate that market-wide returns are driven by lagged order imbalance after controlling for volumes and liquidity measures. Chordia and Subrahmanyam (2004) focus on order imbalance as predictors for individual stocks returns. In particular, they present evidence for order imbalance as a significant force linking market makers' inventories and stock price movements. Andrade et al. (2008) show that order imbalance in one stock affects prices in other stocks. More recent literature focuses on the effect of non-informational order imbalance. Griffin et al. (2003) classify order imbalance for NASDAQ stocks as institutional- vs. retail-originated and document that institutional-driven order imbalance are persistent over several days and have a positive contemporaneous relation to returns. Boehmer and Wu (2008) document that institutional program trades provide liquidity to the market. Finally, Hendershott and Seasholes (2009) find that non-informational market-wide order imbalance affect individual stock returns and their cross-sectional co-movement.

The paper proceeds as follows. Section 1.2 develops a set of testable hypotheses. Section 1.3 describes the data and the variables of interest. Section 1.4 contains the findings from the main empirical analysis. Finally, Section 1.5 reviews our results and concludes.

1.2 Hypotheses Development

In well-functioning and competitive markets the predictability of order imbalance can affect both market quality and prices via several channels. By observing a signal about uninformed order imbalance market makers are able to filter out the informed demand as the residual component of the total contemporaneous order imbalance.

The ability to separate informed and uninformed order flow allows market makers to adjust the level of liquidity in the market as a consequence of a reduced degree of asymmetric information. Market efficiency also improves through new information about future order imbalance being incorporated into the price. In addition, better knowledge of future order imbalance helps market makers to better handle their inventories and further improve liquidity and minimise price pressures due to a reduction of inventory risk.

On the other hand, predictable uninformed order imbalance can also have a detrimental effect on market quality. Sophisticated investors can incorporate knowledge of uninformed order imbalance into their trading strategies and better mask their demand. Moreover, informed traders, in addition to exploiting knowledge of fundamental information, can also front run the uninformed traders to earn additional returns. This reduces profits that market makers can earn to compensate their potential losses against informed traders and forces them to widen the bid-ask spread.

In order to guide the empirical analysis and demonstrate the competing effect of these two forces, we extend the framework of Subrahmanyam (1991) and build an equilibrium model whereby market makers face both adverse selection and inventory concerns. We show that the net effect of predictability on market quality and prices depends on the model parameters. The model is described in detail in Appendix A. For most of the parameter values the positive effects of predictability of order imbalance on market liquidity tend to dominate, whereas for some other values of the parameters the opposite holds. As a result, understanding which one of the two effects described above is dominating in the data becomes ultimately an empirical question.

In the following we outline the main hypotheses that come from the model, which we empirically test later on. The first two hypotheses concern the effect of predictability of uninformed order imbalance on market quality:

Hypothesis 1 (H1) *The predictability of* uninformed *order imbalance in a stock has no effect on the liquidity of that stock.*

Our second testable hypothesis relates price efficiency.

Hypothesis 2 (H2) *The predictability of* uninformed *order imbalance in a stock has no effect on the price efficiency of that stock.*

Our main goal is to establish the effect of predictability of *total* order imbalance on market quality. It is not clear however if predictability of uninformed order imbalance is automatically translated into predictability of total order imbalance. In fact, although the uninformed order flow is a part of the total order flow, the part of the informed traders' demand is endogenously adjusted and might potentially eliminate or reduce the predictability of total order imbalance. Yet, our model establishes that in equilibrium the predictability of uninformed order imbalance generates predictability of total order imbalance (see Appendix A). To test this empirically, we re-formulate H1 and H2 by focusing on total order imbalance. Our re-formulated first hypothesis is again related to market liquidity:

Hypothesis 1b (H1b) *The predictability of* total *order imbalance in a stock has no effect on the liquidity of that stock.*

Similarly to H2 we are interested in the effect of the predictability of total order imbalance on price efficiency. Therefore our re-formulated second hypothesis is defined as

Hypothesis 2b (H2b) *The predictability of* total *order imbalance in a stock has no effect on the price efficiency of that stock.*

Finally, our third hypothesis concerns the asset pricing implications of the predictability of order imbalance. The latter potentially proxies for some source of systematic risk through two channels: adverse selection and inventory management. Easley and O'Hara (2004) provide a theoretical argument that adverse selection costs are significant determinant of the equity cost of capital. In particular, they show that investors demand a higher return to hold stocks with higher degree of private information and that this higher return is a consequence of informed investors being able to better adjust their portfolio to incorporate new information, and as a result disadvantaging uninformed investors. On the other hand, Hendershott and Menkveld (2014) demonstrate that shocks to inventories affect the pricing errors of stocks. In both cases, changes in order imbalance predictability are expected to partly explain the cross-sectional variation of stock returns. Therefore, our third hypothesis reads as follows:

Hypothesis 3 (H3) *The predictability of* total *order imbalance has no effect on the equity cost of capital.*

By testing H3 we can investigate whether the predictability of order imbalance proxies a source of systematic risk and hence contributes to the understanding of the cross-sectional variation of stock returns alongside typical risk factors and anomalies such as liquidity, size, value, and momentum, among others.

1.3 Data and Variable Definitions

We construct total order imbalance using data from the Trades and Quotes tape (TAQ) spanning all market places with reporting obligations. The main sample (we refer to this sample as *Full sample of stocks*) spans the period from January 1, 1995 to December 31, 2013. The TAQ data is pre-processed following Holden and Jacobsen (2014) and then matched with CRSP data using CUSIP / NCUSIP pairs. We exclude all securities with CRSP share code other than 10 or 11, as well as those stocks with average closing price smaller than \$5 and greater than \$1,000. We further require stocks to have more than 5 trades per day on average and a history of quotes longer than three years within our sample period. All variables used in the analysis are winsorised at the 0.1% and 99.9% quantiles so that values smaller / larger than these thresholds are set equal to the quantile value. After accounting for all filters, our sample includes 6,121 firms in total.

We use the algorithm proposed by Lee and Ready (1991) to infer the trade direction as "buy" or "sell" in the TAQ sample and use contemporaneous quotes to sign trades and calculate effective spreads (see, e.g., Holden and Jacobsen, 2014). We define order imbalance oib_{it}^{tot} of stock i on day t as the daily number of buyer-initiated minus seller-initiated dollar volume of transactions scaled by total dollar volume.

We use another sample of stocks in order to proxy for the predictability of uninformed order imbalance. The uninformed order imbalance is approximated by using pension fund trading from the ANcerno database.[6] We obtain institutional

[6] For a discussion of the ANcerno data including fund reporting requirements see Hu et al. (2018).

daily trading data for the period from January 1, 2006 to December 31, 2010.[7] Each institution type identifier distinguishes between clients that are pension plan sponsors vs. money managers. For each execution, the database reports the date of the trade, the execution price of the trade, the stock traded, the number of shares traded, whether the trade was a buy or a sell, and identity codes for the institution making the trade. The granularity of the data makes it particularly useful to compute stock level order imbalance of pension funds (denoted by oib_{it}^{pf}).

In using the pension funds as a proxy for the "uninformed" part of order imbalance we acknowledge that pension funds cover a fraction of the aggregate uninformed order imbalance (that also includes, among others, order imbalance from retail investors and ETFs). Moreover, there are many stocks where pension fund trading is not intense. Sparse trading can lead to spurious evidence of low predictability of order imbalance. To address this issue, we delete stocks with low level of pension fund trading activity. That means, a stock is included in our ANcerno sample if a trade has been executed for at least 50% of the days in the sample.

Our initial ANcerno sample includes 1,560 stocks that record at least one trade by a pension fund. Figure 1.1 presents the empirical cumulative distribution of the fraction of those firms as a function of the number of days during which we observe at least one trade by a pension fund. It shows that about 50% of the firms in the sample experienced a trade by pension funds for at least half their sample length, i.e. on average 791 days of trades. By applying this filter, we retain 789 firms in our *Active pension funds* sample.

We measure liquidity using quoted and effective spreads for both samples. The quoted spread $qspread_{it}$ is the difference between the bid and ask quotes of stock i scaled by the prevailing midpoint. The effective spread $espread_{it}$ is the difference between the midpoint of the bid and ask quotes and the actual transaction price of stock i, expressed as a proportion of the prevailing midpoint. We use all trades and quotes to calculate quoted and effective spreads for each reported transaction and calculate a share-weighted average across all trades in a given day. In addition we compute the realised spread $rspread_{it}$ as the difference between the transaction price and the midpoint five minutes after the trade scaled by the midpoint prevailing immediately before the transaction. The price impact $prcimpact_{it}$ is obtained

[7]The data stops in 2010 since ANcerno database stops providing the client type identifier after 2010.

from the difference of midpoint five minutes after the current trade and the current midpoint scaled by the current midpoint. Finally, the intraday measures of *rspread* and *prcimpact* are averaged within a day.

We use two measures of market efficiency for both samples in our analysis: return autocorrelations and variance ratios. Return autocorrelations have been used to assess price efficiency in a number of papers (see,.e.g. Chordia et al., 2008; Saffi and Sigurdsson, 2011). We construct stock's i daily return autocorrelation ac_{it} based on a past 20 days rolling window basis. We use the past 20 observation days to construct our measure to capture almost a full month of trading activity which is also motivated by earlier literature on long-memory processes in financial markets (see, e.g. Corsi, 2009). Given that we are interested in the magnitude of the autocorrelation, we define an absolute value of a transformed autocorrelation measure as[8]

$$abs\_ac_{it} = \left| \frac{1 + ac_{it}}{1 - ac_{it}} \right|.$$

The variance ratio represents our second measure of price efficiency. The literature shows that the variance of longer horizon returns should equal the variance of shorter horizon returns times the frequency of the short horizon returns in the absence of autocorrelations (see Lo and MacKinlay, 1988). This finding is a property of any random walk process since variance increases linearly with time. In the market micro-structure literature variance ratios have been used by Chordia et al. (2008). Similar to the autocorrelation measure we compute variance ratio using a 20-day rolling window. In the case of a perfect random walk, the variance ratio, defined as $2\,var[r_{it}^{(1)}]/var[r_{it}^{(2)}]$ should be equal to one in expectation, where $var[r_{it}^{(1)}]$ and $var[r_{it}^{(2)}]$ are the variance of one-day and two-day returns, respectively. We use the absolute value of the deviation of this ratio from one as a measure of market efficiency:

$$vratio_{it} = \left| 1 - \frac{2\,var[ret_{it}^{(1)}]}{var[ret_{it}^{(2)}]} \right|.$$

We use a set of order book and market characteristics to forecast order imbalance for both samples. The choice of the predictors is motivated by taking the perspective of a market maker who is interested in forecasting order imbalance. In a typical setting, the market maker has access to past observed order imbalance and

[8] We transform this response variable to increase its variation from 0 to ∞.

further order book measures such as volumes, intraday volatility, etc. Thus, we limit our set of predictors to this class of variables. The daily realised variance $rvar_{it}$ is defined as a sum of squared 5-minute midpoint price returns in stock i during day t. The trade-to-quote ratio ttq_{it} is constructed as a ratio of daily number of quotes (as measured by the changes in bid and ask quotes in the TAQ dataset) with respect to the number of trades in stock i on day t.[9] In addition, we also construct a measure of order book depth imbalance, i.e. the overhang of buy / sell orders in the order book.[10] It is defined a difference between depth on the ask side minus the depth on the bid side of the market divided by the total depth:

$$dib_{it} = \frac{ask\_depth_{it} - bid\_depth_{it}}{ask\_depth_{it} + bid\_depth_{it}}.$$

1.3.1 Predicting Order Imbalance

In order to predict order imbalance we focus on a linear regression approach of the form

$$y_t = \boldsymbol{x}'_{t-1}\beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{1.1}$$

where y_t represents the one-day ahead order imbalance, $\boldsymbol{x}_{t-1}$ a $(k+1) \times 1$ vector of predictors (see description below) in addition to the intercept, and ϵ_t an orthogonal error term with constant variance. A dynamic relationship between order imbalance can be imposed by recursively estimating Eq.(1.1) in a rolling window fashion for a constant window of size n. Although simple to implement such an approach is both ad-hoc and inefficient, as the lack of a more specific parametric form makes testing for time-variation highly dependent on hard-to-justify choices of the window size. As a matter of fact, rolling window estimates $\hat{\beta}_t$ exploit a limited amount of information on order imbalance implicitly assigning equal weight to each observation in the estimation sample, that is, betas are assumed to change (remain constant) across (within) sub-samples with probability one.

In this paper, we follow existing research and assume stochastic dynamics for the model parameters (see Hendershott and Menkveld, 2014; Pascual and Veredas,

[9] Rosu et al. (2018) demonstrate theoretically and empirically an association between order imbalance and trade-to-quote ratio.

[10] Hagströmer (2018) shows that a micro-price effective spread measure, which is closely related to the order book depth imbalance, can be a valuable source of fundamental information for market makers.

2010; Brogaard et al., 2014). In particular, we assume that the dynamics of the slope parameters follow a random walk, resembling the information acquisition process of market makers and informed traders who form beliefs based on the available information, i.e.

$$\boldsymbol{\beta}_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \boldsymbol{\Omega}), \tag{1.2}$$

where $\boldsymbol{\Omega}$ is a $(k+1) \times (k+1)$ full covariance matrix. Such random walk dynamics imply that the weights assigned to past information now decrease exponentially rather than being constant, depending on how much the predictors are informative about future order imbalance, i.e. the signal-to-noise ratio (see West and Harrison, 1997; Koopman and Durbin, 2012 for details). The missing parameters $\boldsymbol{\Omega}$ and σ^2 are estimated by maximum likelihood, whereas the dynamic betas β_t are assumed latent and are extracted through a standard Expectation Maximisation (EM) algorithm. In particular, we can obtain the forecast of the time t observation using data up to time $t-1$ as $y_{t|t-1}$ and the associated forecast variance. The model parameters are estimated by using the entire sample for each stock i.

Predictability Measure

Our objective is to forecast the one-day ahead order imbalance for stock i based on some conditioning information available at time $t-1$, $\boldsymbol{x}_{t-1}$. The predictor variables are the order imbalance of stock i on day $t-1$, $oib_{i,t-1}$, the individual stock return $ret_{i,t-1}$, the return on the S&P 500 index $spret_{t-1}$, the realised variance $rvar_{i,t-1}$, the trade-to-quote ratio $ttq_{i,t-1}$ and the depth imbalance $dib_{i,t-1}$.

The literature has entertained a number of out-of-sample forecast evaluation measures. For instance, Campbell and Thompson (2008) uses out-of-sample R^2 that is evaluating model forecast performance against historical mean forecasts in the context of return predictability. We choose a measure of out-of-sample forecast performance more related to the dynamic linear model. In particular, the model introduced above provides a forecast of the order imbalance $oib_{t|t-1}$ and the associated forecast variance $Var_t(oib_{t|t-1})$. The normality assumptions in Equation (1.1) ensure that the forecasted observations are distributed as $N\big(oib_{t|t-1}, Var_t(oib_{t|t-1})\big)$. Thus, we can obtain a measure of predictability that is akin to a predictive likelihood as

$$\tilde{pl}_t = p\big(oib_t \mid oib_{t|t-1}, \sigma^2(oib_{t|t-1})\big), \tag{1.3}$$

where $p(\mu, \sigma^2)$ is the density function of the Normal distribution with parameters μ and σ^2. This is the value of the normal probability density function evaluated at oib_t with mean $oib_{t|t-1}$ and variance $\sigma^2(oib_{t|t-1})$. One comment is in order. The predictive likelihood evaluated in (1.3) is not bounded from above such as for example empirical R^2. Notwithstanding, as we are interested in the degree of order imbalance predictability in relative terms (across stocks and time) rather than its absolute value, this does not pose any particular challenge for our analysis. Since a daily measure of predictability derived from the predictive likelihood is inherently noisy we opt to smooth the measure by taking its rolling 20-day average. Thus, henceforth pl_t is understood as the 20-day rolling average over $\tilde{pl}_t$ constructed from past observations from $t-19$ to t. We denote the predictability of total order imbalance in firm i on day t by pl_{it}^{tot} and the predictability of pension funds order imbalance in firm i on day t by pl_{it}^{pf}.

Summary Statistics

Table 1.1 reports the summary statistics for both samples: *Active pension funds* (Panel A) and *Full sample of stocks* (Panel B). For both samples we sort stocks based on their market capitalisation using the monthly NYSE breakpoints and present summary statistics for each size quintile.[11] We refer to the smallest value quintile, e.g. small-cap, as Quintile 1 and to the highest value quintile as Quintile 5.

Consistent with prior evidence, smaller stocks tend to be less liquid than large stocks. As a matter of fact, the average values of the four illiquidity variables are monotonically decreasing with size. Stocks traded by active pension funds seem to be more liquid, on average, than stocks in the aggregate sample. This is consistent with the idea that pension funds trade more actively in large stocks. This observation is also confirmed by the number of observations in the Q5 quintile. Both market efficiency measures indicate that large stocks' prices are more efficient than the small stock ones. While for active pension funds this evidence is weaker, average

[11] The monthly NYSE breakpoints are retrieved from Kenneth French's data library `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_me_breakpoints.html`. Accessed 14th May 2020.

price inefficiency monotonically decreases from Q1 to Q5 in the *Full sample of stocks*.

The average daily total order imbalance is -1.87% and monotonically increases with the firm size, i.e. from -6.55% for small stocks to 4.32% for large stocks. The order imbalances from the *Active pension funds* sample are smaller in magnitude (see Panel A). For both samples, the order imbalance is autocorrelated with the first order autocorrelation coefficients being about 15% for the order imbalance in the *Full sample of stocks* and 33% for the order imbalance of *Active pension funds*, respectively.

The *Active pension funds* sample average of the predictive likelihood is 36.62%. This means that the probability of observing the order imbalance value of oib_t^{pf} within an ϵ-neighbourhood of $oib_{t|t-1}$ is approximately $2\epsilon \times 0.3662$ for small $\epsilon > 0$. The sample average of predictive likelihood for the total order imbalance sample is 34.76%. On average, the order imbalances of large stocks appear to be more predictable than the order imbalances of small stocks. This result holds for both samples. The descriptive statistics of the additional predictors, such as trading volume, trade-to-quote ratio, depth imbalance, realised variance and stock returns are summarised in Table A.1 in the Appendix.

Figure 1.2 plots the cross-sectional mean and 95% quantiles of the predictive likelihoods as well as the time series average of the predictive likelihoods across five market capitalisation quintiles, both for the *Active pension funds* sample (Panel A) and the *Full sample of stocks* (Panel B).

Figure 1.2 documents a substantial variation of predictive likelihood in both the cross-section and time series. We observe that the median predictive likelihood (see Panel A of Figure 1.2) drops noticeably during the global financial crisis of 2008. The degree of predictability is the lowest for small cap stocks and is monotonically increasing with size. Panel B of Figure 1.2 also reveals that the slump in predictive likelihoods around the crisis is particularly pronounced for large cap stocks. Unsurprisingly, we find a trending behaviour in the time series of predictive likelihoods. This can originate from the convergence characteristics of the dynamic linear model. Through the accumulation of information the model estimates improve and converge to their long-run levels. Another reason for the existence of trends can come from the sample characteristics due to increasing number of institutional investors and high-frequency trading in the market. In the following analysis, we will address both of those issues.

Table 1.2 reports the sample pairwise correlations of the predictors that are used in the main empirical analysis (see above for a detailed description). With the only exception of the quoted and effective spreads, correlations tend to be low. The absence of contemporaneous correlations across predictors should rule out multicollinearity issues in the battery of panel regressions estimated in the main empirical analysis. There is a potential concern that the degree of order imbalance predictability is related to the direction of order imbalance. It it is worth stressing a low correlation between predictive likelihood of order imbalance and the level of order imbalance itself which mitigates this concern.

1.4 Empirical Results

In this section we test the hypotheses outlined in Section 1.2. Hypothesis 1 and 2 concern the effect of predictability of uninformed order imbalance on market quality. We test both hypotheses by using the order imbalance from active pension funds. Hypothesis 1b and 2b are based on the assumption that posits that the predictability of uninformed order imbalance is translated in the predictability of the total order imbalance. It is evident that this assumption does not hold automatically due to the fact that the uninformed order imbalance is only a part of the total order imbalance. Informed traders can endogenously adjust their trading activity in response to the signal about the uninformed orders which might reduce or completely eliminate the predictability of total order imbalance.

The theoretical model developed in the Appendix A.1 shows that, in equilibrium, there is a clear association between the levels of predictability of the two types of order imbalance. While there is a demand adjustment by informed traders, its extent is insufficient to completely eliminate the predictability of the total order imbalance. The empirical evidence supports the positive association between the predictability of uninformed and total order imbalance. In particular, we merge our two samples and estimate the following regression

$$pl_{it}^{tot} = \alpha_i + \omega_t + \beta pl_{it}^{pf} + u_{it}, \tag{1.4}$$

where α_i contains firm fixed effects, and u_{it} is an idiosyncratic error term. Table 1.3 presents the estimation results for a simple pooled OLS, i.e. without firm and time fixed effects, as well as two panel regressions with firm fixed effects as well as time fixed effects. The estimates of β show that regardless the regression specification there is a positive and statistically significant association between the predictability of total versus uninformed order imbalance. Interestingly, there is a substantial unobservable heterogeneity captured by firm-level fixed effects as suggested by a much higher R^2 that is obtained with respect to a simple pooled OLS regression.

1.4.1 Order Imbalance Predictability and Liquidity

Following, our hypotheses 1 and 1b we directly analyse the relationship between the predictability of order imbalance and market quality by using a variety of liquidity measures. Specifically, we estimate the following regression

$$illiq_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it}, \tag{1.5}$$

where $illiq_{it}$ corresponds to one of several illiquidity variables - $qspread_{it}$, $espread_{it}$, $rspread_{it}$ and $prcimpact_{it}$ - and $pl_{i,t-1}$ stands for either $pl^{pf}_{i,t-1}$ or $pl^{tot}_{i,t-1}$. The set of control variables $\boldsymbol{X}_{it}$ include the log market cap of the firm i ($lsize_{it}$), the inverse of the price ($1/prc_{it}$), the trade-to-quote ratio (ttq_{it}), the share turnover ($turn_{it}$), the value of the VIX index (vix_t), the lagged value of the dependent variable ($illiq_{i,t-1}$) and firm fixed effects dummies. We also control for contemporaneous order imbalance oib^{tot}_{it} to mitigate a potential concern that predictability of order imbalance is related to the level of order imbalance itself. While the regressions are predictive in nature (with respect to the predictability measure $pl_{i,t-1}$) we still control for a number of contemporaneous day t market variables to ensure that our results are not driven by other factors known to impact liquidity (see, e.g., Hendershott et al., 2011). Given that illiquidity levels are persistent on the daily timescale, we include lagged spread as a control to proxy for any omitted but potentially relevant variables. We report least-squares estimates of $\hat{\beta}$ and $\hat{\gamma}$ with the corresponding t-statistics (in brackets) calculated based on double-clustered standard errors.

Panel A of Table 1.4 reports the estimation results using the *Active pension funds* sample and pl^{pf} as the proxy for the predictability of the uninformed component of order imbalance. The empirical evidence supports a significant negative relation between the predictability of uninformed order imbalance and illiquidity. Such negative relationship holds across all of the liquidity measures used. In addition, the economic size of the effect is quite substantial; a one standard deviation increase in predictive likelihoods leads to an average decrease in the quoted spreads by 0.56 basis points and a decrease in the effective spreads by 0.27 basis points.

The coefficients for pl^{pf} are statistically significant at the 1% level. This rejects our Hypothesis 1 in favour of the alternative that an increase in the predictability of uninformed order imbalance improves market liquidity, on average. The estimation results based on the *Full sample of stocks* and pl^{tot} as the proxy for predictability of total order imbalance (see Panel B of Table 1.4) also rejects Hypothesis 1b. In fact, the effect of total order imbalance predictability on liquidity is an order of magnitude larger than the predictability of pension funds order imbalance. A one standard deviation increase in total order imbalance predictive likelihood leads to an average decrease in the quoted and effective spreads by about 3.00 bps, respectively. The coefficients for pl^{tot} are also statistically significant at 1% level. As far as the full sample of stocks is concerned, it is worth noting that the effect of predictability of order imbalance on the adverse selection component of the effective spread (*prcimpact*) is about three times higher than on the realised spreads. This emphasises the claim that predictability of order imbalance is closely related to the cost of adverse selection.

1.4.2 Order Imbalance Predictability and Market Efficiency

Next, we test the effect of predictability of order imbalance on market efficiency (Hypotheses 2 and 2b). To do so we estimate the following regression

$$ineff_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it}, \tag{1.6}$$

where $ineff_{it}$ corresponds to one of the two illiquidity variables $vratio_{it}$ and abs_ac_{it}, whereas $pl_{i,t-1}$ stands for either $pl^{pf}_{i,t-1}$ or $pl^{tot}_{i,t-1}$. Similar to Eq.(1.5), the set of control variables include the log market cap of the firm i ($lsize_{it}$), the inverse of the price ($1/prc_{it}$), the trade-to-quote ratio (ttq_{it}), the share turnover ($turn_{it}$), the value of the VIX index (vix_t). The parameters α_i capture a firm fixed effect. The t-statistics are presented in brackets and are calculated based on double-clustered standard errors.

Panel A of Table 1.5 provides the estimation results for the *Active pension funds* sample and $pl^{pf}_{i,t-1}$ as the independent variable. We find a significant negative relation between the predictability of order imbalance and the degree of market inefficiency. Such negative relationship holds for both measures of market inefficiency. That is, an increase by one standard deviation in the predictive likelihood corresponds to a decrease in the variance ratio by 0.20% (which is about 1.43% of its sample average value) and absolute autocorrelation of daily returns by 0.37% (which is about 1.02% of its sample average). Both of these effects are statistically significant at the 1% level. This results rejects Hypothesis 2 that a higher predictability of uninformed order imbalance does not improve market efficiency, on average.

We obtain similar results for the predictability of order imbalance from the *Full sample of stocks*, i.e. $pl^{tot}_{i,t-1}$. Indeed, the coefficient β is still negative and significant at the 1% level (see Panel B of Table 1.5). An increase by one standard deviation in predictive likelihood of total order imbalance decreases the variance ratio by 0.55% (which is about 3.44% of its sample average value) and the absolute autocorrelation of daily returns by 0.98% (which is about 2.45% of its sample average). This results rejects Hypothesis 2b that a higher predictability of order imbalance does not improve market efficiency, on average.

1.4.3 Order Imbalance Predictability, Market Conditions and Stock Characteristics

Next, we explore how the relation between order imbalance predictability and market quality depends on aggregate market conditions. We interact the predictive likelihoods with a time dummy corresponding to different levels of the VIX index, which is a widely used proxy for aggregate market uncertainty (see, e.g., Bloom, 2009). We define a dummy variable D^{high}_t that takes value one whenever vix_t falls within the upper tercile of its historical distribution drawn from our sample and zero otherwise. Similarly, dummy variables D^{med}_t and D^{low}_t take the value one when vix_t is in the middle and bottom terciles, respectively. Each of these dummy variables are then interacted with the predictive likelihood to study the differential impact of predictability on spreads during periods of high versus low market uncertainty.

Specifically, we estimate a panel regression of the following form:

$$y_{it} = \sum_{j \in \{\text{high,med,low}\}} \beta_j \cdot pl_{i,t-1} \cdot D_t^{(j)} + \alpha_i + \gamma' \boldsymbol{X}_{it} + \epsilon_{it}, \tag{1.7}$$

where y_{it} represents either one of the illiquidity measures $illiq_{it}$ or one of the proxies used for market inefficiency $ineff_{it}$, and α_i represents a firm fixed effect. Here, $pl_{i,t-1}$ stands for either $pl^{pf}_{i,t-1}$ or $pl^{tot}_{i,t-1}$. The set of control variables is the same as in the previous regressions. Standard errors are double clustered by firm and time.

Table 1.6 reports the estimation results. For the ease of exposition we present the results for a measure of illiquidity and a measure of inefficiency, namely the quoted spread $qspread_{it}$ and the variance ratio $vratio_{it}$. Panel A displays the estimation results for the *Active pension funds* sample. For all three levels of the VIX index we confirm the positive effect of the predictive likelihoods on liquidity and market efficiency. The marginal effect of the predictability of uninformed order imbalance on market liquidity tends to increase as aggregate market uncertainty increases. That is, bid-ask spreads decrease by a greater extent during periods of high investors' uncertainty. On the other hand, the effect of predictability of uninformed order imbalance on market efficiency tends to be higher during calmer times.

Panel B reports the results for the *Full sample of stocks*. The effect is stronger by an order of magnitude than for the sub-sample of pension funds. Nevertheless the monotonic relationship of the interaction terms is preserved. Again, the effect of predictability of total order imbalance on market liquidity is more pronounced as aggregate uncertainty increases. On the opposite, the effect of predictability on price efficiency decreases as the level of the VIX index increases. Table A.2 in the Appendix shows further results for additional measures of market liquidity such as the effective spread, $espread_{it}$, the realised spread, $rspread_{it}$, and price impact, $prcimpact_{it}$, as well as returns autocorrelation, abs_ac_{it}, as additional measure of market inefficiency. The results confirm the evidence provided in Table 1.6, namely the higher the market uncertainty the bigger (smaller) is the impact of order imbalance predictability on liquidity (inefficiency).

A potential concern is that the documented effects exist only for a specific group of stocks and is attributed to size and volatility of a firm due to limits to arbitrage, activity of institutional investors or high-frequency traders. We address this question by sorting the stocks in quintiles for a given characteristic and re-estimating Equations (1.5) and (1.6) separately for each quintile.

Table 1.7 reports the estimates of the slope parameter $\hat{\beta}$ on the measure of predictability. Panel A reports the results for the sample of stocks traded by *Active pension funds*, whereas Panel B reports the results for the *Full sample of stocks*. For the ease of exposition we report the results for quoted spreads as a measure of liquidity. A number of aspects emerge: when stocks are sorted by market cap, the coefficient on pl_{t-1}^{pf} (see Panel A) is negative and statistically significant for all five quintiles. The effect is strongest for small stocks and weakest for large stocks. A one standard deviation increase in predictive likelihood pl_{t-1}^{pf} leads to about 2.28% decline in quoted spreads of small stocks and about 1.63% decline in quoted spreads of large stocks relative to their sample averages. We obtain qualitatively similar results when stocks are sorted by trade-to-quote ratio, by idiosyncratic volatility and passive institutional ownership. The highest magnitude of coefficient on pl_{t-1}^{pf} is for stocks with the lowest trade-to-quote ratio, with the highest idiosyncratic volatility and with lowest passive institutional ownership. Our results are also robust to the choice of liquidity measure.

Panel B confirms that the coefficient on the predictability measure pl_{t-1}^{tot} is negative and statistically significant at 1% level for all specifications. The pattern of the estimate across stock characteristics remains the same as for Panel A. These results confirm that an increase in the predictability of order imbalance reduces the extent of adverse selection in the market and overall improves liquidity, conditional on a wide range of stock characteristics.

Table 1.8 confirms the relationship between the predictability of order imbalance and market efficiency that we showed at the aggregate level (see Table 1.5). Panel A reports the results for the sample of stocks traded by *Active pension funds*; in the majority of cases, price efficiency – measured by the variance ratio – substantially improves with an increase of order imbalance predictability for each quintile of stocks sorted by market cap, trade-to-quote ratio, idiosyncratic volatility and passive institutional ownership. The exceptions are mainly for stocks in the top and bottom market cap quintiles, stocks in top trade-to-quote ratio quintile, stocks with the highest idiosyncratic volatility and lowest passive institutional ownership. In these cases the coefficient on pl_{t-1}^{pf} is not statistically significant. Similar to Table 1.7, the results for the *Full sample of stocks* are stronger (see Panel B). The coefficient on pl_{t-1}^{tot} is negative and statistically significant at the 1% level for all characteristic quintiles, with the only exception of the largest stocks.

Table A.3 in the Appendix reports further results on alternative measures of liquidity and market efficiency. Panel A displays the results for the stocks traded by *Active pension funds*. Except few nuances, the effect of the predictability of order imbalance on liquidity tends to be almost monotonically decreasing with the market capitalisation, trade-to-quote ratio, and the amount of passive institutional ownership. Panel B confirms the results for the *Full sample of stocks*; the magnitude of the coefficients is higher and the monotonic behaviour of the beta on predictability for the liquidity measures is retained. As far as *abs_ac* is concerned, both Panel A and Panel B confirm the results on market efficiency outlined in Table 1.8.

1.4.4 Detrending the Predictive Likelihood

In addition to the main empirical analysis we investigate the robustness of the results to the concern of a spurious regression problem due to the trending behaviour of our predictability measure pl_t. To address this issue, we perform a linear detrend of the predictive likelihood and repeat the regression exercises using the detrended series. For each given stock, we define the detrended predictive likelihood pl_t^* as residuals from the regression of pl_t on the corresponding time index t, i.e. $pl_t = \alpha + \beta\, t + pl_t^*$.

Table A.4 in the Appendix contains the results for regressions of quoted spreads onto the detrended predictive likelihood. The coefficient on the detrended predictive likelihood remains negative and statistically significant. Moreover its magnitude increases in absolute value. This confirms the main results outlined above, both in relation to the effect of predictability of order imbalance on market quality and on efficiency. Notice the additional results for the detrended predictive measure hold for both the sample of stocks traded by *Active pension funds* (Panel A) and the *Full sample of stocks* available in our dataset (Panel B).

1.4.5 Asset Pricing Implications

In this section we test Hypothesis 3 which lays out the relation between the predictability of order imbalance and the cross-sectional variation of stock returns. In Sections 1.4.1 - 1.4.3 we have shown that the predictability of pension funds and total order imbalance has similar positive effects on market quality. For the investigation of asset pricing implications of the predictability of order imbalance, we concentrate on the *Full sample of stocks* in order to benefit from the wider cross-section of stocks and longer time period.

We start by presenting the portfolio characteristics across quintiles sorted by the predictability of total order imbalance. To do so, we first aggregate our predictability measure to monthly frequency by averaging the daily predictive likelihood within a given month. Table 1.9 reveals the sorted portfolio characteristics. For instance, stocks with the highest level of predictability in the order imbalance tend to be growth stocks, larger in terms of market capitalisation, and experience past returns which are both lower and less volatile. Although not exhaustive, this evidence may imply that our predictability measure simply captures systematic risk features such as size, value, and momentum. In addition, the fact that the order imbalance of less liquid stocks tends to be more predictable may suggests that our predictability may be subsumed by a liquidity risk factor. Hence, the following analysis will put an emphasis on disentangling the asset pricing effect of order imbalance predictability from the potentially confounding effects mentioned above.

1.4.6 Portfolio Analysis

We test the implications of our predictability measure to capture the cross-sectional variation of stock returns against a variety of competing explanations both based on a portfolio sorting procedure and a regression analysis. We start by presenting univariate portfolio sort results for the predictability of order imbalance. The dark-blue bars in Figure 1.3 show the average future returns for quintile portfolios sorted on the previous month's value of $pl_{i,t-1}^{tot}$. We also present the average return of the trading strategy that buys an equally weighted portfolio of stocks which experienced a low level of predictability of order imbalance during the past month and sells short the portfolio with high predictability instead (red bar, "Low-High").

The average returns are monotonically decreasing with respect to the predictability of order imbalance. Stocks with lowest order imbalance predictability experience a return of 2.14% per month, whereas stocks with highest predictability of order imbalance generate a lower 0.86% monthly return over the period 01:1997-12:2013. The second red bar from the right shows that our low-minus-high Predictability of Order Imbalance portfolio, which we denote by POI, generates statistically significant returns at a conventional 1% credibility level. Such average return is also economically large: the strategy yields on average 1.29% per month with the annualised Sharpe ratio of 1.00. In addition, the first red bar from the right shows that our strategy also generates a substantial risk-adjusted return calculated conditional on several sources of systematic risks, such as market, size, value, momentum and the liquidity factor introduced by Pastor and Stambaugh (2003).

Figure 1.4 plots the cumulative returns of our POI portfolio. The returns are pro-cyclical and experience positive performance for most of the sample except 1999 and 2008 crises. The value-weighted strategy generates lower returns as compared to its equally-weighted counterpart. This is consistent with the evidence that the bulk of the economic magnitude is within small stocks, which are mechanically underweighted in a value-weighted portfolio.

As discussed above, Table 1.9 reveals some monotonic relation of the predictability of order imbalance with other characteristics that are shown to be priced in the empirical finance literature, e.g. size and value. To investigate the pricing implications of POI, we perform a set of double sorts based on $pl_{i,t-1}^{tot}$ and other variables, that are size, bid-ask spread, turnover, trade-to-quote ratio, and volatility of order imbalance. Table 1.10 reports the results.

Panel A reports the returns of 5×5 portfolios sorted on past predictability of order imbalance and the lagged value of market capitalisation. The average return on the low-minus-high predictability portfolio, conditional on small stocks, is positive and statistically significant. Conditional on the smallest stocks, the portfolio yields 1.64% per month. For large stocks the return on the low minus high portfolio is still positive but not statistically significant. When controlling for market, size, value, momentum and liquidity factors, the Jensen's alphas are large and statistically significant at 1% level for all quintiles with the only exception of Q5, i.e. the largest stocks.

Panel B shows the results for a double-sort of 5×5 portfolios based on lagged predictability of order imbalance and past value of the bid-ask spread. A conditional strategy based on low-minus-high predictability and high bid-ask spreads generates the highest performance, with a 2.18% on a monthly basis and a risk-adjusted return, once controlling for the aforementioned risk factors, equal to 2.4%

per month. The first column shows that there is no significant premium captured by our *POI* portfolio for highly liquid stocks. The results for a double sort based on past order imbalance predictability and trade-to-quote ratios (Panel C) are significant across all portfolios. All low-minus-high portfolios generate substantial risk-adjusted returns in the range 1.92% for Q1 to 3.00% for Q5.

Panel D and E show the performance of our strategy conditional on different levels of volatility of order imbalance and share turnover, respectively. Our strategy tends to perform better for stocks with high levels of order imbalance volatility and high turnover. As a matter of fact, when order imbalance volatility is low, i.e. in Q1, our low-minus-high predictability strategy does not generate statistically significant returns. On the opposite, for highly volatile stocks our *POI* strategy generates up to 1.84% risk-adjusted returns on a monthly basis.

Finally, Panel F reports the returns of portfolios sorted on past order imbalance predictability and the order imbalance. The average return on the low-minus-high predictability portfolio, conditional on low order imbalance stocks, is 1.398% per month. For high order imbalance stocks the average return on the high-low portfolio drops to 1.036% per month.

Overall, the results suggest that the predictability of order imbalance proxies some aggregate and non-diversifiable source of risk that is priced in the cross-section of stock returns. Yet, this pricing effect is robust when controlling for other relevant characteristics which are commonly assumed to proxy for systematic risks.

1.4.7 Cross-Sectional Regressions

If low predictability of order imbalance poses a risk to marginal investors due to increased asymmetric information, we should observe a risk-return trade-off. To investigate this we construct 25 portfolios based on size and the predictability of order imbalance, and test the exposure of these test assets to a set of commonly used systematic risk factors as well as the returns on our *POI*.[12] More specifically, we first estimate the following time series regression,

$$ret_{i,t} = \alpha_i + \beta_i POI_t + \gamma_i' \boldsymbol{F}_t + u_{i,t}, \qquad i = 1, \ldots, 25 \tag{1.8}$$

[12]We sort by predictability of order imbalance to ensure that the *POI* betas have high variation across portfolios and we choose size as the second sorting characteristics because it is a hard benchmark to beat given results in Table 1.10.

where $\boldsymbol{F}_t$ contains the excess returns on the market, the size and value factors of Fama and French (1993a), the momentum factor calculated as in Jegadeesh and Titman (1993), the liquidity risk factor from Pastor and Stambaugh (2003), as well as the short-term reversals factor and the returns on high-minus low portfolios sorted by past month trade-to-quote ratio and both the level and volatility of order imbalance.

Table 1.11 reports the estimates of the slope parameter $\hat{\beta}_i$ for each of the 25 test portfolios. Notice that the nature of the regression (1.8) implies that higher estimates $\hat{\beta}_i$ are associated with higher expected returns controlling for other risk factors. Panel A shows the results for equally-weighted test assets. Conditional on the level of predictability, the exposure to POI tends to decrease with the size of a firm. Similarly, conditional on size, the exposure of stock returns to the POI returns decrease with the level of predictability. That is, small size - low predictability stocks tend to be the most correlated to the POI factor, whereas the opposite is true for the large size-high predictability stocks.

Panel B shows the results for value-weighted test assets. The results are largely similar to the equally-weighted portfolio both in terms of sign, magnitude and significance of the slope parameter estimates. As a whole, the results for the value-weighted test portfolios reinforce the evidence of Panel A whereby large size-high predictability stocks tend to have the most negative correlation with the POI factor, while small size-low predictability stocks tend to have the highest correlation with the POI factor.

Although instructive, the results reported in Table 1.11 do not necessarily imply that the risk associated with the time series variation of our POI factor is priced in the cross-section of stock returns. To address this issue, and further investigate the asset pricing implications of the predictability of order imbalance, we perform a set of standard Fama-MacBeth regressions. This allows to estimate the equilibrium price of risk implied by order imbalance predictability while controlling for alternative, and possibly competing, sources of systematic risk (see Fama and MacBeth, 1973). Specifically, we estimate the following cross-sectional regression

$$returns_{i,t} = \lambda_{POI,t}\hat{\beta}_i + \lambda_t'\hat{\gamma}_i + e_{i,t}, \qquad t = 1, \ldots, T \tag{1.9}$$

where the parameters $\hat{\beta}_i, \hat{\gamma}_i$ are estimated for each test portfolio by using the time series regression in Eq.(1.8). The estimates $\hat{\lambda}_{POI}, \hat{\lambda}'$ are obtained as the sample averages of $\hat{\lambda}_{POI,t}, \hat{\lambda}'_t$ (see Ch.12 Cochrane, 2009).[13] As additional control variables, we use the returns on a set of alternative strategies such as a buy-and-hold exposure on the market, long-short portfolios based on size, value, momentum, liquidity as well as short-term reversals. In addition, we include the returns on two alternative high-minus-low portfolio strategies constructed based on the past month trade-to-quote ratio and the volatility of order imbalance (see Chordia et al., 2018).[14] Panel A of Table 1.12 reports the estimation results for 5×5 portfolios sorted on past predictability of order imbalance and the lagged value of market capitalisation, and each portfolio is equally weighted. The baseline specification is a four-factor Carhart model with the addition of our *POI* factor. Then each of the competing risk factors is added one by one to investigate the marginal effect of *POI* conditioning for alternative risk factors which have been investigated in the market microstructure literature. Last column shows an implementation in which all factors have been included.

Two findings emerge: first, the price of risk for our *POI* factor is positive and significant at the 1% confidence level for all regression specifications. This means that the risk associated with the time series variation of our *POI* factor is priced in the cross-section of stock returns. Given that *POI* is a tradable factor, the price of risk estimated from the Fama-MacBeth regressions is statistically indistinguishable from the average return on the *POI* portfolio documented earlier. Second, the prices of risk for both the size factor and a long-short portfolio based on the volatility of order imbalance – which is labelled *VOIB* in the table – remain positive and highly significant across model specifications. The results for *VOIB* are consistent with recent evidence provided by Chordia et al. (2018).

Panel B reports the results whereby the 5×5 test portfolios are still sorted on past predictability of order imbalance and the lagged value of market capitalisation, but now are value-weighted. The value weighting scheme naturally dilutes the effect of small stocks in the time series variation of the test portfolios. Despite the different

[13] To address the time series correlation of the lambda estimates, the standard errors are calculated based on a Newey-West estimator with 12 lags (see Newey and West, 1987a).

[14] Notice the two additional factors constructed based on the past month trade-to-quote ratio and the volatility of order imbalance are built by sorting stocks into the corresponding quintiles on our sample.

weighting scheme, the sign, magnitude and significance of the price of risk for the *POI* factor remain intact. Again, both the *SMB* and the *VOIB* risk proxies turn out to be significant and positive across model specifications.

As a whole, Table 1.12 provides evidence that the risk associated with our *POI* factor is neither subsumed nor substantially affected by other competing (or alternative) risk factors, such as liquidity, short-term reversals, size and value. This result, coupled with the portfolio evidence in Table 1.10, suggests that we can reject Hypothesis 3 and therefore provide significant evidence that the predictability of order imbalance affects the equity cost of capital.

1.5 Conclusion

In this paper, we provide evidence that the predictability of order imbalance is an important determinant of market quality. We take the perspective of a sophisticated market participant who forecasts order imbalances in a dynamic linear model. We show that there is a significant and positive relationship between the predictability of order imbalance and both measures of market liquidity and market efficiency. This positive relationship is robust to controlling for periods of high and low market uncertainty and across different groups of stocks sorted by size, idiosyncratic volatility, trade-to-quote ratio and institutional ownership.

These findings are consistent with the notion that increases in predictability of order imbalance reduce adverse selection costs and inventory risks. As a result, order imbalance predictability significantly affects stocks prices. We document that a strategy that buys stocks with low order imbalance predictability and sells short stocks with high order imbalance predictability produces significant and economically large returns. Theses returns are not explained by a set of traditional risk factors as well as other characteristics that capture different aspects of adverse selection costs. The exposure of stock returns to the returns of this strategy commands a systematic risk premium.

Our results document that features of market microstructure that potentially influence predictability of order imbalance can affect firms' equilibrium return. This suggests that a firm's cost of capital is influenced not only by information, but also by market anonymity, organization of exchanges, and proliferation of high-frequency trading.

Figures and Tables

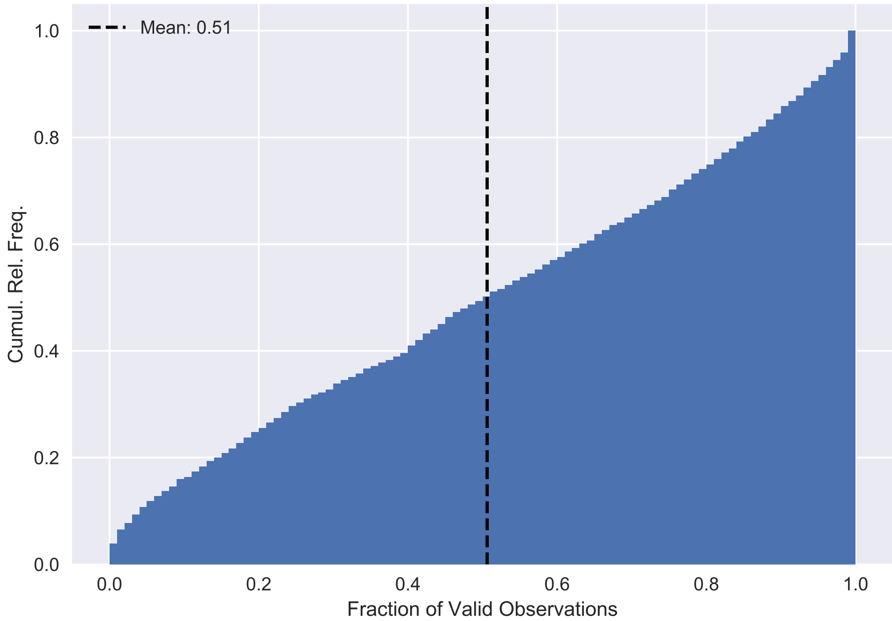
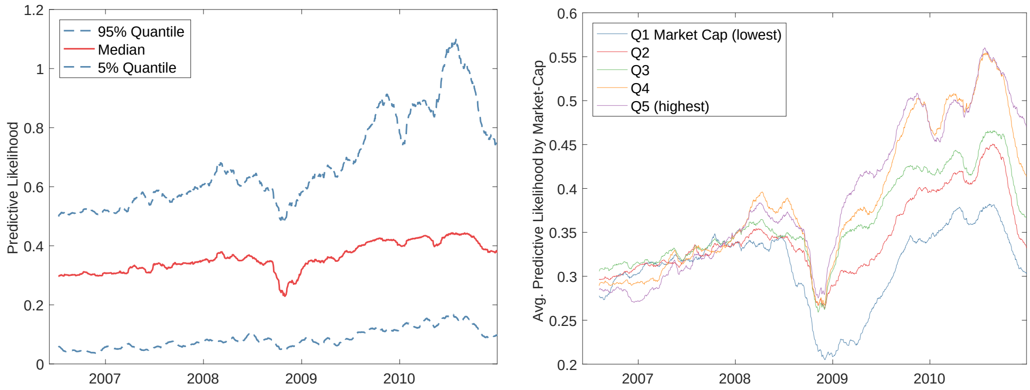


Figure 1.1: Number of Days with a Trade by a Pension Fund.
The figure presents a histogram for the number of observations with a trade from a pension fund in the ANcerno dataset. It shows a fraction of firms in the original dataset (vertical axis) that has a given proportion of trading days (horizontal axis) with a trade in the stock from a pension fund. The sample is from January 1, 2006 to December 31, 2010 (1,144 trading days in total) and 1,560 firms.

Panel A: Predictive Likelihood of Order Imbalance from Pension Funds



Panel B: Predictive Likelihood of Total Order imbalance

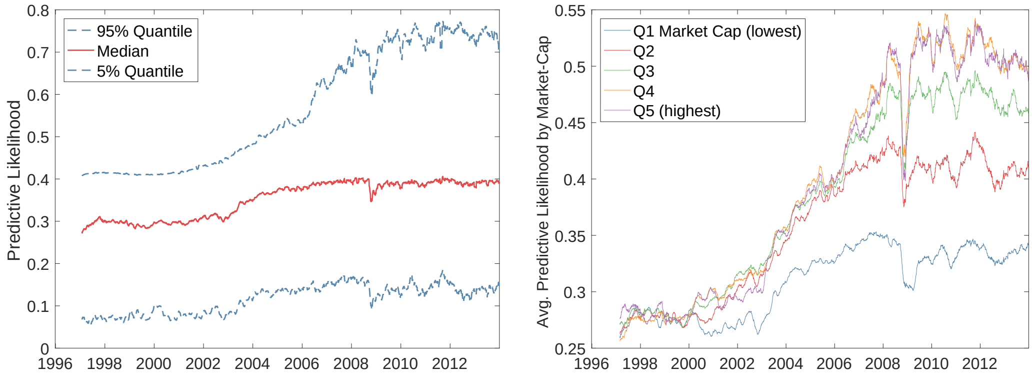


Figure 1.2: Predictive Likelihoods.
This figure shows the daily estimates of the smoothed predictive likelihood as outlined in Section (1.3.1). Top panel shows the results for the order imbalance of those stocks traded by active pension funds, i.e. the ANcerno sample, whereas the bottom panel shows the results for the total order imbalance. The left column of both panels report the 5%, 50% (median) and 95% quintiles across the daily estimates at each time t. The right column shows the average predictive likelihood for stocks sorted into quintiles based on the market capitalisation at each time t. The Ancerno sample is from January 1, 2006 to December 31, 2010 (**Panel A**) and the full sample of stocks is from January 1, 1997 to December 31, 2013 (**Panel B**).

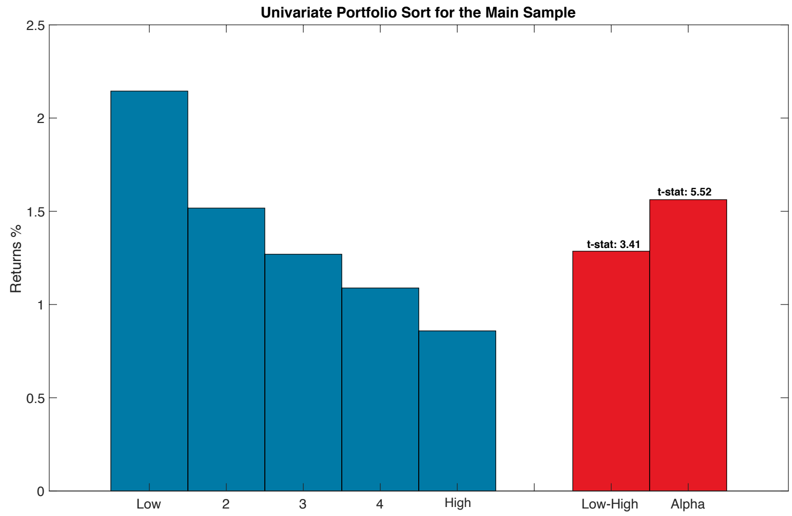


Figure 1.3: Average Returns on Portfolios of Stocks Sorted on pl_t^{tot}.
This figure shows the average returns on portfolios of stocks sorted based on the lagged predictive likelihood of total order imbalance pl_t^{tot}. Every month we sort stocks into quintiles, form five portfolios and rebalance next month. Portfolio 1 corresponds to the quintile of stocks with the lowest predictability of order imbalance pl_t^{tot} and portfolio 5 is for stocks with the highest predictability of order imbalance. The sample period is from January 1, 1997 to December 31, 2013. The red bars report the returns on the low-minus-high portfolio and the Jensen's alpha obtained once controlling for the three Fama-French factors, momentum and liquidity as in Pastor and Stambaugh (2003).

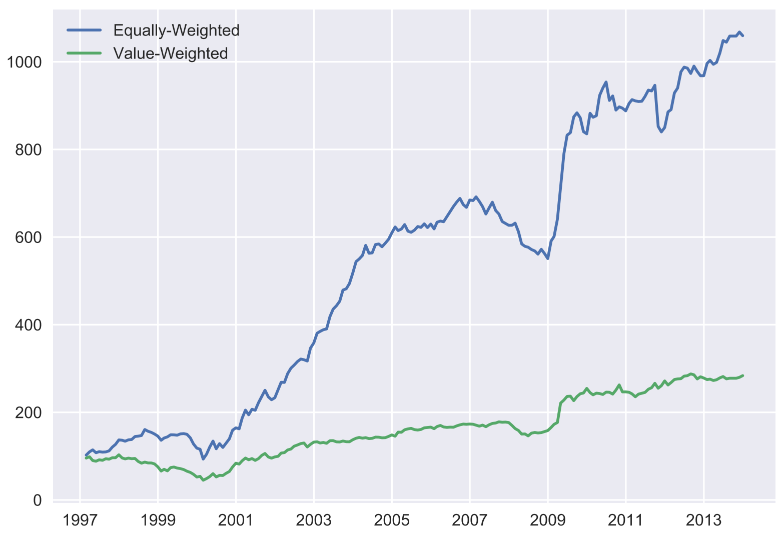


Figure 1.4: Cumulative Returns on the *POI* Portfolio.
This figure presents time series of cumulative returns on equally and value weighted portfolios of stocks sorted based on the lagged predictive likelihood of total order imbalance pl_t^{tot}. Every month we sort stocks into quintiles, form five portfolios and rebalance next month. The returns are generated long-short strategy which goes long on the portfolio of stocks with the lowest level of predictability and short the portfolio of stocks with the highest level of predictability. The blue line shows the cumulative returns when portfolios are equally weighted whereas the green line shows the cumulative returns when portfolios are value weighted. The sample period is from January 1, 1997 to December 31, 2013.

Table 1.1: Summary Statistics.

This table presents means and standard deviations (in squared brackets) for the variables used in this study. We report the summary statistics both for the full sample and across different quintiles of market capitalisation. **Panel A:** summarises the variables for the sample of active pension funds for the period from January 1, 2006 to December 31, 2010. **Panel B:** reports the summary statistics for the full sample of stocks for the period from January 1, 1997 to December 31, 2013. The labels *espread*, *qspread* and *rspread* denote effective, quoted and realized spreads respectively. *prcimpact* is the 5 minutes price impact, *vratio* is the absolute value of the variance ratio minus one, $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns, oib^{pf} is the order imbalance of pension funds, and oib^{pf} autocorr is the first order autocorrelation in the pension funds order imbalance. oib^{tot} and oib^{tot} autocorr are total order imbalance and its autocorrelation coefficient. The variables pl^{pf} and pl^{tot} stand for the cross-sectional average of the predictive likelihoods of pension and total order imbalance, respectively. The predictive likelihoods pl^{pf}_{it} is defined as the 20-day rolling average over $\tilde{pl^{pf}}_{it} = \phi(oib^{pf}_{i,t} \mid oib^{pf}_{i,t|t-1}, \sigma^2(oib^{pf}_{i,t|t-1}))$, where $\phi(\mu, \sigma^2)$ denotes the Normal distribution density function. Here $oib^{pf}_{i,t|t-1}$ denote the out-of-sample forecast for stock i on day t, which is traded by an active pension fund, using data up to day $t-1$ and $\sigma^2(oib^{pf}_{i,t|t-1})$ is its forecast variance. pl^{tot}_{it} is defined analogously using oib^{tot}_{it} instead of oib^{pf}_{it}.

Panel A: Active Pension Funds

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
espread (*bp*)	8.37 [4.36]	21.96 [9.72]	11.65 [4.28]	7.82 [2.87]	5.92 [2.22]	4.04 [1.68]
qspread (*bp*)	13.91 [7.63]	39.85 [18.56]	21.36 [8.93]	13.11 [5.19]	8.64 [3.36]	4.93 [1.93]
rspread (*bp*)	2.61 [3.87]	7.55 [9.93]	3.52 [4.98]	2.34 [3.28]	1.80 [2.45]	1.31 [1.81]
prcimpact (*bp*)	5.64 [4.55]	14.24 [10.62]	8.02 [5.75]	5.37 [3.70]	4.01 [2.87]	2.63 [1.79]
vratio (%)	13.70 [9.18]	15.19 [9.69]	13.76 [9.69]	13.42 [9.39]	13.57 [9.33]	13.61 [9.44]
abs_ac (%)	35.39 [24.01]	38.70 [25.29]	35.57 [25.29]	34.59 [24.73]	35.13 [24.70]	35.32 [25.47]
oib^{pf} (%)	0.01 [6.84]	0.01 [11.12]	0.10 [7.90]	0.07 [6.04]	-0.10 [4.79]	-0.05 [3.70]
oib^{pf} acorr (%)	32.56 [12.71]	31.90 [34.17]	32.71 [22.29]	26.75 [23.80]	23.28 [23.46]	20.69 [21.06]
pl^{pf} (%)	36.62 [9.79]	31.32 [6.09]	34.41 [6.70]	36.26 [6.97]	38.58 [8.55]	38.78 [10.05]
Nr. obs.	780,605	59,141	161,021	201,805	174,132	184,506
Nr. firms	789					

Panel B: Full Sample of Stocks

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
espread (*bp*)	77.11 [82.47]	140.4 [112.1]	38.74 [28.08]	23.62 [16.53]	15.82 [11.57]	9.10 [7.37]
qspread (*bp*)	109.4 [101.6]	199.5 [137.9]	55.52 [33.71]	33.22 [18.91]	21.85 [12.31]	11.86 [7.46]
rspread (*bp*)	45.75 [328.9]	89.42 [392.9]	16.11 [238.3]	8.03 [154.8]	5.24 [97.21]	3.62 [42.18]
prcimpact (*bp*)	32.05 [233.5]	52.29 [280.1]	22.49 [170.0]	15.94 [113.7]	10.78 [71.31]	5.89 [26.46]
vratio (%)	15.89 [11.86]	17.88 [12.09]	15.06 [10.48]	14.19 [9.88]	13.68 [9.74]	13.40 [9.86]
abs_ac (%)	39.58 [30.01]	42.98 [30.64]	38.29 [27.00]	36.61 [25.68]	35.67 [25.18]	35.28 [25.33]
oib^{tot} (%)	-1.87 [34.83]	-6.56 [41.02]	-0.24 [26.58]	2.17 [21.61]	3.54 [18.63]	4.34 [14.76]
oib^{tot} acorr (%)	15.41 [6.73]	13.71 [9.75]	11.22 [15.44]	11.81 [15.95]	13.17 [16.76]	16.76 [17.22]
pl^{tot} (%)	34.73 [11.40]	31.15 [10.46]	35.46 [10.12]	37.87 [10.63]	39.55 [11.35]	39.54 [12.34]
Nr. obs.	13,029,155	5,917,088	2,478,330	1,772,525	1,510,898	1,350,314
Nr. firms	6,122					

Table 1.2: Pairwise Correlations of the Variables of Interest.
This table presents the pairwise sample correlation for each predictors used in the main empirical analysis. We report the correlations both for the full sample and across different quintiles of market capitalisation. **Panel A:** summarises the variables for the sample of active pension funds for the period from January 1, 2006 to December 31, 2010. **Panel B:** reports the summary statistics for the full sample of stocks for the period from January 1, 1997 to December 31, 2013. The labels *espread*, *qspread* and *rspread* denote effective, quoted and realised spreads respectively. *prcimpact* is the 5 minutes price impact, *vratio* is the absolute value of the variance ratio minus one, $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns, oib^{tot} is total order imbalance, ttq_{it} is the trade-to-quote ratio, $rvar_t$ the daily realised variance, ret_t the realised returns, and vol_t the daily realised volume. The variables pl^{pf} and pl^{tot} stand for the cross-sectional average of the predictive likelihoods of pension and total order imbalance, respectively. The predictive likelihoods pl^{pf}_{it} is defined as the 20-day rolling average over $\tilde{pl^{pf}}_{it} = \phi\big(oib^{pf}_{i,t} \mid oib^{pf}_{i,t|t-1}, \sigma^2(oib^{pf}_{i,t|t-1})\big)$, where $\phi(\mu, \sigma^2)$ denotes the Normal distribution density function. Here $oib^{pf}_{i,t|t-1}$ denote the out-of-sample forecast for stock i on day t, which is traded by an active pension fund, using data up to day $t-1$ and $\sigma^2(oib^{pf}_{i,t|t-1})$ is the associated forecast variance. pl^{tot}_{it} is defined analogously using oib^{tot}_{it} instead of oib^{pf}_{it}.

Panel A: Active Pension Funds

	espread	oib^{tot}	*prcimpact*	*rspread*	pl^{pf}	*qspread*	ret_t	*ttq*	vol_t	*abs_ac*	*vratio*	$rvar_t$
espread	1.00	-0.05	0.36	0.37	-0.37	0.84	-0.02	0.21	-0.13	0.03	0.05	0.23
oib^{tot}	-0.05	1.00	0.01	-0.04	0.01	-0.04	0.25	-0.01	0.01	-0.01	-0.01	0.01
prcimpact	0.36	0.01	1.00	-0.51	-0.14	0.31	0.00	0.07	-0.02	0.00	0.01	0.14
rspread	0.37	-0.04	-0.51	1.00	-0.15	0.31	-0.01	0.09	-0.07	0.02	0.02	0.05
pl^{pf}	-0.37	0.01	-0.14	-0.15	1.00	-0.38	-0.01	-0.24	0.18	-0.05	-0.07	-0.04
qspread	0.84	-0.04	0.31	0.31	-0.38	1.00	-0.01	0.17	-0.17	0.03	0.05	0.27
ret_t	-0.02	0.25	0.00	-0.01	-0.01	-0.01	1.00	0.02	0.07	-0.01	0.01	0.01
ttq	0.21	-0.01	0.07	0.09	-0.24	0.17	0.02	1.00	0.10	0.00	0.02	-0.04
vol_t	-0.13	0.01	-0.02	-0.07	0.18	-0.17	0.07	0.10	1.00	-0.04	0.00	0.10
abs_ac	0.03	-0.01	0.00	0.02	-0.05	0.03	-0.01	0.00	-0.04	1.00	0.41	0.00
vratio	0.05	-0.01	0.01	0.02	-0.07	0.05	0.01	0.02	0.00	0.41	1.00	0.02
$rvar_t$	0.23	0.01	0.14	0.05	-0.04	0.27	0.01	-0.04	0.10	0.00	0.02	1.00

Panel B: Full Sample of Stocks

	espread	oib^{tot}	*prcimpact*	*rspread*	pl^{tot}	*qspread*	ret_t	*ttq*	vol_t	*abs_ac*	*vratio*	$rvar_t$
espread	1.00	-0.04	0.56	0.35	-0.21	0.78	-0.03	0.01	0.16	0.01	0.02	0.41
oib^{tot}	-0.04	1.00	-0.01	-0.03	-0.05	-0.05	0.14	0.04	-0.05	-0.01	-0.01	-0.02
prcimpact	0.56	-0.01	1.00	-0.53	-0.11	0.51	0.00	0.01	0.14	-0.01	0.01	0.29
rspread	0.35	-0.03	-0.53	1.00	-0.09	0.21	-0.03	0.01	0.00	0.02	0.01	0.11
pl^{tot}	-0.21	-0.05	-0.11	-0.09	1.00	-0.19	0.00	-0.28	0.00	-0.01	-0.02	-0.12
qspread	0.78	-0.05	0.51	0.21	-0.19	1.00	-0.03	-0.07	0.07	0.01	0.03	0.46
ret_t	-0.03	0.14	0.00	-0.03	0.00	-0.03	1.00	0.03	0.04	-0.01	0.01	-0.01
ttq	0.01	0.04	0.01	0.01	-0.28	-0.07	0.03	1.00	0.40	-0.01	0.01	0.04
vol_t	0.16	-0.05	0.14	0.00	0.00	0.07	0.04	0.40	1.00	-0.02	0.04	0.20
abs_ac	0.01	-0.01	-0.01	0.02	-0.01	0.01	-0.01	-0.01	-0.02	1.00	0.39	0.00
vratio	0.02	-0.01	0.01	0.01	-0.02	0.03	0.01	0.01	0.04	0.39	1.00	0.03
$rvar_t$	0.41	-0.02	0.29	0.11	-0.12	0.46	-0.01	0.04	0.20	0.00	0.03	1.00

Table 1.3: Relation of Total & Uninformed Order Imbalance Predictability.
This table reports the estimation results from the following regression

$$pl_{it}^{tot} = \alpha_i + \omega_t + \beta pl_{it}^{pf} + u_{it},$$

where pl_{it}^{tot} and pl_{it}^{pf} are predictive likelihoods of total and pension funds order imbalance respectively, and α_i, ω_t contain firm and time fixed effects, respectively. The t-statistics are given in parentheses and are based on double-clustered standard errors. The sample period is January 1, 2006 to December 31, 2010.

	pl_t^{tot}	pl_t^{tot}	pl_t^{tot}	pl_t^{tot}
pl_t^{pf}	0.196 (9.67)	0.219 (18.5)	0.150 (6.13)	0.1339 (11.431)
const	0.398 (48.3)			
fixed effects	none	firm	time	firm, time
R^2	3.59%	29.30%	20.71%	18.62%
Nr. obs.	710,633	710,633	710,633	710,633

Table 1.4: Predictability and Market Liquidity.
This table reports the results from the regression

$$illiq_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where $illiq_{it}$ takes one of following illiquidity variables (expressed in basis points): quoted spread *qspread*, effective spread *espread*, realised spread *rspread*, and price impact *prcimpact*. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous total order imbalance ($oib^{tot}_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	qspread	*espread*	*rspread*	*prcimpact*
pl_{t-1}	-0.057 (-11.50)	-0.028 (-7.41)	-0.022 (-11.33)	-0.023 (-9.91)
$illiq_{t-1}$	0.574 (23.70)	0.556 (7.62)	0.210 (8.18)	0.271 (11.73)
oib^{tot}_t	0.001 (0.93)	-0.001 (-0.61)	-0.005 (-3.81)	0.006 (5.36)
$lsize$	-0.017 (-6.01)	-0.011 (-4.68)	0.003 (1.32)	-0.020 (-10.89)
$turn$	-0.069 (-2.20)	0.106 (4.19)	-0.057 (-2.42)	0.162 (4.53)
$1/prc$	0.229 (4.32)	0.242 (3.86)	0.232 (6.53)	0.169 (4.19)
ttq	-14.721 (-11.2)	-3.702 (-5.06)	-2.013 (-3.46)	-1.114 (-1.64)
vix	0.094 (12.02)	0.039 (5.96)	-0.007 (-2.98)	0.072 (22.33)
R^2	54.96%	65.71%	16.67%	32.37%
No. obs.	764,825	764,825	764,825	764,825

Panel B: Full Sample of Stocks

	qspread	*espread*	*rspread*	*prcimpact*
pl_{t-1}	-0.263 (-11.73)	-0.263 (-11.72)	-0.163 (-3.56)	-0.485 (-17.11)
$illiq_{t-1}$	0.688 (171.41)	0.579 (125.56)	0.047 (22.90)	0.036 (21.37)
oib^{tot}_t	-0.052 (-22.12)	-0.073 (-25.77)	-0.086 (-5.99)	-0.006 (-0.61)
$lsize$	-0.168 (-31.93)	-0.159 (-31.44)	-0.262 (-24.97)	-0.097 (-19.52)
$turn$	-1.542 (-9.87)	-1.182 (-9.75)	-2.474 (-9.01)	-0.014 (-0.24)
$1/prc$	0.559 (21.23)	0.611 (22.97)	1.174 (20.71)	0.371 (14.65)
ttq	1.722 (5.67)	7.342 (18.94)	16.460 (9.51)	1.433 (1.24)
vix	0.543 (17.49)	0.485 (19.34)	0.464 (11.61)	0.649 (28.07)
R^2	59.84%	49.50%	1.04%	0.42%
No. obs.	12,906,715	12,906,715	12,906,715	12,906,715

Table 1.5: Predictability and Market Efficiency.
This table reports the results from the regression

$$ineff_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where $ineff_{it}$ corresponds to one of the two market efficiency variables $vratio_{it}$ and abs_ac_{it}. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the contemporaneous order imbalance ($oib^{tot}_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	$vratio$	abs_ac
pl_{t-1}	-0.020 (-3.92)	-0.037 (-3.08)
oib^{tot}_t	-0.002 (-0.97)	-0.002 (-0.47)
$lsize$	0.009 (3.80)	0.010 (2.17)
$turn$	0.245 (7.02)	-0.456 (-6.38)
$1/prc$	0.066 (3.88)	0.133 (5.36)
ttq	-0.410 (-0.40)	-4.222 (-1.76)
vix	0.024 (3.28)	0.040 (2.13)
R^2	0.32%	0.18%
No. obs.	764,825	764,825

Panel B: Full Sample of Stocks

	$vratio$	abs_ac
pl_{t-1}	-0.048 (-14.4)	-0.086 (-11.4)
oib^{tot}_t	-0.002 (-8.60)	-0.006 (-11.6)
$lsize$	-0.008 (-18.9)	-0.014 (-16.2)
$turn$	0.054 (6.45)	-0.440 (-8.35)
$1/prc$	0.013 (8.48)	0.028 (9.98)
ttq	-0.032 (-0.89)	-0.095 (-1.25)
vix	0.035 (7.33)	0.050 (3.83)
R^2	0.85%	0.49%
No. obs.	12,906,715	12,906,715

Table 1.6: Market Uncertainty and Effect of Order Imbalance Predictability. This table reports the results of the following regression

$$y_{it} = \sum_{j\in\{\text{high,med,low}\}} \beta_j \cdot pl_{i,t-1} \cdot D_t^{(j)} + \alpha_i + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where y_{it} represents either one of the illiquidity measures $illiq_{it}$ or one of the proxies used for market inefficiency $ineff_{it}$, and α_i represents a firm fixed effect. For the ease of exposition we present the results for a measure of illiquidity and a measure of inefficiency, namely the quoted spread $qspread_{it}$ and the variance ratio $vratio_{it}$. We define a dummy variable D_t^{high} that takes value one whenever vix_t falls within the upper tercile of its historical distribution drawn from our 1997-2013 sample and zero otherwise. Similarly, dummy variables D_t^{med} and D_t^{low} take the value one when vix_t is in the middle and bottom terciles, respectively. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($oib_{i,t}^{tot}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	qspread	*vratio*
$pl_{t-1}^{pf} \times D_t^{high}$	-0.065 (-10.2)	-0.018 (-2.72)
$pl_{t-1}^{pf} \times D_t^{med}$	-0.054 (-11.2)	-0.022 (-4.26)
$pl_{t-1}^{pf} \times D_t^{low}$	-0.050 (-7.59)	-0.032 (-3.72)
oib_t^{tot}	0.001 (0.79)	-0.001 (-0.53)
$illiq_{t-1}$	0.573 (23.8)	
Controls	Yes	Yes
No. obs.	764,825	764,825
R^2	54.97%	0.33%

Panel B: Full Sample of Stocks

	qspread	*vratio*
$pl_{t-1}^{pf} \times D_t^{high}$	-0.427 (-15.3)	-0.032 (-7.11)
$pl_{t-1}^{pf} \times D_t^{med}$	-0.223 (-9.79)	-0.052 (-16.1)
$pl_{t-1}^{pf} \times D_t^{low}$	-0.179 (-6.74)	-0.053 (-14.6)
oib_t^{tot}	-0.052 (-21.9)	-0.002 (-8.79)
$illiq_{t-1}$	0.680 (171.3)	
Controls	Yes	Yes
No. obs.	12,906,715	12,906,715
R^2	59.85%	0.88%

Table 1.7: Effect of Order Imbalance Predictability on Liquidity by Stock Characteristics.
This table presents the results of the following regression

$$qspread_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where $qspread_{it}$ is the quoted bid-ask spread and pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). We report the estimates $\hat{\beta}$ by sorting stocks in quintiles based on the following characteristics: market cap, trade-to-quote ratio, idiosyncratic volatility, and passive institutional ownership. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote (TTQ) ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($oib^{tot}_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors. Results for analogous regression using the alternative liquidity measures *espread*, *rspread* and *prcimpact* are reported in Table A.3.

Panel A: Active Pension Funds

Sorting variable	Q1	Q2	Q3	Q4	Q5
Market cap	-0.149 (-3.98)	-0.069 (-7.39)	-0.047 (-9.81)	-0.023 (-9.51)	-0.008 (-6.48)
TTQ ratio	-0.072 (-7.20)	-0.055 (-7.38)	-0.046 (-9.80)	-0.047 (-11.4)	-0.051 (-5.47)
Idiosync. vol.	-0.022 (-8.57)	-0.041 (-9.95)	-0.063 (-9.94)	-0.073 (-10.0)	-0.079 (-6.26)
Passive ownersh.	-0.090 (-7.28)	-0.044 (-7.83)	-0.048 (-6.95)	-0.039 (-8.86)	-0.046 (-7.53)

Panel B: Full Sample of Stocks

Sorting variable	Q1	Q2	Q3	Q4	Q5
Market cap	-1.044 (-13.6)	-0.354 (-13.9)	-0.119 (-12.8)	-0.084 (-7.34)	-0.028 (-14.3)
TTQ ratio	-0.702 (-14.6)	-0.383 (-14.6)	-0.354 (-14.3)	-0.406 (-13.7)	-0.417 (-8.41)
Idiosync. vol.	-0.201 (-18.1)	-0.300 (-18.1)	-0.429 (-15.3)	-0.712 (-16.1)	-1.563 (-16.0)
Passive ownersh.	-1.069 (-8.93)	-0.575 (-11.0)	-0.289 (-11.3)	-0.257 (-13.7)	-0.273 (-12.9)

Table 1.8: Effect of Order Imbalance Predictability on Efficiency by Stock Characteristics.
This table presents the results of the regression

$$vratio_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where the variance ratio $vratio_{it}$ is calculated as shown in Section (1.3) and pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). We report the estimates $\hat{\beta}$ by sorting stocks in quintiles based on the following characteristics: market cap, trade-to-quote ratio, idiosyncratic volatility, and passive institutional ownership. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote (TTQ) ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($oib^{tot}_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors. Results using the alternative measure of price efficiency *abs_ac* are detailed in Table A.3.

Panel A: Active Pension Funds

Sorting variable:	Q1	Q2	Q3	Q4	Q5
Market cap	-0.016 (-0.66)	-0.028 (-1.88)	0.015 (0.13)	-0.026 (-3.01)	-0.011 (-1.19)
TTQ ratio	-0.044 (-3.62)	-0.017 (-2.03)	-0.017 (-1.98)	-0.019 (-1.91)	-0.009 (-0.72)
Idiosync. vol.	-0.033 (-3.31)	-0.021 (-2.27)	-0.029 (-3.54)	-0.020 (-2.09)	-0.007 (-0.75)
Passive ownership	0.001 (0.08)	-0.026 (-2.33)	-0.039 (-2.63)	-0.014 (-1.21)	-0.018 (-1.27)

Panel B: Full Sample of Stocks

Sorting variable:	Q1	Q2	Q3	Q4	Q5
Market cap	-0.170 (-19.1)	-0.060 (-8.31)	-0.040 (-7.30)	-0.026 (-5.11)	-0.004 (-0.79)
TTQ ratio	-0.066 (-8.16)	-0.038 (-8.40)	-0.045 (-9.63)	-0.059 (-11.8)	-0.072 (-10.8)
Idiosync. vol.	-0.031 (-6.31)	-0.043 (-9.63)	-0.047 (-10.3)	-0.067 (-13.2)	-0.133 (-16.1)
Passive ownership	-0.191 (-13.7)	-0.076 (-8.88)	-0.039 (-6.44)	-0.028 (-5.07)	-0.029 (-4.98)

Table 1.9: Characteristics of Stocks sorted by the Predictability of Order Imbalance. This table reports the average value of the characteristics of stocks sorted on quintiles based on the predictability of order imbalance as proxied by pl_{it}^{tot}. The predictive likelihood for the full sample of stocks is defined as the 20-day rolling average over $\tilde{pl}_{it}^{tot}$ which is calculated as in Eq.(1.3). Here, $Size$ is the market capitalisation of the firm i (in million of dollars), BE/ME is the book-to-market ratio, ret_t and $\sigma(ret_t)$ are the monthly return and its standard deviation, $qspread$ is the quoted spread, ttq denotes the trade-to-quote ratio, $voib$ is the volatility of order imbalance, and $turn_{it}$ is the share turnover (in percent). Sample period is from January 1, 1997 to December 31, 2013.

Quint. of pl_t^{tot}	Size	BE/ME	ret_t	$\sigma(ret_t)$	$qspread$	ttq	$voib$	$turn$
Q1 (low)	1,756	0.93	1.87	3.07	212.38	0.32	0.43	0.38
Q2	2,565	0.79	1.65	2.90	129.06	0.34	0.32	0.54
Q3	3,373	0.70	1.59	2.89	89.98	0.36	0.26	0.67
Q4	5,643	0.62	1.55	2.80	62.79	0.40	0.21	0.81
Q5 (high)	6,310	0.54	1.58	2.92	41.28	0.47	0.15	1.13

Table 1.10: Double Sorted Portfolio Returns.
This table reports the returns on portfolios sorted on the lagged value of a given stock characteristic and the past predictability of order imbalance for the full sample of stocks. The latter is defined as the 20-day rolling average over $\tilde{pl}_{it}^{tot}$ which is calculated as in Eq.(1.3). The characteristics used for double sorting are the market capitalisation (Panel A), the quoted bid-ask spread (Panel B), share turnover (Panel C), the trade-to-quote ratio (Panel D), the volatility of order imbalance (Panel E), and order imbalance (Panel F). We first sort stocks into quintile portfolios based on a characteristic, and then sort on pl_{it}^{tot} into quintile portfolios in each group at month t. Portfolio returns and return differences (Low-High) in month $t+1$ are reported. All returns are reported in percent. The sample period is from January 1, 1997 to December 31, 2013. The t-statistics for the low-minus-high predictability portfolio are based on Newey-West adjusted standard errors. The Jensen's alphas are calculated conditioning on the three Fama and French (1993a) factors along with the momentum factor of Jegadeesh and Titman (1993) and the Pastor and Stambaugh (2003) traded liquidity factor.

Panel A: Double sort by order imbalance predictability and market cap

	Quintile of $lsize$				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	2.785	1.735	1.394	1.175	0.949
Q2	1.956	1.196	1.220	0.993	0.903
Q3	1.720	1.174	1.170	0.969	0.828
Q4	1.363	1.098	0.765	0.888	0.913
Q5 (high)	1.146	1.043	0.721	0.725	0.691
Low-High	1.639 (4.47)	0.691 (1.84)	0.673 (1.97)	0.450 (1.35)	0.259 (0.71)
Alpha	1.922 (4.64)	1.025 (2.48)	1.023 (2.82)	0.730 (2.05)	0.501 (1.45)

Panel B: Double sort by order imbalance predictability and quoted bid-ask spread

	Quintile of $qspread$				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	0.983	1.418	1.886	2.873	3.216
Q2	1.006	1.178	1.324	1.812	2.009
Q3	0.916	1.020	1.234	1.518	1.810
Q4	0.896	1.078	1.220	1.229	1.485
Q5 (high)	0.832	0.804	1.048	1.196	1.033
Low-High	0.151 (0.46)	0.614 (1.56)	0.838 (2.04)	1.677 (3.58)	2.183 (5.50)
Alpha	0.392 (1.28)	0.975 (2.42)	1.139 (2.48)	2.076 (4.24)	2.369 (5.53)

Panel C: Double sort by order imbalance predictability and trade-to-quote ratio

	Quintile of ttq				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	1.641	1.951	2.028	2.253	3.327
Q2	1.040	1.280	1.487	1.695	2.280
Q3	0.926	1.039	1.128	1.658	1.674
Q4	0.943	0.967	1.193	1.178	1.190
Q5 (high)	0.796	1.025	0.818	0.900	0.589
Low-High	0.845 (3.30)	0.926 (3.26)	1.210 (3.39)	1.353 (3.55)	2.74 (5.50)
Alpha	1.922 (3.82)	1.215 (4.01)	1.366 (3.94)	1.641 (4.29)	3.00 (5.85)

Table 1.10 continued.

Panel D: Double sort by order imbalance predictability and volatility of order imbalance

	Quintile of *voib*				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	1.142	1.753	2.402	2.802	2.621
Q2	1.122	1.466	1.699	1.799	1.650
Q3	1.131	1.160	1.397	1.623	1.291
Q4	0.890	1.061	1.153	1.048	1.032
Q5 (high)	0.650	0.874	1.144	1.136	0.880
Low-High	0.492 (1.19)	0.879 (2.58)	1.257 (3.35)	1.667 (3.80)	1.741 (5.00)
Alpha	0.834 (1.93)	1.188 (3.07)	1.407 (3.88)	1.856 (4.25)	1.835 (5.67)

Panel E: Double sort by order imbalance predictability and share turnover

	Quintile of *turn*				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	1.831	2.719	2.485	2.469	2.631
Q2	1.049	1.705	1.661	1.606	1.832
Q3	0.903	1.157	1.442	1.349	1.249
Q4	0.557	0.915	0.971	1.247	1.098
Q5 (high)	0.759	0.903	1.049	0.663	0.641
Low-High	1.027 (3.71)	1.816 (3.57)	1.436 (3.67)	1.806 (4.46)	1.990 (4.65)
Alpha	1.166 (4.35)	1.968 (3.97)	1.559 (4.11)	1.984 (5.04)	2.176 (5.29)

Panel F: Double sort by order imbalance predictability and order imbalance

	Quintile of *oib*				
Quintile of pl_t^{tot}	Q1	Q2	Q3	Q4	Q5
Q1 (low)	2.439	2.462	2.078	1.872	1.787
Q2	1.813	1.672	1.422	1.441	1.324
Q3	1.438	1.664	1.137	1.137	1.076
Q4	1.346	1.379	1.015	1.037	0.927
Q5 (high)	1.041	1.145	0.838	0.770	0.750
Low-High	1.398 (4.56)	1.317 (2.92)	1.240 (2.96)	1.102 (2.48)	1.036 (3.70)
Alpha	1.165 (4.85)	1.648 (3.52)	1.615 (3.91)	1.366 (3.28)	1.265 (4.81)

Table 1.11: Betas on the Order Imbalance Predictability Factor.
This table reports the betas on the returns on the POI factor calculated from the following regression

$$returns_{i,t} = \alpha_i + \beta_i POI_t + \gamma_i' \mathbf{F}_t + u_{i,t}, \qquad i = 1, \ldots, 25$$

where $\mathbf{F}_t$ contains the excess returns on the market, the size and value factors of Fama and French (1993a), the momentum factor calculated as in Jegadeesh and Titman (1993), the liquidity risk factor from Pastor and Stambaugh (2003), as well as the short-term reversals factor and the returns on high-minus low portfolios sorted by past month trade-to-quote ratio and both the level and volatility of total order imbalance. The returns on the 5×5 test assets are constructed by sorting on both size and the predictability of order imbalance. The POI risk factor is defined as monthly returns on the low minus high portfolio sorted by predictive likelihood of total order imbalance pl_{it}^{tot}. We report the estimates $\hat{\beta}_i$ for each of these test portfolios. **Panel A:** reports the results for test assets which are constructed as equally-weighted portfolios. **Panel B:** reports the results for value-weighted test portfolios. The sample period is from January 1, 1997 to December 31, 2013. The t-statistics are based on Newey-West adjusted standard errors. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Equally-Weighted Portfolios

Decile of pl_t^{tot}	Q1 (small)	Q2	Q3	Q4	Q5 (large)
Q1 (low)	0.662***	0.310***	0.460***	0.556***	0.236***
Q2	0.320***	0.212***	0.208***	0.271***	0.236***
Q3	0.147**	0.073	0.072	0.139**	0.142***
Q4	-0.047	-0.137***	-0.022	0.024	0.021
Q5 (high)	-0.345***	-0.437***	-0.349***	-0.209***	-0.240***

Panel B: Value-Weighted Portfolios

Decile of pl_t^{tot}	Q1 (small)	Q2	Q3	Q4	Q5 (large)
Q1 (low)	0.507***	0.305***	0.343***	0.493***	0.159**
Q2	0.240***	0.204***	0.180***	0.222***	0.189***
Q3	0.110**	0.085*	0.085	0.126**	0.124***
Q4	-0.095*	-0.125**	-0.015	0.027	-0.002
Q5 (high)	-0.376***	-0.433***	-0.325***	-0.208***	-0.239***

Table 1.12: Fama-MacBeth Regressions.
This table reports estimates of the cross-sectional regression

$$returns_{i,t} = \lambda_{POI,t}\hat{\beta}_i + \lambda_t'\hat{\gamma}_i + e_{i,t}, \qquad t = 1, \dots, T$$

where the parameters $\hat{\beta}_i, \hat{\gamma}_i$ are estimated for each test portfolio by using the time series regression in Eq.(1.8). The estimates $\hat{\lambda}_{POI}, \hat{\lambda}'$ are obtained as the sample averages of $\hat{\lambda}_{POI,t}, \hat{\lambda}_t'$. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). **Panel A:** reports the results for test assets which are constructed as equally-weighted portfolios. **Panel B:** reports the results for value-weighted test portfolios. The sample period is from January 1, 1997 to December 31, 2013. The sample period is from January 1, 1997 to December 31, 2013. The t-statistics (in parentheses) are based on Newey-West adjusted standard errors.

Panel A: Equally-Weighted Portfolios

POI	1.21 (2.97)	1.19 (3.48)	1.15 (2.88)	1.18 (3.30)	1.19 (3.29)	1.18 (2.96)	1.30 (2.96)
Mkt	0.60 (1.56)	0.60 (1.64)	0.57 (1.46)	0.62 (1.63)	0.59 (1.58)	0.60 (1.53)	0.58 (1.41)
SMB	0.70 (2.14)	0.54 (1.78)	0.65 (2.15)	0.53 (1.83)	0.66 (2.20)	0.56 (2.17)	0.77 (2.85)
HML	-0.43 (-0.91)	-0.40 (-0.92)	-0.42 (-0.88)	-0.33 (-0.80)	-0.53 (-1.23)	-0.38 (-0.92)	-0.47 (-1.16)
MOM	-0.64 (-0.82)						-0.85 (-0.98)
REV		0.27 (0.40)					-1.41 (-1.38)
LIQ			0.96 (0.89)				0.09 (0.14)
$VOIB$				0.71 (2.09)			0.55 (1.37)
TTQ					-0.91 (-1.70)		-0.80 (-1.62)
OIB						-0.72 (-2.63)	-1.01 (-3.08)

Panel B: Value-Weighted Portfolios

POI	1.43 (3.27)	1.44 (3.41)	1.42 (3.37)	1.44 (3.48)	1.43 (3.30)	1.44 (3.53)	1.57 (3.66)
Mkt	0.61 (1.62)	0.61 (1.62)	0.61 (1.61)	0.62 (1.61)	0.61 (1.62)	0.61 (1.62)	0.60 (1.52)
SMB	0.60 (2.09)	0.61 (2.11)	0.60 (2.21)	0.58 (2.12)	0.60 (2.25)	0.57 (2.08)	0.70 (2.48)
HML	-0.49 (-1.03)	-0.49 (-1.12)	-0.50 (-1.14)	-0.46 (-1.16)	-0.50 (-1.34)	-0.46 (-1.14)	-0.56 (-1.20)
MOM	-1.19 (-1.41)						-1.19 (-1.26)
REV		-0.03(-0.04)					-0.14 (-1.09)
LIQ			0.51(0.62)				-0.24 (-0.32)
$VOIB$				0.82 (2.24)			0.68 (1.52)
TTQ					-1.03 (-1.64)		-0.89 (-1.50)
OIB						-0.86 (-2.50)	-1.07 (-2.48)

Chapter 2

Bond Risk Premia With Machine Learning[1]

> *"There is a popular cliché ... which says that you cannot get out of computers any more than you put in. Other versions are that computers only do exactly what you tell them to, and that therefore computers are never creative. The cliché is true only in the crashingly trivial sense, the same sense in which Shakespeare never wrote anything except what his first schoolteacher taught him to write — words."*
>
> – Richard Dawkins; The Blind Watchmaker (1986)

Recent advancements in the fields of statistics and computer science have spurred interest in dimensionality reduction and model selection techniques, as well as predictive models with complex features, such as sparsity and non-linearity, both in finance and economics.[2] Over the last two decades, however, the use of such methods in the financial economics literature has been mostly limited to data compression techniques, such as principal component and latent factor analysis. A likely explan-

[1]This chapter is based on a research paper jointly authored with Daniele Bianchi and Andrea Tamoni with the same title. Parts of the chapter have been accepted for publication in the *Review of Financial Studies* as Bianchi et al. (2020). The chapter also incorporates results from the accompanying online appendix accessible on the publishers website. After acceptance for publication an issue regarding overlapping observations was brought to our attention that is addressed in a corrigendum which will be published alongside the main paper in the *Review of Financial Studies.* Appendix B.3 presents additional results from the corrigendum.

[2]See, for example, Rapach et al. (2013), Kelly and Pruitt (2013; 2015), Freyberger, Neuhierl and Weber (2020a), Giannone et al. (2017), Giglio and Xiu (2017), Heaton, Polson and Witte (2017), Kozak, Nagel and Santosh (2017), Messmer (2017), Fuster et al. (2018), Gu, Kelly and Xiu (2020), Kelly, Pruitt and Su (2019), Rossi (2018), Sirignano et al. (2018), Chen, Pelger and Zhu (2019), Feng, Giglio and Xiu (2020), Feng, Polson and Xu (2019), and Huang and Shi (2019).

ation for the slow adoption of advances in statistical learning is that these methods are not suitable for structural analysis and parameter inference (see Mullainathan and Spiess, 2017). Indeed, the primary focus of machine learning is prediction, i.e. to produce the best out-of-sample forecast of a quantity of interest based on a potentially large conditioning information set.

The suitability of machine learning methodologies for predictive analysis makes them particularly attractive in the context of financial asset return predictability and risk premia measurement (e.g. Gu, Kelly and Xiu, 2020). As a matter of fact, while many problems in economics rely on the identification of primitive underlying shocks and structural parameters, the quantification of time variation in expected returns is essentially a forecasting problem. This practical view complements the theory-driven approach, which often provides the building blocks for the empirical analysis of financial markets. Modelling the predictable variation in Treasury bond returns, which is the focus of this paper, provides a case in point. Forecasting excess bond returns requires a careful approximation of the a priori unknown mapping between the investors' information set and excess bond returns (e.g. Duffee, 2013, pp. 391-392).

In this paper, we employ machine learning methods to revisit the debate on the presence of predictable variation in bond returns. We work with two traditional frameworks; one that exploits information in the yield curve only, as in Cochrane and Piazzesi (2005), and one that also uses information from a dataset of hundreds of macroeconomic indicators as in Ludvigson and Ng (2009). The research design follows the structure outlined in Gu et al. (2020), whereby a comparison of different machine learning techniques is based on their out-of-sample predictive performance. Methodologically, we consider a variety of machine learning techniques to forecast excess Treasury bond returns across different maturities including partial least squares, penalised linear regressions, boosted regression trees, random forests, extremely randomised regression trees, and shallow and deep neural networks (NNs). All of these methods fall under the heading of "supervised learning" in the computer science literature.[3] Although not exhaustive, this list covers the vast majority of modern

[3]In "supervised" statistical learning the mapping between the quantity of interest y and the predictors $\boldsymbol{x}$ is learned by using information on the joint distribution. Unsupervised learning (e.g. PCA) instead does not explicitly condition on the quantity of interest y to summarise the information content in $\boldsymbol{x}$.

statistical learning techniques (e.g. Friedman et al., 2001).

We also employ more classical dimensionality reduction techniques, such as principal component analysis (PCA), which arguably represent an almost universal approach to regression-based forecasting of Treasury bond returns.

Our contribution to the bond return predictability literature is threefold. First, within each empirical application (i.e. yields-only or yields plus macroeconomic variables), we show that non-linear machine learning methods, such as extreme trees and NNs, are useful to detect predictable variations in bond excess returns, as indicated by out-of-sample predictive R^2s that are significantly higher than those obtained by data compression techniques (e.g. linear combinations of forward rates, as in Cochrane and Piazzesi (2005), and factors extracted from macroeconomic variables, as in Ludvigson and Ng (2009)) and penalised regression techniques. Importantly, a battery of asset allocation exercises confirm that the deviations from the Expectations Hypothesis documented in this paper are economically large. In this regard, our paper contributes to the debate on the statistical evidence supporting bond return predictability (e.g. Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005), for applications with yields-only) or absence thereof (e.g. Thornton and Valente, 2012).

Second, zooming in on non-linear methods, we document that using information from macroeconomic and financial variables improves the predictive accuracy of forecasts based only on (potentially non-linear transformations of) the yield curve.

Indeed, the best-performing NN that exploits macroeconomic and term structure information attains out-of-sample R^2s that are about 10 percentage points larger (for maturities ranging from two to ten years) than the best-performing NN that employs yields only. Similarly, we document that employing the NN forecasts based on macroeconomic and yield information produces significantly higher certainty equivalent return values than those implied by the NN forecasts based only on yield curve information. In this respect, our paper contributes to the debate on whether there is macroeconomic variation not spanned by bond yields that helps forecast excess bond returns (e.g. see Joslin et al., 2014 for evidence in favour of unspanned macroeconomic information, and Bauer and Rudebusch, 2017; Bauer and Hamilton, 2018 for a critical analysis of such evidence, along with a discussion of econometric issues plaguing the "spanning" linear regressions). On the one hand, our analysis reinforces the evidence in favour of unspanned macroeconomic information useful to forecast excess bond returns (e.g. Cooper and Priestley, 2009; Ludvigson and Ng, 2009; Duffee, 2011a; Joslin et al., 2014; Cieslak and Povala, 2015; Coroneo et al., 2016; Gargano et al., 2019). On the other hand, our evidence is novel in three respects. First, we continue to find support for unspanned macroeconomic risk even after accounting for potential non-linearities in interest rates. Second, we find that it is important to account for non-linearities within macroeconomic categories in order to detect information useful for predicting excess bond returns above and beyond the yield curve. Finally, we document substantial heterogeneity in the relative importance of macroeconomic and financial variables across bond maturities: variables pertaining to the stock and labour markets are more important for short-term maturity bonds, whereas variables pertaining to orders and inventories, and output and income are more relevant for variation in long-term bonds. Thus, the type and nature of unspanned factors may depend on bond maturity.

Our third contribution concerns the economic properties of the forecasts implied by deep NNs. First, to provide insight into the origins of the improvements in out-of-sample predictability, we investigate the ability of NNs to forecast the first three principal components of the term structure: level, slope, and curvature. We show that when using yields only as predictors, the NNs improve the forecast of the level of the term structure. However, when both macroeconomic and financial information is used in addition to yields, we find that the factors extracted from the NNs contribute to the ability to predict the level of the yield curve, as well as the slope. This is consistent with the idea that the slope of the yield curve is related to the state of the economy, and an NN is able to extract the relevant information from the large set of macroeconomic variables used. Next, we document that NN forecasts are counter-cyclical and mostly related to variables that proxy for macroeconomic uncertainty and time-varying risk aversion. Thus, our results support models that feature both time variation in risk prices and time-varying risk as in, e.g. Bekaert et al. (2009) and Creal and Wu (2018). However, our statistical measure of expected bond returns contrasts recent survey-based measures like the one proposed by Buraschi et al. (2019), which is mostly related to financial (specifically, bond) volatility.

In the context of machine learning in asset pricing, we document three novel facts.[4] First, our result that extreme trees and NNs constitute the best-performing methods even in the case when only information in the term structure is used to forecast bond returns (i.e. in a low dimensional setting) is new and provides evidence that the gain from non-linear machine learning methods is not relegated to a big data context.

Second, we show that an economically-driven choice of the network structure may perform on par with more data-driven network architectures. More specifically, when macroeconomic data are included as potential bond return predictors, we find that the out-of-sample predictive R^2 increases almost monotonically when we move from shallow specifications (one hidden layer) to deeper networks (up to three hidden layers). However, we also find that economic priors about the role of variables may improve the performance of the network. In particular, grouping variables within economic categories and then training a shallow network within each group – a network structure that we dub "group ensembling"– attains a performance that is on par with the best-performing deep NN where no economic priors are utilised. Thus, the depth of the network and the economic priors used to design the network (e.g. grouping within categories) interact with one another, a result that is new to the empirical finance literature.

Third, the fact that the group ensembled network outperforms more complex and agnostic specifications is important since it highlights what type of non-linearities are important from an economic perspective: Is it the interaction of many variables (across categories) or rather a higher polynomial of the same variable (within a category)? Since our group-ensembled network switches off interactions across categories, our analysis shows that it is the non-linearity *within* a group that is ultimately relevant for the performance of the network. In this respect, our results for Treasury bond returns echo those in Gu et al. (2020) and Chen et al. (2019) for the equity market: the success of NNs lies in their ability to exploit the non-linear mapping between returns and the predictors. However, whereas Chen et al. (2019) emphasise the importance of identifying the relevant interaction between firm characteristics for equity returns, we document that, in the bond market, the interaction across economic categories matters to a lesser extent than the interaction within a

[4]The literature on machine learning and asset pricing is rapidly growing (cf., footnote 1). Except for Huang and Shi (2019), none of these papers explore machine learning methods to forecast excess bond returns.

category. Thus, different types of network structures may be needed for different asset markets.

The remainder of this paper is organised as follows. Section 2.1 provides a discussion why machine learning techniques can prove useful to measure expected bond returns within the context of predictive regression. Section 2.2 outlines the estimation strategy and machine learning methodologies used in the paper. Section 2.4 summarises the results on predictability of bond excess returns. Section 2.5 dissects the predictability of bond excess returns uncovered by machine learning methods along various dimensions. Section 2.6 examines whether gains in predictive accuracy translate into better investment performance. Section 2.7 considers the economic drivers of bond return predictability. Section 2.8 concludes.

2.1 Motivating Framework

In this section, we provide a motivation for the use of machine learning to predict excess Treasury bond returns. The discussion is framed within the context of regression approaches for forecasting treasury yields. We start with the accounting identity of Campbell and Shiller (1991). We consider a zero-coupon bond with maturity $t + n$ and a pay-off of one dollar. We denote its (log) price and (continuously compounded) yield at time t by $p_t^{(n)}$ and $y_t^{(n)} = -\frac{1}{n} p_t^{(n)}$, respectively. The superscript refers to the bond's remaining maturity. The (log) excess return to the n-year bond from t to $t + 1$, when its remaining maturity is $n - 1$, is denoted by $xr_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)}$. Then, it is possible to express the log returns to bonds as:

$$xr_{t+1}^{(n)} = -(n-1)\left(y_{t+1}^{(n-1)} - y_t^{(n)}\right) + \left(y_t^{(n)} - y_t^{(1)}\right). \tag{2.1}$$

The identity states that (after controlling for the slope $y_t^{(n)} - y_t^{(1)}$) any variable that forecasts the change in the bond yield from t to $t + 1$, i.e. $\left(y_{t+1}^{(n-1)} - y_t^{(n)}\right)$, also forecasts the log returns to bonds. Assuming that the investors' information set at time t can be summarised by a latent k-dimensional state vector $\boldsymbol{x}_t$, and exploiting the identity $y_t^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} E_t\left(y_{t+j}^{(1)} \mid \boldsymbol{x}_t\right) + \frac{1}{n}\sum_{j=0}^{n-1} E_t\left(xr_{t+j+1}^{(n-j)} \mid \boldsymbol{x}_t\right)$, we can write:

$$\boldsymbol{y}_t = f(\boldsymbol{x}_t; N),$$

where we stack time-t yields on bonds with different maturities in a vector $\boldsymbol{y}_t$, and the maturities of the bonds in the vector N. Combining the equation above with Equation (2.1), we obtain:

$$E_t\left[xr_{t+1}^{(n)}\right] = g(\boldsymbol{x}_t; N) \ , \tag{2.2}$$

for some function $g(\boldsymbol{x}_t; N)$. Every term structure model reduces to a specific mapping between yields and state variables.

In the simplest case, yields are linear affine functions of the state variables: $\boldsymbol{y}_t = A + B\boldsymbol{x}_t$. The linearity of $f(\cdot)$, together with a dimensionality reduction of the space of yields, gives rise to principal component regression (PCR) where the quantity of interest (excess bond returns) is regressed onto principal components $\boldsymbol{x}_t$ (see Chapter 3.5 in Friedman et al., 2001):

$$E_t\left[xr_{t+1}^{(n)}\right] = \hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{x}_t \qquad \text{with} \qquad \boldsymbol{x}_t = \boldsymbol{W}\boldsymbol{y}_t + b, \tag{2.3}$$

where the columns of $\boldsymbol{W}$ form an orthogonal basis for directions of greatest variance, and b captures the average "reconstruction error" or bias.

Practically, the linear predictive system outlined in Equation (2.3) represents a two-step procedure where researchers extract the latent factors $\boldsymbol{x}_t$, and then learn the regression coefficients $\hat{\theta} = \left(\hat{\alpha}, \hat{\boldsymbol{\beta}}^\top\right)$ by minimising a loss function that depends on the residual sum of squares. In addition to this yields-only specification, researchers have often evaluated the role of macroeconomic variables as an important driver of bond returns. This leads to an augmented predictive regression:

$$E_t\left[xr_{t+1}^{(n)}\right] = \hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{x}_t + \hat{\boldsymbol{\gamma}}^\top \boldsymbol{F}_t \tag{2.4}$$

where $\boldsymbol{F}_t \subset f_t$ and f_t is an $r \times 1$ vector of latent common factors extracted from a $T \times N$ panel of macroeconomic data with elements $m_{it}, i = 1, \ldots, N, \ t = 1, \ldots, T$, and $r \ll N$. This is the framework originally proposed by Ludvigson and Ng (2009).

Equations (2.3) and (2.4) constitute two important applications of unsupervised data compression for bond forecasting.

Another very popular set of reduced-form term structure models consists of Gaussian linear-quadratic models (e.g. Ahn, Dittmar and Gallant, 2002). In this case, the relation between yields and state variables is given by $\boldsymbol{y}_t = A + B\boldsymbol{x}_t + \boldsymbol{x}_t' C\boldsymbol{x}_t$. Note that in this case the mapping between yields and factors is non-linear in the state variables, meaning that standard linear PCA represents a mere approximation and does not necessarily give consistent estimates of the true underlying quadratic factor (see Schölkopf et al., 1998). Non-linearities are also featured in reduced-form term structure models with regime switches (e.g. Dai et al., 2007), in shadow rate models (Black, 1995; Wu and Xia, 2016), and in the model by Feldhutter et al. (2016) where the price of a bond is a time-varying weighted average of bond prices in artificial, affine Gaussian economies.

Interestingly, non-linearity between bond yields and factors also emerges naturally in structural models with habit formation. For example, Buraschi and Jiltsov (2007) use habit formation preferences as a source of time varying market price of risk in fixed income models. In their model, yields are non-linear in the state variables, and depend on the habit stock and the factors affecting the monetary aggregate. For completeness, Appendix B.1 provides a simple habit formation economy that leads to bond yields being a linear-quadratic function of macroeconomic variables like consumption growth, expected inflation, and habit.

Motivated by this literature, we investigate the possibility that a more precise measurement of bond risk premia can be obtained by using non-linear transformations of the data, an avenue that has also been advocated by Stock and Watson (2002, p. 154) within the context of forecasting macroeconomic time series. Differently from the Gaussian linear-quadratic models, we do not postulate a specific functional form connecting bond yields and state variables; instead we use various statistical techniques such as trees and networks to learn about it. Besides being agnostic about the functional form between excess bond returns and macroeconomic and financial variables, the use of machine learning techniques has two additional advantages relative to the principal component regressions in Equations (2.3) and (2.4).

First, the implementation of regression-based forecasts of excess bond returns using principal components as outlined in Equations (2.3) and (2.4) typically implies that no direct use of the response variable (i.e. the excess bond returns) is made to learn about the state variables $\boldsymbol{x}_t$ and $\boldsymbol{F}_t$. This is not surprising as data compression methods such as PCA are a form of "unsupervised learning".

However, Equation (2.2) suggests that excess bond returns play the implicit role of a conditioning argument, namely, one should be able to tailor the extraction of hidden latent states $\boldsymbol{x}_t$ to the response variable $xr_{t+1}^{(n)}$. In this respect, "supervised learning" algorithms, such as the lasso, elastic net, partial least squares, regression trees, random forests, and NNs, that explicitly condition on the response variables to summarise the information in the predictors, may arguably prove useful to overcome the limitations of standard data compression methods.[5]

Second, traditional PCA and factor analysis (FA) are based on the assumption that all variables could bring useful information for the prediction of future excess bond returns, although the impact of some of them could be small. However, PCA or FA does not guarantee that by simply adding any number of predictors, we can be sure that the extracted factors will provide an optimal summary. Boivin and Ng (2006) formalise this argument by providing evidence that the structure of the common components is sensitive to the input variables, and that more data does not always mean there will be more sensible estimates. In this respect, one may want to "select" the variables that actually matter for forecasting excess bond returns. Penalised regressions, such as lasso and elastic net, as well as NNs exploit the entire span of the input variables without imposing that they all carry useful information for determining excess bond returns.

The existing literature on bond return predictability has vastly ignored the potential capability of machine learning techniques to address the issue of non-linearity and variable regularisation. Arguably, this comes at the expense of not fully capturing the extent to which yields and macroeconomic variables are relevant for the measurement of expected excess bond returns. This is the focus of our paper.

2.2 Research Design

In this section, we outline the research design for the empirical analysis. We start with a description of the data, along with the specific applications. We then review

[5]Ludvigson and Ng (2009, p. 5034) acknowledge that "factors that are pervasive for the panel of data [input] need not be important for predicting [the output]" and propose a three-step forecasting procedure where a subset of principal components extracted from a large panel of macroeconomic variables is selected according to the information criteria before running the bond return forecasting regressions. In line with this intuition, we provide evidence that supervised learning methodologies, such as NNs, are useful to exploit the information in predictors other than yields, and to improve the out-of-sample forecast of bond returns.

the methodologies implemented in the main empirical analysis. We conclude with a short discussion of our estimation strategy.

2.2.1 Data and Empirical Applications

The empirical analysis is based on two main benchmark applications. The first application concerns the forecasting of future bond excess returns based on the cross-section of yields as originally proposed by Cochrane and Piazzesi (2005). We use the novel zero-coupon Treasury yield curve dataset constructed by Liu and Wu (2020). This dataset allows us to study bond returns with maturity greater than five years (the longest maturity in the Fama-Bliss dataset). This is important since long maturity yields contain substantial extra predictive power over and above the first five yields (Le and Singleton, 2013, discuss the importance of using long-maturity bond yields in assessing the dynamic properties of risk premiums in Treasury markets). Moreover, Liu and Wu (2020) construct the zero-coupon curve using a non-parametric kernel-smoothing method that does not discard Treasury bills, which is instead the case for the parametric approach adopted by Gurkaynak et al. (2007). This is important since Liu and Wu (2020) find that securities at the short end of the yield curve contain important information in disciplining the overall behaviour of the curve. Using the Liu and Wu (2020) yield curve dataset, we then construct forward rates and excess bond returns as described in Section 2.1. We focus on bonds with maturities up to 10 years. Since the U.S. Treasury started issuing 10-year notes in September 1971, this also defines the start of our sample period. We do not use bonds with longer maturities for two reasons. First, the Treasury began issuing 20-year bonds in July 1981, and 30-year bonds in November 1985. This would force us to start the analysis later, reducing further the out-of-sample period since training the NNs requires a sufficient amount of data.[6] Second, the issuance of long-maturity Treasury notes and bonds occurs at irregular intervals; to compensate for the lack of observations at long-maturities, the Liu and Wu (2020) method pulls information for the 20- and 30-year bonds from maturities that are 10

[6]Contrary to typical machine learning applications, such as image recognition, common signal-to-noise ratios in financial data are low, exacerbating the need for sufficient data. Hence, we delay the start of the out-of-sample period so far that we have at least a handful of observations per weight to be estimated in our smallest NN (i.e. a single hidden layer with three nodes). For larger network architectures in which the number of parameters exceeds the number of available observations, regularisation methods are used to ensure satisfactory training.

years (or more) away.

The second application consists of forecasting future bond excess returns based on both forward rates and a large panel of macroeconomic variables as proposed by Ludvigson and Ng (2009). We consider a balanced panel of $N = 128$ monthly macroeconomic and financial variables. A detailed description of how variables are collected and constructed is provided in McCracken and Ng (2016). The series were selected to represent broad categories of macroeconomic time series: real output and income, employment and hours, real retail, manufacturing and sales data, international trade, consumer spending, housing starts, inventories and inventory sales ratios, orders and unfilled orders, compensation and labour costs, capacity utilisation measures, price indexes, interest rates and interest rate spreads, stock market indicators, and foreign exchange measures. This dataset has been widely used in the literature (e.g. Stock and Watson, 2002, 2006; Ludvigson and Ng, 2009), and permits comparison with previous studies.

2.2.2 Forecasting Methods

Principal Component Regressions and Partial Least Squares

The first method we employ is a linear, dimensionality-reduction technique known as principal component regressions (PCRs). Undoubtedly, PCRs constitute the most common method used to forecast interest rates and Treasury bond returns.

In the classical implementation of PCRs, the target variable is discarded when extracting the latent factors. Thus, we also consider an alternative data compression methodology called partial least squares (PLS). Unlike PCR, with PLS the common components of the predictors are derived by conditioning on the joint distribution of the target variable and the regressors. Section 2.3.1 provides additional details on PLS, while contrasting this method to PCRs and penalised regressions, which we discuss next.

Penalised Regressions: Ridge, Lasso and Elastic Net

Confronted with a large set of predictors, a popular strategy is to impose sparsity / shrinkage in the set of regressors via a penalty term. The idea is that by selecting a subset of variables with the highest predictive power out of a large set of predictors, and discarding the least relevant ones, one can mitigate in-sample overfitting and improve the out-of-sample performance of the linear model.

In its general form, a penalised regression entails adding a penalty term on top of the OLS objective function $\mathcal{L}_{OLS}(\boldsymbol{\theta}) = \frac{1}{t}\sum_{\tau=1}^{t-1}\left(xr_{\tau+1}^{(n)} - \alpha - \boldsymbol{\beta}^\top \boldsymbol{y}_\tau\right)^2$ with $\boldsymbol{\theta} = \left(\alpha, \boldsymbol{\beta}^\top\right)$:

$$\mathcal{L}(\boldsymbol{\theta};\cdot) = \underbrace{\mathcal{L}_{OLS}(\boldsymbol{\theta})}_{\text{Loss Function}} + \underbrace{\phi(\boldsymbol{\beta};\cdot)}_{\text{Penalty Term}} . \tag{2.5}$$

Depending on the functional form of the penalty term, the regression coefficients can be regularised and shrunk towards zero (as in ridge), exactly set to zero (as in lasso), or a combination of the two (as in elastic net). In Section 2.3.2, we describe each method in details.

Penalised regressions still do not account for non-linear relations. To address this issue, we consider a third class of non-linear methods: "shallow learners", such as regression trees and more deep structures, such as neural networks.

Regression Trees

Regression trees are based on a partition of the input space into a set of "rectangles." Then, a simple linear model is fit to each rectangle. Figure 2.1 displays an example of a binary partition (Panel (a)) and the corresponding regression tree (Panel (b)). Regression trees are conceptually simple, yet powerful, and therefore highly popular in the machine learning literature. In addition to a standard regression tree methodology, we consider extensions that employ an ensemble of individual trees, like "random forests' (Breiman, 2001), and, furthermore, take into account the randomness in the predictors' partition process, like "extremely randomised trees" (Geurts et al., 2006). Section 2.3.3 provides additional technical details on the estimation of regression trees (and their extensions).

Neural Networks

Neural networks (NNs) represent a widespread class of supervised learning methods. We focus on traditional "feed-forward" networks or multi-layer perceptrons (MLP). Throughout the paper, we follow Feng et al. (2018) and adopt the convention of counting only the hidden layers, without including the output layer.

For details on the estimation of NNs, see Section 2.3.4. Next, we provide a high level description of the NNs structure we consider in our empirical applications.

Neural networks: yields-only. When we forecast bond returns using only information from the term structure of interest rates, we use variants of the NN architecture depicted in Figure 2.2. That is, we use a classical MLP. An MLP consists of, at least, three layers of nodes (the case displayed in the figure): an input layer, a hidden layer and an output layer.

The result is a learning method that can approximate virtually any continuous function with compact support (Kolmogorov, 1957; Diaconis and Shahshahani, 1984; Cybenko, 1989; Hornik, Stinchcombe and White, 1989). In the empirical application, we study the predictive accuracy of this classical network as we vary the number of hidden layers, as well as the number of nodes per layer.

Neural networks: macro plus yields. When we forecast bond returns using macroeconomic variables in addition to information from the term structure, we consider three alternative specifications that extend the typical MLP structure by taking into account the economic structure of the input data, as well as the nature of the forecasting problem.

The first specification, displayed in Panel (a) of Figure 2.3, can be thought of as a "hybrid" modelling framework in the sense that forward rates are simply included as an additional predictor in the output layer ("fwd rates direct"). This structure simulates the idea of Ludvigson and Ng (2009) in which the latent factors $\boldsymbol{F}_t$ are extracted from a large cross-section of macroeconomic variables and a linear combination of forward rates is included as proposed by Cochrane and Piazzesi (2005). We label this structure where forward rates have been pre-processed a "*hybrid* neural network".

The second specification, displayed in Panel (b) of Figure 2.3, *ensembles* two separate networks at the output layer level: one network is trained for the forward rates ("fwd rates net") and one for the macroeconomic variables. In contrast to the specification in Panel (a), this specification allows for a non-linear transformation of forward rates.

The third specification, displayed in Panel (c) of Figure 2.3, entails a collection of networks, one for each group of macroeconomic variables, which are trained in parallel and ensembled at the output layer level. The groups of macroeconomic variables are constructed following the classification provided by McCracken and Ng (2016).[7] This latter specification, which we call "*group ensembling*" switches off the (non-linear) interactions across groups of macroeconomic variables, which are instead present in the specifications of Panels (a) and (b).

To the best of our knowledge, these specifications have not been proposed before in the empirical asset pricing literature within the context of non-linear predictive methods. We study the predictive accuracy of these networks as we vary the number of hidden layers, and the number of nodes per layer.

2.2.3 Estimation Strategy

Following common machine learning practice, we split the data into three sub-samples: a training set used to train the model, a validation set used to evaluate the estimated model on an independent data set, and a testing set, which represents the out-of-sample period in a typical forecasting exercise.[8]

We follow Gu et al. (2020) and, conditional on the estimates from the training set, we produce forecasting errors over the validation sample. We then use the prediction errors over the validation sample to iteratively search the hyperparameters that optimise the objective function. Thus, the validation sample represents the part of data that is used to provide an unbiased evaluation of a model fit. It is trivial to see that predictions in the validation set are not out-of-sample as they are used to tune the model hyperparameters. The third sub-sample, or the testing sample, contains observations that are not used for estimation or tuning. This third sub-sample is known as the "out-of-sample" period and can be used to test the predictive performance on observations yet unseen by the machine learning model.[9]

There are a variety of splitting schemes that could be considered but the

[7]Specifically, we group 128 predictors into eight categories: i) output (16 series); ii) labour market (31 series); iii) housing sector (10 series); iv) orders and inventories (10 series); v) money and credit (14 series); vi) bond and FX and interest rates or financial (22 series); vii) prices or price indices (16 series); and viii) stock market (5 series). Four series in our sample could not be matched to McCracken and Ng (2016) and are left unclassified.

[8]See Chen et al. (2017) for an in-depth discussion of model fragility (i.e. the tendency of a model to over-fit the data in-sample) at the expense of poor out-of-sample performance.

[9]Note, the out-of-sample period is only "pseudo" out-of-sample in the sense that its observations are available to the researcher at the time of the study. Nevertheless, given that the observations in the out-of-sample period have not been used to train the model itself, it is standard in the machine learning literature (in asset pricing) to refer to the testing sample as the "out-of-sample" period (e.g. see Gu et al., 2020; Feng et al., 2018).

trade-off between the size of the training and validation samples is ultimately an empirical question (see Arlot et al., 2010 for a comprehensive survey of cross-validation procedures for model selection). We keep the fraction of data used for training and validation fixed at 85% and 15% of the in-sample data, respectively. The training and the validation samples are consequential. In this respect, we do not cross-validate by randomly selecting independent subsets of data to preserve the time series dependence of both the predictors and the target variables. Forecasts are produced recursively by using an expanding window procedure, that is we re-estimate a given model at each time t and produce out-of-sample forecasts for one-year holding period excess returns.

Figure 2.4 provides a visual representation of the sample splitting scheme we adopt in the empirical analysis. Notice that for some of the methodologies, validation is not required. For instance, neither standard linear regressions nor PCA require a pseudo out-of-sample period to validate the estimates. In these cases, we adopt a traditional separation between in-sample versus out-of-sample period, where the former consists of both the training data and the validation data.

We recursively forecast bond returns in excess of the short-term rate. We focus on one-year holding period excess returns for comparability with the original settings in Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009).

We recursively fit machine learning methods at each time t (i.e. we increase the in-sample period by one monthly observation). This scheme allows us to incorporate the most recent updates from the yield curve, as well as the set of macroeconomic and financial variables.[10] When enlarging the in-sample period, we roll it forward to include the most recent information in a recursive fashion but keep constant the ratio between the training sample and the validation sample. In this respect, we always retain the entire history of the training sample, thus its window size gradually increases. By keeping the proportion of the training and validation sets fixed, the validation sample gradually increases as well.

The result is a sequence of performance evaluation measures that correspond to each recursive estimate. Although computationally expensive, this leverages more information for prediction. In each empirical application, the sample spans from

[10]Note that this is different from the implementation of machine learning methods for stock returns, where trading signals from firm characteristics are often updated once per year, which means that retraining of the models could be performed with lower frequency (Gu et al., 2020).

1971:08 to 2018:12.

Since our forecasting exercise is iterative, the computational challenge becomes sizeable. Thus, we perform all computations on a high performance computing cluster consisting of 84 nodes with 28 cores each, totalling to more than 2,300 cores. Appendix B.2 provides a complete description of the computational specifications.

2.2.4 Statistical Performance

We compare the forecasts obtained from each methodology to a naive prediction based on the historical mean of excess bond returns. In particular, we calculate the out-of-sample predictive R^2 as suggested by Campbell and Thompson (2008). The R^2_{oos} is akin to the in-sample R^2 and is calculated as

$$R^2_{oos} = 1 - \frac{\sum_{t_0=1}^{T-1}\left(xr^{(n)}_{t+1} - \widehat{xr}^{(n)}_{t+1}\left(\mathcal{M}_s\right)\right)^2}{\sum_{t_0=1}^{T-1}\left(xr^{(n)}_{t+1} - \overline{xr}^{(n)}_{t+1}\right)^2}, \tag{2.6}$$

where $\overline{xr}^{(n)}_{t+1}$ is the prediction obtained based on the historical mean and $\widehat{xr}^{(n)}_{t+1}\left(\mathcal{M}_s\right)$ is the forecast of the excess bond returns for maturity n obtained using model $\mathcal{M}_s$, and t_0 is the date of the first prediction. The first forecast error obtains by comparing the excess holding period return during the February 1989 through January 1990 period and its forecast made on January 1989.

We also build a portfolio-level return forecast from the individual maturity forecasts produced by our models. We construct the forecast of an equally-weighted portfolio by $\widehat{xr}^{(EW)}_{t+1} = \frac{1}{6}\sum_{n=2}^{10}\widehat{xr}^{(n)}_{t+1}\left(\mathcal{M}_s\right)$. We compute $R^2_{oos,EW}$ by constructing forecast errors using the realised return $xr^{(EW)}_{t+1} = \frac{1}{6}\sum_{n=2}^{10} xr^{(n)}_{t+1}$ and comparing to the historical mean.

Testing the null hypothesis, $R^2_{oos} \leq 0$, against the alternative hypothesis, $R^2_{oos} > 0$, is tantamount to testing whether the predictive model has a significantly lower mean squared prediction error (MSPE) than the historical average benchmark forecast. Thus, to test whether R^2_{oos} is significantly greater than zero, we implement the MSPE-adjusted Clark and West (2007) statistic.

A limitation of the R^2_{oos} measure is that it does not explicitly account for the risk borne by an investor over the out-of-sample period. To this, end we also calculate realised utility gains for a mean-variance or power utility investor (see Section 2.6).

2.3 Algorithmic Procedures

This section provides the details of the algorithmic procedures for the main forecasting methods introduced in the previous section. The section starts with the linear approaches, including partial least squares and penalised regressions, before moving on to non-linear methods, including tree-based methods and neural networks. Readers familiar with the details of the aforementioned techniques may safely decide to skip this section.

2.3.1 Partial Least Squares

Following the extant practice (see Ch.3.5 Friedman et al., 2001), Partial Least Squares (PLS) is constructed iteratively as a two-step procedure: in the first step we regress excess bond returns on each predictor $j = 1, \ldots, p$ separately and store the regression coefficient ψ_j. The first partial least squares direction is constructed by multiplying the vector of coefficients by the original inputs, that is $\boldsymbol{x}_1 = \boldsymbol{\psi}'\boldsymbol{y}_t$. Hence the construction of $\boldsymbol{x}_1$ is weighted by the strength of the relationship between the excess bond returns and the predictors. In the second step, excess bond returns are regressed onto $\boldsymbol{x}_1$ giving the coefficient θ_1. Then all inputs are orthogonalised with respect to $\boldsymbol{x}_1$. In this manner, PLS produces a sequence of $l < p$ derived inputs (or directions) orthogonal to each other.[11]

Notice that since the response variable is used to extract features of the input data, the solution path of PLS represents a non-linear function of excess bond returns. Stone and Brooks (1990) and Frank and Friedman (1993) show that, unlike PCA which seeks directions that maximise only the variance, the PLS maximises both variance and correlation with the response variable subject to orthogonality conditions across derived components.[12] PLS does not require the calibration of

[11] It is easy to see that for $l = p$ we go back to usual linear least squares estimates similar to PCR.

[12] In particular, the m-th direction solves:

$$\max_{\gamma} \ \mathrm{Corr}^2\left(xr^{(n)}, \boldsymbol{y}\gamma\right) \cdot \mathrm{Var}\left(\boldsymbol{y}\gamma\right)$$
$$\text{subject to} \ \ \|\gamma\| = 1, \ \ \gamma'\Sigma\hat{\psi}_j = 0, \ \ j = 1, \ldots, m-1$$

hyperparameters as the derived input directions are deterministically obtained by the two-step procedure outlined above. In this respect, unlike penalised regressions no shrinkage / regularisation parameters are required to be calibrated.

2.3.2 Penalised Regressions

We present the algorithms utilised to estimate the penalised regression models, i.e. the ridge, lasso and elastic net regressions. To recall, penalised regressions add a penalty term $\phi(\boldsymbol{\beta}; \cdot)$ to the least squares loss function $\mathcal{L}(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)$. The penalty terms in the individual methods are given by

$$\phi(\boldsymbol{\beta}; \cdot) = \begin{cases} \lambda \sum_{j=1}^{p} \beta_j^2 & \text{Ridge regression} \quad (2.7a) \\ \lambda \sum_{j=1}^{p} |\beta_j| & \text{Lasso} \quad (2.7b) \\ \lambda\mu \sum_{j=1}^{p} \beta_j^2 + \frac{\lambda(1-\mu)}{2} \sum_{j=1}^{p} |\beta_j| & \text{Elastic net} \quad (2.7c) \end{cases}$$

Apart from the estimation of $\boldsymbol{\theta}$, we have to determine the level of the shrinkage / regularization parameters λ and μ. Usually, this is achieved by cross-validation, i.e. λ and μ are chosen from a suitably wide range of values by evaluating the pseudo out-of-sample performance of the model on a validation sample and picking the λ, μ that yield the best validation error. In the context of time series forecasts the validation sample should be chosen to respect the time-dependence of the observed data, meaning that the validation sample is chosen to follow upon the training sample used to obtain $\boldsymbol{\theta}$ in time. In the following we discuss in more detail the algorithms that are used to obtain coefficient estimates for the penalised regression models.

In contrast to ridge, which is discussed further below, lasso and elastic net coefficient estimates cannot be obtained analytically because of the L^1 component that enters their respective penalty terms (cf. Eq. (2.7b) and (2.7c)). Hence, we estimate $\boldsymbol{\theta}$ by means of cyclical coordinate descent proposed by Wu et al. (2008) and extended in Friedman et al. (2010). In our exposition of the algorithm of Friedman et al. (2010) we focus on the elastic net case since the lasso case is contained as a special case therein (i.e. by setting $\mu = 0$).

At a high-level coordinate descent can be described as an optimisation method aimed at minimising a loss function one parameter at a time while keeping all other parameters fixed. More precisely, consider the loss function for the elastic net

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - \alpha - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 + \lambda\mu\sum_{j=1}^{p}\beta_j^2 + \frac{\lambda(1-\mu)}{2}\sum_{j=1}^{p}|\beta_j| \tag{2.8}$$

where the factor two in the denominator in front of the sum is introduced to simplify subsequent expressions for the gradient of the loss function. The minimisation of the loss function is unaffected by multiplication with a scalar. Denote by $\mathcal{L}(\boldsymbol{\theta})^{(k)}$ the loss function after the k-th optimisation step. The gradient of the loss function with respect to β_j evaluated at its current estimate $\hat{\beta}_j^{(k)}$ is given by

$$-\frac{1}{N}\sum_{i=1}^{N}x_{ij}(y_i - \alpha - \boldsymbol{x}_i^\top \hat{\beta}) + \lambda(1-\mu)\beta_j + \lambda\mu \tag{2.9}$$

if $\hat{\beta}_j > 0$. A similar expression can be obtained for the case $\hat{\beta}_j < 0$ and $\hat{\beta}_j = 0$ (cf. Friedman et al., 2007). Then, it can be shown that the optimal $\boldsymbol{\beta}$ is obtained by following the Algorithm (1). Commonly, a "warm-start" approach is used to obtain the parameter estimates over the range for λ and μ during cross-validation, meaning that when moving from one set of regularization parameters λ, μ to the next, the prior estimates $\hat{\boldsymbol{\beta}}$ are utilised as initial parameters for the subsequent coordinate descent optimization.

In contrast to lasso and elastic net regressions, the Ridge regression has a closed-form solution given by (e.g. see Friedman et al., 2001, Ch. 3)

$$\hat{\beta}^{\text{Ridge}} = (\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}^\top \boldsymbol{y} \tag{2.10}$$

where $\boldsymbol{X}$ is the input $N \times p$ matrix of p regressors, $\boldsymbol{I}$ is an $N \times N$ identity matrix and $\boldsymbol{y}$ is the vector of dependent variables.

Although there exists an elegant analytical solution to the ridge regression setup, it is common to apply a matrix decomposition technique to circumvent issues incurred by matrix inversion. Thus, we use a singular value decomposition (SVD)

Algorithm 1: Coordinate Descent

Choose initial estimates for $\hat{\alpha} = \bar{y}$ and $\hat{\boldsymbol{\beta}}^{(0)}$ for given λ and μ, where $\bar{y}$ is the unconditional mean of y.
Standardise the inputs x_{ij} such that $\frac{1}{N}\sum_{i=1}^{N} x_{ij} = 0, \frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 = 1$, for $j = 1, \ldots, p$.
Set ϵ to desired convergence threshold
while *there is an improvement in the loss function, i.e.* $|\mathcal{L}(\boldsymbol{\theta})^{(k+1)} - \mathcal{L}(\boldsymbol{\theta})^{(k)}| > \epsilon$ **do**
  for *all predictors* $j = 1, \ldots, p$ **do**
    $\hat{y}_i^{(j)} = \hat{\alpha} + \sum_{l \neq j} x_{il}\hat{\beta}_l$, i.e. the fitted value when omitting the covariate x_{ij}
    $\hat{\beta}_j \leftarrow \frac{S\left(\frac{1}{N}\sum_{i=1}^{N} x_{ij}(y_i - \hat{y}_i^{(j)}), \lambda\mu\right)}{1+(1-\mu)}$, defines the parameter-wise update, where S, the soft-thresholding operator, is given by

$$S(a,b) = \begin{cases} a - b, \text{ if } a > 0 \vee b < |a| \\ a + b, \text{ if } a < 0 \vee b < |a| \\ 0, b \geq a \end{cases}$$

  end
end
Output: Estimates $\hat{\beta}$ for given level of λ, μ;

of the matrix $\boldsymbol{X}$ with the form

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top \tag{2.11}$$

where $\boldsymbol{U}$ is a $N \times N$ orthogonal matrix, $\boldsymbol{V}$ is a $p \times p$ orthogonal matrix and D is an $N \times p$ diagonal matrix containing the singular values of $\boldsymbol{X}$. Then it can be shown that the fitted values are given as

$$\boldsymbol{X}\hat{\beta}^{\text{Ridge}} = \boldsymbol{U}\boldsymbol{D}\left(\boldsymbol{D}^2 + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{D}\boldsymbol{D}^\top\boldsymbol{y}. \tag{2.12}$$

The shrinkage parameter λ is chosen by cross-validation. Alternative estimation approaches such as conjugate gradient descent (Zou and Hastie, 2005) become relevant when $\boldsymbol{X}$ gets larger in dimension.

2.3.3 Tree-Based Methods

Regression trees can approximate any a priori unknown function while keeping the interpretation from a recursive binary tree. However, with more than two inputs, the interpretation is less obvious as trees like the one depicted in Figure 2.1 grow exponentially in size. Nevertheless the algorithmic procedure is equivalent. Suppose one deals with a partition of M regions $\mathcal{A} = \{A_1, \ldots, A_M\}$ of the vector of yields $\boldsymbol{y}_t$ such that

$$g(\boldsymbol{y}_t; N) = \sum_{m=1}^{M} \beta_m \mathbb{I}\left(\boldsymbol{y}_t \in A_m\right) \ .$$

By minimising the sum of squared residuals, one can show that the optimal estimate $\hat{\beta}_m$ is just the average of the excess bond returns in that region, i.e. $\hat{\beta}_m = E\left[xr_{t+1}^{(1)} \middle| \boldsymbol{y}_{1:t} \in A_m\right]$. Finding the optimal partition by using a least squares procedure is generally infeasible, however. We thus follow Friedman (2001) and implement a gradient boosting procedure. Gradient boosting in a tree context boils down to combining several weak trees of shallow depth.

Boosting is a technique for reducing the variance of the model estimates and increasing precision. However, trees are "grown" in an adaptive way to reduce the bias, and thus are not identically distributed. An alternative procedure would be to build a set of *de-correlated* trees which are estimated separately and then averaged out. Such modelling framework is known in the machine learning literature as "Random Forests" (see Breiman, 2001). It is a substantial modification of bagging (or bootstrap aggregation) whereby the outcome of independently drawn processes is averaged to reduce the variance estimates. Bagging implies that the regression trees are identically distributed – as the number of simulated trees increases, the variance of the average estimates depends on the variance of each tree times the correlation among the trees. Random forests aim to minimise the variance of the average estimate by minimising the correlation among the simulated regression trees.

We also consider an extended version of the random forest procedure which is called "Extremely Randomised Trees" (Geurts et al., 2006) . While similar to ordinary random forests in that they still represent an ensemble of individual trees, extreme trees have two main distinguishing features: first, each tree is trained using the whole training sample (rather than a bootstrap sample); and second, the top-down splitting in the tree learner is randomised. That means that instead of computing the optimal cut-point locally for each input variable under consideration, a random cut-point is selected. In other words, with extreme trees the split of the trees is stochastic; with random forests the split is instead deterministic.

Tree-based methods such as Gradient Boosted Regression Trees or Random Forests are essentially modifications of a universal underlying algorithm utilised for the estimation of regression trees, commonly, that is the Classification and Regression Tree (CART) algorithm (Breiman et al., 1984) presented in Algorithm (2).

Algorithm 2: Classification and Regression Trees

Initialise tree $T(D)$ where D denotes the depth; denote by $R_l(d)$ the covariates in branch l at depth d.

for $d = 1, \ldots, D$ **do**

  for $\tilde{R}$ in $\{R_l(d), l = 1, \ldots, 2^{d-1}\}$ **do**

    Given splitting variable j and split point s define regions

$$R_{\text{left}}(j,s) = \{X \mid X_j \leq s, X_j \cap \tilde{R}\} \quad \text{and}$$
$$R_{\text{right}}(j,s) = \{X \mid X_j > s, X_j \cap \tilde{R}\}$$

    In the splitting regions set

$$c_{\text{left}}(j,s) \leftarrow \frac{1}{|R_{\text{left}(j,s)}|} \sum_{x_i \in R_{\text{left}(j,s)}} y_i(x_i)$$
$$c_{\text{right}}(j,s) \leftarrow \frac{1}{|R_{\text{right}(j,s)}|} \sum_{x_i \in R_{\text{right}(j,s)}} y_i(x_i)$$

    Find j^*, s^* that optimise

$$j^*, s^* = \underset{j,s}{\text{argmin}} \Big[\sum_{x_i \in R_{\text{left}(j,s)}} (y_i - c_{\text{left}}(j,s))^2 + \sum_{x_i \in R_{\text{right}(j,s)}} (y_i - c_{\text{right}}(j,s))^2 \Big]$$

    Set the new partitions

$$R_{2l}(d) \leftarrow R_{\text{right}}(j^*, s^*) \quad \text{and} \quad R_{2l-1}(d) \leftarrow R_{\text{left}}(j^*, s^*)$$

  end

end

Output: A fully grown regression tree T of depth D. The output is given by

$$f(x_i) = \sum_{k=1}^{2^L} \text{avg}(y_i \mid x_i \in R_k(D)) \mathbf{1}_{\{x \in R_k(D)\}},$$

i.e. the average response in each region R_k at depth D.

Next, we present the Algorithm (3) used to populate random forests as suggested by Breiman (2001). Random Forests consist of trees populated following an algorithm like CART, but randomly select a sub-set of predictors from the original data. In this manner, the individual trees in the forest are de-correlated and overall predictive performance relative to a single tree is increased. The hyperparameters to be determined by cross-validation include first and foremost the number of trees in the forest, the depth of the individual trees and the size of the randomly selected sub-set of predictors. Generally, larger forests tend to produce better forecasts in terms of predictive accuracy. Finally, Algorithm (4) delivers the gradient boosted regression tree (GBRT) (Friedman, 2001). GBRTs are based on the idea of combining the forecasts of several weak learners. The GBRT comprises of trees of shallow depth that produce weak predictions stand-alone, however, tend to deliver powerful forecasts when aggregated adequately.

Algorithm 3: Random Forest

Determine forest size F

for $t = 1, \ldots, F$ **do**

  Obtain bootstrap sample Z from original data.

  Grow full trees following Algorithm (2) with the following adjustments:

  1. Select $\tilde{p}$ variables from the original set of p variables.
  2. Choose the best combination (j, s) (cf. Algorithm (2)) from $\tilde{p}$ variables
  3. Create the two daughter nodes

  Denote the obtained tree by T_t

end

Output: Ensemble of F many trees. The output is the average over the trees in the forest given as

$$f(x_i) = \frac{1}{F} \sum_{t=1}^{F} T_t(x_i)$$

2.3.4 Neural Networks

A commonly used algorithm to fit neural networks is stochastic gradient descent (SGD). For this paper we make use of a modified form of gradient descent by adding Nesterov momentum (Nesterov, 1983). In comparison to plain SGD which is often affected by oscillations between local minima, Nesterov momentum (also known as Nesterov accelerated gradient) accelerates SGD in the relevant direction. Algorithm (5) outlines the procedure. It is best practice to initialise neural network parameters with zero mean and unit variance or variations thereof such as He et al. (2015) like we do in this paper. Over the course of the training process this normalisation vanishes and a problem referred to as covariate shift occurs. Thus, we apply batch normalisation (Ioffe and Szegedy, 2015) to the activations after the last ReLU layer.

Algorithm 4: Gradient Boosted Regression Trees

Initialise a gradient boosted regression tree $f_0(x) = 0$ and determine number of learners F. Let $\mathcal{L}(y, f(x))$ be the loss function associated with tree output $f(x)$.

for $t = 1, \ldots, F$ **do**

  for $i = 1, \ldots, N$ **do**

    Compute negative gradient of loss function evaluated for current state of regressor $f = f_{t-1}$

$$r_{it} = -\frac{\partial \mathcal{L}(y_i, f_{t-1}(x_i))}{\partial f_{t-1}(x_i)}.$$

  end

  Using the just obtained gradients grow a tree of depth D (commonly, $D \ll p$ where p is the number of predictors) on the original data replacing the dependent variable with $\{r_{it}, \forall i\}$. Denote the resulting predictor as $g_t(x)$.

  Update the learner f_t by

$$f_t(x) \leftarrow f_{t-1}(x) + \nu g_t(x)$$

  where $\nu \in (0, 1]$ is a hyperparameter.

end

Output: $f_F(x)$ is the gradient boosted regression tree output.

Algorithm 5: Stochastic Gradient Descent with Nesterov Momentum

Initialise the vector of neural network parameters θ_0 and choose momentum parameter γ. Determine the learning rate η and set $v_0 = 0$.

while *No convergence of* θ_t **do**

  $t \leftarrow t + 1$

  $v_t = \gamma v_{t-1} + \eta \nabla_{\theta_t} \mathcal{L}(\theta_t)$

  $\theta_t \leftarrow \theta_{t-1} - v_t$

end

Output: The parameter vector θ_t of the trained network.

Batch normalisation reduces the amount of variability of predictors by adjusting and scaling the activations. This increases the stability of the neural network and the speed of training. The output of a previous activation is normalised by subtracting the batch mean and dividing by the batch standard deviation. This is particularly advantageous if layers without activation functions, i.e. the output layer, follow layers with non-linear activations, such as ReLU, which tend to change the distributions of the activations. Since the normalisation is applied to each individual mini-batch, the procedure is referred to as batch normalisation. The SGD optimisation remains largely unaffected; in fact, by using batch normalisation the structure of the weights is more parsimonious. Algorithm (6) outlines the procedure from the original paper.

Algorithm 6: Batch Normalisation per mini-batch

Let $\mathcal{B} = x_{1,\ldots,m}$ be a mini-batch of batch-size m. Set parameters λ, β.
$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$
$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2$
$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$
$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i)$
Output: The normalised mini-batch $\mathrm{BN}_{\gamma,\beta}(x_i)$.

Algorithm (7) presents the early stopping procedure that is used to abort the training process early when the loss on the validation sample has not improved for a specific number of consecutive iterations. Early stopping is used to improve the performance of the trained models and reduce over-fitting. By evaluating the validation error, it prevents the training procedure from simply memorising the training data (see Bishop, 1995 and Goodfellow et al., 2016). More specifically, by means of early stopping the training process is stopped prematurely if the loss on the validation sample has not improved for a number of consecutive epochs. In detail, our algorithm is stopped early if any of the following is true: maximum number of epochs reached the value of 1000, gradient of loss function falls below a specified threshold, or the MSE on validation set has not improved for 20 consecutive epochs. When early stopping occurs, we retrieve the model with the best validation performance. Early stopping has two effects. Firstly, early stopping prevents over-fitting by aborting the training when the pseudo out-of-sample performance starts to deteriorate, hence it reduces over-fitting. Secondly, since the optimal number of weight updates is unknown initially, early stopping helps to keep the computational cost at a minimum by potentially stopping the training far before the maximum number of iterations is reached.

Algorithm 7: Early Stopping

Initialise the validation error $\epsilon = \text{inf}$ and define a patience ϕ, also set $j = 0$
while $j < \phi$ **do**
  Update θ using Algorithm (5) to get $\theta^{(j)}$, i.e. the parameter vector at iteration j
  Compute loss function on validation sample $\epsilon' = \mathcal{L}_{val}(\theta; \cdot)$ **if** $\epsilon' > \epsilon$ **then**
    $j \leftarrow j + 1$
  end
  else
    $j \leftarrow 0$
    $\epsilon \leftarrow \epsilon'$
    $\theta' = \theta$
  end
end
Output: The early-stopping optimised parameter vector θ'

Finally, it is important to highlight that we use a form of forecast averaging / ensembling, i.e. we train multiple copies of networks with different seeds for the randomly drawn initial network weights. Using fixed seeds will in general lead to replicable results. Nevertheless, different seeds will produce different forecasts as discussed also in Gu et al. (2020). Therefore, in order to reduce prediction variance, we average over forecasts from networks initialised with different seeds. To be precise, for each time t we initialise 100 models with different but fixed seeds. The 100 models are then trained and as part of the training we obtain the validation sample loss. The validation sample loss is then used to select the 10 out 100 models with the smallest validation sample error. Finally, we average the forecasts of those 10 in-sample best performing models.

2.4 An Empirical Study of U.S. Treasury Bonds

2.4.1 Bond Return Predictability and the Yield Curve

We start by forecasting the excess returns of Treasury bonds with the yield curve. In this case, the classical specification is given by the principal component regression (PCR) in Equation (2.3).

In words, excess returns are regressed on PCs of the Treasury term structure, i.e. $\boldsymbol{x}_t = [PC_{1,t} \ldots, PC_{k,t}]$. We use the first three, five, or ten PCs. The case with ten PCs essentially corresponds to the setting in Cochrane and Piazzesi (2005), where excess returns are regressed on a linear combination of short-rate, $y_t^{(1)}$, and nine forward rates for loans between $t+n-1$ and $t+n$, $f_t^{(n)}, n = 2, \ldots, 10$. Table 2.1 displays the out-of-sample R^2_{oos} (and its p-value) for different bond maturities; we also report the R^2_{oos} for an equally-weighted portfolio.

Panel A of Table 2.1 displays the results for PCRs and partial least squares (PLS). The first three rows show the predictive performance of PCRs for $k = 3, 5$, and 10 PCs. The predictive R^2 are negative across different maturities. A parsimonious representation with only three PCs significantly outperforms the specification with five and ten PCs, particularly at long maturities. Further, adding simple forms of non-linearities such as squared PCs worsens the performance. Perhaps surprisingly, a linear supervised learning method like PLS does not lead to any improvement relative to PCR.

Panel B of Table 2.1 displays the results from various configurations of the linear penalised regressions. Ridge regression performs poorly out-of-sample with predictive R^2_{oos} that are mostly negative across bond maturities. The second and third rows of Panel B show that sparse modelling improves the forecasting performance of the current term structure relative to ridge: the R^2_{oos} for both the lasso and elastic net are positive for maturities greater than four years, as well as for the equally-weighted bond portfolio. However, the performance of elastic net is on par to that obtained from a principal component regression with three PCs, which proves to be a tough benchmark.

Panel C of Table 2.1 shows the results for boosted regression trees, random forests, extremely randomised trees, and neural networks. All these methods attain good performance with significantly positive R^2_{oos} across maturities. With respect to trees, the randomisation of the feature split locations (i.e. for extreme trees) turns out to improve the out-of-sample performance over random forests, particularly for long maturities. Turning to NNs, we observe that a shallow network with a single hidden layer and three nodes (cf. Figure 2.2) performs on par with the best, deeper network with two hidden layers and seven nodes. Interestingly, further increasing the depth of the network deteriorates its performance. This continues to be the case even when we consider alternative structures, like an NN with three hidden layers and pyramidal node architecture.

The last row of Panel C presents the results of an interesting case. In their paper, Cochrane and Piazzesi (2005) conclude that lags of forward rates (dated $t-1$ and earlier) contain information about excess returns that is not spanned by month t forward rates. They note that this result is inconsistent with the logic that the time-t term structure contains all information relevant to forecasting future yields and excess returns (cf. Equation (2.2)). We therefore ask whether the flexibility of a NN can help reconcile the theoretical assumption that yields at time t already incorporate all information about the term structure that is needed to understand bond risk premia. To this end, the last row in Panel C reports the results obtained by feeding the NN with the ten forward rates at time t and lagged forward rates from time $t - 11$ to $t - 1$. By comparing the last row to the NN with one hidden layer and three nodes, we find no evidence that we can improve upon a NN that uses just the month-t forward rates.[13]

The evidence in Table 2.1 confirms that, even in a small dimensional setting using only information in the yield curve, we can improve bond risk premia measurement by acknowledging that (1) the function $g(\boldsymbol{y}_t; N)$ in Equation (2.2) can be non-linear, and that (2) such improvement depends on the neural network specification; a shallow NN with one hidden layer performs on par with a network with two layers, but deeper networks worsen the performance. Finally, consistent with Equation (2.2), we find that lagged values of the yield curve cannot improve the forecast obtained using just the time-t term structure when we account for non-linearities.

2.4.2 Bond Return Predictability and Macroeconomic Variables

Next, we consider the set-up where information embedded in the yield curve does not necessarily subsume information contained in macro variables. In this case, the classical specification is given by Equation (2.4), where the factors F_t now have the potential to serve as the model's state vector beyond yields only. To ensure comparability with the literature, we adopt the specification proposed by Ludvigson and Ng (2009), whereby $\boldsymbol{F}_t$ is a subset of the first eight PCs extracted from a large cross-section of macroeconomic variables and $\boldsymbol{x}_t$ represents a linear combination of forward rates as proposed by Cochrane and Piazzesi (2005), a.k.a. the CP factor.

[13]We report only the best specification with lagged forward rates (i.e. a shallow NN with seven nodes). NNs with more layers or a different number of nodes underperform.

Panel A in Table 2.2 displays results from a simple principal component regression with eight PCs, the specification proposed in Equation (8) of Ludvigson and Ng (2009) (i.e. $\boldsymbol{F}_t = \left(F_{1t}, F_{1t}^3, F_{3t}, F_{4t}, F_{8t}\right)$), and partial least squares (PLS). Panel B shows the results from two alternative implementations of sparse and regularised linear regressions. In the first implementation ("using CP factor"), we employ the CP factor as an additional regressor; this specification ensures a closer comparability to Ludvigson and Ng (2009). In the second implementation ("using fwd rates directly"), we treat the whole set of forward rates as additional regressors with respect to macroeconomic variables.

The results in Panels A and B shows that: (1) dense modelling, such as data compression techniques and ridge regression, tends to perform poorly out-of-sample; and (2) sparse modelling with both regularization and shrinkage (i.e. elastic net regressions), perform well, particularly when restricting the linear combination of forward rates. Comparing Panel B in Table 2.1 to that in Table 2.2, it seems apparent that there is information beyond the term structure of interest rates that can be used to predict bond returns.

Turning to non-linear machine learning methods, Panel C in Table 2.2 shows the results from the three alternative network specifications discussed in Subsection 2.2.2: (1) a "*hybrid*" framework in which the forward rates enter linearly as additional predictors in the output layer ("fwd rates direct"; Figure 2.3, Panel (a)); (2) a specification that ensembles one network for the forward rates ("fwd rates net") and one for the macroeconomic variables (Figure 2.3, Panel (b)); and (3) a specification that entails a collection of networks, one for each group of macroeconomic variables (Figure 2.3, Panel (c)). This latter specification, dubbed "*group ensembling*", switches off the (non-linear) interactions across groups of macroeconomic variables, which are present in the hybrid network, as well as in the specification that ensembles separately forwards and macroeconomic variables.

The performance of hybrid networks stands out. Interestingly, and differently from Table 2.1, increasing the depth of the NN from one- to three-layers improves its accuracy.

However, a careful choice of network structure based on prior economic information exerts a great impact on performance. In particular, a one-layer *group ensembled* model (see third-to-last row) performs on par with the three-layer *hybrid*

NN for short- and medium-term maturities, and attains the highest predictive accuracy for the 7- and 10-year bonds. Interestingly, adding more layers is detrimental to the performance of the group ensembled NN.

Panel C in Table 2.2 also shows that the performances of boosted regression trees, random forests, and extreme trees improve substantially when using a large panel of macroeconomic information. In fact, the extreme trees perform better than shallow (hybrid and ensembled) NNs but worse than the (best performing) one-layer NN with group ensembling.

The results in Table 2.2 show that macroeconomic variables carry information that is not contained in the yield curve. When we compare the best performing (one layer and three nodes) NN in Table 2.1 to the 1-layer group ensembled NN, we observe approximately a 10 percentage point increase in R^2_{oos} for each maturity. The results also show that the depth and structure of the network interact with one another: having a separate network for each group of macroeconomic variables compensates for the need of a deep NN when macroeconomic variables are processed together without further classification.

2.4.3 Bond Return Predictability in Monthly Holding Periods

Bollerslev et al. (2009) show that short-horizon stock returns are very hard to predict. Similarly, the monthly predictability of bond excess returns tends to be much lower than predictability of annual holding period returns (see, e.g. Gargano et al., 2019). To address this issue we now focus on the predictability of monthly holding period returns in this section. Table 2.3 reports the results for the sample period from 1964:10 to 2016:12.[14]

We observe that the ranking displayed in Tables 2.1 is largely preserved. That is, the performance (averaged across maturities) of non-linear models, specifically NNs, is higher than the one obtained by data compression methods and penalised linear regressions. For instance, at longer maturities neural networks attain an R^2_{oos} which is up to 50% bigger than the parsimonious PCA with three factors. The statistical results from Tables 2.1 and 2.3 provide evidence that non-linear models

[14]The test assets are the Center for Research in Security Prices Treasury Fama bond portfolios. These very same assets have been used, e.g. by Duffee (2002). Only non callable, non flower notes and bonds are included in the portfolios. The portfolio returns are an equal-weighted average of the unadjusted holding period return for each bond in the portfolios in excess of the risk-free rate. We consider four buckets: bonds with maturity of 1-2 years, 2-3 years, 3-4 years, and 4-5 years.

are able to detect substantial bond return predictability with the R^2_{oos}s for NN models are in positive territory for both monthly and annual holding period returns.

For the case in which macroeconomic information is used in addition to the yield curve, Table 2.4 reports the performance in monthly holding period. The results obtained for annual bond excess returns are largely confirmed. In particular, Panel C shows that imposing a sensible economic structure through group ensembling significantly increases the out-of-sample predictive performance of the (two-layer hybrid) NNs. I.e., grouping variables within economic categories and then training a shallow network within each group – a network structure that we dub "group ensembling"– attains a performance that is better than the best-performing deep NN where no economic priors are utilised. Thus, the depth of the network and the economic priors used to design the network (e.g. grouping within categories) interact with one another.

Further, comparing Table 2.4 to Table 2.3, we find that a group ensemble NN that exploits macroeconomic and term structure information attains out-of-sample R^2s that are more than three times larger than the best-performing NN that employs yields only for maturities ranging from two to five years.

2.5 Dissecting Predictability

2.5.1 Bond Return Predictability in Expansions and Recessions

We start by investigating whether bond return predictability varies over the economic cycle. To this end, we split the data into recession and expansion periods using the NBER recession indicator.

Table 2.5 shows the R^2_{oos} values computed separately for the recession and expansion sub-samples. For yields-only PCA, we recover a classical result: predictability is concentrated in economic recessions (in particular for long maturity bonds) and is absent during expansions.

Turning to NNs, we continue to observe R^2_{oos} that are generally higher during recessions than in expansions. However, the difference in R^2_{oos} values decreases with bond maturity. More importantly, a formal test confirms that the bond return prediction from NNs is statistically different from that of the expectations hypothesis (EH) model both in expansion and recession. In contrast to NNs, the return predictability implied by trees is actually stronger during expansions. A formal test confirms that the predictive accuracy of trees is significantly better than that generated by the EH benchmark only for expansion periods. Finally, a pairwise test (untabulated) confirms that the improvement of NNs over trees is mainly due to the better predictive accuracy of networks in recessionary periods; however, the predictions from trees and NNs are indistinguishable in expansions.

Our finding that the predictability of bond returns implied by machine learning methods is not concentrated exclusively in bad times, but is present also in expansions is novel to the literature and contrasts with evidence for equities (Rapach et al., 2010; Dangl and Halling, 2012) and bonds (Gargano et al., 2019). In Section 2.5.2, we analyse the models' performance in different periods using the cumulative sum of squared errors and confirm that the out-performance of machine learning-based forecasts versus the EH benchmark is not concentrated in isolated events.

Interestingly, however, it is possible to relate, ex post, NN forecasts to specific patterns of the yield curve. Table 2.6 shows that NNs, and in particular the group ensemble network that exploits macroeconomic and financial information in addition to interest rates, predict high excess bond returns when there is a steep slope in the yield curve (e.g. right after recessions) and when the level of the yield curve is high. Further, these NN predictions (conditional on specific shapes of the yield curve) are highly correlated with realised returns, thus leading to high R^2s.

2.5.2 Bond Return Predictability and Cumulative SSE

To identify the periods in which the models perform best, we follow Welch and Goyal (2008) and compute the difference in the cumulative sum of squared errors (SSE) for the EH model versus the machine learning model of interest, $\Delta CumSSE$. Positive and increasing values of $\Delta CumSSE$ suggest that the model under consideration generates more accurate point forecasts than the EH benchmark.

Figure 2.5 plot the $\Delta CumSSE$ for the best performing regression tree specification, i.e. extreme tree (Panels (a) and (c)), and for the best performing neural network, namely the *NN 1 Layer (3 nodes)* – when forecasting with only the forward rates (Panel (b)) – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables (Panel (d)) – (see Tables 2.1-2.2 for reference). We focus on the ten-year bond maturity. In general, the plots show that the various machine learning models perform well relative to the EH model as manifested by lines that are increasing for most periods. Indeed, only when we consider exclusively yields-based predictors, we then observe occurrence of decreasing graphs (i.e. periods where the models underperform against this benchmark). However, these are few isolated occasions (e.g. around 2001 and 2008). Interestingly, adding macroeconomic variables improves the predictive accuracy of the models relative to the benchmark. Comparing panel (c) to (d), we note that: (1) the performance of extreme trees is particularly effective in few, rather prolonged, periods, namely 1992 to 1996 and the aftermath of the financial crisis; (2) the performance of the NN with group ensembling is instead characterised by an almost steady improvement in performance.

Another interesting aspect to investigate is whether expected bond returns could have evolved differently in response to availability of technology.[15] However, referring to our plots of out-of-sample R^2 over time, even in later years when adoption of neural networks spread, we still notice the outperformance of neural networks vis-a-vis the expectation hypothesis.

Overall, we take these patterns as rather reassuring that the value-added through machine learning based forecasts is rather pervasive and not concentrated in isolated events.

2.5.3 Understanding the Performance of Neural Networks: Level, Slope or Both?

In this subsection we provide an heuristic interpretation of the performance of the NNs based on the Campbell and Shiller (1991) accounting identity (Equation (2.1)). Such identity posits that the forecasts of future yields (or their PCs) using current yields are necessarily also forecasts of expected log returns to bonds.

[15] The theory necessary to apply our networks was available in the 1990s when our out-of-sample period starts. The overarching concept of back-propagation in multi-layer neural networks was introduced in Werbos (1974). The use of automatic differentiation (AD) that is necessary for fast computation of gradients during back-propagation was used in Werbos (1982). An early example of a study using neural networks for prediction in a finance context, on gas markets to be precise, with similar models as used by us is Werbos (1988) and others exist before 1990. What is less clear is whether the necessary computational power was available to estimate the models in a reasonable amount of time. Also, while the most fundamental building blocks were in place in the late 1980s, regularisation concepts such as dropout (Srivastava et al., 2014) and batch normalization (Ioffe and Szegedy, 2015) were only introduced later.

Thus, we investigate the ability of the latent factors extracted by the NN to predict the year-on-year changes in the first three PCs extracted from the cross-section of forward rates. We denote these principal components as $PC_{1,t}, PC_{2,t}$, and $PC_{3,t}$. The auxiliary forecasting regressions are:

$$PC_{i,t+1} - PC_{i,t} = b_0 + \boldsymbol{b}_1^\top \mathcal{P}_t + \boldsymbol{b}_2^\top \boldsymbol{x}_t + \epsilon_{i,t+1} \qquad \text{for} \quad i = 1, 2, 3,$$

where we stack the first three PCs of the term structure in the vector $\mathcal{P}_t$ and denote by $\boldsymbol{x}_t$ the hidden factors extracted by the NN.[16]

Table 2.7 reports the in-sample R^2 of such predictive regressions. The first row provides the benchmark results based on the sole vector $\mathcal{P}_t$. In support of Duffee (2011b, 2013), we find weak evidence that changes in the first PC (level) are forecastable ($R^2 = 9.28\%$ being the lowest), whereas the slope and curvature are unquestionably forecastable with 21.66% (48.70%) of the variation in slope (curvature) that is predictable.

Next, we add to the regression the hidden factors extracted from the two best performing NNs in Table 2.1-2.2: the *NN 1 Layer (3 nodes)* – when forecasting only with the forward rates – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables. We observe that the factors extracted from NNs that use yields-only (second row in Table 2.7) contribute substantially to the predictability of the level and curvature. On the other hand, the statistical evidence for slope forecasts - after controlling for the standard three principal components - is weak. We conclude that there is substantial information in the time-t term structure not only about future values of slope, but also about the level (and curvature). Standard PCs are not entirely able to extract all the relevant information about the level. A shallow NN is successful in extracting such information about the future level of the curve, an information which leads to excess returns being more predictable out-of-sample.

The last row in Table 2.7 focuses on factors extracted from a NN that exploits macroeconomic variables in addition to forward rates. We observe that the factors extracted from the group ensembled NN not only contribute to the ability to predict

[16] Take a shallow network with $L = 1$ hidden layers as an example, i.e.: $E_t\left[xr_{t+1}^{(n)}\right] = \hat{\alpha}_n + \hat{\boldsymbol{\beta}}_n^\top \boldsymbol{x}_t$, where, $\boldsymbol{x}_t = h\left(\boldsymbol{W}\boldsymbol{y}_t + b\right)$. The latent factor $\boldsymbol{x}_t$ is extracted at each time t conditional upon estimates of the weights $\boldsymbol{W}$ and bias b from the vector of inputs y_t.

the level of the yield curve, but also the slope. This suggests that the slope of the yield curve is related to the state of the economy, and our NN is able to extract the relevant information from our large set of macroeconomic variables.

2.5.4 Relative Importance of Macroeconomic Variables

In this subsection, we investigate which variables drive the performance of NNs studied in Tables 2.1-2.2. To this end, we examine the marginal relevance of single variables based on the partial derivative of the target variable, $xr_{t+1}^{(n)}$, with respect to each input, where the gradient is evaluated at the in-sample mean value of the input, i.e.

$$\mathbb{E}\left[\frac{\partial}{\partial y_{it}} xr_{t+1}^{(n)} \middle| y_{it} = \overline{y}_i\right], \tag{2.13}$$

where $\overline{y}_i$ represents the in-sample mean of the input variable i. The partial derivative represents the sensitivity of the output to the ith input evaluated at its sample mean, conditional on the network structure and the average value of the other input variables (see Dimopoulos et al., 1995). Further, we focus on the magnitude of the gradients by taking their absolute values.

Figure 2.6 shows the relative importance of each input variable based on the gradient in Equation (2.13). The analysis is carried out for the best-performing NN in Table 2.2, the NN with one hidden layer where macroeconomic variables and forward rates are modelled separately through ensembling at the output layer level. For ease of exposition, we report only the top 20 most relevant predictors. Panels (a) and (b) display results for the 2- and 10-year bond maturities, respectively. To gain further intuition about the systematic patterns in the drivers of expected excess bond returns, we also calculate the relative importance from the absolute gradients averaged for each class of input variables as labelled in McCracken and Ng (2016). The results, displayed in Panels (c) and (d), provide an indication of which economic category dominates.

Comparing Panel (c) to (d) in Figure 2.6, we observe that the variables pertaining to inflation, and money and credit are important independently from the maturity considered. However, the results also show that the effect of other classes of predictors is heterogeneous over the term structure. For instance, variables related to the stock and labour market are more important for the short-end of the yield curve, while variables pertaining to the categories output & income and orders & inventories become more relevant for the long-end of the curve.

This analysis is important for two reasons. First, it suggests that inflation has a level-like effect on bond yields, whereas variables pertaining to the labour market (order & inventories) are likely to have a slope effect acting mostly on short-term (long-term) bonds, while leaving long-term (short-term) bonds unaffected. This evidence can therefore provide guidance for theoretical models that include macro risk factors as drivers of bond risk premia by highlighting their permanent or transient nature. Second, our analysis suggests that the use of excess bond returns averaged across maturities is unlikely to flesh out the true impact of macro risk factors on bond risk premia.[17]

In all, our results show that there is information in macroeconomic and financial variables beyond that conveyed by the yield curve, and this information improves the predictions of bond returns (Tables 2.1-2.2). In addition, the type of unspanned (by the yield curve) information may vary across different bond maturities. To our knowledge, this fact is novel and provides a new angle to revisit a central question in the term structure literature, which is whether yields data contain all the relevant information to predict future bond returns.

2.5.5 Interactions Within or Across Categories?

The finding that a shallow group ensembled network performs on par with (or better than) a deep, three layer NN that models all macroeconomic and financial variables together is novel to the literature on machine learning and asset prices, and is important for two reasons. First, our results on group ensembled NNs show that the depth of the network and the economic priors used to design it (e.g. grouping variables that pertain to the same category) interact with one another. In particular, for the application at hand, group ensembling can compensate for the depth of the network. Second, our results highlight what type of non-linearities are important from

[17] Our evidence of a level-like effect of inflation on bond yields is in line with the analysis in Joslin et al. (2014, Section VI). Furthermore, within the orders & inventories category, we find that "New Orders for Durable Goods" is of single-most importance for forecasts at the far end of the curve (see Panel (b) in Figure 2.6). Yang (2011) provides theoretical and empirical evidence that is consistent with our finding by showing that the impact of durable consumption growth on the yield curve strengthens with bond maturity.

an economic perspective: Is it the interaction of many variables (across categories) or a higher polynomial of the same variable (within a category)? Since our group ensembled network switches off interactions across categories, our results show that it is the non-linearity within a group that drives the outperformance of the network. Of course, this is true only in so far as cross-group network weights are not already small in the fully connected network.

Table 2.8 shows that this is indeed not the case. More precisely, we calculate the second order derivatives of the output with respect to each input conditional on the inputs being in different groups of predictors.[18] We estimate cross-group partial derivatives for both a fully connected network and a network with group-ensembling.

The results in Panel A show that the absolute value of the sum of interaction derivatives in the fully connected network is orders of magnitude larger than the value obtained from a group-ensembled NN (i.e. the cross-group interactions are indeed large in the fully connected network). In Panel B of Table 2.8, we provide the within-group second order partial derivatives of the outputs with respect to the inputs conditional on being in the same group. We find that the magnitude of the within-group effects is similar between the fully connected and group-ensembled NNs. Hence, the performance of the group-ensembled NN is driven by imposing the absence of interactions across categories while allowing for non-linearity within an economic category.

2.5.6 Model Uncertainty

Faced with multiple NN estimates, the question of how to best exploit ex-ante different forecasting specifications immediately arises. In particular, should we rely on a single, ex post, dominant model specification or should a combination of different forecasts be used to produce a better forecast? From a pure theoretical perspective, unless the best forecasting model can be identified ex-ante, forecast combinations may offer some diversification benefits (see Clemen, 1989, for a discussion). However, it may also be the case that a carefully designed validation procedure is able

[18]That is, we calculate:

$$\mathbb{E}\left[\frac{\partial^2}{\partial y_i \partial y_j} xr_{t+1}^{(n)} \middle| y_i \in G_A, y_j \in G_B\right], \tag{2.14}$$

where G_A and G_B are two non-overlapping groups of variables defined as in McCracken and Ng (2016), and we sum the absolute value of Equation (2.14) for each interaction of variables that do not belong to the same group.

to systematically pick the best out-of-sample model specification.

To answer this question, we first compare the best-performing NN within the context of forecasting bond returns with both yields and macroeconomic variables – the *NN 1 Layer Group Ensem + fwd rate net* (see Table 2.2) – against a combined forecast of the form,

$$\hat{xr}^{(n)}_{c,t+1} = \sum_{i=1}^{\mathcal{M}} \omega_{i,t} \cdot \hat{xr}^{(n)}_{i,t+1} \tag{2.15}$$

where $\hat{xr}^{(n)}_{c,t+1}$ denotes the one-step ahead combined forecast for maturity n, $\omega_{i,t}$ is the weight assigned to each individual prediction, $\hat{xr}^{(n)}_{i,t+1}$, and $i = 1, \dots, \mathcal{M}$ are the forecasts from the set of NNs $\mathcal{M}$ in Table 2.2. We choose two representative model combination schemes: (1) an equal weight assigned to each forecast, i.e. $\omega_{i,t} = 1/\mathcal{M}$, and (2) a linear combination of forecasts based on the validation losses, i.e. $\omega_{i,t} = \frac{1/L(e_{i,t}|\theta_i)}{\sum_{i=1}^{\mathcal{M}}(1/L(e_{i,t}|\theta_i))}$, where $L\left(e_{i,t}|\theta_i\right)$ is the validation loss obtained from the cross-validation prediction error $e_{i,t}$ given the network hyperparameters θ_i.[19]

In addition to an equal-weight and a relative-performance combination scheme, we also compare our best-performing NN forecasts against a full-blown cross-validated network. In particular, we expand the set of hyperparameters that are cross-validated and selected every five years; that is, we let optimisation procedures not only select the dropout rate and the L1/L2 penalties, but also the number of hidden layers, the nodes per group of macroeconomic variables, and the nodes in the forward rate network (see Appendix Table B.1 for details on the hyperparameters).[20]

The logic for comparing our best-performing model against two representative forecast combination schemes and a full-blown cross-validated network is to make sure that our results are robust to more flexible and adaptive modelling strategies. Table 2.9 reports the results. Two interesting aspects emerge from the table. First, the group-ensembled NN (see first row) outperforms both forecast combination schemes (second and third rows), the sole exception being at the two-year maturity. Second, our group-ensemble network specification tends also to perform on par with, or better than, the full-blown cross-validated network for maturities

[19]Note that the loss function we use is a simple mean squared error plus a penalty to induce regularisation in the weights. This means that the weighting scheme reflects the performance of each model relative to the performance of the average model (e.g. Bates and Granger, 1969; Newbold and Granger, 1974; Stock and Watson, 1998; Elliott and Timmermann, 2004).

[20]We thank an anonymous referee for suggesting this exercise.

greater than three years. Hence, we conclude that the optimal structure of layers and nodes, which is endogenously chosen through an adaptive cross-validation exercise, does not improve (except for the very short-end) upon a more parsimonious and economically motivated network structure, like our one-layer group-ensemble NN.

We next examine the recursive performance, meaning cross-validation error, of the top performing neural network. In principle, the performance of the *NN 1 Layer Group Ensem + fwd rate net* specification could be justified by the fact that such network is consistently chosen through cross-validation and across time. In Table 2.10, we compare how often the four on average best-performing NN structures from Table 2.2 are selected throughout the out-of-sample period. Two interesting facts emerge. First, our best-performing group-ensembled NN generates the smallest validation error (and thus it would be chosen through cross-validation) for about a half of the out-of-sample period. This could explain the similar performance between the best-performing (group-ensemble) NN in Table 2.2 and the full-blown cross-validated model, which contains the benchmarking specification in the model set. Second, shallow NNs tend to consistently deliver lower validation errors. This reinforces our result that network depth and structure interact with one another: in fact, a carefully designed network outperforms a deeper, and more data-driven, network structure.

2.6 Economic Value of Excess Bond Return Forecasts

So far, our analysis concentrated on statistical measures of predictive accuracy. Next we evaluate whether the apparent gains in predictive accuracy translate into better investment performance relative to the no-predictability alternative. This is important since Thornton and Valente (2012) find that yield-based predictors, when used to guide the investment decisions of an investor with mean-variance preferences, do not lead to higher out-of-sample Sharpe ratios compared with investments based on a no-predictability expectations hypothesis (EH) model. Sarno et al. (2016) reach a similar conclusion. However, the large time variation in expected bond returns that is detectable in real time by machine learning methods naturally calls for revisiting these findings.

2.6.1 The Asset Allocation Framework

In order to assess the economic importance of machine learning methods (particularly trees and NNs) in forecasting bond returns, we use a classic portfolio choice problem (Della Corte et al., 2008; Thornton and Valente, 2012). Specifically, we consider an investor who optimally invests in a portfolio comprising $K+1$ bonds: a risk-free one-period bond and K risky n-period bonds.

We consider both univariate and multivariate asset allocation exercises. In the univariate case, the investor selects between an n-year bond and the risk-free return based on the expected return implied by a given model. We focus on the results for $n=2$ and $n=10$ years. In the joint asset allocation exercise, the investor selects bonds with maturities of two- to ten-years, and the risk-free return. We present the asset allocation decisions of a mean-variance and power utility investor.

Mean-Variance Utility Investor At each time t, the decision-maker selects the weights on the risky n-period bonds $\boldsymbol{w}_t = \left[w_t^{(2)} \dots w_t^{(10)}\right]$ to maximise the quadratic utility:

$$\max_{\boldsymbol{w}_t} \; E\left[R_{p,t+1}\right] - \frac{\gamma}{2}\mathrm{Var}\left(R_{p,t+1}\right),$$

where γ is the risk aversion coefficient of the mean-variance investor, $R_{p,t+1} = 1 + y_t^{(1)} + \boldsymbol{w}_t'\boldsymbol{xr}_{t+1}$ is the gross return on the portfolio, $E\left[R_{p,t+1}\right]$ is the sample mean portfolio return, and $\mathrm{Var}\left(R_{p,t+1}\right)$ is the sample variance portfolio return. Then the solution of the above optimisation is $\boldsymbol{w}_{t,s} = \frac{1}{\gamma}\Sigma_{t+1|t}^{-1}\widehat{\boldsymbol{xr}}_{t+1}\left(\mathcal{M}_s\right)$, where $\widehat{\boldsymbol{xr}}_{t+1}\left(\mathcal{M}_s\right)$ is the vector of bond returns' forecast obtained using model $\mathcal{M}_s$, and $\Sigma_{t+1|t} = Var_t\left(\boldsymbol{xr}_{t+1} - E_t\left[\boldsymbol{xr}_{t+1}\right]\right)$. For the univariate allocation exercise we have: $w_{t,s}^{(n)} = \frac{\widehat{xr}_{t+1}^{(n)}(\mathcal{M}_s)}{\gamma\sigma_{t+1|t}^{(n)}}$ where $\widehat{xr}_{t+1}^{(n)}\left(\mathcal{M}_s\right)$ is the bond returns' forecast for maturity n given model $\mathcal{M}_s$, and $\sigma_{t+1|t}^{(n)}$ the diagonal element of $\Sigma_{t+1|t}$ relative to the bond with n-year maturity.

To proxy for $\Sigma_{t+1|t}$, we employ a rolling sample variance estimator as in Thornton and Valente (2012): $\widehat{\Sigma}_{t+1|t} = \sum_{l=0}^{\infty}\Omega_{t-l}\odot\epsilon_{t-l}\epsilon_{t-l}'$, where $\epsilon_t = \left[\epsilon_t^{(2)} \dots \epsilon_t^{(10)}\right]'$ are forecast errors, $\Omega_{t-l} = \alpha\exp\left(-\alpha\right)\mathbf{1}\mathbf{1}'$ is a symmetric matrix of weights, $\odot$ denotes element-by-element multiplication, and we set the decay rate α to 0.05 (same value as in Thornton and Valente (2012) and within the range of those reported in studies like Fleming et al. (2001)). We also winsorise the weights for each of the

n-period bonds to $-1 \leq w_t^{(n)} \leq 2$ to prevent extreme investments; however, we evaluate the robustness of our results to alternative assumptions about the portfolio weights. Finally, to make our results directly comparable to other studies (e.g. Thornton and Valente, 2012; Gargano et al., 2019), we assume a coefficient of risk aversion of five.

Given the Markowitz optimal weights on the risky bonds, we compute the realised utilities. Then, following Fleming et al. (2001), we obtain the certainty equivalent gains (annualised and in percentages) by equating the average utility of the EH model with the average utility of any of the alternative models.

To test whether the certainty equivalent return (CER) values are statistically greater than zero, we use a Diebold and Mariano (1995) test. Specifically, to evaluate the allocation implied by the NN forecasts, we estimate the following regression:

$$u_{t+1,NN} - u_{t+1,EH} = \alpha^{(n)} + \varepsilon_{t+1} \ ,$$

where $u_{t+1,s} = \boldsymbol{w}'_{t,s}\boldsymbol{xr}_{t+1} - \frac{\gamma}{2}\boldsymbol{w}'_{t,s}\Sigma_{t+1}\boldsymbol{w}_{t,s}$ and $s = \{EH, NN\}$, i.e. we use the optimal weights together with the realised returns.

Power Utility Investor Next, we consider the investment decision of a representative investor that has a power utility of the form,

$$U\left(\boldsymbol{w}_t, \boldsymbol{xr}_{t+1}\right) = \frac{\left[(1 - \boldsymbol{w}'_t\iota)\exp\left(y_t^{(1)}\right) + \boldsymbol{w}'_t\exp\left(y_t^{(1)}\iota + \boldsymbol{xr}_{t+1}\right)\right]^{1-\gamma}}{1-\gamma}, \qquad \gamma > 0 \tag{2.16}$$

where γ captures the investor's risk aversion and ι is a vector of ones. We follow Campbell and Viceira (1999), Campbell and Viceira (2004) and Gargano et al. (2019) and assume excess bond returns are jointly log-normal distributed so that the excess returns on a portfolio of treasury bonds can be approximated by

$$R_{p,t+1} = 1 + y_t^{(1)} + \boldsymbol{w}'_t\boldsymbol{xr}_{t+1} + \frac{1}{2}\boldsymbol{w}'_t\boldsymbol{\sigma}^2_{t+1|t} - \frac{1}{2}\boldsymbol{w}'_t\Sigma_{t+1|t}\boldsymbol{w}_t \tag{2.17}$$

where $\Sigma_{t+1|t}$ denotes the covariance matrix of the excess bond returns, and we denote with $\boldsymbol{\sigma}_{t+1|t}$ its diagonal elements. Campbell and Viceira (2004) show that under log-normality of excess returns, the optimal allocation on a maturity-specific bond can be defined as

$$w_t^{(n)} = \frac{1}{\gamma\left(\sigma_{t+1|t}^{(n)}\right)^2}\left[\widehat{xr}_{t+1}^{(n)} + \left(\sigma_{t+1|t}^{(n)}\right)^2/2\right] \tag{2.18}$$

Given these optimal set of weights, the realised utility for, say, the univariate case can be computed by plugging $\boldsymbol{w}_t$ into Equation (2.16).

Similar to the mean-variance case presented above we proxy for $\Sigma_{t+1|t}$ by using a rolling sample variance estimator as in Thornton and Valente (2012), and set the coefficient of relative risk aversion equal to $\gamma = 5$. Further, to test if the annualised CER values are statistically greater than zero, we employ again a Diebold and Mariano (1995) test. Specifically, for a power utility investor that selects a single risky asset with maturity n using when the forecast from a NN, we estimate the regression

$$u_{t+1,NN}^{(n)} - u_{t+1,EH}^{(n)} = \alpha^{(n)} + \varepsilon_{t+1}$$

where

$$u_{t+1,j}^{(n)} = \frac{1}{1-\gamma}\left[(1-\omega_{t,j}^{(n)})\exp(y_t^{(1)}) + \omega_{t,j}^{(n)}\exp(y_t^{(1)} + xr_{t+1}^{(n)})\right]^{1-\gamma}$$

and $j = \{EH, NN\}$.

2.6.2 Asset Allocation: Results

Table 2.11 shows the annualised CER values computed relative to the EH model. Positive values indicate that the predictive model performs better than the EH model. We focus on the best predictive models from Tables 2.1-2.2, the extreme trees and the *NN 1 Layer (3 nodes)* – when forecasting only with forward rates – or *NN 1 Layer Group Ensem + fwd rate net* – when including the macroeconomic variables.

With the sole exception of the mean-variance investor selecting the two-year bond, the remaining CER values for the trees and NNs are significantly higher than those generated by the EH benchmark. The CER values increase with bond maturity, but the highest CER values are found in the multivariate setting, which suggests that the economic gains associated with NNs forecasts are not limited to specific maturities.

Interestingly, when the (mean-variance or power utility) investor makes no use of information beyond the term structure of interest rates, then trees deliver CER values that are 0.2%-0.7% greater than those obtained using NNs. However, when the investor also considers information from macro and financial variables, then NNs outperform trees by 0.6% (power utility and 10-year bond) to 1% (multivariate setting). A pairwise test confirms that this improvement of NNs over trees is statistically significant. Furthermore, the results in Table 2.11 also show that (for the univariate and multivariate allocation, independently from utility) the group-ensemble NN that exploits macroeconomic information produces significantly higher CER values than those implied by the best-performing NN using yields-only. Overall, our machine learning-based forecasts of bond returns provide support for the hypothesis that a (statistically and economically) significant portion of macroeconomic information is not captured by the yield curve, even after accounting for non-linearity in interest rates.[21]

In summary, we find that a NN that exploits the non-linearities within groups of macroeconomic variables delivers high predictive accuracy (see Table 2.2), which, in turn, translates into investment strategies with large economic value (see Table 2.11).

2.7 Economic Drivers of Bond Return Predictability and Portfolio Performance

In this section, we investigate whether our forecasts of excess bond returns are consistent with explanations based on time-varying risk premia. We then examine the economic drivers of bond return predictability and portfolio performance.

2.7.1 Cyclical Patterns of Expected Excess Bond Returns

We start by investigating the cyclicality of our forecasts of excess bond returns. Indeed, standard asset pricing models featuring habit persistence like Wachter (2006) suggest that bond risk premia are counter-cyclical.

[21]The p-values based on power utility (Panel B) are lower than those reported for mean-variance (Panel A). This is because the power utility setting generates less volatile CER series. To address the higher persistence of power utility CER, we compute Newey-West standard errors using a larger truncation parameter equal to 20 lags (see Lazarus et al., 2018). Even in this case, we continue to find statistical support for our conclusions.

In Panels (a) and (b) in Figure 2.7, we plot the forecast of 10-year bond returns obtained from the best-performing NNs against the industrial production index growth. The results are similar for alternative maturities. We report the prediction based on yields only (Panel (a)), as well as the prediction obtained by adding macroeconomic variables to forward rates (Panel (b)). In Panels (c) and (d), we overlay our forecasts with the realised ten-year excess bond returns series.[22]

Independently of the set of predictors employed, Panels (a) and (b) in Figure 2.7 reveal that the bond risk premium obtained from NNs displays a clear counter-cyclical pattern. In particular, the contemporaneous correlations between forecasts of the ten-year excess bond returns and industrial production is -12.4% (p-value of 0.07) when only information in the term structure is used (Panel (a)). This correlation almost doubles to -24.6% (p-value of 0.01) when we add macroeconomic variables to forward rates (Panel (b)).[23] Thus, using macroeconomic variables greatly improves the estimates of the risk premium.

This prima facie evidence suggests that our forecasts may be consistent with the fact that investors want to be compensated for bearing recession-related risks. To the extent that our forecasts of excess bond returns reflect time-varying risk premia, we would also expect higher Sharpe ratios in recessions. To this end, Table 2.12 reports, for the 2- and 10-year bond maturities, the Sharpe ratios computed separately for recession and expansion periods. We find that, across all maturities and forecasting models, the Sharpe ratios are substantially higher during recessions than in expansions.

2.7.2 Economic Drivers of Expected Excess Bond Returns

Having established that our forecasts of excess bond returns, and the associated Sharpe ratios, move counter-cyclically, we next investigate whether these forecasts are linked to key drivers of bond risk premia suggested by asset pricing theory and previous evidence. In particular, we regress the forecasts of ten-year excess bond returns obtained from the best-performing NNs in Tables 2.1-2.2 on a set of

[22] Relative to yields-only, the addition of macroeconomic variables leads to: (1) NN forecasts that are higher in the recession of 2007-2009, and (2) better predictive performance. In Section 2.5.2, and in particular Panels (b) and (d) of Figure 2.5, we examine the predictive accuracy of NNs throughout our sample period.

[23] The correlation p-values are computed using Newey and West (1987b) standard errors with 12 lags.

structural risk factors that arise in equilibrium models and generate time-varying bond risk premia. Each row in Table 2.13 corresponds to a different specification.

Motivated by the literature on the role of disagreement in asset prices (e.g. Buraschi and Jiltsov, 2007; Dumas et al., 2009), we examine the role played by differences in beliefs for the dynamics of excess bond returns. The results are in row (i) in Table 2.13. We proxy for real disagreement (DiB(g)) and nominal disagreement (DiB(π)) using the interquartile range of four-quarter-ahead forecasts of GDP and consumer prices (CPI), respectively, obtained from the Survey of Professional Forecasters (SPF).

We investigate the link between time-varying risk aversion and excess bond returns in rows (ii) and (iii) of Table 2.13. Asset pricing models featuring habit persistence suggest that risk premia should be higher during recessions due to a reduced surplus consumption ratio. Following Wachter (2006), we proxy for risk aversion using (the negative of) a weighted average of 10 years of quarterly consumption growth rates (dubbed −Surplus). We also employ the new measure of time-varying risk aversion proposed by Bekaert et al. (2019) (dubbed RAbex). This risk aversion measure is calculated from observable financial information at high frequencies.

We next examine the role played by economic growth and inflation uncertainty, UnC(g) and UnC(π), for expected bond returns in row (iv). This link can be motivated by long-run risk models like Bansal and Shaliastovich (2013) or by habit-models that allow for time variation in quantities of risk like Creal and Wu (2018).[24]

Finally, we examine the link between bond volatility and our forecasts of excess bond returns in row (v) of Table 2.13. To assess this link, we employ two proxies: (1) the intra-month sum of squared yield changes (returns) on a constant maturity 10-year zero-coupon bond (denoted as $\sigma(n)$); and (2) the one month implied 10-year maturity bond risk-neutral volatility published by the CME (denoted as $TYVIX$).[25]

[24]To proxy for uncertainty, we adapt the procedure of Bansal and Shaliastovich (2013). In the first step, we use our SPF on consensus expectation of four-quarter GDP growth and inflation and fit a bivariate VAR(1). In a second step, we compute a GARCH(1,1) process on the VAR residuals to estimate the conditional variance of expected real growth and inflation.

[25]`http://www.cboe.com/products/vix-index-volatility/volatility-on-interest-rates/cboe-cbot-10-year-u-s-treasury-note-volatility-index-tyvix`. Accessed 10th February 2020.

Several conclusions emerge from the results in Table 2.13. First, the link between structural risk factors and realised returns is generally weak. The sole factor that is statistically linked to realised bond returns is the risk neutral volatility (Panel A, row v).

Our forecasts of excess bond returns paint a completely different picture. Independently from the set of predictors we use, we find a strongly positive coefficient on uncertainty about economic growth but not on inflation uncertainty (Panels B and C, row iv). We also find strong support for the prediction of equilibrium models based on habit preferences (Panels B and C, row ii). Adding macroeconomic information strengthens this conclusion: in this case, the slope coefficient on the risk aversion measure proposed by Bekaert et al. (2019) is also positive and statistically significant (Panel C, row iii). Finally, in line with Duffee (2002), we find only a weak link between expected bond returns and bond volatility (Panels B and C, row v; the link is marginally significant in Panel B, but the R^2 is small).

There are minor differences between Panel B and C. In particular, the addition of macroeconomic variables leads to a positive and statistically significant slope coefficient on nominal disagreement (Panel C, row i). However, in a horse race only (habit) risk aversion and macroeconomic uncertainty continue to stay significant, leading to a large R^2 of about 25%. Instead, (nominal) disagreement is driven out (Panel C, row vi). This is also the case in Panel B. Comparing row (vi) in Panels B and C to Panel A, it is apparent that using the measure of excess bond returns implied by NNs instead of realised returns leads to stronger support of the predictions of equilibrium models.[26]

The evidence in Table 2.13 for bond returns forecasts confirm that the variation in expected bond returns implied by NNs can be understood in terms of time variation in risk prices and time-varying (macroeconomic) risk. Overall, our results support models that feature both channels, such as Bekaert et al. (2009) and Creal and Wu (2018).

Our evidence stands in stark contrast to the recent finding of Buraschi et al. (2019). They find that the quantity of risk as measured by bond volatility has a

[26] A saturated regression that includes all variables simultaneously leads to the same conclusion: only habit-based risk aversion and macroeconomic uncertainty remain significant. The R^2 from the saturated regressions are 12%, 25.51%, and 25.60%, adding little explanatory power to the specification (vi) in Panels A, B, and C.

strong role, whereas habit-based risk aversion matters little. Thus, our statistical measure of bond risk premia likely captures a potentially different channel component from the subjective bond risk premia of Buraschi et al. (2019). We investigate this point further in the next subsection.

2.7.3 Statistical vs. Subjective Forecasts

Table 2.14 reports the correlations between our forecasts of ten-year excess bond returns obtained from the best-performing tree and NNs in Tables 2.1-2.2 with three recent proxies for risk premia that rely on interest rates forecasts as surveyed by the Blue Chip Financial Forecasts (BCFF): (1) the measure of subjective bond risk premia (EBR^*) proposed by Buraschi et al. (2019) based on aggregation of expectations of (the top decile of) professional forecasters; (2) the Piazzesi et al. (2015) consensus measure of subjective bond risk premia constructed as the difference between subjective and VAR interest-rate expectations, $E_t^*\left[i_{t+h}^{(n-h)}\right] - E_t\left[i_{t+h}^{(n-h)}\right]$; and (3) the forecasts by Giacoletti et al. (2016) based on a learning rule that updates beliefs using the history of bond yields and disagreement among forecasters.[27]

The results in Table 2.14 show that, across all bond maturities and model specifications, the correlation between our forecasts and the subjective bond risk premia of Buraschi et al. (2019) is small, and not statistically significant. This is in line with our previous analysis of economic drivers of bond return predictability (Table 2.13): our forecasts are associated with proxies for time-varying risk aversion and macroeconomic uncertainty whereas Buraschi et al. (2019) find a strong link between the quantity of risk channel (as proxied by bond volatility) and their proxy for bond risk premia.

We instead find a strong and positive association between our forecasts and the Piazzesi et al. (2015) measure of bond risk premia, in particular when yields-only based forecasts are considered. This is perhaps not surprising given the evidence in Buraschi et al. (2019): the consensus is not a sufficient statistic for the cross-section of expectations so that aggregation of subjective bond risk premia for each

[27]We measure $E_t^*\left[i_{t+h}^{(n-h)}\right]$ using the median survey forecast of $i_{t+h}^{(n-h)}$ for the 10-year Treasury bond from the Survey of Professional Forecasters. Importantly, Piazzesi et al. (2015) find that "median forecasts from the SPF are similar to those from the Bluechip survey." The statistical forecasts follow Piazzesi et al. (2015): we compute the forecasts by running OLS directly on the system $Y_{t+h} = \mu + \phi Y_t + \varepsilon_{t+h}$, so that we can compute the h-horizon forecast simply as $\mu + \phi Y_t$. The vector of interest rates Y includes the 1-, 2-, 3-, 4-, 5-, 7-, and 10-year maturity.

single contributor (as in EBR^*) may differ from measures that rely on the simple arithmetic average of the cross-section of forecasters (as in Piazzesi et al., 2015). However, we note that adding macroeconomic variables weakens the correlation between our forecasts and subjective measures based on consensus.

Finally, the correlation between our forecasts and those of Giacoletti et al. (2016) are quite large and mostly significant. This effect is generally stronger for long maturity bonds and when yields-only based forecasts are considered. Overall, dynamic learning effects could account for some of our findings of bond return predictability.

2.8 Conclusion

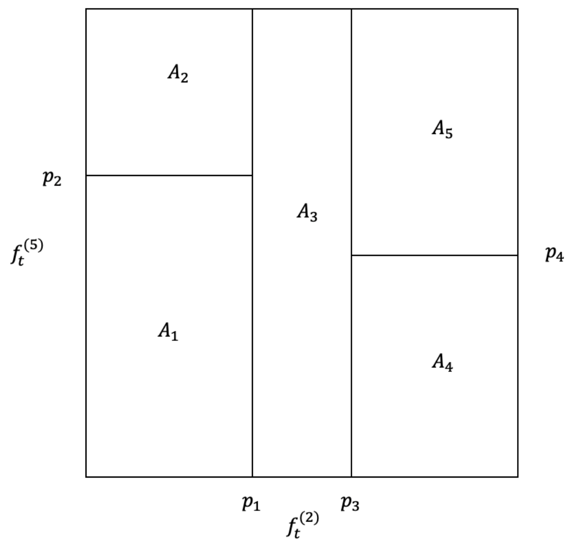
In this paper we evaluate the benefits of using machine learning methods for understanding bond price fluctuations. Three main findings emerge from our analysis.

First, we show that non-linear machine learning techniques, such as extreme trees and neural networks, detect predictable variations in bond returns that are statistically large; importantly, the forecasts implied by these methods translate into similarly large out-of-sample economic gains. Second, we document that employing the NN forecasts based on macroeconomic and yield information produces significantly higher certainty equivalent return values than those implied by the NN forecasts based on yields-only variables, thus providing support for information that is unspanned by (potentially non-linear transformations of) the yield curve and yet useful to forecast bond returns. We also provide evidence of a significant heterogeneity in the relative importance of macroeconomic variables across bond maturities. Hence, the type and nature of unspanned factors may depend on the bond maturity. Finally, we document that NN forecasts are counter-cyclical and mostly related to variables that proxy for macroeconomic uncertainty and time-varying risk aversion. Our results provide support for models that feature both time variation in risk prices and time-varying risk as in, for example, Bekaert et al. (2009) and Creal and Wu (2018). However, our statistical measure of expected bond returns contrasts with recent survey-based measures like the one proposed by Buraschi et al. (2019), which is mostly related to financial (specifically, bond) volatility.

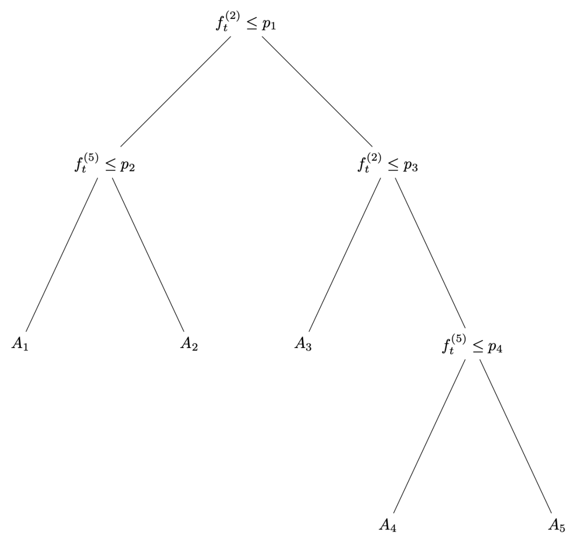
From a pure machine learning perspective in the context of asset pricing, we make three contributions. First, we find that NNs perform well even when, in the context of bond return regressions, we employ just yield-based variables (i.e. in a low dimensional setting). This finding emphasises that the success of NNs is largely due to their ability to capture complex non-linearities in the data. Second, we document that non-linearities within macroeconomic categories (output, inflation, labor market, etc.) are more important than interactions across categories. Finally, we document that a carefully chosen structure of the network (like group ensembling) may compensate for the depth of the network.

Overall machine learning methods that dispense with the linearity assumption in the return-predicting function may prove useful to improve our empirical understanding of asset price movements.

Figures & Tables



(a) Recursive Binary Splitting



(b) Partition Tree

Figure 2.1: Example of a Regression Tree
This figure shows an example of a regression tree for a predictive regression with a univariate target variable, e.g., the holding period excess return of a one-year treasury bond, and two predictors, e.g., the two-year and the five-year forward rates, which we label $y_t^{(2)}$ and $y_t^{(5)}$. The left panel shows the partition of the two-dimensional regression space by recursive splitting. The right panel shows the corresponding regression tree.

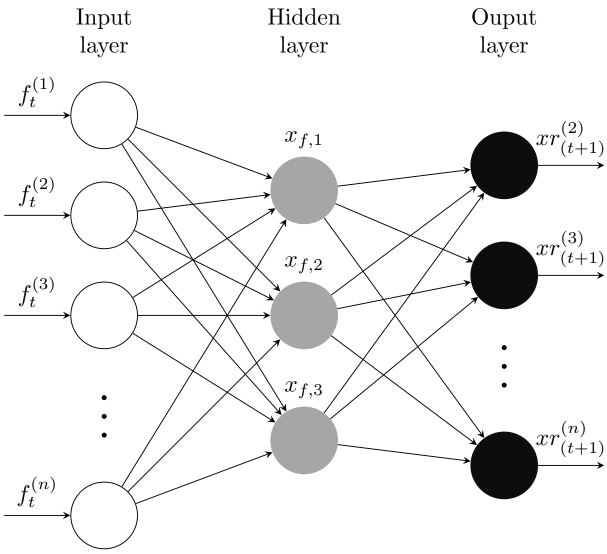
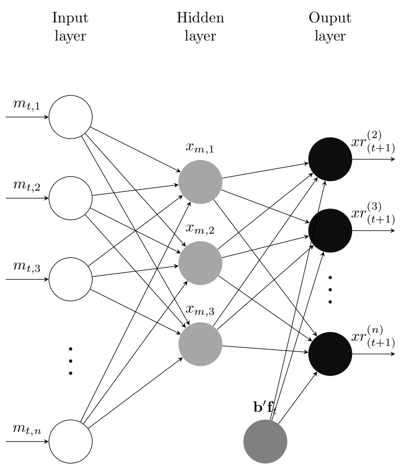
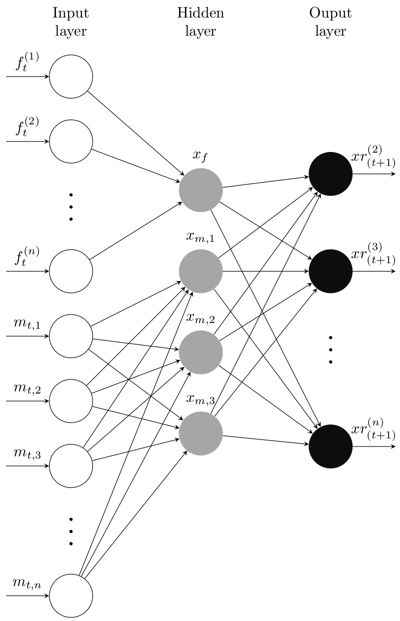


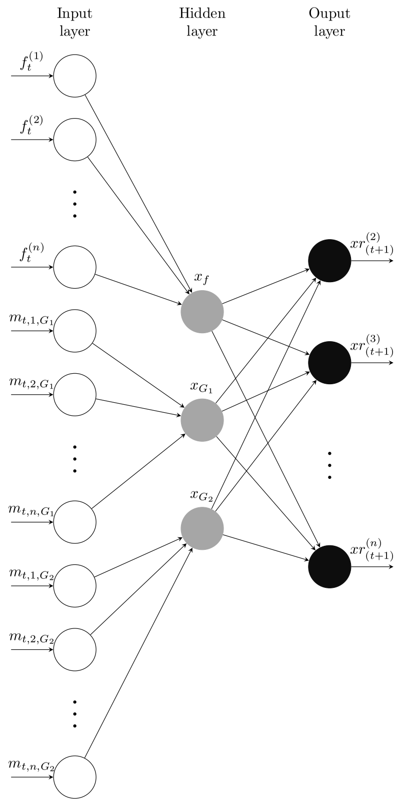
Figure 2.2: Examples of a Neural Network with only Forward Rates
This figure shows a neural network with one hidden layer when forecasts are based only on forward rates as e.g. in Cochrane and Piazzesi (2005).



(a) *Hybrid* network with fwd rates direct



(b) Macro network + fwd rate net



(c) Group-ensemble network

Figure 2.3: Examples of Neural Networks with Forward Rates and Macro Variables
This figure shows examples of the network structures used in the paper. The left panel shows the neural network with a linear combination of forward rates, $b'f_t$, that is included as an exogenous regressor (in the paper such specification is called macro + fwd rates direct). This structure simulates the idea of Ludvigson and Ng (2009) in which the latent macro factors are extracted from a large cross-section of macroeconomic variables and forward rates are included as a linear combination as proposed by Cochrane and Piazzesi (2005). The center panel displays a network structure whereby macro variables, $m_{t,i}$, and forward rates, $f_t^{(n)}$, define two separate groups (in the paper such specification is called macro + fwd rates net). The right panel shows the group-ensemble network whereby groups of macroeconomic variables, m_{t,i,G_1} and m_{t,i,G_2}, and forward rates, $f_t^{(n)}$, define a separate network. The collection of networks is then ensembled at the output layer.

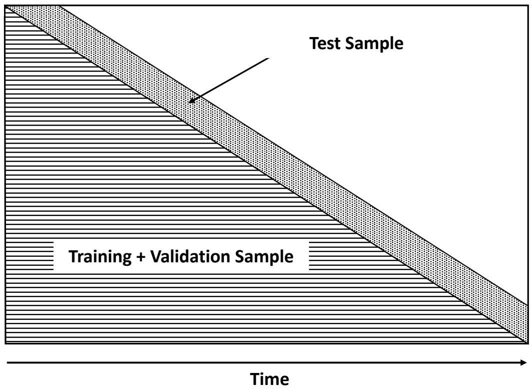
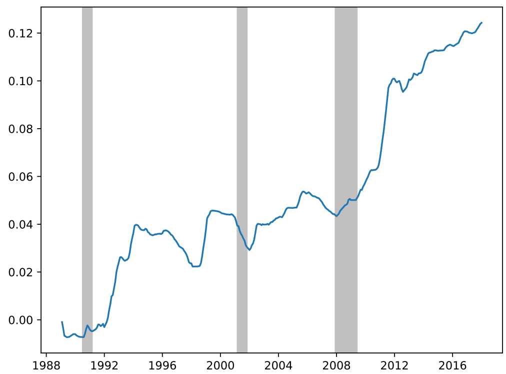
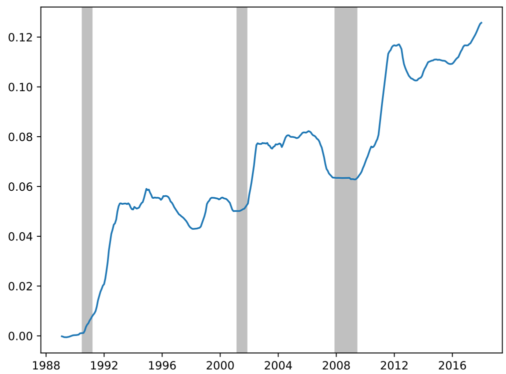


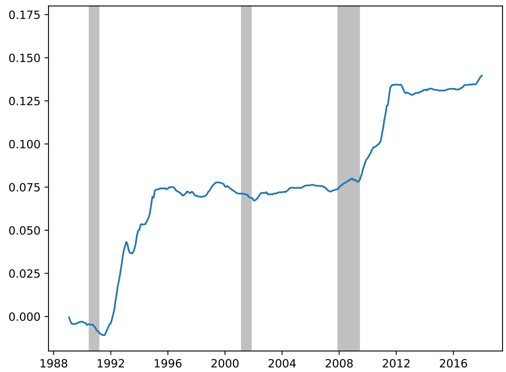
Figure 2.4: Sample Splitting Strategy
This figure shows the sample splitting used for cross-validation of the hyper-parameters of the penalised regressions, i.e. lasso, elastic net, ridge, and the neural networks. The forecasting exercise involves an expanding window that starts in January 1990. The full sample period is from 1971:08 to 2018:12. The training sample consists of the first 85% of the observations while the validation sample consists of the final 15% of observations. The training and the validation samples are consequential and not randomly selected in order to preserve the time series dependence. The testing sample consists of observations on one-year holding period excess Treasury bond returns.



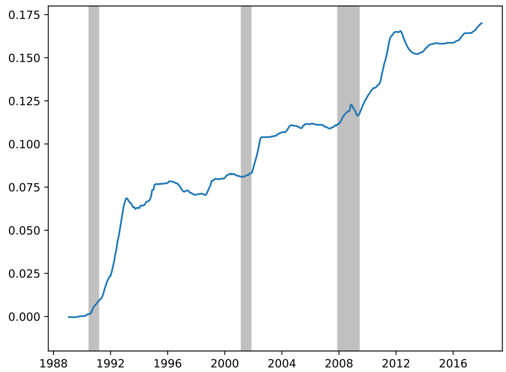
(a) Extreme tree, forward rates, 10-year



(b) Neural net, forward rates, 10-year

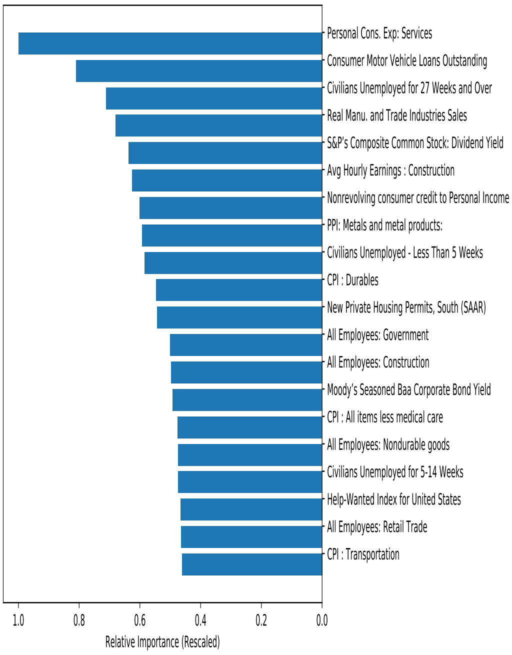


(c) Extreme tree, macro + fwd rates, 10-year

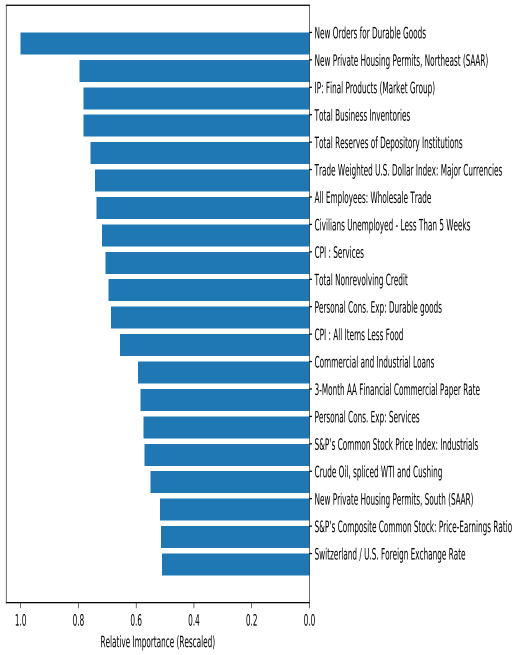


(d) Neural net, macro + fwd rates, 10-year

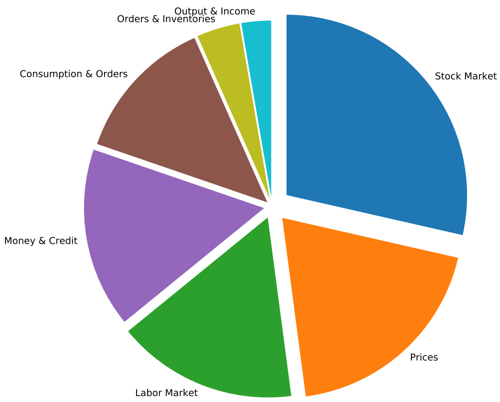
Figure 2.5: Squared Forecast Errors Across Time
This figure plots the time series of difference in squared forecast errors from a given model versus the expectations hypothesis. We scale the forecast errors by the variance of the dependent variable times $T-1$, i.e. each month t we plot the value attained by $\frac{(\hat{\epsilon}^{EH}_{t+1})^2-(\hat{\epsilon}^{Model}_{t+1})}{(T-1)\mathrm{Var}(rx_{t+1})}$. The out-of-sample period starts in 1990. The expectation hypothesis uses all data starting from the in-sample period in 1971:08. The figures present results for the 10-year maturity and focus on the best performing regression tree and neural network when forecasts are generated either based on forward rates only or by macroeconomic variables + forward rates.



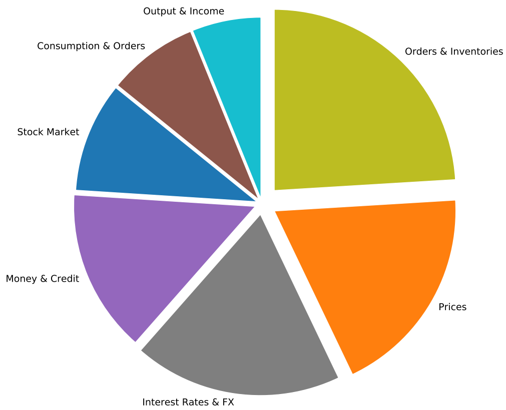
(a) 2-year maturity, individual variables



(b) 10-year maturity, individual variables

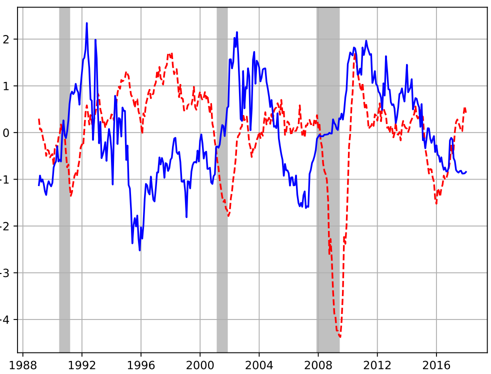


(c) 2-year maturity, groups of variables

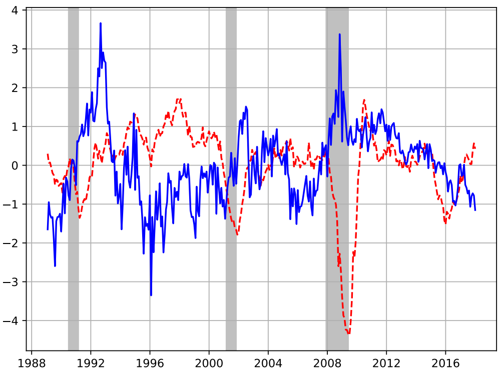


(d) 10-year maturity, groups of variables

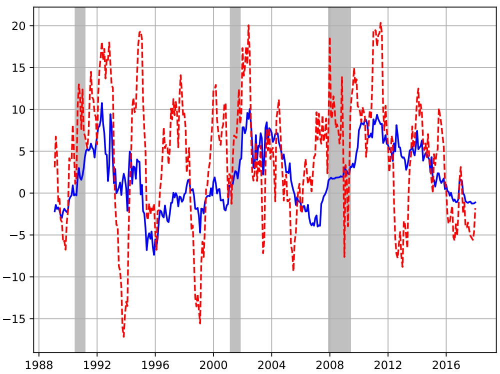
Figure 2.6: Relative Importance of Macroeconomic Variables
This figure shows the relative importance of different macroeconomic variables used to forecast bond excess returns. Panels (a) and (b) show results for individual variables, while panels (c) and (d) present results for groups of macro variables. The groups are labelled according to McCracken and Ng (2016). The relative importance of an input variable is computed based on the absolute value of the gradient of the network outputs with respect to the input variable. The gradient is evaluated at the in-sample mean of the input variable. The gradient-at-the-mean is calculated for each time t of the recursive forecasting exercise and then averaged over the out-of-sample period. For the grouped results in panels (c) and (d) the relative importance is averaged within groups.



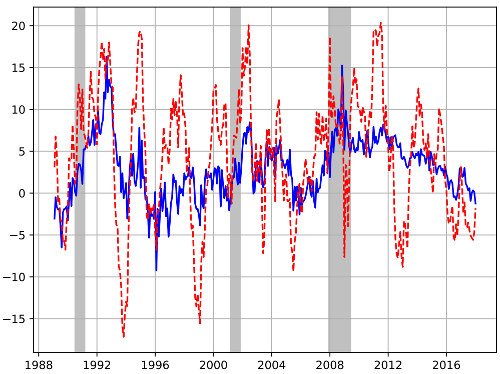
(a) Fwd only: expected return vs. IP



(b) Fwd + Macro: expected return vs. IP



(c) Fwd only: realised vs. expected return



(d) Fwd + Macro: realised vs. expected return

Figure 2.7: Bond Excess Returns, Model-implied Risk Premia, and Economic Growth Panels (a) and (b) plot the model-implied expected bond excess returns for the 10-year maturity (solid lines) against the annual growth rate of industrial production (IP) in the US (dashed line). Panels (c) and (d) display the time series of annual realised (dashed line) and expected (solid line) ten-year bond excess returns (in percentage terms). We report the two best performing forecasts from the neural nets, that is the *NN 1 Layer (3 nodes)* – when forecasting with only the forward rates – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables (see Table 2.1-2.2 for reference). The left panels report the results for the expected bond excess returns obtained by using the forward rates, whereas the right panels report the results for the expected bond excess returns obtained by using a large set of macroeconomic variables in addition to the forward rates.

Table 2.1: Forecasting Annual Holding Period Returns with Forward Rates.
This table reports the out-of-sample R^2_{oos} obtained using forward rates to predict annual excess bond returns for different maturities and across methodologies. To compute the out-of-sample R^2_{oos} we compare the forecasts obtained from each methodology to the expectation hypothesis (i.e. to the prediction based on the historical mean). In addition to the R^2_{oos}, we report the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice that we report a p-value only when the R^2_{oos} is positive. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Models	R^2_{oos}						p-value						R^2_{oos} EW	p-value EW
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(EW)}_{t+1}$	$xr^{(EW)}_{t+1}$
Panel A: PCA and PLS														
PCA (10 components)	-53.0%	-36.2%	-27.5%	-20.3%	-12.4%	-0.7%							-17.7%	
PCA (5 components)	-54.9%	-38.8%	-30.4%	-20.3%	-15.3%	-3.1%							-20.0%	
PCA (3 components)	-17.6%	-10.2%	-5.3%	1.0%	2.9%	10.2%				5.2%	2.8	0.8%	1.0%	3.6%
PCA-Squared (5 Components)	-55.0%	-42.1%	-32.5%	-22.6%	-16.4%	-5.5%							-22.4%	
PCA-Squared (3 Components)	-46.8%	-38.8%	-30.6%	-21.0%	-17.0%	-8.9%							-22.5%	
Partial Least Squares (5 components)	-56.3%	-40.5%	-33.4%	-25.7%	-18.6%	-9.5%							-24.8%	
Partial Least Squares (3 components)	-57.9%	-40.7%	-31.9%	-22.9%	-14.6%	-1.8%							-20.2%	
Panel B: Penalised Linear Regressions														
Ridge	-40.1%	-25.8%	-18.9%	-11.4%	-6.0%	4.6%						1.1%	-9.9%	
Lasso	-11.5%	-8.1%	-2.5%	0.2%	2.3%	9.3%				7.2%	4.4%	1.0%	2.3%	4.1%
Elastic Net	-10.7%	-8.4%	-5.1%	0.4%	0.3%	7.0%				6.4%	6.0%	1.8%	0.7%	5.2%
Panel C: Regression Trees and Neural Networks														
Gradient Boosted Tree	4.9%	4.9%	7.0%	10.8%	7.1%	11.2%	0.2%	0.7%	0.4%	0.1%	0.3%	0.0%	8.9%	0.3%
Random Forest	12.2%	13.1%	15.2%	17.4%	14.9%	16.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.9%	0.0%
Extreme Tree	3.9%	10.1%	14.3%	16.9%	19.7%	24.6%	1.4%	0.9%	0.4%	0.1%	0.1%	0.0%	17.9%	0.1%
NN - 1 Layer (3 nodes)	12.7%	16.4%	19.5%	21.6%	23.1%	26.4%	2.8%	1.4%	0.7%	0.3%	0.2%	0.1%	23.0%	0.2%
NN - 1 Layer (5 nodes)	7.0%	10.7%	14.6%	17.2%	18.9%	22.9%	2.3%	1.3%	0.8%	0.4%	0.3%	0.2%	18.2%	0.3%
NN - 1 Layer (7 nodes)	-3.2%	3.0%	7.1%	12.1%	13.6%	17.1%		2.6%	0.9%	0.3%	0.4%	0.4%	11.5%	0.4%
NN - 2 Layer (3 nodes each)	7.4%	10.8%	12.9%	14.7%	16.3%	19.3%	5.3%	3.1%	1.9%	1.2%	0.8%	0.4%	16.4%	0.9%
NN - 2 Layer (5 nodes each)	8.9%	13.0%	16.0%	18.5%	20.4%	23.6%	2.7%	1.3%	0.7%	0.4%	0.3%	0.2%	20.2%	0.3%
NN - 2 Layer (7 nodes each)	9.0%	15.2%	18.1%	21.2%	23.7%	27.4%	1.3%	0.4%	0.2%	0.1%	0.1%	0.1%	23.1%	0.1%
NN - 3 Layer (3 nodes each)	3.4%	8.9%	10.9%	11.2%	12.7%	15.1%	9.5%	4.0%	2.6%	1.7%	1.2%	0.6%	12.9%	1.3%
NN - 3 Layer (5 nodes each)	3.6%	7.4%	8.9%	10.6%	11.6%	13.7%	11.8%	7.0%	5.7%	3.7%	2.8%	1.8%	11.7%	3.1%
NN - 3 Layer (7 nodes each)	3.9%	6.7%	8.3%	8.9%	11.1%	13.2%	8.8%	6.1%	4.3%	3.1%	2.3%	1.4%	10.8%	2.5%
NN - 3 Layer (5,4,3 nodes each)	-0.9%	2.8%	5.6%	6.5%	8.3%	10.6%		13.9%	7.9%	6.1%	4.0%	2.0%	8.0%	4.4%
NN - 1 Layer (7 nodes), lagged inputs: $t-1:t-11$	6.4%	14.4%	17.0%	19.5%	20.3%	24.4%	2.8%	0.9%	0.6%	0.3%	0.1%	0.1%	20.6%	0.2%

Table 2.2: Forecasting Annual Holding Period Returns with Forward Rates and Macroeconomic Variables.
This table reports the out-of-sample R^2_{oos} obtained using forward rates and a large panel of macroeconomic variables to predict annual excess bond returns for different maturities. To compute the out-of-sample R^2_{oos} we compare the forecasts obtained from each methodology to the expectation hypothesis (i.e. prediction based on the historical mean). In addition to the R^2_{oos}, we report the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice that we report a p-value only when the R^2_{oos} is positive. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12. Penalised regressions are estimated including macroeconomic variables plus either raw forward rates or a linear combination of forward rates as introduced by Cochrane and Piazzesi (2005) (CP). Similarly, neural networks are estimated either adding the CP factor as an additional regressor in the output layer ("fwd rates direct") or by estimating a separate network for forward rates and ensembling both macro and forward rates networks in the output layer ("fwd rates net").

Models	R^2_{oos}						p-value						R^2_{oos} EW	p-value EW
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(EW)}_{t+1}$	$xr^{(EW)}_{t+1}$
Panel A: PCA and PLS														
PCA - first 8 PC'S	-9.8%	-2.9%	0.3%	3.0%	3.3%	4.5%			0.4%	0.4%	0.3%	0.2%	1.8%	0.3%
PCA as in Ludvigson and Ng (2009)	-3.4%	0.2%	1.6%	1.6%	-1.4%	-4.7%		0.7%	0.6%	0.7%			-1.3%	
PLS - 8 components	-40.7%	-19.7%	-12.0%	-8.2%	-2.7%	3.4%						0.0%	-6.4%	
Panel B: Penalised Linear Regressions														
Ridge (using CP factor)	-45.3%	-23.6%	-16.7%	-13.2%	-3.1%	5.3%						0.0%	-5.6%	
Lasso (using CP factor)	6.4%	11.2%	12.9%	14.4%	19.6%	23.7%	0.8%	0.4%	0.2%	0.2%	0.1%	0.0%	21.0%	0.1%
Elastic Net (using CP factor)	6.4%	11.0%	14.3%	15.7%	21.7%	29.1%	0.7%	0.4%	0.2%	0.1%	0.1%	0.0%	22.0%	0.1%
Ridge (using fwd rates directly)	-52.2%	-28.7%	-22.7%	-18.3%	-13.1%	-3.5%							-15.4%	
Lasso (using fwd rates directly)	11.0%	12.0%	12.3%	16.4%	19.9%	23.6%	0.3%	0.3%	0.2%	0.6%	0.7%	0.4%	20.7%	0.4%
Elastic Net (using fwd rates directly)	10.2%	14.2%	16.0%	13.2%	19.9%	23.6%	0.5%	0.3%	0.3%	0.3%	0.2%	0.2%	21.0%	0.2%
Panel C: Regression Trees and Neural Networks														
Gradient Boosted Tree	13.1%	15.9%	18.1%	23.8%	22.5%	25.5%	0.4%	0.6%	0.5%	0.3%	0.4%	0.1%	26.2%	0.2%
Random Forest	26.7%	21.5%	22.0%	24.4%	20.0%	25.0%	0.3%	0.3%	0.2%	0.1%	0.4%	0.2%	26.6%	0.2%
Extreme Tree	23.0%	23.4%	22.3%	23.7%	29.9%	29.6%	0.4%	0.5%	0.5%	0.4%	0.1%	0.3%	29.2%	0.2%
NN 1 Layer (32 nodes), fwd rates direct	6.0%	13.6%	17.8%	22.0%	22.5%	26.5%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%	22.4%	0.0%
NN 2 Layer (32, 16 nodes), fwd rates direct	16.6%	21.5%	24.9%	28.0%	29.4%	32.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	29.3%	0.0%
NN 3 Layer (32, 16, 8 nodes), fwd rates direct	24.8%	26.3%	29.7%	32.0%	31.7%	33.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	32.5%	0.0%
NN 1 Layer (32 nodes), fwd rates net (1 layer: 3 nodes)	8.4%	19.0%	23.8%	25.6%	27.2%	29.5%	0.3%	0.1%	0.1%	0.1%	0.1%	0.0%	27.2%	0.001
NN 2 Layer (32,16, nodes), fwd rates net (1 layer: 3 nodes)	12.1%	15.7%	20.0%	23.5%	25.4%	28.1%	0.8%	0.4%	0.2%	0.1%	0.1%	0.1%	25.1%	0.001
NN 3 Layer (32,16, 8 nodes), fwd rates net (1 layer: 3 nodes)	7.6%	16.3%	20.2%	23.7%	25.0%	28.1%	1.4%	0.5%	0.4%	0.2%	0.2%	0.1%	25.0%	0.2%
NN 1 Layer Group Ensem (1 node per group), fwd rates direct	12.6%	17.3%	21.6%	24.2%	25.9%	29.6%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	25.9%	0.000
NN 1 Layer Group Ensem (1 node per group), fwd rates net (1 layer: 3 nodes)	20.0%	25.6%	29.5%	31.2%	33.6%	36.3%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	34.0%	0.0%
NN 2 Layer Group Ensem (2,1 nodes per group / hidden layer), fwd rates net (2 layer: 3 nodes)	17.3%	23.6%	27.8%	29.8%	31.0%	33.0%	0.6%	0.1%	0.1%	0.0%	0.0%	0.0%	31.6%	0.0%
NN 3 Layer Group Ensem (3, 2, 1 nodes per group / hidden layer), fwd rates net (3 layer: 3 nodes)	13.6%	20.0%	23.7%	26.1%	27.5%	30.7%	1.1%	0.5%	0.3%	0.2%	0.2%	0.1%	27.9%	0.2%

Table 2.3: Forecasting Monthly Holding Period Returns with Forward Rates.
This table reports the out-of-sample R^2_{oos} obtained using forward rates to predict monthly excess bond returns for different maturities and across methodologies. To compute the out-of-sample R^2_{oos} we compare the forecasts obtained from each methodology to the expectation hypothesis (i.e., to the prediction based on the historical mean). In addition to the R^2_{oos} we report the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice that we report a p-value only when the R^2_{oos} is positive. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Models	R^2_{oos}				p-value				R^2_{oos} EW	p-value EW
Maturity (months)	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	$xr^{(EW)}_{t+1}$	$xr^{(EW)}_{t+1}$
Panel A: PCA and PLS										
PCA (5 of 5 components)	-5.3%	-3.7%	-1.3%	-0.3%					-1.8%	
PCA (3 of 5 components)	-0.2%	0.2%	1.0%	1.0%		11.3%	5.8%	5.1%	0.7%	7.1%
PCA (1 of 5 components)	-4.4%	-3.3%	-2.8%	-2.5%					-3.0%	
PCA-Squared (5 of 5 Components)	-7.7%	-4.0%	-1.6%	-0.5%					-2.1%	
PCA-Squared (3 of 5 Components)	-3.9%	-2.0%	-0.8%	0.0%				4.8%	-0.9%	
Partial Least Squares (5 Components)	-5.3%	-3.7%	-1.3%	-0.3%					-1.8%	
Partial Least Squares (3 Components)	-4.0%	-2.7%	0.3%	0.6%			10.5%	7.9%	-0.5%	
Panel B: Penalised Linear Regressions										
Ridge	-1.9%	-1.1%	0.0%	0.3%				12.6%	-0.3%	
LASSO	-1.1%	-1.3%	-0.9%	-0.5%					-0.6%	
ElasticNet	0.4%	-0.5%	-0.3%	-0.2%	24.5%				-0.2%	
Panel C: Regression Trees and Neural Networks										
Gradient Boosted Tree	-0.2%	-1.0%	-1.0%	-0.3%					-0.5%	
Random Forest	-15.9%	-17.0%	-16.0%	-14.7%					-14.8%	
Extreme Tree	-16.3%	-24.5%	-27.3%	-21.9%					-22.8%	
NN - 1 layer (3 nodes)	0.8%	1.0%	1.1%	1.1%	9.3%	5.1%	2.1%	1.1%	1.1%	2.3%
NN - 1 layer (5 nodes)	0.7%	1.0%	1.1%	1.1%	12.1%	4.5%	2.3%	1.0%	1.1%	2.4%
NN - 2 layer (3 nodes each)	0.7%	1.1%	1.2%	1.2%	10.3%	4.3%	2.0%	1.1%	1.2%	2.2%
NN - 2 layer (5 nodes each)	0.8%	1.3%	1.4%	1.5%	10.1%	3.5%	1.2%	0.5%	1.4%	1.5%
NN - 3 layer (3 nodes each)	0.9%	1.1%	1.2%	1.2%	8.9%	5.2%	2.5%	1.2%	1.2%	2.6%
NN - 3 layer (5 nodes each)	0.7%	1.2%	1.3%	1.3%	11.9%	4.5%	2.1%	0.9%	1.3%	2.3%
NN - 3 layer (5,4,3 nodes each)	0.7%	1.1%	1.3%	1.3%	11.7%	5.1%	2.1%	1.0%	1.2%	2.4%

Table 2.4: Forecasting Monthly Holding Period Returns with Forward Rates and Macroeconomic Variables.
This table reports the out-of-sample R^2_{oos} obtained using forward rates and a large panel of macroeconomic variables to predict monthly excess bond returns for different maturities. To compute the out-of-sample R^2_{oos} we compare the forecasts obtained from each methodology to the expectation hypothesis (i.e., prediction based on the historical mean). In addition to the R^2_{oos} we report the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice that we report a p-value only when the R^2_{oos} is positive. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12. Penalised regressions are estimated including macroeconomic variables plus either raw forward rates or a linear combination of forward rates as introduced by Cochrane and Piazzesi (2005) (CP). Similarly, neural networks are estimated either adding the CP factor as an additional regressor in the output layer ("fwd rates direct") or by estimating a separate network for forward rates and ensembling both macro and forward rates networks in the output layer ("fwd rates net").

Models	R^2_{oos}				p-value				R^2_{oos} EW	p-value EW
Maturity (months)	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	$xr^{(EW)}_{t+1}$	$xr^{(EW)}_{t+1}$
Panel A: PCA and PLS										
PCA - first 8 PC'S	-12.1%	-4.2%	-0.3%	3.5%				0.3%	-0.4%	
PCA as in LN	-4.9%	-1.1%	0.3%	1.9%			0.5%	0.8%	0.3%	0.6%
PLS - 8 components	-68.8%	-54.5%	-39.6%	-30.8%					-40.6%	
Panel B: Penalised Linear Regressions										
Ridge (using CP factor)	-61.2%	-49.1%	-35.5%	-27.3%					-36.5%	
Lasso (using CP factor)	4.2%	3.2%	2.6%	2.6%	0.5%	0.3%	0.3%	0.4%	3.2%	0.4%
Elastic net (using CP factor)	4.6%	3.5%	2.0%	2.5%	0.4%	0.3%	0.4%	0.3%	3.1%	0.4%
Ridge (using fwd rates directly)	-61.8%	-49.8%	-36.0%	-27.8%					-37.1%	
Lasso (using fwd rates directly) 5.1%	2.9%	1.1%	2.0%	0.2%	0.5%	0.6%	0.4%	2.6%	0.4%	
Elastic Net (using fwd rates directly)	4.8%	3.5%	2.0%	2.3%	0.3%	0.3%	0.5%	0.5%	3.0%	0.4%
Panel C: Regression Trees and Neural Networks										
Gradient Boosted Regression Tree	-1.8%	-3.4%	-2.8%	-5.8%					-2.1%	
Random Forest	3.9%	1.2%	2.4%	1.3%	0.5%	0.5%	0.2%	3.4%	2.8%	0.9%
Extreme Tree	6.2%	3.7%	2.3%	3.6%	0.1%	0.1%	0.7%	0.2%	4.1%	0.2%
NN 1 Layer (32 nodes), fwd rates direct	-1.5%	0.4%	2.0%	4.3%		0.9%	0.7%	2.2%	2.6%	1.3%
NN 2 Layer (32, 16 nodes), fwd rates direct	3.6%	3.4%	3.8%	4.7%	1.0%	0.7%	0.6%	1.7%	4.3%	1.0%
NN 3 Layer (32, 16, 8 nodes), fwd rates direct	-2.8%	-2.1%	-1.8%	-1.6%					-2.0%	
NN 1 Layer (32 nodes), fwd rates net (1 layer: 3 nodes)	-1.0%	-0.6%	-0.8%	0.2%				6.2%	-0.3%	
NN 2 Layer((32,16, nodes), fwd rates net (1 layer: 3 nodes)	0.6%	1.0%	0.9%	1.9%	1.4%	0.7%	0.6%	0.7%	1.4%	0.5%
NN 3 Layer((32,16, 8 nodes), fwd rates net (1 layer: 3 nodes)	0.8%	1.1%	1.2%	1.2%	8.0%	5.2%	3.4%	2.4%	1.2%	3.3%
NN 1 Layer Group Ensem (1 node per group), fwdrates direct	6.1%	5.1%	4.6%	4.6%	0.1%	0.1%	0.1%	0.5%	5.1%	0.2%
NN 2 Layer Group Ensem (2,1 nodes per group / hidden layer), fwdrates direct	1.4%	0.8%	0.7%	0.8%	6.3%	9.9%	10.5%	9.2%	0.9%	8.9%
NN 1 Layer Group Ensem (1 node per group), fwdrate Net (1 layer: 3 nodes)	5.9%	5.2%	5.0%	5.1%	0.1%	0.1%	0.1%	0.3%	5.5%	0.2%
NN 2 Layer Group Ensem (2, 1 node per group), fwdrate Net (1 layer: 3 nodes)	1.5%	1.3%	1.2%	1.1%	3.9%	3.6%	3.9%	3.9%	1.3%	3.4%

Table 2.5: Forecasting Performances in Expansions and Recessions.
This table reports the out-of-sample performances, measured by R^2_{oos}, separately for expansions (Exp) and recessions (Rec) as defined by the NBER recession index. For ease of exposition, we report the results for the Principal Component Regression with three PCs and for the best performing non-linear methodologies, that is extreme trees and the *NN 1 Layer (3 nodes)* – when forecasting with only the forward rates – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables – (see Table 2.1-2.2 for reference). We focus on the prediction exercise with two- and ten-year maturity bonds. We denote in boldface values for which the predictive accuracy of a given model is better than that obtained from the EH benchmark at the 5% level (p-value calculated as in Clark and West (2007)). The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

	Forward rates		Fwd rates + Macro	
	R^2_{oos}		R^2_{oos}	
	Exp	Rec	Exp	Rec
PCA (10-year maturity)	7.69%	**34.88%**	-1.08%	-41.19%
PCA (2-year maturity)	-15.27%	-30.04%	-8.30%	**22.98%**
Extreme tree (10-year maturity)	**26.80%**	2.85%	**33.38%**	-8.74%
Extreme tree (2-year maturity)	**7.19%**	-14.18%	**25.08%**	11.63%
Neural net (10-year maturity)	**26.28%**	**27.33%**	**35.70%**	**42.54%**
Neural net (2-year maturity)	**9.42%**	**30.31%**	**17.34%**	**34.23%**

Table 2.6: Conditional Forecast Accuracy: Double Sorts on Prevailing Yield Curve Level & Slope

This table reports the forecast accuracy when conditioning on the prevailing shape of the yield curve, i.e. its level and slope. As a measure of the yield curve level we use the 2-year yield, while the measure of yield curve slope is the difference between the 10-year and 2-year yield. We sort all observations in our out-of-sample period based on the median level and slope of the yield curve prevailing at the start of the holding period. The double sort is performed unconditionally. We denote observations below the median by "Low" and observations above the median by "High". For example, "Level Low - Slope High" refers to all observations for which the prevailing yield curve level was below the median, while the yield curve slope was above the median. The median yield curve level over our out-of-sample period is 3.84% and the median yield curve slope is 1.33%. Forecast accuracy is proxied by the R^2 in regressions of realised returns, $xr_{t+1}^{(i)}$, on the predicted bond risk premium, $\hat{xr}_{t+1}^{(i)}$: $xr_{t+1}^{(i)} = \alpha + \beta \hat{xr}_{t+1}^{(i)} + \epsilon_{t+1}^{(i)}$, where the regressions are performed using all the observations that fall into the four distinguished yield curve shape cases. Predicted bond risk premia stem from forecasting either with only the forward rates (**Panel A**) – using *NN 1 Layer (3 nodes)* – or with forward rates plus macroeconomic variables (**Panel B**) – using *NN 1 Layer Group Ensem + fwd rate net* – (see Table 2.1-2.2 for reference). The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12. For each case, we report R^2, the mean fitted value, the p-values for the hypothesis tests $H_0: \alpha = 0$ and $H_0: \beta = 1$, and the fraction of the sample period falling into the respective case.

Panel A: Forecasting with Forward Rates

	R^2 (%)	Mean Fitted Value	p-val ($\alpha = 0$)	p-val ($\beta = 1$)	Sample Fraction
All	21.97	4.33%	0.69	0.10	100.0%
Level Low - Slope Low	7.79	-2.27%	0.00	0.91	9.8%
Level Low - Slope High	16.18	6.26%	0.76	0.70	40.2%
Level High - Slope Low	23.79	3.60%	0.00	0.15	40.2%
Level High - Slope High	58.84	5.96%	0.09	0.00	9.8%

Panel B: Forecasting with Forward Rates + Macro

	R^2 (%)	Mean Fitted Value	p-val ($\alpha = 0$)	p-val ($\beta = 1$)	Sample Fraction
All	27.95	4.33%	0.91	0.21	100.0%
Level Low - Slope Low	14.06	-2.27%	0.00	0.99	9.8%
Level Low - Slope High	18.48	6.26%	0.83	0.42	40.2%
Level High - Slope Low	28.16	3.60%	0.01	0.12	40.2%
Level High - Slope High	65.75	5.96%	0.06	0.01	9.8%

Table 2.7: Ex-post Diagnostics Based on Principal Components Forecasts.

This table reports the in-sample R^2 of a predictive regression where the dependent variable is the year-on-year change in the first three principal components extracted from the term structure of interest rates. The first row reports results when the independent variables are the lagged first three principal components. The second and third rows display the in-sample R^2 when, in addition to the first three principal components, we include the factors extracted from the best performing neural networks obtained using either forward rates only or forward rates plus a large set of macroeconomic variables (see Table 2.1-2.2 for reference).

	Level	Slope	Curvature
PCA	9.28%	21.66%	48.70%
Neural net (fwd rates only)	36.67%	22.05%	70.52%
Neural net (fwd rates + macro)	30.98%	30.91%	65.43%

Table 2.8: Magnitude of Cross- and Within-group Interactions
This table reports the sum of the absolute value of the cross-group (/**Panel A**) and within-group (**Panel B**) interactions obtained from an ensembled neural network with one hidden layer ("Groups ensemble") and a (non-ensembled) neural network with three hidden layers ("Fully connected network"). See Table 2.2 for their performance (row "NN 3 Layer (32, 16, 8 nodes), fwd rates direct" and "NN 1 Layer Group Ensem (1 node per group), fwd rates Net (1 layer: 3 nodes)"). Interactions magnitudes are computed by numerically approximating the network Hessian's, i.e. the second derivatives of network outputs with respect to two distinct network inputs. The NNs take as inputs both macroeconomic variables and forward rates. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Panel A: Sum of Cross-Group Hessian Absolute Values

Model	2y	3y	4y	5y	7y	10y
Fully connected network	60.24	108.61	153.22	184.43	247.84	336.49
Groups ensemble	0.04	0.07	0.09	0.11	0.14	0.18

Panel B: Sum of Within-Group Hessian Absolute Values

Model	2y	3y	4y	5y	7y	10y
Fully connected network	11.93	21.30	30.12	36.03	48.51	65.77
Groups ensemble	16.93	32.05	45.09	54.70	78.85	117.05

Table 2.9: Alternative Model Combination Strategies
This table reports the out-of-sample R^2_{oos} obtained using a large panel of macroeconomic variables and forward rates to predict annual excess bond returns for different maturities. In addition to the best performing neural network – *NN 1 Layer Group Ensem + fwd rate net* (see Table 2.2), we compute the R^2_{oos} for three alternative model combination strategies. The first is a recursive full cross-validation that selects every five years not only the dropout rate and the L1/L2 penalties, but also the number of layers, the nodes per macro group, and the nodes in fwd rate net. The second and third model combination are based on a weighted average of each neural network forecasts with weights that are calculated as the inverse of the validation loss or, alternatively, simply as $1/N$ where N is the number of neural networks estimated. The out-of-sample R^2_{oos} is calculated by considering the expectations hypothesis as a benchmark. That is, we compare the forecasts obtained from each methodology to the prediction based on the historical mean of bond excess returns. In addition to the R^2_{oos} we report the p-value calculated as in Clark and West (2007) in parentheses. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Models	R^2_{oos}						R^2_{oos} EW
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(EW)}_{t+1}$
NN 1 Layer Group Ensem + fwd rate net	20.0%	25.6%	29.5%	31.2%	33.6%	36.3%	34.0%
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Inverse Val. Loss Weighted Model (across all NNs)	22.3%	25.6%	29.1%	30.8%	32.0%	34.4%	32.6%
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Equally Weighted Model (across all NNs)	22.0%	25.2%	28.6%	30.4%	31.6%	33.9%	32.1%
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Full CV Model	24.7%	27.6%	30.4%	31.7%	32.3%	34.7%	33.4%
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 2.10: Stability of the Neural Network Ranking
This table reports how often the four best performing neural networks are selected throughout the sample. More specifically, we select the top 4 models based on the unconditional average of (inverse) validation loss. Then we count how often the four models rank 1st, 2nd, 3rd and 4th, in terms of their inverse validation error. The sum of the values in each column equals the number of periods in our out-of-sample period. This gives an approximate measure of how often one model ranks on top in terms of validation loss. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

	Ranking			
Model	1st place	2nd place	3rd place	4th place
NN 1-Layer Group Ensem + Fwd. Rate Net	170	42	115	21
NN 1-Layer Group Ensem + Fwd. Rates	137	50	154	7
NN 3-Layer + Fwd. Rates	31	112	16	189
NN 2-Layer Group Ensem + Fwd. Rates Net	10	144	63	131

Table 2.11: Economic Significance of Bond Predictability.
This table reports the annualised certainty equivalent values (in %) for portfolio decisions based on the out-of-sample forecasts of bond excess returns for an investor with either mean-variance (**Panel A**) or power utility (**Panel B**) and a coefficient of risk aversion equal to five. The table reports two asset allocation exercises. In the univariate asset allocation case, the investor selects either the two- or the ten-year bond, along with the one-year short-rate. In the multivariate case, the investor selects bonds across the six maturities, two- to five-, seven- and ten-years. The asset allocation decision is based on the predictions implied by either the best performing regression tree specification, i.e. extreme tree, or the best performing neural network, namely the *NN 1 Layer (3 nodes)* – when forecasting with only the forward rates – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables – (see Table 2.1-2.2 for reference). The row Δ reports the value added by NN relative to extreme tree within each application ("Fwd rates" and "Fwd + Macro"). The column Δ reports the value added by "Fwd+Macro" relative to "Fwd rates" within each model (NN and extreme tree). The models are benchmarked against the expectation hypothesis. The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12. Statistical significance is based on a one-sided Diebold and Mariano (1995) test as extended by Harvey et al. (1997) to account for autocorrelation in the forecasting errors. We flag in bold those values that are statistically significant at the 5% confidence level.

Panel A: Mean-Variance Utility

	2-year maturity			10-year maturity			All		
	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ
Neural net	-0.044	-0.002	0.042	**2.622**	**4.194**	**1.571**	**3.555**	**5.015**	**1.461**
p-value	(0.552)	(0.981)	(0.422)	(0.002)	(0.000)	(0.000)	(0.017)	(0.050)	(0.000)
Extreme tree	-0.151	-0.022	0.128	**2.881**	**2.955**	0.074	3.266	**3.961**	0.695
p-value	(0.464)	(0.675)	(0.355)	(0.001)	(0.000)	(0.864)	(0.078)	(0.078)	(0.112)
Δ	0.106	0.020		-0.258	**1.239**		0.289	**1.054**	
p-value	(0.421)	(0.680)		(0.510)	(0.001)		(0.722)	(0.024)	

Panel B: Power Utility

	2-year maturity			10-year maturity			All		
	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ
Neural net	**0.056**	**0.111**	**0.054**	**2.714**	**3.077**	**0.363**	**3.152**	**4.829**	**1.678**
p-value	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.020)	(0.000)	(0.000)	(0.000)
Extreme tree	0.022	**0.092**	**0.072**	**3.301**	**2.511**	**-0.795**	**3.831**	**3.943**	0.113
p-value	(0.486)	(0.002)	(0.000)	(0.000)	(0.000)	(0.021)	(0.000)	(0.000)	(0.835)
Δ	0.034	0.017		**-0.591**.	**0.567**		**-0.680**	**0.886**	
p-value	(0.212)	(0.267)		(0.023)	(0.024)		(0.030)	(0.029)	

Table 2.12: Sharpe Ratios in Expansions and Recessions.
This table reports the out-of-sample annualised Sharpe ratio, separately for expansions (Exp) and recessions (Rec) as defined by the NBER recession index. We report the results for the benchmarking regression that employ the first three principal components of the yield curve, and for the best performing non-linear methodologies, that is extreme trees and the *NN 1 Layer (3 nodes)* – when forecasting with only the forward rates – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables – (see Table 2.1-2.2 for reference). We focus on the prediction of two- and ten-year bonds. The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

	Forward rates		Fwd rates + Macro	
	Exp	Rec	Exp	Rec
PCA (10-year maturity)	0.087	1.364	-0.118	0.190
PCA (2-year maturity)	0.037	1.524	0.403	1.356
Extreme tree (10-year maturity)	0.261	0.521	0.491	0.555
Extreme tree (2-year maturity)	0.253	1.688	0.740	1.541
Neural net (10-year maturity)	0.506	1.769	0.749	1.707
Neural net (2-year maturity)	1.077	2.384	1.093	2.098

Table 2.13: Drivers of Bond Risk Premia.
This table reports the regression estimates of realised (**Panel A**) and expected (**Panel B and C**) bond excess returns on 10-year bonds on a set of structural determinants of risk premia (see discussion in the paper for details). The expected bond return (dependent variable) is based on the predictions implied by the best performing neural network, namely the *NN 1 Layer (3 nodes)* – when forecasting with only the forward rates – and *NN 1 Layer Group Ensem + fwd rate net* – when including also macroeconomic variables – (see Table 2.1-2.2 for reference). We standardise both left- and right-hand side variables, so that a 1-standard deviation change in the right hand variables implies a β-standard deviation in the dependent variable. We report the regression estimates as well as Newey-West p-values. Bold font indicates significance at the 5% level. The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Panel A: 10-year realised bond excess returns

	$DiB(g)$	$DiB(\pi)$	$-Surplus$	$RAbex$	$UnC(g)$	$UnC(\pi)$	$TYVIX$	$\sigma_B^{(n)}$	$R^2(\%)$
(i)	-0.19	0.29							6.43
	(0.41)	(0.19)							
(ii)			0.01						0.25
			(0.90)						
(iii)				0.01					0.25
				(0.81)					
(iv)					-0.14	0.28			5.82
					(0.42)	(0.14)			
(v)							**0.25**	-0.22	6.36
							(0.02)	(0.38)	
(vi)		0.27	0.04	-0.03	0.09				7.01
		(0.08)	(0.66)	(0.67)	(0.14)				

Panel B: 10-year expected bond excess returns (fwd rates)

	$DiB(g)$	$DiB(\pi)$	$-Surplus$	$RAbex$	$UnC(g)$	$UnC(\pi)$	$TYVIX$	$\sigma_B^{(n)}$	$R^2(\%)$
(i)	0.02	0.31							3.06
	(0.96)	(0.22)							
(ii)			**0.35**						9.85
			(0.01)						
(iii)				0.15					1.58
				(0.21)					
(iv)					**0.40**	-0.11			10.34
					(0.00)	(0.72)			
(v)							-0.11	**0.74**	3.30
							(0.53)	(0.04)	
(vi)		0.01	**0.43**	0.04	**0.28**				23.69
		(0.95)	(0.00)	(0.77)	(0.00)				

Panel C: 10-year expected bond excess returns (fwd rates + macro)

	$DiB(g)$	$DiB(\pi)$	$-Surplus$	$RAbex$	$UnC(g)$	$UnC(\pi)$	$TYVIX$	$\sigma_B^{(n)}$	$R^2(\%)$
(i)	-0.35	**0.55**							8.22
	(0.43)	(0.01)							
(ii)			**0.32**						11.80
			(0.02)						
(iii)				**0.27**					6.06
				(0.02)					
(iv)					**0.38**	-0.15			8.49
					(0.01)	(0.71)			
(v)							0.05	0.59	5.44
							(0.80)	(0.14)	
(vi)		0.19	**0.35**	0.15	0.21				25.22
		(0.35)	(0.00)	(0.14)	(0.09)				

Table 2.14: Statistical vs. Subjective Forecasts of Bond Risk Premia.
This table reports the correlation between our machine learning implied forecasts and existing measures of bond risk premia based on subjective forecasts or on asset pricing models with learning dynamics. **Panel A:** shows the results for the forecasts generated using only the forward rates, whereas **Panel B:** shows the results for the forecasts generated using both forward rates and a large panel of macroeconomic variables. Correlations are computed with respect to the subjective bond risk premia in Buraschi et al. (2019) (EBR^*_{10y} and EBR^*_{2y}), the subjective risk premia measure proposed by Piazzesi et al. (2015) (SUBJ_BRP), and the out-of-sample bond returns forecasts in Giacoletti et al. (2016) (GLS). In parentheses we report Newey-West p-values with 12 lags. Bold font indicates significance at the 5% level. The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Panel A: Forecasting with forward rates

	10-year maturity		
	EBR^*_{10y}	SUBJ_BRP	GLS
Extreme tree	-7.5%	**49.5%**	**52.3%**
	(0.45)	(0.00)	(0.00)
NN - 1 Layer (3 nodes)	3.3%	**56.3%**	**59.5%**
	(0.75)	(0.00)	(0.00)

	2-year maturity		
	EBR^*_{2y}	SUBJ_BRP	GLS
Extreme tree	-18.7%	**42.7%**	**63.8%**
	(0.17)	(0.00)	(0.00)
NN - 1 Layer (3 nodes)	4.1%	**53.8%**	**50.3%**
	(0.78)	(0.00)	(0.00)

Panel B: Forecasting with forward rates and macro variables

	10-year maturity		
	EBR^*_{10y}	SUBJ_BRP	GLS
Extreme tree	-1.6%	**40.9%**	**48.3%**
	(0.88)	(0.00)	(0.00)
NN - 1 Layer Group Ensem + fwd rate net	13.4%	**38.0%**	**47.2%**
	(0.26)	(0.00)	(0.00)

	2-year maturity		
	EBR^*_{2y}	SUBJ_BRP	GLS
Extreme tree	3.5%	21.3%	20.0%
	(0.80)	(0.13)	(0.16)
NN - 1 Layer Group Ensem + fwd rate net	6.4%	**27.5%**	**30.3%**
	(0.61)	(0.03)	(0.03)

Chapter 3

A Factor Model for Option Returns[1]

"Order and simplification are the first steps towards mastery of a subject. The actual enemy is the unknown."

– Thomas Mann; The Magic Mountain (1924)

3.1 Introduction

Asset pricing aims to understand the risk-reward trade-off that investors face in financial markets. The most common empirical approach for evaluating this trade-off is to model returns with a low-dimensional common factor structure. The structure of options contracts makes it difficult to model their returns in this way. Instead, the literature has studied the risk-return trade-off in options using parametric no-arbitrage pricing models. These models are parametric in the sense that they require a full specification of underlying distributions and dynamics of the underlying. While this approach benefits from arbitrage-free pricing and mathematical elegance, it is prone to model misspecification and likely too simplistic to describe empirically observed patterns of options returns. In fact, when applying no-arbitrage models in practice the researcher finds that they fail to account for a large part of the empirically observed variation in option returns (e.g. see Israelov and Kelly, 2017).

[1]This chapter is based on a similarly titled research paper jointly authored with Bryan Kelly. The paper has been presented at the FMA Derivatives & Volatility Workshop 2019, Western Finance Association Meeting 2020, the Econometric Society Word Congress 2020, and the Warwick University Finance Brown Bag Seminar.

Our objective is to develop an understanding for the risk-return trade-off in option markets using the factor pricing approach commonly applied in other asset classes. The motivation for factor modelling is independent of the asset class studied: from the asset pricing Euler equation and the assumption of no-arbitrage, a stochastic discount factor, m_{t+1}, exists that satisfies $E_t[m_{t+1}r_{i,t+1}] = 0$, and hence

$$E_t[r_{i,t+1}] = -\frac{\mathrm{Cov}_t\left(m_{t+1}, r_{i,t+1}\right)}{\mathrm{Var}_t\left(m_{t+1}\right)} \frac{\mathrm{Var}_t\left(m_{t+1}\right)}{\mathrm{E}_t\left[m_{t+1}\right]}. \tag{3.1}$$

For the first ratio in Equation (3.1) we adopt the notation $\beta_{i,t}$ and note that it describes the exposure of asset i to systematic risk factors. For the second ratio we adopt the notation λ_t. It can be interpreted as the price of risk associated with factors. When m_{t+1} is linear in factors f_{t+1}, as assumed in many asset pricing studies, the cross-section of excess returns satisfies a linear factor model:

$$r_{i,t+1} = \alpha_{i,t} + \beta'_{i,t} f_{t+1} + \epsilon_{i,t+1}, \tag{3.2}$$

where for all i and t we have $\mathrm{E}_t[\epsilon_{i,t+1}] = \mathrm{E}[\epsilon_{i,t+1}f_{t+1}] = 0$, $\mathrm{E}_t[f_{t+1}] = \lambda_t$, and $\alpha_{i,t} = 0$.

Typical factor models used in asset pricing are significantly less restrictive than parametric no-arbitrage option pricing models since they forgo the need to fully specify underlying distributions and dynamics. However, the common factor modelling approaches frequently rely on pre-ordained factors and require a reasonably long time series of asset returns to estimate betas. This factor modelling approach is difficult to apply to options data for a few reasons. First, the short lives and rapid migration of option risk attributes (such as moneyness and maturity) make it hard to estimate betas with time series regression. Second, the factors are difficult to ascertain a priori. If instead one were to consider estimating a latent factor model, the same complications that make it difficult to estimate option betas take the de facto method of principal components analysis off the table due to its reliance on static betas.

In this paper we take a different tack by treating risk factors as latent, but allowing for time variation in the factor structure. We use the instrumented principal components analysis (IPCA) methodology of Kelly, Pruitt and Su (2019), which explicitly accounts for time variation in individual asset behaviour by allowing risk factor loadings, $\beta_{i,t}$, to depend on observable characteristics of options. These characteristics function as instrumental variables for conditional betas and avoid the limitations of estimating betas with time series regression.

The riskiness of options changes constantly as the contract maturity winds down and as the underlying price and volatility evolve. Previous literature studying the cross-section of option returns, which has dealt with this complication by using empirical techniques such as portfolio sorts to construct ad hoc pre-specified factors, has seen limited success in describing option returns.[2] The fraction of variation in option returns explained by such approaches is comparatively small. This is primarily due to standard equity-based asset pricing techniques being ill-equipped to deal with time-varying factor loadings. For example, rolling regressions are too slow to capture changes in contract risk. In contrast, IPCA offers an internally consistent approach to estimate conditional loadings and factors simultaneously. And option contracts are special in that some of their characteristics, like moneyness and maturity, are easy to measure and unambiguously relevant to the contract's factor risk. IPCA is in this way particularly well-suited to the analysis of option returns.

3.1.1 Findings

We assess the performance of option return factor models in two dimensions. The first is how well the estimated factors and betas capture the contemporaneous variation in realised returns. In particular, we evaluate candidate models in terms of their "total R^2" which measures the fraction of variance in individual option contract returns $r_{i,t+1}$ explained by $\hat{\beta}'_{i,t}\hat{f}_{t+1}$, where $\hat{\beta}_{i,t}$ are estimates of conditional loadings on estimated latent risk factors $\hat{f}_{t+1}$.

Second, we inspect a model's match of differences in expected returns across assets, which describes the accuracy of the model-implied risk-return trade-off. We measure this as a "predictive R^2" or the fraction of variance in realised returns $r_{i,t+1}$ explained by the model-implied conditional expected returns $\hat{\beta}'_{i,t}\hat{\lambda}$, where $\hat{\lambda}$ denotes the model's estimated risk prices.[3] We compute both measures by aggregating over

[2] These papers include among others Coval and Shumway (2001); Goyal and Saretto (2009); Frazzini and Pedersen (2012); Cao et al. (2015); Karakaya (2013)

[3] Alternatively, one might entertain time-varying λ_t. However, the gains from allowing for time-varying risk prices, if any, are small.

all option contracts and time periods.

We apply the IPCA methodology to the panel of monthly S&P 500 option returns from 1996 to 2017. Most prior literature restricts their analysis of option returns to particular subsets of option contracts, for example focusing on at-the-money contracts with one month to maturity, because it makes their analyses less sensitive to misspecification biases that are exacerbated when looking across contracts with different moneyness and maturities. In contrast, our goal is to achieve an accurate description of return variation and risk compensation for a wide heterogeneity of contracts in a single model.

We begin with a one-factor IPCA specification and with a set of characteristics consisting of option moneyness, time to maturity, implied volatility, embedded leverage, and Black-Merton-Scholes (BMS) "Greeks." A single latent factor is sufficient to explain around 73% of the return variation in delta-hedged option returns. None of the pre-specified factor models that we study, including variations of the Fama-French-Carhart model and two models that additionally include the Frazzini and Pedersen (2012) betting-against-beta (or embedded leverage) factor and the Coval and Shumway (2001) option straddle factor, achieve a total R^2 close to that of IPCA. Additional factors in IPCA improve the models ability to describe joint fluctuations in the panel of option returns. The total R^2 for the model with five factors reaches 91%.

Further, the IPCA model allows us to test whether average option returns are explained by factor risk or if some contracts earn additional compensation (i.e., alpha) above and beyond that warranted by their risks. Across all numbers of factors examined we find, on average, no evidence of alpha in option returns. However, note that this finding does not rule out existence of significant alphas in subsets of contracts, similar to the findings of Jones (2006). Indeed, in an analysis of alphas in subsets of our sample (i.e. characteristics managed portfolios) we find evidence for some significant managed portfolio alphas.

In terms of predictive R^2, the IPCA model generally outperforms pre-specified factor models by producing more accurate model-based option return forecasts (both in-sample and out-of-sample). IPCA is particularly dominant when the competing pre-specified factors models use static betas. When we introduce dynamic betas on pre-specified factors as functions of contract characteristics similar to those in the

IPCA specification, the predictive R^2 for pre-specified factor models rises to match what we find for IPCA, though the total R^2 remains substantially lower than in IPCA. Overall, we find that the IPCA model provides the most accurate description of the risk-return trade-off for index options – its portfolio alpha is on average about one percentage point per annum smaller than those in the model with pre-specified factors.

We arrive at an interpretation of the factors recovered by IPCA through a comprehensive correlation analysis with a number of relevant financial time series, including returns on moneyness and maturity-sorted option portfolios, option return factors studied in prior literature, and proxies of liquidity and intermediary capital constraints. Our analysis shows that IPCA recovers a tail risk factor (Factor 1) that is highly correlated with returns on deep out-of-the-money options and a variance risk premium factor (Factor 5) that behaves similarly to returns on a strategy that sells at-the-money straddles.[4] The interpretation of the other factors is more challenging. We find evidence for a maturity-risk factor (Factor 3) that captures differences in returns of short- and long-dated contracts. This factor also correlates with the volatility term structure factor of Karakaya (2013). The next factor (Factor 4), exhibits similarity to the "value" factor in Karakaya (2013) that is constructed as the spread between historical and BMS implied volatility. Following the interpretation attempt by Karakaya (2013), this factor might capture institutional demand pressures. In support of this idea, we find evidence that this factor is related to measures of option market liquidity such as bid-ask spreads and open interest. The final factor (Factor 2), is most challenging to interpret. It correlates most strongly with the returns on a portfolio of options with one month to maturity that shorts high moneyness contracts and longs low moneyness contracts.

IPCA factor Sharpe ratios range from 0.2 for the tail risk factor to around 1.9 for the variance risk premium factor. Furthermore, we calculate out-of-sample Sharpe ratios for the tangency portfolio of IPCA factors and compare with the Sharpe ratios corresponding to observable factor models. We find that for any number of factors studied, the IPCA factors generate out-of-sample tangency portfolio Sharpe ratios that exceed those of the observable factor models. The out-of-sample tangency portfolio Sharpe ratio peaks for the IPCA model with $K = 3$ factors at

[4]For ease of reference, we adopt the convention to order factors by their variances, i.e. "Factor 1" corresponds to the factor with highest time series variance, while "Factor 5" has lowest variance.

1.30. Along with evidence for three priced factors in the in-sample analysis, we conclude that the empirical evidence points towards three priced factors in delta-hedged option returns as being most likely.

We examine which characteristics are the most important contributors to the accurate fits of IPCA. Option moneyness and implied volatility matter most for describing time variation in betas and produce the largest contributions to total R^2. Options' BMS Gamma, a measure of sensitivity to jump risk, also enters prominently. In addition to studying the importance of characteristics in our baseline model specification, we assess our choice of characteristics against an economically motivated choice of characteristics from an option return decomposition. We find that although the economically motivated set of characteristics is able to capture a similar amount of variation in option returns, our more agnostic baseline specification yields a superior risk description in terms of predictive R^2.

While most of our analysis uses monthly data, we also evaluate the IPCA model performance for daily returns. This analysis is motivated by Carr and Wu (2020), who propose a decomposition of daily option returns using a BMS approximation that attributes return performance to each of the BMS Greeks. While this attribution reliably captures returns of short-dated at-the-money contracts (where the approximation is most apt), it struggles to describe returns of out-of-the money contracts and of contracts with longer maturities. We show that, in contrast to the Carr and Wu (2020) attribution, IPCA offers a uniformly accurate description of risk and return for options throughout the moneyness-maturity spectrum.

3.1.2 Related Literature

Our paper relates to three strands of literature. The first seeks to understand option returns by sorting options into portfolios based on a small number of characteristics, then measuring the full sample average returns of these portfolios. This literature includes among others Coval and Shumway (2001), Bakshi and Kapadia (2003), Goyal and Saretto (2009), Frazzini and Pedersen (2012), Cao and Han (2013), Karakaya (2013), Cao, Han, Tong and Zhan (2015),Vasquez (2017), and Israelov and Tummala (2017).

A second strand of literature models option prices directly by enforcing no-arbitrage restrictions and specifying the full distributional properties and dynamics of the underlying asset. Starting from the seminal Black and Scholes (1973) model for vanilla options, this literature develops refinements that allow for various forms of stochastic volatility and jumps (Heston, 1993; Duffie et al., 2000; Carr and Wu, 2004).[5] While this approach has the advantage of imposing economically meaningful no-arbitrage restrictions that guarantee consistent pricing across strikes and maturities, they ultimately lead the researcher to sacrifice some realism for the sake of mathematical tractability. In fact, earlier empirical research demonstrates that arbitrage opportunities in option markets exist.[6] As a result these models often have difficulty matching the empirical behaviour of option returns (Israelov and Kelly, 2017).

The third, and most closely related, literature acknowledges the limitations of no-arbitrage option pricing models, and directly models option returns with pre-specified factors. Jones (2006) estimates non-linear factor models for short-term deep out-of-the-money S&P 500 index options, allowing for potentially latent factors that eventually manifest as volatility and jump risks factors. Karakaya (2013) is focused on single-name equity options and proposes a three factor model consisting of a level, slope and value factor that capture much of the variation in delta-hedged option returns. Israelov and Kelly (2017) propose an approach for estimating conditional option return distributions using semi-parametric time series techniques. Carr and Wu (2020) develop a valuation framework that attributes daily option returns to variation in the first and second moments of the underlying price and underlying volatility. Brooks, Chance and Shafaati (2018) employ the LASSO estimator in the context of individual equity options to uncover the characteristics that provide independent information for the cross-section of option returns. Finally, Christoffersen, Fournier and Jacobs (2018) investigate a factor structure in single-name equity options using principal components of equity volatility, skews and term structures and document that these principal components explain a sizeable fraction of the cross-sectional variation in option returns.

Our paper differs from prior literature by proposing an internally consistent factor-based approach to modelling option returns. It tackles the challenges of estim-

[5]Further papers studying models with price and volatility jump specifications are Eraker et al. (2003); Eraker (2004); Broadie et al. (2007).

[6]For example, see Ofek and Richardson (2003); Ofek, Richardson and Whitelaw (2004); Constantinides, Jackwerth and Perrakis (2009); Chambers, Foy, Liebner and Lu (2014)

ating factor betas by leveraging contract characteristics as conditioning instruments. Apart from delivering a richer factor model for returns through characteristics, our IPCA approach estimates latent factors without ex ante knowledge of the cross-section of returns, hence eliminating the need for the researcher to take a prior stance on the nature of the factors. And, in contrast to the prior literature, our approach models conditional factor loadings, which is critical due to the rapidly evolving risks at the individual option level. Notably, our approach does not enforce no-arbitrage across contracts, which allows us more flexibility in modelling the cross-section of index option returns and results in a better fit for realised returns. Importantly, the benefits of this flexibility are *not* an artefact of overfit. We show that our model continues to excel in purely out-of-sample assessments.

The paper proceeds as follows. Section 3.2 introduces the data and variables used in the analysis, Section 3.3 recaps the instrumented principal components methodology, Section 3.4 summarises our empirical results, and Section 3.5 concludes.

3.2 Data

In this paper we focus on options on the S&P 500 index.[7] Daily option data is obtained from OptionMetrics for the period of January 1996 to December 2017. The information provided by OptionMetrics includes contract specifications (exercise date, strike, etc.) as well as underlying index values, historical dividend yields, and option sensitivity measures such as the BMS Delta, Gamma, Vega, and Theta. Data for the VIX index is obtained through CBOE.

To introduce notation, each option contract i is defined by its strike price K_i and maturity date T_i. Hence, the time-to-maturity is given as $T_i - t$. As a measure of the position of an option contract i relative to its strike K_i, we compute moneyness standardised by implied volatility as

$$m_{i,t} = \frac{log(K_i/S_{i,t})}{IV_{i,t}\sqrt{T_i - t}},$$

where $IV_{i,t}$ denotes the BMS implied volatility of the contract.[8]

[7]Note, the framework employed here also lends itself to the study of single-name options in a characteristics-rich environment. However, for the sake of simplicity, in this paper, we focus on index options.

[8]In our definition of contract moneyness we follow Israelov and Kelly (2017).

Our results focus on monthly holding period returns that are delta-hedged daily.[9] Option returns are computed for periods defined by the expiration date in a given month which usually is the third Friday in a month, i.e. we compute monthly holding period returns from daily data over periods starting on the next trading day after the expiration date and ending with the expiration date.

Variation in the price of the underlying is the most important driver of option returns. By delta-hedging the option contracts daily, we obtain an option return in excess of the fraction explained by variation in the underlying. This choice is common in the academic context (e.g., see Cao and Han, 2013). The delta-hedged option gain for a contract with value F over a period $t = 1, \ldots, T$ is given by

$$\begin{aligned} \Pi_{[1,T]} = & \sum_{t=1}^{T-1} F_{t+1} - F_t \\ & - \sum_{t=1}^{T-1} \Delta_{F,t} \left(S_{t+1} - S_t \right) \\ & - \sum_{t=1}^{T-1} \frac{a_{t,t+1} r_t}{365} \left(F_t - \Delta_{F,t} S_t \right), \end{aligned} \tag{3.3}$$

where the first term is the raw option profit / loss, the second term captures the adjustment from delta-hedging the position, and the last term adjusts for the cost of funding the delta-hedged portfolio at the risk-free rate where $a_{t,t+1}$ is the number of calendar days between trading dates t and $t + 1$.

We compute returns not against the prevailing option mid-price, but against the mid-price of the underlying. This ensures that returns are well-behaved even in situations when the option mid-price is close to zero as is common for deep out-of-the-money (OTM) options. We denote the delta-hedged return against the spot price by r^{Δ}_{Spot}.

We apply a number of filters to the OptionMetrics data. We remove observations with negative bid-prices, observations where the bid-price exceeds the ask-price, and observations that violate no-arbitrage conditions. In a small number of cases OptionMetrics has missing information for contract implied volatility, we delete those observations. Furthermore, numerous contracts have a zero observation for open interest, i.e. the number of open positions in the option contract. This is

[9] In Section 3.4.8, we perform a robustness exercise and assess the performance of our model at daily frequency.

particularly true for deep in-the-money options for which liquidity is generally low. We remove observations with zero open interest. Following Karakaya (2013), we remove the most extreme contracts as measured by embedded leverage

$$\Omega = |\Delta \cdot S/F| \,. \tag{3.4}$$

Specifically, we drop observations below (above) the 1st (99th) percentile of the distribution of embedded leverage on our portfolio formation dates.

After applying the aforementioned filters, we limit the set of options to at-the-money (ATM) and out-of-the-money (OTM) contracts that are generally more liquid and hence at the centre of most academic and practitioner research. To do so, we limit the moneyness range to 0 to 2 for call options and -2.5 to 0 for put options. Furthermore, we limit our sample to options with time-to-maturity (TTM) between 1 month and 12 months. Although, we "limit" our sample in this way, our sample still encompasses a far larger fraction of the outstanding contracts than most papers in the literature that limit their sample to, say, at-the-money contracts. The main motivation behind our choice of filters is to remove outliers and contracts with little liquidity that are of little economic relevance in the market place.

Table 3.1 contains summary statistics for our sample. After applying all aforementioned filters, we arrive at a sample of approximately 77,000 option-month observations. Out of those observations a little more than two-thirds are put options. This is a consequence of our choice of filters that considers a wider range of moneyness for puts than for calls. Also, since over the course of our sample period the underlying S&P 500 index appreciates considerably, OTM put options are naturally more abundant than OTM calls. In addition, the put / call ratio for options on the S&P 500 is commonly higher than one as market participants demand more put options for insurance purposes.

Construction of the Implied Volatility Surface For replication of the Carr and Wu (2020) P&L attribution in Section 3.4.8, we construct an implied volatility surface to obtain moments of the underlying and its volatility under the risk-neutral measure. We construct implied volatility surfaces using the daily OptionMetrics volatility file that contains volatilities over a prespecified moneyness / maturity grid. For the construction of the volatility surface we carry out a second order smooth bivariate spline interpolation in the time-to-maturity vs. moneyness space.[10] Since it is common for implied volatilities of put and call options at the same moneyness / maturity to disagree as a consequence of differences in price pressure, we construct three different volatility surfaces: the first is constructed solely from observations of put options in the OptionMetrics volatility files, the second is constructed from observations of call options, and the third is constructed using both put and calls. In the latter case, the implied volatilities of put and call options are weighted by their market capitalisation, i.e. open interest times option mid-price.

3.3 Methodology

In this paper, we aim to recover a factor structure in realised option returns. To this end, we use the instrumented principal components model of Kelly et al. (2019). The model is specified for a general excess return $r_{i,t+1}$ as follows

$$\begin{gathered} r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1} \\ \alpha_{i,t} = z'_{i,t}\Gamma_\alpha + \nu_{\alpha,i,t}, \quad \beta_{i,t} = z'_{i,t}\Gamma_\beta + \nu_{\beta,i,t} \end{gathered} \tag{3.5}$$

The system is estimated over a total of N assets and T periods. The loadings, $\beta_{i,t}$, are time-varying and partially depend on an $L \times 1$ vector of (option) characteristics $z_{i,t}$. For ease of notation we assume that $z_{i,t}$ includes a constant. The vector of factors, f_{t+1}, is of dimensionality $K \times 1$, where the number of factors, K, is to be determined in a data-driven approach. Following Kelly et al. (2019), the IPCA model can be estimated by means of an alternating least squares approach that effectively switches back and forth between the first order conditions of Γ[11] and f_{t+1}.[12]

The application of IPCA in the context of option returns can be motivated in several ways. First, observable characteristics feature centrally in the IPCA framework as they make the estimation of factors more efficient and improve model performance. Contrary to common equity, option contracts are defined through

[10] For details, see the SciPy function *SmoothBivariateSpline*, documented at `https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.SmoothBivariateSpline.html`.

[11] For brevity, we denote Γ_α and Γ_β jointly as just Γ. The IPCA model can equivalently be stated as $r_{i,t+1} = z'_t\Gamma\tilde{f}_{t+1} + \epsilon_{i,t+1}$ where $\Gamma = [\Gamma_\beta, \Gamma_\alpha]$ is a horizontally stacked matrix and $\tilde{f}_{t+1} = [f_{t+1}, 1]$.

[12] For the Python implementation used to carry out the estimation we refer the reader to `https://github.com/bkelly-lab/ipca`.

a set of precisely measured characteristics such as their strike, time-to-maturity, implied volatility, their location on the implied volatility surface, and market / dividend yields. Through the dimensionality reduction that is integral to IPCA, a large number of potential characteristics can be entertained simultaneously. IPCA will then form a linear mapping Γ_β between characteristics and factors that isolates the informative signal coming from characteristics and averages out the noise. This eliminates the need for the researcher to take an ad-hoc stance on the characteristics that matter. Second, time-varying loadings $\beta_{i,t}$ that depend on characteristics $z_{i,t}$ allow us to model the conditional behaviour of option returns. The "identity" of an option contract is to a good extent determined by its location in the time-to-maturity / moneyness space. IPCA tracks the migration of the asset in this space through the conditional betas. Furthermore, time-varying betas allow the researcher to entertain a large number of characteristics even if they are are only relevant during specific periods. Third, by restricting the IPCA model (3.5) such that $\Gamma_\alpha = 0$, we can test whether risk compensation in option returns solely arises from exposure to systematic factors, f_t, or whether returns partially line up with characteristics directly (i.e. $\Gamma_\alpha \neq 0$), hence constituting compensation without risk. Fourth, existing approaches in the literature that construct observable risk factors for option returns have had limited success and usually explain only a fraction of the empirically observed variation in option returns. IPCA forms factors in a statistical approach. The researcher is then tasked with interpreting the estimated factors by relating them to economic theory.

Asset Pricing Performance To assess the performance of the IPCA factors in pricing option returns, we use two measures following Kelly, Pruitt and Su (2019). Similar to a common R^2 goodness-of-fit measure we compute

$$R^2_{total} = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1})\right)^2}{\sum_{i,t} r^2_{i,t+1}}. \tag{3.6}$$

This total R^2 measures how well the set of factors and loadings captures realised returns. As a second measure, we compute

$$R^2_{pred} = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - z'_{i,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{\lambda})\right)^2}{\sum_{i,t} r^2_{i,t+1}}, \tag{3.7}$$

where $\hat{\lambda}$ denotes the unconditional time series mean of the factors. This predictive R^2 captures how well differences in average returns are explained through the model's description of conditional expected returns, i.e. the model's ability to describe risk.

In addition to studying the pricing performance for individual option contracts, it is insightful to examine the pricing performance for portfolios. The IPCA methodology elegantly incorporates a portfolio notion that circumvents a common problem in the asset pricing literature that is the choice of relevant test assets. To this end, Kelly et al. (2019) employ the notion of characteristics managed portfolios. Let Z_t be an $N \times L$ matrix of characteristics at time t. Then managed portfolios can be constructed via

$$x_{t+1} = \frac{Z_t' r_{t+1}}{N_{t+1}}, \tag{3.8}$$

where N_{t+1} is the number of outstanding options at time $t+1$. The managed portfolios, x_{t+1}, are a weighted average of option returns where the weights are determined by the characteristics in Z_t. Analogously to total and predictive R^2 at the individual option level, we can define performance measures for managed portfolios

$$R^2_{total,x} = 1 - \frac{\sum_{l,t}\left(x_{l,t+1} - z_{l,t}' z_{l,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1})\right)^2}{\sum_{l,t} x_{l,t+1}^2}, \tag{3.9}$$

$$R^2_{pred,x} = 1 - \frac{\sum_{l,t}\left(x_{l,t+1} - z_{l,t}' z_{l,t}(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{\lambda})\right)^2}{\sum_{l,t} x_{l,t+1}^2}, \tag{3.10}$$

where $l = 1, \ldots, L$.

3.4 Empirical Results

3.4.1 Data

From the data set described in Section 3.2, we collect a number of option characteristics and market variables. Option-level variables include moneyness (`mness`), time-to-maturity (`ttm`), embedded leverage (`embed_lev`), BMS implied volatility (`impvol`), BMS Theta (`theta`), BMS Gamma (`gamma`), BMS Vega (`vega`), BMS Volga (`volga`).[13] In all specifications we interact aforementioned variables with

[13]The sensitivity of Vega to changes in volatility, here referred to as Volga = $\frac{\partial Vega}{\partial \sigma}$, is not part of the set of Greeks computed by OptionMetrics as standard and therefore computed from the

both a constant and an indicator variable that is equal one for put options and equal zero for calls, i.e. we allow for the possibility that characteristics determine option returns differently for calls and puts. In order to limit the impact of outliers and aid the interpretation of the IPCA results, we re-scale all characteristics to the range $[-0.5, 0.5]$. By scaling the characteristics in this manner, we focus on the ranking of characteristics in the cross-section which is the main determinant of differences in returns across contracts.[14]

3.4.2 IPCA Performance

To begin with, we estimate IPCA for a set of common option characteristics that have both academic and practical relevance. The baseline set of characteristics includes `mness`, `ttm`, `embed_lev`, `impvol`, `theta`, `gamma` and `vega`, yielding a total of 15 characteristics after interaction with the put / call dummy and addition of a constant. The dependent variable is the monthly holding period return that is delta-hedged daily, r^{Δ}_{Spot}, as defined in Section 3.2.

Table 3.2 details the performance of the restricted ($\Gamma_\alpha = 0$) and unrestricted ($\Gamma_\alpha \neq 0$) IPCA models with $K = 1, \ldots, 5$ factors. With a single factor, IPCA explains more than 72% of the observed variation in option returns in the restricted model ($\Gamma_\alpha = 0$) and more than 74% in the unrestricted model ($\Gamma_\alpha \neq 0$). When including an additional IPCA factor, the R^2_{total} increases by about 10% to around 80% for the restricted and 82% for the unrestricted specification. Adding further factors, the increases become more gradual. We find that five factors are needed to explain more than 90% of the variation in monthly delta-hedged index option returns. To disentangle the gain from allowing for latent factors from the fraction of return variation explained by the characteristics on their own, we run a panel regression of option returns on the same set of characteristics as used in IPCA. These regressions are identical to the estimation of IPCA with a single pre-specified constant factor. Table 3.3 details the results of the panel regression. We find that characteristics on their own explain around 3.4% of the empirically observed variation in returns.

The estimation of IPCA through a least squares criterion targets the total

Black-Scholes pricing formula.

[14] Rescaling is carried out before interacting the characteristics with the put / call dummy variable. This ensures that the scaling on both the non-interacted and the interacted characteristic is comparable.

R^2. This makes the factors identified by IPCA optimal to capture return variation across options. However, the IPCA factors are not ex ante designed to yield a good description of expected returns as measured through the predictive R^2. We find that the monthly predictive R^2 for a single factor is 5.48% in the restricted model and 7.22% in the unrestricted model. When including additional factors, the predictive R^2 increases in the restricted model, while it decreases slightly in the unrestricted model.

It is common practice in the empirical asset pricing literature to examine the explanatory power of asset pricing models using portfolios such as, for example, the 5×5 Fama-French size and book-to-market sorted portfolios as test assets.[15] The IPCA framework can be approximately stated in terms of managed portfolios $x_t = Z'_{t-1} r_t / N_{t-1}$ (see Section 3.3 for a discussion), where N_t is the number of assets in the cross-section at time t. This construction yields an $L \times 1$ vector x_t, where L corresponds to the number of considered characteristics. We then proceed to compute the performance measures for the managed portfolios as test assets. We notice that the R^2_{total} is markedly higher for managed portfolios than for individual contracts. For example, using just a single factor the total R^2 is 96.00% in the unrestricted model and 95.47% in the restricted model. When moving from individual options to managed portfolios we find that the predictive R^2 increases for the restricted model, while it slightly decreases for the unrestricted model with $\Gamma_\alpha \neq 0$.

In Panel C of Table 3.2 we display p-values from a bootstrap with 1000 draws for the hypothesis test $H_0 : \Gamma_\alpha = 0$. The bootstrap follows the residual bootstrap procedure outlined in Kelly et al. (2019). In all specifications of the IPCA model with $K = 1, \ldots, 5$ factors, we find that we cannot reject the null. We interpret this result as evidence that characteristics help to explain risk through systematic factors, but not on their own. It is important to point out that Γ_α captures the aggregate effect that characteristics have on returns. Hence, it is still a possibility that characteristics on their own explain returns for a subset of contracts such as for example deep OTM options that are commonly considered mispriced. This finding is related to Jones (2006) who finds that on average alpha's are close to zero. However, he argues that this finding is a consequence of alpha's averaging out in aggregate.

[15]The asset pricing literature has entertained a large range of potential test assets. This is a sign for the struggle that this literature has long faced in determining the set of assets that need to be priced in order for a model to be accepted.

Indeed, Jones finds significant mispricing of specific sub-groups of option contracts as for example deep OTM options. We examine mispricing of particular sub-groups of options in a more detailed asset pricing study in Section 3.4.3. Furthermore, Jones notes that mispricing is reduced by increasing the number of factors. We find evidence in a similar spirit: as we increase the number of factors the discrepancy in terms of total / predictive R^2 between the unrestricted model with $\Gamma_\alpha = 0$ and the restricted model with $\Gamma_\alpha \neq 0$ vanishes. This suggests that as we allow for more factors, the incremental contribution of characteristics lining up with returns on their own becomes smaller. For the model with $K = 5$ factors, the relative difference in total R^2 between the restricted and unrestricted models is a mere 0.2% at the individual asset level and effectively zero at the portfolio level.

Performance in the Cross-Section of Contracts The IPCA framework is designed to target the aggregate R^2_{total} across all option contracts simultaneously. While a small number of factors explains a high degree of empirically observed option returns on aggregate, it is less clear how IPCA fares in the cross-section when we focus on groups of contracts that are of particular interest to practitioners as for example out-of-the-money options.

For the following analysis, we focus on the restricted IPCA model with $K = 5$ factors. We choose this specification for two reasons. First, the earlier bootstrap test of the hypothesis W_α motivates our choice of the restricted model in favour of the unrestricted counterpart. Second, our focus on the specification with $K = 5$ factors is motivated by our finding of increasing predictive R^2 in the restricted specifications. Also, the increase in total R^2 when including additional factors one at a time appears to level off at around four to five factors as is noticeable in the modest increase in total R^2 when moving from the model with $K = 4$ to the model with $K = 5$. Table 3.4 breaks down the IPCA performance by bins sorted on moneyness, time-to-maturity, and VIX index level. The results by bin are obtained by computing the total and predictive R^2 only from those option-month observations corresponding to a particular bin.

Panel A of Table 3.4 shows that IPCA performance is particularly strong in terms of both total and predictive R^2 for ATM options. Panel B of Table 3.4 details results sorted on the option's time-to-maturity. We find that IPCA performs best for options with maturities greater than one month in terms of the total R^2, while the predictive R^2 peaks for contracts with two months to maturity. Performance is noticeably weaker for short-dated contracts with 1 month to expiry, highlighting that these contracts present themselves as particularly challenging. In order to assess the model's performance in different market environments (low vs. high volatility regimes), Panel C dissects the performance by level of the CBOE VIX index. We find that the IPCA model captures return variation across a range of market regimes. Only for very extreme market conditions as indicated by VIX levels greater than 50%, the predictive R^2 turns negative.

These results suggest that IPCA captures not only observed return variation in a subset of the cross-section of index options but in the overwhelming majority of them. Due to their abundance in our sample, at-the-money contracts have a relatively higher weight during the IPCA estimation and as evidenced above they are the sort of contracts best captured by our model. Nevertheless, option contracts more extreme in moneyness and implied volatility are reasonably well captured in an IPCA model using only standard option characteristics and a small number of factors.

Comparison with Extant Factor Models

We now turn to comparing the performance of the IPCA framework to a number of observable risk factor models entertained in earlier literature. We consider five different observable factors. As a model with a single factor, the market factor, we use the capital asset pricing model (CAPM). The Fama-French (1993) three factor model (FF3) includes additionally the size (SMB) and value factors (HML). We add the momentum factor (UMD) to obtain the Carhart four factor model (FFC4). As demonstrated in earlier literature, the standard risk factors used to price equities are insufficient to price delta-hedged option returns.[16] For this reason, we include two additional factors constructed specifically from S&P 500 index options.

The first is the betting against the beta (*bab*) factor of Frazzini and Pedersen (2012) constructed from embedded leverage, Ω, as defined in Equation (3.4). Specifically, *bab* is constructed by sorting options into a high- and low-embedded leverage

[16] See, e.g., Coval and Shumway (2001); Cao and Han (2013); Frazzini and Pedersen (2012); Karakaya (2013) for factor models entertained in the earlier literature.

portfolio using the cross-sectional median of embedded leverage as the breakpoint. We weight returns in the two portfolios using the option's market capitalisations, open interest times price. Then the self-financing, zero-beta factor return is given by

$$bab_t = r_t^L/\Omega_{t-1}^L - r_t^H/\Omega_{t-1}^H,$$

where Ω_t^L and Ω_t^H are the weighted average embedded leverage of the low and high embedded leverage portfolios, respectively. The *bab* factor is constructed separately for call and put options. Finally, we average the bab factor of calls and puts to give an aggregate factor as follows

$$bab_t = (bab_t^{Call} + bab_t^{Put})/2.$$

The second option risk factor we include is the *straddle* factor of Coval and Shumway (2001) as constructed in Frazzini and Pedersen (2012). For construction of the straddles we limit the set of considered contracts to those with 1 month to expiry. Then, we construct straddles from all pairs of put and call options with absolute BMS Delta between 0.4 and 0.6. The straddle factor return is obtained by weighting the delta-hedged straddle returns with the market capitalisations of the constituting options as measured by open interest times price.

We add the *bab* factor to the "FFC4" four factor model and refer to the resulting five factor model as "FFCB5". Finally, we add the straddle factor to the "FFCB5" model and refer to the resulting model as "FFCBS6".

The factors just introduced are examined in two different model implementations. The first version places the observable risk factors in a setting akin to the IPCA latent factor model by allowing for characteristics as instruments (see Equation (14) in Kelly et al. (2019)). We can estimate this implementation by pre-specifying the factors in the IPCA model with one of the observable factor models above, leaving us only with an estimation of the loadings matrix Γ_β by evaluating the associated first order condition. The second implementation follows a rather standard empirical asset pricing procedure in which betas are estimated from time series regressions. Following Kelly et al. (2019), we implement this approach by means of a panel regression of option returns onto the set of factors (without an intercept).

Table 3.5 compares the different implementations. In Panel A we recap the performance for the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 1, \ldots, 5$ factors for both individual option returns as well as characteristics managed portfolios.

Panel B reports results for the implementation of IPCA with prespecified observable risk factors and characteristics as instruments, i.e. here we allow for characteristics to instrument time-varying loadings for observable risk factors. We find that the IPCA model with a single factor outperforms the fits of the observable factor models in terms of the total R^2 for all combinations of observable factors that we examine. We note that adding the observable factors constructed from option returns (models "FFCB5" and "FFCBS6") considerably improves the model fit. The R^2_{total} increases by around 16% versus the Carhart model ("FFC4") when the embedded leverage factor is included and by around 62% when also the straddle factor is included. With respect to the observable factor models' ability to describe risk across assets (predictive R^2) the IPCA model with $K = 5$ factors generates a predictive R^2 that is on par with the best performing observable factor model "FFCBS6".

In Panel C, we summarise the fits from panel regressions without instruments. It is easy to see that static betas are detrimental to model fit: total R^2s for individual options as test assets shrink by about one third, while the effect is less strong when characteristics managed portfolios are examined. Similarly, the models ability to explain average returns in terms of predictive R^2 is noticeably weaker such that the best performing observable factor model "FFCBS6" is now outperformed by all IPCA model specifications.

Taken together, this evidence suggests that the IPCA framework captures differences in average returns across assets (R^2_{pred}) similarly well as a number of benchmark observable factor models instrumented with characteristics. However, IPCA is able to better capture the empirically observed variation (R^2_{total}) in option returns. Furthermore, when observable factor models are not instrumented with characteristics (Panel C), both total and predictive R^2 are noticeably lower. This suggests that introducing time-varying loadings via option characteristics plays a major role in describing risk differences across options.

3.4.3 Unconditional Asset Pricing Performance

The IPCA model is a conditional asset pricing model since loadings are instrumented by a set of conditioning characteristics. Thus, it is not clear ex ante whether the set of IPCA factors is unconditionally mean-variance efficient, i.e. whether they price assets unconditionally. So far, when testing for the significance of Γ_α, we have found that characteristics on their own do not align with returns. This suggests that our factors are conditionally mean-variance efficient. In order to test the unconditional mean-variance efficiency of our factors, we carry out a number of asset pricing tests.

We test the model's ability to price test portfolios unconditionally by running time series regressions of test portfolio returns on the set of IPCA factors. As a benchmark, we pick the best performing observable factor model that includes the Fama-French three factor, momentum, the straddle factor as well as the betting against beta factor constructed from embedded leverage ("FFCBS6"). We study two sets of anomaly portfolios. First, the managed portfolio notion in IPCA immediately provides us with a set of tests assets that weights assets by their characteristics. Second, we use a set of 20 moneyness / maturity double-sorted portfolios of call and put options. We re-leverage all portfolios to yield 10% annualised volatility for comparability.

Figure 3.1 shows portfolio alphas from time series regressions of managed portfolio returns onto the set of factors. Significant alphas are denoted by filled-in diamond markers. We find that both in the observable factor model FFCBS6 and the IPCA model (Panels (a) and (b)) there is only one out of 15 portfolios with significant alpha, indicating that the unconditional pricing performance of IPCA is at least on par with a benchmark observable factor model. Furthermore, we find that the average portfolio alphas in the IPCA case are more than a percentage point lower than those in the observable factor model FFCBS6. For comparison, we plot the portfolio alphas from using IPCA in its originally intended form, that is, using conditional loadings (Panel (c) and (d)). Portfolio alphas are obtained as the time series averages of period-by-period portfolio residuals in the conditional IPCA model. We find that both for the observable factor model and IPCA four portfolios have significant alphas. However, the average absolute alpha in IPCA is more than two absolute percentage points lower.

We check the robustness of our previous results on the unconditional mean-variance efficiency of our factors by using a set of 20 BMS Delta / maturity double-sorted portfolios that could provide a more challenging test case.[17] Figure 3.2 shows the results of this test. We find that using the observable factor model, FFCBS6, six portfolios have significant alphas, while eight portfolios have significant alphas using the five IPCA factors. Although IPCA appears to be slightly weaker at pricing the 20 double sorted portfolios, its overall performance is on par with the observable factor model: average portfolio alphas are around 1 percentage point lower when using the IPCA factors than when the observable factors are used. Furthermore, when visually assessing the distribution of alphas, we find that portfolio alphas from the FFCBS6 model exhibit a clear pattern: portfolio alphas increase with raw portfolio returns. This suggests a systematic shortcoming of the observable factors in pricing the test portfolios. In fact, except for one portfolio out of the six portfolios with significant alphas in panel (a) of Figure 3.2, all significant portfolio alphas occur for the portfolios with highest average returns. In contrast, significant portfolio alphas from using IPCA pricing factors are more evenly distributed (panel (b)).

In the previous analysis of mean-variance efficiency of our IPCA factors we have focused on the model with $K = 5$ factors. In order to inspect the behaviour of portfolio alphas as we increase the number of factors, we now re-run the previous tests for the models with $K = 1, \dots, 5$ factors and compute the conditional and unconditional portfolio alphas as before. Table 3.6 summarises the results. We find that unconditional and conditional alphas decrease as we increase the number of factors. Note, this finding is *not* a guaranteed result of increasing the number of factors. Further, this result is in line with Jones (2006) who finds decreasing alphas in option portfolios with increasing number of estimated factors.

3.4.4 Out-of-Sample Performance

So far, we have demonstrated that the IPCA model achieves a description of option return variation in-sample that is at least on par with observable factors in terms of its asset pricing performance and likely better in terms of its ability to describe variation in realised returns. However, it is not clear whether this performance

[17]Double sorting on moneyness (Delta) and maturity creates a set of portfolios that is relevant in practice: according to their risk preferences investors hold options of varying riskiness and maturity. For example, investors that require portfolio insurance create a considerable demand for out-of-the money options.

translates into meaningful fits out-of-sample. Hence, we now turn to analysing IPCA's out-of-sample fits.

Procedurally, we carry out a recursive estimation of the IPCA model in a backward-looking fashion. The IPCA model is estimated for an expanding window starting at half the available sample length, i.e. the first forecast is made in January 2007. The computation of the out-of-sample realised factor returns, $\hat{f}_{t+1}$, exactly follows Kelly et al. (2019, see Section 4.4).[18]

We evaluate the out-of-sample fit of the restricted IPCA models with $K = 1, \ldots, 5$ factors using the same performance metrics introduced above, i.e. the total and predictive R^2 using both individual options as well as managed portfolios as test assets. Table 3.7 details the results of our analysis. We find that the strong in-sample performance of IPCA in terms of capturing observed return variation carries over into the out-of-sample exercise. For the model with $K = 5$ factors that most of our analysis focuses on, the total R^2 only reduces from 90.61% to 88.88%. Despite this small deterioration, IPCA still outperforms all of the in-sample fits of observable factor models such as FFCBS6. However, with reference to the predictive R^2 the deterioration is more pronounced. At the individual contract level the predictive R^2 shrinks from 6.40% to 4.37%. The predictive R^2 at the portfolio level drops from 6.51% to 2.73%. Furthermore, we notice that a model with a single factor performs best out-of-sample with respect to predictive R^2. This finding is in line with the typical bias-variance trade-off involved in fitting models. Essentially, the researcher needs to decide whether the objective is to better capture in-sample variation by allowing for more model flexibility through an increased number of factors, or whether the objective is to create a model that generalises well to unseen data, hence favouring relatively sparse models with fewer factors. While the model with a single factor has a somewhat higher predictive R^2 out-of-sample than the five factor model, the five factor model has a considerably higher total R^2.

Out-of-Sample Factor Tangency Portfolios Over and above the out-of-sample predictive accuracy of the IPCA model stands the question of whether the mean-

[18]We briefly summarise the computation of the factor realizations here: starting from an estimation of $\hat{\Gamma}_{\beta,t}$ using the available in-sample data up to time t, the out-of-sample factor realization is computed as

$$\hat{f}_{t+1} = \left(\hat{\Gamma}'_{\beta,t} Z'_t Z_t \hat{\Gamma}_{\beta,t}\right)^{-1} \hat{\Gamma}'_{\beta,t} Z'_t.$$

variance efficiency of the IPCA factors carries over into an out-of-sample setting. We assess the out-of-sample mean-variance efficiency of both the observable factor models entertained earlier and the IPCA model by computing out-of-sample Sharpe ratios of the individual factors and tangency portfolios formed from the sets of factors. We rescale portfolio weights using only backward-looking information such that a volatility target of 1% per month is attained. Table 3.8 collects the out-of-sample Sharpe ratios. The table reports in column k the univariate Sharpe ratio of the respective factor as well as the Sharpe ratio of the tangency portfolio formed from the factors 1 through k. Panel A contains the results for the IPCA factors. A single factor ($K = 1$), yields a Sharpe ratio of 0.60. Adding further factors, the Sharpe ratio peaks for the model with $K = 3$ factors at 1.30, before declining again when moving to higher numbers of factors. Despite the decline in the Sharpe ratio for more than three factors, they still exceed those of the observable factor models such as the FFCBS6 model that generates a Sharpe ratio of 0.51. Joint with the evidence of three significant in-sample factor Sharpe ratios (see Table 3.9), this suggests that out of the five factors recovered by IPCA three factors are priced factors.

3.4.5 Interpreting the IPCA Factors

Arguably, understanding the drivers of the IPCA model's performance is as crucial as a strong fit performance itself. To this end, we now examine both the mapping between characteristics and factors Γ_β as well as the time series of the IPCA factors themselves. As earlier, we focus on the IPCA model with $K = 5$ factors. We adopt the convention to order factors by their variance such that "Factor 1" corresponds to the factor with highest time series variance, and "Factor 5" corresponds to the factor with lowest time series variance.

In order to give interpretation to the factors, we primarily draw on correlations of the factors with portfolios of options sorted on maturity and moneyness. Figure 3.5 shows the correlation for each of the five factors with those portfolios. For additional evidence we draw on the correlations of the factors with the observable factors mentioned above and two further sets of factors. The first, dubbed "OPT3", follows Karakaya (2013) and consists of three factors constructed as follows: a *level* factor defined as the return on ATM delta-hedged contracts of all maturities; a *slope* factor defined as the difference in returns of short-dated options with at most 2 months to expiry and longer-dated options with between four to six months to expiry; and a *value* factor defined as the difference in returns of options with high value and options with low value, where value is defined as the difference between BMS implied volatility and historical, realised volatility. In addition, we construct a set of factors constructed based on the shape of the implied volatility surface: a short-dated skew factor ("opt skew - short") is constructed as the difference in returns of options with moneyness between -1 and -0.5, and options with moneyness between 0.5 and 1, where all options have one month to expiry. Similarly, we construct a long-dated skew factor ("opt skew - long") using the same moneyness ranges, but options with six to 12 months to expiry. The skew twist factor ("opt skew twist") is constructed as the difference between the short-dated and long-dated skew factors (opt skew - short / long). The correlations between the IPCA factors and the sets of observable factors constructed as above are summarised in Table 3.10.

We now interpret the factors in turn. The first factors' correlation with moneyness / maturity sorted portfolios reveals that the factor strongly relates to the returns on deep OTM options (correlation peaks at 76% for the 3- to 6-months portfolio). Deep OTM contracts are often perceived as lottery-like due to their extreme returns in rare events and periods of crisis. Indeed, a time series plot of cumulative return on the first factor (see Figure 3.3) reveals a striking pattern: factor returns are predominantly related to crisis periods such as the dot-com bubble in the early 2000s, during the downturn of 2002, and during the great financial crisis of 2007/08. Furthermore, the factor exhibits a correlation in excess of 50% with changes in the VIX index (see Table 3.11). In terms of the contract characteristics instrumenting the conditional loadings on this factor, moneyness and BMS Gamma are most notable. In conjunction, this evidence suggests an interpretation of the first factor in line with a (volatility) jump factor, similar to those documented in Eraker et al. (2003) and Broadie et al. (2007). However, the factor's Sharpe ratio is small with only 0.21 (see Table 3.9).

The second factor is challenging to interpret. A multitude of characteristics enter its loadings (see Figure 3.4). The factor's correlation pattern with moneyness / maturity sorted portfolios is to some extent similar to that of factor one: high positive correlations with OTM contracts and smaller (or even negative) correlations with ITM contracts. However, given that the factors recovered by IPCA are orthogonal, the second factor likely captures an effect that although related to factor 1 in some regards (e.g., see the large positive factor return during 2008 in Figure 3.3) is ultimately distinct (negative return during the early 2000s). The factor's Sharpe ratio is comparable to that of the first factor with 0.30.

The third factor's correlations with portfolios sorted on moneyness / maturity (see Figure 3.5), reveal a distinct effect related to contract maturity: while the shortest contracts (1-month) exhibit primarily positive correlation with this factor, longer maturity contracts are almost entirely negatively correlated with this factor. This suggests that the factor captures a form of term-structure or maturity related effect. In addition, we find a 54% correlation with the OPT3 model slope factor, confirming the notion of a maturity factor. The factors Sharpe ratio is sizeable with 0.96 on an annualised basis.

The fourth factor correlates most strongly with the returns on high moneyness contracts. A correlation of 68% with the OPT3 model's value factor stands out. Karakaya (2013) motivates his value factor by reference to differences in institutional demand pressures. In the context of index options, investor attention is typically focused on put options for forms of portfolio insurance / volatility selling, leading to a relatively smaller demand for call options, an effect potentially captured in this factor.

Finally, the fifth factor exhibits another distinct pattern in its correlation with moneyness / maturity sorted portfolios: the factor is most strongly correlated with ATM contracts and relatively less so with OTM contracts. The effect is more pronounced for shorter maturities. In addition, the factor loadings (see Figure 3.4) are dominated by a positive coefficient on the BMS Theta. In terms of correlations with observable factors a -80% correlation with the returns on the Coval and Shumway (2001) straddle factor stands out. In conjunction, this suggests that the factor is strongly related to a variance risk premium, i.e. the return to shorting straddles. As expected, the time series plot of the cumulative factor return (see Figure 3.3) shows that the factor loses during typical tail risk events such as the dot-com bubble and the 2007/08 crisis. The loading on BMS Theta is consistent with this idea: when buying straddles, traders have high negative Theta exposure, i.e. the position quickly loses value as time passes. In addition, the fifth factor has the largest annualised Sharpe ratio in our set of IPCA factors with 1.88, strongly suggesting this factor to be priced.

3.4.6 Which Characteristics Matter?

The IPCA model allows us to test for the incremental contribution of individual characteristics. To this end, Kelly et al. (2019) develop a bootstrap approach that tests the significance of an individual characteristic by setting it to zero, while keeping all other characteristic loadings fixed at their estimated values. We briefly outline the procedure now. Characteristics enter the IPCA model through the loadings matrix Γ_β that contains the loadings for specific characteristics in each of its rows. The bootstrap test essentially tests the magnitude of the difference of a given row in Γ_β from zero. We start by estimating the unrestricted model, (i.e. the model that does not force the coefficients of a given characteristic in Γ_β to zero) and save the estimated model parameters for $\hat{\Gamma}_\beta$, $\left\{\hat{f}_t\right\}_{t=1}^{T}$, as well as the residuals from the fitted model $\{d_t\}_{t=1}^{T}$. Using the residuals we generate a bootstrap sample under the null hypothesis that the l-th characteristic does not affect loadings.[19] We can then compare a Wald-type test statistic of the form $W_{\beta,l} = \hat{\gamma}'_{\beta,l}\hat{\gamma}_{\beta,l}$ from the alternative model with unrestricted Γ_β, to the analogous test statistic $\tilde{W}^b_{\beta,l}$ for the model estimated on the bootstrapped data, where the super-script b indexes the bootstrap draw. Finally, p-values can be computed as the fraction of test statistics $\tilde{W}^b_{\beta,l}$ that exceed $W_{\beta,l}$.

Table 3.12 summarises the results from the significance tests. We measure the reduction in total R^2 when setting the row pertaining to a specific characteristic to zero and report the associated p-values from the bootstrap exercise.

We find that moneyness, implied volatility and BMS Gamma appear at the top of the variable importance ranking. Comparing with our earlier discussions of the loadings matrices Γ_β in Figure 3.4, the bootstrap exercise confirms that the majority of characteristics that factors load on strongly, are also the ones that are most relevant in terms of generating the models ability to capture the panel variation in option returns.

[19]For details, we refer the reader to Kelly et al. (2019), Section 3.3.

3.4.7 Alternative Sets of Characteristics

While the IPCA model potentially allows for a large number of characteristics to be entertained as instruments for conditional loadings simultaneously, the researcher ultimately has to take a stance on the set of characteristics taken into account. In our analysis so far the set of characteristics included moneyness, time-to-maturity, implied volatility, embedded leverage, Gamma and Vega, and all characteristics were interacted with both a constant and the put / call dummy variable. In order to put our choice of characteristics into perspective, we run two sets of analysis.

Firstly, we compare the performance of the baseline model against a model in which the choice of characteristics is motivated economically from an option return decomposition. A standard option return decomposition is obtained via a simple Taylor expansion of the option price F_t along its most relevant dimensions as follows

$$\Delta F_t = \frac{\mathrm{d}F_t}{\mathrm{d}S_t}\Delta S_t + \frac{\mathrm{d}F_t}{\mathrm{d}IV_t}\Delta IV_t + \frac{1}{2}\frac{\mathrm{d}^2 F_t}{\mathrm{d}{S_t}^2}(\Delta S_t)^2 + \dots . \tag{3.11}$$

This can be re-arranged to yield the relative change in the option price

$$\frac{\Delta F_t}{F_t} = \underbrace{\frac{\mathrm{d}F_t}{\mathrm{d}S_t}\frac{S_t}{F_t}}_{\Delta^{\text{Decomp}}}\frac{\Delta S_t}{S_t} + \underbrace{\frac{\mathrm{d}F_t}{\mathrm{d}IV_t}\frac{IV_t}{F_t}}_{\mathcal{V}^{\text{Decomp}}}\frac{\Delta IV_t}{IV_t} + \underbrace{\frac{1}{2}\frac{\mathrm{d}^2F_t}{\mathrm{d}{S_t}^2}\frac{S_t^2}{F_t}}_{\Gamma^{\text{Decomp}}}\frac{\Delta S_t^2}{S_t^2} + \dots . \tag{3.12}$$

The terms Δ^{Decomp}, $\mathcal{V}^{\text{Decomp}}$, and Γ^{Decomp} can be understood as coefficients on a set of factors that attribute the option return variation to a Delta, Gamma and Vega component. Table 3.13 details the performance of an IPCA model that uses the three model-implied coefficients as characteristics. We find that the model with characteristics following the return decomposition in Eq.(3.12), achieves only slightly lower total R^2 than the baseline model (e.g. for $K = 5$ and $\Gamma_\alpha = 0$, 85.3% vs. 90.6%), but at higher numbers of factors, the predictive R^2 of the decomposition based model is notably lower (e.g. for $K = 5$ and $\Gamma_\alpha = 0$, 4.7% vs. 6.4%). This may provide evidence that although the decomposition based characteristics provide roughly similar ability to capture empirically observed variation in option returns (measured by total R^2), the wider set of characteristics in our baseline specification is yielding a better risk description (measured by predictive R^2). Overall, this evidence shows that the choice of characteristics in our baseline specification stacks up well against a more economically driven choice of characteristics.

Secondly, we estimate a battery of models with varying sets of characteristics to assess further the dependence of our baseline model's performance on the choice of characteristics. Although the setting of index options naturally constrains the potential set of characteristics in contrast to individual equity options, there are a number of additional characteristics that can plausibly be used. Table 3.14 contains an overview of IPCA specifications using different sets of characteristics as conditioning information. Starting from a first specification that only includes moneyness, maturity and embedded leverage as characteristics, we find that adding characteristics such as Theta and implied volatility one by one gradually increase the total R^2. This finding is somewhat expected as allowing for additional characteristics increases the model's number of free parameters. Consequentially, the models ability to capture observed return variation improves. However, increases in total R^2 do not necessarily entail increases in predictive R^2 which measures the model's ability to describe differences in average returns. In fact, we find that a model which excludes BMS Gamma and Vega from our baseline specification produces even higher predictive R^2 at both the individual contract and portfolio level. This finding is likely explained by the correlation structure of the characteristics that can potentially affect a model's generalisation ability in a detrimental way.

3.4.8 Model Performance at Daily Frequency

The analysis so far has demonstrated that IPCA offers a favourable description of realised returns and risk at monthly frequency when compared to observable factor models that include factors such as the Fama-French three factors, momentum, and option factors such as straddle returns as well as the embedded leverage / betting against beta factor. In this section, we estimate the IPCA model for daily frequency, delta-hedged returns. This horizon is of particular interest to market makers and traders that often only hold contracts for short periods of time and that place an emphasis on risk managing a book of derivatives.

As a benchmark, we compare IPCA against the profit and loss attribution framework proposed by Carr and Wu (2020) that is developed particularly with shorter horizon returns in mind. Their framework attributes option returns to variation in the first and second moments of the underlying process and the implied volatility process. This framework is a formalisation of the typical trader's "rule-of-thumb" model for managing option risk. The framework is obtained by performing a Taylor series expansion of the Black-Merton-Scholes (BMS) pricing formula, taking expectations under the risk-neutral measure and enforcing a no-arbitrage condition. A particularly neat version of their P&L attribution is obtained by scaling option gains by their cash Gamma

$$\begin{aligned}\frac{\mathrm{d}F}{F_{SS}S_t} =&\frac{F_S}{F_{SS}}\frac{\mathrm{d}S_t}{S_t} + IV_t^2 S_t\tau\left(\frac{\mathrm{d}IV_t}{IV_t} - \mu_t\mathrm{d}t\right) + \frac{1}{2}S_t\left(\left(\frac{\mathrm{d}S_t}{S_t}\right)^2 - \sigma_t^2\mathrm{d}t\right)\\ &+ \frac{1}{2}S_t z_+ z_-\left(\left(\frac{\mathrm{d}IV_t}{IV_t}\right)^2 - \omega_t^2\mathrm{d}t\right) + z_+ S_t\left(\left(\frac{\mathrm{d}S_t}{S_t}\frac{\mathrm{d}IV_t}{IV_t}\right) - \gamma_t\mathrm{d}t\right),\end{aligned} \tag{3.13}$$

where F_S denotes the BMS option Delta, F_{SS} denotes the BMS option Gamma, $\tau = (T_i - t)/365$ is the standardised time-to-maturity, and $z_\pm = (log(K_i/S_t) \pm \frac{1}{2}IV_t^2\tau)$. Further, the moments under the risk-neutral measure are

$$\mu_t = \mathbb{E}_t\left[\frac{\mathrm{d}IV_t}{IV_t}\right]/\mathrm{d}t, \qquad \sigma_t^2 = \mathbb{E}_t\left[\left(\frac{\mathrm{d}S_t}{S_t}\right)^2\right]/\mathrm{d}t,$$

$$\omega_t^2 = \mathbb{E}_t\left[\left(\frac{\mathrm{d}IV_t}{IV_t}\right)^2\right]/\mathrm{d}t, \qquad \gamma_t^2 = \mathbb{E}_t\left[\left(\frac{\mathrm{d}S_t}{S_t}, \frac{\mathrm{d}IV_t}{IV_t}\right)\right]/\mathrm{d}t,$$

i.e. μ_t is the expected rate of change in the BMS implied volatility, σ_t^2 and ω_t^2 are the conditional variance rate of underlying and implied volatility, respectively, and γ_t is the conditional covariance between the two processes. Note that in contrast to the BMS framework itself, Carr and Wu (2020) do not require us to fully specify the dynamics of the underlying.

The risk-neutral moments above can be obtained from the implied volatility surface. However, since we are not limiting the analysis to ATM options on a grid of fixed maturities as in Carr and Wu (2020), a few generalisations are necessary. We estimate μ_t from the implied volatility term-structure, but instead of using fixed maturity points to obtain the term structure slope, we approximate the slope by taking a symmetric 60-day window around the location of a given contract on the implied volatility surface. For contracts close to the edges of the surface, we shift the window from which we compute the slope to ensure a window size of 60 days. The variance rate ω_t is estimated from a 63-day rolling window variance of daily implied volatility changes where we ensure the availability of at least 21 observations. The covariance rate γ_t is estimated from the rolling 21-day covariance between spot changes and implied volatility changes and we ensure availability of at least 10 observations. Finally, we estimate σ_t^2 from the ATM implied volatility at a maturity of one month.

Asset Pricing Performance We measure the performance of the Carr and Wu (2020) approach individually for each of the risk attributions. To this end we define the set of series

$$
\begin{aligned}
R_0 &= \frac{\mathrm{d}F}{S_t} - F_S \frac{\mathrm{d}S_t}{S_t} \\
R_1 &= R_0 - IV_t^2 S_t F_{SS} \tau \left(\frac{\mathrm{d}IV_t}{IV_t} - \mu_t \mathrm{d}t \right) \\
R_2 &= R_1 - \frac{1}{2} S_t F_{SS} \left(\left(\frac{\mathrm{d}S_t}{S_t} \right)^2 - \sigma_t^2 \mathrm{d}t \right) \\
R_3 &= R_2 - \frac{1}{2} S_t F_{SS} z_+ z_- \left(\left(\frac{\mathrm{d}IV_t}{IV_t} \right)^2 - \omega_t^2 \mathrm{d}t \right) \\
R_4 &= R_3 - z_+ S_t F_{SS} \left(\left(\frac{\mathrm{d}S_t}{S_t} \frac{\mathrm{d}IV_t}{IV_t} \right) - \gamma_t \mathrm{d}t \right).
\end{aligned}
$$

We then compute the fraction of variation in delta-hedged option returns r_{Spot}^{Δ} explained by sequentially accounting for Vega risk, Gamma risk and so on as

$$
R_i^2 = 1 - Var[R_i]/Var[r_{Spot}^{\Delta}], \qquad i = 1, \ldots, 4. \tag{3.14}
$$

Since there is no direct analogue of R_{pred}^2 in this framework, we focus our analysis on the total R^2.

Results We summarise the results from a comparison of the Carr & Wu model and the restricted IPCA frame work in Table 3.16. Before diving into the comparison of the two approaches, we note that IPCA also performs well at daily frequency and delivers high total R^2 across a large range of option contracts. In analogy to our earlier exercise, we present results for $K = 1, \ldots, 5$ factors to highlight how the IPCA performance increases as we allow for more factors to mirror progressively hedging additional sources of risk in the Carr & Wu model. For completeness, Table 3.15 reports a more detailed set of results for IPCA model.

Comparing with the results from the no-arbitrage model, we find that when allowing for a sufficient number of factors, IPCA almost always outperforms (Table

3.16, Panel A), and also when sorting the sample on moneyness, time-to-maturity, and VIX (Panel B-D). Here we note that while the improvement in performance of the IPCA model with increasing number of factors is almost mechanical, Table 3.15 demonstrates that the models ability to describe risk (as measured by the predictive R^2) does not decrease when allowing for more factors. While the Carr & Wu model performs reasonably well for options close to the money, performance deteriorates when moving to OTM options. Furthermore, we find that the no-arbitrage model exhibits relatively poor performance for short-dated options with one month to maturity (Panel C). Finally, Panel D demonstrates that both models are able to capture option return variation, albeit to varying degrees of accuracy, over a wide range of market conditions as proxied by the VIX index.

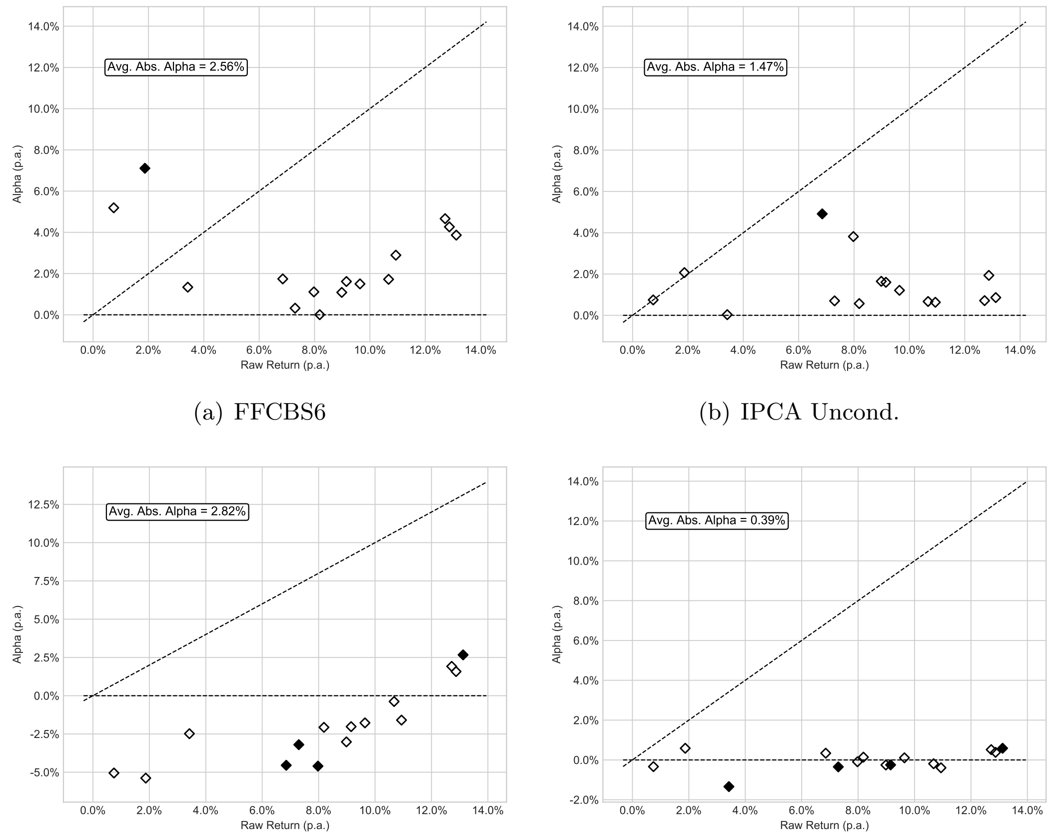
3.5 Conclusion

In this paper we study a latent factor approach that uses characteristics as conditioning information for time-varying betas to describe empirically observed option returns. Our findings from this study are threefold. First, we find that a low dimensional latent factor model is successful in capturing variation in option returns and describing differences in risk across a wide range of options. Especially, we confirm that our approach, instrumented principal components analysis (IPCA), performs notably better than observable factor models in terms of its ability to capture the variation in realised option returns, and at least as well in terms of its ability to describe risk. Importantly and in contrast to earlier findings in the literature, we demonstrate that when using the IPCA factors, there are *on average* no significant intercepts associated with the characteristics managed portfolios. Furthermore, we test the robustness of our findings in an out-of-sample exercise and find that IPCA's performance remains strong, still outperforming the in-sample results from observable factor models.

Second, the factors recovered by IPCA are interpretable and relate to jump risk, maturity risk, volatility risk and the shape of the implied volatility surface. One of the recovered factors, correlates highly with the returns on ATM straddles that have often been understood as a proxy for market implied volatility risk. The Sharpe ratios on the IPCA factors reach as high as 1.88. Furthermore, we examine the characteristics that are important in recovering the latent IPCA factors. Naturally, due to the study of index options the set of potential option characteristics is limited to contract specific characteristics such as moneyness, time-to-maturity, option Greeks, etc. and cannot make use of underlying characteristics as is the case when studying single-name options. Nevertheless, we find that only about a half of the total 14 characteristics used are statistically relevant to describe returns. Among the most important characteristics we find option moneyness, implied volatility and option Gamma.

Third, we test the IPCA model with shorter horizon returns at daily frequency and compare to a no-arbitrage model. The findings show that IPCA not only captures return variation well at monthly frequency, but also robustly describes shorter horizon returns across a wide range of option contracts, outperforming a benchmark no-arbitrage model which implements a trader's "rule-of-thumb" approach to P&L attribution.

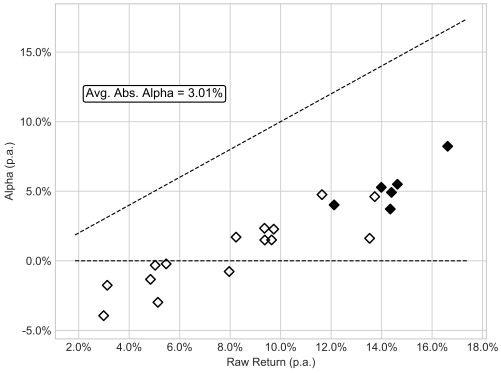
Figures & Tables



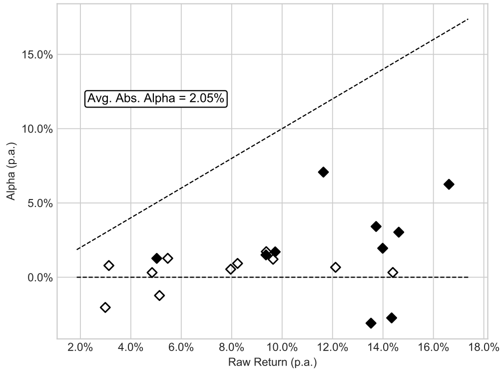
(a) FFCBS6 (b) IPCA Uncond.

(c) FFCBS6 Cond. (d) IPCA Cond.

Figure 3.1: Test of Mean-Variance Efficiency using Managed Portfolios.
This figure plots portfolio alphas from a time series regression of portfolio returns on the set of factors from either an observable factor model (FFCBS6) and the IPCA model with $K = 5$ factors. The test portfolios are the managed portfolios constructed from the set of characteristics included in the estimated IPCA model: moneyness, maturity, implied volatility, embedded leverage, gamma and vega interacted with both a constant and the put / call dummy variable. This yields a total of 15 portfolios including the equally weighted portfolio. Portfolio alphas are plotted against the raw portfolio returns. All portfolios are re-leverage to yield 10% annualised volatility. Filled diamond markers correspond to alphas with t-statistic greater than 2.0. Panels (a) and (b) show results from unconditional asset pricing tests, while panel (c) and (d) show results from the conditional FFCBS6 / IPCA model, i.e. alphas are computed as time series averages of period-by-period portfolio residuals.



(a) FFCBS6



(b) IPCA

Figure 3.2: Test of Unconditional Mean-Variance Efficiency using Double Sorted Portfolios.
This figure plots portfolio alphas from a time series regression of portfolio returns on the set of factors from either an observable factor model (FFCBS6) and the IPCA model with $K = 5$ factors. The test portfolios are constructed by double sorting all option contracts on their time-to-maturity and option delta. Specifically, we construct portfolios of options with 1 month, 2 months, 3 to 6 months and 7 to 12 months to maturity and absolute option deltas between zero and 0.1, 0.1 and 0.2, and so on up to an absolute delta of 0.5. This generates 20 double sorted portfolios. Portfolio alphas are plotted against the raw portfolio returns. All portfolios are re-leverage to yield 10% annualised volatility. Filled diamond markers correspond to alphas with t-statistic greater than 2.0. Panels (a) and (b) show results from unconditional asset pricing tests, while panel (c) shows results from the conditional IPCA model, i.e. alphas are computed as time series averages of period-by-period portfolio residuals.

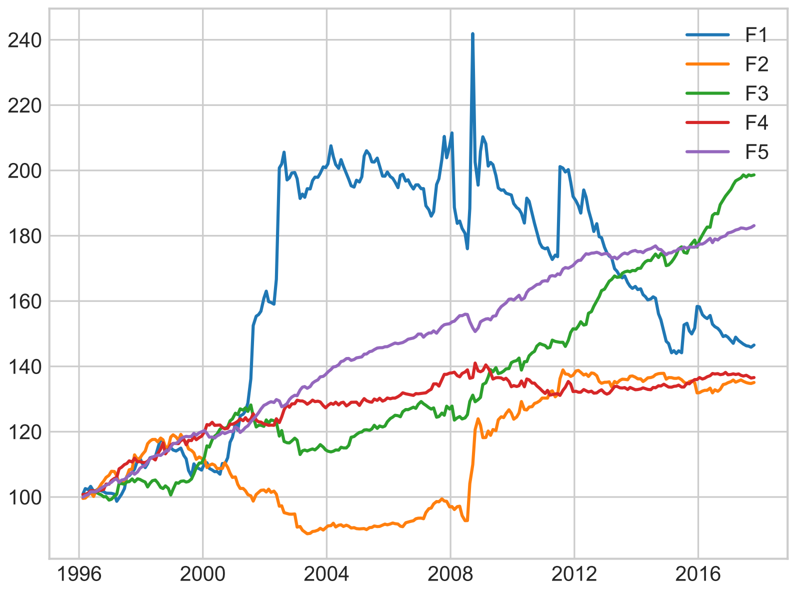
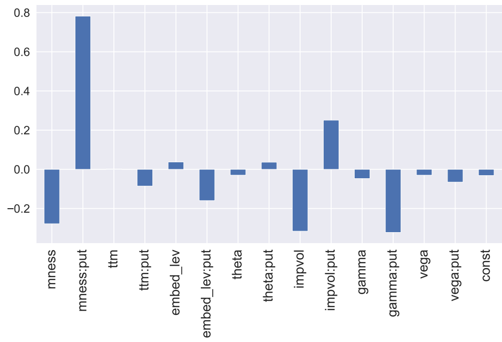
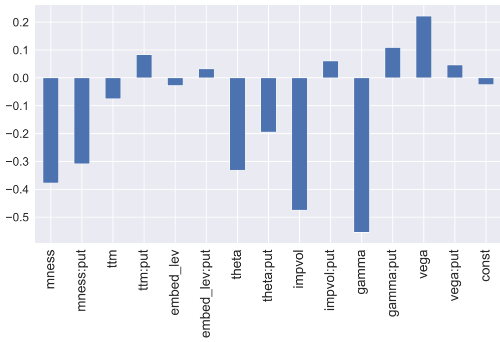


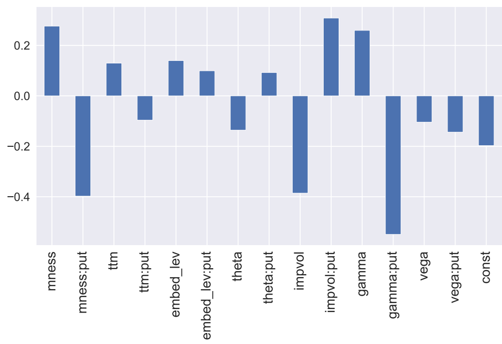
Figure 3.3: Cumulative Factor Returns
This figure shows the cumulative returns on the IPCA factors for the restricted IPCA model with $K = 5$ factors. The factors are ordered by their variance, from highest to lowest, i.e. the first factor "F1" has the highest time series variance and "F5" has the lowest time series variance.



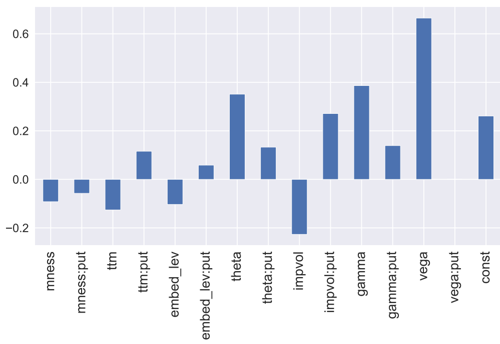
(a) Factor 1



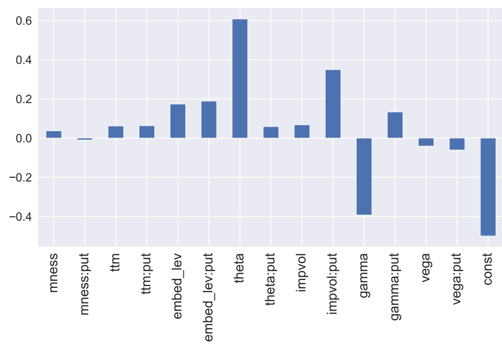
(b) Factor 2



(c) Factor 3

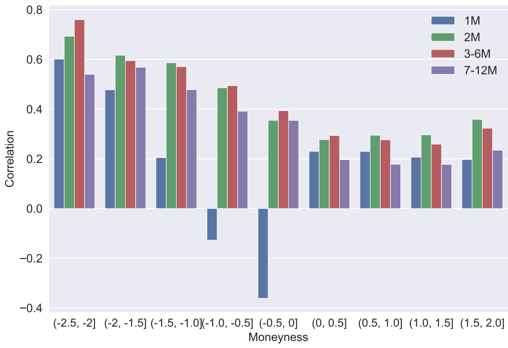


(d) Factor 4

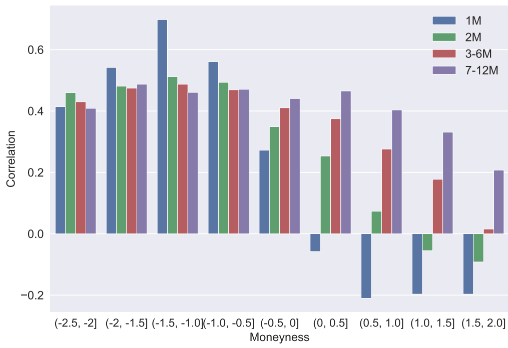


(e) Factor 5

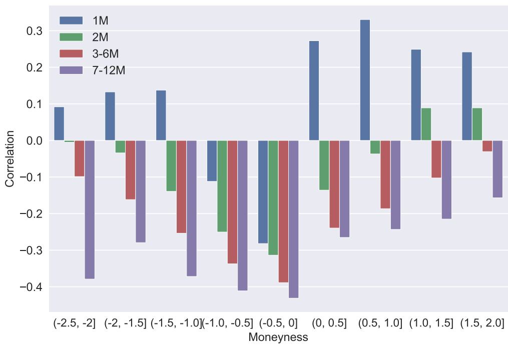
Figure 3.4: Plots of loadings Γ_β
This figure shows the loadings, Γ_β, that create the mapping between characteristics and factors. Each sub-plot corresponds to one column of Γ_β, i.e. one IPCA factor. The IPCA model fitted is the restricted model with $K = 5$. The colon notation is used to indicate interaction with another variable, e.g. ttm:put refers to the interaction between an option's time-to-maturity with an indicator variable that is equal one for puts and zero for calls.



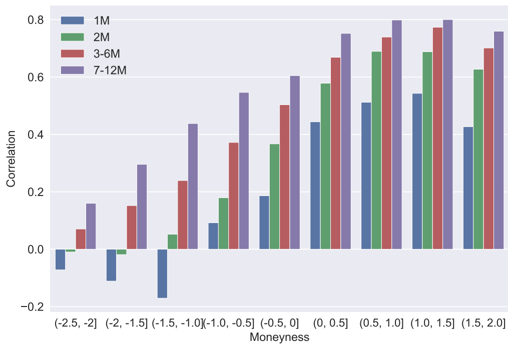
(a) Factor 1



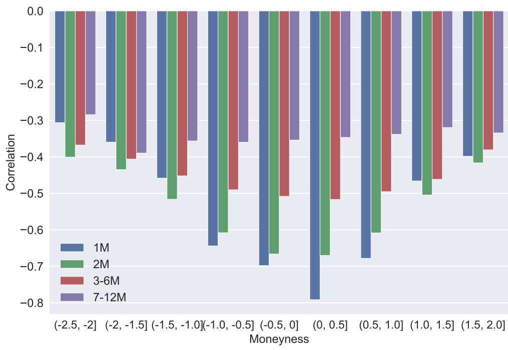
(b) Factor 2



(c) Factor 3



(d) Factor 4



(e) Factor 5

Figure 3.5: Factor Correlations with Moneyness / Maturity Portfolios
This figure displays correlations of the latent factors extracted by the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 5$ factors with returns of double sorted portfolios on moneyness and maturity. The portfolios are constructed from equally weighting the returns on all options that fall in a given bin on the portfolio formation date. The breakpoints for portfolio construction are as follows: moneyness bins start at -2.5 going up to 2.0 in increments of 0.5; time-to-maturity (TTM) bins are 1 month, 2 months, 3 to 6 months and 7 - 12 months to expiry.

Table 3.1: Summary Statistics of Option Level Variables.
This table reports summary statistics for the sample of monthly S&P 500 index option contracts. **Panel A** contains results for call option contracts, while **Panel B** contains results for put option contracts. The sample period is 1996 to 2017. For option level variables the table shows the time-to-maturity ttm, moneyness mness computed as mness = $ln(K/S)/(IV \cdot \sqrt{ttm})$ with strike K and spot price S, Black-Scholes implied volatility IV, option Delta, option Gamma, option Vega, option Theta, and the annualised delta-hedged return versus spot price r^{Δ}_{Spot}.

Panel A: Call Option Contracts

	`ttm`	`mness`	`embed_lev`	`iv`	`delta`	`gamma`	`vega`	`theta`	r^{Δ}_{Spot}
Mean	132	0.86	28.93	0.16	0.24	0.003	215.81	-63.95	-3.60%
Median	91	0.79	23.01	0.15	0.23	0.002	181.31	-52.45	-3.35%
Std. Dev.	100	0.57	19.82	0.07	0.16	0.002	146.06	47.10	25.24%
No. Obs.	22,067	22,067	22,067	22,067	22,067	22,067	22,067	22,067	22,067

Panel B: Put Option Contracts

	`ttm`	`mness`	`embed_lev`	`iv`	`delta`	`gamma`	`vega`	`theta`	r^{Δ}_{Spot}
Mean	116	-1.32	19.30	0.28	-0.13	0.001	141.50	-64.83	-4.72%
Median	63	-1.36	17.28	0.26	-0.08	0.001	96.81	-51.75	-3.35%
Std. Dev.	95	0.71	10.75	0.11	0.13	0.002	137.34	50.08	23.33%
No. Obs.	54,881	54,881	54,881	54,881	54,881	54,81	54,881	54,881	54,881

Table 3.2: IPCA Performance.
This table reports the IPCA fit performance as measured by R^2_{total} and R^2_{pred} using both the restricted model ($\Gamma_\alpha = 0$) and the unrestricted model ($\Gamma_\alpha \neq 0$) with $K = 1, \ldots, 5$. The employed option characteristics `mness`, `ttm`, `embed_lev`, `theta`, `impvvol`, `gamma`, `vega` are each interacted with a constant and an indicator variable that is equal one if the option is a put and is zero otherwise. Additionally, a constant is included in the set of characteristics. In **Panel A** performance measures are computed with respect to individual option contracts, while in Panel B performance measures are computed with respect to the characteristics-managed portfolios. Panel C reports p-values for the test with null hypothesis $H_0 : \Gamma_\alpha = 0$ from a bootstrap with 1000 draws for each time t.

		No. Factors				
		1	2	3	4	5
		Panel A: Individual Options				
R^2_{total}	$\Gamma_\alpha = 0$	72.69%	80.18%	85.31%	89.15%	90.61%
	$\Gamma_\alpha \neq 0$	74.24%	81.61%	86.09%	89.53%	90.83%
R^2_{pred}	$\Gamma_\alpha = 0$	5.48%	5.29%	6.11%	6.31%	6.40%
	$\Gamma_\alpha \neq 0$	7.22%	7.34%	7.25%	7.08%	6.98%
		Panel B: Managed Portfolios				
R^2_{total}	$\Gamma_\alpha = 0$	95.47%	97.54%	99.08%	99.63%	99.73%
	$\Gamma_\alpha \neq 0$	96.00%	98.03%	99.25%	99.65%	99.73%
R^2_{pred}	$\Gamma_\alpha = 0$	6.23%	6.06%	6.38%	6.49%	6.51%
	$\Gamma_\alpha \neq 0$	6.71%	7.18%	7.04%	6.79%	6.71%
		Panel C: Bootstrap Test ($\boldsymbol{H_0 : \Gamma_\alpha = 0}$)				
W_α p-value		99.9%	99.4%	94.6%	76.9%	62.4%

Table 3.3: Panel Regression of Option Returns on Option Characteristics.
The dependent variable is the monthly delta-hedged return versus the prevailing spot price r^{Δ}_{Spot}. Regression specification (1) uses a common intercept for all observations, while specification (2) uses time-fixed effects. As regressors we include the moneyness computed as mness $= ln(K/S)/(IV \cdot \sqrt{ttm})$, time-to-maturity ttm, Black-Merton-Scholes (BMS) gamma, vega, theta and the BMS implied volatility (IV). In addition to the characteristics themselves, we include the interaction of the characteristics with a put / call dummy that has value one if the option contract is a put and is zero otherwise. For example, the interaction of moneyness and the put / call indicator is denoted mness:put. All predictors are standarised by their sample standard deviation. Standard errors are clustered by option contract and in parentheses we report t-statistics.

	r^{Δ}_{Spot}	
	(1)	(2)
Intercept	-0.0006	
	(-0.3642)	
mness	0.1212	0.0654
	(1.3761)	(0.8360)
ttm	0.0397	0.0505
	(2.1013)	(1.9910)
embed_lev	-0.0219	-0.0058
	(-0.8734)	(-0.4110)
gamma	0.0105	0.0217
	(0.4761)	(1.5546)
vega	-0.0015	0.0016
	(-0.0507)	(0.0711)
theta	-0.0538	-0.0240
	(-1.7241)	(-0.8829)
IV	-0.1311	-0.1545
	(-1.5084)	(-2.7647)
mness:put	-0.1765	-0.1282
	(-1.1220)	(-1.2194)
ttm:put	-0.0237	-0.0400
	(-1.5188)	(-1.7509)
embed_lev	0.0381	0.0256
	(1.9839)	(1.1326)
gamma:put	-0.0321	-0.0384
	(-1.8850)	(-2.6538)
vega:put	-0.0148	-0.0091
	(-0.9335)	(-0.6323)
theta:put	0.0872	0.0776
	(2.3337)	(3.7570)
IV:put	0.1071	0.1248
	(1.7136)	(2.8630)
Effects		Time
R^2	3.41%	6.77%
No. Obs.	76948	76948

Table 3.4: IPCA Performance by Bins of Moneyness, Maturity, and VIX.
The table details the total and predictive R^2 by bins sorted on option characteristics for the restricted IPCA model with $K = 5$ factors. Note, total and predictive R^2 is computed using only observations from the respective bin examined.

Panel A: Moneyness Bin

Range	R^2_{total}	R^2_{pred}	No. Obs.
-2.5 to -2	65.81%	1.46%	12,178
-2 to -1.5	86.50%	3.21%	12,017
-1.5 to -1.0	86.28%	6.41%	11,155
-1.0 to -0.5	91.87%	8.25%	10,253
-0.5 to 0.0	89.92%	9.02%	9,278
0.0 to 0.5	89.70%	2.43%	7,460
0.5 to 1.0	90.78%	3.16%	5,895
1.0 to 1.5	79.70%	4.20%	4,848
1.5 to 2.0	57.50%	2.85%	3,864

Panel B: Time-to-Maturity Bin

Range	R^2_{total}	R^2_{pred}	No. Obs.
1 Month	77.76%	6.19%	17,778
2 Months	92.14%	8.25%	19,280
3 to 6 Months	94.65%	7.75%	22,446
6 to 12 Months	93.40%	3.19%	17,444

Panel C: VIX

Range	R^2_{total}	R^2_{pred}	No. Obs.
10% to 20%	87.66%	3.18%	48,318
20% to 30%	86.50%	18.64%	21,012
30% to 50%	94.61%	3.82%	5,578
50% to 90%	83.71%	-5.61%	665

Table 3.5: IPCA versus Observable Factor Models.
The table compares the total and predictive R^2 of the restricted IPCA model with $K = 1, \ldots, 5$ factors (**Panel A**) and a number of observable factor models (**Panel B / C**). Starting from the Fama-French three factor model (FF3), we add the Carhart momentum factor (FFC4), the Frazzini and Pedersen (2012) 'betting against beta' factor (FFCB5), and the Coval and Shumway (2001) straddle factor (FFCBS6). Panel B contains the results when loadings are dynamic by instrumenting with characteristics, while in Panel C loadings are static corresponding to a panel regression. We report performance measures for individual options as test assets R^2_{total} / R^2_{pred} and for characteristics managed portfolios as test assets $R^2_{total,x}$ / $R^2_{pred,x}$.

Panel A: IPCA

	K				
	1	2	3	4	5
R^2_{total}	72.69%	80.18%	85.31%	89.15%	90.61%
R^2_{pred}	5.48%	5.29%	6.11%	6.31%	6.40%
$R^2_{total,x}$	95.47%	97.54%	99.08%	99.63%	99.73%
$R^2_{pred,x}$	6.23%	6.06%	6.38%	6.49%	6.51%

Panel B: Observable Factors - With Instruments

	CAPM	FF3	FFC4	FFCB5	FFCBS6
R^2_{total}	25.64%	29.31%	31.19%	36.49%	50.59%
R^2_{pred}	2.42%	2.58%	2.75%	4.10%	6.55%
$R^2_{total,x}$	29.98%	37.46%	40.39%	43.67%	58.85%
$R^2_{pred,x}$	3.12%	3.30%	3.59%	4.90%	6.70%

Panel C: Observable Factors - No Instruments

	CAPM	FF3	FFC4	FFCB5	FFCBS6
R^2_{total}	18.86%	19.41%	19.44%	24.40%	31.71%
R^2_{pred}	2.51%	2.46%	2.53%	3.35%	4.06%
$R^2_{total,x}$	25.75%	27.91%	28.23%	32.57%	45.56%
$R^2_{pred,x}$	2.62%	2.65%	2.79%	4.38%	5.66%

Table 3.6: IPCA Portfolio Alphas.
The table summarises the conditional and unconditional portfolio average absolute alphas when the factors come from the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 1, \ldots, 5$ factors. The test portfolios are the characteristics managed portfolios. Unconditional portfolio alphas are obtained from time series regressions of portfolio returns onto the set of factors. Conditional alphas are obtained as the time series averages of period-by-period portfolio residuals in the main IPCA model. The reported values are the average absolute alphas across the set of test portfolios.

	K=1	K=2	K=3	K=4	K=5
Unconditional	3.39%	2.11%	1.83%	1.55%	1.47%
Conditional	3.24%	2.87%	0.54%	0.38%	0.39%

Table 3.7: Out-of-Sample Performance
The table shows the out-of-sample performance of the restricted IPCA model with $K = 1, \ldots, 5$ factors. The out-of-sample exercise starts at 50% of our sample length, i.e. the first forecast is made for January 2007. **Panel A** contains results for individual option contracts, while **Panel B** examines the out-of-sample performance of managed portfolios.

	K=1	K=2	K=3	K=4	K=5
			Panel A: Individual Options		
R^2_{total}	71.99%	75.75%	82.56%	88.10%	89.28%
R^2_{pred}	4.77%	4.39%	4.15%	4.38%	4.37%
			Panel B: Managed Portfolios		
R^2_{total}	95.76%	96.79%	98.47%	99.50%	99.52%
R^2_{pred}	3.31%	2.82%	2.19%	2.76%	2.73%

Table 3.8: Out-of-Sample Factor Portfolio Sharpe Ratios.
The table summarises the out-of-sample Sharpe ratios of the individual factors ("Univariate") and mean-variance optimal portfolios ("Tangency"). We assume a portfolio volatility target of 1% per month and rescale the portfolio weights accordingly. The observable factor models start with the CAPM and progressively add further factors: size ("FF2"), value ("FF3"), momentum ("FFC4"), betting-against-beta for options ("FFCB5"), and straddles ("FFCBS6"). **Panel A** details the results for the IPCA model, while **Panel B** reports results for the observable factor models. The out-of-sample period starts at 50% of the total available sample length, i.e. in January 2007.

			Panel A: IPCA		
	K=1	K=2	K=3	K=4	K=5
Univariate	0.60	0.46	0.44	0.88	1.18
Tangency	0.60	1.04	1.30	0.96	0.73

			Panel B: Observable Factors			
	CAPM	FF2	FF3	FFC4	FFCB5	FFCBS6
Univariate	0.46	0.13	-0.09	0.12	0.36	0.55
Tangency	0.46	0.41	0.10	0.27	0.32	0.51

Table 3.9: IPCA Factor Summary Statistics. This table reports summary statistics of the restricted IPCA model, estimated using five factors and as characteristics `mness`, `ttm`, `embed_lev`, `impvol`, `gamma`, `theta`, `vega`. Reported measures are annualised. Note, interpreting factor returns comes with the caveat that we compute delta-hedged option returns using the prevailing spot price as demnominator, i.e. we compare the excess return on the delta-hedged option position to the value of the underlying. Significance of the Sharpe ratio is obtained from a regression of the factor return series on a constant where the regression is using Newey-West standard errors with up to 10 lags. The significance levels are: $p < 0.10$ - *, $p < 0.05$ - **, $p < 0.01$ - ***. The sample period is from 1996 to 2017 covering 261 months.

	F1	F2	F3	F4	F5
Mean	0.02	0.02	0.03	0.01	0.03
Std. Dev	0.11	0.05	0.03	0.02	0.02
Sharpe Ratio	0.21	0.30	0.96***	0.60***	1.88***
Skewness	3.63	3.34	-0.66	0.18	-0.75
Kurtosis	32.62	25.11	3.29	2.71	1.81

Table 3.10: IPCA Factors versus Asset Pricing Factors: Correlations
The table relates the IPCA factors extracted by the restricted IPCA model with $K = 5$ factors to a set of asset pricing factors studied in previous empirical option returns literature. The observable factors include the Fama-French three factors (mktrf, smb, hml), the Carhart momentum factor (umd), the Frazzini and Pedersen (2012) betting against beta factor (bab), and the Coval and Shumway (2001) straddle factor (straddle). We construct the level (opt3 level), slope (opt3 slope) and value factors (opt3 value) of Karakaya (2013). We enlarge the set of factors constructed in the spirit of Karakaya (2013) and construct a set of implied volatility skew and skew twist factors. The short-dated skew factor (opt skew - short) is constructed as the difference in returns of options with moneyness between -1 and -0.5, and options with moneyness between 0.5 and 1, where all options have one month to expiry. Similarly, we construct the long-dated skew factor (opt skew - long) using the same moneyness ranges, but options with six to 12 months to expiry. The skew twist factor (opt skew twist) is constructed as the difference between the short-dated and long-dated skew factors (opt skew - short / long). The table contains pairwise correlations of the two sets of factors and p-values from the associated F-statistic. Bold font indicates correlations significant at the 5% level or better.

		F1	F2	F3	F4	F5
mktrf	corr.	**-46.16%**	3.98%	**18.27%**	**-37.55%**	**17.46%**
	p-value	0.00	0.52	0.00	0.00	0.00
smb	corr.	-8.81%	-8.85%	4.30%	**-21.68%**	**22.05%**
	p-value	0.16	0.16	0.49	0.00	0.00
hml	corr.	1.82%	6.73%	6.03%	-10.55%	6.18%
	p-value	0.77	0.28	0.33	0.09	0.32
umd	corr.	4.27%	-3.10%	-2.08%	**27.63%**	-4.46%
	p-value	0.49	0.62	0.74	0.00	0.47
bab_call	corr.	**-24.96%**	**15.48%**	-2.37%	**-13.91%**	**37.34%**
	p-value	0.00	0.01	0.70	0.03	0.00
bab_put	corr.	**-35.97%**	-11.65%	**-13.62%**	**39.85%**	**23.03%**
	p-value	0.00	0.06	0.03	0.00	0.00
bab	corr.	**-39.89%**	0.61%	-11.04%	**20.33%**	**37.81%**
	p-value	0.00	0.92	0.08	0.00	0.00
straddle	corr.	-5.54%	-2.39%	-1.86%	**27.57%**	**-79.83%**
	p-value	0.37	0.70	0.77	0.00	0.00
opt3 level	corr.	**28.30%**	**38.87%**	**-33.17%**	**65.39%**	**-49.05%**
	p-value	0.00	0.00	0.00	0.00	0.00
opt3 slope	corr.	**-32.69%**	**-34.00%**	**54.46%**	**-45.63%**	-13.67%
	p-value	0.00	0.00	0.00	0.00	0.07
opt3 value	corr.	-6.68%	**-16.21%**	**-21.93%**	**68.19%**	**-26.16%**
	p-value	0.28	0.01	0.00	0.00	0.00
opt skew - short	corr.	**21.96%**	**47.14%**	5.52%	**-22.12%**	0.31%
	p-value	0.00	0.00	0.38	0.00	0.96
opt skew - long	corr.	**26.81%**	**32.03%**	-2.72%	**-30.08%**	-5.71%
	p-value	0.00	0.00	0.67	0.00	0.36
opt skew twist	corr.	11.23%	**35.03%**	8.56%	-9.44%	1.95%
	p-value	0.07	0.00	0.17	0.13	0.76

Table 3.11: Correlations of IPCA Factors and Measures of Liquidity Provision.
This table contains correlations of IPCA factors for the model with $K = 5$ factors with measures of liquidity, funding and volatility risks. All correlations are computed using the monthly first differences in the liquidity measures. The variables are as follows: TED_spread is the TED spread obtained from the FRED database, intermed_capital is the intermediary capital factor from He et al. (2017), openint_puts is the open interest in put options, openint_calls is the open interest in call options, openint_tot is the total open interest in both puts and calls, baspread is the bid-ask spread in the option contract, volume is the traded volume in the option contract in a given month, pcratio is the ratio of open interest in puts to calls, and vix is the CBOE VIX index. Values in bold font indicate significance at the 5% level or better.

	F1	F2	F3	F4	F5
TED_spread	**34.6%**	7.2%	-2.0%	1.1%	**-23.3%**
intermed_capital	**-36.0%**	-8.2%	-10.3%	8.0%	0.1%
openint_puts	6.7%	**12.7%**	-4.0%	**13.0%**	-11.0%
openint_calls	**15.3%**	10.3%	-2.3%	**22.3%**	**-17.6%**
openint_tot	10.5%	12.2%	-3.4%	**17.4%**	**-14.2%**
baspread	0.4%	-6.4%	**17.2%**	**-18.3%**	2.5%
volume	9.3%	11.6%	**-12.3%**	8.4%	-11.3%
pcratio	2.3%	-4.4%	2.5%	-4.8%	10.8%
vix	**56.1%**	10.2%	**-26.3%**	**36.9%**	**-28.8%**

Table 3.12: IPCA Instrument Significance.
This table contains results from the bootstrap test for individual characteristics contributions to overall fit in the restricted IPCA specification with $K = 5$ factors. The test sets all elements in Γ_β related to a given characteristics equal to zero and compares the overall model fit against the model with fully specified Γ_β. The first column summarises the absolute reduction in total R^2 from setting the row in the matrix Γ_β pertaining to a given characteristic to zero. The second column contains p-values for the bootstrap test W_β with 1000 draws that tests $H_0 : \Gamma_\beta = [\gamma_{\beta,1}, \ldots, \gamma_{\beta,l-1}, \mathbf{0}_{K\times 1}, \gamma_{\beta,l+1}, \ldots, \gamma_{\beta,L}]$ against the alternative $H_1 : [\gamma_{\beta,1} \ldots, \gamma_{\beta,L}]$. The table is sorted by the reduction in total R^2 from largest to smallest.

	Reduction R^2_{total} (abs.)	W_β p-value
`mness:put`	1.24	0.00
`impvol`	0.90	0.00
`gamma:put`	0.44	0.00
`mness`	0.43	0.00
`gamma`	0.40	0.00
`theta`	0.16	0.00
`vega`	0.16	0.00
`impvol:put`	0.13	0.00
`embed_lev:put`	0.10	0.10
`ttm:put`	0.05	0.53
`ttm`	0.03	0.46
`theta:put`	0.03	0.12
`embed_lev`	0.03	0.17
`vega:put`	0.02	0.64

Table 3.13: IPCA Performance - Robustness This table reports the IPCA fit performance as measured by R^2_{total} and R^2_{pred} using both the restricted model ($\Gamma_\alpha = 0$) and the unrestricted model ($\Gamma_\alpha \neq 0$) with $K = 1, \ldots, 5$. In contrast to Table 3.2, the model estimated in this table uses charachteristics as implied by a Taylor expansion of the BMS model price as shown in Equation (3.12). The resulting characteristics are each interacted with a constant and an indicator variable that is equal one if the option is a put and is zero otherwise. Additionally, a constant is included in the set of characteristics. In **Panel A** performance measures are computed with respect to individual option contracts, while in **Panel B** performance measures are computed with respect to the characteristics-managed portfolios.

		No. Factors				
		1	2	3	4	5
		Panel A: Individual Options				
R^2_{total}	$\Gamma_\alpha = 0$	71.27%	77.83%	81.81%	83.98%	85.26%
	$\Gamma_\alpha \neq 0$	71.76%	78.26%	82.03%	84.09%	85.28%
R^2_{pred}	$\Gamma_\alpha = 0$	5.77%	5.85%	5.23%	4.92%	4.71%
	$\Gamma_\alpha \neq 0$	6.14%	6.08%	5.28%	4.93%	4.76%
		Panel B: Manged Portfolios				
R^2_{total}	$\Gamma_\alpha = 0$	98.43%	99.47%	99.65%	99.78%	99.91%
	$\Gamma_\alpha \neq 0$	98.43%	99.50%	99.68%	99.77%	99.90%
R^2_{pred}	$\Gamma_\alpha = 0$	7.14%	7.23%	6.74%	6.73%	6.67%
	$\Gamma_\alpha \neq 0$	7.03%	7.03%	6.77%	6.71%	6.67%

Table 3.14: IPCA Performance for Varying Sets of Instruments.
The table details the performance of different IPCA specifications for both the restricted and unrestricted models where the number of factors is $K = 1, \ldots, 5$. The performance measures include total and predictive R^2. The characteristics and market variables included in a given specification are marked by an "x", while fields left empty indicate that a characteristic was not included in the specification.

Model		Performance				Characteristics and Market Variables							
K	Unrestr.	R^2_{total}	R^2_{pred}	$R^2_{total,x}$	$R^2_{pred,x}$	mness	ttm	embed_lev	theta	IV	gamma	vega	volga
1	y	72.03	6.35	97.23	6.75	x	x	x					
2	y	77.94	6.31	98.55	6.80	x	x	x					
3	y	81.34	6.22	99.61	6.68	x	x	x					
4	y	83.25	6.14	99.82	6.63	x	x	x					
5	y	84.40	6.13	99.92	6.62	x	x	x					
1	n	71.25	5.76	97.03	6.81	x	x	x					
2	n	77.28	5.83	98.44	6.92	x	x	x					
3	n	80.89	5.98	99.51	6.65	x	x	x					
4	n	83.08	6.02	99.83	6.65	x	x	x					
5	n	84.35	6.09	99.92	6.62	x	x	x					
1	y	72.31	6.80	96.82	6.82	x	x	x	x				
2	y	79.15	6.41	98.16	6.75	x	x	x	x				
3	y	83.04	6.44	99.23	6.73	x	x	x	x				
4	y	85.87	6.37	99.72	6.73	x	x	x	x				
5	y	87.37	6.34	99.84	6.71	x	x	x	x				
1	n	71.30	5.69	96.48	6.53	x	x	x	x				
2	n	78.13	5.98	97.93	6.77	x	x	x	x				
3	n	82.63	6.39	99.15	6.75	x	x	x	x				
4	n	85.60	6.31	99.73	6.85	x	x	x	x				
5	n	87.17	6.27	99.84	6.84	x	x	x	x				
1	y	73.35	7.20	97.00	6.75	x	x	x	x	x			
2	y	80.67	7.19	98.47	6.98	x	x	x	x	x			
3	y	84.66	7.16	99.47	6.95	x	x	x	x	x			
4	y	87.52	7.13	99.72	6.97	x	x	x	x	x			
5	y	88.71	7.00	99.82	7.00	x	x	x	x	x			
1	n	71.91	5.47	96.56	6.28	x	x	x	x	x			
2	n	79.25	5.33	98.08	6.13	x	x	x	x	x			
3	n	83.90	6.40	99.35	6.73	x	x	x	x	x			
4	n	87.03	6.66	99.72	7.04	x	x	x	x	x			
5	n	88.44	6.88	99.79	7.08	x	x	x	x	x			

Table 3.14 continued.

Model		Performance				Included Option Characteristics and Market Variables							
K	Unrestr.	R^2_{total}	R^2_{pred}	$R^2_{total,x}$	$R^2_{pred,x}$	mness	ttm	embed_lev	theta	IV	gamma	vega	volga
1	y	73.72	7.26	96.61	6.71	x	x	x	x	x	x		
2	y	81.13	7.35	98.23	7.05	x	x	x	x	x	x		
3	y	85.54	7.21	99.37	6.90	x	x	x	x	x	x		
4	y	88.86	7.07	99.69	6.66	x	x	x	x	x	x		
5	y	90.26	6.98	99.78	6.72	x	x	x	x	x	x		
1	n	72.22	5.54	96.14	6.36	x	x	x	x	x	x		
2	n	79.66	5.43	97.74	6.28	x	x	x	x	x	x		
3	n	84.79	6.22	99.24	6.44	x	x	x	x	x	x		
4	n	88.41	6.35	99.67	6.46	x	x	x	x	x	x		
5	n	90.02	6.34	99.77	6.44	x	x	x	x	x	x		
1	y	74.24	7.22	96.00	6.71	x	x	x	x	x	x	x	
2	y	81.61	7.34	98.03	7.18	x	x	x	x	x	x	x	
3	y	86.09	7.25	99.25	7.04	x	x	x	x	x	x	x	
4	y	89.53	7.08	99.65	6.79	x	x	x	x	x	x	x	
5	y	90.83	6.98	99.73	6.71	x	x	x	x	x	x	x	
1	n	72.69	5.48	95.47	6.23	x	x	x	x	x	x	x	
2	n	80.18	5.29	97.54	6.06	x	x	x	x	x	x	x	
3	n	85.31	6.11	99.08	6.38	x	x	x	x	x	x	x	
4	n	89.15	6.31	99.63	6.49	x	x	x	x	x	x	x	
5	n	90.61	6.40	99.73	6.51	x	x	x	x	x	x	x	
1	y	74.50	7.54	95.41	7.18	x	x	x	x	x	x	x	x
2	y	81.85	7.54	98.07	7.56	x	x	x	x	x	x	x	x
3	y	86.48	7.44	99.26	7.47	x	x	x	x	x	x	x	x
4	y	89.79	7.28	99.65	7.21	x	x	x	x	x	x	x	x
5	y	90.76	7.23	99.74	7.15	x	x	x	x	x	x	x	x
1	n	72.86	5.35	94.93	6.33	x	x	x	x	x	x	x	x
2	n	80.32	5.25	97.59	6.30	x	x	x	x	x	x	x	x
3	n	85.45	5.92	99.08	6.45	x	x	x	x	x	x	x	x
4	n	89.30	6.11	99.64	6.55	x	x	x	x	x	x	x	x
5	n	90.77	6.22	99.74	6.65	x	x	x	x	x	x	x	x

Table 3.15: IPCA Performance - Daily Frequency.
This table reports the performance as measured by R^2_{total} and R^2_{pred} using both the restricted model ($\Gamma_\alpha = 0$) and the unrestricted model ($\Gamma_\alpha \neq 0$) with $K = 1, \ldots, 5$. Option characteristics `mness`, `ttm`, `embed_lev`, `theta`, `iv`, `gamma`, `vega` are each interacted with a constant and an indicator variable that is equal one if the option is a put and is zero otherwise. In **Panel A** performance measures are computed with respect to individual option contracts, while in **Panel B** performance measures are computed with respect to the characteristics managed portfolios. **Panel C** reports p-values for the test $\Gamma_\alpha = 0$ from a bootstrap with 1000 draws for each time t.

		No. Factors				
		1	2	3	4	5
		Panel A: Individual Options				
R^2_{total}	$\Gamma_\alpha = 0$	69.89%	85.57%	91.11%	92.73%	93.57%
	$\Gamma_\alpha \neq 0$	70.33%	85.72%	91.23%	92.84%	93.56%
R^2_{pred}	$\Gamma_\alpha = 0$	0.37%	0.41%	0.39%	0.40%	0.42%
	$\Gamma_\alpha \neq 0$	0.66%	0.64%	0.62%	0.61%	0.61%
		Panel B: Managed Portfolios				
R^2_{total}	$\Gamma_\alpha = 0$	94.34%	97.35%	99.31%	99.67%	99.76%
	$\Gamma_\alpha \neq 0$	94.31%	97.37%	99.31%	99.67%	99.77%
R^2_{pred}	$\Gamma_\alpha = 0$	0.35%	0.34%	0.34%	0.35%	0.37%
	$\Gamma_\alpha \neq 0$	0.24%	0.37%	0.35%	0.33%	0.35%
		Panel C: Bootstrap Test ($\boldsymbol{H_0 : \Gamma_\alpha = 0}$)				
W_α p-value		99.8%	95.6%	90.2%	44.4%	99.4%

Table 3.16: Comparison of IPCA against a No-Arbitrage Model at Daily Frequency. This table details the performance of both the Carr and Wu (2020) model and the restricted IPCA model with $K = 1, \ldots, 5$ factors at daily frequency. The daily returns are delta-hedged. For the no-arbitrage model the total R^2 is computed as follows: for a series R_i, $1, \ldots, 4$ as specified in Section 3.4.8 the R^2 is computed as $R^2_{total,i} = 1 - \text{Var}(R_i)/\text{Var}(R_0)$ where R_0 is the series of delta-hedged daily returns. The computation of the total R^2 in the IPCA model follows the usual form.

Panel A: Average Performance

	Carr & Wu - R^2_{total}				IPCA - R^2_{total}				
	R_1	R_2	R_3	R_4	K=1	K=2	K=3	K=4	K=5
All Options	74.2%	83.0%	79.8%	83.1%	69.9%	85.6%	91.1%	92.7%	93.6%

Panel B: Average Performance by Moneyness Bin

	Carr & Wu - R^2_{total}				IPCA - R^2_{total}				
mness	R_1	R_2	R_3	R_4	K=1	K=2	K=3	K=4	K=5
-2.5 to -2	33.4%	38.4%	-55.5%	-45.8%	33.1%	28.5%	29.2%	47.6%	48.7%
-2 to -1.5	53.5%	62.9%	37.4%	40.5%	65.2%	67.2%	72.1%	82.1%	83.3%
-1.5 to -1.0	67.3%	78.6%	73.4%	71.9%	77.7%	82.5%	88.0%	91.1%	92.1%
-1.0 to -0.5	72.0%	85.1%	84.2%	83.1%	77.0%	87.2%	93.4%	94.0%	94.7%
-0.5 to 0.0	71.2%	85.0%	84.9%	84.3%	66.9%	87.0%	93.3%	94.2%	94.9%
0.0 to 0.5	77.8%	85.3%	85.1%	86.7%	65.6%	88.2%	92.9%	93.9%	94.7%
0.5 to 1.0	81.1%	83.8%	80.8%	90.4%	72.9%	87.1%	91.7%	93.2%	94.2%
1.0 to 1.5	83.7%	77.4%	63.4%	89.9%	72.4%	79.4%	84.7%	88.6%	90.1%
1.5 to 2.0	87.5%	68.2%	25.4%	84.9%	57.6%	47.8%	59.8%	68.2%	70.2%

Panel C: Average Performance by Time-to-Maturity Bin

	Carr & Wu - R^2_{total}				IPCA - R^2_{total}				
ttm	R_1	R_2	R_3	R_4	K=1	K=2	K=3	K=4	K=5
1 Month	46.3%	62.1%	52.0%	60.1%	69.5%	80.2%	91.3%	94.0%	95.0%
2 Months	79.1%	92.6%	91.7%	94.6%	78.3%	91.8%	93.4%	94.9%	95.6%
3 to 6 Months	85.8%	92.8%	92.2%	93.8%	72.6%	90.2%	92.4%	93.5%	94.0%
6 to 12 Months	85.6%	88.4%	88.2%	88.6%	56.8%	78.2%	86.5%	87.8%	89.0%

Panel D: Average Performance by VIX bin

	Carr & Wu - R^2_{total}				IPCA - R^2_{total}				
VIX	R_1	R_2	R_3	R_4	K=1	K=2	K=3	K=4	K=5
10% to 20%	65.7%	78.1%	74.9%	76.9%	63.5%	85.2%	90.3%	91.6%	92.6%
20% to 30%	78.0%	85.4%	82.5%	84.7%	71.3%	84.8%	90.7%	92.4%	93.3%
30% to 50%	76.2%	84.3%	80.8%	85.8%	83.8%	89.4%	95.0%	96.4%	97.0%
50% to 90%	76.0%	83.6%	80.3%	87.4%	83.3%	89.4%	94.1%	96.1%	96.4%

Chapter 4

What Drives Asset Holdings? Commonality in Investor Demand[1]

"The special sphere of finance within economics is the study of allocation and deployment of economic resources, both spatially and across time, in an uncertain environment. To capture the influence and interaction of time and uncertainty effectively requires sophisticated mathematical and computational tools."

– Robert C. Merton; Nobel Lecture 1997

4.1 Introduction

Asset pricing models rely on market clearing to equate the equilibrium demands of all investors. Akin to a zero-sum game, for every quantity of an asset sold an equal quantity has to be purchased. Hence, market clearing introduces an almost mechanical correlation structure in investor demands. In this paper, I investigate the economic and statistical properties of this correlation structure. Specifically, I study whether changes in investor holdings exhibit a factor structure, what a potential set of factors in holdings changes captures economically, and what the existence of a factor structure in holdings can mean for asset pricing.

[1]Parts of the chapter have been presented at the 2020 AEFIN PhD Mentoring Day, and the Warwick University Finance Brown Bag Seminar.

Contrary to most asset pricing models that focus on return and consumption data, this paper studies holdings data. Holdings data complements price / return data by allowing the researcher to explicitly track investor demand across assets and time. Using this data, I revisit the investor portfolio choice problem and empirically model asset demand as measured through *changes in holdings*. Studying holdings changes in a factor structure approach is both novel to the literature and important for a number of reasons. Firstly, there is a notion that financial markets are risk-sharing markets. This means asset supply and demand shocks should reflect changes in risks. Hence, to first order, a factor modelling approach should allow us to obtain a better understanding for the factors that investors consider in their investment decisions. Secondly, a factor modelling approach could help in discerning how different investor types act in different market conditions (e.g., mutual funds vs. hedge funds). Given market clearing, it is obvious that not all investors can trade in the same direction at the same time. A factor model can capture differences in investor behaviour by means of signed, investor-specific factor loadings. Thirdly, a suitably constructed factor model can also aim to be comprehensive in the sense that it describes not only a single asset or investor type, but many assets and investors simultaneously.

My modelling approach has two essential ingredients: aggregate factors that capture the shocks to asset supply / demand and investor specific exposure to the factors. This empirical approach reflects an economy with aggregate shocks in the presence of which heterogeneous investors choose from a pool of differentiated assets (cf. Campbell et al., 2003). By extracting factors in a data-driven manner, I do not take an *ex ante* stance on the nature of the shocks, that is, whether they originate on the demand or supply side. Rather, I will shed light on the nature of the extracted factors by relating them to relevant economic time series.

Methodologically, I build on *Instrumented Principal Components Analysis (IPCA)* (Kelly et al., 2019, 2017), a latent factor modelling approach originally applied in the context of equity returns. After an extension to allow for a third dimension in the input data, i.e. the investor dimension, IPCA provides a comprehensive approach to jointly model latent, systematic factors as well as investor-specific, time-varying factor loadings that partially depend on observed asset characteristics. Asset characteristics enter IPCA akin to instruments allowing for efficient and consistent recovery of latent factors and loadings.

This paper is linked to a recent strand of literature that draws on investor holdings data to better understand asset prices. Koijen and Yogo (2019a) propose a particular way in which we can think about holdings data, i.e. a characteristics-based demand system. In their paper the optimal portfolio choice differs across investors with heterogeneous beliefs and depends on observed asset characteristics and latent demand. Contrary to their work, this paper explicitly models the correlated trading of different investors. Thus, this paper effectively relaxes the commonly made assumption of atomistic investors. However, this paper accommodates a characteristics-based demand notion of Koijen and Yogo, in which investor holdings are determined by asset characteristics alone, alongside a set of systematic factors common to all investors (albeit with differing loadings).

4.1.1 Findings

Using a data set of nearly forty years of changes in U.S. equity holdings across seven broad investor types, I estimate an extended IPCA model. My extension of IPCA, dubbed "IPCA3D" for distinction, allows for heterogeneous characteristics loadings across different investor types, but assumes a latent factor structure common to all investors. Initially, without any latent factors, I find that a set of six characteristics as used in Koijen and Yogo (2019a) explains less than 1% of the variation in holdings changes in my sample. Allowing for a wider set of characteristics as in Kelly et al. (2019) improves the model fit performance by a factor of six. I then proceed to allow for latent factors in the relation between asset characteristics and changes in holdings. With a single latent factor the fraction of explained variance in holdings more than doubles relative to the model that only allows for a characteristics based intercept term. Adding more than three factors contributes little to model fit performance. For parsimony I focus on a model with $K = 2$ factors, and find a stock level R^2 of roughly 10%. Instead of focusing on variation in holdings at stock level alone, I also form characteristics-managed portfolios for each of the 36 characteristics in the model. At portfolio level, the model with $K = 2$ factors has an R^2 in excess of 90%, highlighting IPCA3D's ability to accurately track changes in investor holdings as a function of asset characteristics. In a sub-sample analysis I assess the model's performance in explaining the changes in holdings for particular investor types. The model performs particularly well in explaining variation in holdings for mutual funds, investment advisors (which includes many hedge funds), as well as the small investors sector. The holdings of insurance companies and pension funds are explained least well.

To understand the economic character of the factors recovered by IPCA3D, I run a comprehensive correlation analysis relating the factors to relevant economic time series. Starting with a set of principal components from a well-known macro data set as used in Ludvigson and Ng (2009) & Stock and Watson (2012), I find that the first latent factor clearly relates to the macro state of the economy (including industrial production, employment, productivity, etc.). Importantly, the factor exhibits pro-cyclicality. The second factor in my benchmark specification is related to household and business sector balance sheet variables, especially measures of (household) debt to income. Therefore, I interpret this second factor as a measure of financial constraints of investors.

To gain an understanding for what differentiates the investment behaviour of different investor types, I study the time-varying, sector-specific loadings obtained alongside the factors. Analysing the loadings on the first factor (business cycle), I establish that banks and mutual funds act in a pro-cyclical manner, while hedge funds and pension funds act in a counter-cyclical manner. These patterns broadly match those found in Timmer (2018). The loadings on the second factor exhibit a clear split between the institutional and small investor (household) sector. Given the correlation of this factor with measures of investors' financial constraints, this suggests that as smaller investors become more financially constrained they divest, while larger institutions act as counter-parties.

Next, I assess the overall importance of the 36 characteristics used in the baseline specification. I find that among the eight characteristics significant at the 1% level in a bootstrap exercise, momentum and liquidity related variables stand out. This suggests that in addition to the characteristics used in Koijen and Yogo (2019a) there is evidence that investors take into account past returns and liquidity when choosing their portfolios. In a set of panel regressions of the decomposed changes in holdings onto past returns, I find that past returns are particularly important in explaining the characteristics-based part of the holdings changes, while their impact on systematic holdings changes is far less pronounced. This finding also helps to further pin down the source of the cyclicality captured in the systematic factors mentioned earlier: it mostly derives from macroeconomic channels such as industrial production, employment, etc., and less from past returns. Overall, the relevance of past returns as drivers of portfolio holdings is in line with studies such as Grinblatt et al. (1995); Nofsinger and Sias (1996); Sias (2004) that investigate different drivers of mutual fund herding. Liquidity considerations are consistent with Gompers and Metrick (2001), who document a particular institutional demand for larger, more liquid stocks.

The paper concludes by investigating relations between asset prices and changes in holdings. In particular, using the fitted values of an IPCA3D model estimated in expanding window fashion, I decompose the holdings changes for each investor sector into three components: intercept, systematic part, idiosyncratic part. Then, in a host of predictive regressions, I analyse which parts of investors' holdings changes predict future returns. Changes in the holdings of mutual funds in response to systematic shocks are negatively associated with future returns, consistent with the idea of institutional price pressure effects (see, e.g., Coval and Stafford, 2007). A similar effect obtains for banks. Further, I find that the changes in holdings of investment advisors, unrelated to systematic shocks are predictive of future returns. Both the characteristics-based and idiosyncratic changes in investment advisor holdings predict returns for up to two quarters ahead. The relevance of idiosyncratic investment advisor demand suggests that in addition to the conventional characteristics entertained by investors (and used in the estimation of IPCA3D), investment advisors possess superior information that is useful for market-timing.

4.1.2 Related Literature

This paper relates to three streams of literature. First, this paper builds on a growing literature in demand-based asset pricing. In Koijen and Yogo (2019a) (KY hereafter), log-utility maximising investors face an inter-temporal budget constraint and a short sale restriction. Assets are differentiated by characteristics and their risk attributes are fully determined through characteristics loadings in a low-dimensional factor structure of returns. At the heart of their paper, there is an estimation which in stylised form can be summarised as follows. For investor i at time t the portfolio weight $w_{i,j,t}$ is regressed on a set of asset characteristics $z_{j,t}$

$$\frac{w_{i,j,t}}{w_{i,0,t}} = \underbrace{\sum_{l=1}^{L} \beta_{l,i,t}\, z_{l,j,t}}_{\text{Characteristics-Based Demand}} + \underbrace{\epsilon_{i,j,t}}_{\text{Latent Demand}}, \tag{4.1}$$

where $w_{i,0,t}$ denotes the weight of investor i on the *outside asset*, i.e. the aggregate holding in all assets not modelled immediately in their study.[2] KY refer to latent demand as the part of demand related to characteristics unobserved by the econometrician and demonstrate that latent demand explains the vast majority of stock returns.[3] My paper examines more closely the structure and economics of what KY term "latent demand", albeit this paper pursues a reduced-form, factor modelling approach to synthesise cross-sector and aggregate macroeconomic effects.

Second, this paper's modelling relates to a large literature concerned with state-/ regime-based asset pricing going back to at least Harrison and Kreps (1978). Brennan et al. (1997) develop a model in which investors choose portfolios in the presence of aggregate state variables for a mean-variance optimising investor. In related research, Campbell et al. (2003) solve a more complex version of the portfolio choice problem when the investor has Epstein-Zin utility. Guidolin and Timmermann (2007) examine asset allocation in the presence of regime switches in asset returns, finding evidence for four separate regimes: crash, recovery, slow growth and bull markets. Investors' asset allocation varies over these states and depends on their heterogeneous preferences, e.g., a short-horizon investor's portfolio weights are far less stable than those of a long-horizon investor. Brennan and Xia (2002) extends Brennan et al. (1997) and accounts for inflation. My paper contributes an empirical perspective on this literature by recovering the main macroeconomic channels that drive asset demand and can therefore provide a useful guide for further theory research.

Third, since this paper's results are based on data pertaining to a hand-

[2]Note, this stylised summary is intended to give a simplified idea of the core estimation in Koijen and Yogo. Their estimation requires an instrumental variable approach that addresses the issue that demand shocks across investors are likely correlated and that latent demand and asset prices are potentially endogenous.

[3]In a related paper Koijen and Yogo (2019b) estimate an international demand system for 36 countries. Using their demand system they decompose exchange rates, yields and equity prices into three channels: macro variables, policy variables, and latent demand. Finally, Koijen et al. (2019) use a demand system approach to study how different investors contribute to the price formation process and which characteristics matter most for prices.

ful of broad investor types (banks, pension funds, insurance firms, etc.), it relates to a whole range of literatures that have examined investment behaviour in these different sectors. Perhaps most prominently and due to data availability, a large literature has focused on understanding mutual fund performance and manager behaviour. This literature has on the one hand examined fund investment styles as drivers of their asset demand (e.g., see Brown and Goetzmann, 1997; Chan et al., 2002; Berk and Green, 2004; Barber et al., 2016). On the other hand, momentum based trading and herding have been found to explain some mutual funds' behaviour (e.g., see Grinblatt et al., 1995; Nofsinger and Sias, 1996; Wermers, 1999; Sias, 2004). In addition, a number of papers focused on debt holdings provide evidence for cyclicality in investor holdings. Timmer (2018) finds banks and investment funds behaving pro-cyclically, while insurance firms and pension funds act counter-cyclically. Abbassi et al. (2016) document differences in bank trading behaviour in debt markets during the great financial crisis driven by differing trading expertise of banks. Another growing literature looks at the investment behaviour of individual investors and households. Behavioural factors such as disposition effects and overconfidence are key characteristics of individual investor behaviour.[4] Odean (1998) demonstrates an individual investor disposition effect that makes them reluctant to realise investment losses / gains, i.e. holding losing investments too long and selling winning investments too soon. Relatedly, Barber and Odean (2000) shows that individual investor portfolios with high turnover underperform those of investors with less inclination to trade by a significant margin. My paper adds to this literature, by studying the interplay between different sectors of investors that have commonly been studied in isolation from each other. In this regard, my paper shows which investors "take the other end of the trade" on average.

This paper proceeds as follows. In Section 4.2, I introduce instrumented principal components analysis and extend the method to allow for an additional dimension in the data. Section 4.3 describes the construction of the data and its properties. Section 4.4 presents results for the factor structure in holdings changes, interprets the recovered factors, and investigates differences in investor exposure to the recovered factors. Section 4.5 documents a number of asset pricing results following from the factor structure in holdings. Finally, Section 4.6 concludes.

[4]Barber and Odean (2013) perform a comprehensive and detailed review of this literature.

4.2 Methodology

In this section, I introduce the methodology used in the later analysis. I start by briefly summarising the original IPCA model of Kelly et al. (2019), before demonstrating an extension of IPCA that allows me to handle the additional investor dimension in the holdings data.

4.2.1 Recap: Instrumented Principal Components

Instrumented Principal Components Analysis (IPCA) (Kelly et al., 2019) simultaneously estimates common factors and dynamic loadings from a panel of data.[5] In particular, IPCA allows the researcher to include conditioning information that aids recovery of latent factors. For example, in the case examined by Kelly et al. (2019) latent factors are extracted from a panel of asset returns using asset characteristics as conditioning information. As suggested by the name of the method, the conditioning information effectively functions as a set of instruments for the conditional loadings.

The IPCA model specification for an excess return $r_{i,t+1}$ is given as

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1} \tag{4.2}$$

$$\alpha_{i,t} = z'_{i,t}\Gamma_\alpha + \nu_{\alpha,i,t}, \qquad \beta_{i,t} = z'_{i,t}\Gamma_\beta + \nu_{\beta,i,t},$$

for $i = 1, \ldots, N$ assets and $t = 1, \ldots, T$ periods. The dynamic factor loadings $\beta_{i,t}$ are a $1 \times K$ vector, and f_{t+1} is a $K \times 1$ vector of latent factors. Conditioning information such as assets characteristics enter the model through a $L \times 1$ vector $z_{i,t}$ containing L characteristics for asset i. In order to uniquely pin down factors and loadings, Kelly et al. (2019) impose that $\Gamma'_\beta \Gamma_\beta = \mathbb{I}_K$ and $\Gamma'_\alpha \Gamma_\beta = \mathbf{0}_{K\times 1}$, that the unconditional second moment matrix of f_t is diagonal with descending entries, and that the mean of f_t is non-negative.

The role of characteristics as conditioning information in IPCA is two-fold. On the one hand observable instruments increase estimation efficiency, on the other hand time-variation in characteristics gives rise to dynamic factor loadings $\beta_{i,t}$. Including characteristics as conditioning information follows the idea of asset migra-

[5] A Python package implementing the IPCA method of Kelly et al. (2019) that I am jointly maintaining with Leland Bybee is available on GitHub: `https://github.com/bkelly-lab/ipca`.

tion, i.e. assets changing their "identity" as time passes by. Characteristics are mapped onto the set of latent factors f_{t+1} via a matrix Γ_β. This mapping implicitly carries out a dimensionality reduction that isolates characteristics that provide independently relevant information about expected returns and averages out those characteristics that do not. Furthermore, the IPCA feature of dimensionality reduction provides the researcher with an opportunity to entertain a large number of characteristics simultaneously, hence foregoing the need to take a potentially sub-optimal stance on the set of relevant characteristics ex-ante.

4.2.2 Extending IPCA to the Third Dimension

The three-dimensional instrumented principal components methodology (IPCA3D) introduced here, extends the instrumented principal components approach of Kelly et al. (2019) to allow for an additional dimension of the dependent variable. I assume a setup with I investors, N_t assets, and T periods. The dependent variable $\Delta h_{i,t}$ is a vector of holdings changes for investor i at time t. As in the original model, IPCA3D assumes that dynamic factor loadings are potentially dependent on a set of L asset characteristics collected in a matrix Z_t of dimensions $N_t \times L$. The IPCA3D model is stated as follows

$$\begin{aligned} \Delta h_{i,t} &= \alpha_{i,t} + \beta_{i,t} f_t + \epsilon_{i,t}, \\ \alpha_{i,t} &= Z_t \Gamma_{\alpha,i} + \nu_{\alpha,i,t}, \\ \beta_{i,t} &= Z_t \Gamma_{\beta,i} + \nu_{\beta,i,t}, \end{aligned} \tag{4.3}$$

where $h_{i,t}$ is a $N_t \times 1$ vector of the holdings (portfolio weights) that investor group i assigns to the N_t assets at time t. In this way, IPCA3D has two central assumptions. First, IPCA3D assumes that the cross-section of assets which each investor group i holds is the same for all investors, hence ensuring that $h_{i,t}$ is of dimension $N_t \times 1$ for all i. This assumption is made workable by including zero holdings in the vector of holdings. The decision of not holding an asset may well be as informative as the decision of holding an asset. Second, IPCA3D assumes that each of the investor groups i is present throughout the entire sample period. This assumption will later allow me to re-state IPCA3D analogous to the original IPCA model, thereby facilitating efficient estimation of the model. Furthermore, this assumption of panel balancedness at the investor group level resolves a dimensionality issue resulting from the usually large number of individual investors that typically reaches into the thousands, making estimation at the individual investor level very challenging. The assignment of individual investors to the i groups is at the discretion of the researcher and should be guided by the investor characteristics that distinguish different sets of investors. For example, investor groups could be created by classifying the type of an investor as pension funds, hedge funds, banks, etc. as in Koijen and Yogo (2019a). Alternative classifications could see funds labelled according to their investment style such as “value”, “growth”, or “momentum”.

Estimation I now turn to the procedure for estimating IPCA3D. The estimation of IPCA3D can be greatly simplified by re-phrasing the model as a special case of IPCA. This is achieved by stacking the I equations for all investors. I begin by vertically stacking the vectors of holdings $h_{i,t}$ such that $\tilde{h}_t = [h_{1,t}; \ldots; h_{I,t}]$ is an $N_t I \times 1$ vector. Then, denoting the diagonally stacked matrix of characteristics as $\tilde{Z}_t = \mathbb{I}_{I\times I} \otimes Z_t$, IPCA3D can be re-stated as

$$\Delta\tilde{h}_t = \tilde{Z}_t\tilde{\Gamma}_\alpha + \tilde{Z}_t\tilde{\Gamma}_\beta f_t + \tilde{\epsilon}_t, \tag{4.4}$$

where $\tilde{\Gamma}_\alpha = [\Gamma_{\alpha,1}; \ldots; \Gamma_{\alpha,I}]$ and $\tilde{\Gamma}_\beta = [\Gamma_{\beta,1}; \ldots; \Gamma_{\beta,I}]$ are the vertically stacked matrices of the investor mapping matrices $\Gamma_{\alpha,i}$ and $\Gamma_{\beta,i}$. This re-statement fully maps IPCA3D into the original IPCA specification of Kelly et al. (2019). Estimation of the model then proceeds by an alternating least squares (ALS) procedure that switches back and forth between the first order condition of $\tilde{\Gamma}_\alpha$ / $\tilde{\Gamma}_\beta$ and f_t, in which the original IPCA matrices are replaced with the re-stated objects introduced above. Note, the identification restrictions applied in IPCA above now apply to the matrices $\tilde{\Gamma}_\alpha$ and $\tilde{\Gamma}_\beta$.

Evaluating Model Fit To assess the performance of the IPCA3D factors in capturing variation in holdings, building on Kelly et al. (2019), I use two measures, the total R^2 at investor-asset level, as well as the total R^2_x at investor-characteristics managed portfolio level. The investor-asset level R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i,j,t}\left(\Delta h_{i,j,t} - z'_{j,t}(\hat{\Gamma}_{\alpha,i} + \hat{\Gamma}_{\beta,i}\,\hat{f}_t)\right)^2}{\sum_{i,j,t}\Delta h^2_{i,j,t}}, \tag{4.5}$$

while the characteristics-managed portfolio R_x^2 is defined as

$$R_x^2 = 1 - \frac{\sum_{i,t}\left(\Delta x_{i,t} - Z_t' Z_t(\hat{\Gamma}_{\alpha,i} + \hat{\Gamma}_{\beta,i}\hat{f}_t)\right)^2}{\sum_{i,t}\Delta x_{i,t}^2}, \tag{4.6}$$

where the characteristics managed portfolio holdings changes are given by $\Delta x_{i,t} = Z_t'\Delta h_{i,t}$. Essentially, the characteristics managed portfolio is a characteristics weighted average of holdings across assets. These two formulations of the R^2 allow assessment of model fit at different levels of granularity. While the general R^2 takes into account individual assets at investor level, the R_x^2 instead focuses on matching holdings changes in the characteristics managed portfolios of each investor.

4.3 Data & Variables

This paper makes use of two main data sets. Firstly, stock information including prices and returns from the Center of Security Prices (CRSP) and stock characteristics (accounting data) from Compustat. Secondly, institutional holdings data provided in quarterly fillings of form 13F at the Securities & Exchange Commission (SEC). I will describe both data sets in more detail now.

4.3.1 Stock Characteristics

The construction of monthly stock characteristics data follows Freyberger et al. (2020b) and Kelly et al. (2019). This baseline set of characteristics deviates from KY, by many additional characteristics (6 characteristics in KY vs 36 in Kelly et al. (2019)). In particular, I include variables constructed on past returns and trading frictions.[6] The characteristics are: assets-to-market-equity (`a2me`), total assets (`at`), sales-to-assets (`ato`), book-to-market (`beme`), market beta (`beta`), cash-to-short-term-investment (`c`), capital turnover (`cto`), ratio of change in property, plants and equipment to the change in total assets (`delta_pi2a`), earnings-to-price (`e2p`), cash flow-to-book (`freecf`), idiosyncratic volatility with respect to the FF3 model (`idiovol`), investment (`investment`), dividend-price-ratio (`ldp`), leverage (`lev`), market capitalisation (`lme`), turnover (`lturnover`), net operating assets (`noa`), operating accruals (`oa`), operating leverage (`ol`), price-to-cost margin (`pcm`), profit

[6] KY exclude past return variables as they could pose a threat to identification in their framework.

margin (pm), gross profitability (prof), Tobin's Q (q), price relative to its 52-week high (rel_to_high_price), return on net operating assets (rna), return on assets (roa), return on equity (roe), momentum (r12_2), intermediate momentum (r12_7), short-term reversal (r2_1), long-term reversal (r36_13), sales-to-price (s2p), capital intensity (sat), the ratio of sales and general administrative costs to sales (sga2s), bid-ask spread (spread), and unexplained volume (suv).

The above set of characteristics will form my baseline set for the later analysis. Using an initially large set of characteristics leverages IPCA3D's ability to filter out those characteristics from the set that provide independently relevant information to describe the left-hand-side variable. However, in order to compare with results from KY, I will also make use of the original set of six characteristics used by KY. The KY set of characteristics includes market beta, book equity, market equity, dividend-to-book-equity, profitability, and investment.

With regards to the assets I study, I follow KY and focus on ordinary common shares (CRSP share codes 10 and 11) with non-missing characteristics.

Asset characteristics are merged with the holdings data (described in the next section), using their CRSP PERMNO identifier. The characteristics are lagged such that they are available at the beginning of the quarter over which holdings are observed in order to mitigate concerns around a look-ahead bias. An obvious caveat of lagging asset characteristics relative to the holdings is that some characteristics will likely change over the quarter, while the IPCA3D loadings are assumed fixed for the quarter. For simplicity, I choose not to dive deeper into this issue, but future research could investigate mixed frequency approaches in the spirit of MIDAS (Ghysels et al., 2005, 2007; Andreou et al., 2010, 2013) for more efficient pairing of holdings and characteristics within the quarter.

4.3.2 Institutional Holdings

The construction of the data on institutional holdings follows Koijen and Yogo (2019a).[7] The holdings data under Form 13F covers all institutional investment firms exercising investment discretion over more than \$100 million in aggregate market value and is available from the first quarter of 1980 onwards. Access to Form 13F

[7] Replication code, including code for sample construction, is provided online at the *Journal of Political Economy*: https://www.journals.uchicago.edu/doi/10.1086/701683.

is provided through the Thomson Reuters Institutional Holdings Database (s34 file). Contrary to the s12 file that covers individual mutual fund holdings, s34 is aggregated to the investment firm level and hence does not allow to distinguish between individual funds run by the same management company. I follow the institutions grouping in KY and group institutions into six types: banks, insurance companies, investment advisors, mutual funds, pension funds, and other 13F institutions. The investment advisors category includes many hedge funds. The group of "other" 13F institutions includes university endowments, foundations and non-financial corporations. Given the total number of outstanding shares as provided by CRSP, I infer the residual holdings by non-13F institutions. These non-13F institutions include both retail investors and institutions below the $100 Million filing threshold. In the rest of the paper the non-13F sector will be dubbed "small investor" sector.

For a brief inspection of the sample properties, Figure 4.1 shows the market share held by the seven different institutional sectors in the sample. At the start of the sample period, small investors which includes households, account for about two-thirds of the total outstanding market capitalisation. Over the course of the sample period from 1980 to 2017, this share decreases to around one third. Accordingly, the mutual fund and investment advisor sectors gain in size, with mutual funds making up around one third of the market share and investment advisors contributing about 20%. The size of the market share held by banks declines only slightly over the forty-year sample period. The pension fund sector holds less than 5% of the equity market with little fluctuation over the sample period.

Changes in Holdings The quantity of interest studied in this paper, i.e. changes in holdings, differs from Koijen and Yogo (2019a) in two ways. Firstly, KY perform their analysis looking at the level of holdings, while I focus on changes in holdings to capture dynamic effects. Secondly, I focus on the number of shares held by an investor instead of U.S. dollar holdings as in Koijen and Yogo (2019a). This choice allows me to single out the part of variation in holdings that comes from active trading, i.e. changing the quantity of an asset owned, opposed to changes in holdings driven by changes in market prices.

For each stock and investor, I compute the change in the number of shares held by the investor relative to the outstanding number of shares of the stock. Using shares outstanding as a normalising constant has the benefit that it is independent of the investor, hence, providing a way to focus on variation in holdings across investors instead of variation across assets. Precisely, I compute for investor i and asset j

$$\Delta h_{i,j,t} = \frac{(\tilde{S}_{i,j,t} - \tilde{S}_{i,j,t-1})}{\mathbb{S}_{j,t}} - \tag{4.7}$$

where $\tilde{S}_{i,j,t}$ is the split adjusted number of shares of investor i in asset j at time t and $\mathbb{S}_{j,t}$ is the total number of (split-adjusted) shares outstanding.

Note that the panel of holdings is highly unbalanced at the investor level due to investors entering and exiting the sample over time. In order to estimate the IPCA3D model in Equation (4.3), I require the data of holdings to be balanced at the investor level. This is achieved by aggregating holdings for each asset in each of the investor types / sectors of the 13F data. This means instead of studying changes in the holdings of individual investors (e.g. an individual mutual fund), my paper focusses on changes in the holdings at the sector level (e.g. the mutual fund sector defined as the aggregate of all shares held by mutual funds). The shares not held by 13F institutions are assigned to the small investor sector.[8] Naturally, the study of sector-level holdings changes shifts the focus of the paper from the intra-sector to the inter-sector margin, i.e. I investigate asset exchange behaviour between institutional sectors rather than individual investors in those sectors. While this restriction reduces the granularity of my data, it also eases a concern around the dimensionality of the individual investor level data that would require the estimation of thousands of characteristics-loadings matrices $\Gamma_{\beta,i}$.

Since the choice of focusing on sectors instead of individual investors is potentially restrictive, it is necessary to understand to which extent sector level holdings reflect holdings of individual investors within the sector. I assess this question by performing a principal components analysis (PCA) of the panel of individual investor holdings changes on an asset-by-asset basis. Given the missing observations in this panel, standard PCA is not feasible. Instead, following Tipping and Bishop (1999), I use a probabilistic PCA procedure that allows to handle missing data. Table 4.2 details the portion of variation in holdings changes explained by the first to third

[8] Koijen and Yogo (2019a) assign shares not held by 13F investors to a household sector. However, note that rather than being held by households a sizeable part of these shares are in fact held by institutions smaller than the 13F threshold or foreign institutional investors without reporting obligations.

principal component on average across all assets. I find that across all investor types the first principal component explains around two thirds of the variation in holdings changes of individual investors. The second and third principal component explain progressively less with about 18% and 7%. Given that principal components can be regarded as particular linear combinations of the inputs, this result suggests that (up-to an approximation) the sector level holdings changes are representative of holdings changes at individual investor level.

Next, I inspect the properties of the sector-level holdings changes by testing their stationarity in an Augmented Dickey and Fuller (1979) (ADF) test. Figure 4.3 shows the cumulative distribution (CDF) of p-values from the ADF tests performed individually for each time series $\Delta h_{i,j}$. Comparing Panels (A) and (B) in Figure 4.3, I find that the first difference is sufficient to reject the presence of a unit root for the vast majority of investors and stocks in my sample.

Summary Statistics In order to provide some initial intuition for the properties of the holdings changes constructed in this paper, Figure 4.2 plots the trading activity for two exemplary cases: the IBM and Pfizer stocks. The plots detail the quantity of assets bought and sold by each sector every quarter in terms of the market capitalisation outstanding. Consistent across the two examples, the time series of holdings changes appear stationary and show spikes around the typical distress periods in the dot-com bubble and the great financial crisis. To complement the exemplary results, Table 4.1 provides summary statistics of the changes in holdings by investor type split by quarters in which an investor type was an overall buyer / seller of a stock. The summary statistics indicate that the small investor sector that is comprised of non-13F investors is the largest buyer and seller on average, to be followed by mutual funds and investment advisors. This finding is not surprising given that in the early parts of the sample the 13F investors only hold about one third of the market (see Koijen and Yogo, 2019a, Table 2), indicating that the residual holdings of small investors are relatively larger. However, over the course of the sample period, this fact is reversed with more than two thirds of the market being held by 13F institutions. Econometrically, this change in the size of the individual sectors is handled by allowing for time-varying loadings, $\beta_{i,t}$, that are sector-specific. Nevertheless, this fact rules out comparison of the absolute sizes of the time-varying loadings across investors, instead, calling for comparison in relative terms.

Studying the unconditional sample summary statistics as above, masks interesting exchange behaviour *between* institutional types. One way to gain a more detailed understanding of the broad patterns of asset exchange between investors is to condition the computation of summary statistics on certain institutional types buying / selling a given stock and then to observe the contemporaneous behaviour of the other types of investors. Table 4.3 details those results. The left half of the table conditions on a certain institutional type buying a given asset, while the right half conditions on that institutional type selling the asset. I find that the buying of small investors is on average met primarily by mutual funds, investment advisors and banks selling the asset. Similarly, when small investors are selling, mutual funds and investment advisors acquire the lions share of the assets sold. This finding suggests that one interesting exchange margin could be between the institutional and small investor sector. In fact, in Section 4.4.3 I find that one of the latent holdings factors essentially captures this interaction. Another interesting finding emerges for investment advisors. While the majority of their buying is met by small investors that sell, the majority of their selling is met not only by small investors but also mutual funds that buy.

4.4 Empirical Results: Dissecting Investor Demand

4.4.1 Estimation

I estimate the IPCA3D system (Equation (4.3)) using the data described in Section 4.3. In particular, I use two sets of characteristics. First, I follow Koijen and Yogo and use six characteristics that include market equity, book equity, profitability, investment, dividends to book equity, and market beta. Second, in order to assess the impact from using a wider set of characteristics, I employ the set of 36 asset characteristics as used in Kelly et al. (2019) and described in Section 4.3.1. Note, in order to mitigate the influence of outliers in characteristics, I cross-sectionally rank-transform each characteristic at a given observation date to the interval $[-0.5, 0.5]$. This transformation focuses on the cross-sectional ordering of characteristics instead of their absolute magnitude. Furthermore, I winsorise the period-by-period distribution of changes in holdings at the 1% and 99% quantile to mitigate concerns around erroneous holdings records.

4.4.2 Model Fit Performance

Using the entire sample period of available data, i.e. 1980:Q1 to 2017:Q4, I fit the IPCA3D model to the data of holdings changes in Section 4.3. Table 4.4 summarises the fit performance for models with $K = 0, \dots, 5$ factors for the two sets of characteristics.

I start by comparing the impact from using a wider set of characteristics than originally used in KY. To that end, I focus on a specification that does not include any latent factors (i.e. $K = 0$) which is akin to running a panel regression of changes in holdings onto asset characteristics. Comparing the first row of Panels A and B of Table 4.4 ("Intercept Only"), I find that allowing for more characteristics increases the R^2 at stock level by a factor of six. The increase in R_x^2 is more moderate with an increase from around 9% to about 16%. This evidence suggests that it is necessary to consider a wider set of characteristics to explain variation in holdings across investors and time. The result is robust to accounting for the higher number of model parameters by means of an adjusted R^2 at both stock and portfolio level. Hence, in the following I will focus on models that make use of the larger set of 36 characteristics.

Next, I turn to models that allow for latent factors in the mapping between characteristics and holdings changes (see Equation (4.3)). Allowing for a single latent factor, i.e. $K = 1$, the R^2 at stock level more than doubles from 3.5% to 8.3% (Table 4.4, Panel B). At the portfolio level, the R_x^2 increases by more than a factor of five. Increasing the number of latent factors further, the fraction of explained variation in changes of holdings keeps increasing, however, gains become quite gradual after two to three factors. Interestingly, comparing the results between the models with few vs. many characteristics, i.e, Panel A vs. Panel B in Table 4.4, I find that the impact of allowing for more characteristics becomes relatively smaller once latent factors are employed. For example, with $K = 2$ latent factors, the difference in R_x^2 is a mere 0.3%.

In the interest of parsimony, my following analysis will focus on the model with $K = 2$ factors. This is motivated by the fact that allowing for more than two factors only marginally increases the R^2 at single stock level. Simultaneously, two factors appear to be sufficient to capture the vast majority of variation in holdings changes at the level of the 36 characteristics-managed portfolios as evidenced by an R_x^2 of 92.4% (see Table 4.4, Panel A).

Since little is known from earlier literature about changes in holdings as modelled here, it is interesting to put the fit performance of the IPCA3D model into perspective by comparing with IPCA model fit for stock returns. Kelly et al. (2019) find that for monthly returns, on a sample of stocks comparable to the one used in this paper, the stock level R^2 ranges from about 15% to 20%, depending on the number of latent factors (see Table 1 in their paper). Recall, the R^2 for changes in holdings discussed above reaches roughly 10% at the stock level when using the enlarged set of characteristics (see 4.4, Panel B). Hence, the amount of variation in holdings that can be explained in a latent factor model with time-varying loadings is of comparable magnitude to that of returns.

Fit Performance by Investor Type

In addition to studying the average fit performance of the IPCA3D model pooled across investor types, it is interesting to dissect the fit performance conditional on investor type. Therefore, I compute the stock level R^2 and characteristics-managed portfolio level R_x^2 only from those observations pertaining to each investor type. Tables 4.5 and 4.6 report the results. For brevity, I focus on the results at stock level in Table 4.5.

Focusing on the model with $K = 2$ and the enlarged set of characteristics as motivated above, IPCA3D best captures the variation in holdings of the small investor sector ($R^2 = 13.4\%$), closely followed by mutual funds ($R^2 = 10.1\%$) and banks ($R^2 = 3.1\%$). With exception of the "Other" category, holdings changes of insurance companies are least well captured ($R^2 = 1.3\%$). The latter could partially be a consequence of fewer observations with trading activity as evidenced in Table 4.1, suggesting that insurance companies tend to leave their positions untouched more often than other investor types. Similar reasoning could explain the equally low fit performance for the pension fund category and the "Other" category that includes, e.g., university endowments and foundations. It could equally be the case that pension fund / endowments holdings move in a more idiosyncratic manner detached from the systematic factors captured by IPCA3D.

4.4.3 What are the Factors Capturing?

Based on the analysis in the previous section, it is reasonable to focus on a parsimonious model with a low number of factors since increases in explanatory power become marginal after allowing for a couple of factors. However, the empirical evidence is in favour of a higher number of characteristics than used in KY being necessary to more accurately track changes in holdings. Therefore, the following analysis of the economic properties of the latent holdings factors will focus on the model with the large set of 36 characteristics and $K = 2$ latent factors. In order to aid interpretation of the factors, the factors are rotated such that the banking sector always exhibits an on average positive factor loading (see Figure 4.6). This means that a positive shock to any of the factors will on average increase bank sector stock holdings. A more detailed analysis of the resulting factor loadings is performed in Section 4.4.4.

In the following, I will provide evidence on the economics behind the recovered factors in holdings changes by relating them to proxies of economic activity, interest rates and balance sheet measures.

Macroeconomic Activity Following economic theories with heterogeneous beliefs and imperfect risk sharing, the demand for risky assets should primarily be related to the business cycle and the wider state of the economy (e.g., see Harrison and Kreps, 1978). I start my investigation in this direction by relating the factors to a well-known proxy of economic activity, i.e. the Chicago FED National Activity Index (CFNAI).[9] Figure 4.4 plots the first holdings factor versus the CFNAI measure. The figure shows that the first-factor is clearly pro-cyclical. A regression of the first factor onto the CFNAI measure in Table 4.10 confirms that the positive correlation is statistically significant.

Next, I estimate multiple regressions of the estimated factors onto principal components of a large panel of macroeconomic variables, i.e. the well known FRED data set (McCracken and Ng, 2016). Specifically, following Ludvigson and Ng (2009) and Stock and Watson (2012), I estimate principal components from a subset of the quarterly FRED data set that consists of 132 individual time series.[10]

[9]The Chicago FED National Activity Index is obtained from `https://www.chicagofed.org/publications/cfnai/index`. Accessed 11th February 2020.

[10]The series used are listed in the FRED-QD documentation appendix (column "SW FACTORS") provided on Michael McCracken's website `https://research.stlouisfed.org/econ/mccracken/fred-databases/`. Accessed 11th February 2020.

Ludvigson and Ng (2009) find that the first few principal components of this macro data set can be linked to distinct economic concepts. They dub the first factor the "real factor" due to its heavy loadings on measures of employment, production, capacity utilisation and manufacturing orders, while exhibiting only low correlation with financial variables. The second factor is strongly associated with interest rate spreads. The third and fourth factors exhibit very little correlation with the aforementioned concepts, and instead are strongly associated with measures of inflation and (commodity / consumer) prices. Table 4.7 displays the results of the regression of the holdings change factors on the first four principal components of the FRED data. The first factor exhibits strongly significant covariation with the first principal component (PC) of the FRED data. The first FRED PC is low in recessions and high in expansions. Hence, the positive sign in the regression confirms the previous evidence on the pro-cyclical behaviour of the first latent factor. In addition to the covariation with the first FRED PC, I find a significant relation of the factor with the third and fourth PC that pertain to concepts such as inflation and price pressure. When regressing the second factor onto the same set of macroeconomic principal components, I find no significant relationship with the first FRED PC, but a strongly significant association with the fourth PC. Furthermore, comparing the regression R^2s between the regression of the first and second factor in holdings changes, it is evident that while a sizeable part of the variation in the first factor is explained through measures of the macroeconomy, the second factor is clearly less well explained through the first four FRED data principal components.

For a more granular study of the macroeconomic drivers of holdings changes, I construct principal components individually for groups of macroeconomic variables. The grouping of the variables follows the FRED-QD appendix (see Footnote 10). Table 4.8 displays the results of univariate regression of the holdings factors on each of the first principal components extracted from the 14 macro groups.

Consistent with the full sample FRED principal components analysed earlier, I find that the first holdings factor exhibits strongly positive associations with variable groups pertaining to the real economy including industrial production, employment, inventories, orders & sales, as well as earnings & productivity. However, a strongly significant covariation with measures of the stock market (R^2 of 12.4%) stands out. Drilling down into the individual variables in the stock market category, I find positive and significant correlations with stock market indices including S&P 500 and Nikkei 225 (untabulated).

Interest & Exchange Rates It is to be expected that the demand for equity correlates with the demand for safer assets. In the spirit of "risk on - risk off" episodes, the demand for equity versus fixed income should primarily be related to the expected rates of return on both types of assets. Therefore, it is natural to ask how the holdings factors relate to measures of interest rates. Panel A in Table 4.9 details results of univariate regressions of the holdings factors onto a set of common interest rate measures. The three interest rate measures I focus on are the three-month T-Bill in excess of the risk free rate, the three-month commercial paper rate in excess of the risk free rate and the yield spread on Moody's BAA rated 10-year corporate bonds minus 10-year treasury bills. Across all measures I find a negative correlation with the first and second holdings factor. This means that, in general, increases in interest rates will lead to decreased equity demand. The relation is particularly strong for the commercial paper rate that is a relevant interest rate in the context of corporate short-term funding. The explanatory power of interest rates for asset demand is slightly higher for the first latent holdings factor with R^2 ranging from 6.9% to 10.4%, while R^2 range from 1.9% to 10.3% for the second holdings factor.

Following from earlier evidence in Table 4.8, the second holdings factor is related to foreign exchange rates. I provide more evidence in this direction in Panel B of Table 4.9. In univariate regressions on log-changes in exchange rate measures, I find that equity demand is positively correlated with USD/EUR and USD/GBP exchange rate pairs. An explanation for this finding could be that a weaker U.S. dollar, incentivises more foreign investor demand for U.S. stocks. Conversely, a negative correlation with changes in the trade-weighted U.S. index backs up this explanation.

Initial Public Offerings A large literature has demonstrated cyclical patterns in the market for initial public offerings (IPOs) (e.g., see Ritter and Welch, 2002; Loughran and Ritter, 2002; Pastor and Veronesi, 2005). IPOs are commonly timed to coincide with times of positive economic outlook and therefore exhibit a pro-cyclicality. Hence, given the cyclical properties of the first factor, I ask whether asset demand is potentially connected to the timing of IPOs. To that end, I correlate the holdings factors with two measures of IPO volume:[11]. Gross IPO volume which includes penny-stocks, units, closed-end funds, etc., and net IPO which volume excludes those kinds of offerings. The results are detailed in Panel A of Table 4.10. Consistent with the cyclical patterns established above, I find significant positive correlations of the first holdings factor with measures of IPO volume.

Financial Constraints Large literatures focusing on both households (Atif and Sufi, 2009; Mian and Sufi, 2011; Mian et al., 2013) and intermediaries (Adrian and Shin, 2010; Adrian et al., 2014; He and Krishnamurthy, 2013; He et al., 2017) emphasise the relevance of financial constraints, especially leverage, for different economic agents. Hence, it is likely that, in addition to changes in the perceived riskiness of assets, the risk-taking capacity of households and businesses plays an important role in explaining the demand for risky assets. I find clear evidence consistent with this idea. Table 4.8 shows that the second holdings factor exhibits a strongly positive association with measures of household and non-household balance sheets (R^2 of 14.2% and 14.6%, respectively). More specifically, Table 4.10 (Panel B) shows strong covariation of the second holdings factor with measures of household liabilities (R^2 of 11.6%) and business sector assets (R^2 of 11.1%). Therefore, the second holdings factor appears to be related to the capitalisation of institutions and households.

In summary, the evidence suggests that the first holdings factor is closely related to the state of the macroeconomy and interest rates, while the second holdings factor relates to financial conditions including leverage constraints of households as well as exchange rates. These findings align with extant theories of asset pricing such as Campbell et al. (2003), which think about asset choice in the presence of aggregate state variables: the state of the business cycle (inflation), interest rates, and dividend expectations.

4.4.4 How are Different Investors Exposed to the Factors?

In addition to interpreting the economic nature of the factors in holdings changes, it is insightful to study the exposure of different investor types to these factors. As a first pass, I assess average factor loadings for each institutional type. To that end, I

[11]IPO Volume is obtained from Jay Ritter's website `https://site.warrington.ufl.edu/ritter/ipo-data/`. Accessed 12th November 2019.

compute averages of the conditional loadings, $\beta_{i,t}$, in Equation (4.3) across all assets for each investors, yielding a time series of average factor loadings for each investor type. In Figure 4.6, I plot the time series averages of those factor loadings as well as their time series standard deviation. Recall that the factors are rotated to ensure an average positive loading of "Banks" on each of the factors. I now describe the results for each factor in turn.

For the first factor, I find interesting heterogeneity in loadings across institutional types: investment advisors (i.e. hedge funds) and pension funds display on average negative loadings, while banks (by construction) and mutual funds exhibit on average positive loadings. Given that the first holdings factor is broadly pro-cyclical, these results suggest that investment advisors and pension funds act in a counter-cyclical manner, i.e. they sell risky assets after high returns and buy after low returns. An explanation for this finding is that hedge funds and pension funds often are deep-pocketed, and hence these investors have the ability to withstand immediate selling pressure when the aggregate market declines. On the contrary, bank and mutual fund trading behaviour appears to be pro-cyclical. This in line with stricter investment mandates for banks and mutual funds that force them to hold portfolios that are on average closer to market weights, and therefore suffer more in market downturns. Overall, these findings on the cyclicality of investment behaviour across different institutional types are consistent with Timmer (2018).

Loadings on the second factor clearly separate the small investor sector from the institutional sectors. This evidence is consistent with the earlier study of the conditional buying / selling behaviour of small investors presented in Table 4.3 and discussed in Section 4.3.

For a more detailed glimpse into the investor loadings, I next turn to analysing the characteristics loadings matrices $\Gamma_{\beta,i}$. Figures 4.7 and 4.8 plot the entries of the loadings matrix separately for each factor and investor type.

Starting with the first holdings factor (Figure 4.7), one characteristic stands out in particular for banks and investment advisors: leverage. Bank holdings have a positive coefficient on leverage, while investment advisors have a negative coefficient on leverage. Given the notion that non-financial firm leverage is pro-cyclical, this finding provides an explanation for the pro-cyclicality of the first factor demonstrated above. A positive shock to the first factor (i.e. during an economic expansion) leads to a relative increase of holdings in higher leverage firms for banks, while the opposite is true for investment advisors. Given the inverse relation between leverage and assets-to-market-equity, the negative (positive) coefficient for banks (investment advisors) on assets-to-market-equity confirms the remarks regarding leverage. For the other investor types, in particular mutual funds and small investors, the number of characteristics that appear relevant is relatively larger, therefore making an interpretation based on leverage alone challenging. Following leverage related characteristics in magnitude is profitability (`prof`) which presents itself as the third largest characteristic loading in absolute value for banks, investment advisors and mutual funds, while being the fourth to fifth largest for small investors. Other interesting loadings are a positive coefficient on momentum (`r12_2`) for small investors while insurance firms and mutual funds exhibit a negative loading on momentum.

Turning to the second factor (Figure 4.8), I find the sales-to-assets (`sat`, i.e. asset turnover) and operating leverage (`ol`) characteristics to stand out for small investors and mutual funds. While small investors exhibit a positive loading on operating leverage and negative loading on asset turnover, investment advisors and mutual fund holdings exhibit a negative loading on operating leverage and positive loading on asset turnover. For ease of interpretation, it is insightful to think about these two characteristics, i.e. operating leverage and asset turnover, in conjunction. The ratio of asset turnover to operating leverage yields a measure of operating efficiency, that is, the ratio of sales to operating expenses (OPEX). In general, a firm with a higher ratio of sales to OPEX, will be regarded as more efficient. Therefore, while investment advisors and mutual funds exhibit positive coefficients on this measure operating efficiency, small investors are negatively correlated with it.

4.4.5 Which Characteristics Matter Most?

In addition to studying the mapping between characteristics and factors, it is relevant to know which characteristics contribute most to overall model fit, that is, across all components of the model. Characteristics enter the model through the loadings matrices $\tilde{\Gamma}_\alpha$ and $\tilde{\Gamma}_\beta$ (see, Eq. (4.3)). Denote by $\tilde{\Gamma}$ the matrix which horizontally stacks $\tilde{\Gamma}_\beta$ and $\tilde{\Gamma}_\alpha$ (in that order). Then the entries pertaining to the characteristic l-th in $z_{j,t}$ are in the rows $l + l \cdot i$, $i = 0, \ldots, I$ of $\tilde{\Gamma}$. I partition the matrix $\tilde{\Gamma}$ as follows

$$\tilde{\Gamma} = [\gamma_{1,1}; \ldots; \gamma_{1,L}; \gamma_{2,1}; \ldots; \gamma_{2,L}; \ldots; \gamma_{I,L}].$$

The hypotheses I test are

$$\begin{aligned}\mathbb{H}_0 : \tilde{\Gamma} =& [\gamma_{1,1}; \ldots; \gamma_{1,l-1}; \mathbf{0}_{(K+1)\times 1}; \gamma_{1,l+1}; \ldots; \\ & \gamma_{2,l-1}; \mathbf{0}_{(K+1)\times 1}; \gamma_{2,l+1}; \ldots; \gamma_{I,l-1}; \mathbf{0}_{(K+1)\times 1}; \gamma_{I,l+1}; \ldots; \gamma_{I,L}]\end{aligned}$$

versus

$$\mathbb{H}_1 : \tilde{\Gamma} = [\gamma_{1,1}; \ldots; \gamma_{1,L}; \gamma_{2,1}; \ldots; \gamma_{2,L}; \ldots; \gamma_{I,L}].$$

That is, the null hypothesis assumes that the l-th characteristic has no impact on the model and therefore cannot impact any of the factors / the intercept. Hypothesis testing relies on a residual bootstrap as outlined in Kelly et al. (2019).

The earlier analysis above demonstrated that including additional characteristics in addition to those used in Koijen and Yogo (2019a) notably improved model fit. The bootstrap exercise helps in getting a better understanding for the characteristics that matter most for model fit. Figure 4.5 reports the results from the bootstrap exercise.

Out of the 36 characteristics used, only 8 are significant at the 1% level. These characteristics are total assets, turnover, closeness to 52-week highest price, profitability, momentum (12-2), idiosyncratic volatility and unexplained volume. This set of characteristics is markedly different from the smaller set of six characteristics used in Koijen and Yogo (2019a). In terms of their contribution to model fit, characteristics based on past returns stand out. They make up three out of the top five variables in the importance ranking in Figure 4.5. The importance of past returns is in line with an earlier literature including Grinblatt et al. (1995); Nofsinger and Sias (1996); Sias (2004) that investigates institutional herding and demonstrates that institutional investors trade on momentum more strongly than individual investors. In addition to past returns, liquidity characteristics such as spread and turnover feature high up in the importance ranking. The importance of liquidity related measures is consistent with studies such as Gompers and Metrick (2001), who find that institutional investors have a particular demand for more liquid assets in order to satisfy their institutional constraints.

4.5 Holdings Factors & Asset Prices

A tight link exists between equilibrium demands and prices in asset pricing models. While the previous section has investigated the economic fundamentals behind the common drivers of demand, I now turn to investigating the relation of the components of demand identified by IPCA to asset prices. Firstly, I study if the factor structure in holdings relates to the factor structure in returns as captured in popular risk factor models and in IPCA itself. Secondly, I study the time series relations between asset prices and demand components in a two-part exercise. First, I provide additional evidence on the cyclical patterns in investor behaviour by relating demand components to past returns, and second, I demonstrate the predictive power of demand components for future returns.

4.5.1 Factors in Holdings vs. Factors in Returns

As motivated in the introduction, the factors in holdings changes could be related to changes in risk. Therefore, it is insightful to ask whether the commonly found factor structure in returns corresponds to the factor structure in holdings. This line of inquiry relates to Barber et al. (2016) who show that mutual fund flows of sophisticated investors covary with common risk factors. In order to investigate this aspect I draw on two sets of risk factors.

Firstly, IPCA factors from returns present a natural benchmark for the IPCA factors from holdings. To that end, I estimate an IPCA model on asset returns following Kelly et al. (2019). Since holdings data is only observed at quarterly frequency in this paper, I re-sample the corresponding returns at quarterly frequency to estimate the IPCA factors. In line with Kelly et al., I find that increases in total / predictive R^2 level off for about five factors. At quarterly frequency, the restricted IPCA model with five factors explains 32.3% of the cross-sectional variation in returns.

I run regressions of the holdings factors onto the set of returns factors. Table 4.11 details the result. In general, I find that the explanatory power of the factors in returns for factors in holdings is small with adjusted R^2 measures of 7.6% and 6.44% for the first and second holdings factor, respectively.[12]

[12] For the study of the regression coefficients it is necessary to understand what the returns factors capture. Since the nature of the IPCA returns factors is potentially different from the findings in

Secondly, I investigate how the factors in holdings relate to popular risk factors studied in the asset pricing literature (Fama and French, 2015; Carhart, 1997; Pastor and Stambaugh, 2003). Table 4.12 reports results from univariate regressions of the two holdings factors onto individual risk factors. For the first factor, no coefficients are significant at the 1% and 5% levels, and only borderline significance at the 10% level is found for the regression on the value factor and short-term reversal. For the second factor, a significant coefficient is found only for the size factor and the Pastor and Stambaugh (2003) liquidity factor, however, the overall explanatory power in these regressions is low with R^2s of 2.7% an 5.5%, respectively.

Overall, the evidence from these regressions suggests that the set of factors in holdings has a rather small contemporaneous overlap with factors in returns, both of fundamental and statistical nature. This finding is important for the following reason. It suggests that the set of factors explaining returns and the set of factors explaining changes in portfolio holdings are mostly distinct. Further, it means that while the broad patterns in portfolio choice can be explained along well-understood economic dimensions (business cycle & financial frictions), they do not necessarily correspond to the risk compensations captured in the common factor models. This result further suggests that while (mutual) fund flows correspond to risk factors (cf. Barber et al., 2016), there is only weak evidence that aggregate sector level changes in portfolio holdings correspond to common risk factors.

4.5.2 Time Series Evidence

In this section, I investigate time series relations between (components of) changes in holdings and returns. In the first part of the section, I establish a link between past returns and changes in holdings. In the second part, I demonstrate the predictive relations between specific components of demand and future returns over longer horizons.

Kelly et al. following the change from monthly to quarterly frequency, I compute correlations with common risk factors and identify them as follows. The first factor has a 52% correlation with the market as well as a 40% correlation with the Fama-French size factor. The second factor has a -56% correlation with 12-2 momentum. The third factor has a 45% correlation with the Fama-French value factor. The fourth factor has a -50% correlation with value and a -44% correlation with short-term reversal. Finally, the fifth factor exhibits smaller correlations with a number of different factors and is therefore challenging to interpret clearly.

Past Returns as Drivers of Investor Demand

It is a well-known fact that past performance is an important driver of future flows into mutual funds and, more generally, the demand for assets (Grinblatt et al., 1995; Nofsinger and Sias, 1996; Wermers, 1999; Sias, 2004; Timmer, 2018). The decomposition of holdings changes via IPCA3D provides a novel take on this idea by giving a more detailed understanding for which components of holdings changes are driven by past returns. To that end, I estimate the IPCA3D model using an expanding window starting in 1990:Q1 up to 2018:Q1.[13] From the estimated model, I recover the intercept, systematic part and residual as follows:

$$\underbrace{\Delta h_{i,t}}_{\text{Realised}} = \underbrace{\underbrace{\hat{\alpha}_{i,t-1}}_{\text{Intercept}} + \underbrace{\hat{\beta}_{i,t-1}\hat{f}_t}_{\text{Syst. Part}}}_{\text{Fitted}} + \underbrace{\hat{\epsilon}_{i,t}}_{\text{Residual}} .$$

Then, separately for each sector of investors i, I estimate a panel regression of the following form

$$\widehat{\Delta h}_{i,j,t}^{(comp)} = \gamma_i \, Return_{j,t-1} + b_j + b_t + \epsilon_{j,t}, \tag{4.8}$$

where $comp \in \{\text{Realised}, \text{Fitted}, \text{Intercept}, \text{Syst. Part}, \text{Residual}\}$ is a component of the decomposed holdings changes. Table 4.13 reports the results from the panel regressions.

The first finding from these regressions is that the sector-specific cyclicalities established earlier using the loadings on the recovered factors are broadly consistent with the way in which different investors react to past returns. Column "Realised" of Table 4.13 reports how the overall sector holdings change in response to past returns. Banks, investment advisors, mutual funds and insurance companies expand their stock holdings after recent positive returns, suggesting that their holdings are overall pro-cyclical. In contrast, pension funds and small investor holdings respond counter-cyclically to past returns. The finding that the small investor category (which includes many households) reacts inversely to past returns is in line with Odean (1998) who finds that individual investors tend to stick to losing investments too long and sell winning investments too early. With the exception of the investment

[13] The use of an expanding window in this exercise limits the look-ahead bias. However, note that the IPCA3D model is estimated using information from time t, while returns are observed at time $t-1$.

advisor sector, these findings confirm the cyclicalities established in Section 4.4.4 and Figure 4.6 (Panel (a)). Furthermore, this finding broadly agrees with Timmer (2018), who investigates sector level holdings in response to past returns, albeit in a fixed income setting. In addition, I find that the small investor sector appears to exhibit properties of a deep-pocketed long-horizon investor, that is, sector-level holdings increase after recent negative returns. This is also in line with the idea that smaller investors can tilt their portfolios away from market weights more strongly than larger institutions which are constrained by their institutional mandates (Koijen and Yogo, 2019a).

Secondly, the panel regressions of decomposed holdings changes on past returns in Table 4.13 yield a better understanding for *how* investors respond to past returns: via characteristics alone (Intercept), through systematic factors in holdings, or both. For most sectors the response of holdings to past returns is mainly driven by characteristics on their own, i.e. the intercept term. Across almost all sectors, the regressions of the intercept on past returns attract the highest R^2 measures, with past returns explaining between around one fifth to one third of the variation in the part of the holdings change that is unrelated to common factors. In contrast, past returns only explain between about 0.1% to roughly 8% of the variation in systematic holdings changes. This result provides further evidence on the nature of the cyclicality of the factors established earlier in Section 4.4.3: while past returns matter to some extent (see column "Syst. Part"), the bulk of the impact of past returns on changes in holdings is captured in the intercept part of IPCA3D. Therefore, the cyclicality in the first latent factor discussed earlier, is primarily related to macroeconomic channels such as industrial production / output, employment, and productivity, etc., rather than past returns.

In this context, it is important to recall that this finding is not simply a mechanical result from past returns entering the set of characteristics in IPCA3D. Ex ante, a large number of characteristics is allowed to explain changes in holdings. IPCA3D then "picks" those characteristics that are most relevant in explaining the dependent variable.

Demand Components & Future Price Changes

In this section I investigate how the different components of changes in holdings relate to future asset returns. Literature has demonstrated that future returns respond differently to investment behaviour of different sectors (see, e.g., Timmer, 2018). The decomposition of investment behaviour using IPCA3D therefore allows to ask a more granular question: *which parts* of demand from *which investors* relate to future prices?

As a first step, I present results for regressions that use the total changes in holdings, $\Delta h_{i,j,t}$ to predict future prices. Table 4.14 summarises the results from the following panel regressions that are estimated separately for each investor sector i and each horizon k

$$Return_{j,t+k} = \gamma_i \, \Delta h_{i,j,t} + b_t + \epsilon_{j,t+k}, \tag{4.9}$$

where $Return_{j,t+k}$ is stock j's cumulative excess log-return over the period t to $t+k$. The results suggest that over the next quarter the prices of stocks that have been bought by banks and insurance companies fall, while the prices of stocks that have been bought by investment advisors increase. No immediate impact on prices of stocks bought by the mutual fund and pension fund sectors can be found. For the stocks bought by investment advisors the positive relation to future returns holds for horizons up to one year, after which coefficients become insignificant. One interesting result obtains for the small investor sector. Although no impact on prices in the immediately following period is noticeable, at longer horizons stocks bought by small investors experience positive returns, supporting the idea that the small investor sector exhibits properties of a long-horizon investor.

When considering these results it is important to recall that the relation between sector-level holdings changes and prices is likely different from the impact that certain groups of funds within each sector have on prices. Furthermore, the regressions do not reflect rebalancing of portfolios in the future. Nevertheless, the regressions provide an idea for the direction in which prices move following the purchases / sales by certain sectors.

Building on this preliminary evidence, I now use the changes in holdings as decomposed by IPCA3D to shed some light on which "parts" of changes in holdings are potentially predictive of future prices. Table 4.15 summarises the results from panel regressions estimated separately for each investor sector and horizon as follows

$$Return_{j,t+k} = \gamma_1 \widehat{\Delta h}_{i,j,t}^{\text{(Intercept)}} + \gamma_2 \widehat{\Delta h}_{i,j,t}^{\text{(Syst. Part)}} + \gamma_3 \widehat{\Delta h}_{i,j,t}^{\text{(Residual)}} + b_t + \epsilon_{j,t+k}. \quad (4.10)$$

The regression includes a time-fixed effect to capture market wide events. For ease of economic interpretation, I divide the right-hand-side variables by their pooled standard deviations.

The previous set of regressions of future returns on total changes in holdings revealed that bank sector holdings changes have a negative relation to future returns. The regressions (4.10) reveal that over the next quarter this relation is primarily driven by the part of the holdings change that is systematic. A one standard deviation sized systematic holdings change is associated with a negative log-return of 1.14% over the next quarter. For the mutual funds sector, the above analysis using total changes in holdings demonstrated no significant relation between their sector-level holdings changes and future returns. However, decomposing mutual fund sector holdings changes in IPCA3D reveals that the total change masks an interesting, more granular effect: while the characteristics-based component of demand positively predicts future returns, I find that the systematic part of the holdings change is negatively associated with future returns. The latter finding is consistent with the idea of price pressures created by correlated institutional demand (Coval and Stafford, 2007), i.e. prices become dislocated in response to a systematic shock to investor holdings, hence leading to a subsequent reversal. In contrast, the positive return of assets bought by investment advisors demonstrated above appears to be unrelated to systematic demand shocks. Rather, the purely characteristics-based change in holdings of investment advisors as well as their residual demand component appear to predict future prices. This effect persists for up to two quarters ahead. The persistence beyond the first quarter is particularly important from a practical perspective, since 13F holdings are released with a delay, hence, potentially rendering the first quarter ahead return unattainable in practice. Finally, while the total change in holdings of small investors was not found to be predictive of future returns, I find that decomposed changes in holdings of the small investor sector are: a one standard deviation systematic shock to their holdings is associated with a

1.43% log-return over the next quarter. This effect is partially counterbalanced by a negative return from the characteristics-based demand component.

4.6 Conclusion

I study changes in sector-level equity holdings through the lens of a factor model with time-varying loadings. Using an extended version of instrumented principal components analysis dubbed "IPCA3D", I recover aggregate factors (shocks) common to all investors as well as investor-specific exposure to the aforementioned factors. I find that two factors capture the bulk of systematic variation in holdings changes. The recovered factors relate to the state of the real economy (business cycle) and investors' financial conditions. Sector-specific loadings allow me to distinguish the effects of demand shocks on particular investor groups. I find evidence for banks and mutual funds partially acting in a pro-cyclical manner, while investment advisors (i.e. hedge funds) and pension funds partially act counter-cyclical. Further, I find that the set of asset characteristics relevant for explaining changes in holdings is likely wider than those entertained in common risk factor models. Finally, I investigate asset pricing effects of changes in holdings as decomposed by IPCA3D. I find evidence consistent with a return reversal induced by institutional price pressures in response to systematic shocks to holdings as well as evidence for market-timing ability of investment advisors that is unrelated to common asset characteristics.

Overall, this paper provides evidence supporting the existence of a low-dimensional factor structure in sector-level holdings changes. In doing so, the paper connects the literatures that have studied behaviour of specific, single investor types, such as mutual funds or individual investors, in isolation with each other. The findings contribute to our understanding of asset exchange patterns across different institutional types. A promising direction for future research lies in extending the modelling approach presented in this paper to allow for investor specific characteristics, e.g. their investment style, fee structure, etc., to enter the conditional IPCA loadings, in addition to the asset characteristics incorporated already.

Figures & Tables

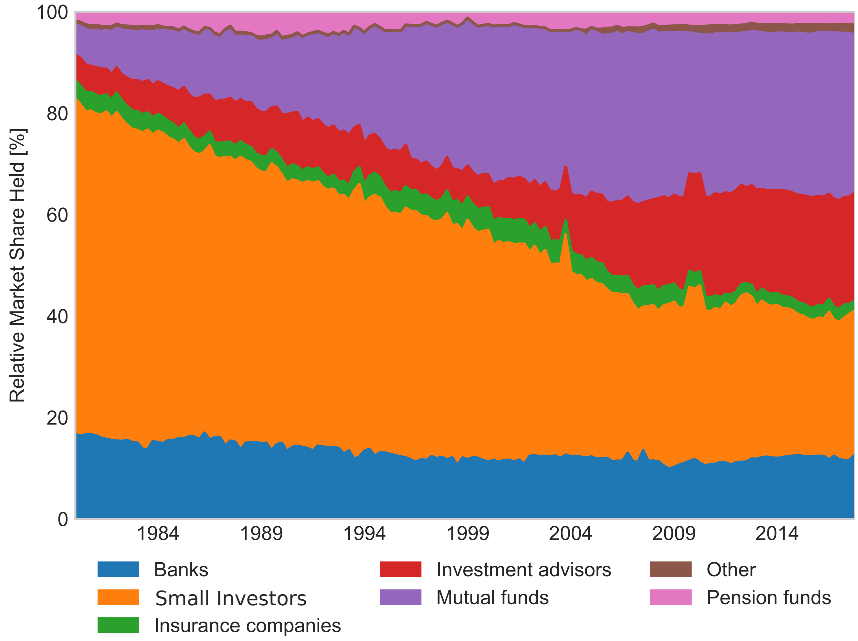
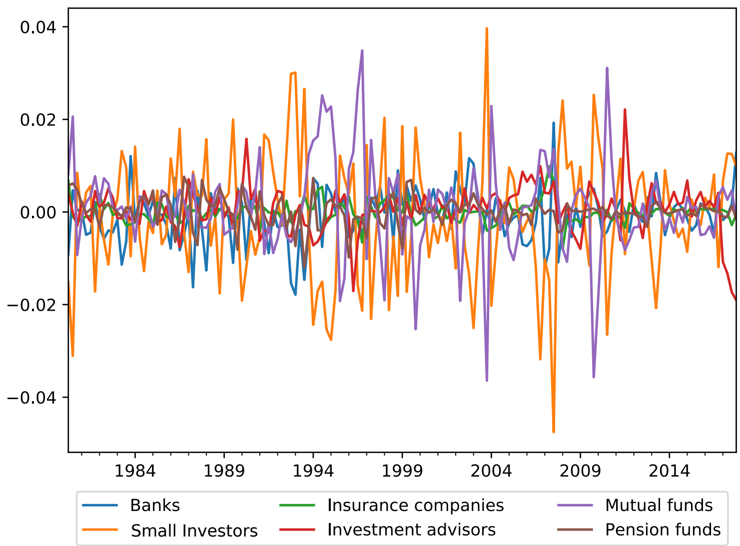
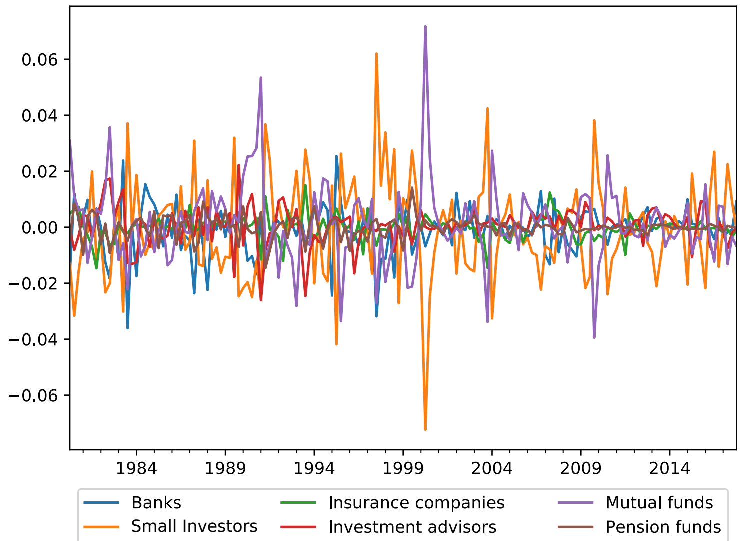


Figure 4.1: Sample Composition by Market Share. This figure shows sample composition for the 13F institutions studied in this paper. For each institutional type (banks, mutual funds, etc.) the figure shows the share of the total outstanding market capitalisation held by that sector.

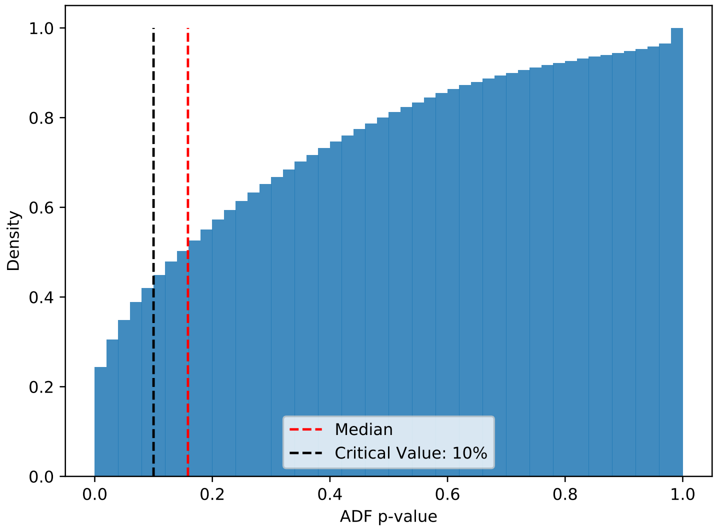


(a) Holdings Changes: IBM Co.

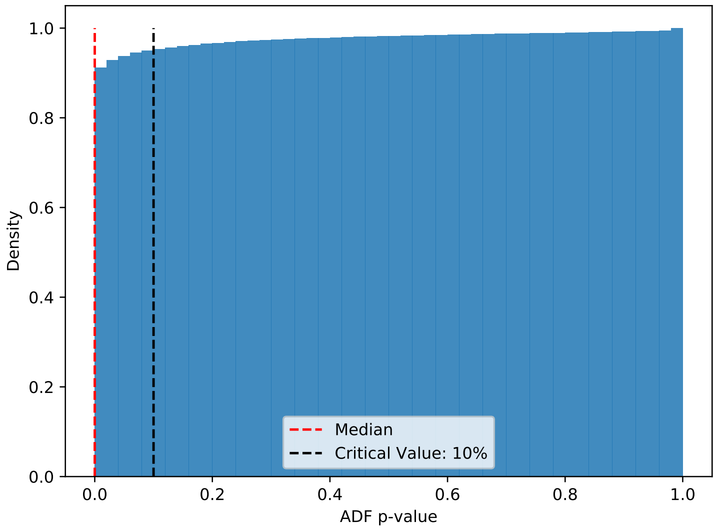


(b) Holdings Changes: Pfizer Inc.

Figure 4.2: Example of Holdings Changes for IBM & Pfizer. This figure shows the changes in holdings computed as in Equation (4.7) by investor type / sector for two exemplary cases, i.e. the IBM and Pfizer stocks. The holdings changes are constructed as quarter on quarter change in the number of shares held by a given sector divided by the number of shares outstanding. For ease of exposition, the figure does not report the holdings changes for the investor type “Other”. The “Small Investor” sector is composed of all residual holdings by non-13F insitutions.

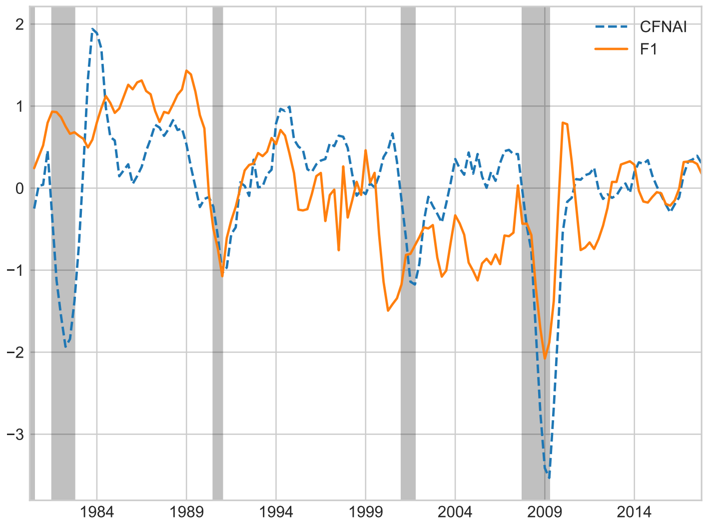


(a) Cumulative Distribution of ADF p-values - Level



(b) Cumulative Distribution of ADF p-values - First Difference

Figure 4.3: Augmented Dickey-Fuller Test of Portfolio Weights. This figure shows the cumulative distribution of p-values from an Augmented Dickey and Fuller (1979) (ADF) test performed individually for each investor-by-asset time series of holdings. The ADF test null hypothesis is that a unit root is present in the time series of portfolio weights $w_{i,j,t}$. The test is performed using a constant and no trend. The number of ADF lags employed is determined by means of an Akaike Information Criterion (AIC). Panel (A) shows the cumulative distribution of p-values for the null hypothesis that a unit root is present when the *level* of the relative holdings $w_{i,j,t}$ is used, i.e. shares held over number of shares outstanding. Panel (B) shows the cumulative distribtuion of p-values when the first difference of the portfolio weight $\Delta w_{i,j,t} = w_{i,j,t} - w_{i,j,t-1}$ is used. In both panels, the red dashed line indicates the location of the distribution median and the black dashed line indictes the 10% critical value threshold.



(a) Factor 1

Figure 4.4: Cyclical Properties of First Demand Factor. This figure plots the first factor of holdings changes (solid, orange line) against a business cycle proxy (dashed, blue line), i.e. the Chicago FED National Activity Index (CFNAI). The IPCA3D factors are rotated such that the bank sector exhibits an on average positive loading on each factor. The correlation measure is robust to autocorrelation and heteroskedasticity. Vertical, grey bands indicate NBER recession periods.

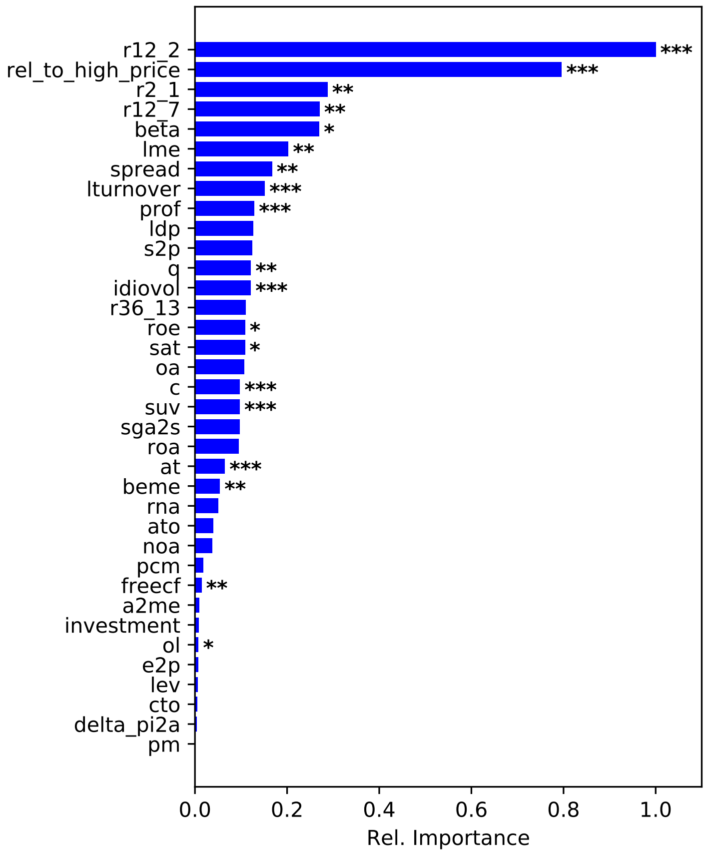
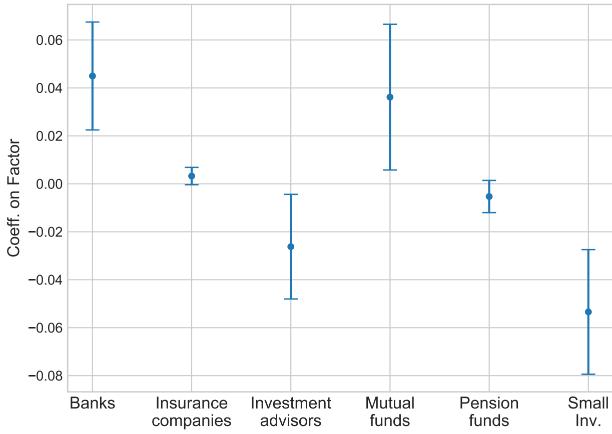
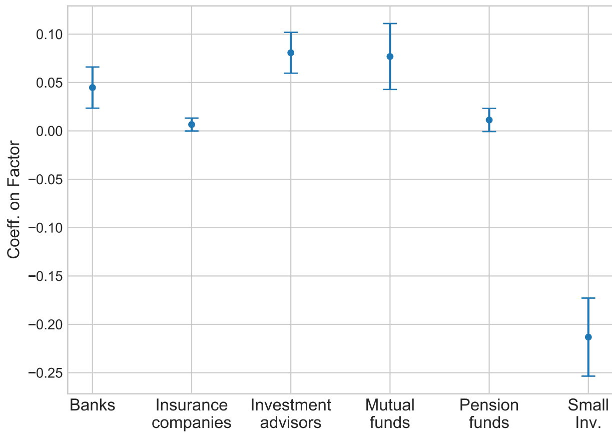


Figure 4.5: Characteristics Importance. This figure shows the contribution of the 36 characteristics to model performance. The performance contribution is measured by the change in total R^2 for the full model containing all characteristics versus a model that drops one characteristic each from its set of characteristics. For ease of interpretation, the change in model performance is rescaled to the range 0 to 1. The characteristic with the biggest contribution to model fit attains a value of 1, the characteristic with the smallest contribution attains a value of zero, i.e. I focus on the ranking of the characteristics relative to each other. In addition, the figure reports the significance of the characteristics contribution from the bootstrap exercise outlined in Section 4.4.5. The significance thresholds are: $p < 0.1$ - *, $p < 0.05$ - **, and $p < 0.01$ - ***.

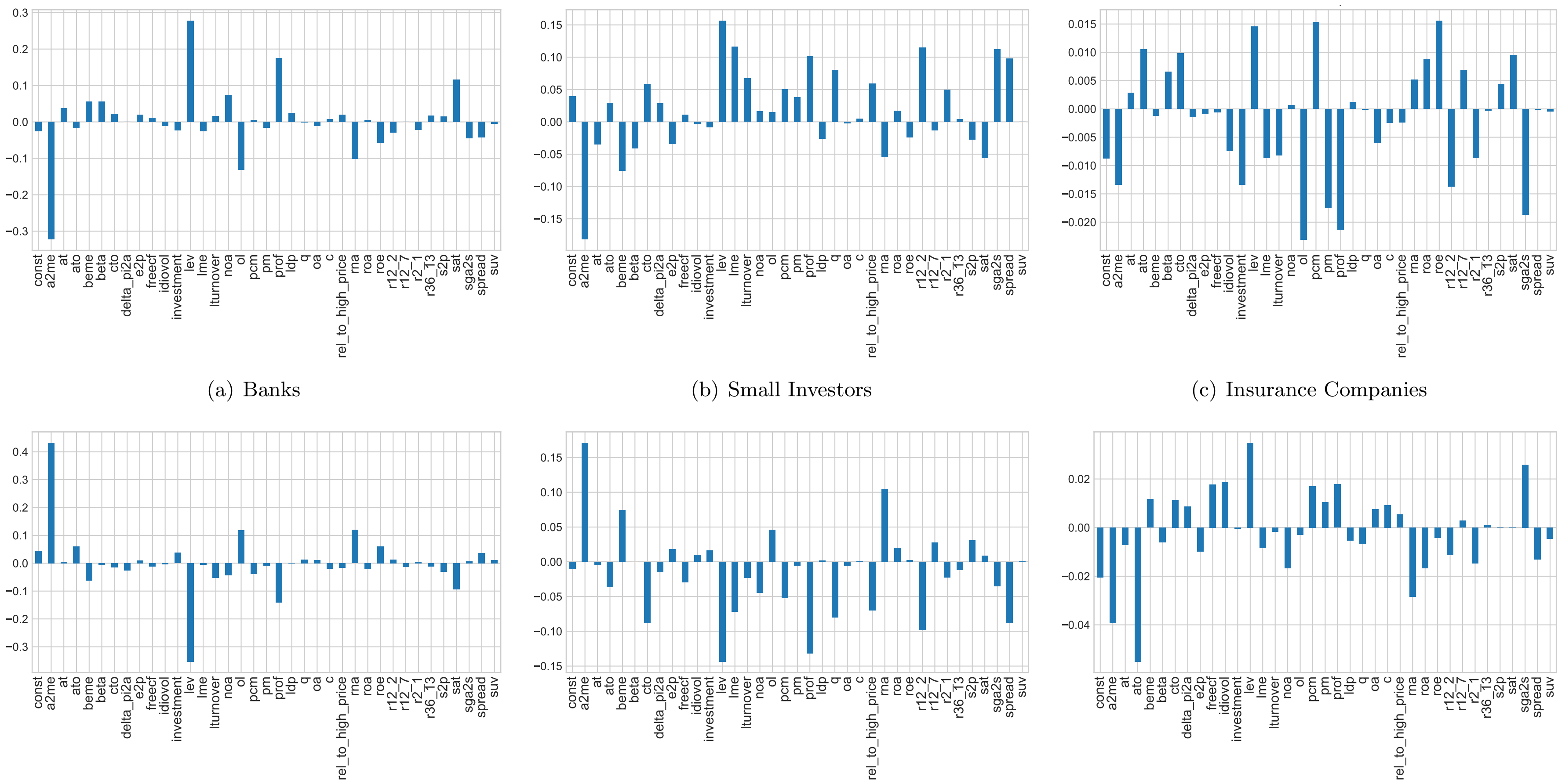


(a) Factor 1



(b) Factor 2

Figure 4.6: Average Factor Loadings by Sector. This figure shows the average factor loadings on each of the latent demand factors by investor type / sector. The investor specific factor loading $\beta_{i,t}$ in the IPCA model equation (4.3) is an $N_t \times K$ dimensional object, where N_t is the number of investable assets at time t and K is the number of factors. Let $\bar{\beta}_{i,t}$ be the factor loading of investor type i averaged across all assets, i.e. $\bar{\beta}_{i,t} = \sum_{j=1}^{N_t} \beta_{i,j,t}/N_t$. The plotted value is the time series mean of $\bar{\beta}_{i,t}$ computed separately for each factor. The error bars report one standard deviation of the time series variation of $\bar{\beta}_{i,t}$. The IPCA factors are ordered by their time series variance from largest to smallest. In order to aid interpretation, the factors are rotated such that the average factor loading of "Banks" is always positive (see Figure 4.6).



(a) Banks

(b) Small Investors

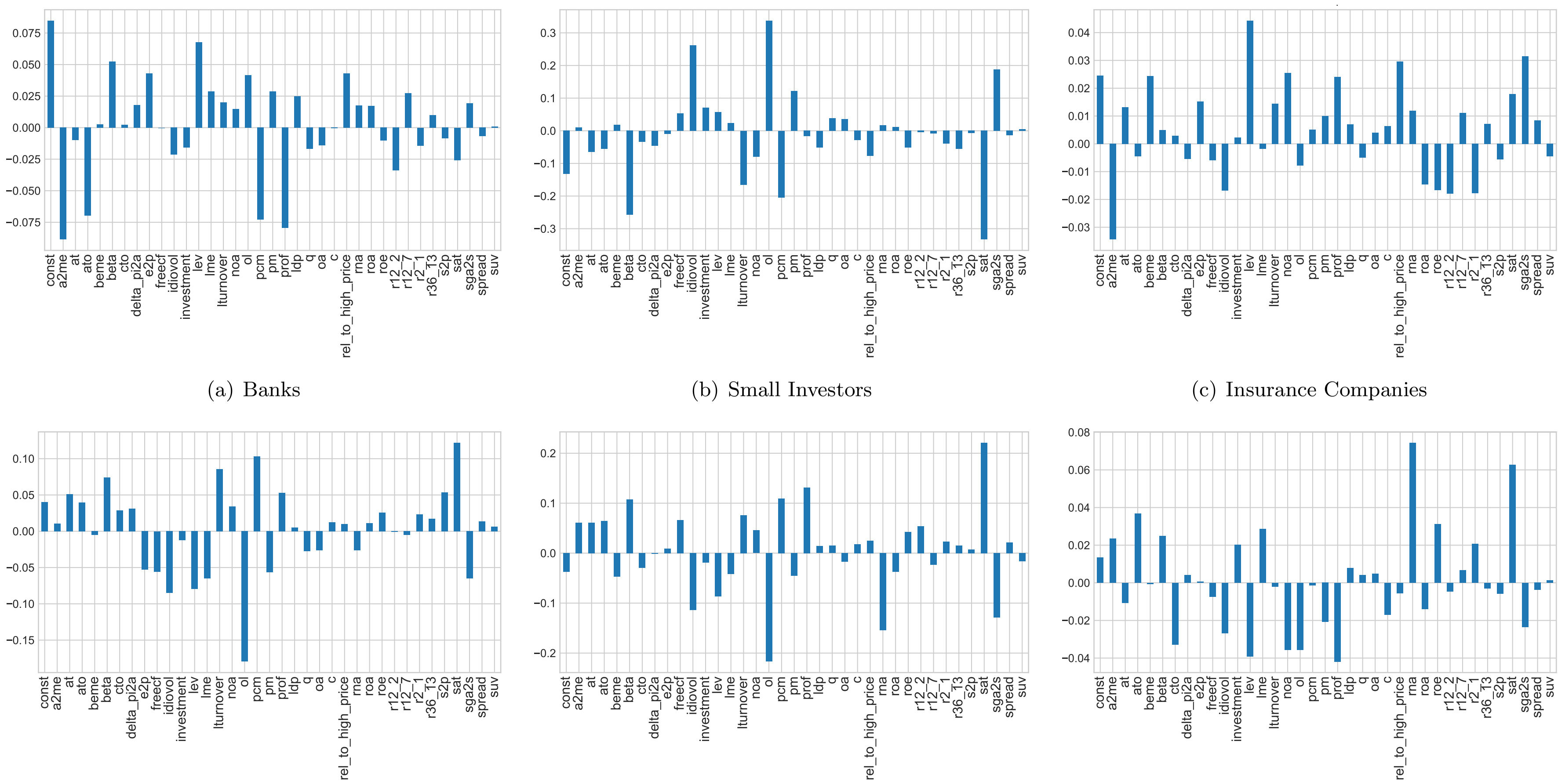
(c) Insurance Companies

(d) Investment Advisors

(e) Mutual Funds

(f) Pension Funds

Figure 4.7: Characteristics Loadings, $\Gamma_{\beta,i}$, by Sector - Factor 1. This figure shows the estimated characteristics loadings mapping separately for each investor. That is, each panel plots the $\Gamma_{\beta,i}$ corresponding to the first factor for one of the investor types. A description of the characteristics is provided in Section 4.3.1. The IPCA factors are ordered by their time series variance from largest to smallest. In order to aid interpretation, the factors are rotated such that the factor loading of "Banks" is always positive.



(a) Banks

(b) Small Investors

(c) Insurance Companies

(d) Investment Advisors

(e) Mutual Funds

(f) Pension Funds

Figure 4.8: Characteristics Loadings, $\Gamma_{\beta,i}$, by Sector - Factor 2. This figure shows the estimated characteristics loadings mapping separately for each investor. That is, each panel plots the $\Gamma_{\beta,i}$ corresponding to the second factor for one of the investor types. A description of the characteristics is provided in Section 4.3.1. The IPCA factors are ordered by their time series variance from largest to smallest. In order to aid interpretation, the factors are rotated such that the factor loading of "Banks" is always positive.

Table 4.1: Summary Statistics of Changes in Holdings.
This table presents summary statistics of the quarterly changes in holdings normalised by market capitalisation on a stock by stock basis. Each column corresponds to one particular investor type, i.e. Banks, Small Investors (13F residual), etc. **Panel A** presents summary statistics for stocks for which an investor type was an overall buyer in a given quarter, while **Panel B** shows summary statistics for stocks for which an investor type was an overall seller in a given quarter.

	Banks	Small Inv.	Insurance companies	Investment advisors	Mutual funds	Other	Pension funds
	Panel A: Buying Quarters						
Mean	0.65%	1.68%	0.40%	1.20%	1.38%	0.29%	0.27%
Median	0.33%	0.98%	0.16%	0.67%	0.82%	0.09%	0.12%
Std. Dev.	0.89%	1.91%	0.68%	1.43%	1.56%	0.61%	0.45%
No. Obs.	251,105	244,149	198,475	260,682	275,652	133,457	206,674
	Panel B: Selling Quarters						
Mean	-0.58%	-1.80%	-0.31%	-0.96%	-1.32%	-0.16%	-0.23%
Median	-0.24%	-1.18%	-0.08%	-0.46%	-0.67%	-0.02%	-0.07%
Std. Dev.	0.91%	1.86%	0.64%	1.33%	1.66%	0.46%	0.47%
No. Obs.	279,424	300,208	267,091	257,168	246,417	215,229	228,670

Table 4.2: Principal Components Analysis of Holdings Changes.
This table details the results from a principal components analysis (PCA) of the panel of holdings changes. The principal components are computed asset-by-asset from the holdings changes of all investors falling into a particular group, e.g., for the bank sector. The reported means are the fractions of explained variance for the first to third principal component averaged across assets. Respectively, the reported standard deviations are computed across all assets. Due to missing observations in the panel of holdings, principal components are computed using a probabilistic variant of PCA proposed in Tipping and Bishop (1999). In order to reduce the number of missing observations in the holdings panel, investors that are present in the sample for less than two years are dropped. The sample period is 1980:Q1 to 2017:Q4.

	1st PC		2nd PC		3rd PC	
	Mean	Std. Dev	Mean	Std. Dev.	Mean	Std. Dev.
Banks	60.79%	20.64%	18.63%	9.60%	8.84%	5.69%
Insurance companies	67.31%	18.85%	18.82%	10.85%	7.39%	5.88%
Investment advisors	66.08%	19.52%	18.16%	10.15%	7.52%	5.73%
Mutual funds	55.71%	21.04%	18.70%	8.75%	9.36%	5.35%
Other	78.29%	16.57%	15.10%	11.66%	4.23%	5.02%
Pension funds	58.64%	20.67%	18.90%	9.33%	9.31%	5.72%

Table 4.3: Conditional Summary Statistics.
This table presents the summary statistics of the changes in holdings when conditioning on a particular investor type buying / or selling a given stock. For example, the first panel details the average changes in holdings in single stocks for each investor type conditional on Banks buying the stock (in the left of the panel), or selling the stock (in the right of the panel). The change in single stock holdings is computed as the quarter on quarter change in the number of shares held times the quarter average price over the market capitalisation of the stock. The sample period is 1980:Q2 to 2017:Q4.

	Banks - Buying						Banks - Selling					
	Banks	SI	IC	IA	MF	PF	Banks	SI	IC	IA	MF	PF
Mean	0.65%	-0.92%	0.02%	0.15%	0.20%	0.02%	-0.58%	0.35%	-0.03%	0.09%	0.02%	-0.01%
Std. Dev.	0.33%	-0.59%	0.00%	0.00%	0.05%	0.00%	0.91%	2.50%	0.68%	1.75%	2.07%	0.48%
	Small Investors (SI) - Buying						Small Investors (SI) - Selling					
	Banks	SI	IC	IA	MF	PF	Banks	SI	IC	IA	MF	PF
Mean	-0.31%	1.68%	-0.11%	-0.46%	-0.85%	-0.05%	0.26%	-1.80%	0.08%	0.59%	0.88%	0.05%
Std. Dev.	1.07%	1.91%	0.70%	1.63%	1.91%	0.49%	1.01%	1.86%	0.67%	1.62%	1.84%	0.44%
	Insurance Companies (IC) - Buying						Insurance Companies (IC) - Selling					
	Banks	SI	IC	IA	MF	PF	Banks	SI	IC	IA	MF	PF
Mean	0.08%	-0.80%	0.40%	0.15%	0.27%	0.02%	-0.06%	0.14%	-0.31%	0.11%	-0.01%	0.00%
Std. Dev.	1.14%	2.63%	0.68%	1.77%	2.18%	0.51%	1.09%	2.65%	0.64%	1.82%	2.17%	0.49%
	Investment Advisors (IA) - Buying						Investment Advisors (IA) - Selling					
	Banks	SI	IC	IA	MF	PF	Banks	SI	IC	IA	MF	PF
Mean	0.01%	-1.05%	-0.02%	1.20%	-0.01%	0.01%	-0.01%	0.56%	0.00%	-0.96%	0.23%	0.00%
Std. Dev.	1.09%	2.48%	0.71%	1.43%	2.10%	0.47%	1.08%	2.48%	0.69%	1.33%	2.10%	0.48%
	Mutual funds (MF) - Buying						Mutual funds (MF) - Selling					
	Banks	SI	IC	IA	MF	PF	Banks	SI	IC	IA	MF	PF
Mean	0.03%	-1.30%	0.00%	0.02%	1.38%	0.01%	-0.03%	0.92%	-0.01%	0.24%	-1.32%	0.00%
Std. Dev.	1.08%	2.30%	0.69%	1.65%	1.56%	0.47%	1.09%	2.41%	0.71%	1.83%	1.66%	0.49%
	Pension funds (PF) - Buying						Pension funds (PF) - Selling					
	Banks	SI	IC	IA	MF	PF	Banks	SI	IC	IA	MF	PF
Mean	0.07%	-0.66%	0.00%	0.19%	0.21%	0.27%	-0.07%	0.11%	-0.02%	0.08%	0.01%	-0.23%
Std. Dev.	1.13%	2.64%	0.74%	1.79%	2.19%	0.45%	1.12%	2.72%	0.72%	1.86%	2.25%	0.47%

Table 4.4: IPCA Fit Performance.
This table summarises the in-sample performance of the fitted IPCA model as per Equation 4.3. **Panel A** uses the six characteristics used in Koijen and Yogo (2019a), the six characteristics are market equity, book equity, profitability, investment, dividends to book equity, and market beta. **Panel B** employs the 36 characteristics also used in Kelly et al. (2019) and described in Section 4.3.1. Each row corresponds to a different model starting with a model that only includes the intercept term, $K = 0$. Subsequent rows allow for a number of latent factors, $K = 1, ..., 5$. For each specification the table reports the R^2 measured at the individual stock level, as well as the R^2 measured at the characteristics-managed portfolio level. The sample period is 1980:Q1 to 2017:Q4.

Panel A: 6 Characteristicis

K	R^2 (Stock Level)	R^2_{adj} (Stock Level)	R^2 (Portf. Level)	R^2_{adj} (Portf. Level)
Intercept Only	0.62%	0.62%	9.38%	8.78%
1	5.20%	5.20%	78.90%	78.61%
2	6.13%	6.12%	92.34%	92.19%
3	6.53%	6.53%	96.17%	96.06%
4	6.81%	6.81%	97.05%	96.94%
5	7.06%	7.06%	98.15%	98.07%

Panel B: 36 Characteristics

K	R^2 (Stock Level)	R^2_{adj} (Stock Level)	R^2 (Portf. Level)	R^2_{adj} (Portf. Level)
Intercept Only	3.50%	3.50%	15.97%	15.41%
1	8.27%	8.26%	81.95%	81.71%
2	9.29%	9.28%	92.58%	92.43%
3	9.85%	9.84%	96.03%	95.92%
4	10.31%	10.30%	97.50%	97.41%
5	10.63%	10.62%	98.09%	98.02%

Table 4.5: IPCA Fit Performance at Stock Level - By Sector.
This table summarises the in-sample performance of the fitted IPCA model as per Equation 4.3 for individual investor groups / sectors. **Panel B** employs the 36 characteristics also used in Kelly et al. (2019), described in Section 4.3.1; **Panel A** uses the six characteristics used in Koijen and Yogo (2019a), the six characteristics are market equity, book equity, profitability, investment, dividends to book equity, and market beta. Each column corresponds to a different model with $K = 1, ..., 5$ latent factors. For each specification the table reports the R^2 measured at the individual stock level only from those observations falling into the specified investor group / sector. The sample period is 1980:Q1 to 2017:Q4.

Panel A: 6 Characteristicis

	K				
Inst. Type	1	2	3	4	5
Banks	0.6%	1.8%	3.4%	3.7%	4.8%
Small Investors	8.0%	8.3%	8.7%	9.0%	9.3%
Insurance companies	0.4%	0.5%	0.5%	0.7%	0.9%
Investment advisors	1.4%	3.2%	3.7%	3.7%	3.9%
Mutual funds	5.7%	7.1%	7.2%	7.5%	7.7%
Other	0.3%	0.5%	0.5%	0.6%	0.6%
Pension funds	0.4%	0.9%	2.0%	2.4%	2.9%

Panel B: 36 Characteristics

	K				
Inst. Type	1	2	3	4	5
Banks	1.9%	3.1%	5.1%	5.5%	5.9%
Small Investors	12.9%	13.4%	13.9%	14.5%	14.9%
Insurance companies	1.2%	1.3%	1.4%	1.6%	1.7%
Investment advisors	2.2%	3.8%	4.2%	4.6%	4.8%
Mutual funds	8.6%	10.1%	10.5%	10.9%	11.2%
Other	0.5%	0.6%	0.8%	0.9%	0.9%
Pension funds	0.8%	1.9%	2.5%	3.8%	4.5%

Table 4.6: IPCA Fit Performance at Portfolio Level - By Sector.
This table summarises the in-sample performance of the fitted IPCA model as per Equation 4.3 for individual investor groups / sectors. **Panel B** employs the 36 characteristics also used in Kelly et al. (2019), described in Section 4.3.1; **Panel A** uses the six characteristics used in Koijen and Yogo (2019a), the six characteristics are market equity, book equity, profitability, investment, dividends to book equity, and market beta. Each column corresponds to a different model with $K = 1, ..., 5$ latent factors. For each specification the table reports the R_x^2 measured at the characteristics-managed portfolio level only from those observations falling into the specified investor group / sector. The sample period is 1980:Q1 to 2017:Q4.

Panel A: 6 Characteristicis

	K				
Inst. Type	1	2	3	4	5
Banks	11.5%	36.8%	72.9%	80.2%	95.8%
Small Investors	96.6%	99.1%	99.4%	99.7%	99.8%
Insurance companies	25.5%	29.3%	27.6%	40.7%	48.6%
Investment advisors	37.0%	84.2%	96.1%	97.6%	98.6%
Mutual funds	81.2%	98.5%	99.4%	99.5%	99.6%
Other	18.3%	28.1%	28.9%	32.9%	32.1%
Pension funds	5.6%	16.5%	30.2%	38.3%	42.4%

Panel B: 36 Characteristics

	K				
Inst. Type	1	2	3	4	5
Banks	22.1%	46.9%	83.1%	87.3%	92.6%
Small Investors	97.2%	98.7%	99.3%	99.5%	99.6%
Insurance companies	39.9%	46.4%	53.2%	58.7%	61.7%
Investment advisors	45.8%	82.1%	92.1%	96.8%	97.4%
Mutual funds	83.6%	97.8%	98.2%	99.4%	99.5%
Other	25.6%	32.2%	36.4%	41.4%	44.3%
Pension funds	10.4%	30.2%	39.6%	54.1%	63.6%

Table 4.7: Covariation with FRED-QD Principal Components.
This table presents results from a regression of the first two IPCA factors onto the first four principal components of the FRED-QD data (Ludvigson and Ng, 2009; McCracken and Ng, 2016). For purposes of interpreation, all variables have been standarised using their standard deviation. The regressions are heteroskedasticity and autocorrelation robust. Standard errors are reported in parentheses. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	Factor 1	Factor 2
PC_1	0.3115***	0.2036
	(0.0577)	(0.1172)
PC_2	0.0206	-0.0235
	(0.0854)	(0.0987)
PC_3	-0.1603**	-0.0877
	(0.0773)	(0.0803)
PC_4	-0.2122***	-0.3466***
	(0.0793)	(0.0800)
N	151	151
R^2 (adj.)	25.55%	12.67%

Table 4.8: Covariation with Macro Groups.
This table shows regression results for univariate regressions of the first two IPCA factors onto the first principal components extracted from variables in each of 14 macroeconomic groups. Each row corresponds to one univariate regression of one of the factors onto the first principal components of one of the groups. The variable assignments to the macroeconomic groups follow McCracken and Ng (2016). Transformations of macroeconomic variables ensuring stationarity also follow McCracken and Ng. For purposes of interpretation, all variables have been standarised using their standard deviation. The regressions are heteroskedasticity and autocorrelation robust. The sample period is 1980:Q2 to 2017:Q4, corresponding to 151 observations in each univariate regression. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	Factor 1			Factor 2		
1st Princip. Comp.	Coeff.	t-Statistic	R^2 (adj.)	Coeff.	t-Statistic	R^2 (adj.)
NIPA	0.28**	2.49	7.1%	0.21	1.25	3.9%
Ind. Prod	0.29***	2.91	8.0%	0.22	1.35	4.3%
Employment	0.31***	2.94	9.2%	0.11	0.67	0.6%
Housing	0.28**	2.43	7.1%	0.21	1.36	3.8%
Invent., Orders, Sales	0.27**	2.41	6.4%	0.35***	2.59	11.6%
Prices	0.08	0.59	0.0%	0.19	1.29	2.9%
Earnings & Productivity	0.09	1.12	0.2%	-0.05	-0.49	-0.4%
Interest Rates	-0.08	-0.76	-0.1%	-0.05	-0.56	-0.4%
Money & Credit	0.03	0.32	-0.6%	0.03	0.26	-0.6%
Household Balance Sheets	0.19	1.41	2.9%	0.38***	3.08	14.2%
Exchange Rates	0.08	0.68	0.0%	0.26**	2.32	6.1%
Other	-0.06	-0.63	-0.3%	0.07	0.75	-0.1%
Stock Markets	0.36***	3.56	12.4%	0.16	0.91	1.8%
Non-Household Balance Sheets	-0.01	-0.07	-0.7%	0.39***	3.18	14.6%

Table 4.9: Determinants of Asset Demand: Interest & Exchange Rates.
This table presents results from univariate regression of the first demand factor on a set of interest rate measures (**Panel A**) and a set of foreign exchange rates (**Panel B**). The variables in Panel A are: the 3-month treasury rate - Fed. funds, the commercial paper rate minus 3-month Fed. funds, and the 10-year Moody's BAA corporate bond yield minus 10-year treasury. The variables in Panel B are: the trade-weighted USD exchange rate index, the USD/GBP exchange rate, and the USD/EUR exchange rate. All variables in are taken from the FRED-QD panel and are transformed as described in the FRED-QD appendix documentation (see Footnote 10). For purposes of interpretation, all variables have been standardised using their standard deviation. All regressions are heteroskedasticity and autocorrelation robust. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

Panel A: Interest Rates

	Factor 1			Factor 2		
	Coeff.	t-Statistic	R^2 (adj.)	Coeff.	t-Statistic	R^2 (adj.)
3-Month T-Bill	-0.30***	-4.26	8.1%	-0.16**	-2.03	1.9%
Comm. Paper Rate	-0.33***	-3.31	10.4%	-0.33***	-2.70	10.3%
BAA Corp Bond	-0.27**	-2.42	6.9%	-0.28*	-1.84	7.3%

Panel B: Exchange Rates

	Factor 1			Factor 2		
	Coeff.	t-Statistic	R^2 (adj.)	Coeff.	t-Statistic	R^2 (adj.)
Trade-Weighted USD	-0.06	-0.54	-0.3%	-0.21**	-1.99	3.6%
US/UK FX Rate	0.05	0.36	-0.5%	0.26**	2.09	6.2%
US/EUR FX Rate	0.14	1.47	1.3%	0.24**	2.36	5.2%

Table 4.10: Determinants of Latent Demand: Business Cycle Proxies & Sector Assets / Liabilities.
This table presents results from univariate regression of the first latent demand factor on a set of business cycle proxies (**Panel A**) and second latend demand factor on a set of proxies for sector assets / debt (**Panel B**). The variables in Panel A are: the gross and net IPO volume given as the number of IPOs per quarter, and the Chicago FED National Activity Index (CFNAI) as well as its three-month moving average (CFNAI - MA(3)). The variables in Panel B are taken from the FRED-QD panel and are transformed as described in the FRED-QD appendix documentation (see Footnote 10). The variables in Panel B are: real non-financial non-corporate business sector assets (Real Business Sector Assets, FRED MNEMONIC: TABSNNBx), real total liabilities of households and non-profits (FRED MNEMONIC: TLBSHNOx), and liabilities of households and non-profits relative to personal diposable income (FRED MNEMONIC: LIABPIx). For purposes of interpretation, all variables have been standardised using their standard deviation. All regressions are heteroskedasticity and autocorrelation robust. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

Panel A: Interest Rates & Business Cycle Proxies

	Factor 1		
	Coeff.	t-Statistic	R^2 (adj.)
Gross IPO Volume	0.38***	3.90	13.8%
Net IPO Volume	0.25**	2.20	5.5%
CFNAI	0.31***	3.00	9.2%
CFNAI - MA(3)	0.33***	3.14	10.0%

Panel B: Sector Assets / Debt

	Factor 2		
	Coeff.	t-Statistic	R^2 (adj.)
Real Business Sector Assets	0.34***	2.72	11.1%
Real Household Liabilities	0.27***	2.99	6.5%
Household Liabilities / Disposable Income	0.35***	3.15	11.6%

Table 4.11: Relation to IPCA Returns Factors.
This table details regression results from regressing the IPCA factors extracted from holdings onto a set of the five IPCA factors extracted from the underlying sample of asset returns as in Kelly et al. (2019). In order to match the sampling frequency of the returns and holdings data, returns are computed at quarterly frequency. The IPCA returns factors are obtained from a model with $K = 5$ factors consistent with Kelly et al. (2019). The return factors are ordered by their time series variance from largest to smallest. The sample period is 1980:Q2 to 2017:Q4. In parentheses, I report standard errors. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	Factor 1	Factor 2
IPCA Ret F1	-0.0040	-0.0026
	(0.0041)	(0.0028)
IPCA Ret F2	-0.0135***	0.0089
	(0.0052)	(0.0059)
IPCA Ret F3	-0.0041	-0.0165**
	(0.0095)	(0.0078)
IPCA Ret F4	-0.0385***	0.0001
	(0.0110)	(0.0092)
IPCA Ret F5	-0.0015	-0.0407***
	(0.0184)	(0.0141)
N	151	151
R^2 (adj.)	7.60%	6.44%

Table 4.12: Common Risk Factors vs. Demand Factors.
This table presents results from univariate regressions of the first and second IPCA3D demand factors onto a set of common risk factors. The risk factors are the five factors from Fama and French (2015) (market excess return, size, value, profitability, investment), the momentum factor (12-2), short term reversal (2-1), long term reversal (36-13), and the Pastor and Stambaugh (2003) liquidity factor. Returns on the risk factors are contemporaneous to the quarterly periods over which holdings changes are observed. Left- and right-hand side variables are standardized using the sample standard deviation. The regression are heteroskedasticity and autocorrelation robust. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	Factor 1			Factor 2		
Risk Factor	Coeff.	t-Statistic	R^2 (adj.)	Coeff.	t-Statistic	R^2 (adj.)
Market	0.14	1.36	1.5%	0.15	1.40	1.5%
Size	-0.09	-1.35	0.5%	0.15***	2.58	2.7%
Value	0.20*	1.82	2.9%	0.08	0.74	-0.2%
Profitability	-0.08	-0.71	-0.1%	0.01	0.04	-0.7%
Investment	0.07	0.77	-0.1%	0.04	0.52	-0.5%
Momentum	-0.04	-0.53	-0.5%	0.11	1.32	0.7%
ST Reversal	0.13*	1.66	1.7%	0.08	1.01	0.1%
LT Reversal	0.02	0.45	-0.6%	-0.03	-0.61	-0.5%
Liquidity (Pastor & Stambaugh)	0.01	0.00	-0.7%	0.25**	2.55	5.5%

Table 4.13: Past Returns & Components of Demand.
This table details results from a panel regression of components of investor demand dissected as follows

$$\Delta h_{i,t} = \underbrace{\hat{\alpha}_{i,t-1}}_{\text{Intercept}} + \underbrace{\hat{\beta}_{i,t-1}\hat{f}_t}_{\text{System. Part}} + \underbrace{\hat{\epsilon}_{i,t}}_{\text{Residual}} .$$

onto the previous quarter return ($Return_{t-1}$). The panel regression is run separately for each investor type. The regression controls for time and stock fixed effects. Standard errors are double clustered at the stock and time level. T-statistics are reported in parentheses. The dependent variable is standardised by its standard deviation for ease of interpretation. The sample period is 1990:Q1 to 2017:Q4. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	Realised	Fitted	Syst. Part.	Intercept	Residual
			Banks		
$Return_{t-1}$	0.30***	0.81***	0.08***	1.50***	0.12***
	(11.79)	(26.89)	(3.38)	(27.77)	(4.91)
R^2	0.52%	10.02%	0.14%	21.69%	0.23%
			Small Inv.		
$Return_{t-1}$	-0.73***	-1.41***	-0.42***	-2.49***	-0.23***
	(-17.59)	(-31.27)	(-9.48)	(-24.54)	(-4.89)
R^2	6.68%	26.76%	4.84%	35.50%	4.58%
			Pension funds		
$Return_{t-1}$	-0.12***	-0.05*	0.05*	-0.23***	-0.11***
	(-3.81)	(-1.77)	(1.75)	(-7.84)	(-3.40)
R^2	0.25%	0.07%	0.03%	0.87%	0.21%
			Investment advisors		
$Return_{t-1}$	0.05	0.98***	0.06***	1.77***	-0.14***
	(1.42)	(33.29)	(3.21)	(33.02)	(-4.18)
R^2	0.41%	20.71%	0.52%	33.16%	0.44%
			Mutual funds		
$Return_{t-1}$	0.65***	1.27***	0.49***	1.9***	0.29***
	(19.27)	(26.45)	(9.62)	(19.27)	(7.71)
R^2	5.03%	26.43%	7.70%	32.98%	3.60%
			Insurance companies		
$Return_{t-1}$	0.38***	1.25***	0.43***	1.53***	0.24***
	(15.04)	(18.93)	(6.99)	(24.68)	(9.61)
R^2	0.75%	15.02%	2.14%	21.83%	0.42%
Observations	112419	112419	112419	112419	112419

Table 4.14: Future Returns & Total Demand.
This table details results from a panel regression of the return over the period t to $t + k$, $k = 1, 2, 3, 4, 6, 8, 12$ quarters ahead onto the realised changes in holdings. The panel regression is run separately for each investor type. The panel regression uses time-fixed effects to capture market-wide events. Standard errors are clustered by time. T-statistics are reported in parentheses. The independent variable is standardised using its standard deviation. The sample period is 1990:Q1 to 2017:Q4. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	t+1	t+2	t+3	t+4	t+6	t+8	t+12
				Banks			
$\Delta h_{i,t}$	-0.0027***	-0.0038***	-0.0049***	-0.0063***	-0.0074***	-0.0051**	-0.0049*
	(-3.95)	(-3.65)	(-3.71)	(-4.06)	(-3.81)	(-2.21)	(-1.83)
R^2	0.02%	0.02%	0.02%	0.03%	0.03%	0.01%	0.01%
				Small Inv.			
$\Delta h_{i,t}$	-0.0001	0.0023*	0.0061***	0.0093***	0.015***	0.0136***	0.0131***
	(-0.07)	(1.85)	(3.79)	(4.89)	(5.99)	(4.64)	(3.77)
R^2	0.00%	0.01%	0.03%	0.05%	0.10%	0.07%	0.05%
				Pension Funds			
$\Delta h_{i,t}$	-0.001	-0.0027**	-0.004***	-0.0038**	-0.003	-0.0029	-0.0048
	(-1.39)	(-2.29)	(-2.75)	(-2.18)	(-1.35)	(-1.01)	(-1.47)
R^2	0.00%	0.01%	0.01%	0.01%	0.00%	0.00%	0.01%
				Investment Advisors			
$\Delta h_{i,t}$	0.0047***	0.0043***	0.0042***	0.0035**	-0.0008	0.0023	0.0004
	(6.67)	(3.68)	(2.89)	(2.01)	(-0.34)	(0.92)	(0.14)
R^2	0.05%	0.02%	0.01%	0.01%	0.00%	0.00%	0.00%
				Mutual funds			
$\Delta h_{i,t}$	-0.0002	-0.0011	-0.004**	-0.0063***	-0.0095***	-0.0084***	-0.0068**
	(-0.23)	(-0.88)	(-2.53)	(-3.34)	(-4.04)	(-3.16)	(-2.14)
R^2	0.00%	0.00%	0.01%	0.03%	0.04%	0.03%	0.01%
				Insurance companies			
$\Delta h_{i,t}$	-0.003***	-0.0051***	-0.0071***	-0.0071***	-0.0052**	-0.007***	-0.0081***
	(-4.25)	(-4.65)	(-4.99)	(-4.31)	(-2.54)	(-3.01)	(-3.06)
R^2	0.02%	0.03%	0.04%	0.04%	0.01%	0.02%	0.02%
Observations	112,433	104,953	98,577	92,845	83,104	74,636	60,661

Table 4.15: Future Returns & Decomposed Demand.
This table details results from a panel regression of the cumulative excess log return over the period t to $t+k$, $k = 1,2,3,4,6,8,12$ quarters ahead onto the decomposed changes in holdings. The decomposition is based on IPCA3D as follows

$$\Delta h_{i,t} = \underbrace{\hat{\alpha}_{i,t-1}}_{\text{Intercept}} + \underbrace{\hat{\beta}_{i,t-1}\hat{f}_t}_{\text{System. Part}} + \underbrace{\hat{\epsilon}_{i,t}}_{\text{Residual}} .$$

The IPCA3D system is estimated using an expanding window in order to avoid a look-ahead bias. The panel regression is run separately for each investor type. The panel regression uses time-fixed effects to capture market-wide events. Standard errors are clustered by time. T-statistics are reported in parentheses. The independent variables are standardised using their sample standard deviation. The sample period is 1990:Q1 to 2017:Q4. The p-value thresholds are: * - $p < 0.1$, ** - $p < 0.05$, *** - $p < 0.01$.

	t+1	t+2	t+3	t+4	t+6	t+8	t+12
				Banks			
Intercept	-0.0018*	-0.0055***	-0.0108***	-0.0136***	-0.0150***	-0.0166***	-0.0039
	(-1.65)	(-2.89)	(-4.06)	(-4.12)	(-3.47)	(-3.07)	(-0.50)
Syst. Part	-0.0114***	-0.0203***	-0.0298***	-0.0289***	-0.0381***	-0.0488***	-0.0485***
	(-6.18)	(-7.29)	(-8.57)	(-7.15)	(-7.76)	(-7.99)	(-6.42)
Residual	-0.0022***	-0.0029***	-0.0034***	-0.0046***	-0.0054***	-0.0029	-0.0033
	(-3.43)	(-2.92)	(-2.72)	(-3.14)	(-2.92)	(-1.31)	(-1.31)
R^2	0.06%	0.11%	0.18%	0.17%	0.18%	0.21%	0.13%
				Small Inv.			
Intercept	-0.0072***	-0.0079***	-0.0043**	-0.0001	0.0044	0.006	-0.0021
	(-7.94)	(-5.05)	(-2.02)	(-0.04)	(1.32)	(1.50)	(-0.39)
Syst. Part	0.0143***	0.0284***	0.0444***	0.0530***	0.0574***	0.0748***	0.0690***
	(7.18)	(9.55)	(11.93)	(11.41)	(10.36)	(11.01)	(8.09)
Residual	0.0002	0.0018	0.0042***	0.0063***	0.0109***	0.0085***	0.0094***
	(0.21)	(1.58)	(2.84)	(3.69)	(4.86)	(3.25)	(2.97)
R^2	0.14%	0.17%	0.25%	0.30%	0.31%	0.39%	0.24%
				Pension funds			
Intercept	-0.0028**	-0.0110***	-0.0220***	-0.0285***	-0.0415***	-0.0463***	-0.0444***
	(-2.29)	(-5.03)	(-7.08)	(-7.11)	(-7.28)	(-6.37)	(-4.32)
Syst. Part	-0.0049***	-0.0103***	-0.0186***	-0.0182***	-0.0163***	-0.0270***	-0.0443***
	(-3.98)	(-5.41)	(-7.50)	(-6.07)	(-4.22)	(-5.60)	(-6.91)
Residual	-0.0008	-0.0019*	-0.0026*	-0.0022	-0.0012	-0.0006	-0.0022
	(-1.07)	(-1.72)	(-1.88)	(-1.34)	(-0.56)	(-0.23)	(-0.71)
R^2	0.02%	0.07%	0.19%	0.22%	0.32%	0.34%	0.31%

Table 4.15 continued.

	t+1	t+2	t+3	t+4	t+6	t+8	t+12
				Investment advisors			
Intercept	0.0093***	0.0093***	0.0042	-0.0015	-0.0054	-0.0097*	0.0029
	(7.51)	(4.28)	(1.42)	(-0.43)	(-1.14)	(-1.70)	(0.38)
Syst. Part	-0.0021	-0.0049	-0.0104*	-0.0187***	-0.0242***	-0.0395***	-0.0118
	(-0.74)	(-1.08)	(-1.74)	(-2.68)	(-2.87)	(-3.78)	(-0.89)
Residual	0.0046***	0.0042***	0.0043***	0.0037**	-0.0004	0.0028	0.0005
	(6.67)	(3.72)	(3.02)	(2.20)	(-0.18)	(1.16)	(0.17)
R^2	0.14%	0.07%	0.03%	0.03%	0.03%	0.07%	0.01%
				Mutual funds			
Intercept	0.0098***	0.0129***	0.0104***	0.0072**	0.0028	0.0011	0.0075
	(8.83)	(6.75)	(4.04)	(2.28)	(0.71)	(0.22)	(1.07)
Syst. Part	-0.0059***	-0.0179***	-0.0350***	-0.0510***	-0.0554***	-0.0689***	-0.0568***
	(-2.61)	(-5.09)	(-7.78)	(-9.16)	(-8.24)	(-8.40)	(-5.53)
Residual	-0.0007	-0.0013	-0.0032**	-0.0045***	-0.0071***	-0.0054**	-0.0048
	(-0.97)	(-1.13)	(-2.16)	(-2.59)	(-3.29)	(-2.22)	(-1.64)
R^2	0.11%	0.13%	0.16%	0.23%	0.21%	0.25%	0.13%
				Insurance companies			
Intercept	0.0091***	0.0119***	0.0111***	0.0094***	0.0066	0.0036	0.0074
	(8.10)	(6.09)	(4.10)	(2.80)	(1.53)	(0.67)	(0.97)
Syst. Part	-0.0161***	-0.0293***	-0.0391***	-0.0374***	-0.0410***	-0.0427***	-0.0331***
	(-11.27)	(-12.57)	(-12.59)	(-10.16)	(-9.57)	(-8.45)	(-5.26)
Residual	-0.0026***	-0.0043***	-0.0057***	-0.0058***	-0.0036*	-0.0053**	-0.0070***
	(-3.82)	(-3.97)	(-4.13)	(-3.59)	(-1.78)	(-2.30)	(-2.65)
R^2	0.35%	0.51%	0.57%	0.40%	0.31%	0.27%	0.15%
Observations	112,433	104,953	98,577	92,845	83,104	74,636	60,661

Appendix A

Predictability of Order Imbalance, Market Quality and Equity Cost of Capital

A.1 A Simple Microstructure Model

We present a simple equilibrium model that guides our empirical analysis. Before describing it formally, we outline its main ingredients and discuss why these are required for our analysis. In our model, one risky security is traded by three types of traders: informed traders, liquidity traders, and a competitive, risk-averse market maker. The model is in spirit of Subrahmanyam (1991) with one substantial difference: order imbalance of liquidity traders is predictable to both informed traders and market makers. Our main goal it to establish the effect of order imbalance predictability on market liquidity and efficiency. In the model, the market maker absorbs the net demand of the other traders and sets the price such that he expects to earn zero utility conditional upon observing the net total market order imbalance. Predictability of uninformed order imbalance has two effects on liquidity. On the one hand predictable order imbalance helps the market maker to better estimate the total order imbalance and reduce costly inventory. This effect improves market liquidity. On the other hand it might exacerbate the adverse selection problem leading to deterioration of liquidity. Our model aims to answer which effect and when dominates in the market and helps us to formulate testable hypotheses.

There are three types of agents: an informed trader, uninformed / liquidity traders, and a competitive risk-averse market maker.[1] The informed trader is assumed to be risk-neutral. Market maker has risk aversion coefficient A_m and CARA utility. The agents trade in one risky security and market makers absorb the residual demand of the other types of traders. Market makers set prices such that their expected profit conditional on their signals is zero. Trading taking place at time 0 and liquidation of the security taking place at time 1. The security is liquidated at time 1 for value δ, where δ is a random variable with $\delta \sim \mathcal{N}\left(0, \sigma_\delta^2\right)$. The informed trader observes the fundamental value δ. Uninformed traders submit an order z, $z \sim \mathcal{N}\left(0, \sigma_z^2\right)$, and both the informed trader and the market maker observe a noisy signal of the uninformed order imbalance in form of $z + y$, $y \sim \mathcal{N}\left(0, \sigma_y^2\right)$. The total order imbalance is denoted by ω.

We focus on characterising the unique linear Nash equilibrium. The pricing rule of the market maker has the form

$$p = \lambda_1 \omega + \lambda_2 \left(z + y\right), \tag{A.1}$$

where λ_1 and λ_2 are measures of price impact (inverse market depth / liquidity) for the total order imbalance ω and the predictable portion of the order imbalance, respectively. The term $\lambda_2 \left(z + y\right)$ captures the partial price impact originating from the predictability of uninformed order imbalance z.

Let us denote the order of the informed trader as x. The informed trader's profit is

$$x\left(\delta - p\right) = x\left(\delta - \lambda_1 \omega - \lambda_2 \left(z + y\right)\right). \tag{A.2}$$

We conjecture the linear demand function of the informed trader in the form

$$x = \beta_1 \delta + \beta_2 \left(z + y\right). \tag{A.3}$$

This assumption is the linear extension of the original response function from Subrahmanyam (1991) to account for the impact of order imbalance predictability. Given x, the total order imbalance ω can be written as $\omega = x + z = \beta_1 \delta + \beta_2 \left(z + y\right) +$

[1] Introducing a number of competing informed traders as in Subrahmanyam (1991) does not change the results qualitatively, so we consider only a single informed trader for simplicity of exposition.

z.

The following proposition summarises the price at $t = 2$ obtained in equilibrium and the informed trader's orders, given the linear pricing rule and the publicly observed dividend signal.

Proposition 1. The price impact coefficients λ_1 and λ_2 and informed traders' demand response coefficients β_1 and β_2 are the solutions to the following system of equations:

$$\beta_1 = \frac{1}{2\lambda_1}, \tag{A.4}$$

$$\beta_2 = -\frac{\sigma_z^2}{2(\sigma_z^2 + \sigma_y^2)} - \frac{\lambda_2}{2\lambda_1}, \tag{A.5}$$

$$\lambda_1 = -\frac{\lambda_2}{\beta_2 + \sigma_z^2/(\sigma_z^2 + \sigma_y^2)} + \frac{A_m}{2} Var[\delta \mid \omega, z + y] \tag{A.6}$$

$$\lambda_2 = -\frac{\beta_1 \sigma_\delta^2 (\beta_2 + \sigma_z^2/(\sigma_z^2 + \sigma_y^2))}{\beta_1^2 \sigma_\delta^2 + \sigma_z^2 \sigma_y^2/(\sigma_z^2 + \sigma_y^2)}, \tag{A.7}$$

where

$$Var[\delta \mid \omega, z + y] = Var[\delta \mid p] = \sigma_\delta^2 \left(1 - \left(0.5\sqrt{\sigma_\delta^2 / Var[p]}\right)^2\right), \tag{A.8}$$

$$Var[p] = \lambda_1^2 \left[\beta_1^2 \sigma_\delta^2 + \beta_2^2 \left(\sigma_z^2 + \sigma_y^2\right) + \sigma_z^2 + 2\beta_2 \sigma_z^2\right] + \lambda_2^2 \left(\sigma_z^2 + \sigma_y^2\right) + 2\lambda_1 \lambda_2 \beta_2 \left(\sigma_z^2 + \sigma_y^2\right) \tag{A.9}$$

Proof of Proposition 1. The informed trader derives his utility from the profit as[2]

$$\begin{aligned} E\big[U\left(x\left(\delta - p\right)\right) \mid \delta, z + y\big] = {} & E\big[x\left(\delta - \lambda_1 \omega - \lambda_2 \left(z + y\right)\right) \mid \delta, z + y\big] \\ & - \left(\frac{A}{2}\right) Var\big[x\left(\delta - \lambda_1 \omega - \lambda_2 \left(z + y\right)\right) \mid \delta, z + y\big]. \end{aligned} \tag{A.10}$$

Maximising the expected utility with respect to x yields

$$x = \frac{\delta - \lambda_1 \frac{\sigma_z^2 (z+y)}{\sigma_z^2 + \sigma_y^2} - \lambda_2 (z + y)}{2\lambda_1 + A\lambda_1^2 \frac{\sigma_z^2 \sigma_y^2}{\sigma_z^2 + \sigma_y^2}}. \tag{A.11}$$

[2]For completeness the proof allows for a risk-aversion coefficient of the informed trader, A.

The second order condition for the maximisation is

$$\lambda_1 > \lambda_1^2 \frac{\sigma_z^2 \sigma_y^2}{\sigma_z^2 + \sigma_y^2}. \tag{A.12}$$

We obtain the unique linear Nash equilibrium by setting $x = \beta_1 \delta + \beta_2(z + y)$ in Equation (A.11) and solving for β_1 and β_2. We obtain

$$\beta_1 = \frac{1}{2\lambda_1 + A\lambda_1^2 \frac{\sigma_z^2 \sigma_y^2}{\sigma_z^2+\sigma_y^2}}, \tag{A.13}$$

$$\beta_2 = -\frac{\lambda_1 \frac{\sigma_z^2}{\sigma_z^2+\sigma_y^2} + \lambda_2}{2\lambda_1 + A\lambda_1^2 \frac{\sigma_z^2 \sigma_y^2}{\sigma_z^2+\sigma_y^2}}. \tag{A.14}$$

The market maker's objective function is

$$E\big[U_m\left(\omega(p-\delta)\right) \mid \omega, z+y\big] = E\big[\omega(p-\delta) \mid \omega, z+y\big] - \left(\frac{A_m}{2}\right) Var\big[\omega(p-\delta) \mid \omega, z+y\big]. \tag{A.15}$$

Following standard procedure we assume Bertrand competition between market makers and, thus, market makers set the prices so that their expected utility equals zero. Rearranging the previous equation we obtain an equation of bivariate regression layout

$$E\left[\delta \mid \omega, z+y\right] = \underbrace{\left(\lambda_1 - \left(\frac{A_m}{2}\right) Var\left[\delta \mid \omega, z+y\right]\right)}_{=:\gamma_1} \omega + \lambda_2(z+y). \tag{A.16}$$

Comparing this equation with the general bivariate regression design equations we obtain the coefficients as

$$\gamma_1 = \frac{Var[z+y]\, Cov[\delta,\omega] - Cov[\omega, z+y]\, Cov[z+y,\delta]}{Var[\omega]\, Var[z+y] - Cov^2[\omega, z+y]} \tag{A.17}$$

$$\lambda_2 = \frac{Var[\omega]\, Cov[\delta, z+y] - Cov[\omega, z+y]\, Cov[\omega,\delta]}{Var[\omega]\, Var[z+y] - Cov^2[\omega, z+y]}, \tag{A.18}$$

where

$$Var[z+y] = \sigma_z^2 + \sigma_y^2, \tag{A.19}$$

$$Var[\omega] = \beta_1^2\sigma_\delta^2 + \beta_2^2\left(\sigma_z^2+\sigma_y^2\right) + \sigma_z^2 + 2\beta_2\sigma_z^2, \tag{A.20}$$

$$Cov[\delta,\omega] = \beta_1\sigma_\delta^2, \tag{A.21}$$

$$Cov[\omega, z+y] = \beta_2\left(\sigma_z^2+\sigma_y^2\right) + \sigma_z^2 \tag{A.22}$$

$$Cov[z+y,\delta] = 0. \tag{A.23}$$

Finally,

$$\begin{aligned} Var[\delta \mid \omega, z+y] &= Var[\delta - \gamma_1\omega - \lambda_2(z+y)] \\ &= Var[\delta - \gamma_1\omega - \lambda_2(z+y)] = \sigma_d^2 + \gamma_1^2\sigma_\omega^2 + \lambda_2^2(\sigma_z^2+\sigma_y^2) \\ &\quad - 2\gamma_1 Cov[\delta,\omega] + 2\gamma_1\lambda_2 Cov[\omega, z+y] \end{aligned}$$

Q.E.D.

Price Efficiency We follow Subrahmanyam (1991) and compute a measure of price efficiency as

$$Q = \left(Var[\delta \mid P]\right)^{-1}, \tag{A.24}$$

where we can use the result from Equation (A.9). This definition of price efficiency can be interpreted as the uncertainty about the fundamental value conditional on observing the price. The smaller the uncertainty about the fundamental value given all available information, the greater price efficiency.

Total Price Impact One of our main variables of interest is the market quality as measured by the price impact of the total order imbalance unconditional of the predicted part of uninformed order imbalance. Writing $p = \alpha + \lambda_{total}\,\omega$, we get

$$\lambda_{total} = \frac{\text{cov}[p,\omega]}{Var[\omega]} = \frac{Cov[\lambda_1\omega + \lambda_2(z+y),\omega]}{Var[\omega]} \tag{A.25}$$

$$= \lambda_1 + \frac{\lambda_2\left(\beta_2\left(\sigma_z^2+\sigma_y^2\right)+\sigma_z^2\right)}{\beta_1^2\sigma_\delta^2 + \beta_2^2\left(\sigma_z^2+\sigma_y^2\right) + \sigma_z^2(1+2\beta_2)}. \tag{A.26}$$

Order Imbalance Predictability Predictability of the order imbalance of uninformed traders affects the optimal strategy of the informed trader who adjusts his optimal demand to extract maximum revenue. Despite this, the total order imbalance remains predictable.

We define the predictability of order imbalance in two ways. Firstly, we consider how the market quality (as measured by total price impact and market efficiency) is affected by the predictability in the uninformed order imbalance. We define the degree of uninformed order imbalance predictability as the ratio of the variance of predicted uninformed order imbalance to the total variance of the uninformed order imbalance:

$$R_u^2 = \frac{Var[\hat{z}]}{Var[z]} = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_y^2}. \tag{A.27}$$

Secondly, we consider how the market quality is affected by the predictability in the total order imbalance (resulted from the predictability of the uninformed orders). We define the degree of total order imbalance predictability as the ratio of the variance of predicted total order imbalance to the variance of the total order imbalance:

$$R^2 = \frac{Var[\hat{\omega}]}{Var[\omega]} = \frac{\left(\beta_2 + \frac{\sigma_z^2}{\sigma_z^2+\sigma_y^2}\right)^2 (\sigma_z^2 + \sigma_y^2)}{\beta_1^2\sigma_\delta^2 + \beta_2(\sigma_z^2 + \sigma_y^2) + \sigma_z^2(1 + 2\beta_2)}, \tag{A.28}$$

where we have used that $Var(\hat{\omega}) = \gamma^2(\sigma_z^2 + \sigma_y^2)$ with γ given as $\gamma = Cov[\omega, z + y]/Var[z + y]$.

Numerical solution & comparative statics The system of four Equations (A.13), (A.14), (A.17), (A.18) with respect to variables $\lambda_1, \lambda_2, \beta_1, \beta_2$ cannot be solved analytically. Instead, we present numerical solutions to the system and perform comparative statistics for a set of parameter values. The focus of this exercise is to gain an intuition for the model reaction to changes in the predictability of total and uninformed order imbalance R^2 and R_u^2 respectively.

We set our baseline scenario set of parameters as $\sigma_z^2 = 1$, $\sigma_\delta^2 = 1$ and let the noise of the uninformed order imbalance component vary according to $\sigma_y^2 = (1, \ldots, 100)$. The predictability variables R^2 and R_u^2 are then computed from the specific combination of parameters in each of the scenarios analysed. We perform our analysis for two different values of market maker's risk aversion: $A_m = 0$ and $A_m = 1$.[3]

[3]We keep the risk-aversion parameter A_m rather small since earlier empirical work by Hende-

Figure A.1 plots the relation between the two measures of predictability: R^2 and R_u^2. It reveals that for a wide range of parameter values, R^2 is monotonically increasing in R_u^2. This goes in line with a common belief that predictability of uninformed order imbalance is translated into predictability of total order imbalance despite the endogenous nature of informed traders' demand.

The optimal demand of informed traders increases with fundamental value of the asset (see Figure A.2). Moreover, the sensitivity of informed orders to fundamental value monotonically declines with predictability of order imbalance. Higher degree of order imbalance predictability makes it harder for informed traders to hide behind uninformed orders. This leads to optimal reduction of trading aggressiveness as R^2 increases. This is true regardless of the degree of market-makers' risk aversion. The response of informed demand to the predictive part of uninformed order imbalance is zero when market maker is risk neutral (Panel A of Figure A.2). Price impact of total order imbalance increases (Panel A of Figure A.3) while price response to the predicted part of uninformed order imbalance decreases with R^2 (Panel B of Figure A.3).

Figure A.4 presents numerical comparative statics of λ_{total} – price impact of total order imbalance ω unconditional of predicted uninformed order imbalance. In general the relationship between the liquidity and order imbalance predictability is non-monotonic and has an inverse U-shape. When the degree of order imbalance predictability is low, marginal increase in informed trader's profit from front-running the uninformed orders is higher than marginal decrease in profit due to revelation of private information (price impact). For higher degrees of order imbalance predictability marginal profit from front-running drops and eventually a positive effect on market quality prevails.

Figure A.4 reveals that the sensitivity of the total price impact to predictability of order imbalance increases in magnitude with σ_δ^2. This suggests that order imbalance predictability reduces the degree of asymmetric information the most where such asymmetries are the most pronounced: during periods of high uncertainty and in stocks with high volatility of the fundamental values. In Figure A.4, we also show that price impact sensitivity of the total price impact to order imbalance predictability exhibit similar patterns for different values of uninformed order

rshott and Seasholes (2009) concludes that there is little evidence for overly limited risk-bearing capacity of market makers.

imbalance volatility σ_z^2.

Similarly, Figure A.5 presents numerical comparative statics of the level and the market efficiency measure Q with respect to predictability of order imbalance R_l^2. In general efficiency increases with order imbalance predictability. The higher the predictability, the easier it is for the market maker to infer the information from the total order imbalance. Although the informed trader tends to reduce the amount of trading as predictability of order imbalance increases, the positive effect of market efficiency still dominates.

Finally, Figures A.6 and A.7 present comparative statics of λ_{uncond} and Q with respect to R_u^2. The results are similar qualitatively to those presented in Figures A.4 and A.5.

Figures and Tables

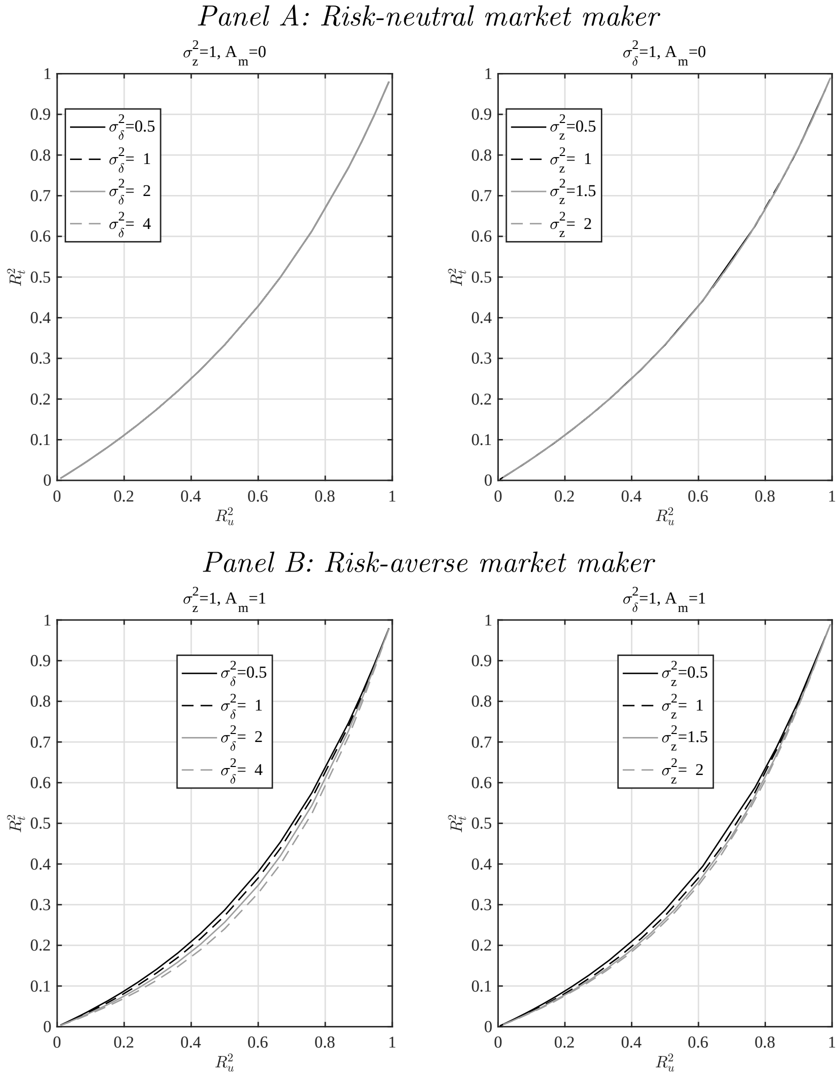


Figure A.1: Comparative Statics: R^2_{tot} vs. R^2_u.
This figure plots the values of R^2 coefficient of total order imbalance predictability as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2_u derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

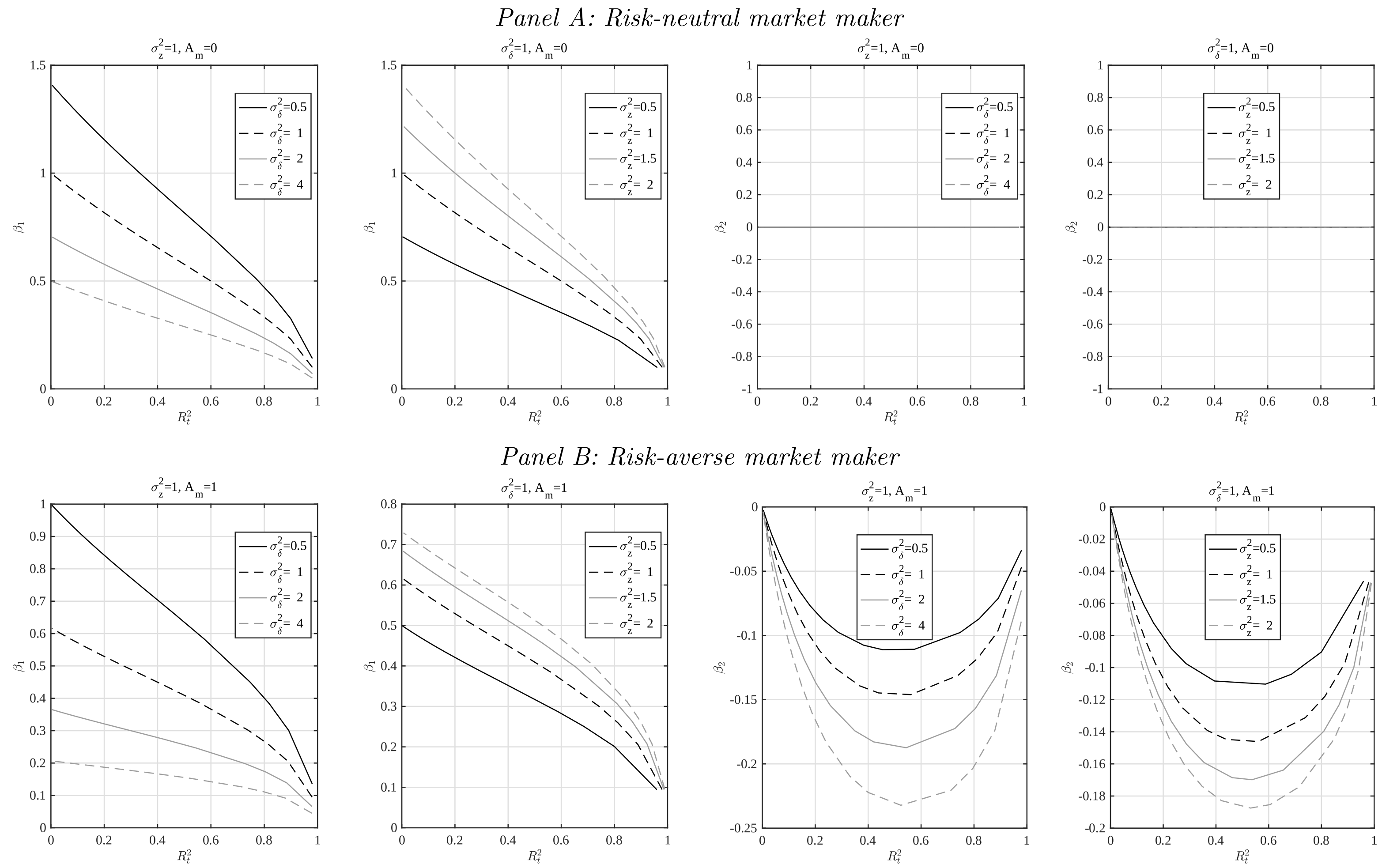
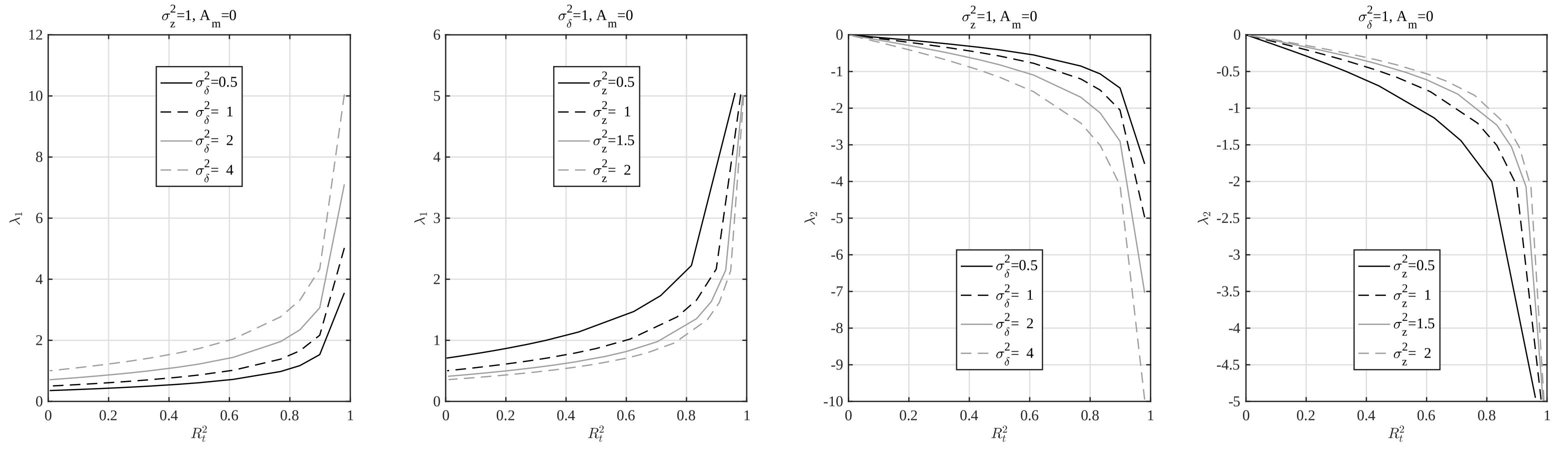


Figure A.2: Comparative Statics: β_1 and β_2 Coefficients.
This figure plots the values of parameters β_1 and β_2 determining the demand function of informed traders as functions of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

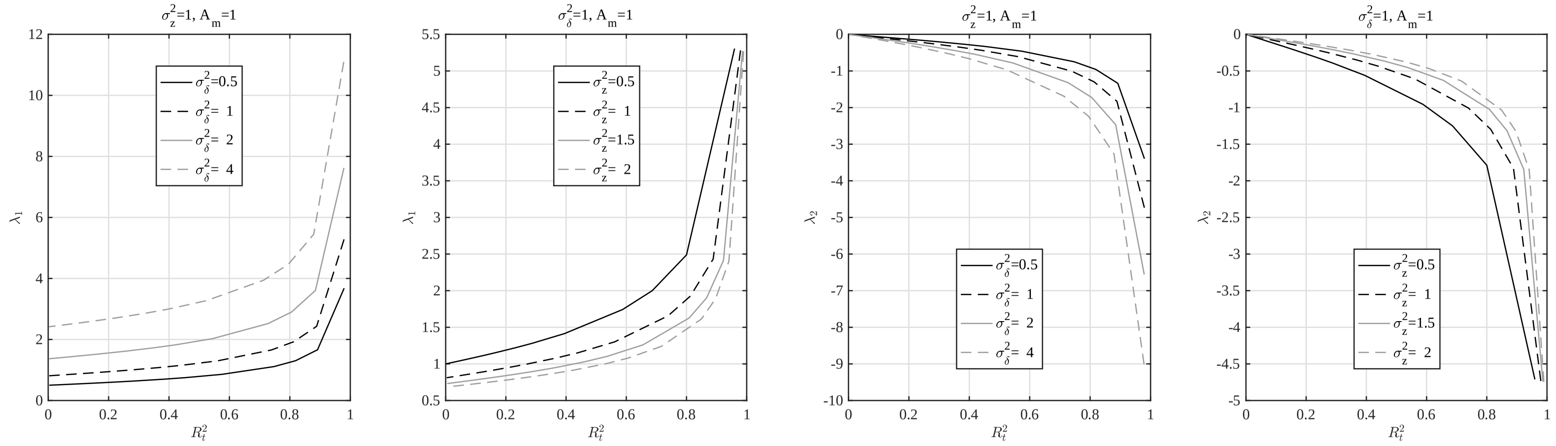
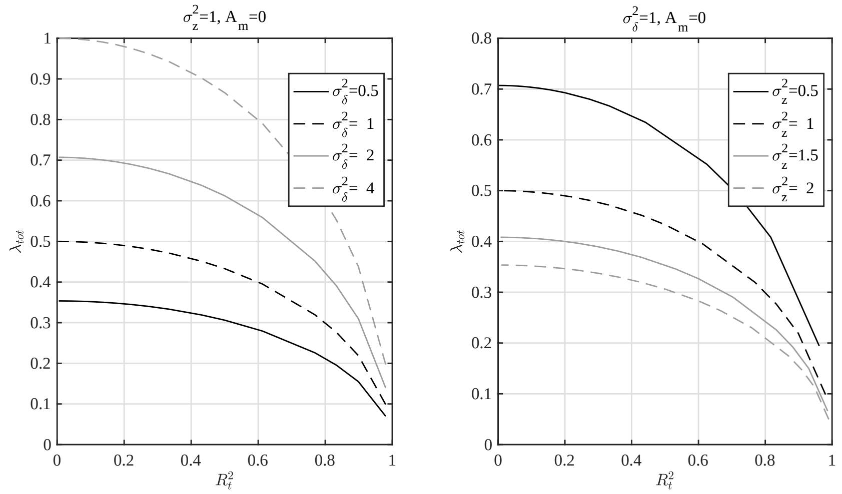


Figure A.3: Comparative Statics: λ_1 and λ_2 Coefficients.
This figure plots the values of price impact parameters λ_1 and λ_2 as functions of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

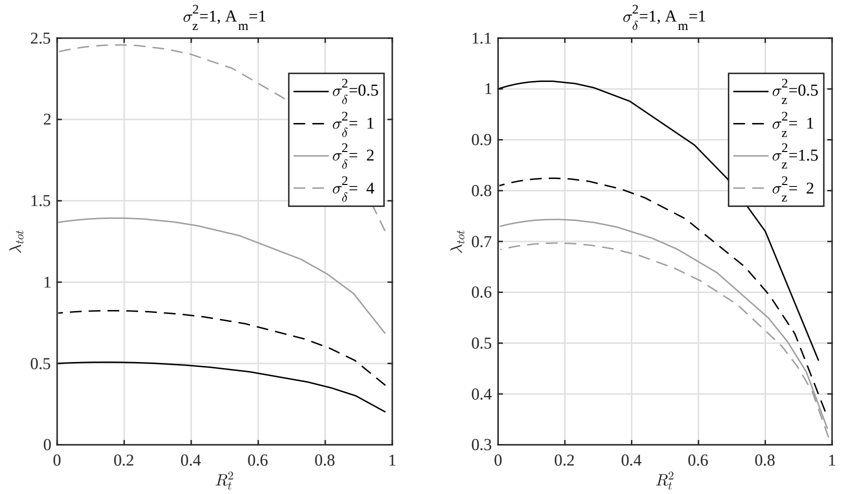
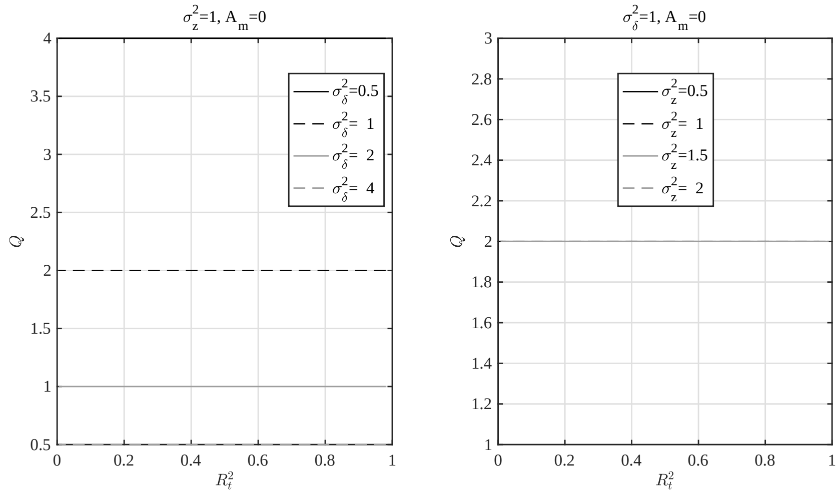


Figure A.4: Comparative Statics: Price Impact vs. Predictability of Total Order Imbalance.

This figure plots the value of total price impact λ_{uncond} as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

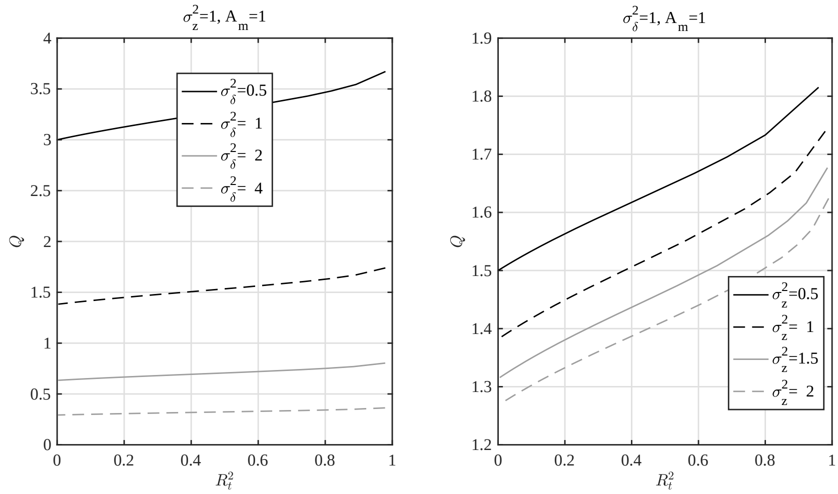
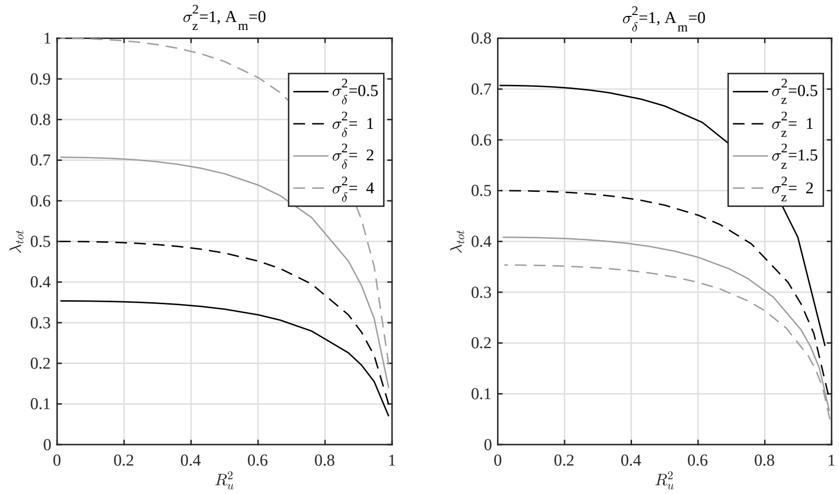


Figure A.5: Comparative Statics: Efficiency vs. Predictability of Total Order Imbalance.

This figure plots the value of price efficiency Q as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

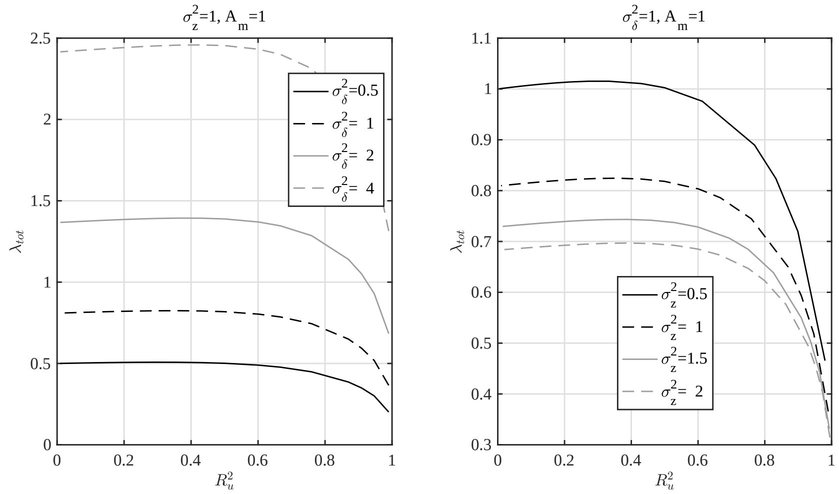
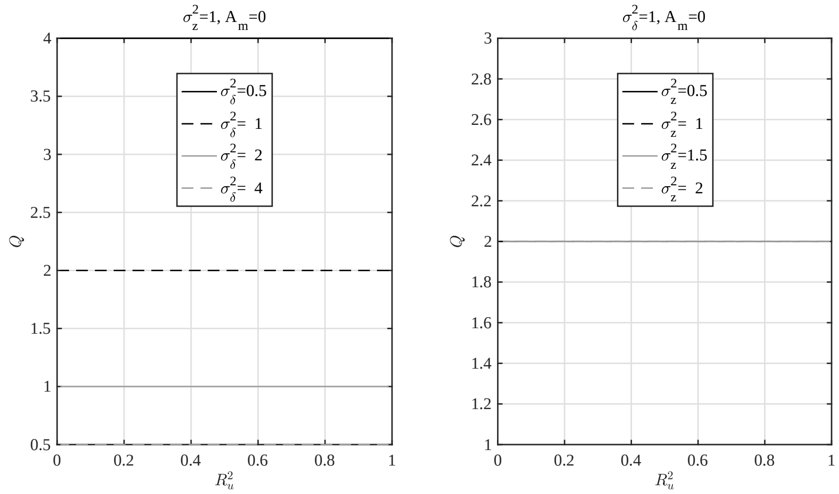


Figure A.6: Comparative Statics: Price Impact vs. Predictability of Uninformed Order Imbalance.
This figure plots the value of total price impact λ_{uncond} as a function of the predictability of total order imbalance R_u^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R_u^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Panel A: Risk-neutral market maker



Panel B: Risk-averse market maker

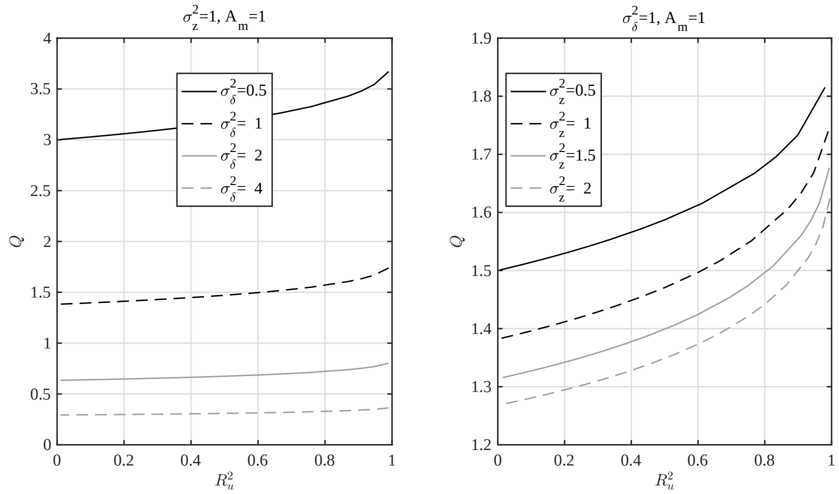


Figure A.7: Comparative Statics: Efficiency versus Predictability of Uninformed Order Imbalance.

This figure plots the value of price efficiency Q as a function of the predictability of total order imbalance R_u^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R_u^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

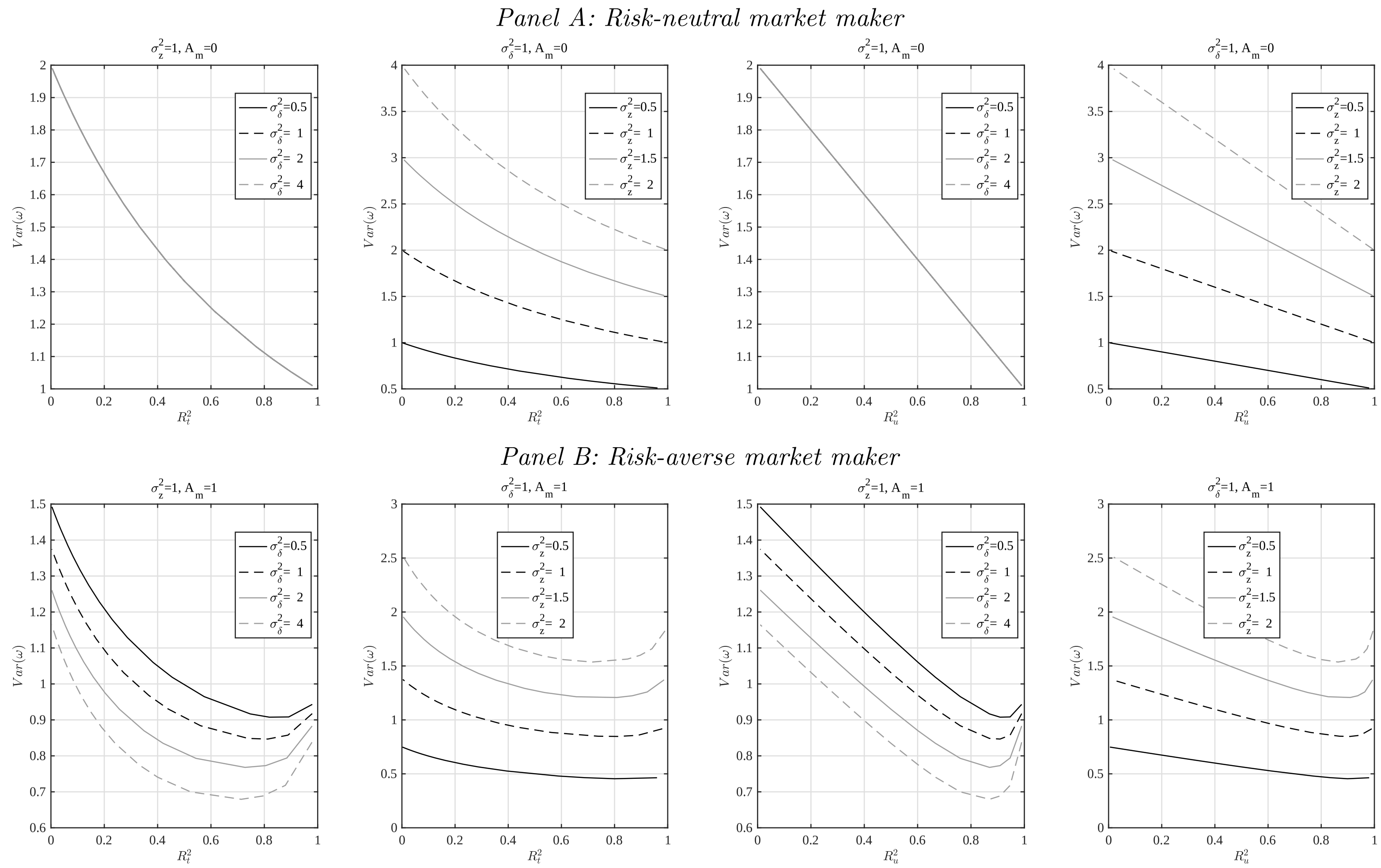


Figure A.8: Comparative Statics: $Var[w]$ **versus** R^2 **and** R_u^2**.**
This figure plots the variance of total order imbalance as a function of the predictability of total order imbalance R^2. The data points correspond to individual numerical solutions of the non-linear system (A.13)-(A.18) using the parameter set as indicated by the figure titles and legends. Simultaneously, σ_y^2 is varied in the range $0, \ldots, 100$. The dependent variable is plotted over the R^2 derived from the chosen parameter set. Panel A corresponds to the case of risk-neutral market maker ($A_m = 0$) and Panel B presents the case of risk-averse market maker.

Table A.1: Summary Statistics for the Predictors of Order Imbalance.
This table presents means and standard deviations (in squared brackets) for the predictors used to forecast order imbalance. We report the summary statistics both for the full sample and across different quintiles of market capitalisation. **Panel A:** summarise the variables for the sample of active pension funds for the period from January 1, 2006 to December 31, 2010. **Panel B:** reports the summary statistics for the full sample of stocks for the period from January 1, 1997 to December 31, 2013. Here *ttq* is the trade-to-quote ratio, *ret* is the daily stock return, *rvar* is the realised variance based on 5 minutes individual stock returns, *size* is the market capitalisation in billions of dollars, *prc* denotes the price, and *trvol* is the daily trading volume in millions of dollars. Finally, *dib* denotes the depth imbalance defined as $dib_{it} = (ask\_depth - bid\_depth)/(ask\_depth + bid\_depth)$, where *ask_depth* and *bid_depth* are daily average depth levels at the best bid and ask quotes respectively.

Panel A: Active Pension Funds

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
ttq	0.12 [0.05]	0.12 [0.05]	0.11 [0.05]	0.12 [0.04]	0.12 [0.04]	0.14 [0.05]
ret (%)	0.06 [3.09]	-0.02 [5.11]	0.05 [4.08]	0.08 [3.55]	0.08 [3.09]	0.07 [2.58]
dib (%)	1.12 [14.52]	0.11 [19.65]	0.79 [15.65]	1.31 [13.74]	1.63 [12.67]	1.05 [11.19]
rvar (%)	0.15 [0.56]	0.38 [1.12]	0.21 [0.60]	0.15 [0.51]	0.11 [0.37]	0.06 [0.20]
size (B$)	9.70 [2.06]	0.37 [0.08]	0.90 [0.18]	1.92 [0.34]	4.40 [0.78]	33.87 [5.06]
prc ($)	36.82 [9.69]	14.99 [2.57]	23.80 [4.34]	32.54 [5.62]	43.53 [7.33]	53.55 [10.82]
trvol (M$)	78.24 [46.78]	3.73 [5.76]	10.92 [10.99]	25.00 [21.99]	55.31 [43.86]	240.7 [107.7]

Panel B: Full Sample of Stocks

	Full Sample	Q1 (small)	Q2	Q3	Q4	Q5 (large)
ttq	0.38 [0.41]	0.38 [0.41]	0.38 [0.32]	0.38 [0.27]	0.37 [0.25]	0.41 [0.26]
ret (%)	0.08 [3.74]	0.05 [4.61]	0.10 [4.03]	0.10 [3.68]	0.10 [3.37]	0.09 [2.90]
dib (%)	2.32 [29.51]	0.14 [33.70]	2.42 [24.45]	4.10 [21.12]	5.31 [18.96]	5.97 [16.41]
rvar (%)	0.36 [1.36]	0.65 [1.87]	0.17 [0.55]	0.11 [0.37]	0.09 [0.38]	0.05 [0.39]
size (B$)	3.83 [1.10]	0.17 [0.07]	0.67 [0.15]	1.53 [0.30]	3.75 [0.76]	28.80 [5.34]
prc ($)	24.81 [9.38]	12.50 [4.35]	23.57 [5.17]	31.98 [6.81]	41.37 [9.43]	53.06 [15.63]
trvol (M$)	24.80 [17.37]	1.08 [2.44]	6.23 [7.52]	15.22 [16.06]	36.51 [31.82]	162.3 [80.24]

Table A.2: Aggregate Market Uncertainty and the Effect of Order Imbalance Predictability: Additional Results.
This table reports the results of the following regression

$$y_{it} = \sum_{j\in\{\text{high,med,low}\}} \beta_j \cdot pl_{i,t-1} \cdot D_t^{(j)} + \alpha_i + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where y_{it} represents either some of the additional liquidity measures, such as the effective spread (*espread*), the realised spread (*rspread*) and price impact (*prcimpact*), or the additional measure of market inefficiency defined as $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\widetilde{pl}_{it}$ which is calculated as in Eq.(1.3). **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($oib^{tot}_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	espread	*rspread*	*prcimpact*	*abs_ac*
$pl^{pf}_{t-1} \times D^{high}_t$	-0.031 (-6.57)	-0.018 (-6.55)	-0.032 (-9.92)	-0.013 (-0.74)
$pl^{pf}_{t-1} \times D^{med}_t$	-0.025 (-7.19)	-0.022 (-11.7)	-0.019 (-8.31)	-0.047 (-3.88)
$pl^{pf}_{t-1} \times D^{low}_t$	-0.010 (-1.97)	-0.015 (-3.23)	-0.004 (-1.05)	-0.076 (-4.02)
oib^{tot}_t	-0.002 (-1.47)	-0.006 (-4.05)	0.005 (4.39)	0.0001 (0.01)
$illiq_{t-1}$	0.554 (7.58)	0.209 (8.14)	0.268 (11.7)	
$lsize$	-0.011 (-4.22)	0.003 (1.33)	-0.021 (-10.5)	0.011 (2.37)
$turn$	0.118 (4.33)	-0.053 (-2.09)	0.176 (4.67)	-0.484 (-6.79)
$1/prc$	0.244 (3.78)	0.233 (6.40)	0.169 (4.17)	0.135 (5.49)
ttq	-4.032 (-4.82)	-2.156 (-3.04)	-1.475 (-2.07)	-3.596 (-1.52)
vix	0.051 (4.87)	-0.008 (-1.63)	0.091 (15.3)	-0.003 (-0.11)
No. obs.	764,825	764,825	764,825	764,825
R^2	65.76%	16.67%	32.49%	0.21%

Panel B: Full Sample of Stocks

	espread	*rspread*	*prcimpact*	*abs_ac*
$pl^{pf}_{t-1} \times D^{high}_t$	-0.410 (-15.9)	-0.429 (-8.41)	-0.580 (-18.8)	-0.054 (-4.39)
$pl^{pf}_{t-1} \times D^{med}_t$	-0.229 (-10.1)	-0.098 (-2.13)	-0.464 (-16.4)	-0.096 (-13.0)
$pl^{pf}_{t-1} \times D^{low}_t$	-0.179 (-7.19)	-0.021 (-0.42)	-0.424 (-13.9)	-0.096 (-11.6)
oib^{tot}_t	-0.073 (-25.7)	-0.086 (-5.95)	-0.006 (-0.59)	-0.006 (-11.8)
$illiq_{t-1}$	0.578 (125.1)	0.047 (22.8)	0.036 (21.2)	
$lsize$	-0.161 (-31.5)	-0.264 (-24.9)	-0.099 (-19.6)	-0.014 (-16.1)
$turn$	-1.177 (-9.74)	-2.462 (-9.00)	-0.010 (-0.17)	-0.442 (-8.35)
$1/prc$	0.606 (22.7)	1.166 (20.5)	0.367 (14.5)	0.029 (10.2)
ttq	7.241 (18.3)	16.22 (9.32)	1.378 (1.19)	-0.041 (-0.55)
vix	0.755 (16.3)	0.956 (12.8)	0.828 (21.1)	-0.001 (-0.11)
No. obs.	12,906,715	12,906,715	12,906,715	12,906,715
R^2	49.58%	1.05%	0.43%	0.51%

Table A.3: Effect of Order Imbalance Predictability on Liquidity and Market Efficiency by Stock Characteristics: Additional Results. This table presents the results of the following regression

$$y_{it} = \alpha_i + \beta pl_{i,t-1} + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where y_{it} represents either some of the additional liquidity measures, such as the effective spread (*espread*), the realised spread (*rspread*) and price impact (*prcimpact*), or the additional measure of market inefficiency defined as $abs\_ac = |(1+ac)/(1-ac)|$, where ac is the autocorrelation of daily returns. The predictive likelihood pl_{it} is defined as the 20-day rolling average over $\tilde{pl}_{it}$ which is calculated as in Eq.(1.3). For the ease of exposition, only the loading and t-statistics on pl_t are reported. The sample stocks are sorted into quintiles Q1 to Q5 based on NYSE market capitalisation breakpoints, trade-to-quote ratio, idiosyncratic volatility and passive institutional ownership. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($oib_{i,t}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	Q1	Q2	Q3	Q4	Q5
			Sorting by Market Cap		
estpread	-0.05 (-3.96)	-0.03 (-8.06)	-0.04 (-4.47)	-0.02 (-8.28)	-0.01 (-5.16)
rspread	-0.03 (-1.69)	-0.02 (-7.33)	-0.02 (-9.42)	-0.02 (-11.42)	-0.01 (-6.10)
prcimpact	-0.07 (-4.11)	-0.02 (-5.61)	-0.02(-7.77)	-0.01 (-5.57)	-0.01 (-4.13)
abs_ac	-0.11 (-2.07)	-0.02 (-0.97)	0.00 (0.17)	-0.05 (-2.56)	-0.03 (-1.35)
			Sorting by Trade-to-Quote ratio		
estpread	-0.04 (-6.09)	-0.02 (-10.30)	-0.03 (-4.80)	-0.03 (-7.58)	-0.02 (-4.71)
rspread	-0.02 (-6.44)	-0.02 (-10.39)	-0.02 (-9.34)	-0.02 (-10.11)	-0.03 (-5.21)
prcimpact	-0.03 (-6.81)	-0.02 (-7.92)	-0.02 (-7.87)	-0.02 (-8.67)	-0.02 (-3.65)
abs_ac	-0.06 (-2.43)	-0.04 (-1.98)	-0.02 (-1.32)	-0.04 (-2.02)	-0.01 (-0.24)
			Sorting by Idiosyncratic Volatility		
estpread	-0.02 (-7.33)	-0.03 (-5.94)	-0.04 (-5.43)	-0.04 (-5.95)	-0.03 (-7.40)
rspread	-0.01 (-10.07)	-0.01 (-11.47)	-0.02 (-9.10)	-0.03 (-9.99)	-0.03 (-5.50)
prcimpact	-0.01 (-5.80)	-0.02 (-8.68)	-0.03 (-9.11)	-0.03 (-8.77)	-0.03 (-6.48)
abs_ac	-0.05 (-2.22)	-0.06 (-2.53)	-0.03 (-1.35)	-0.04 (-1.69)	-0.01 (-0.013)
			Sorting by Passive Institutional Ownership		
estpread	-0.03 (-7.80)	-0.02 (-6.71)	-0.02 (-6.61)	-0.03 (-5.87)	-0.02 (-4.19)
rspread	-0.03 (-7.63)	-0.01 (-5.79)	-0.01 (-6.34)	-0.01 (-6.51)	-0.01 (-6.00)
prcimpact	-0.03 (-6.47)	-0.02 (-6.26)	-0.02 (-5.91)	-0.02 (-4.77)	-0.02 (-4.58)
abs_ac	-0.00(-0.05)	-0.04 (-1.51)	-0.05 (-1.54)	-0.05 (-2.12)	-0.03 (-1.56)

Table A.3 continued.

Panel B: Full Sample of Stocks

	Q1	Q2	Q3	Q4	Q5
			Sorting by Market Cap		
estpread	-1.25 (-16.81)	-0.40 (-9.98)	-0.15 (-5.81)	-0.14 (-5.49)	-0.14 (-5.99)
rspread	-1.61 (-10.82)	-0.40 (-5.79)	0.02 (0.35)	0.01 (0.24)	-0.05 (-1.82)
prcimpact	-1.15 (-15.07)	-0.44 (-9.47)	-0.29 (-6.19)	-0.23 (-6.25)	-0.13 (-6.24)
abs_ac	-0.27 (-16.30)	-0.12 (-7.73)	-0.07 (-5.63)	-0.05 (-4.09)	-0.010 (-0.85)
			Sorting by Trade-to-Quote ratio		
estpread	-0.80 (-16.10)	-0.30 (-12.43)	-0.28 (-11.41)	-0.38 (-13.09)	-0.41 (-9.34)
rspread	-0.78 (-10.92)	-0.15 (-3.74)	-0.34 (-4.94)	-0.39 (-5.19)	-0.47 (-4.54)
prcimpact	-0.59 (-16.52)	-0.37 (-12.68)	-0.30 (-6.35)	-0.48 (-10.08)	-0.47 (-8.19)
abs_ac	-0.09 (-5.65)	-0.07 (-7.10)	-0.09 (-9.34)	-0.10 (-8.90)	-0.12 (-8.86)
			Sorting by Idiosyncratic Volatility		
estpread	-0.15 (-12.56)	-0.26 (-15.31)	-0.39 (-14.23)	-0.70 (-16.06)	-1.54 (-16.50)
rspread	-0.26 (-4.10)	-0.16 (-2.99)	-0.36 (-6.17)	-0.77 (-9.04)	-2.16 (-12.86)
prcimpact	-0.08 (-1.75)	-0.36 (-9.01)	-0.46 (-11.86)	-0.67 (-12.82)	-1.25 (-14.98)
abs_ac	-0.06 (-5.17)	-0.05 (-5.38)	-0.06 (-6.34)	-0.09 (-8.92)	-0.23 (-14.19)
			Sorting by Passive Institutional Ownership		
estpread	-1.21 (-10.02)	-0.61 (-11.64)	-0.28 (-11.05)	-0.23 (-12.91)	-0.23 (-11.57)
rspread	-1.67 (-7.26)	-0.58 (-4.19)	-0.22 (-2.73)	-0.16 (-3.00)	-0.18 (-2.92)
prcimpact	-0.94 (-8.60)	-0.83 (-8.91)	-0.44 (-7.77)	-0.37 (-9.74)	-0.32 (-7.98)
abs_ac	-0.32 (-12.93)	-0.12 (-7.32)	-0.08 (-6.44)	-0.05 (-4.40)	-0.065 (-5.18)

Table A.4: De-trended Predictive Likelihood.
This table reports the results for the main regression analysis but with the predictive likelihood replaced by its de-trended transformation. For each given stock, we define the de-trended predictive likelihood pl_t^* as residuals from the regression of pl_t on the corresponding time index t, i.e. $pl_t = \alpha + \beta\, t + pl_t^*$. As a result, for instance, the effect of order imbalance (de-trended) predictability is investigated by the following regression

$$illiq_{it} = \alpha_i + \beta pl_{i,t-1}^* + \gamma' \boldsymbol{X}_{it} + \epsilon_{it},$$

where $illiq_{it}$ takes one of following illiquidity variables (expressed in basis points): quoted spread *qspread*, effective spread *espread*, realised spread *rspread*, and price impact *prcimpact*. We run the same regression for market inefficiency measures, such as the variance ratio $vratio_{it}$ and a measure of absolute correlation abs_ac_{it}. **Panel A:** reports the results for the order imbalance of active pension funds. **Panel B:** shows the results for the full sample of stocks. The vector of control variables $\boldsymbol{X}_{it}$ includes: the log-market cap $lsize_{it}$ of the firm i (scaled by 100), the inverse of the price $1/prc_{it}$ (multiplied by 100), the trade-to-quote ratio ttq_{it}, the share turnover $turn_{it}$ (in percent), the value of the VIX index vix_t (in percent), and the lagged value of the dependent variable ($illiq_{i,t-1}$), and the contemporaneous order imbalance ($oib_{i,t}^{tot}$). Sample period in Panel A is from January 1, 2006 to December 31, 2010 and in Panel B from January 1, 1997 to December 31, 2013. In parentheses we report the t-statistics computed based on double-clustered standard errors.

Panel A: Active Pension Funds

	qspread	*espread*	*rspread*	*prcimpact*	*vratio*	*abs_ac*
pl_{t-1}^*	-0.02 (-6.00)	-0.01 (-4.91)	-0.01 (-3.77)	-0.01 (-4.60)	-0.02 (-2.94)	-0.03 (-2.21)
$illiq_{t-1}$	0.58 (23.88)	0.56 (7.75)	0.21 (8.24)	0.27 (11.81)		
oib_t^{tot}	0.003 (2.08)	0.001 (0.06)	-0.001 (-3.16)	0.07 (5.99)	-0.001 (-0.63)	-0.001 (-0.23)
$lsize$	-0.02 (-5.79)	-0.01 (-4.22)	0.003 (1.11)	-0.02 (-10.36)	0.01 (3.72)	0.01 (2.09)
$turn$	-0.16 (-4.31)	0.06 (2.22)	-0.09 (-3.70)	0.12 (3.49)	0.22 (6.27)	-0.50 (-6.80)
$1/prc$	0.23 (4.28)	0.24 (3.76)	0.23 (6.37)	0.17 (4.17)	0.07 (3.94)	0.13 (5.45)
ttq	-10.50 (-8.35)	-1.60 (-1.70)	-0.22 (-0.30)	0.55 (0.78)	0.75 (0.70)	-1.99 (-0.76)
vix	0.09 (9.45)	0.04 (4.67)	-0.01 (-1.79)	0.07 (15.24)	0.02 (3.00)	0.04 (1.99)
No. obs.	764,825	764,825	764,825	764,825	764,825	764,825
R^2	54.67%	65.52%	16.44%	32.21%	0.30%	0.16%

Panel B: Full Sample of Stocks

	qspread	*espread*	*rspread*	*prcimpact*	*vratio*	*abs_ac*
pl_{t-1}^*	-0.43 (-15.27)	-0.42 (-15.52)	-0.65 (-10.80)	-0.43 (-12.42)	-0.09 (-18.55)	-0.17 (-16.99)
$illiq_{t-1}$	0.68 (170.8)	0.58 (125.21)	0.05 (22.87)	0.04 (21.34)		
oib_t^{tot}	-0.05 (21.44)	-0.07 (-25.16)	-0.08 (-5.90)	-0.01 (-0.28)	-0.001 (7.13)	-0.001 (-10.51)
$lsize$	-0.18 (-36.23)	-0.17 (-35.37)	-0.27 (-26.61)	-0.13 (-25.92)	-0.01 (-23.29)	-0.02 (-19.82)
$turn$	-1.61 (-9.86)	-1.27 (-9.75)	-2.45 (-9.11)	-0.21 (-3.60)	0.04 (5.61)	-0.46 (-8.38)
$1/prc$	0.53 (20.54)	0.59 (22.24)	1.15 (20.39)	0.33 (13.10)	0.01 (5.43)	0.02 (6.90)
ttq	2.27 (7.46)	7.89 (20.50)	16.61 (9.76)	2.63 (2.31)	0.07 (1.62)	0.08 (0.89)
vix	0.53 (17.11)	0.48 (18.69)	0.49 (11.46)	0.63 (26.41)	0.03 (6.92)	0.05 (3.57)
No. obs.	12,906,715	12,906,715	12,906,715	12,906,715	12,906,715	12,906,715
R^2	59.84%	49.50%	1.05%	0.41%	0.88%	0.51%

Appendix B

Bond Risk Premia With Machine Learning

B.1 A Simple Motivating Framework

Relying on the large literature on time varying risk premia (see e.g. Campbell and Cochrane, 1999, Wachter, 2006 and Buraschi and Jiltsov, 2007) we present a simple model with external habit formation which leads to the Quadratic Linear model.[1] The representative agent maximises

$$E\left[\int_0^\infty u(C_t, X_t, t)dt\right] \ ,$$

where the instantaneous utility function is given by

$$u(C_t, X_t, t) = \begin{cases} e^{-\rho t}\frac{(C_t - X_t)^{1-\gamma}}{1-\gamma} & \text{if } \gamma > 1 \\ e^{-\rho t}\log\left(C_t - X_t\right) & \text{if } \gamma = 1 \end{cases}$$

where X_t is an external habit level as in Campbell and Cochrane (1999). Consider now the Surplus Consumption Ratio

$$S_t = \frac{C_t - X_t}{C_t} \ .$$

[1]This material is based on the 2015 Version of Pietro Veronesi's lecture notes on "Topics in Dynamic Asset Pricing".

We model external habit formation as in Pastor and Veronesi (2005), i.e. we assume:

$$S_t = e^{s_t}$$
$$s_t = a_0 + a_1 z_t + a_2 z_t^2$$
$$dz_t = k_z \left(\bar{z} - z_t\right) dt + \sigma_z dW_{c,t} \ .$$

Pastor and Veronesi (2005) show that by choosing a_i appropriately (in particular, $a_2 < 0$), then $s_t < 0 \rightarrow S_t \in [0, 1]$. In addition, we must have $\frac{\partial s(y)}{\partial y} = a_1 + 2a_2 y_t > 0$ so that positive shocks to consumption $dW_{c,t}$ translate into positive shocks to the surplus consumption ratio S_t. The rest of the model is defined by:

$$dc_t = g_t dt + \sigma_c dW_{c,t}$$
$$dq_t = i_t d_t + \sigma_q dW_{q,t}$$

where we let $c_t = \log C_t$ and $q_t = \log Q_t$ be log consumption and log inflation. Finally assume that $\boldsymbol{X}_t = (g_t, i_t, z_t)'$ follows the process

$$d\boldsymbol{X}_t = K(\Theta - \boldsymbol{X}_t)dt + \Sigma d\boldsymbol{W}_t \ , \tag{B.1}$$

where $d\boldsymbol{W}_t$ is a vector of Brownian motions.
In this economy, the SDF is given by $M_t = e^{\eta t - \gamma\left(c_t + a_0 + a_1 z_t + a_2 y_t^2\right) - q_t}$ and the interest rate has a linear quadratic structure

$$r_t = \delta_0 + \gamma g_t + i_t + \delta_z z_t + \delta_{zz} z_t^2 \tag{B.2}$$

Finally, denote the zero coupon bond price by $P(\boldsymbol{X}_t, t; T)$. Given the specification of the model (B.1), the price of the zero coupon bond $P(\boldsymbol{X}_t, t; T)$ is the solution to the Partial Differential Equation (PDE)

$$rZ = \frac{\partial P}{\partial t} + \frac{\partial P}{\partial \boldsymbol{X}} K(\Theta - \boldsymbol{X}_t) + \frac{1}{2} tr \left(\frac{\partial^2 P}{\partial \boldsymbol{X} \partial \boldsymbol{X}'} \Sigma \Sigma' \right)$$

subject to the final condition $P(\boldsymbol{X}_T, T; T) = 1$. Using the method of undetermined coefficients and exploiting the risk free rate equation (B.2), we can verify that the log bond price is given by

$$\log P(\boldsymbol{X}_t, t; T) = A(t;T) + \mathbf{B}(t,T)'\mathbf{X}_t + \mathbf{X}_t'\mathbf{C}(t;T)\mathbf{X}_t$$

where $A(t;T), \mathbf{B}(t,T)$ and $\mathbf{C}(t;T)$ satisfy a set of ODEs - see Ahn et al. (2002) and Leippold and Wu (2003).[2] Hence, the bond pricing formula is also linear-quadratic with factors given by consumption growth g_t, expected inflation i_t and habit z_t.

For a fairly general framework with non-linear dynamics under the historical measure, and encompassing many equilibrium models with recursive preferences and habit formation see Le et al. (2010). Finally, note that besides habit-based term structure models, non-linearities are also featured in state-dependent, learning-based models (see, e.g. Veronesi, 2004).

[2] More precisely, we conjecture $P(\boldsymbol{X}_t, t; T) = e^{A(t;T)+\mathbf{B}(t,T)'\mathbf{X}_t+\mathbf{X}_t'\mathbf{C}(t;T)\mathbf{X}_t}$, we compute derivatives $\frac{\partial P}{\partial t}$, $\frac{\partial P}{\partial \boldsymbol{X}}$, and $\frac{\partial^2 P}{\partial \boldsymbol{X} \partial \boldsymbol{X}'}$, we substitute r and partial derivatives in the PDE, and we collect terms.

B.2 Computational Details

For our implementation of the various machine learning techniques in Python we utilise the well-known packages `Scikit-Learn`[3] and `Tensorflow`[4] in the `Keras`[5] wrapper. `Scikit-Learn` provides the functionality to estimate regression trees (both gradient boosted regression trees and random forest), partial least squares and penalised regressions (ridge, lasso, elastic net). Furthermore, we make use of numerous auxiliary functions from `Scikit-Learn` for data pre-processing such as input standardisation / scaling and train-test splits. A particularly useful `Scikit-Learn` function is `GridSearchCV`, which allows streamlined systematic investigation of neural network hyperparameters. Our neural networks are trained using `Keras` and Google's `Tensorflow`. The `Keras` wrapper provides two distinct approaches to construct neural networks, i.e. a sequential API and a functional API. The sequential API is sufficient to construct relatively simple network structures that do not require merged layers, while the functional API is used to build those networks that require merged layers as for example in the case of the exogenous addition of forward rates into the last hidden layer. `Keras` also implements a wide range of regularisation methods applied in this paper, i.e. early-stopping by cross-validation, L1 / L2 penalties, drop-out, and batch normalisation.

B.2.1 Setup

Since the forecasting exercise in this paper is iterative and since we use model averaging, the computational challenge becomes sizeable. For that reason, we perform all computations on a high performance computing cluster consisting of 84 nodes with 28 cores each, totalling to more than 2300 cores. We parallelise our code using the Python `multiprocessing`[6] package. Specifically, we parallelise our code execution at the point of model averaging such that for each forecasting step a large number of models can be estimated in parallel and averaged before moving to the next time step. Although it is common in applications such as image recognition to perform neural network training on GPUs, we refrain from doing so since the speed-up from GPU computing would be eradicated by the increased communica-

[3] `http://scikit-learn.org/stable/`, as of 26th October 2018
[4] `https://www.tensorflow.org/`, as of 26th October 2018
[5] `https://keras.io/`, as of 26th October 2018
[6] `https://docs.python.org/3.4/library/multiprocessing.html`, as of 26th October 2018

tion over-head between CPU and GPU as the computational effort of training an individual neural network is relatively small in our exercise.

B.2.2 Full Cross-Validated Neural Network vs. Group-Ensemble

In Section 2.5.6 we compare our best performing group-ensemble model (labelled as NN 1 Layer Group Ensem (1 node per group), fwd rate net (1 layer: 3 nodes)) against two different types of model averaging schemes, i.e. weighting based on the inverse of the validation loss and an equal-weighted model as well as a fully cross-validated (CV) network. While the logic of the weighting schemes for the model-averaging may be intuitive, the specifications for the fully CV network may be not. Table B.1 gives an outline of these specifications vis-a-vis the choice made for the group-ensemble structure.

Table B.1: Specifics of Full Cross-Validation vs. Group-Ensembling
This table reports the hyperparameters for the full cross-validated network vs. the shallow network with group ensemble (see Table 2.9). The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Hyperparameter Set	Full CV	CV for group ensemble
Number of Layers	1, 2	1
Nodes Per Group	1, 2, 3	1
Nodes in Fwd Rate Net	1, 3	3
Dropout - Group Net	0.1, 0.3, 0.5	0.1, 0.3, 0.5
Dropout - Fwd Rate Net	0, 0.3	0
L1/L2 Penal - Group Net	0.01, 0.001, 0.0001	0.01, 0.001
L1/L2 Penal - Fwd Rate Net	0.001, 0.0001	0.0001
Combinations per retraining	432	6
CV frequency	5 years	5 years

A number of aspects should be discussed; in the fully CV setting we recursively choose the number of hidden layers (between 1 and 2) as well as the number of nodes for each group of macroeconomic variables. Similarly, the number of nodes in the forward rate net is allowed to change whereas for the group-ensemble it is kept fixed. In addition, the full CV allows for a bit more flexibility in the L1/L2 penalty terms for both the group-specific networks as well as for the forward rates. Increasing the flexibility comes at the cost of increased computational expense (423 possible combinations vs. 6). To keep the computational cost manageable we employ a procedure referred to as randomised cross-validation that randomly draws combinations of hyperparameters from the set of available combinations. Specifically, we perform randomised cross-validation over 60 specifications from the set of hyperparameters above. Note, even under randomised cross-validation as we employ it here the effective training time increases by a factor of 10 (60 specifications vs. 6 specifications).

B.3 Accounting for Overlapping Observations

In this section we address an aspect related to the empirical execution of Bianchi et al. (2020) which has been brought to our attention after the acceptance for publication.[7] To set the stage, it is important to redefine notation. Specifically, define the one-year holding period return as $r^{(n)}_{t+12} = \log \frac{P^{(n-12)}_{t+12}}{P^{(n)}_{t}}$ where the time units are in months (i.e., 12 refers to twelve months). In forecasting $r^{(n)}_{t+12}$, Bianchi et al. (2020) made use of information brought about by $r^{(n)}_{t+11} = \log \frac{P^{(n-12)}_{t+11}}{P^{(n)}_{t-1}}$ (this was the case also for the benchmark historical returns).[8] Due to the high persistence in bond returns, it is important to re-evaluate the findings in Bianchi et al. (2020) when one calculates the forecasts of $r^{(n)}_{t+12}$ exploiting the return series only up to $r^{(n)}_{t}$.

We revisit the main empirical analysis in the original paper. Specifically, we consider both a framework that exploits information in the yield curve only, as in Cochrane and Piazzesi (2005), as well as a framework that uses information from a dataset of a large set of macroeconomic indicators as in Ludvigson and Ng (2009). We employ the same list of models analysed in the main paper to ensure full comparison of the results.

Moreover, in addition to the analysis of one-year holding period excess returns, we also report evidence for the one-month holding period returns. This ana-

[7]We thank Tobias Hoogteijling for bringing this aspect to our attention.

[8]Note that defining $r_{t,t+1} = \log \frac{P_{t+1}}{P_t}$ for stock returns, we have $r^s_{t,t+12} = r_{t,t+1} + r_{t+1,t+2} + r_{t+2,t+3} + r_{t+3,t+4} + r_{t+4,t+5} + r_{t+5,t+6} + r_{t+6,t+7} + r_{t+7,t+8} + r_{t+8,t+9} + r_{t+9,t+10} + r_{t+10,t+11} + r_{t+11,t+12}$, and $r^s_{t-1,t+11} = r_{t-1,t} + r_{t,t+1} + r_{t+1,t+2} + r_{t+2,t+3} + r_{t+3,t+4} + r_{t+4,t+5} + r_{t+5,t+6} + r_{t+6,t+7} + r_{t+7,t+8} + r_{t+8,t+9} + r_{t+9,t+10} + r_{t+10,t+11}$; there are clearly 11 observations that are exactly the same between $r^s_{t-1,t+11}$ and $r^s_{t,t+12}$. However, for bonds this is not immediately the case since the n-year bond becomes an $n-1$ month bond, one month later. Indeed,

$$r^{(n)}_{t+12} = \log \left(\frac{P^{(n-1)}_{t+1}}{P^{(n)}_{t}} \frac{P^{(n-2)}_{t+2}}{P^{(n-1)}_{t+1}} \frac{P^{(n-3)}_{t+3}}{P^{(n-2)}_{t+2}} \frac{P^{(n-4)}_{t+4}}{P^{(n-3)}_{t+3}} \frac{P^{(n-5)}_{t+5}}{P^{(n-4)}_{t+4}} \frac{P^{(n-6)}_{t+6}}{P^{(n-5)}_{t+5}} \frac{P^{(n-7)}_{t+7}}{P^{(n-6)}_{t+6}} \frac{P^{(n-8)}_{t+8}}{P^{(n-7)}_{t+7}} \frac{P^{(n-9)}_{t+9}}{P^{(n-8)}_{t+8}} \frac{P^{(n-10)}_{t+10}}{P^{(n-9)}_{t+9}} \frac{P^{(n-11)}_{t+11}}{P^{(n-10)}_{t+10}} \frac{P^{(n-12)}_{t+12}}{P^{(n-11)}_{t+11}} \right)$$

vs.

$$r^{(n)}_{t+11} = \log \left(\frac{P^{(n-1)}_{t}}{P^{(n)}_{t-1}} \frac{P^{(n-2)}_{t+1}}{P^{(n-1)}_{t}} \frac{P^{(n-3)}_{t+2}}{P^{(n-2)}_{t+1}} \frac{P^{(n-4)}_{t+3}}{P^{(n-3)}_{t+2}} \frac{P^{(n-5)}_{t+4}}{P^{(n-4)}_{t+3}} \frac{P^{(n-6)}_{t+5}}{P^{(n-5)}_{t+4}} \frac{P^{(n-7)}_{t+6}}{P^{(n-6)}_{t+5}} \frac{P^{(n-8)}_{t+7}}{P^{(n-7)}_{t+6}} \frac{P^{(n-9)}_{t+8}}{P^{(n-8)}_{t+7}} \frac{P^{(n-10)}_{t+9}}{P^{(n-9)}_{t+8}} \frac{P^{(n-11)}_{t+10}}{P^{(n-10)}_{t+9}} \frac{P^{(n-12)}_{t+11}}{P^{(n-11)}_{t+10}} \right)$$

with no immediate evidence of overlapping observations. The issue arises because of the time series autocorrelation in bond returns (e.g., for the 10-year bond, the autocorrelation of the annual holding period return series is ca. 92% when using monthly, overlapping observations).

lysis appeared in an early draft of the paper, but was dropped during the revision process. However, this evidence is important because it provides a cleaner setup to test the effect of overlapping returns on the main results of the paper.

The empirical evidence from the revised analysis confirms the main message of the paper, that is: (1) within each empirical application (i.e., yields-only or yields plus macroeconomic variables), non-linear machine learning methods, in particular neural networks (NNs), are helpful to detect predictable variations in bond excess returns, as indicated by out-of-sample predictive R^2s that are significantly higher than those obtained by data compression techniques (e.g., linear combinations of forward rates, as in Cochrane and Piazzesi (2005), and factors extracted from macroeconomic variables, as in Ludvigson and Ng (2009)) and penalised regression techniques; (2) zooming in on non-linear methods, we document that using information from macroeconomic and financial variables improves the predictive accuracy of forecasts based only on (potentially non-linear transformations of) the yield curve; (3) NN gains in predictive accuracy translate into better investment performance compared to a naive strategy based on the recursive mean; (4) the economic gains from NN forecasts based on macroeconomic and yield information are significantly larger than those obtained using yields alone.

Research Design

We consider as a burn-in sample the period from 1971:08 to 1989:01.[9] Using only information up until the end of this period we estimate the forecasting model for individual bond excess return and lagged predictors (macro and/or yields). Due to the predictive nature of the regression, the last observation in the right-hand-side variables is that of January 1988. We use the estimates and the value of the predictors on January 1989 to produce out-of-sample forecasts of one-year excess returns for each maturity. The first forecast error obtains by comparing the excess holding period return during the February 1989 through January 1990 period and its forecast made on January 1989. We then include the February 1989 information and follow the same procedure to produce forecasts of the March 1989 through February

[9] We use the Liu and Wu (2020) dataset, for which the full set of ten forward rates is available starting from 1971:08.

1990 returns, and so on until the end of 2018.[10]

Bond Return Predictability and the Yield Curve

Table B.2 displays results when we forecast excess returns of Treasury bonds with the yield curve.

Panel A of Table B.2 displays the results for PCRs and partial least squares (PLS). Consistent with the original paper, the predictive R^2 are negative across different maturities. A parsimonious representation with only three PCs significantly outperforms the specification with five and ten PCs, particularly at longer maturities.

Panel B displays the results from various configurations of linear penalised regressions. Differently from the paper, we now see that sparse modelling performs poorly out-of-sample, independently from the bond maturity. In fact, they perform even worse than a parsimonious data compression method (PCA) for almost all maturities.

Panel C of Table B.2 shows the results for the non-linear methods. With the only difference of regression tree methods, the main results of the paper holds. That is, neural networks continue to attain good performance with significantly positive R^2_{oos} across maturities. We also observe that a shallow network with a single hidden layer and three nodes performs only marginally worse than the best, deeper network with two hidden layers and three nodes. Further increasing the depth of the network deteriorates its performance. This continues to be the case even when we consider alternative structures, like a NN with three hidden layers and pyramidal node architecture.

One may argue that the forecasting results can depend on a particular structure of neural network. To address this issue, we complement and extend some of the results reported in the original paper and report the R^2_{oos} for two model combination strategies that are based on two alternative model averaging schemes: the first model combination scheme combines forecasts of each neural network based on the inverse of the validation loss, which is akin to combine forecasts based on a

[10] In the exercise with one-month holding periods, we use the Center for Research in Security Prices Treasury Fama bond portfolios. In this case the sample period is from 1964:01 to 2016:12, as in the draft version of the paper. The start of the out-of-sample period remains the same, i.e. January 1990.

pseudo-out-of-sample prediction error. The second combination scheme is more agnostic and simply gives equal weights to the forecasts of each neural network. While an equal-weight strategy may seem overly simplistic, the forecasting literature has shown that it may outperform the optimal weights based on log-score or in-sample calibration (see, e.g., Elliott and Timmermann, 2004, Smith and Wallis, 2009, and Diebold and Shin, 2017). The results show rather unequivocally that even when averaged out, neural networks perform best in out-of-sample forecasting of annual holding period returns.

To summarise, our analysis correcting for non-overlapping returns confirms that NNs constitute the best-performing methods even in the case when only information in the term structure is used to forecast bond returns (i.e., in a low dimensional setting). This, again, suggests that the gain from non-linear machine learning methods may not necessarily be relegated to a big data context. Overall, we continue to confirm the main results in Section 3.1 of Bianchi et al. (2020).

Bond Return Predictability and Macroeconomic Variables

Next, we consider the setup where information embedded in the yield curve does not necessarily subsume information contained in macro variables.

The results in Panels A and B in Table B.3 show that: (1) dense modelling, such as data compression techniques and ridge regression, tends to perform poorly out-of-sample; and (2) sparse modelling with both regularisation and shrinkage (i.e., elastic net regressions), perform well at long maturities, particularly when restricting the linear combination of forward rates.

Turning to non-linear machine learning methods in Panel C of Table B.3, we observe that the performance of hybrid networks stands out. The only difference relative to the original paper is that increasing the depth of the NN from one- to three-layers does not improve its accuracy.

We also confirm that a careful choice of network structure based on prior economic information exerts a great impact on performance. In particular, a one-layer *group ensembled* model performs better than the *hybrid* NN for the 2-, 3-, 4-years maturities and attains slightly lower predictive accuracy for long-term bonds. Also for group ensembling NN, adding more layers is detrimental to the performance.

In line with the analysis reported for the case with yields only above, we also compute the performance of model combination strategies that are based on a weighted average of each neural network's forecasts (see also Section 4.5 of Bianchi et al. (2020)). This is a natural strategy to exploit, since we have observed that different network structures perform differently at short- and long-maturity (e.g., the 1-layer NN with group ensembling and the hybrid 1-layer network perform best at short- and long-maturity, respectively). As before, we consider a forecast combination strategy whereby each neural network is weighted based on the inverse of the validation loss, as well as a second, more agnostic, combination scheme which gives equal weight to each neural network prediction. The results show that as a class of model, on average, neural networks unequivocally outperform both data compression, penalised regression and shallow non-linear models, therefore confirming the main results of the original paper.

Finally, comparing Table B.2 to Table B.3, we confirm that macroeconomic variables carry information that is not contained in the yield curve. For instance, a group ensembled NN that exploits macroeconomic and term structure information attains out-of-sample R^2s that are about twice as large as the best-performing NN that employs yields only for maturities ranging from two to ten years.

Economic Value of Return Forecasts from Neural Networks

We re-examine the findings in Section 5 of Bianchi et al. (2020). Again, since our goal is solely to correct for the persistence in one-year holding period returns, we consider an asset allocation framework that is identical to that described in Section 5.1 of the main paper.

Table B.4 shows the annualised certainty equivalent return (CER) values computed relative to the EH model. Positive values indicate that the predictive model performs better than the EH model. For the sake of completeness, we focus on the predictions implied by (1) the best performing neural network as in Tables B.2-B.3, (2) a model-average neural-network based on equal weights, (3) a model-average neural network where the weights of each model are based on the validation loss, and (4) for the exercise with macro + fwd rates the group-ensemble approach outlined in the main paper. We flag in bold those values that are statistically significant at the 5% confidence level.

Two facts emerge; first, there is strong evidence for the economic value of using neural network forecasts both for the univariate case exploiting the long maturity bond, and for the multivariate asset allocation strategy. Second, including macroeconomic information in the conditioning information set significantly increases the economic performance of the forecasts; specifically, the NN forecasts that exploit macroeconomic information produce significantly higher CER values than those implied by a NN using yields-only (for the univariate and multivariate allocation, independently of the form of utility).

Overall, with now the exception of regression trees, we largely confirm the analysis in Section 5.2 of Bianchi et al. (2020) that neural network-based forecasts of bond returns provide support for the hypothesis that a (statistically and economically) significant portion of macroeconomic information is not captured by the yield curve, even after accounting for non-linearity in interest rates.

Table B.2: Forecasting Annual Holding Period Returns with Forward Rates

This table reports the out-of-sample R^2_{oos} obtained using forward rates to predict annual excess bond returns for different maturities and across methodologies. The monthly one-year holding period returns have been lagged by 11 months in order to mitigate the effect of overlapping returns on the forecasting performance. To compute the out-of-sample R^2_{oos} we compare the forecasts obtained from each methodology to the expectation hypothesis (i.e., to the prediction based on the historical mean). In addition to the R^2_{oos} we report the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice that we report a p-value only when the R^2_{oos} is positive. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12.

Models	R^2_{oos}						p-value						R^2_{oos} EW	p-value EW
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(EW)}_{t+1}$	$xr^{(EW)}_{t+1}$
Panel A: PCA and PLS														
PCA (10 of 10 components)	-92.8%	-71.0%	-59.5%	-49.9%	-37.4%	-20.8%							-45.2%	
PCA (5 of 10 components)	-78.3%	-59.4%	-49.1%	-36.9%	-30.1%	-14.7%							-36.0%	
PCA (3 of 10 components)	-30.7%	-22.2%	-15.9%	-8.1%	-5.3%	3.4%							-7.8%	
PCA-Squared (5 of 10 Components)	-99.3%	-83.8%	-70.6%	-57.1%	-47.0%	-31.1%							-55.7%	
PCA-Squared (3 of 10 Components)	-86.8%	-77.4%	-66.3%	-52.9%	-47.0%	-37.5%							-55.5%	
Partial Least Squares (5 components)	-93.4%	-73.9%	-64.8%	-54.1%	-43.5%	-30.0%							-52.0%	
Partial Least Squares (3 components)	-85.2%	-65.0%	-54.3%	-43.4%	-32.2%	-15.7%							-39.3%	
Panel B: Penalised Linear Regressions														
Ridge	-67.0%	-49.3%	-40.8%	-31.4%	-24.5%	-10.8%							-29.5%	
LASSO	-22.9%	-23.0%	-20.0%	-19.7%	-21.2%	-15.6%							-19.4%	
Elastic Net	-22.5%	-22.5%	-24.2%	-18.1%	-23.3%	-18.1%							-21.2%	
Panel C: Regression Trees and Neural Networks														
Gradient Boosted Tree	-64.9%	-73.8%	-75.1%	-64.4%	-62.4%	-53.2%							-64.2%	
Random Forest	-18.4%	-13.7%	-11.7%	-7.4%	-8.2%	-5.0%							-9.0%	
Extreme Tree	-26.6%	-22.9%	-20.6%	-17.2%	-18.3%	-17.0%							-20.0%	
NN - 1 layer (3 nodes)	2.6%	4.3%	4.5%	4.4%	4.4%	3.8%	16.8%	10.1%	9.5%	8.8%	10.6%	12.6%	4.4%	10.4%
NN - 1 layer (5 nodes)	2.8%	4.3%	4.9%	4.6%	4.5%	3.8%	15.6%	8.3%	6.7%	5.9%	7.6%	9.5%	4.6%	7.6%
NN - 1 layer (7 nodes)	2.8%	4.0%	4.3%	3.9%	3.8%	3.2%	12.4%	6.0%	5.1%	5.1%	7.3%	10.3%	3.9%	6.8%
NN - 2 layer (3 nodes each)	5.5%	6.5%	6.8%	5.9%	6.2%	5.1%	2.8%	1.1%	1.0%	1.0%	1.4%	2.6%	6.2%	1.4%
NN - 2 layer (5 nodes each)	5.9%	6.6%	6.5%	5.7%	5.8%	5.1%	2.7%	1.5%	1.5%	1.5%	2.4%	3.9%	6.1%	2.1%
NN - 2 layer (7 nodes each)	6.8%	6.8%	6.4%	5.5%	5.6%	4.6%	1.4%	1.0%	1.0%	1.2%	1.7%	3.3%	5.9%	1.5%
NN - 3 layer (3 nodes each)	4.0%	4.5%	4.3%	3.9%	3.2%	2.3%	1.1%	0.5%	0.3%	0.4%	0.2%	0.7%	3.6%	0.3%
NN - 3 layer (5 nodes each)	4.5%	4.5%	4.4%	3.5%	3.3%	2.6%	1.1%	0.3%	0.2%	0.2%	0.1%	0.5%	3.6%	0.1%
NN - 3 layer (7 nodes each)	5.0%	4.7%	4.9%	4.1%	3.9%	3.0%	0.9%	0.1%	0.1%	0.1%	0.1%	0.5%	4.1%	0.1%
NN - 3 layer (5,4,3 nodes each)	4.6%	5.1%	4.9%	4.0%	3.9%	2.8%	0.6%	0.1%	0.1%	0.2%	0.1%	0.4%	4.1%	0.1%
NN - 2 layer (5 nodes), incl. 3 lags	3.0%	3.6%	3.5%	3.2%	3.5%	3.1%	12.1%	6.6%	7.9%	6.8%	7.5%	9.4%	3.4%	7.9%
NN - 3 layer (5 nodes), incl. 3 lags	5.1%	4.8%	4.5%	3.6%	3.5%	2.8%	0.3%	0.1%	0.0%	0.0%	0.0%	0.2%	3.8%	0.0%

Table B.3: Forecasting Annual Holding Period Returns with Forward Rates and Macroeconomic Variables

This table reports the out-of-sample R^2_{oos} obtained using forward rates and a large panel of macroeconomic variables to predict annual excess bond returns for different maturities. The monthly one-year holding period returns have been lagged by 11 months in order to mitigate the effect of overlapping returns on the forecasting performance. To compute the out-of-sample R^2_{oos} we compare the forecasts obtained from each methodology to the expectation hypothesis (i.e., prediction based on the historical mean). In addition to the R^2_{oos} we report the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice that we report a p-value only when the R^2_{oos} is positive. The out-of-sample prediction errors are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12. Penalised regressions are estimated including macroeconomic variables plus either raw forward rates or a linear combination of forward rates as introduced by Cochrane and Piazzesi (2005) (CP). Similarly, neural networks are estimated either adding the CP factor as an additional regressor in the output layer ("fwd rates direct") or by estimating a separate network for forward rates and ensembling both macro and forward rates networks in the output layer ("fwd rates net").

Models	R^2_{oos}						p-value						R^2_{oos} EW	p-value EW
	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(2)}_{t+1}$	$xr^{(3)}_{t+1}$	$xr^{(4)}_{t+1}$	$xr^{(5)}_{t+1}$	$xr^{(7)}_{t+1}$	$xr^{(10)}_{t+1}$	$xr^{(EW)}_{t+1}$	$xr^{(EW)}_{t+1}$
Panel A: PCA and PLS														
PCA - first 8 PC'S	-24.3%	-12.8%	-7.8%	-3.8%	-2.7%	-1.8%							-4.7%	
PCA as in LN	-9.0%	-4.7%	-3.1%	-2.9%	-6.0%	-10.3%							-5.9%	
PLS - 8 components	-105.0%	-69.0%	-55.4%	-46.3%	-33.3%	-20.8%							-40.7%	
Panel B: Penalised Linear Regressions														
Ridge (using CP factor)	-122.6%	-82.8%	-64.4%	-51.7%	-35.2%	-18.7%							-41.3%	
Lasso (using CP factor)	-31.6%	-16.6%	-10.4%	-9.3%	-5.9%	6.3%						0.1%	-2.9%	
Elastic Net (using CP factor)	-33.4%	-14.3%	-8.5%	-4.0%	5.7%	13.1%					0.2%	0.0%	2.4%	0.3%
Ridge (using fwd rates directly)	-146.9%	-111.6%	-99.8%	-87.7%	-73.5%	-52.9%							-79.5%	
Lasso (using fwd rates directly)	-22.2%	-15.1%	-8.8%	-7.3%	-1.4%	1.5%						0.9%	-3.3%	
Elastic Net (using fwd rates directly)	-27.8%	-15.3%	-10.0%	-4.9%	2.2%	5.7%					1.1%	0.4%	-0.6%	
Panel C: Regression Trees and Neural Networks														
Random Forest	-4.2%	-12.0%	-8.9%	-7.0%	-6.3%	-2.3%							-5.9%	
Gradient Boosted Tree	-17.4%	-14.0%	-9.6%	-6.3%	-7.7%	-11.4%							-7.9%	
Extreme Tree	-3.2%	-1.3%	-1.5%	-0.7%	-3.3%	2.0%						8.3%	-0.7%	
NN 1 Layer (32 nodes), fwd rates direct	3.4%	3.2%	8.6%	16.7%	21.1%	28.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	19.6%	0.0%
NN 2 Layer (32, 16 nodes), fwd rates direct	-37.5%	-6.2%	-5.5%	1.5%	6.9%	14.5%				0.0%	0.0%	0.0%	3.6%	0.0%
NN 3 Layer (32, 16, 8 nodes), fwd rates direct	-5.7%	2.4%	5.5%	9.0%	9.5%	13.8%		0.5%	0.3%	0.2%	0.1%	0.1%	8.8%	0.2%
NN 1 Layer (32 nodes), fwd rates net (2 layer: 3,3 nodes)	0.1%	15.0%	15.5%	16.7%	16.0%	22.1%	1.2%	0.3%	0.2%	0.1%	0.1%	0.0%	18.2%	0.1%
NN 2 Layer((32,16, nodes), fwd rates net (2 layer: 3,3 nodes)	-9.3%	1.9%	2.1%	4.0%	7.0%	11.5%		0.2%	0.1%	0.1%	0.0%	0.0%	6.5%	0.1%
NN 3 Layer((32,16, 8 nodes), fwd rates net (2 layer: 3,3 nodes)	-13.4%	-1.2%	-1.2%	1.8%	2.3%	6.1%				0.3%	0.1%	0.1%	2.2%	0.2%
NN 1 Layer Group Ensem (1 node per group), fwd rates direct	-9.2%	2.4%	8.1%	12.5%	14.1%	16.4%		0.6%	0.2%	0.1%	0.0%	0.0%	12.1%	0.0%
NN 2 Layer Group Ensem (2,1 nodes per group), fwd rates direct	-2.1%	6.2%	9.5%	12.2%	13.8%	16.7%		2.1%	0.9%	0.3%	0.1%	0.0%	12.6%	0.2%
NN 1 Layer Group Ensem (1 node per group), fwd rates net (1 layer: 3 nodes)	3.6%	8.9%	11.3%	10.9%	12.9%	17.0%	0.9%	0.4%	0.3%	0.1%	0.1%	0.0%	13.3%	0.1%
NN 2 Layer Group Ensem (2, 1 node per group), fwd rates net (2 layer: 3, 3 nodes)	-9.5%	2.1%	1.7%	5.8%	8.4%	13.7%		1.5%	1.2%	0.6%	0.3%	0.1%	8.0%	0.3%

Table B.4: Economic Significance of Forecasts for Non-Overlapping Annual Returns.
This table reports the annualised certainty equivalent values (in %) for portfolio decisions based on the out-of-sample forecasts of bond excess returns for an investor with either mean-variance (Panel A) or power utility (Panel B) and a coefficient of risk aversion equal to five. The table reports two asset allocation exercises. In the univariate asset allocation case, the investor selects either the two- or the ten-year bond, along with the one-year short-rate. In the multivariate case, the investor selects bonds across the six maturities, two- to five-, seven- and ten-years. The asset allocation decision is based on the predictions implied by (1) the best performing neural network as in Tables 2.1-2.2, (2) a model-average neural-network based on equal weights, (3) a model-average neural network where the weights of each model are based on the validation loss, and (4) for the exercise with macro + fwd rates the group-ensemble approach outlined in the main paper. The models are benchmarked against the expectation hypothesis. We also test for the difference in each model performance between a forecast using only fwd rates (CP) and both fwd rates and macroeconomic variables (LN). The out-of-sample predictions are obtained by a recursive forecast which starts in January 1990. The sample period is from 1971:08-2018:12. Statistical significance is based on a one-sided Diebold and Mariano (1995) test as extended by Harvey et al. (1997) to account for autocorrelation in the forecasting errors. We flag in bold those values that are statistically significant at the 5% confidence level.

Panel A: Mean-Variance Utility

	2-year maturity			10-year maturity			All		
	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ
Best performing NN	0.29	0.30	0.01	**0.77**	**6.20**	**5.44**	1.08	**3.45**	**2.37**
p-value	(0.085)	(0.108)	(0.901)	(0.044)	(0.000)	(0.000)	(0.377)	(0.038)	(0.045)
Group-ensemble NN		0.30	0.01		**4.48**	**3.71**		3.03	1.95
p-value		(0.077)	(0.191)		(0.000)	(0.000)		(0.114)	(0.075)
NN Model averaged EW	0.30	**0.40**	0.11	**0.66**	**5.54**	**4.88**	0.04	**2.06**	**2.02**
p-value	(0.078)	(0.024)	(0.202)	(0.004)	(0.000)	(0.000)	(0.150)	(0.000)	(0.002)
NN Model averaged VL	0.30	**0.41**	0.11	**0.67**	**5.43**	**4.76**	0.03	**1.92**	**1.89**
p-value	(0.078)	(0.022)	(0.210)	(0.004)	(0.000)	(0.000)	(0.130)	(0.000)	(0.001)

Panel B: Power Utility

	2-year maturity			10-year maturity			All		
	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ	Fwd rates	Fwd + Macro	Δ
Best performing NN	0.26	0.27	0.01	**0.78**	**6.14**	**5.35**	1.25	**3.81**	**2.56**
p-value	(0.103)	(0.135)	(0.454)	(0.046)	(0.000)	(0.000)	(0.209)	(0.031)	(0.032)
Group-ensemble NN		0.27	0.01		**4.46**	**3.68**		3.33	1.98
p-value		(0.096)	(0.191)		(0.000)	(0.000)		(0.088)	(0.065)
NN Model averaged EW	0.27	**0.37**	0.10	**0.67**	**5.49**	**4.82**	0.17	**2.22**	**2.05**
p-value	(0.093)	(0.030)	(0.230)	(0.004)	(0.000)	(0.000)	(0.175)	(0.000)	(0.003)
NN Model averaged VL	0.27	**0.37**	0.11	**0.69**	**5.36**	**4.69**	0.16	**2.05**	**1.88**
p-value	(0.098)	(0.028)	(0.210)	(0.005)	(0.000)	(0.000)	(0.168)	(0.000)	(0.001)

Appendix C

What Drives Asset Holdings? Commonality in Investor Demand

C.1 Data

The construction of the asset characteristics data mostly follows Freyberger et al. (2020b). In order to ensure that balance sheet data is available at the time market price information is available, I follow the Fama and French (1993b) timing convention and use balance sheet information from fiscal year $t-1$ for returns from July of year t to June of year $t-1$.

- **a2me:** Following Bhandari (1988), assets-to-market equity can be defined as total assets over the market capitalisation as of December in year $t-1$.
- **at:** Total firm assets as in Gandhi and Lustig (2015).
- **ato:** Net sales over lagged net operating expenses as in Soliman (2008), where net operating assets are the diferrence between operating assets and operating liabilities.
- **beme:** Book value of equity over market value of equity.
- **beta:** Following Frazzini and Pedersen (2014), CAPM beta can be defined as the product of correlations between excess returns of a stock and market

excess returns over the ratio of their volatilities. Correlations are estimated using overlapping three-day windows over five-year periods with at least 750 non-missing observations. Volatilities are computed from standard deviation of at least 120 observations of daily log-excess returns over a one-year horizon.

- **c:** Ratio of cash plus short-term investments to total assets as in Palazzo (2012).

- **cto:** Following Haugen et al. (1996), capital turnover can be defined as ratio of net sales to lagged total assets.

- **delta_pi2a:** Change in property, plants, and equipment over lagged total assets as in Lyandres et al. (2008).

- **e2p:** Following Basu (1981), earnings to price ratio can be defined as income before extraordinary items to market capitalisation as of December in year $t-1$.

- **freecf:** Free cash flow to book value of equity can be defined as net income of the firm plus depreciation and amortization minus change in working capital and capital expenditure divided by book value of equity.

- **idiovol:** Idiosyncratic volatility defined as in Ang et al. (2006) is the standard deviation of residuals from a regression of excess returns on the Fama and French (1993b) three factors where one month of daily data with at least 15 non-missing observations is used.

- **investment:** Following Cooper et al. (2008), investment can be defined as the year-on-year growth rate in total assets in percent.

- **lev:** Following Bhandari (1988), leverage can be defined as the ratio of long-term debt plus debt in current liabilities over market equity.

- **ldp:** Following Litzenberger and Ramaswamy (1979), dividend-price ratio can be defined as a measure of annual dividends over last month's price. Annual dividend is measured as the sum of dividends over the past 12 months.

- **lme:** Total market capitalisation defined as price times number of shares outstanding.

- **lturnover:** Following Datar et al. (1998), trading turnover can be defined as traded volume in the previous month over number of shares outstanding.
- **noa:** Following Hirshleifer et al. (2004), net operating assets can be defined as operating asset minus operating liabilities divided by lagged total assets.
- **oa:** Following Sloan (1996), operating accruals can be defined as changes in non-cash working capital minus depreciation divided by lagged total assets.
- **ol:** Following Novy-Marx (2011), operating leverage can be defined as the sum of cost of goods sold plus selling, general and administrative expense over total assets.
- **pcm:** Following Gorodnichenko and Weber (2016) and D'Acunto et al. (2018), price-to-cost margin can be defined as the difference between net sales and cost of goods sold over net sales
- **pm:** Following Soliman (2008), profit margin can be defined as operating income after depreciation over lagged sales.
- **prof:** Following Ball et al. (2015), profitability can be defined as gross profit divided by book value of equity.
- **q:** Tobin's Q is defined as total assets plus market equity minus cash and short-term investments minus deferred taxes divided by total assets.
- **rel_to_high_price:** Current return relative to the previous 52 week high.
- **rna:** Following Soliman (2008), return on net operating assets is given as operating income after depreciation divided by nlagged net operating assets.
- **roa:** Following Balakrishnan et al. (2010), return-on-assets is given as income before extraordinary items divided by total assets.
- **roe:** Following Haugen et al. (1996), return-on-equity can be defined as income before extraordinary items divided by book value of equity.
- **r12_2:** Momentum defined as return over past 12 months excluding the most recent two months like in Fama and French (1996).

- **r12_7:** Intermediate momentum defined as return over past 12 months excluding most recent seven months as in Novy-Marx (2012).
- **r2_1:** Short-term reversal defined as return over past two months excluding the most recent month as in Jegadeesh (1990).
- **r36_13:** Long-term reversal defined as return over the past three years excluding the most recent 13 months as in De Bondt and Thaler (1985).
- **s2p:** Following Lewellen (2015), sales-to-price is the ratio of net sales to market capitalisation as of last December.
- **sat:** Following Soliman (2008), asset turnover can be defined as the ratio of sales to lagged total assets.
- **sga2s:** Ratio of selling, general and administrative expenses to net sales.
- **spread:** Following Chung and Zhang (2014), a measure of bid-ask spread can be computed as the average daily spread in the previous month.
- **suv:** Following Garfinkel (2009), standarised unexplained volume is computed as actual volume minus a measure of predicted volume. Predicted volume is the fitted value from a regression of daily volume on a constant and the absolute value of returns. Standardization is achieved by dividing by the standard deviation of residuals from the regression.

Bibliography

Abbassi, Puriya, Rajkamal Iyer, José Luis Peydró, and Francesc R. Tous (2016) "Securities trading by banks and credit supply: Micro-evidence from the crisis", *Journal of Financial Economics*, Vol. 121, No. 3, pp. 569–594.

Admati, A. (1985) "A noisy rational expectations equilibrium for multi-asset securities markets", *Econometrica*, Vol. 53, pp. 629–658.

——— (1993) "A model of intertemporal asset prices under asymmetric information", *Review of Economic Studies*, Vol. 60, pp. 249–282.

Admati, A. and P. Pfleiderer (1991) "Sunshine trading and financial market equilibrium", *Review of Financial Studies*, Vol. 4, pp. 443–481.

Adrian, Tobias and Hyun Song Shin (2010) "Liquidity and leverage", *Journal of Financial Intermediation*, Vol. 19, No. 3, pp. 418–437.

Adrian, Tobias, Erkko Etula, and Tyler Muir (2014) "Financial Intermediaries and the Cross-Section of Asset Returns", *Journal of Finance*, Vol. 69, No. 6, pp. 2557–2596.

Ahn, Dong Hyun, Robert F. Dittmar, and Andrew Ronald Gallant (2002) "Quadratic Term Structure Models: Theory and Evidence", *Review of Financial Studies*, Vol. 15, No. 1, pp. 243–288.

Aït-Sahalia, Y. and M. Saglam (2017) "High Frequency Market Making: Optimal Quoting", Working Paper.

Akbas, F., R. Petkova, and W. Armstrong (2011) "The Volatility of Liquidity and Expected Stock Returns", Working Paper.

Andrade, Sandro C., Charles Chang, and Mark S. Seasholes (2008) "Trading imbalances, predictable reversals, and cross-stock price pressure", *Journal of Financial Economics*, Vol. 88, No. 2, pp. 406–423.

Andreou, Elena, Eric Ghysels, and Andros Kourtellos (2010) "Regression models with mixed sampling frequencies", *Journal of Econometrics*, Vol. 158, No. 2, pp. 246–261.

——— (2013) "Should Macroeconomic Forecasters Use Daily Financial Data and How?", *Journal of Business and Economic Statistics*, Vol. 31, No. 2, pp. 240–251.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang (2006) "The cross-section of volatility and expected returns", *Journal of Finance*, Vol. 61, No. 1, pp. 259–299.

Arlot, Sylvain, Alain Celisse et al. (2010) "A survey of cross-validation procedures for model selection", *Statistics surveys*, Vol. 4, pp. 40–79.

Atif, Mian and Amir Sufi (2009) "The consequences of mortgage credit expansion", *Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1449–1496.

Bakshi, Gurdip and Nikunj Kapadia (2003) "Delta-Hedged Gains and the Negative Market Volatility Risk Premium", *Review of Financial Studies*, Vol. 16, No. 2, pp. 527–566.

Balakrishnan, Karthik, Eli Bartov, and Lucile Faurel (2010) "Post loss/profit announcement drift", *Journal of Accounting and Economics*, Vol. 50, No. 1, pp. 20–41.

Ball, Ray, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri V. Nikolaev (2015) "Deflating profitability", *Journal of Financial Economics*, Vol. 117, No. 2, pp. 225–248.

Bansal, Ravi and Ivan Shaliastovich (2013) "A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets", *Review of Financial Studies*, Vol. 26, No. 1, pp. 1–33, 01.

Barber, Brad M. and Terrance Odean (2000) "Trading is hazardous to your wealth: The common stock investment performance of individual investors", *Journal of Finance*, Vol. 55, No. 2, pp. 773–806.

——— (2013) "The behavior of individual investors", in *Handbook of the Economics of Finance*, Vol. 2: Elsevier, pp. 1533–1570.

Barber, Brad M., Xing Huang, and Terrance Odean (2016) "Which Factors Matter to Investors? Evidence from Mutual Fund Flows", *Review of Financial Studies*, Vol. 29, No. 10, pp. 2601–2642.

Basu, Sanjoy (1981) "The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence", *Journal of Financial Economics*, Vol. 12, No. 1, pp. 129–156.

Bates, John M. and Clive W.J. Granger (1969) "The combination of forecasts", *Journal of the Operational Research Society*, Vol. 20, No. 4, pp. 451–468.

Bauer, Michael D. and James D. Hamilton (2018) "Robust Bond Risk Premia", *Review of Financial Studies*, Vol. 31, No. 2, pp. 399–448.

Bauer, Michael D. and Glenn D. Rudebusch (2017) "Resolving the Spanning Puzzle in Macro-Finance Term Structure Models", *Review of Finance*, Vol. 21, No. 2, pp. 511–553.

Bekaert, Geert, Eric Engstrom, and Yuhang Xing (2009) "Risk, uncertainty, and asset prices", *Journal of Financial Economics*, Vol. 91, No. 1, pp. 59–82.

Bekaert, Geert, Eric C. Engstrom, and Nancy R. Xu (2019) "The Time Variation in Risk Appetite and Uncertainty", NBER Working Paper No. 25673, National Bureau of Economic Research.

Berk, Jonathan B. and Richard C. Green (2004) "Mutual fund flows and performance in rational markets", *Journal of Political Economy*, Vol. 112, No. 6, pp. 1269–1295.

Bessembinder, H., A. Carrion, L. Tuttle, and K. Venkataraman (2015) "Liquidity, resiliency, and market quality around predictable trades: Theory and Evidence", *Journal of Financial Economics*, Vol. 7, pp. 1–20.

Bhandari, Laxmi Chand (1988) "Debt/equity ratio and expected common stock returns: Empirical evidence", *Journal of Finance*, Vol. 43, No. 2, pp. 507–528.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni (2020) "Bond Risk Premiums with Machine Learning", *The Review of Financial Studies*, 05. hhaa062.

Bishop, Christopher M. (1995) "Regularization and complexity control in feed-forward networks", in *Proceedings International Conference on Artificial Neural Networks ICANN'95*, pp. 141–148: EC2 et Cie.

——— (2006) *Pattern recognition and machine learning*: Springer.

Black, Fischer (1995) "Interest Rates as Options", *Journal of Finance*, Vol. 50, No. 5, pp. 1371–1376.

Black, Fischer and Myron Scholes (1973) "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, Vol. 81, No. 3, pp. 637–654.

Bloom, N. (2009) "The impact of uncertainty shocks", *Econometrica*, Vol. 77, No. 3, pp. 623–685.

Boehmer, Ekkehart and Juan Julie Wu (2008) "Order Flow and Prices", Working Paper.

Boivin, Jean and Serena Ng (2006) "Are more data always better for factor analysis?", *Journal of Econometrics*, Vol. 132, No. 1, pp. 169–194.

Bollerslev, Tim, George Tauchen, and Hao Zhou (2009) "Expected stock returns and variance risk premia", *Review of Financial Studies*, Vol. 22, No. 11, pp. 4463–4492.

Breiman, Leo (2001) "Random forests", *Machine learning*, Vol. 45, No. 1, pp. 5–32.

Breiman, Leo, Jerome H. Friedman, Charles J. Stone, and R.A. Olshen (1984) *Classification and Regression Trees*: Taylor & Francis.

Brennan, Michael J. and Yihong Xia (2002) "Dynamic asset allocation under inflation", *Journal of Finance*, Vol. 57, No. 3, pp. 1201–1238.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado (1997) "Strategic Asset Allocation", *Journal of Economic Dynamics and Control*, Vol. 21, pp. 1377–1403.

Broadie, Mark, Mikhail Chernov, and Michael Johannes (2007) "Model specification and risk premia: Evidence from futures options", *Journal of Finance*, Vol. 62, No. 3, pp. 1453–1490.

Brogaard, J., A. Carrion, Th. Moyaert, R. Riordan, A. Shkilko, and K.n Sokolov (2018) "High frequency trading and extreme price movements", *Journal of Financial Economics*, Vol. 128, No. 2, pp. 253–265.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014) "High-frequency trading and price discovery", *Review of Financial Studies*, Vol. 27, No. 8, pp. 2267–2306.

Brooks, Robert, Don Chance, and Mobina Shafaati (2018) "The Cross-Section of Individual Equity Option Returns", Working Paper.

Brown, Stephen J. and William N. Goetzmann (1997) "Mutual fund styles", *Journal of Financial Economics*, Vol. 43, No. 3, pp. 373–399.

Brunnermeier, M. and L. Pedersen (2005) "Predatory trading", *Journal of Finance*, Vol. 60, pp. 1825–1863.

Buraschi, Andrea and Alexei Jiltsov (2007) "Habit Formation and Macroeconomic Models of the Term Structure of Interest Rates", *Journal of Finance*, Vol. 62, No. 6, pp. 3009–3063.

Buraschi, Andrea, Ilaria Piatti, and Paul Whelan (2019) "Subjective Bond Risk Premia and Belief Aggregation", Working Paper, Said Business School.

Campbell, John Y. and John H. Cochrane (1999) "By force of habit: A consumption-based explanation of aggregate stock market behavior", *Journal of Political Economy*, Vol. 107, No. 2, pp. 205–251.

Campbell, John Y. and Robert J. Shiller (1991) “Yield Spreads and Interest Rate Movements: A Bird’s Eye View”, *Review of Economic Studies*, Vol. 58, No. 3, pp. 495–514.

Campbell, John Y. and Samuel B. Thompson (2008) “Predicting excess stock returns out of sample: Can anything beat the historical average?”, *Review of Financial Studies*, Vol. 21, No. 4, pp. 1509–1531.

Campbell, John Y. and Luis M. Viceira (1999) “Consumption and portfolio decisions when expected returns are time varying”, *Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 433–495.

——— (2004) “Long-horizon mean-variance analysis: A user guide”, *Manuscript, Harvard University, Cambridge, MA*.

Campbell, John Y., Yeung Lewis Chan, and Luis M. Viceira (2003) “A multivariate model of strategic asset allocation”, *Journal of Financial Economics*, Vol. 67, No. 1, pp. 41–80.

Cao, Jie and Bing Han (2013) “Cross-Section of Option Returns and Idiosyncratic Stock Volatility”, *Journal of Financial Economics*, Vol. 108, No. 1, pp. 231–249.

Cao, Jie, Bing Han, Qing Tong, and Xintong Zhan (2015) “Option Return Predictability”, Working Paper 2698267, Rotman School of Management.

Carhart, Mark M. (1997) “On Persistence in Mutual Fund Performance”, *Journal of Finance*, Vol. 52, No. 1, pp. 57–82.

Carlin, B., M. Lobo, and S. Viswanathan (2007) “Episodic liquidity crises: cooperative and predatory trading”, *Journal of Finance*, Vol. 62, pp. 2235–2274.

Carr, Peter and Liuren Wu (2004) “Time-changed Lévy Processes and Option Pricing”, *Journal of Financial Economics*, Vol. 71, No. 1, pp. 113–141.

——— (2020) “Option profit and loss attribution and pricing: A new framework”, *Journal of Finance*.

Chambers, Donald R., Matthew Foy, Jeffrey Liebner, and Qin Lu (2014) “Index Option Returns: Still Puzzling”, *Review of Financial Studies*, Vol. 27, No. 6, pp. 1915–1928.

Chan, Louis K. C., Hsiu-Lang Chen, and Josef Lakonishok (2002) “On Mutual Fund Investment Styles”, *Review of Financial Studies*, Vol. 15, No. 5, pp. 1407–1437, oct.

Chen, Hui, Winston Dou, and Leonid Kogan (2017) “Measuring the ’Dark Matter’ in Asset Pricing Models”, NBER Working Paper No. 26418, National Bureau of Economic Research.

Chen, H., G. Noronha, and V. Singal (2006) "Index changes and losses to index fund investors", *Financial Analysts Journal*, Vol. 62, pp. 31–47.

Chen, Luyang, Markus Pelger, and Jason Zhu (2019) "Deep Learning in Asset Pricing", SSRN Working Paper 3350138.

Cheng, M. and A. Madhavan (2009) "The dynamics of leveraged and inverse exchange-traded-funds", *Journal of Investment Management*, Vol. 7, pp. 1–20.

Choi, J., L. Jin, and H. Yan (2016) "Informed Trading and Expected Returns", Working Paper.

Chordia, Tarun and Avanidhar Subrahmanyam (2004) "Order imbalance and individual stock returns: Theory and evidence", *Journal of Financial Economics*, Vol. 72, No. 3, pp. 485–518, jun.

Chordia, Tarun, J. Hu, A. Subrahmanyam, and Q. Tong (2018) "Order Flow Volatility and Equity Costs of Capital", *Management Science*, Vol. forthcoming.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2002) "Order imbalance, liquidity, and market returns", *Journal of Financial Economics*, Vol. 65, No. 1, pp. 111–130, jul.

——— (2008) "Liquidity and market efficiency", *Journal of Financial Economics*, Vol. 87, No. 2, pp. 249–268.

Christoffersen, Peter, Mathieu Fournier, and Kris Jacobs (2018) "The factor structure in equity options", *Review of Financial Studies*, Vol. 31, No. 2, pp. 595–637.

Chung, Kee H. and Hao Zhang (2014) "A simple approximation of intraday spreads using daily data", *Journal of Financial Markets*, Vol. 17, pp. 94–120.

Cieslak, Anna and Pavol Povala (2015) "Expected Returns in Treasury Bonds", *Review of Financial Studies*, Vol. 28, No. 10, pp. 2859–2901.

Clark, Todd E. and Kenneth D. West (2007) "Approximately normal tests for equal predictive accuracy in nested models", *Journal of Econometrics*, Vol. 138, No. 1, pp. 291–311.

Clemen, Robert T. (1989) "Combining forecasts: A review and annotated bibliography", *International Journal of Forecasting*, Vol. 5, No. 4, pp. 559–583.

Cochrane, John H. (2009) *Asset Pricing: Revised Edition*: Princeton University Press.

Cochrane, John H. and Monika Piazzesi (2005) "Bond Risk Premia", *American Economic Review*, Vol. 95, No. 1, pp. 138–160, March.

Comerton-Forde, C., T. Hendershott, C. Jones, P. Moulton, and M. Seasholes (2010) “Time Variation in Liquidity: The Role of Market Maker Inventories and Revenues”, *Journal of Finance*, Vol. 65, pp. 295–331.

Constantinides, George M., Jens Carsten Jackwerth, and Stylianos Perrakis (2009) “Mispricing of S&P 500 Index Options”, *Review of Financial Studies*, Vol. 22, No. 3, pp. 1247–1277.

Cooper, Ilan and Richard Priestley (2009) “Time-Varying Risk Premiums and the Output Gap”, *Review of Financial Studies*, Vol. 22, No. 7, pp. 2801–2833.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill (2008) “Asset growth and the cross-section of stock returns”, *Journal of Finance*, Vol. 63, No. 4, pp. 1609–1651.

Coroneo, Laura, Domenico Giannone, and Michele Modugno (2016) “Unspanned Macroeconomic Factors in the Yield Curve”, *Journal of Business & Economic Statistics*, Vol. 34, No. 3, pp. 472–485.

Corsi, Fulvio (2009) “A Simple Approximate Long-Memory Model of Realized Volatility”, *Journal of Financial Econometrics*, Vol. 7, No. 2, pp. 174–196.

Coval, Joshua D. and Tyler Shumway (2001) “Expected Option Returns”, *Journal of Finance*, Vol. 56, No. 3, pp. 983–1009.

Coval, Joshua and Erik Stafford (2007) “Asset fire sales (and purchases) in equity markets”, *Journal of Financial Economics*, Vol. 86, No. 2, pp. 479–512.

Creal, Drew D. and Jing Cynthia Wu (2018) “Bond Risk Premia in Consumption-based Models”, NBER Working Paper 22183, National Bureau of Economic Research.

Cybenko, George (1989) “Approximation by superpositions of a sigmoidal function”, *Mathematics of control, signals and systems*, Vol. 2, No. 4, pp. 303–314.

Dai, Qiang, Kenneth J. Singleton, and Wei Yang (2007) “Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields”, *Review of Financial Studies*, Vol. 20, No. 5, pp. 1669–1706.

Dangl, Thomas and Michael Halling (2012) “Predictive regressions with time-varying coefficients”, *Journal of Financial Economics*, Vol. 106, No. 1, pp. 157–181.

Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe (1998) “Liquidity and stock returns: An alternative test”, *Journal of Financial Markets*, Vol. 1, No. 2, pp. 203–219.

De Bondt, Werner F.M. and Richard Thaler (1985) “Does the stock market overreact?”, *Journal of finance*, Vol. 40, No. 3, pp. 793–805.

Degryse, Hans, F. de Jong, and V. van Kervel (2013) "Does order splitting signal uninformed order flow?", Working Paper.

Della Corte, Pasquale, Lucio Sarno, and Daniel Thornton (2008) "The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value", *Journal of Financial Economics*, Vol. 89, No. 1, pp. 158–174.

Diaconis, P. and M. Shahshahani (1984) "On Nonlinear Functions of Linear Combinations", *SIAM Journal on Scientific and Statistical Computing*, Vol. 5, No. 1, pp. 175–191.

Dickey, D. A. and W. A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association1*, Vol. 74, No. 366, pp. 427–431.

Diebold, Francis X. and Robert S. Mariano (1995) "Comparing predictive accuracy", *Journal of Business & Economic Statistics*, Vol. 20, pp. 134–144.

Diebold, Francis X. and Minchul Shin (2017) "Beating the simple average: Egalitarian LASSO for combining economic forecasts".

Dimopoulos, Yannis, Paul Bourret, and Sovan Lek (1995) "Use of some sensitivity criteria for choosing networks with good generalization ability", *Neural Processing Letters*, Vol. 2, No. 6, pp. 1–4.

Dow, James and Gary Gorton (1953) "Profitable informed trading in a simple general equilibrium model of asset pricing", *Journal of Economic Theory*, Vol. 67, pp. 327–369.

Duffee, Gregory R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models", *Journal of Finance*, Vol. 57, No. 1, pp. 405–443.

——— (2011a) "Information in (and not in) the Term Structure", *Review of Financial Studies*, Vol. 24, No. 9, pp. 2895–2934.

——— (2011b) "Forecasting with the term structure: The role of no-arbitrage restrictions", Economics Working Paper Archive No. 576, The Johns Hopkins University.

——— (2013) *Forecasting Interest Rates*, Vol. 2 of Handbook of Economic Forecasting, Chap. 0, pp. 385–426: Elsevier.

Duffie, Darrel (2010) "Presidential address: asset price dynamics with slow-moving capital", *Journal of Finance*, Vol. 65, pp. 1237–1267.

Duffie, Darrell, Jun Pan, and Kenneth Singleton (2000) "Transform Analysis and Asset Pricing for Affine Jump-Diffusions", *Econometrica*, Vol. 68, No. 6, pp. 1343–1376.

Dumas, Bernard, Alexander Kurshev, and Raman Uppal (2009) "Equilibrium Portfolio Strategies in the Presence of Sentiment Risk and Excess Volatility", *Journal of Finance*, Vol. 64, No. 2, pp. 579–629.

D'Acunto, Francesco, Ryan Liu, Carolin Pflueger, and Michael Weber (2018) "Flexible prices and leverage", *Journal of Financial Economics*, Vol. 129, No. 1, pp. 46–68.

Easley, David and Maureen O'Hara (2004) "Information and the Cost of Capital", *Journal of Finance*, Vol. 59, No. 4, pp. 1553–1583.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara (2002) "Is information risk a determinant of asset returns?", *Journal of Finance*, Vol. 57, pp. 2185–2221.

Elliott, Graham and Allan Timmermann (2004) "Optimal forecast combinations under general loss functions and forecast error distributions", *Journal of Econometrics*, Vol. 122, No. 1, pp. 47–79.

Eraker, Bjørn (2004) "Do stock prices and volatility jump? Reconciling evidence from spot and option prices", *Journal of Finance*, Vol. 59, No. 3, pp. 1367–1404.

Eraker, Bjørn, Michael Johannes, and Nicholas Polson (2003) "The Impact of Jumps in Volatility and Returns", *Journal of Finance*, Vol. 58, No. 3, pp. 1269–1300.

Fama, Eugene F. and Robert R. Bliss (1987) "The information in long-maturity forward rates", *American Economic Review*, pp. 680–692.

Fama, Eugene F. and Kenneth R. French (1993a) "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, Vol. 33, pp. 3–56.

——— (1993b) "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, Vol. 33, No. 1, pp. 3–56.

——— (1996) "Multifactor explanations of asset pricing anomalies", *Journal of Finance*, Vol. 51, No. 1, pp. 55–84.

——— (2015) "A five-factor asset pricing model", *Journal of Financial Economics*, Vol. 116, No. 1, pp. 1–22.

Fama, Eugene F. and James D. MacBeth (1973) "Risk, return, and equilibrium: Empirical tests", *Journal of political economy*, Vol. 81, No. 3, pp. 607–636.

Feldhutter, Peter, Christian Heyerdahl-Larsen, and Philipp Illeditsch (2016) "Risk Premia and Volatilities in a Non-linear Term Structure Model", *Review of Finance*, Vol. 22, No. 1, pp. 337–380.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2020) "Taming the factor zoo: A test of new factors", *Journal of Finance*, Vol. 75, No. 3, pp. 1327–1370.

Feng, Guanhao, Jingyu He, and Nicholas G. Polson (2018) "Deep Learning for Predicting Asset Returns", *arXiv preprint arXiv:1804.09314*.

Feng, Guanhao, Nicholas Polson, and Jianeng Xu (2019) "Deep Learning Alpha", Chicago Booth Research Paper 23527, Chicago Booth.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek (2001) "The Economic Value of Volatility Timing", *Journal of Finance*, Vol. 56, No. 1, pp. 329–352, February.

Frank, Ldiko E. and Jerome H. Friedman (1993) "A statistical view of some chemometrics regression tools", *Technometrics*, Vol. 35, No. 2, pp. 109–135.

Frazzini, Andrea and Lasse Heje Pedersen (2012) "Embedded Leverage", NBER Working Paper 18558, National Bureau of Economic Research.

——— (2014) "Betting against beta", *Journal of Financial Economics*, Vol. 111, No. 1, pp. 1–25.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020a) "Dissecting Characteristics Nonparametrically", *Review of Financial Studies*, Vol. 33, No. 5, pp. 2326–2377.

——— (2020b) "Dissecting characteristics nonparametrically", *Review of Financial Studies*, Vol. 33, No. 5, pp. 2326–2377.

Friedman, Jerome H. (2001) "Greedy function approximation: a gradient boosting machine", *Annals of statistics*, pp. 1189–1232.

Friedman, Jerome H., Trevor Hastie, and Robert Tibshirani (2001) *The elements of statistical learning*, Vol. 1: Springer, New York, NY, USA.

——— (2010) "Regularization paths for generalized linear models via coordinate descent", *Journal of statistical software*, Vol. 33, No. 1, p. 1.

Friedman, Jerome H., Trevor Hastie, Holger Höfling, and Rob Tibshirani (2007) "Pathwise Coordinate Optimization", *The Annals of Applied Statistics*, Vol. 1, No. 2, pp. 302–332.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2018) "Predictably unequal? the effects of machine learning on credit markets", SSRN Working Paper 3072038.

Gandhi, Priyank and Hanno Lustig (2015) "Size anomalies in US bank stock returns", *Journal of Finance*, Vol. 70, No. 2, pp. 733–768.

Garfinkel, Jon A. (2009) "Measuring investors' opinion divergence", *Journal of Accounting Research*, Vol. 47, No. 5, pp. 1317–1348.

Gargano, Antonio, Davide Pettenuzzo, and Allan Timmermann (2019) "Bond Return Predictability: Economic Value and Links to the Macroeconomy", *Management Science*, Vol. 65, No. 2, pp. 508–540.

Geurts, Pierre, Damien Ernst, and Louis Wehenkel (2006) "Extremely randomized trees", *Machine learning*, Vol. 63, No. 1, pp. 3–42.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov (2005) "There is a risk-return trade-off after all", *Journal of Financial Economics*, Vol. 76, No. 3, pp. 509–548.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov (2007) "MIDAS regressions: Further results and new directions", *Econometric Reviews*, Vol. 26, No. 1, pp. 53–90.

Giacoletti, Marco, Kristoffer Laursen, and Kenneth J Singleton (2016) "Learning, dispersion of beliefs, and risk premiums in an arbitrage-free term structure model", Working Paper.

Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2017) "Economic predictions with big data: The illusion of sparsity", Working Paper, Federal Reserve Bank of New York.

Giglio, Stefano and Dacheng Xiu (2017) "Inference on Risk Premia in the Presence of Omitted Factors", NBER Working Paper 23527, National Bureau of Economic Research.

Gompers, Paul A. and Andrew Metrick (2001) "Institutional investors and equity prices", *Quarterly Journal of Economics*, Vol. 116, No. 1, pp. 229–259.

Goodfellow, Ian, Yoshua Bengio, Aaron Courville, and Yoshua Bengio (2016) *Deep learning*, Vol. 1, Cambridge, MA, USA: MIT Press.

Gorodnichenko, Yuriy and Michael Weber (2016) "Are sticky prices costly? Evidence from the stock market", *American Economic Review*, Vol. 106, No. 1, pp. 165–99.

Goyal, Amit and Alessio Saretto (2009) "Cross-Section of Option Returns and Volatility", *Journal of Financial Economics*, Vol. 94, No. 2, pp. 310–326.

Griffin, John M., Jeffrey H. Harris, and Selim Topaloglu (2003) "The Dynamics of Institutional and Individual Trading", *Journal of Finance*, Vol. 58, No. 6, pp. 2285–2320.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers (1995) "Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior", *American Economic Review*, Vol. 85, No. 5, pp. 1088–1105.

Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu (2019) "Autoencoder asset pricing models", *Journal of Econometrics, Forthcoming.*

——— (2020) "Empirical Asset Pricing via Machine Learning", *Review of Financial Studies*, Vol. 33, No. 5, pp. 2223–2273.

Guidolin, Massimo and Allan Timmermann (2007) "Asset allocation under multivariate regime switching", *Journal of Economic Dynamics and Control*, Vol. 31, No. 11, pp. 3503–3544.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007) "The U.S. Treasury yield curve: 1961 to the present", *Journal of Monetary Economics*, Vol. 54, No. 8, pp. 2291–2304, November.

Hagströmer, Björn (2018) "Overestimated Effective Spreads: Implications for Investors", Working Paper.

Harrison, J. Michael and David M. Kreps (1978) "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations", *Quarterly Journal of Economics*, Vol. 92, No. 2, pp. 323–336.

Harvey, David, Stephen Leybourne, and Paul Newbold (1997) "Testing the equality of prediction mean squared errors", *International Journal of forecasting*, Vol. 13, No. 2, pp. 281–291.

Haugen, Robert A., Nardin L. Baker et al. (1996) "Commonality in the determinants of expected stock returns", *Journal of Financial Economics*, Vol. 41, No. 3, pp. 401–439.

He, Zhiguo and Arvind Krishnamurthy (2013) "Intermediary Asset Pricing", *American Economic Review*, Vol. 103, No. 2, pp. 732–770.

He, Zhiguo, Bryan Kelly, and Asaf Manela (2017) "Intermediary Asset Pricing: New Evidence from many Asset Classes", *Journal of Financial Economics*, Vol. 126, No. 1, pp. 1–35.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015) "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.

Heaton, J. B., N. G. Polson, and J. H. Witte (2017) "Deep learning for finance: deep portfolios", *Applied Stochastic Models in Business and Industry*, Vol. 33, No. 1, pp. 3–12.

Hendershott, Terrence and Albert Menkveld (2014) "Price pressures", *Journal of Financial Economics*, Vol. 114, pp. 405–423.

Hendershott, Terrence and Mark S. Seasholes (2009) "Market Predictability and Non-Informational Trading", Working Paper.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld (2011) "Does Algorithmic Trading Improve Liquidity?", *Journal of Finance*, Vol. 66, No. 1, pp. 1–33, feb.

Heston, Steven L. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *Review of Financial Studies*, Vol. 6, No. 2, pp. 327–343.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang (2004) "Do investors overvalue firms with bloated balance sheets?", *Journal of Accounting and Economics*, Vol. 38, pp. 297–331.

Holden, Craig W. and Stacey Jacobsen (2014) "Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions", *Journal of Finance*, Vol. 69, No. 4, pp. 1747–1785.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989) "Multilayer feedforward networks are universal approximators", *Neural networks*, Vol. 2, No. 5, pp. 359–366.

Hu, Gang, Koren M. Jo, Yi Alex Wang, and Jing Xie (2018) "Institutional trading and Abel Noser data", *Journal of Corporate Finance*, Vol. 52, pp. 143–167.

Huang, Jing-Zhi and Zhan Shi (2019) "Determinants of Bond Risk Premia: A Machine-Learning-Based Resolution of the Spanning Controversy", Working Papers 2019-12, Penn State.

Hwang, C. and X. Qian (2011) "Is Information Risk Priced? Evidence from the Price Discovery of Large Trades", Working Paper.

Ioffe, Sergey and Christian Szegedy (2015) "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *arXiv preprint arXiv:1502.03167*.

Israelov, Roni and Bryan T. Kelly (2017) "Forecasting the Distribution of Option Returns", SSRN Working Paper 3033242.

Israelov, Roni and Harsha Tummala (2017) "Which Index Options Should You Sell?", SSRN Working Paper 2990542.

Jegadeesh, Narasimhan (1990) "Evidence of predictable behavior of security returns", *Journal of finance*, Vol. 45, No. 3, pp. 881–898.

Jegadeesh, Narasimhan and Sheridan Titman (1993) "Returns to buying winners and selling losers: Implications for stock market efficiency", *Journal of Finance*, Vol. 48, No. 1, pp. 65–91.

Johnson, Travis L. and Eric C. So (2018) "A simple multimarket measure of information asymmetry", *Management Science*, Vol. 64, No. 3, pp. 1055–1080.

Jones, Christopher S. (2006) "A Nonlinear Factor Analysis of S&P 500 Index Option Returns", *Journal of Finance*, Vol. 61, No. 5, pp. 2325–2363.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton (2014) "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks", *Journal of Finance*, Vol. 69, No. 3, pp. 1197–1233.

Karakaya, Mete (2013) "Characteristics and Expected Returns in Individual Equity Options", Working Paper, University of Chicago.

Kelly, Bryan and Seth Pruitt (2013) "Market Expectations in the Cross-Section of Present Values", *Journal of Finance*, Vol. 68, No. 5, pp. 1721–1756.

——— (2015) "The three-pass regression filter: A new approach to forecasting using many predictors", *Journal of Econometrics*, Vol. 186, No. 2, pp. 294–316.

Kelly, Bryan T., Seth Pruitt, and Yinan Su (2017) "Instrumented Principal Component Analysis", SSRN Working Paper 2983919.

——— (2019) "Characteristics Are Covariances: A Unified Model of Risk and Return", *Journal of Financial Economics, forthcoming*, Vol. 134, No. 3, pp. 501–524.

Koijen, Ralph S. J. and Motohiro Yogo (2019a) "A Demand System Approach to Asset Pricing", *Journal of Political Economy*, Vol. 127, No. 4, pp. 1475–1515.

——— (2019b) "Exchange Rates and Asset Prices in a Global Demand System", NBER Working Paper 27342, National Bureau of Economic Research.

Koijen, Ralph S. J., Robert J. Richmond, and Motohiro Yogo (2019) "Which Investors Matter for Global Equity Valuations and Expected Returns?", *Working Paper*.

Kolmogorov, A. K. (1957) "On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of One Variable and Addition", *Doklady Akademii Nauk SSSR*, Vol. 114, pp. 369–373.

Koopman, S.J. and J. Durbin (2012) *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press, 2nd edition.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2017) "Shrinking the Cross Section", NBER Working Paper 24070, National Bureau of Economic Research.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson (2018) "HAR Inference: Recommendations for Practice", *Journal of Business & Economic Statistics*, Vol. 36, No. 4, pp. 541–559, October.

Le, Anh and Kenneth J. Singleton (2013) "The Structure of Risks in Equilibrium Affine Models of Bond Yields", Working Paper, University of North Carolina at Chapel Hill.

Le, Anh, Kenneth J. Singleton, and Qiang Dai (2010) "Discrete-Time AffineQ Term Structure Models with Generalized Market Prices of Risk", *Review of Financial Studies*, Vol. 23, No. 5, pp. 2184–2227.

Lee, Charles M. C. and Mark J. Ready (1991) "Inferring Trade Direction from Intraday Data", *Journal of Finance*, Vol. 46, No. 2, pp. 733–746.

Leippold, Markus and Liuren Wu (2003) "Design and Estimation of Quadratic Term Structure Models", *Review of Finance*, Vol. 7, No. 1, pp. 47–73.

Lewellen, Jonathan (2015) "The cross section of expected stock returns", *Critical Finance Review*, Vol. 4, No. 1, pp. 1–44.

Litzenberger, Robert H. and Krishna Ramaswamy (1979) "The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence", *Journal of financial economics*, Vol. 7, No. 2, pp. 163–195.

Liu, Yan and Jing Cynthia Wu (2020) "Reconstructing the yield curve", NBER Working Paper No. 27266, National Bureau of Economic Research.

Lo, Andrew W. and A. Craig MacKinlay (1988) "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test", *Review of Financial Studies*, Vol. 1, No. 1, pp. 41–66.

Loughran, Tim and Jay R. Ritter (2002) "Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs?", *Review of Financial Studies*, Vol. 15, No. 2 SPEC., pp. 413–443.

Ludvigson, Sydney C. and Serena Ng (2009) "Macro Factors in Bond Risk Premia", *Review of Financial Studies*, Vol. 22, No. 12, pp. 5027–5067.

Lyandres, Evgeny, Le Sun, and Lu Zhang (2008) "The new issues puzzle: Testing the investment-based explanation", *Review of Financial Studies*, Vol. 21, No. 6, pp. 2825–2855.

Madhavan, A. and S. Smidt (1993) "An analysis of changes in specialist inventories and quotations", *Journal of Finance*, Vol. 48, pp. 1595–1628.

McCracken, Michael W. and Serena Ng (2016) "FRED-MD: A Monthly Database for Macroeconomic Research", *Journal of Business and Economic Statistics*, Vol. 34, No. 4, pp. 574–589.

Messmer, Marcial (2017) "Deep Learning and the Cross-Section of Expected Returns", SSRN Working Paper 3081555.

Mian, Atif and Amir Sufi (2011) "House Prices, Home Equity–Based Borrowing, and the US Household Leverage Crisis", *American Economic Review*, Vol. 101, No. 5, pp. 2132–2156.

Mian, Atif, Kamalesh Rao, and Amir Sufi (2013) "Household balance sheets, consumption, and the economic slump", *Quarterly Journal of Economics*, Vol. 128, No. 4, pp. 1687–1726.

Mohanram, P. and S. Rajgopal (2009) "Is PIN priced risk?", *Journal of Accounting and Economics*, Vol. 47, pp. 226–243.

Mou, Y. (2011) "Limits to arbitrage and commodity index investment: front-running the Goldman Roll", Working paper.

Mullainathan, Sendhil and Jann Spiess (2017) "Machine learning: an applied econometric approach", *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 87–106.

Murphy, Kevin P. (2012) *Machine learning: a probabilistic perspective*: MIT Press.

Nesterov, Yuri (1983) "A method for solving the convex programming problem with convergence rate O(1/k$\hat{2}$)", *Dokl. Akad. Nauk SSSR*, Vol. 269, pp. 543–547.

Newbold, Paul and Clive W.J. Granger (1974) "Experience with forecasting univariate time series and the combination of forecasts", *Journal of the Royal Statistical Society: Series A (General)*, Vol. 137, No. 2, pp. 131–146.

Newey, Whitney K. and Kenneth D. West (1987a) "Hypothesis testing with efficient method of moments estimation", *International Economic Review*, pp. 777–787.

——— (1987b) "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, Vol. 55, No. 3, pp. 703–708.

Nofsinger, John R. and Richard W. Sias (1996) "Herding and Feedback Trading by Institutional and Individual Investors", *Financial Management Association Doctoral Consortium*, Vol. 54, No. 6.

Novy-Marx, Robert (2011) “Operating leverage”, *Review of Finance*, Vol. 15, No. 1, pp. 103–134.

——— (2012) “Is momentum really momentum?”, *Journal of Financial Economics*, Vol. 103, No. 3, pp. 429–453.

Odean, Terrance (1998) “Are Investors Reluctant to Realize Their Losses?”, *The Journal of Finance*, Vol. 53, No. 5, pp. 1775–1798.

Ofek, Eli and Matthew Richardson (2003) “DotComMania: The Rise and Fall of Internet Stock Prices”, *Journal of Finance*, Vol. 58, No. 3, pp. 1113–1137.

Ofek, Eli, Matthew Richardson, and Robert F. Whitelaw (2004) “Limited Arbitrage and Short Sales Restrictions: Evidence from the Options Markets”, *Journal of Financial Economics*, Vol. 74, No. 2, pp. 305–342.

Palazzo, Berardino (2012) “Cash holdings, risk, and expected returns”, *Journal of Financial Economics*, Vol. 104, No. 1, pp. 162–185.

Pascual, Roberto and David Veredas (2010) “Does the Open Limit Order Book Matter in Explaining Informational Volatility?”, *Journal of Financial Econometrics*, Vol. 8, No. 1, pp. 57–87.

Pastor, Lubos and Robert F. Stambaugh (2003) “Liquidity risk and expected stock returns”, *Journal of Political Economy*, Vol. 111, No. 3, pp. 642–685.

Pastor, Lubos and Pietro Veronesi (2005) “Rational IPO Waves”, *Journal of Finance*, Vol. 60, No. 4, pp. 1713–1757.

Petajisto, A. (2011) “The index premium and Its hidden cost for index funds”, *Journal of Empirical Finance*, Vol. 18, pp. 271–288.

Piazzesi, Monika, Juliana Salomao, and Martin Schneider (2015) “Trend and cycle in bond premia”, Staff Report 424, Federal Reserve Bank of Minneapolis.

Raman, V., M. Robe, and P. Yadav (2016) “Liquidity Provision Under Stress: The Fast, the Slow, and the Dead”, Working Paper.

Rapach, David E., Jack K. Strauss, and Guofu Zhou (2010) “Out-of-sample equity premium prediction: Combination forecasts and links to the real economy”, *Review of Financial Studies*, Vol. 23, No. 2, pp. 821–862.

——— (2013) “International Stock Return Predictability: What Is the Role of the United States?”, *Journal of Finance*, Vol. 68, No. 4, pp. 1633–1662.

Ritter, Jay R. and Ivo Welch (2002) "A review of IPO activity, pricing, and allocations", *Journal of Finance*, Vol. 57, No. 4, pp. 1795–1828.

Rossi, Andrea G. (2018) "Predicting Stock Market Returns with Machine Learning", Working Paper, University of Maryland.

Rosu, I., E. Sojli, and W.W. Tham (2018) "Quoting Activity and the Cost of Capital", Working Paper.

Saffi, Pedro A.C. and Kari Sigurdsson (2011) "Price efficiency and short selling", *Review of Financial Studies*, Vol. 24, No. 3, pp. 821–852.

Saglam, Mehmet (2020) "Order anticipation around predictable trades", *Financial Management*, Vol. 49, No. 1, pp. 33–67.

Sarno, Lucio, Paul Schneider, and Christian Wagner (2016) "The economic value of predicting bond risk premia", *Journal of Empirical Finance*, Vol. 37, pp. 247–267.

Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller (1998) "Nonlinear component analysis as a kernel eigenvalue problem", *Neural Computation*, Vol. 10, No. 5, pp. 1299–1319.

Sias, Richard W. (2004) "Institutional Herding", *Review of Financial Studies*, Vol. 17, No. 1, pp. 165–206.

Sirignano, Justin, Apaar Sadhwani, Kay Giesecke et al. (2018) "Deep Learning for Mortgage Risk", Working Paper.

Skjeltorp, J., E. Sojli, and W.W. Tham (2017) "Flashes of trading intent at NASDAQ", *Journal of Financial and Quantitative Analysis*, Vol. 51, pp. 165–196.

Sloan, Richard G. (1996) "Do stock prices fully reflect information in accruals and cash flows about future earnings?", *Accounting review*, Vol. 71, No. 3, pp. 289–315.

Smith, Jeremy and Kenneth F. Wallis (2009) "A simple explanation of the forecast combination puzzle", *Oxford Bulletin of Economics and Statistics*, Vol. 71, No. 3, pp. 331–355.

Soliman, Mark T. (2008) "The use of DuPont analysis by market participants", *Accounting Review*, Vol. 83, No. 3, pp. 823–853.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014) "Dropout: a simple way to prevent neural networks from overfitting", *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958.

Stock, James H. and Mark W. Watson (1998) “A comparison of linear and non-linear univariate models for forecasting macroeconomic time series”, NBER Working Paper 6607, National Bureau of Economic Research.

——— (2002) “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business & Economic Statistics*, Vol. 20, No. 2, pp. 147–162, April.

——— (2006) “Forecasting with Many Predictors”, Vol. 1: Elsevier, 1st edition, Chap. 10, pp. 515–554.

——— (2012) “Disentangling the channels of the 2007-09 recession”, *Brookings Papers on Economic Activity*, No. 1, pp. 81–135.

Stone, Mervyn and Rodney J Brooks (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)*, pp. 237–269.

Subrahmanyam, Avanidhar (1991) “Risk Aversion, Market Liquidity, and Price Efficiency”, *Review of Financial Studies*, Vol. 4, No. 3, pp. 417–441.

Thornton, Daniel L. and Giorgio Valente (2012) “Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective”, *Review of Financial Studies*, Vol. 25, No. 10, pp. 3141–3168.

Timmer, Yannick (2018) “Cyclical investment behavior across financial institutions”, *Journal of Financial Economics*, Vol. 129, No. 2, pp. 268–286.

Tipping, Michael E. and Christopher M. Bishop (1999) “Probabilistic principal component analysis”, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, Vol. 61, No. 3, pp. 611–622.

Tuzan, T. (2013) “Are leveraged and inverse ETFs the new portfolio insurers?”, Working paper.

Vasquez, Aurelio (2017) “Equity Volatility Term Structures and the Cross-Section of Option Returns”, *Journal of Financial and Quantitative Analysis*, Vol. 52, No. 6, pp. 2727–2754.

Veronesi, Pietro (2004) “Belief-dependent Utilities, Aversion to State-Uncertainty and Asset Prices”, CRSP Working Papers, Center for Research in Security Prices, Graduate School of Business, University of Chicago.

Wachter, Jessica A. (2006) “A consumption-based model of the term structure of interest rates”, *Journal of Financial Economics*, Vol. 79, No. 2, pp. 365–399.

Weigand, Alois (2019) "Machine learning in empirical asset pricing", *Financial Markets and Portfolio Management*, Vol. 33, No. 1, pp. 93–104.

Welch, Ivo and Amit Goyal (2008) "A comprehensive look at the empirical performance of equity premium prediction", *Review of Financial Studies*, Vol. 21, No. 4, pp. 1455–1508.

Werbos, Paul J. (1974) "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences", Ph.D. dissertation, Harvard University.

——— (1982) *Applications of advances in nonlinear sensitivity analysis*, pp. 762–770: Springer.

——— (1988) "Generalization of backpropagation with application to a recurrent gas market model", *Neural networks*, Vol. 1, No. 4, pp. 339–356.

Wermers, Russ (1999) "Mutual fund herding and the impact on stock prices", *Journal of Finance*, Vol. 54, No. 2, pp. 581–622.

West, Mike and Jeff Harrison (1997) *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag New York, 2nd edition.

Wu, Jing Cynthia and Fan Dora Xia (2016) "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound", *Journal of Money, Credit and Banking*, Vol. 48, No. 2-3, pp. 253–291.

Wu, Tong Tong, Kenneth Lange et al. (2008) "Coordinate descent algorithms for lasso penalized regression", *The Annals of Applied Statistics*, Vol. 2, No. 1, pp. 224–244.

Yang, Wei (2011) "Long-run risk in durable consumption", *Journal of Financial Economics*, Vol. 102, No. 1, pp. 45–61.

Yang, Yung Chiang, Bohui Zhang, and Chu Zhang (2020) "Is information risk priced? Evidence from abnormal idiosyncratic volatility", *Journal of Financial Economics*, Vol. 135, No. 2, pp. 528–554.

Zou, Hui and Trevor Hastie (2005) "Method of Conjugate Gradients for Solving Linear Systems", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 67, No. 2, pp. 301–320.