Predicting Real Economic Activity with Individual Option-Implied Expected Returns[1]

Eirini Konstantinidi[2], Mina Mirshahi[3], and George Skiadopoulos[4]

December 2024

PRELIMINARY AND INCOMPLETE - PLEASE DO NOT QUOTE OR CIRCULATE

Abstract

We propose a new predictor to forecast U.S. real economic activity (REA) by utilising the information embedded in equity option prices. We construct our equity option-based predictor by applying standard and recent data reduction methods, to the cross-section of computed option-implied expected returns of the underlying stocks. Our predictor forecasts REA both in- and out-of-sample setting even after controlling for common REA predictors and considering their persistence. We find a robust negative relationship between the option-implied predictor and REA. We show that individual stocks contain some additional predictive power that is not being captured neither by the index option-implied expected return, nor by standard factors.

[1]The authors thank the participants of the 7th International Young Finance Scholars' Conference.
[2]Alliance Manchester Business School, University of Manchester, UK, Email: eirini.konstantiidi@manchester.ac.uk
[3]School of Economics and Finance, Queen Mary University of London, UK, Email: mina.mirshahi@qmul.ac.uk.
[4]School of Economics and Finance, Queen Mary University of London, UK, and Department of Banking and Financial Management, University of Piraeus, Greece. Also Senior Visiting Fellow at Cass Business School, City University of London. g.skiadopoulos@qmul.ac.uk and gskiado@unipi.gr

1 Introduction

The provision of reliable forecasts of real economic activity (REA) has long been one of the principal challenges for economists and of primary importance to corporates, investors and policymakers. There has been considerable research on forecasting economic activity using measures based on the prices of stocks, bonds and commodities. These variables, however, failed to predict the economic downturn from 2007 to 2009, highlighting the necessity of developing a precise REA predictor using financial markets' information (Ng and Wright (2013)).

In this paper, we propose a new option-based predictor which uses for the first time information from *individual* equity option prices, that is, option-implied underlying stocks' expected returns. The stock's expected return is a natural candidate as a predictor of REA. Campbell and Cochrane (1999) show that in a standard log-linear representative agent model, the stock's expected returns are counter-cyclical. Surprisingly, the previous literature has not examined its predictive ability, possibly due to the econometric challenges in estimating stocks' expected returns using historical data.[1]

Considering expected returns extracted from option prices bypasses the obstacle of estimating expected returns using backward looking data. Options are forward-looking and are expected to possess informational content about the future state of the economy, thus, being natural candidates for forecasting REA. Bakshi et al. (2011), and Faccini et al. (2019) have recently proposed using the information from *index* option prices to predict REA and found that index option prices should be incorporated in the list of variables that predict economic growth. We investigate the predictive ability of the expected

[1]Typically, the expected returns are estimated using historical time series of assets return (Damodaran (2019)). However, this is not optimal since the historical approach is backwards-looking and only relies on the belief that the future will be similar to the past, which is not always the case. As outlined by Duan and Zhang (2014), investors are subject to higher uncertainty in the volatile phase- i.e., during a crisis. Hence, the forward-looking risk premium will increase. However, historical average returns cannot adequately reflect market conditions when such an increase in market volatility is temporary. During the crisis period, the realized stock returns are negative; therefore, the historical approach produces a negative risk premium.

returns of the individual equity options as opposed to index options. The information content of individual equity options may differ from that of index options, as they serve different purposes and trading patterns.[2] To the best of our knowledge, our paper is the first that investigates whether individual equity option prices may also convey useful information to predict REA.

Following Martin and Wagner (2019), we extract forward-looking expected returns on stocks from individual equity options. In this case, the expected return on a stock is a function of the risk-neutral variance of the market portfolio, the risk-neutral variance of the individual constituent stocks and the value-weighted average of the constituent stocks' risk-neutral variances.[3] We estimate all three quantities of risk-neutral variance from the volatility surface of individual equity options, exploiting thus the forward-looking information content of individual equity options.

This approach to estimating expected returns is distinct from more conventional methods, such as the risk premium estimates using historical excess returns, the CAPM and the Fama-French three-factor model.[4] First, it does not require accounting data. Second, it is based on observed market prices and can be implemented in real-time. Third, it is model-free and does not need any historical information to estimate its parameters. We focus our analysis on stocks that have been constituents of the S&P 500 index from July 1998 to May 2019. We compute the individuals' option-implied expected returns for different horizons (h =1, 2, 3, 6 and 12 months).

Once we compute the stocks' option-implied expected returns, we construct our option-based predictor by considering the Instrumented Principal Component analysis (IPCA)

[2]Bollen and Whaley (2004) document that trading in S&P 500 index options contributes to the hedging demand of investors- i.e., buying index puts as portfolio insurance against market declines. While Lakonishok et al. (2007) show that trading in stock options involves naked positions.

[3]We also compute the individual stocks' expected returns via Chabi-Yo et al. (2022)s' method, which takes into account the entire shape of risk-neutral density using higher moments of the risk-neutral distribution. The predictive power of the option-based expected returns predictor remains regardless of the method used to compute the option-implied expected returns.

[4]see Lewellen (2015) and Duan and Zhang (2014).

proposed by Kelly and Pruitt (2019). The IPCA considers individual stock characteristics as instrumental variables which help predict factor loadings and estimates factors which inherently depend on observable firms' characteristics.[5] Mapping the characteristics on loadings provide a link between expected returns and firms' characteristics. Therefore, IPCA enables us to accommodate the conditional information of stock characteristics into the expected returns.

Next, we investigate whether the first IPCA factor predicts future U.S. REA both in and out of the sample. We test the predictability of this factor over different forecasting horizons while controlling for well-known predictors of REA. To evaluate the robustness of our results, we employ eight different proxies of REA, namely industrial production (IP), non-farm payroll (NFP), retail sales (RS), housing starts (HS), capacity utilisation (CU), unemployment rate (UR), Chicago Fed National Activity Index (CFNAI) and Aruoba–Diebold–Scotti (ADS).

Our results from the in-sample analysis suggest that the first principal component of expected returns is a statistically significant predictor of REA, especially for the short-term horizons and contains information that other financial predictors have not already incorporated. We reject the null hypothesis of no predictability, even after considering the instrumental variable test of Kostakis et al. (2015). We also find a negative relationship between the first IPCA factor and future REA. This is in line with expected real returns being counter-cyclical due to consumption smoothing (e.g., Balvers et al. (1990); Chen (1991) and Loflund and Nummelin (1997)).

Next, we evaluate the predictive content of our options-implied IPCA factor in an out-of-sample (OOS) setting. Having constructed out-of-sample forecasts for each one of the REA proxies using an expanding window, we consider Campbell and Thompson (2008)'s

[5]The estimation objective in the IPCA method is similar to PCA- i.e., both estimate factors and loadings by concentrating on the common variation of the system of variables. The objective function in both models is to minimise the sum of squared model errors. However, the weight applied to each squared residual is static in PCA whereas it is time-varying and depends on the characteristics of each firm in IPCA.

out-of-sample R^2 as a metric of predictive ability. We compare our forecasts obtained from our predictive model (which contains our options-implied IPCA factor along with a set of standard predictors of REA) to a restricted model (which does not contain the option-implied factor in the set of its predictors). Our results suggest that the option-implied IPCA factor improves OOS forecasts and has predictive content over a set of standard REA predictors.

We further investigate the robustness of our findings by examining whether the option-implied IPCA factor outperforms (i) the previously established pricing factors, such as those from Fama and French (2015) and (ii) the index option-implied expected return. To address this, we compare the OOS forecasts of the full model with the IPCA factor and show that this model outperforms the restricted model, which contains the five risk factors of Fama and French (2015) or the market risk premium (SVIX) measure of Martin (2017). Our findings suggest that even though the derived IPCA factor may be interpreted as a "level" factor that proxies the market, individual stocks contain some additional predictive power that is not captured by the index expected return nor by standard factors that describe the cross-section of expected returns. This is in line with equity and index options being dissimilar and serving different purposes and trading patterns.[6]

The rest of the paper is organised as follows. Section 2 reviews the related literature. Section 3 describes our data. Section 4 describes the procedure we employ to extract the expected return factors. Section 5 explains our prediction models to forecast REA. Section 6 summarises our results. Section 7 reports our additional analysis results and Section 8 concludes.

[6] Lemmon and Ni (2014) document that trading activity in equity options is linked to individual investors' beliefs and past market returns, whereas a hedging demand motivates index options trades.

2 Related Literature

This paper relates to three strands of literature. First, it relates to papers that employ financial variables to estimate real economic activity. Several studies have found that the spread between the short and long-term maturity U.S. Treasury bills (term-spread) has a strong predictive relation to aggregate economic activity. (e.g., Harvey (1989), Chen (1991), Stock and Watson (1989), Friedman and Kuttner (1991), Estrella and Hardouvelis (1991)). Some papers document the predictive power of credit spreads to forecast REA (e.g., Bernanke (1983), Friedman and Kuttner (1991), Gilchrist and Zakrajsek (2012)). Ng and Wright (2013) find that term spreads were reasonable predictors of economic activity during 1970s and 1980s, whereas credit spreads show better predictive ability in recent years. Asset pricing factors have also been documents to contain significant information about future REA growth (Liew and Vassalou 2000). Other financial variables that have been found to predict REA include but are not limited to asset prices (Stock and Watson 2003), the Baltic dry index growth rate (Bakshi et al. 2012), commodity futures open interest (Hong and Yogo 2012), and firm-level price crashes (Kelly and Jiang 2014).

Another strand of literature deals with the estimation of expected returns. The average of historical realised returns are often used for estimating expected return. Damodaran (2019) documents that this method does not reflect transient rises in the market's volatility and produces a negative risk premium, particularly during a crisis when the realised stock return is negative. On the other hand, several studies, such as Welch (2000) and Fernández (2009) interviewed academics, traders or financial managers to obtain their opinions on expected returns. However, this approach is time-consuming, requires a very long prediction horizon, and faces sample selection bias.

Other studies extract the information embedded in option prices to estimate expected returns. For example, Santa-clara and Yan (2010) derive the risk premia as a function of volatility and jump intensity implied from options while Duan and Zhang (2014) derive it as a function of physical moments and option-implied risk aversions. Buss and Vilkov (2012) employ information extracted from equity and index option prices to estimate correlations and build option-implied betas. Using the implied betas, they derive linear factor models and hence expected returns. Martin (2017) derives a lower bound on the market premium in terms of simple risk-neutral variance (SVIX) that can be computed from index option prices. Ross (2015) and Schneider and Trojani (2019) estimate market return from state prices i.e., the product of risk aversion (the pricing kernel) and the risk-neutral probability distribution using the recovery theorem. These studies focus on the market return.

Extracting the information embedded in individual options, Martin and Wagner (2019) and Kadan and Tang (2019) propose a forward-looking lower bound on the expected rates of return on the individual assets. Their proposed lower bounds reflect idiosyncratic and systematic risk and rely on risk-neutral variances. Chabi-Yo et al. (2022) consider higher moments of the stocks' risk-neutral distribution to extract option-implied expected returns. González-Urteaga et al. (2021) combine the results of Martin (2017) and the literature on the stochastic discount factor and extract the expected returns of stocks.

We choose the method proposed by Martin and Wagner (2019) to extract individual stocks' expected returns using option prices. This estimation is computed from the observed current index and stock option prices and does not require any historical data. More importantly, this method is parsimonious in terms of the required inputs and can be computed precisely from the cross-section of market and individual option prices that are easily available. In addition, expected returns obtained from this method correspond to the market risk and indicate a significantly positive market risk premium. They are also align with several stock characteristics, including value, size, profitability, and momentum.[7]

Finally, our paper is also related to studies that have investigated the information content of option-implied variables. Surprisingly, there is very little literature on whether the option-implied information predicts REA. To the best of knowledge, Bakshi et al. (2011), Buss et al. (2019) and Faccini et al. (2019) are the only papers which explore this up to date.[8] Bakshi et al. (2011) show the forward variance extracted from S&P 500 index options predicts economic activity. Faccini et al. (2019) find that investor's implied relative risk aversion extracted from index option prices predicts REA. Buss et al. (2019) document that implied correlation acts as a leading procyclical state variable and negatively forecasts unemployment. We add to this literature by analysing the information embedded in options at the stock level, namely individual stocks expected returns.

3 Data

3.1 Option Data

We compute the individual stocks' expected returns via Martin and Wagner (2019) formula. This formula defines the expected return on a stock in terms of the risk-neutral

[7] In the robustness analysis, we compare the predictive ability of expected returns extracted from Martin and Wagner (2019) with those obtained from Chabi-Yo et al. (2022). Our results suggest that Chabi-Yo et al. (2022) 's method, which takes into account the higher moments of risk-neutral distribution, does not improve the predictive power of implied-expected returns.

[8] Other studies analyse the forecasting power of option-implied information for stock returns. For example, An et al. (2014) find that increases in implied volatilities of ATM put and call options have different indications, forecasting low and high future stock returns, respectively. Conrad et al. (2013) examine the relationship between RN moments and realised returns and find a negative, albeit not statistically significant. Xing et al. (2010) find that stocks with the steepest implied volatility smirks show worst earnings shocks in subsequent months. Cremers and Weinbaum (2010) show that the deviation from put-call parity arising from expensive puts possesses information about future prices of underlying stocks and predicts abnormally negative returns. Rehman and Vilkov (2012) and Stilger et al. (2016) show that the RN skewness is positively related to future stock return. Gkionis et al. (2021) document that high RNS, arising from expensiveness OTM calls relative to OTM puts, contains positive information regarding the underlying stock. Other studies focus on options trading activities. For example, Pan and Poteshman (2006) show a negative relationship between the put-call ratio and future stock returns. Hu (2013) find that the stock exposure imbalance generated by option trading is positively related to next-day stock returns. Johnson and So (2012) find that a low volume trading in options markets compared to the stock market predicts low stock performance. Kostakis et al. (2011) show that the information option-implied distributions can predict market timing.

variance of the market, the risk-neutral variance of the individual stock and the value-weighted average of stocks' risk-neutral variance. These measures are computed directly using observed prices of options on index and individual stocks. We collect the required variables from OptionMetrics Ivy DB (OM) via the Wharton Research Data Services (WRDS) and Compustat. Our sample spans January 1996 to May 2019.

We first obtain information on the S&P500 index and its constituents from Compustat. Using the lists of index constituents, we searched the OptionMetrics database for the entire universe of firms that formed the S&P 500 index during our sample period. Where available, we obtain the time series of implied volatility surface data for these individual firms from the Optionsmetrics volatility surface (VS) file. The VS file contains implied volatilities for standardised equity options for standard maturities- i.e., 30, 60, 91, 182, and 365 calendar days, on a delta grid, accompanied by implied strike prices. We remove options that their deltas or implied volatilities are missing. Moreover, we collect daily index and stock prices as well as number of outstanding shares of each firm from OptionMetrics's underlying price file to compute their market capitalisations. We collect forward prices for each firm from OM standardised option price file. In addition, Zero yield curve data and projected dividend yield are also obtained from the OM Zero Curve File and Index Dividend Yield file.

Using equity prices and volatility surface data, we compute the risk-neutral variance (SVIX) of all individual firms that have been constituents of S&P 500 index from 1996 to 2019 for horizons of one, two, three, six, and twelve months. From these time series, on each day, we only focus on the SVIX of firms that have been included in SPX and calculate the value-weighted average of the risk-neutral variance of the market. As summarised in Panel A of Table 1, we calculate almost 2.5 million individual stocks's SVIX for each horizon; We cover 1,092 firms over our sample period from January 1996 to May 2019. For each horizon, we collect data for 482 firms on average per day, suggesting that we capture slightly more than 96% of the firms included in the S&P 500 index.

[Table 1 about here.]

3.2 Real Economic Activity Data

We obtain monthly data on seven different measures U.S. REA proxies from the Federal Reserve Bank of St. Louis (FRED) for a sample period from July 1998 to May 2019. First, we use Industrial Production (IP), a measure which quantifies the output of the industrial sector of the economy (manufacturing, mining, and utilities). Second, we use on Nonfarm payrolls (NFP) which is the number of employees in the nonfarm sectors in the U.S. economy. Third, we consider real retail sales (RS), defined as the consumer demand for finished goods and is measured by the purchases of durable and non-durable goods over a specific period. Fourth, we employ housing starts (HS), defined as the total of new private-owned houses. Fifth, we use capacity utilisation (CU). CU is the portion of resources in corporations, and factories employed to produce goods in manufacturing, mining, and gas and electric utilities in the United States. Sixth, we consider the unemployment rate (UR), which is the number of unemployed people as a ratio of the labour force. The latter is the sum of unemployed and those in paid or self-employment. Our seventh proxy for the U.S. REA is the Chicago Fed National Activity Index (CFNAI), which is a monthly index created to measure overall economic activity and related inflationary pressure.

Finally, we obtain monthly data on our last U.S. REA proxy from the Philadelphia Fed webpage from July 1998 to May 2019, namely the Aruoba–Diebold–Scotti (ADS) Business Conditions index (Aruoba et al. 2009). This tracks real business conditions at high frequency and the six economic indicators underlying the ADS index are: monthly payroll employment, weekly initial jobless claims, personal income fewer transfer payments, manufacturing and trade sales, industrial production, and quarterly real GDP.

We calculate the h-month overlapping log growth rates of IP, NFP, HS, RS, CU and UR, $REA_{t \rightarrow t+h,i} = lnREA_{t+h,i} - lnREA_{t,i}$, for $h = 1, 2, 3, 6, 12$ months and $i = 1$ for IP, 2 for NFP, 3 for RS, 4 for HS, 5 for CU, and 6 for UR. By construction, the values of

CFNAI and ADS denote growth or recession and therefore, we do not need to calculate the growth rates for these two proxies.

3.3 Control Variables

We obtain data on various variables that have been studied by previous literature and have documented to predict REA. We use these variables as control variables in our predictive regression models. First, we consider lagged REA growth $REA_{t \to t+h,i}$ over a period of h-monthsfor $i = 1, ...6$ and the lagged REA for $i = 7$ (CFNAI), 8 (ADS). Second, we obtain monthly treasury bond and bill rates from the FRED website to measure the term spread (TS), as the difference between the 10-year Treasury bond rate and the three-month Treasury bill rate. Third, we consider credit spread (CR) as a control proxied by the difference between the monthly the Moody's AAA and BAA corporate bonds yields obtained from FRED.

Fourth, we use OTM options data to calculate the value of the forward variance $FV_{t \to t+1}$ between t and $t+1$. We follow Bakshi et al. (2011) and calculate $FV_{t,t+1}$ as follows:

$$H_{(t,t+1)} = e^{-r(h/12)} + \int_{K<S_t} \omega(K) Put_t(K) dK + \int_{K>S_t} \omega(K) Call_t(K) dK \quad (1)$$

where $C_t(K)$ and $P_t(K)$ are the prices of a call and put options on SPX index, respectively, at time t, with time to expiry (h) month(s) and strike price K. Moreover, $e^{r(h/12)}$ is the price at time t of a riskless discount bond with unit face value and time-to-maturity of h month(s).

$$\omega(K) = -\frac{\frac{8}{\sqrt{14}} cos(arctan(1/\sqrt{7}) + \frac{\sqrt{7}}{2} ln(\frac{K}{S_t})}{\sqrt{S_t} K^{3/2}} \quad (2)$$

We define the forward variance at time t between t and $t+1$, as:

$$FV_{t,t+1} = -lnH(t, t+1) \tag{3}$$

Our fifth control variable is investor's implied relative risk aversion (IRRA). We extract IRRA using Kang et al. (2010) formula:

$$\frac{\sigma_{p,t}^2(\tau) - \sigma_{q,t}^2(\tau)}{\sigma_{q,t}^2(\tau)} \approx \gamma \sigma_{q,t}(\tau)\theta_{q,t}(\tau) + \frac{\gamma^2}{2}\sigma_{q,t}^2(\tau)(k_{q,t}(\tau) - 3) \tag{4}$$

where γ is the coefficient of the relative risk aversion of the representative agent. The formula estimates the spread between the risk-neutral variance ($\sigma_{q,t}^2(\tau)$) and physical variance ($\sigma_{q,t}^2(\tau)$) as a function of the investor's relative risk aversion (γ) by considering a power utility function. In this formula $\theta_{q,t}(\tau)$ and $k_{q,t}(\tau)$ are the RN skewness and kurtosis, respectively. We calculate the S&P 500 risk-neutral moments with τ =1-month horizon following Bakshi et al. (2003) formulae (see Appendix A). Following Faccini et al. (2019), we apply the generalised method of moments employing a 30-months rolling window. Provided that our options data set is from January 1996 to May 2019, we obtain the IRRA time series from July 1998 to May 2019.

The left axis of Figure 1 depicts the movements of forward variance, and the right axis displays the changes of the U.S. implied risk aversion (IRRA) over July 1998 to May 2019. Two remarks are in order in case of IRRA. First, IRRA values span from 3 to 6.5. This is similar to range of IRRA computed by the previous literature. For example, Faccini et al. (2019) report IRRA values ranging from 2.27 to 9.55. Second, we can see that the U.S. IRRA and forward variance rises substantially in 2008 during the financial crisis and begin declining after that. These patterns could result from the quantitative easing monetary policy implemented by the Fed from 2008 to 2014.

[Figure 1 about here.]

We consider five factors of Fama-French (5FF) as our sixth set of control variables.[9] These factors include the size of firms (SMB), book-to-market values (HML), expected return on the market (mrkt), profitability factor (RMW) and investment factor (CMA). These factors capture the different dimensions of systematic risk and are expected to explain the cross-section of stock's average returns with a linear relationship. We mainly use these factors to assess whether the sources of systematic risks in individual stocks' returns are adequate to predict the REA or whether the forward-looking option-implied risk factors contain additional predictive information.

Finally, we consider the market risk premium (SVIX) measure of Martin (2017) as a control variable. We extract this at the end of each month by using volatility surface data from OptionMetrics.

4 Constructing the Option-Implied Predictor

In this section, we discuss the procedure to construct our option-implied predictor. We follow Martin and Wagner (2019) formula, which derive the expected return on a stock in terms of the stock's excess risk-neutral variance and the risk-neutral variance of the market. We then apply instrumented principal component analysis (IPCA) to produce our proposed option-implied predictor.

4.1 Formula for the Individual's Expected Return

The inputs to the Martin and Wagner (2019) formula are the following three measures of risk-neutral variance:

[9] We obtain 5FF from https://mba.tuck.dartmouth.edu

$$\begin{aligned} SVIX^2_{m,t} &= var^*_t(\frac{R_{m,t+1}}{R_{f,t+1}}) \\ SVIX^2_{i,t} &= var^*_{i,t}(\frac{R_{i,t+1}}{R_{f,t+1}}) \\ \overline{SVIX}^2_t &= \Sigma_i \omega_{i,t} SVIX^2_{i,t} \end{aligned} \tag{5}$$

where $R_{m,t+1}$ is the market gross return from time t to $t+1$, $R_{i,t+1}$ is the gross return on stock i from time t to $t+1$ and $R_{f,t+1}$ is the gross riskless rate from time t to $t+1$. The length of the period from time t to time $t+1$ depends on the forecasting horizon of interest-i.e., h month(s). $SVIX^2_{m,t}$ denotes the measure of market risk-neutral variance, $SVIX^2_{i,t}$ is the RN variance at the stock level and $\overline{SVIX}^2_t$, measures the average risk-neutral stock volatility.

Martin (2017) claims that the $SVIX^2_{m,t}$ index can forecast the equity premium. The prices of index options determine this measure:

$$SVIX^2_{m,t} = \frac{2}{R_{f,t+1} S^2_{m,t}} [\int_0^{F_{m,t}} Put_{m,t}(K) dK + \int_{F_{m,t}}^{\infty} Call_{m,t}(K) dK] \tag{6}$$

where $S_{m,t}$ denotes the price of the market portfolio. $F_{m,t} = R_{f,t+1}(S_t - \bar{D}_{m,t})$ is the forward price of stock as of time t for delivery at time $t+1$ and $\bar{D}_{m,t}$ is the the present value of dividends paid between times t and $t+1$. $Put_{m,t}(K)$ and $Call_{m,t}(K)$ are the time t prices of European index call and put options with strike prices K for delivery at time $t+1$.

The corresponding SVIX measure at the individual stock level is obtained using individual stock option prices:

$$SVIX^2_{i,t} = \frac{2}{R_{f,t+1} S^2_{i,t}} [\int_0^{F_{i,t}} Put_{i,t}(K) dK + \int_{F_{i,t}}^{\infty} Call_{i,t}(K) dK] \tag{7}$$

where the subscripts i denote that the reference asset is stock i.

Finally, using $SVIX_{i,t}^2$ for all firms in the index available at time t and $\omega_{i,t}$ be the market-capitalization weight of stock i in the index, we compute the risk neutral average stock variance index as $\overline{SVIX}_t^2 = \Sigma_i \omega_{i,t} SVIX_{i,t}^2$

We compute the three measures of risk-neutral variance for horizons (i.e., option maturities) of 1, 3, 6, 12, and 24 months. We then remove a few extreme outliers in our data that do not meet the monotonicity property of $SVIX_{i,t}$ across horizons described above[10].

Then, for each optionable stock in our sample, we construct the expected returns in excess of the market following Martin and Wagner (2019).

$$\frac{\mathbb{E}_t\, R_{i,t+1} - R_{m,t+1}}{R_{f,t+1}} = \frac{1}{2}(SVIX_{i,t}^2 - \overline{SVIX}_t^2) \tag{8}$$

Exploiting the results of Martin (2017), specifically, $E_t R_{m,t+1} - R_{f,t+1} = R_{f,t+1} SVIX_{m,t}^2$, we have:

$$\frac{\mathbb{E}_t\, R_{i,t+1} - R_{f,t+1}}{R_{f,t+1}} = SVIX_{m,t}^2 + \frac{1}{2}(SVIX_{i,t}^2 - \overline{SVIX}_t^2) \tag{9}$$

Finally, we define the expected return of stock i as:

$$\mathbb{E}_t\, R_{i,t+1} = R_{f,t+1} \times [SVIX_{m,t}^2 + \frac{1}{2}(SVIX_{i,t}^2 - \overline{SVIX}_t^2)] + R_{f,t+1} \tag{10}$$

[10]If the underlying asset does not pay dividends, call prices are rising with time-to-maturity. Assuming this is not cancelled by the opposing effect of increased interest rates $R_{f,t+1}$ over longer horizons, $SVIX_{i,t}$ should be expected to be monotonic in horizon length. In the daily data, we come up with 2,837,127 firm-day observations after withdrawing 21,438 observations based on nonmonotonicity. In the case of end-of-the-month data, we end up with 135,343 observations after removing 1,012 observations based on nonmonotonocity

4.2 Instrumented Principal Component Analysis

Kelly and Pruitt (2019) proposed a novel technique for data reduction called instrumented principal components analysis or IPCA. This approach introduces observable firms' characteristics as instrumental variables and brings the informational content of characteristic to the factor models through latent time-varying loadings.[11]

The general IPCA model specification is a system comprising N assets with L firms' characteristics over T periods. We assume that the expected return on asset i from time t to $t+1$ ($r_{i,t+1} = \mathbb{E}_t R_{i,t+1}$) computed from Equation (10) maps to a factor model proposed by Kelly and Pruitt (2019) as follows:

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1} \tag{11}$$

where f_{t+1} is a $(K \times 1)$ vector of K latent factors and inherently depends on lagged observed asset characteristics incorporated in the $(L \times 1)$ instrument vector $z_{i,t}$. The specification of $\beta_{i,t}$ is the main feature of the IPCA analysis and has two roles. First, it uses observable characteristics which serve as instrumental variables in the estimation of the latent factor loadings. Second, the time-varying instruments facilities the estimation of dynamic factor loadings. $\beta_{i,t}$ is estimated following the equation below:

$$\alpha_{i,t} = z'_{i,t}\Gamma_\alpha + \nu_{\alpha,i,t} \text{ and } \beta_{i,t} = z'_{i,t}\Gamma_\beta + \nu_{\beta,i,t} \tag{12}$$

The matrix Γ_β is a $(L \times K)$ vector and maps a large number of characteristics to a few risk factors. To find the Γ_β and the factor, we use the vector form of Equations (13) and (12)[12]:

$$r_{t+1} = Z_t \tilde{\Gamma} \tilde{f}_{t+1} + \epsilon_{t+1} \tag{13}$$

[11]Characteristics are factors that provide independent information about average returns.

[12]We consider the unrestricted model of Kelly and Pruitt (2019) in which $\Gamma_\alpha \neq 0$.

where r_{t+1} is an $(N \times 1)$ vector of individual firm returns, Z_t is the $(N \times L)$ matrix that stores the lagged characteristics of each firm, and ϵ^*_{t+1} is the residuals of individual firms. $\tilde{\Gamma} = [\Gamma_\alpha, \Gamma_\beta]$ and $\tilde{f}_{t+1} = [1, f_{t+1}]$. The estimation goal is to minimise the sum of squared model errors:

$$min_{\tilde{\Gamma},\tilde{f}} \sum_{t=1}^{T-1} (r_{t+1} - Z_t \tilde{\Gamma} \tilde{f}_{t+1})'(r_{t+1} - Z_t \tilde{\Gamma} \tilde{f}_{t+1}) \tag{14}$$

The values of IPCA latent factor and Γ that minimize the above equation will satisfy the first order conditions:

$$f_{t+1} = (\Gamma'_\beta Z'_t Z_t \Gamma_\beta)^{-1} \Gamma'_\beta Z'_t (r_{t+1} - Z_t \Gamma_\alpha), \tag{15}$$

and

$$vec(\tilde{\Gamma}') = \left(\sum_{t=1}^{T-1} [Z_t \otimes \tilde{f}'_{t+1}]'[Z_t \otimes \tilde{f}'_{t+1}] \right)^{-1} \left(\sum_{t=1}^{T-1} [Z_t \otimes \tilde{f}'_{t+1}]' r_{t+1} \right) \tag{16}$$

where $\tilde{\Gamma}$ is $(L \times (K+1))$ vector and is the coefficients of mapping the L instruments to the K factor loadings. We consider $K = 1$ in our analysis. These equations suggest that the IPCA factor is achieved by dynamic cross-section regression of r_{t+1} on the latent loading matrix β_t. Similarly, Γ_β, is the regression coefficient obtained from regressing returns on factors that interact with firm characteristics. There is no closed-form solution for the system of first-order conditions above, and therefore, must be solved numerically. To this end, following Kelly and Pruitt (2019) we choose an initial guess for $\tilde{\Gamma}$ equal to the second moment matrix, $\sum_t x_{t+1} x'_{t+1}$, where $x_{t+1} = Z'_t r_{t+1}$ is the time t observed returns on a set of L managed portfolios. The l^{th} component of x_t is a weighted average of stock returns with weights specified by the value of l^{th} characteristic of each firm at time t.

Provided with starting guess for $\tilde{\Gamma}$, we estimate the least squares regression based on the first-order condition (15) for all t. Next, having the consequent solutions for f_{t+1}'s, we estimate the least squares regression with respect to first-order condition (16). We repeat between estimations of (15) and (16) until convergence. The convergence is described as the point at which the maximum absolute change in any element of $\tilde{\Gamma}$ for $\tilde{f}_{t,T}$ (for all t) is smaller than 10^{-6}.

We empirically instrument the estimation of the first IPCA factor with observable characteristics data of the firms in our sample. Specifically, we focus on 125 firms in our sample that have constantly been included in the SPX index. Our sample spans July 1996 - May 2019. For each firm, we calculate 33 characteristics. These characteristics are defined exactly following Freyberger et al. (2020) and include market beta (beta), assets-to-market (a2me), total assets (assets), sales-to- assets (ato), book-to-market (bm), cash-to-short-term- investment (c), ratio of change in property, capital turnover (cto), capital intensity (d2a), earnings-to-price (e2p), fixed costs-to-sales (fc2y), cash flow-to-book (freecf), idiosyncratic volatility with respect to the FF3 model (idiovol), investment (invest), leverage (lev), market capitalization (mktcap), turnover (turn), net operating assets (noa), plants and equipment to the change in total assets (dpi2a), operating accruals (oa), operating leverage (ol), price-to-cost margin (pcm), profit margin (pm), gross profitability (prof), Tobin's Q (q), price relative to its 52-week high (w52h), return on net operating assets (rna), return on assets (roa), return on equity (roe), momentum (mom), intermediate momentum (intmom), short-term reversal (strev), long-term reversal (ltrev), sales-to-price (s2p), the ratio of sales and general administrative costs to sales (sga2s), bid-ask spread (bidask), and unexplained volume (suv). Section C in the appendix describes these characteristics and their construction.

Figure 2 shows the evolution of the first factor obtained from the first factor obtained from the IPCA model, as well as the SVIX measure. The first IPCA factor is highly correlated with SVIX, yet the correlation is not perfect. This suggests that the informational content of the IPCA factor extracted from equity options and SVIX extracted from index options may differ. Specifically, the correlation between SVIX and IPCA factor is 0.86%.

[Figure 2 about here.]

5 Predicting REA

Stock and Watson (2003) argue that the price of a stock is equal to the expected discounted value of future earnings, therefore, stock returns are useful variables for predicting the output growth. We address this argument in this section. Specifically, we investigate whether the option-implied IPCA factor extracted from the stock level expected returns predicts the subsequent REA growth.

5.1 Empirical Setup

To identify whether stocks' $h-$month(s) expected returns predict REA growth over h forecasting horizons (h = 1, 2, 3, 6, 12 months), we consider the following predictive regression:

$$REA_{t\to t+h,i} = \beta_{i,0} + \beta_{i,1}REA_{t-h\to t,i} + \beta_{i,2}F_{t,h} + \beta'_{i,3}X_t + \epsilon_{i,t+h} \tag{17}$$

where $REA_{t\to t+h,i}$ is the growth rate of the $i-th$ REA proxy ($i = 1$ for IP, 2 for NFP, 3 for RS, 4 for HS, 5 for CU, 6 for UR, 7 for CFNAI and 8 for ADS) over the period t to $t+h$. $F_{t,h}$ is the first IPCA factor using h-month(s) expected returns. X_t is a (4×1) vector of additional predictors at time t and includes credit spread, term spread, forward variance and IRRA. $\beta_{1,i}$, $\beta_{2,i}$ and $\beta'_{3,i}$ are (1×1), (1×1) and (1×4) vectors of regression coefficient parameters respectively and the ith REA proxy is the dependent variable.

In our further analysis, we also analyse the predictive power of our proposed predictor in the presence of common risk factors, which explain the cross-section of expected returns, namely the Fama and French (2015) five-factor model. This will tell us if the additional workload of extracting option-based equity returns adds some value compared to simply using asset pricing factors to predict REA. To test this, we add Fama and French (2015) five factors to the set of our control variables mentioned above and evaluate the predictive ability of our predictor. We fully discuss this model in Section 7.1.

Among different economic concerns that our predictive model may have, we first checked for the correlation matrix of the explanatory variables employed in our regression models. The covariance matrices of our control variables are reported in Table 2. As indicated in Table 2, the first IPCA factors obtained from h-month horizon(s) expected returns have a high correlation with forward variance and credit spread.

[Table 2 about here.]

To alleviate multicollinearity concerns that arise from these correlations, we orthogonalise variables that have correlations bigger than 0.5 with the h-month IPCA factor, namely forward variance and credit spread. More specifically, we regress these variables on a constant and the h-month first factor and obtain the residuals of these regressions. These residuals represent the part of the control variables orthogonal (uncorrelated) with the factor. Then, we run our predictive regressions using the first factor and residual terms as predictors.

Secondly, to consider the possible econometric concerns arising from small samples or time-series properties of the regressors, we report the p-values for the regression coefficients in two ways: (i) Two-sided p-values of Newey and West (1994) which considers heteroscedasticity and autocorrelation of error terms and (ii) p-values of the IVX-Wald test of Kostakis et al. (2015), which considers time-series properties of predictors. The IVX-Wald test does not presume an initial assumption for the degree of persistence and allows different persistence levels for predictors. We describe the details of IVX estimator and test in Appendix D.1 and D.2.

6 Results

6.1 In-Sample Evidence

Table 3 shows the results of estimating Equation (17) for forecasting horizons $h = 1$; 2; 3; 6; and 12 months employing the first IPCA factor in our predictive regression model. To recall, our predictive model considers lagged REA, credit and term spread, forward and IRRA in the set of its control variables. This table reports the standardised ordinary-least-squares (OLS) coefficient estimates of the h-month(s) first IPCA factor, the Newey-West and IVX-Wald test p-values of the regressor and the adjusted R^2 of the regression. We reject the null hypothesis of a zero coefficient (no predictability) based on the p-values of the IVX-Wald test at the 1%, 5%, and 10% levels and denote it by triple, double, and single asterisks, respectively.

We can obtain three key findings from Table 3. First, our results suggest that the first IPCA is a significant predictor of most REA proxies at the short-term horizon (i.e., 1 and 2 months horizon). Specifically, the first IPCA factor predicts all but one REA proxies (i.e., RS) at a one-month horizon. It also predicts four out of eight REA proxies, including IP, CU, UR, and ADS, at a 2-months horizon. At three months horizon, the IPCA factor also predicts UR and ADS. Second, we could not find predictive power of option-implied expected return for horizons greater than six months. Third, the sign of the first IPCA factor coefficient is negative in all cases except for the unemployment rate. UR is a counter-cyclical measure of economic growth, and a positive regression coefficient of the first IPCA factor also explains this property. These results suggest that a rise in expected return forecasts a decrease in REA.

The negative relation between expected returns and REA is in an agreement with the prediction of consumption-based asset pricing models such as Fama and French (1989) and Campbell and Cochrane (1999), which propose that the risk premia and, hence, the expected returns are high during recessions; investors become more risk-averse and require higher premiums to hold stocks during recessions.[13]

[Table 3 about here.]

6.2 Out-of-Sample Evidence

This section assesses the forecasting ability of the IPCA factor in a real time out-of-sample (OOS) setting from October 2007 to May 2019. This period is interesting because it covers the global financial crisis over 2007 to 2009 and includes the explosion of the U.S. housing bubble. It also includes the substantial monetary policies led by the Fed.

We construct the IPCA factor from $h-$month(s) option-implied expected returns. Employing an expanding window, we recursively estimate Equation (17).[14] Our first training sample window covers July 1998 to September 2007. Next, at each month (t), we construct $h = 1, 2, 3, 6, 12-$months ahead forecasts. For every month $t \geq$ September 2007, we employ entire data through t to estimate the regression model (17) which includes our proposed factor[15] in the set of its predictor. Based on (17), we define our full-forecasting model which contains the first IPCA factor in the set of predictors:

$$\mathbb{E}_t(REA^{full}_{t\to t+h,i}) = b_{0,i} + b_{i,1}REA_{t-h\to t,i} + b_{i,2}F_{t,h} + b'_{i,3}X_t \tag{18}$$

where $\mathbb{E}_t(REA^{full}_{t\to t+h})$ denotes the $h-$ months ahead out-of-sample forecasts from the full-forecasting model, $b_{1,i}$, $b_{2,i}$ and $b'_{3,i}$ are the (1×1), (1×1) and (1×4) vectors of estimated regression coefficients, respectively, $F_{t,h}$ is the IPCA factor obtained from h-months expected returns. Similar to the in-sample analysis, X_t is the vector of control variables and consists of credit spread, term spread, forward variance and IRRA. Clearly,

[13] Our results remain unchanged when we use risk premia of stocks instead of expected returns. This suggests that the effect of the risk-free rate is negligible.

[14] We also perform the analysis using a rolling window, and our results remain unchanged.

[15] obtained recursively from h-month(s) expected returns from July 1998 to $t - h$

$\mathbb{E}_t(REA^{full}_{t\to t+h,i})$ forecasts does not rely on the information beyond time t.

We also estimate regression (17) in a restricted form -i.e., with models which do not contain our proposed factor within the set of predictors. Specifically, we consider two alternative model specifications for benchmark (restricted) forecasts:

1. $benchmark_1$: This is an $AR(1)$ model which considers only lagged REA as the predictor of REA.

$$E_t(REA^{benchmark_1}_{t\to t+h,i}) = b_{0,i} + b_{1,i}REA_{t-h\to t,i} \tag{19}$$

2. $benchmark_2$: This model is in the form of:

$$E_t(REA^{benchmark_2}_{t\to t+h,i}) = b_{0,i} + b_{1,i}REA_{t-h\to t,i} + b_{i,2}X_t \tag{20}$$

where X_t includes the credit and term spread, forward variance, and IRRA.

Next, we compute the out-of-sample R^2 measure following Campbell and Thompson (2008) to examine the OOS forecasting performance of our proposed estimator (individual stocks expected returns IPCA factor). The out-of-sample R^2 indicates whether the variance explained by forecasts of the full model, which includes the IPCA factor in the set of its predictors, is more or less than the variance explained by alternative benchmark models ($benchmark_m$, $m = 1, 2$) which do not contain IPCA factor in the set of its predictors. We define the out-of-sample $R^2_{i,h}$ acquired from predicting the ith REA proxy as:

$$R^2_{i,h} = 1 - \frac{var[E_t(REA^{Full}_{t\to t+h,i}) - REA_{t\to t+h,i}]}{var[E_t(REA^{benchmark_m}_{t\to t+h,i}) - REA_{t\to t+h,i}]} \tag{21}$$

A positive out-of-sample R^2 indicates that the full model performs better than the benchmark model; therefore, our proposed factor has superior OOS predictive ability.

Table 4 reports the out-of-sample R^2 obtained from Equation (21) where the full model includes the IPCA factor, respectively.[16] Panels A and B of these tables report results when we consider the $AR(1)$ model in Equation (19) and the nested model in Equation (20), respectively.

Table 4 shows that the out-of-sample R^2 is mostly positive when we consider the IPCA factor in our model. This suggests that the full regression model with the IPCA factor outperforms the restricted models. Therefore, incorporating the IPCA factor in REA prediction models is statistically significant in an out-of-sample setting. More precisely, we cans see in Panel A for the case of the AR(1) model, the out-of-sample R^2 is positive in all proxies except for three cases. Exceptions occur at the 2, 3 and 6 months horizons for CFNAI. In addition, Panel B of this table suggests that the full model with the IPCA factor beats the nested model ($benchmark_2$) for NFP, HOUS, UR and ADS at a one-month horizon.

[Table 4 about here.]

Our findings on OOS analysis complement our results on in-sample analysis and suggest that our proposed equity-level factor contains additional information in predicting REA. This information cannot be captured by other option-implied measures such as IRRA and FV, especially at the short-term horizon. Therefore, the inclusion of our predictor is statistically significant in forecasting REA and the existing list of REA predictors should be extended by option-implied equity factors.

[16]Our results remain robust if we calculate the OOS-R^2 measure using the sum of squared errors instead of the variance.

7 Further Analysis

7.1 Option-Implied Factors versus Equity Factors

The IPCA model is an alternative way to explain the variation in expected returns. A natural question is whether or not the option-implied IPCA factor outperform the traditional pricing factors such as those from Fama and French (2015). We answer this question by extending our analysis to both in- and out-of-sample settings.

First, we assess the predictive performance of option-implied factor in an in-sample setting. To this end, we augment the control variables in the predictive regression model (17) with 5 Fama-French factors (5FF) proposed by Fama and French (2015). More precisely, our new predictive model considers a constant, lagged REA, the extracted first IPCA factor from h-month ($h = 1, 2, 3, 6$ *or* 12 *-months*) expected returns, credit and term spread, forward variance and IRRA plus five factors of Fama-French including the size of firms (SMB), book-to-market values (HML), expected return on the market (mrkt), profitability factor (RMW) and investment factor (CMA).[17] A significant regression coefficient suggests that our proposed forward-looking stock-level option-implied risk factor contains additional predictive information relative to the 5FF.

Table 5 reports the in-sample regression coefficients of the IPCA factor. This table suggests that our results are robust, and the IPCA factor has predictive power even after controlling for the five Fama-French factors. We can see that at a one-month horizon the IPCA factor predicts 7 out of 8 REA proxies, including IP, NFP, HOUS, CU, UR, CFNAI and ADS. At 2-months horizon, the IPCA factor remains predictive for four proxies. However, at three months horizon, the IPCA factor predict only UR. These results highlight the forward-looking nature of options and show that the expected returns implied from options contain additional information over the systematic risk

[17] Fama-French factors capture the different dimension of systematic risk in stock prices and are expected to explain the cross-section of stock's average returns with a linear relationship.

factors obtained from the stock market at short horizons. This extra information is useful in predicting REA, especially at short-term horizons.

[Table 5 about here.]

Second, we compare the performance of our proposed predictor (option-implied stocks' expected returns IPCA factor) with five factors of Fama-French (5FF) in an out-of-sample setting. To this end, we define the benchmark predictive regression model ($benchmark_3$) which does not contain the IPCA factor and instead contain 5FF:

$$E_t(REA_{t\to t+h,i}^{benchmark_3}) = b_{0,i} + b_{1,i}REA_{t-h\to t,i} + b_{i,2}FF5_t + b'_{i,3}X_t \tag{22}$$

where X_t includes the credit and term spread as well as the forward variance and IRRA.

We compare the forecasts of our full predictive model explained in Equation (18) with forecasts of Equation (22) and compute the OOSR^2 following Equation (21). Table 6 Panel A reports the OOS results and suggest that the predictive models with our proposed predictor outperform sthe alternative model with 5 Fama-French factors in most of the cases. More specifically, the full model with IPCA outperforms the alternative model with 5FF factors for NFP, RS, HOUS, CFNAI and ADS at all horizons. Our findings imply that although the derived IPCA factors may be interpreted as "level" factor that proxies for the market, individual stock options contain some additional predictive power that is not captured by standard risk factors.

7.2 Individual Equity Options versus Index Options

Lemmon and Ni (2014) suggest that individual equity options and index options are dissimilar because they attract different types of investors and therefore react differently to risk factors. For instance, equity options are actively traded by individual investors and is positively related to investors' sentiment, while index option tradings are mostly motivated by the hedging demands of sophisticated traders. Therefore, the information extracted from index options maybe different from the information implied from a particular stock option.

In this section, we compare the predictive performance of our proposed predictor (option-implied stocks' expected returns IPCA factor) with the index expected return (SVIX) in an out-of-sample setting. Doing so, we examine whether the option-implied expected return IPCA factor contains predictive power in forecasting REA growth that cannot be captured by info obtained from index options. To this end, we define an alternative predictive regression model ($benchmark_4$), which does not contain the IPCA factor and contains SVIX instead. More specifically, we compare the forecasts of our full predictive model explained in Equation (18), which includes the expected return IPCA factor [18], with the forecast obtained from the following model:

$$E_t(REA^{benchmark_4}_{t\to t+h,i}) = b_{0,i} + b_{1,i}REA_{t-h\to t,i} + b_{i,2}SVIX_t + b'_{i,3}X_t \tag{23}$$

We compute the OOSR^2 following Equation (21) and report the results in Table 6. Panel B of Table 6 reports the OOS results when the full models includes the IPCA factor, and the benchmark model contains the SVIX measure (Equation 23). Panel B of the table shows that the comparative performance of a predictive model with IPCA factor is better than the predictive model with the market SVIX measure for NFP, HOUS, UR, CFNAI and ADS at one-month horizon.

Our findings imply that the individual stocks' option-implied standard IPCA factors contain some useful information in predicting REA, especially at short-term horizon, that the index options cannot capture. Therefore, the list of REA predictors should be extended by individual's option-implied IPCA factor.

[Table 6 about here.]

[18] Augmented by a set of control variables (X_t)- namely, lagged REA, credit spread, term spread, forward variance and IRRA

8 Conclusion

Campbell and Cochrane (1999) suggests that expected return is an appropriate state variable for forecasting output growth and is countercyclical. In addition, Fama and French (1989) suggest that the risk premia and, hence, the expected returns are high during recessions -i.e., investors become more risk-averse and require higher premiums to hold stocks during recessions.[19] Due to the econometric challenges in estimating expected returns, previous literature has not examined its predictive ability. We bypass these obstacles by employing methods which use information from option prices. Doing so, we exploit the information embedded in equity option prices and investigate the predictive ability of the expected returns of the individual stocks extracted from the equity option prices to forecast the growth of U.S. REA. As expected by theory, we find a significant predictive ability of our proposed option-implied factor for forecasting REA both in and out-of-sample.

To build our proposed predictor, first, we follow Martin and Wagner (2019) to compute option-implied expected returns on stocks that have been constituents of the S&P500 index from July 1998 to May 2019. We then employ a dimension reduction technique, namely instrumented principal component analysis, to construct expected return factor. We document that our expected return factor is a new predictor of U.S. real economic activity (REA) and has significant predictive power even after controlling for a well-known set of variables that has been documented by the previous literature to predict REA- namely, term spread credit spread, forward variance and IRRA. Our results are robust after addressing some econometric concerns, such as dealing with the unknown persistence degree of predictors and heteroscedasticity and auto-correlation of error terms. In addition, our findings remain robust even after contorting for the systematic risk factors of Fama and French (2015) in the in-sample analysis and imply that individual expected

[19]The rationale is that as business conditions and investment opportunities change, expected returns change because investors attempt to smooth consumption over multiple periods, which generally predict a countercyclical risk premium.

returns contain some predictive power in addition to the standard systematic risk factors.

We also examine the out-of-sample performance of our proposed predictive model, which contains the expected return factor within the set of its predictor. Our findings suggest that incorporating the expected return factors in predictive regression models is also statistically significant in an out-of-sample setting. In the case of forecasts obtained by the regression models, the evidence is relatively weaker for longer horizons. We also document that the OOS forecasts obtained from a predictive model with our proposed predictor are better than the OOS forecasts obtained from benchmark models, which contain the market risk premium (SVIX) measure of Martin (2017) or the five systematic risk factors of Fama and French (2015). Our findings imply that individual stocks contain some additional predictive power that is not captured by the index expected return nor by standard factors that explain the cross-section of expected returns.

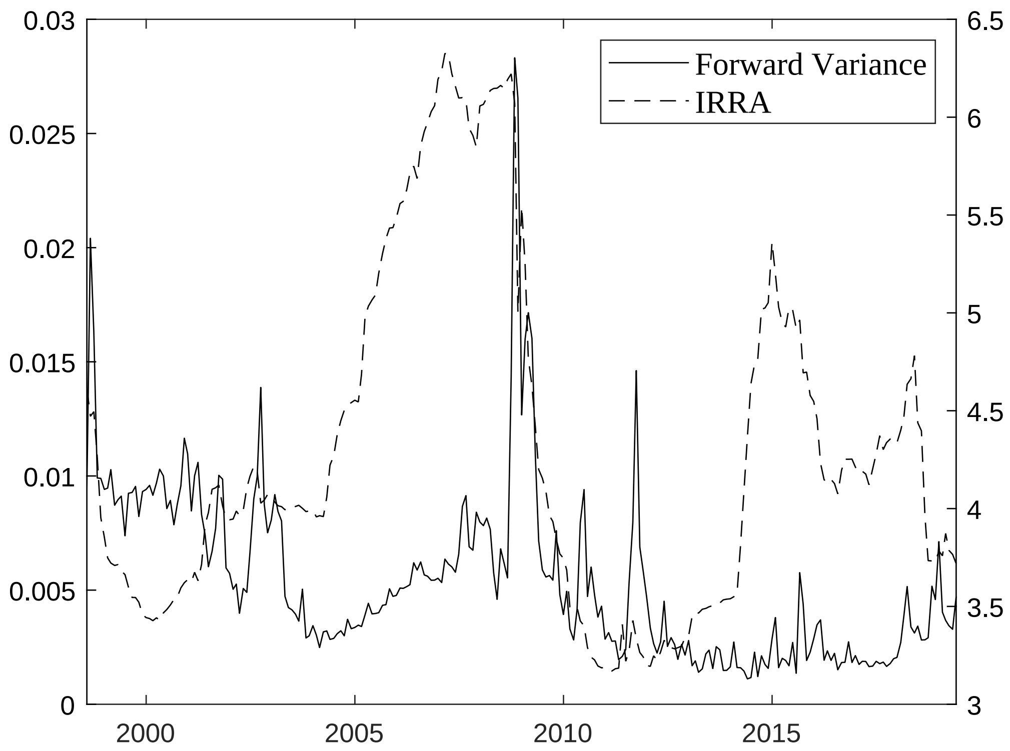

Figure 1. Evolution of the U.S. Forward Variance and Implied Risk Aversion (IRRA): The Right axis shows the evolution of IRRA while the left axis shows the evolution of forward variance over July 1998 to May 2019. We compute the IRRA time series following Faccini et al. (2019) by performing a generalised method of moments (GMM) using a rolling window estimation. We employ a rolling window with a size of 30 months to obtain the respective U.S. IRRA time series. Forward variance is constructed using index options prices as explained in section 3.3

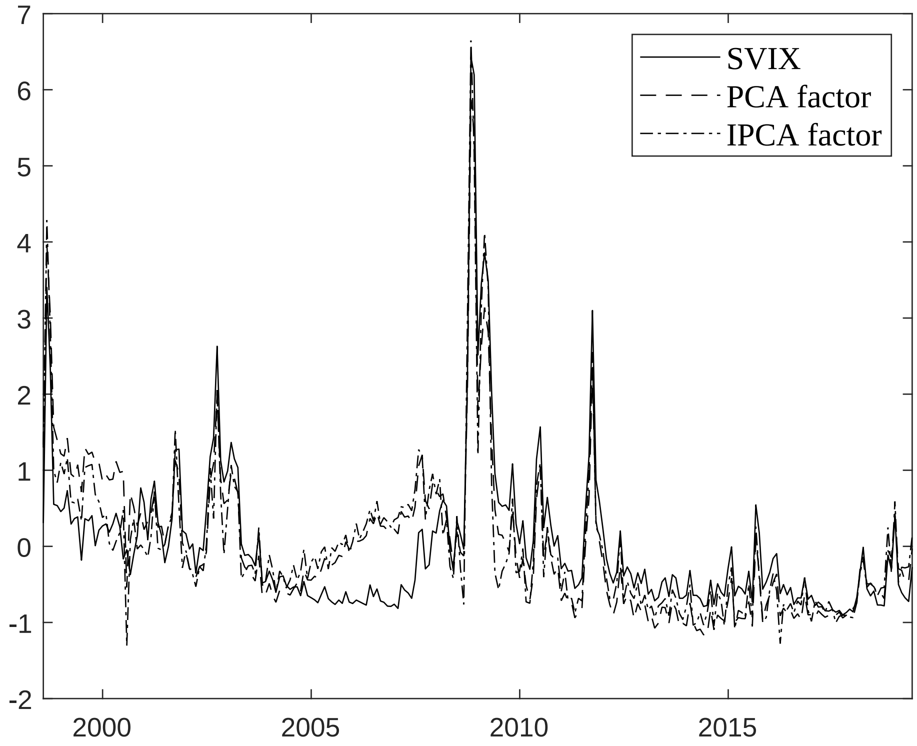

Figure 2. Evolution of the SVIX measure and the IPCA factor: The figure shows the evolution of SVIX and the first IPCA factor obtained from expected returns over July 1998 to May 2019. We extract the SVIX time series via Martin (2017) formula using volatility surface data. We construct the IPCA factor as described in section 4.2.

Panel A: Daily Data for SPX firms					
horizon	30	60	90	182	365
Observations	2,837,127	2,837,127	2,837,127	2,837,127	2,837,127
Sample days	5892	5892	5892	5892	5892
Sample firms	1092	1092	1092	1092	1092
Average firms/day	482	482	482	482	482
Panel A: Monthly Data for SPX firms					
horizon	30	60	90	182	365
Observations	135,343	135,343	135,343	135,343	135,343
Sample months	281	281	281	281	281
Sample firms	1092	1092	1092	1092	1092
Average firms/month	482	482	482	482	482

Table 1. Sample Data: This table reports a summary of the data used in the empirical analysis. We searched the OptionMetrics database for all firms included in the S&P 500 from January 1996 to May 2019 and obtained all available volatility surface data. Using these data, we compute firms' risk-neutral variances ($SVIX^2_{i,t}$) for horizons of 1, 2, 3, 6 and 12 months. Panel A reports the total number of observations, the number of unique days and firms in our sample, and the average number of firms for which options data are available per day. Panel B reports monthly observations subsets for constituents of the S&P 500.

Panel A - First IPCA factor

	TS	CS	mrkt	SMB	HML	RWM	CMA	FV	SVIX	IRRA
1-months Factor	-0.15	0.52	-0.07	0.17	0.24	-0.03	0.02	0.92	0.87	0.25
2-months Factor	-0.18	0.52	-0.06	0.17	0.22	-0.03	0.00	0.92	0.85	0.25
3-months Factor	-0.22	0.51	-0.06	0.19	0.21	-0.05	0.01	0.93	0.83	0.24
6-months Factor	-0.27	0.46	-0.04	0.20	0.15	-0.09	0.00	0.89	0.75	0.22
12-months Factor	-0.30	0.39	-0.02	0.21	0.11	-0.12	-0.02	0.86	0.68	0.19

Panel B - Control variables

	TS	CS	mrkt	SMB	HML	RWM	CMA	FV	SVIX	IRRA
TS	1.00	0.16	-0.01	-0.09	0.02	0.02	-0.04	-0.15	0.15	-0.38
CS	0.16	1.00	0.09	0.07	0.15	-0.08	-0.07	0.51	0.70	0.15
mrkt	-0.01	0.09	1.00	-0.04	-0.02	-0.34	-0.43	-0.06	-0.08	-0.02
SMB	-0.09	0.07	-0.04	1.00	-0.04	-0.29	0.03	0.22	0.16	-0.01
HML	0.02	0.15	-0.02	-0.04	1.00	-0.06	0.44	0.24	0.25	0.17
RWM	0.02	-0.08	-0.34	-0.29	-0.06	1.00	0.17	-0.11	-0.05	0.02
CMA	-0.04	-0.07	-0.43	0.03	0.44	0.17	1.00	0.04	0.03	-0.04
FV	-0.15	0.51	-0.06	0.22	0.24	-0.11	0.04	1.00	0.90	0.14
SVIX	0.15	0.70	-0.08	0.16	0.25	-0.05	0.03	0.90	1.00	-0.01
IRRA	-0.38	0.15	-0.02	-0.01	0.17	0.02	-0.04	0.14	-0.01	1.00

Table 2. Correlation coefficient between predictors: Panel B reports the correlation coefficients between the first IPCA factor extracted from the h months (i.e., h =,1, 2, 3, 6, 12) individual's expected returns and a set of control variable, namely: term spread (TS), credit spread (CR), forward variance (FV) and Fama and French 5 factors (market, SMB, HML, RMW and CMA). Panel B shows the correlation coefficient between control variables.

		$h = 1$ month	$h = 2$ months	$h = 3$ months	$h = 6$ months	$h = 12$ months
IP	β	-0.28***	-0.26**	-0.12	0.06	0.05
		[0.00]	[0.03]	[0.41]	[0.94]	[0.82]
		(0.01)	(0.01)	(0.04)	(0.47)	(0.65)
	R2	0.20	0.32	0.42	0.41	0.36
NFP	β	-0.24***	-0.15	-0.14	-0.06	-0.14
		[0.00]	[0.11]	[0.27]	[0.64]	[0.43]
		(0.00)	(0.01)	(0.03)	(0.36)	(0.16)
	R2	0.67	0.78	0.79	0.75	0.57
RS	β	-0.01	0.07	0.05	0.19	0.05
		[0.70]	[0.37]	[0.48]	[0.50]	[0.88]
		(0.84)	(0.26)	(0.62)	(0.14)	(0.71)
	R2	0.09	0.17	0.26	0.31	0.36
HOUS	β	-0.17**	-0.09	-0.04	-0.17	-0.13
		[0.02]	[0.21]	[0.79]	[0.51]	[0.91]
		(0.10)	(0.37)	(0.67)	(0.06)	(0.11)
	R2	0.21	0.19	0.18	0.36	0.57
CU	β	-0.31***	-0.28**	-0.13	-0.04	-0.15
		[0.00]	[0.02]	[0.34]	[0.71]	[0.38]
		(0.00)	(0.01)	(0.02)	(0.61)	(0.20)
	R2	0.22	0.35	0.44	0.39	0.32
UR	β	0.34**	0.34**	0.32**	0.12	0.17
		[0.00]	[0.01]	[0.04]	[0.65]	[0.47]
		(0.00)	(0.00)	(0.00)	(0.25)	(0.16)
	R2	0.19	0.34	0.43	0.58	0.53
CFNAI	β	-0.33***	-0.13	-0.02	0.04	-0.11
		[0.00]	[0.22]	[0.66]	[0.99]	[0.83]
		(0.00)	(0.03)	(0.74)	(0.63)	(0.43)
	R2	0.48	0.49	0.44	0.30	0.14
ADS	β	-0.08**	-0.15**	-0.11*	0.06	-0.12
		[0.01]	[0.01]	[0.09]	[0.83]	[0.82]
		(0.02)	(0.01)	(0.02)	(0.50)	(0.35)
	R2	0.86	0.73	0.65	0.44	0.22

Table 3. In-sample prediction of REA proxies with the first IPCA factor: This table presents the in-sample estimated regression coefficients of the first IPCA factor for various U.S. real economic activity (REA) proxies. We matched the forecasting horizon with the horizon of expected returns. For each REA under our consideration, we consider predictive regression model 17 as explained in Section 5 and employ the lagged REA, term-spread, credit-spread, forward variance and IRRA in the set of our control variables. Reported entries are the standardized ordinary-least-squares (OLS) coefficient estimates, IVX-Wald (within squared brackets) and Newey-West (within brackets) p-values of each one of the predictors as well as the in-sample adjusted R^2 for the given model. The sample covers July 1998 to May 2019. We reject the null hypothesis of a zero coefficient (no predictability) based on the p-values of the IVX-Wald test at the 1%, 5%, and 10% levels and denote it by triple, double, and single asterisks, respectively.

Panel A: $OOS - R^2$: full model with IPCA factor versus AR(1)								
	IP	NFP	RS	HOUS	CU	UR	CFNAI	ADS
$h = 1$ month	0.20	0.20	0.01	0.03	0.13	0.26	0.27	0.10
$h = 2$ months	0.17	0.17	0.12	0.08	0.11	0.29	-0.02	0.21
$h = 3$ months	0.10	0.21	0.20	0.11	0.02	0.30	-0.07	0.14
$h = 6$ months	0.13	0.28	0.14	0.27	0.06	0.18	-0.09	0.07
$h = 12$ months	0.24	0.26	0.32	0.39	0.12	0.26	0.11	0.15

Panel B: $OOS - R^2$: full model with IPCA factor versus the nested model								
	IP	NFP	RS	HOUS	CU	UR	CFNAI	ADS
$h = 1$ month	-0.04	0.02	-0.05	0.01	-0.03	0.01	-0.01	0.04
$h = 2$ months	0.01	0.05	0.04	0.01	-0.01	0.02	-0.01	0.07
$h = 3$ months	0.01	0.10	0.08	0.02	0.01	0.04	0.04	0.06
$h = 6$ months	0.06	0.12	0.01	0.01	0.05	-0.02	0.01	0.06
$h = 12$ months	0.04	0.08	-0.02	-0.01	0.02	-0.05	-0.01	0.01

Table 4. Out-of-sample predictability of U.S. REA with IPCA factor: Panel A shows the out-of-sample R^2 obtained from the forecasts of the full model in Equation (18) (which contains IPCA factor, lagged REA and the control variables as predictors) versus the forecasts of the AR(1) model in Equation (19) (which considers only lagged REA as predictors). Panel B shows out-of-sample R^2 obtained from the forecasts of the full model versus the forecasts of the nested model ($benchmark_2$) in Equation (20). The nested model does not include the IPCA factor and considers only lagged REA and the control variables in the set of its regressors. Control variables in the full and nested model include credit spread, term spread, forward variance and IRRA.
For each U.S. REA proxy, we estimate the relevant form of Equation (17) for the respective full and benchmark models by considering an expanding window; the first estimation sample window covers July 1998 to September 2007. At each point in time, we estimates $h =$ 1-,3-, 6-, 9-, and 12-months-ahead REA forecasts. The positive (negative) sign of the out-of-sample R^2 denotes that the full model that incorporates the PCA factor in the set of its predictors outperforms (underperforms) the benchmark model.

		$h=1$ month	$h=2$ months	$h=3$ months	$h=6$ months	$h=12$ months
IP	β	-0.27***	-0.26**	-0.12	0.04	0.03
		[0.00]	[0.02]	[0.34]	[0.92]	[0.52]
		(0.00)	(0.01)	(0.03)	(0.55)	(0.78)
	R2	0.19	0.32	0.42	0.41	0.37
NFP	β	-0.25***	-0.15	-0.14	-0.08	-0.15
		[0.00]	[0.12]	[0.20]	[0.50]	[0.50]
		(0.00)	(0.01)	(0.03)	(0.20)	(0.05)
	R2	0.68	0.78	0.79	0.76	0.58
RS	β	-0.01	0.07	0.04	0.18	0.02
		[0.71]	[0.35]	[0.60]	[0.51]	[0.56]
		(0.83)	(0.26)	(0.70)	(0.17)	(0.88)
	R2	0.08	0.17	0.26	0.31	0.38
HOUS	β	-0.17**	-0.09	-0.04	-0.19	-0.14
		[0.01]	[0.17]	[0.63]	[0.34]	[0.88]
		(0.05)	(0.31)	(0.60)	(0.02)	(0.06)
	R2	0.21	0.20	0.20	0.35	0.56
CU	β	-0.30***	-0.28**	-0.13	-0.05	-0.17
		[0.00]	[0.01]	[0.27]	[0.63]	[0.34]
		(0.00)	(0.00)	(0.02)	(0.57)	(0.18)
	R2	0.20	0.35	0.45	0.39	0.33
UR	β	0.34***	0.35**	0.33**	0.15	0.20
		[0.00]	[0.01]	[0.03]	[0.52]	[0.48]
		(0.00)	(0.00)	(0.00)	(0.10)	(0.06)
	R2	0.18	0.35	0.44	0.59	0.55
CFNAI	β	-0.33***	-0.12	-0.03	0.02	-0.16
		[0.00]	[0.17]	[0.53]	[0.84]	[0.63]
		(0.00)	(0.02)	(0.67)	(0.80)	(0.13)
	R2	0.48	0.51	0.46	0.30	0.19
ADS	β	-0.07**	-0.14**	-0.11	0.05	-0.15
		[0.02]	[0.04]	[0.11]	[0.82]	[0.95]
		(0.04)	(0.01)	(0.03)	(0.55)	(0.17)
	R2	0.86	0.73	0.66	0.44	0.23

Table 5. In-sample prediction of REA proxies with IPCA factor in the presence of 5 Fama-French factors. This table presents the in-sample estimated regression coefficients of the first IPCA factor for various U.S. real economic activity (REA) proxies. We matched the forecasting horizon with the horizon of expected returns. For each REA under our consideration, we consider the predictive regression model (17) with the IPCA factor plus a set of control variables which includes the lagged REA, credit and term spread, forward and IRRA and is augmented by 5 factors of Fama-French including the size of firms (SMB), book-to-market values (HML), expected return on the market (mrkt), profitability factor (RMW) and investment factor (CMA). Reported entries are the standardized ordinary-least-squares (OLS) coefficient estimates, IVX-Wald (within squared brackets) and Newey-West (within brackets) p-values of each one of the predictors as well as the in-sample adjusted R^2 for the given model. The sample covers July 1998 to May 2019. We reject the null hypothesis of a zero coefficient (no predictability) based on the p-values of the IVX-Wald test at the 1%, 5%, and 10% levels and denote it by triple, double, and single asterisks, respectively.

Panel A: $OOSR^2$ from model with IPCA factor versus a model with 5 FF factors								
	IP	NFP	RS	HOUS	CU	UR	CFNAI	ADS
h = 1 month	-0.01	0.09	0.01	0.01	-0.01	0.01	0.03	0.15
h = 2 months	-0.02	0.09	0.06	0.02	-0.02	-0.02	0.02	0.13
h = 3 months	-0.02	0.14	0.11	0.03	-0.02	0.05	0.02	0.04
h = 6 months	0.09	0.13	0.06	0.03	0.07	-0.01	0.08	0.12
h = 12 months	0.10	0.10	0.08	0.05	0.07	-0.02	0.03	0.12

Panel B: $OOSR^2$ from model with IPCA factor versus a model with market svix								
	IP	NFP	RS	HOUS	CU	UR	CFNAI	ADS
h = 1 month	-0.05	0.04	-0.09	0.01	-0.06	0.01	0.01	0.05
h = 2 months	0.00	0.05	0.01	-0.06	-0.02	-0.02	0.01	0.05
h = 3 months	0.00	0.09	0.03	-0.02	-0.01	-0.04	0.04	0.03
h = 6 months	0.05	0.06	-0.09	-0.09	0.02	-0.15	-0.01	0.01
h = 12 months	-0.03	-0.03	-0.09	-0.11	-0.09	-0.25	0.02	0.00

Table 6. Individual equity options versus standard risk factors: Panel A shows the OOS-R^2 obtained from the full predictive model explained in Equation (18) which contain the IPCA factor versus the benchmark model which contains the 5 Fama-French factors Equation (22). Panel B reports OOR^2 obtained from the full model with the IPCA factor versus an alternative predictive model which contains the SVIX measure Equation (23). All predictive models are augmented by a set of control variables which consists of the lagged REA, credit spread, term spread, forward variance and IRRA.
For each U.S. REA proxy, we estimate the relevant form of Equation (17) for the respective full and benchmark models by considering an expanding window; the first estimation sample window covers July 1998 to September 2007. At each point in time, we estimates h = 1-,3-, 6-, 9-, and 12-months-ahead REA forecasts. The positive (negative) sign of the out-of-sample R^2 denotes that the full model that incorporates the IPCA factor in the set of its predictors outperforms (underperforms) the benchmark model.

A Computation of the Risk-Neutral Moments

We compute the S&P500 risk-neutral moments following the methodology of Bakshi et al. (2003) using the volatility surface data of S&P500 index options. At time t the τ period risk-neutral skewness ($\theta_{q,t}(\tau)$) and kurtosis ($kq, t(\tau)$) of the log-return $R(t, \tau)$ distribution with horizon τ are given by

$$\theta_{q,t}(\tau) = \frac{e^{r\tau} M(3)_{t,T} - 3e^{r\tau} \mu_{t,T} M(2)_{t,T} + 2\mu_{t,T}^3}{[e^{r\tau} M(2)_{t,T} - \mu_{t,T}^2]^{3/2}} \tag{A.1}$$

$$k_{q,t}(\tau) = \frac{e^{r\tau} M(4)_{t,T} - 4e^{r\tau} \mu_{t,T} M(3)_{t,T} + 6e^{r\tau} \mu_{t,T}^2 M(2) - 3\mu_{t,T}^4}{[e^{r\tau} M(2)_{t,T} - \mu_{t,T}^2]^2} \tag{A.2}$$

where $M(n)_{t,T}(n = 2, 3, 4)$ is given by:

$$M(n)_{t,T} = \int_{S_t}^{\infty} \eta(K, S_t, n) C_t(K, T) dK + \int_0^{S_t} \eta(K, S_t, n) P_t(K, T) dK \tag{A.3}$$

$$\eta(K, S_t, n) = \frac{n}{K^2} [(n-1) log(\frac{K}{S_t})^{n-2} - log(\frac{K}{S_t})^{n-1}] \tag{A.4}$$

and

$$\mu_{t,T} = e^{r\tau} - 1 - e^{r\tau} [M(2)_{t,T}/2 + M(3)_{t,T}/6 + M(6)_{t,T}/24] \tag{A.5}$$

r is the riskless rate. S_t denotes the price of the asset at time t and $P_t(K, T)$ and $C_t(K, T)$ are the OTM put and call prices expiring at T.

B SVIX Empirical Computation

A spectrum of option prices with respect to the strike price is needed to compute the integrals in Equation (5). However, the required continuum of option prices is not available in empirical applications, and we have to make some choices regarding the implementation. Firstly, we follow Conrad et al. (2013) and estimate SVIX for days that a stock has more than two OTM calls and two OTM puts with the same maturity. Moreover, following Dennis and Mayhew (2002), we employ equal numbers of OTM puts and calls for each asset on each day. Therefore, If there are only n OTMP available on day t, we need only n OTMC prices and if $N > n$ OTMC prices are available on the day t, we only use the n OTMC that are the least out-of-the-money. We also discard a few day-stock IV surface data if their OptionsMetrics's strike prices are not monotonic in deltas.

Next, we follow Jiang and Tian (2005) to impose a structure for implied volatilities. To do so, we define moneyness level as the implied strike prices provided by OM, divided by the stock price (K/S_t).

Since the SVIX formula requires only OTM-forward option prices, we follow Christoffersen et al. (2012) and keep only calls (puts) with moneyness that are greater (less) than or equal to the forward-moneyness (F/S_t). Then, we sort the puts and calls in increasing order by moneyness:

$$\frac{K_1^P}{S_t} < \frac{K_2^P}{S_t} < \frac{K_3^P}{S_t} < \dots < \frac{K_{np}^P}{S_t} < \frac{F}{S_t} < \frac{K_1^C}{S_t} < \frac{K_2^C}{S_t} < \frac{K_3^C}{S_t} < \dots < \frac{K_{nc}^C}{S_t},$$

where nc (np) denote the number of call (put) options.

Then, we interpolate implied volatilities using cubic piece-wise Hermite polynomial interpolation in the moneyness-IV metric to obtain a continuum of implied volatilities on a fine grid of 1,001 implied volatilities for moneyness levels between 1/3 and 3. Cubic piece-wise Hermite polynomial interpolation is adequate for interpolating between the available moneyness levels. Therefore, for values outside the available ranges, we use the implied volatility of the lowest strike price (K_1^P) for moneyness below the lowest available moneyness level. We use the implied volatility of the highest strike (K_{nc}^C) for moneyness above the highest available moneyness.

Once the continuous Implied Volatility surface is formed, we compute the BS options prices at 1,001 strike points (equally-spaced) over the moneyness from 1/3 to 3. Moneyness levels smaller than $(F_{t,T}/St)$ are used to generate OTM put prices and moneyness levels greater than $(F_{t,T}/St)$ are used to generate call prices. In addition, to convert IVs to option prices, we derive the projected dividend payments $D_{t,T}$ from forward price as $D_{t,T} = R_{f;t;T}S_t - F_{t;T}$. Next, we deduct the present value of $D_{t,T}$ from the stock price at time t and replace it in the modified Black-Scholes formula.

Finally, we follow Christoffersen et al. (2012) we employ trapezoidal approximation rule to numerically approximate the integrals of Equation (7). More specifically, on each day and for each maturity we assume that:

$$\int_0^\infty Q_{t,T}(K)dK = \Sigma_{i=1}^N \frac{Q_{t,T}(K_{i+1}) + Q_{t,T}(K_i)}{2}(K_{i+1} - Ki) \tag{B.6}$$

where $K_1, ..., K_N$ are the strikes of observable options, $Q_{t,T}(K_i)$ is the mid bid-ask price of an OTM option with strike K_i.[20]

C Characteristic Data

In this section, we define characteristic variables used in the IPCA method. We use CRSP and Compustat variable names in parentheses and report the relevant references. All these definitions[21] are directly obtained from Freyberger et al. (2020). We mainly use balance-sheet data from the fiscal year ending in year $t-1$ for returns from July of year t to May of year $t+1$ following the Fama and French (2015) timing convention.

1. "**A2ME**: Following Bhandari (1988), we define assets-to-market cap as total assets (AT) over market capitalisation as of December t-1. Market capitalisation is shares

[20]In Equation (B.6), we use put prices when their strike prices are less than the forward price and call prices when their strikes are greater than or equal to the forward price.

[21]The definition of all 33 characteristic variables are directly brought from Freyberger et al. (2020). However, we compute all measures ourselves.

outstanding (SHROUT) times price (PRC). AT Total assets (AT) as in Gandhi and Lustig (2015).

2. **ATO**: Following Soliman (2008), ATO is Net sales over lagged net operating assets. Net operating assets are operating assets minus operating liabilities. Operating assets are total assets (AT) minus the investment, minus cash and short-term investments (CHE), and other advances (IVAO). Operating liabilities are total assets (AT) minus long-term debt (DLTT), minus common equity (CEQ), minus debt in current liabilities (DLC), minus preferred stock (PSTK), minus minority interest (MIB).

3. **BEME**: BEME is the ratio of the book value of equity to the market value of equity. Book equity is deferred taxes and investment tax credit (TXDITC) plus shareholder equity (SH) minus preferred stock (PS). SH is shareholders' equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for PS. The market value of equity is as of December t-1. The market value of equity is shares outstanding (SHROUT) times price (PRC). See Rosenberg et al. (1985) and James L. Davis and French (2000).

4. **Beta**: We define the CAPM beta as correlations between the expected return of stock i and the market expected return times volatilities following Frazzinia and Pedersen (2014). We calculate correlations from overlapping three-day log expected returns over a five-year period requiring at least 750 non-missing observations. We estimate volatilities using the standard deviations of daily log expected returns over a one-year horizon requiring at least 120 observations.

5. **C**: Following Palazzo (2012), cash to short-term investment is defined as cash and

short-term investments (CHE) divided by total assets (AT).

6. **CTO**: Capital turnover is defined as the ratio of net sales (SALE) to lagged total assets (AT) following Haugen and Baker (1996).

7. **D2A**: Capital intensity is the ratio of depreciation and amortization (DP) to total assets (AT) as in Gorodnichenko and Weber (2016).

8. **DPI2A**: Following Evgeny Lyandres and Zhang (2008), the change in property, equipment, and plants is defined as the changes in property, plants, and equipment (PPEGT) and inventory (INVT) over lagged total assets (TA).

9. **E2P**: Following Basu (1983), we define earnings to price as the ratio of income before extraordinary items (IB) to the market capitalisation as of December t-1. Market capitalisation is the shares outstanding (SHROUT) times price (PRC).

10. **FC2Y**: Following D'Acuntoa et al. (2018), Fixed costs to sales are defined as the ratio of selling, general, and administrative expenses (XSGS), advertising expenses (XAD), and research and development expenses (XRD) to net sales (SALE).

11. **Free CF**: Cash flow to book value of equity is the ratio of depreciation and amortisation (DP), net income (NI) less change in capital expenditure (CAPX) and working capital (WCAPCH), divided by the book value of equity defined as in the construction of BEME following Houa et al. (2012).

12. **Idio vol**: Following Ang et al. (2006), Idiosyncratic volatility is the standard deviation of the residuals from a regression of expected returns on the Fama and French (1993) three-factor model. We use one month of daily data and require at least fifteen non-missing observations.

13. **investment**: Investment is the percentage year-on-year growth rate in total assets (AT) following Cooper et al. (2008).

14. **Lev**: Following Lewellen (2015), we define Leverage as the ratio of long-term debt (DLTT) and debt in current liabilities (DLC) to the sum of long-term debt, debt in current liabilities, and stockholders' equity (SEQ).

15. **LME**: Following Fama and French (1992), Size is shares outstanding (SHROUT) times the total market capitalisation of the previous month defined as price (PRC).

16. **LTurnover**: Following Datar et al. (1998), lagged Turnover is last month's volume (VOL) divided by shares outstanding (SHROUT).

17. **NOA**: Net operating assets are the difference between operating assets minus operating liabilities scaled by lagged total assets as in Hirshleifer et al. (2004). Operating assets are total assets (AT) minus the investment and other advances (IVAO) minus cash and short-term investments (CHE). Operating liabilities are total assets (AT), minus long-term debt (DLTT), minus debt in current liabilities (DLC), minus preferred stock (PSTK), minus minority interest (MIB), minus common equity (CEQ).

18. **OA**: Following Sloan (1996), we define operating accruals as changes in non-cash working capital minus depreciation (DP) scaled by lagged total assets (TA). Non-cash working capital is the difference between non-cash current assets and current liabilities (LCT), income taxes payable (TXP) and debt in current liabilities (DLC). Non-cash current assets are current assets (ACT) minus cash and short-term investments (CHE).

19. **OL**: Operating Leverage is defined as the sum of administrative expenses (XSGA) and cost of goods sold (COGS) over total assets as in Novy-Marx (2011).

20. **PCM**: The price-to-cost margin (PCM) is the difference between net sales (SALE) and costs of goods sold (COGS) divided by net sales (SALE) as in Bustamant and Donangelo (2017).

21. **PM**: Profit margin is defined as the ratio of operating income after depreciation (OIADP) to net sales (SALE) as in Soliman (2008).

22. **Prof**: Following Balla et al. (2015) Profitability is defined as gross profitability (GP) divided by the book value of equity.

23. **Q**: Tobin's Q is total assets (AT), the market value of equity (SHROUT times PRC), minus deferred taxes (TXDB) scaled by total assets (AT) minus cash and short-term investments (CEQ).

24. **Rel to High**: Closeness to 52-week high is the stock price (PRC) at the end of the previous calendar month divided by the previous 52-week high price following George and Hwang (2004).

25. **RNA**: The return on net operating assets is operating income after depreciation divided by lagged net operating assets following Soliman (2008). Net operating assets are the difference between operating assets minus operating liabilities. Operating assets are total assets (AT) minus the investment and other advances (IVAO) minus cash and short-term investments (CHE). Operating liabilities are total assets (AT) minus common equity (CEQ), minus minority interest (MIB), minus debt in current liabilities (DLC), minus preferred stock (PSTK), minus long-term debt (DLTT).

26. **ROA**: Return-on-assets is income before extraordinary items (IB) to lagged total assets (AT) following Balakrishnan et al. (2010).

27. **ROE**: Return-on-equity is income before extraordinary items (IB) to lagged book-value of equity as in Haugen and Baker (1996). r_{12-2} : We define momentum as cumulative return from 12 months before the return prediction to two months before, as in Fama and French (1992).

28. r_{12-7}: The intermediate momentum is a cumulative return from 12 months before the return prediction to seven months before, as in Novy-Marx (2012).

29. r_{2-1} : Short-term return is the lagged one-month return as in Jagadeesh (1990).

30. r_{36-13} : Long-term return is the cumulative return from 36 months before the return prediction to 13 months before, as in Bondt and Thaler (1985).

31. **S2P**: Sales-to-price is the ratio of net sales (SALE) to the market capitalisation as of December following Lewellen (2015).

32. **SGA2S**: SG&A to sales is the ratio of selling, general and administrative expenses (XSGA) to net sales (SALE).

33. **Spread**: The bid-ask spread is the average daily bid-ask spread in the previous months as in Jagadeesh (1990)."

D IVX-Wald test

D.1 IVX estimator

Assume that the following multivariate system of predictive regressions consists of regressors with arbitrary degree of persistence:

$$y_t(K) = \mu + Ax_{t-1} + \epsilon_t \tag{D.7}$$

where $y_t(K) = \Sigma_{i=0}^{K-1} y_{t+i}$, A is a $(m \times r)$ coefficient matrix and

$$x_{t+1} = R_n x_t + u_{t+1} \tag{D.8}$$

where $x_t = (x_{1t}, x_{2t}, ..., x_{rt})$ is the vector of predictors in (D.7), $R_n = I_r + C/n^{\alpha}$ is the autoregressive matrix with degree of persistence equal to $\alpha \geq 0$, $C = diag(c1, ..., cr)$ and n is the sample size.

The IVX method proposed by Kostakis et al. (2011) estimates Equation (D.7) using two-step least square equations which uses the near-stationary instruments in z_t rather than the initial predictors x_t. The intuition behind it is to build an instrumental variable with a known degree of persistence using the initial predictors of x_t with an unknown degree of persistence. To construct the IVX estimator, first, we estimate Equations (D.7) and (D.8) using ordinary least squares. Next, we build the near-stationary instruments Z_t by differencing the regressor x_t, initializing at $z_0 = 0$. More specifically, IVX construct instruments using a first-order autoregressive process with the autoregressive artificial matrix of R_{nz} and innovations Δx_t as follows:

$$z_t = R_{nz} z_{t-1} + \Delta x_t \tag{D.9}$$

where $R_{nz} = I_r + \frac{C_z}{n^\beta}$. We select $C_z = -I_r$ and $\beta = 0.95$ following Kostakis et al. (2015) .

Once we construct our instrument variable, we apply standard instrumental variable estimation and continue with IVX estimation of A using regression system in (D.7)

$$\bar{A}_{IVX(K)} = \bar{Y}(K)'Z[\bar{X}(K)'Z'] = \sum_{r=1}^{n}(y_t - \bar{y}_n)\bar{z}'_{t-1}[\sum_{j=1}^{n}(x_j - \bar{x}_{n-1})\bar{z}'_{j-1}]^{-1} \tag{D.10}$$

where $\bar{y}_n = (1/n)\sum_{t=1}^{n} y_t$, $\bar{x}_{n-1} = (1/n)\sum_{t=1}^{n} x_{t-1}$, $\bar{Y} = (Y'_1, ..., Y'_n)$ and $\bar{X} = (X_0, ..., X'_{n-1})$ are the predictive regression matrices which are demeaned and Z_{n_k} is the instrument matrix.

D.2 IVX-Wald test

The asymptotic feature of the IVX method suggests that linear restrictions on the coefficients A rendered by the system of predictive regression (D.7) can be examined by a standard Wald test using the IVX estimator. We test for the predictive ability of x_t with the following null hypothesis:

$$H_0 : Hvec(A) = 0 \tag{D.11}$$

where $vec(A)$ is the vectorisation of A and H is a known $r \times r$ matrix, which its (i, i) element is one and its other elements are zero. This helps to test for the significance of each regressor individually.

We consider the following IVX-Wald test statistic for testing the H_0

$$W_{IVX} = (Hvec\bar{A}_{IVX)})'Q^{-1}(Hvec\bar{A}_{IVX)}) \tag{D.12}$$

where $vec\bar{A}_{IVX}$ is the vectorization of IVX estimator in (D.10) and:

$$Q_H = H[(\bar{Z}'X)^{-1} \otimes I_m]\mathbb{M}[(X'\bar{Z})^{-1} \otimes I_m]H' \tag{D.13}$$

$$\mathbb{M} = \bar{Z}'\bar{Z} \otimes \hat{\Sigma}_{\epsilon\epsilon} - n\bar{z}_{n-1}\bar{z}'_{n-1} \otimes \hat{\Omega}_{FM} \tag{D.14}$$

$$\hat{\Omega}_{FM} = \hat{\Sigma}_{\epsilon\epsilon} - \hat{\Omega}_{\epsilon u}\hat{\Omega}_{uu}^{-1}\hat{\Omega}'_{\epsilon u} \tag{D.15}$$

Denoting $\hat{\epsilon}_t$ the OLS residuals from regression (D.7) and $\hat{u}_t$ residuals from regression (D.8), the covariance matrices $\hat{\Sigma}_{\epsilon\epsilon}$, $\hat{\Omega}_{\epsilon u}$ and $\hat{\Omega}_{uu}$ can be estimated as follows:

$$\hat{\Sigma}_{\epsilon\epsilon} = \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t\hat{\epsilon}'_t,\ \ \hat{\Sigma}_{\epsilon u} = \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t\hat{u}'_t,\ \ \hat{\Sigma}_{uu} = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t\hat{u}'_t, \tag{D.16}$$

and

$$\hat{\Lambda}_{uu} = \frac{1}{n}\sum_{i=1}^{M_n}(1 - \frac{i}{M_n+1})\sum_{t=i+1}^{n} \hat{u}_t\hat{u}'_{t-i},\ \ \hat{\Omega}_{uu} = \hat{\Sigma}_{uu} + \hat{\Lambda}_{uu} + \hat{\Lambda}'_{uu} \tag{D.17}$$

$$\hat{\Lambda}_{u\epsilon} = \frac{1}{n}\sum_{i=1}^{M_n}(1 - \frac{i}{M_n+1})\sum_{t=i+1}^{n} \hat{u}_t\hat{\epsilon}'_{t-i},\ \ \hat{\Omega}_{u\epsilon} = \hat{\Sigma}_{uu} + \hat{\Lambda}'_{u\epsilon} \tag{D.18}$$

where M_n is a bandwidth parameter satisfying $Mn \to 1$ and $M_n/\sqrt{n} \to 0$ as $n \to \infty$. We set $M_n = n^{1/3}$ following Kostakis et al. (2015).

E Improving REA Predictability Using SVIX

In this section, we investigate the predictive ability of SVIX, as inferred from the index option, to predict future growth in real economic activity.

In order to examine whether SVIX forecasts REA growth over h forecasting horizons, we set up our empirical analysis based on the following predictive regression:

$$REA_{i,t+h} = \beta_{i,0} + \beta_{i,1} REA_{i,t} + \beta_{i,2} SVIX_t + \beta_{i,2} X_t + \epsilon_{i,t+h} \tag{E.19}$$

where $REA_{i,t+h}$ is the growth rate of a different measure of economic activity as explained in Section 3.2. $SVIX_{t,h}$ is the h-month SVIX measure defined in Equation (5) , and X_t is a vector of additional predictors at time t as explained in Section 3.3. More specifically, we consider all of our control variables: term spread, credit spread, forward variance and IRRA.

Our results are reported in Table E.1 and show two main findings. First, we can see that SVIX predict 8 out of 9 REA proxies at 1-month horizon (all proxies except RS). At two-months horizon, SVIX predicts 6 out of 9 proxies (All proxies except RS and CFNAI). At three-months horizon, SVIX predicts IP, NFP, HOUS, UR and ADS. In addition, SVIX predicts NFP and IP at all forecasting horizon (1 to 12 month). The sign of the SVIX coefficient is significantly negative for NFP, IP, CNFAI, ADS and CU. We found a positive SVIX coefficient for the unemployment rate, which explains its countercyclical behaviour. This suggests that a rise in SVIX forecasts a decrease in REA.

[Table E.1 about here.]

		$h=1$	$h=2$	$h=3$	$h=6$	$h=12$
IP	β	-0.38***	-0.34***	-0.17*	-0.05***	-0.19**
		[0.00]	[0.00]	[0.07]	[0.47]	[0.03]
		(0.00)	(0.00)	(0.04)	(0.67)	(0.28)
	R^2	0.20	0.33	0.40	0.31	0.27
NFP	β	-0.30***	-0.19***	-0.17**	-0.16**	-0.21*
		[0.00]	[0.01]	[0.02]	[0.03]	[0.10]
		(0.00)	(0.02)	(0.06)	(0.10)	(0.19)
	R^2	0.67	0.76	0.76	0.71	0.55
RS	β	-0.05	-0.02	-0.07	0.07	0.01
		[0.45]	[0.92]	[0.44]	[0.95]	[0.96]
		(0.41)	(0.83)	(0.56)	(0.57)	(0.91)
	R^2	0.07	0.11	0.17	0.26	0.36
HOUS	β	-0.12**	-0.06	-0.01	-0.08	0.07
		[0.04]	[0.18]	[0.51]	[0.18]	[0.87]
		(0.17)	(0.52)	(0.84)	(0.15)	(0.43)
	R^2	0.20	0.20	0.20	0.36	0.55
CU	β	-0.37***	-0.32***	-0.14	-0.03	-0.14
		[0.00]	[0.00]	[0.14]	[0.65]	[0.16]
		(0.00)	(0.00)	(0.06)	(0.75)	(0.29)
	R^2	0.21	0.35	0.43	0.32	0.29
UR	β	0.38***	0.40***	0.37**	0.15	0.10
		[0.00]	[0.00]	[0.01]	[0.13]	[0.31]
		(0.00)	(0.00)	(0.00)	(0.21)	(0.37)
	R^2	0.17	0.33	0.41	0.56	0.53
CFNAI	β	-0.43***	-0.17	-0.06	0.05	-0.09
		[0.00]	[0.12]	[0.38]	[0.69]	[0.61]
		(0.00)	(0.02)	(0.54)	(0.67)	(0.46)
	R^2	0.48	0.50	0.43	0.27	0.19
ADS	β	-0.11***	-0.19***	-0.14**	0.06	-0.03
		[0.00]	[0.00]	[0.05]	[0.80]	[0.96]
		(0.02)	(0.01)	(0.10)	(0.69)	(0.83)
	R^2	0.85	0.71	0.62	0.38	0.22

Table E.1. Predicting REA via SVIX: This table reports results from the estimated multiple predictor regressions as in Equation (E.19) for various U.S. real economic activity (REA) proxies and for an h-month forecasting horizon ($h = 1, 3, 6, 9$, and 12 months). The REA proxies considered are explained in Section 3.2. The multiple predictor model includes the lagged REA and the h-month SVIX measure as predictors and is augmented by a set of control variables: term spread (TS), credit spread (CR), forward variance (FV), IRRA and 5 factors of Fama and French. We report the standardized ordinary-least-squares (OLS) coefficient estimates, IVX-Wald (within squared brackets) and Newey-West (within brackets) p-values of each one of the predictors as well as the in-sample adjusted R^2 for any given model. The sample spans July 1998 to May 2019. We reject the null hypothesis of a zero coefficient (no predictability) based on the p-values of the IVX-Wald test at the 1%, 5%, and 10% levels and denote it by triple, double, and single asterisks, respectively.

References

An, B., Ang, A., Bali, T. G. and Cakici, N. (2014), 'The joint cross section of stocks and options', *The Journal of Finance* **69**(5), 2279–2337.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006), 'The cross-section of volatility and expected returns', *Journal of Finance* **61**(1), 259–299.

Aruoba, S., Diebold, F. and Scotti, C. (2009), 'Real-time measurement of business conditions', *Journal of Business Economic Statistics* **27**, 417–427.

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

Bakshi, G., Panayotov, G. and Skoulakis, G. (2011), 'Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios', *Journal of Financial Economics* .

Bakshi, G., Panayotovb, G. and Skoulakisc, G. (2012), The baltic dry index as a predictor of global stock returns,commodity returns, and global economic activity. Chicago meeting.

Balakrishnan, K., Bartov, E. and Faurel, L. (2010), 'Post loss/profit announcement drift', *Journal of Accounting and Economics* **50**(1), 20–41.

Balla, R., Gerakosa, J., Linnainmaaa, J. T. and Nikolaev, V. V. (2015), 'Deflating profitability', *Journal of Financial Economics* **115**, 225–248.

Balvers, R., Cosimano, T. and McDonald, B. (1990), 'Predicting stock returns in an efficient market', *Journal of Finance* .

Basu, S. (1983), 'The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence', *Journal of Financial Economics* **12**(1), 129–156.

Bernanke, B. (1983), 'Nonmonetary effects of the financial crisis in the propagation of the great depression,', *American Economic Review* **73**(3), 257–276.

Bhandari, L. C. (1988), 'Debt/equity ratio and expected common stock returns: Empirical evidence', *The Journal of Finance* **43**(2), 507–528.

Bollen, N. P. B. and Whaley, R. E. (2004), 'Does net buying pressure affect the shape of implied volatility functions?', *Journal of Finance* **59**, 711–753.

Bondt, W. F. M. D. and Thaler, R. (1985), 'Does the stock market overreact?', *The Journal of Finance* **40**(3), 793–805.

Buss, A., Schoenleber, L. and Vilkov, G. (2019), 'Expected correlation and future market returns', *CEPR Discussion Paper* **DP12760**.

Buss, A. and Vilkov, G. (2012), 'Measuring equity risk with option-implied correlations', *Review of Financial Studies* **25**(10).

Bustamant, M. C. and Donangelo, A. (2017), 'Product market competition and industry returns', *The Review of Financial Studies* **30**(12), 4216–4266.

Campbell, J. Y. and Cochrane, J. H. (1999), 'A consumption-based explanation of aggregate stock market behavior', *Journal of Political Economy* .

Campbell, J. Y. and Thompson, S. B. (2008), 'Predicting excess stock returns out of sample: Can anything beat the historical average?', *The Review of Financial Studies* **21**(4), 1509–1531.

Chabi-Yo, F., Dim, C. and Vilkov, G. (2022), 'Generalized bounds on the conditional expected excess return on individual stocks', *Management Science* .

Chen, N.-F. (1991), 'Financial investment opportunities and the macroeconomy', *The Journal of Finance* **46**(2), 529–554.

Christoffersen, P., Jacobs, K. and Chang, B. Y. (2012), 'Forecasting with option-implied information', Handbook of Economic Forecasting.

Conrad, J., Dittmar, R. F. and Ghysels, E. (2013), 'Ex ante skewness and expected stock returns', *The Journal of Finance* **68**, 85–124.

Cooper, M. J., Gulen, H. and Schill, M. J. (2008), 'Asset growth and the cross-sectionof stock returns', *The Journal of Finance* **63**(4), 1609–1651.

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

Damodaran, A. (2019), 'Equity risk premiums (erp): Determinants, estimation and implications – the 2019 edition', *New York University Stern School of Business Research Paper Series* .

Datar, V. T., Naik, N. Y. and Radcliffe, R. (1998), 'Liquidity and stock returns: An alternative test', *Journal of Financial Markets* **1**(2), 203–2019.

Dennis, P. and Mayhew, S. (2002), 'Risk-neutral skewness: Evidence from stock options', *Journal Of Financial And Quantitative Analysis* **37**(3).

Duan, J.-C. and Zhang, W. (2014), 'Forward-looking market risk premium', *Management Science* **60**, 521–538.

D'Acuntoa, F., CarolinPfluegerc, R. and Weberde, M. (2018), 'Flexible prices and leverage', *Journal of Financial Economics* **129**(1), 46–68.

Estrella, A. and Hardouvelis, G. (1991), 'The term structure as a predictor of real economic activity', *The Journal of Finance* **46**(2), 557–576.

Evgeny Lyandres, L. S. and Zhang, L. (2008), 'The new issues puzzle: Testing the investment-based explanation', *Review of Financial Studies* **21**(6), 2825–2855.

Faccini, R., Konstantinidi, E., Skiadopoulos, G. and Sarantopoulou-Chiourea, S. (2019), 'A new predictor of real economic activity: The s&p 500 option implied risk aversion', *Management Science* **65**(10).

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

Fama, E. F. and French, K. R. (1992), 'The cross-section of expected stock returns', *The Journal of Finance* **47**(2), 427–465.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1).

Fama, E. and French, K. (2015), 'A five-factor asset pricing model', *Journal of Financial Economics* **116**(1), 1–22.

Fernández, P. (2009), 'Market risk premium used in 2008 by professors: A survey with 1,400 answers', *Political Economy: International Political Economy eJournal* .

Frazzinia, A. and Pedersen, L. H. (2014), 'Betting against beta', *Journal of Financial Economics* **111**, 1–25.

Freyberger, J., Neuhierl, A. and Weber, M. (2020), 'Dissecting characteristics nonparametrically', *Review of Financial Studies* **35**(5), 2326–2377.

Friedman, B. M. and Kuttner, K. N. (1991), 'Why does the paper-bill spread predict real economic activity?', *NBER Working Papers 3879, National Bureau of Economic Research, Inc* .

Gandhi, P. and Lustig, H. (2015), 'Size anomalies in u.s. bank stock returns', *The Journal of Finance* **79**, 733–768.

George, T. J. and Hwang, C. (2004), 'The 52-week high and momentum investing', *The Journal of Finance* **59**(5), 2145–2176.

Gilchrist, S. and Zakrajsek, E. (2012), 'Credit spreads and business cycle fluctuations', *American Economic Review* **102**, 1692–1720.

Gkionis, K., Kostakis, A., Skiadopoulos, G. and Stilger, P. S. (2021), 'Positive stock information in out-of-the-money option prices', *Journal of Banking & Finance* .

González-Urteaga, A., Nieto, B. and Rubio, G. (2021), 'Extracting expected stock risk premia from option prices and the information contained in non-parametric-out-of-sample stochastic discount factors', *Quantitative Finance* **21**(5), 713–727.

Gorodnichenko, Y. and Weber, M. (2016), 'Are sticky prices costly? evidence from the stock market', *American Economic Review* **106**(1).

Harvey, C. (1989), 'Forecasts of economic growth from the bond and stock markets', *Financial Analysts Journal* **45**(5), 38–45.

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

Hirshleifer, D., Hou, K., Teoh, S. H. and Zhang, Y. (2004), 'Do investors overvalue firms with bloatedbalance sheets?', *Journal of Accounting and Economics* **38**, 297–331.

Hong, H. and Yogo, M. (2012), 'What does futures market interest tell us about the macroeconomy and asset prices?', *Journal ofFinancialEconomics* **105**.

Houa, K., van Dijkb, M. A. and Zhang, Y. (2012), 'The implied cost of capital: A new approach', *Journal of Accounting and Economics* **53**, 504–526.

Hu, J. (2013), 'Does option trading convey stock price information?', *Singapore Management University Lee Kong Chian School of Business Research Paper Series* .

Jagadeesh, N. (1990), 'Evidence of predictable behavior of security returns', *The Journal of Finance* **45**(3), 881–898.

James L. Davis, E. F. F. and French, K. R. (2000), 'Characteristics, covariances, and average returns: 1929 to 1997', *Journal of Finance* **55**, 389–406.

Jiang, G. J. and Tian, Y. S. (2005), 'The model-free implied volatility and its information content', *Review of Financial Studies* **18**(4), 1305–1342.

Johnson, T. L. and So, E. C. (2012), 'The option to stock volume ratio and future returns', *Journal of Financial Economics* **106**, 262–286.

Kadan, O. and Tang, X. (2019), 'A bound on expected stock returns', *The Review of Financial Studies* **33**(4).

Kang, B. J., Kim, T. S. and Yoon, S. (2010), 'Information content of volatility spreads', *The Journal of Futures Markets* **30**(6), 533–558.

Kelly, B. and Jiang, H. (2014), 'Tail risk and asset prices', *The Review of Financial Studies* **27**(10).

Kelly, B. and Pruitt, S. (2019), 'Characteristics are covariances', *Journal of Financial Economics* **134**, 501–524.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M. P. (2015), 'Robust econometric inference for stock return predictability', *The Review of Financial Studies* **28**(5), 1506–1553.

Kostakis, A., Panigirtzoglou, N. and Skiadopoulos, G. (2011), 'Market timing with option-implied distributions: A forward-looking approach', *Management Science* **57**(7), 1231–1249.

Lakonishok, J., Lee, I., Pearson, N. D. and Poteshman, A. M. (2007), 'Option market activity', *Review of Financial Studies* **20**, 813–857.

Lemmon, M. and Ni, S. X. (2014), 'Differences in trading and pricing between stock and index options', *Management Science* **60**.

Lewellen, J. (2015), 'The cross-section of expected stockreturns', *Critical Finance Review* **4**, 1–44.

Liew, J. and Vassalou, M. (2000), 'Can book-to-market, size and momentum berisk factors that predict economic growth?', *Journal of Financial Economics* **57**.

Loflund, A. and Nummelin, K. (1997), 'On stocks, bonds and business conditions', *Applied Financial Economics* **7**(2), 137–146.

Martin, I. (2017), 'What is the expected return on the market', *The Quarterly Journal of Economics* **132**, 367–433.

Martin, I. and Wagner, C. (2019), 'What is the expected return on a stock?', *The Journal of Finance* .

Newey and West (1994), 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies* **61**(4), 631–653.

Ng, S. and Wright, J. H. (2013), 'Facts and challenges from the great recession for forecasting and macroeconomic modeling', *Journal of Economic Literature* **51**(4), 1120–1154.

Novy-Marx, R. (2011), 'Operating leverage', *Review of Finance* **15**(1), 103–134.

Novy-Marx, R. (2012), 'Is momentum really momentum?', *Journal of Financial Economics* **103**(3), 429–453.

Palazzo, B. (2012), 'Cash holdings, risk, and expected returns', *Journal of Financial Economics* **104**, 162–185.

Pan, J. and Poteshman, A. M. (2006), 'The information in option volume for future stock prices', *Review of Financial Studies* **19**, 871–908.

Rehman, Z. U. and Vilkov, G. (2012), 'Risk-neutral skewness: Return predictability and its sources', *Working Paper* .

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

Ross, S. A. (2015), 'The recovery theorem', *Journal of Finance* **70**(2), 615–648.

Santa-clara, P. and Yan, S. (2010), 'Crashes, volatility, and the equity premium: Lessons from s&p 500 options', *The Review of Economics and Statistics* **92**, 435–451.

Schneider, P. and Trojani, F. (2019), '(almost) model-free recovery', *Journal of Finance* **74**(1), 323–370.

Sloan, R. G. (1996), 'Do stock prices fully reflect information in accruals and cash flows aboutfuture earnings?', *The Accounting Review* **71**(3), 289–315.

Soliman, M. T. (2008), 'The use of dupont analysis by market participants', *The Accounting Review* **83**(3), 823–853.

Stilger, P. S., Kostakis, A. and Poon, S.-H. (2016), 'What does risk-neutral skewness tell us about future stock returns?', *Management Science* **63**, 1814–1834.

Stock, J. H. and Watson, M. W. (1989), 'New indexes of coincident and leading economic indicators', *New Indexes of Coincident and Leading Economic Indicators," in NBER Macroeconomics Annual* .

Stock, J. H. and Watson, M. W. (2003), 'Forecasting output and inflation: The role of asset prices.', *Journal of Economic Literature* **41**(3), 788–829.

Welch, I. (2000), 'Views of financial economists on the equity premium and on professional controversies', *Yale School of Management Research Paper Series* **73**(4), 501–537.

Xing, Y., Zhang, X. and Zhao, R. (2010), 'What does the individual option volatility smirk tell us about future equity returns?', *Journal of Financial and Quantitative Analysis* **45**, 641–662.