Butterfly Implied Returns*

October 2023

Abstract

This paper introduces a new approach that infers the individual stock return during market crashes from the options market. The approach relies on the correlation between the VIX and the prices of butterflies at different strikes. Applying it to the cross-section of S&P 500 stocks yields a strategy that hedges the market downturn while earning an annualized alpha of approximately 4%. The aggregation produces a measure for the severity of market crashes, which is shown to be an important determinant of both the equity risk premium and the survey-based expectation of return.

Keywords: Predictability of stock returns; Cross-section of stock returns; Downside risk

JEL classification: G11, G12

*We thank Yakov Amihud, Martijn Boons, Jonathan Brogaard, James O'Donovan, Menna El Hefnawy (discussant at 30th Finance Forum in Málaga), Paul Hsu, Yi Huang, Tao Li, Xingguo Luo, Yue Ma, Philippe Mueller, Matthew Ringgenberg, Junbo Wang, Xueping Wu, Qi Xu, Wayne Yu, Qifei Zhu, and seminar participants at City University of Hong Kong and Zhejiang University for their many helpful comments and suggestions. Di Wu gratefully acknowledges the financial support from the HK Research Grant Council (GRF#11503421 and ECS #21504220).

1 Introduction

In recent years, researchers have made significant efforts to infer the distribution of the stock market return from the index option prices. While the option prices contain forward-looking information on the market return, it is necessary to impose restrictions on the stochastic discount factor to recover the physical distribution. This often leads to significant measurement errors.[1] In this paper, we explore a new direction based on the simple idea that the market return in each state is the weighted average of the returns of the constituents in the same state. Therefore, we can start with a noisy measure in the cross-section of stocks, and let the measurement errors cancel out themselves in the aggregate.

We apply this idea to examine the time-varying severity of a market crash, starting with a noisy measure of individual stock returns in the crash. In this context, we propose the following procedure to recover the latter from option prices: One input is the prices of the butterfly options at different strikes. A butterfly pays off at expiration only if the stock price falls into a particular range. Naturally, the price of the butterfly goes up when investors consider this range to be more likely. The other input is a high-frequency variable closely related to the probability of a market crash, and we adopt the Cboe Volatility Index (VIX), or "fear index," which rises as a market crash becomes more likely. For each stock, we calculate the rolling correlation between changes in VIX and the prices of butterflies at different strikes. The butterfly that co-moves most strongly with VIX reveals the expectation of a stock's return in a market crash, which equals the strike of this butterfly divided by the current stock price, minus one. We refer to this return the butterfly implied return (BIR). Using market capitalization as the weight, we can calculate the BIR of the market, named BIRM, which describes the expectation of market return during a market crash. Although our goal is to analyze BIRM, we start by exploring the information in the dispersion of BIR. We present the following results.

[1] For example, Ross (2015) proposes a procedure by restricting the martingale component of the stochastic discount factor to be a constant one (Borovička, Hansen, and Scheinkman, 2016). As a result, the inferred distribution cannot predict the mean or variance of the return (Jackwerth and Menner, 2020).

First, an easy way to hedge against a future crash is to short vulnerable stocks with low BIR and long resilient stocks with high BIR. We call this cross-sectional strategy betting with butterfly (BwB). The BwB time series demonstrates a clear pattern of profit during a market crash, such as the dot-com crash or the Great Recession. In addition to hedging against a market crash, BwB earns a statistically significant alpha, ranging from 0.28% to 0.39% per month using common factor models. This may seem unexpected, given the hedging benefit. A possible explanation for the positive alpha is that options prices reflect information not incorporated into the stock price. However, we find it unlikely that the expected return in a market crash is insider information. Nevertheless, we interact BIR with variables that measure the relative informativeness of options and stock, such as a higher option-to-stock volume ratio, which suggests more insider trading in the options market. In Fama-MacBeth regressions, BIR itself is significant, with a t-value of around three across different specifications. However, we find no evidence of stronger predictability for the subsample of stocks with more insider activities in the options market.

After examining the pattern of BwB closely, it becomes clear that there is a rational explanation: when the market turns around, BwB realizes a significant loss. This loss can be attributable to stocks with a low BIR, which are hit the hardest by price drops but potentially benefit the most when the market rebounds. BwB shorts these stocks, resulting in a loss at the bottom of the crash. From the point of view of a representative investor, the loss of BwB coincides with the point of highest marginal utility. For this reason, BwB earns a positive alpha, which can be useful for investors with different consumption processes than the representative investor.

Using the demand system approach, we investigate whether BIR affects the portfolio choice of different types of investors. We focus on three relatively flexible investor types: investment advisors (including hedge funds), mutual funds, and households. As BIR positively predicts future return, the weight of a stock in a profit-maximizing investor's portfolio should have a positive loading on BIR. This holds true for investment advisors, whose loading predicts the future 12-month return of BwB with high accuracy. For mutual funds, the loading predicts the future BwB return before 2008. Since then, interestingly, mutual funds have started taking the other side of the trade. Finally, the households seem to be on the other side of BwB in most of the sample periods, as BwB is considered risky in the eyes of a representative household.

Following the rational explanation that shows BwB is risky, one immediate implication is that a stock that co-moves positively with BwB is also risky and thus requires compensation in the form of a higher return. This implication allows us to expand our sample from stocks with liquid options to all stocks. We estimate the beta to the return of BwB using a rolling window, then test the pricing of BwB-beta with Fama-MacBeth regression. The result suggests that a two-standard deviation difference in BwB-beta creates a return spread of 3.9 percent per annum.

We then explore the implications of BIRM, which describes the expectation of market return in the crash state. When investors fear a severe crash, indicated by a very negative BIRM, the price of stocks decreases, and the risk premium increases. This means that the equity risk premium loads negatively on BIRM, which we verify using the SVIX-based measure of equity risk premium (Martin, 2017). Our empirical findings show that the loading is approximately -0.07, a reasonable amount that reflects the difference between the physical and risk-neutral probability of a market crash over the following twelve months. Across various horizons, BIRM accounts for a significant portion of the mean and variations of the equity risk premium, indicating that the severity of a market crash is an important factor in determining the equity risk premium. On the other hand, we discover a strong and positive empirical relationship between BIRM and survey-based expectation of return (Greenwood and Shleifer, 2014). We argue that survey participants may answer questions about expected return with a return that they consider likely. As a result, the survey-based expectation of return is negatively correlated with the theory-based expected return. BIRM provides a rational explanation for the differences between these two returns.

The evidence also suggests that BIRM is a natural candidate for a state variable that describes the investment opportunity. If the investment opportunity worsens, as indicated by a low BIRM, stocks with positive BIRM exposure perform poorly, thus commanding a higher expected return. To estimate the price of risk of BIRM, we use the GMM framework and assume that the stochastic discount factor is linear in the market return and changes in BIRM. The test assets include portfolios sorted by characteristics (such as size, book-to-market ratio, and idiosyncratic volatility) and covariance (beta). As expected, we find a positive and significant price of risk for both the market return and changes in BIRM. The estimated price of risk for the market return in the second stage is 3.5, which is close to a typical coefficient of relative risk aversion.

Related Literature

This paper adds to the existing literature on cross-sectional stock returns by exploring information from the options market. Various methods, including differences in implied volatility, risk-neutral skewness, and risk-neutral variance, have been used in previous works. For example, Cremers and Weinbaum (2010) use the average difference in implied volatility across put-call pairs; Xing, Zhang, and Zhao (2010) use the difference in implied volatility between out-of-money puts and at-the-money calls; and An, Ang, Bali, and Cakici (2014) use changes in the implied volatility of short maturity options. Stilger, Kostakis, and Poon (2017) use risk-neutral skewness; Martin and Wagner (2019) derive a formula that expresses the expected return on a stock in terms of the risk-neutral variance of the market, the risk-neutral variance of the individual stock, and the value-weighted average of the stocks' risk-neutral variance; and Kadan and Tang (2020) show that the risk-neutral variance predicts the future return on a subset of stocks. This paper focuses on the expectation of return in a particular future state, which makes the predictability found in this study robust to controlling for existing measures.

Furthermore, this paper introduces a new aggregate variable called BIRM, which quantifies the market return during a crash. This variable marks a significant step towards recovering the entire physical distribution of returns (Ross, 2015). The approach in this paper does not require restrictions on the stochastic discount factor (Borovička, Hansen, and Scheinkman, 2016; Schneider and Trojani, 2019) or a large number of assets with different maturities (Jensen, Lando, and Pedersen, 2019). In this paper, we focus on market crashes and use VIX as an input variable. However, any other series that is closely linked to a specific state, such as digital options on the index prices (Breeden and Litzenberger, 1978) or new CME event contracts, can be used in place of VIX.

Additionally, this paper finds a significant and positive price of risk for BIRM, making it another option-based risk factor, adding to the list that includes the VIX as used by Ang, Hodrick, Xing, and Zhang (2006b) and risk-neutral skewness as used by Chang, Christoffersen, and Jacobs (2013). Relatedly, Cremers, Halling, and Weinbaum (2015) mimic jump risk using two at-the-money straddles with different maturities; Lu and Murray (2019) construct a collar option that only pays off when market return falls below a threshold.

Finally, this paper contributes to the literature on survey-based subjective expectations of return, offering a rational explanation for the negative correlation between subjective expectation and theory-based expected returns (Greenwood and Shleifer, 2014). The severity of the crash, as measured by BIRM, is shown to negatively affect the expected return, and survey respondents may overemphasize the probability of a market crash, similar to the distorted probability measure of unemployment and inflation used by Baqaee (2020) and Bhandari, Borovička, and Ho (2022).

As the construction of BIRM depends on BIR, we first focus on the cross-sectional analysis of BIR in Section 2, which includes the standard asset pricing tests and the effect of BIR on the portfolio choices of different types of investors. Thereafter, we move on to the time series analysis of BIRM in Section 3, which relates BIRM to both the survey-based return expectation and the theory-based expected return. We conclude this paper in Section 4, with updated results using data that became available after the circulation of the first draft.

2 BIR and its pricing in the cross-section

In a typical jump-diffusion setting, the value of a put option can be attributable to both the normal Brownian motion, which gradually moves the stock price below the strike, and the possibility of crashes. The rationale behind our approach can be understood using the simple case with just market crashes. In this case, a market crash occurs with probability p^M. The gross return of stock i conditional on a market crash is a point mass R_i^M, which is the target of this paper. If no crash occurs, the stock price follows a diffusion process (everywhere continuous). Ignoring the time discount, the value of a put option $put_{i,t}(k)$ with strike k (as a fraction of the stock price) can be approximately decomposed into

$$put_{i,t}(k) = p^M B^{-\gamma} \max\left(k - R_i^M, 0\right) + \left(1 - p^M\right) bs_i(k). \tag{1}$$

The first term is multiplied by $B^{-\gamma} > 1$ to recognize that a market crash tends to happen in bad times with higher marginal utility. $bs_i(k)$ represents the value from the diffusion part. Gabaix (2012) shows that this simple structure is sufficient to quantitatively account for the well-known puzzles in the derivatives market, such as the volatility skew and the predictability of variance risk premium.

The presence of crash creates a kink (not twice-differentiable) around R_i^M. Therefore, as suggested by the title of this paper, we use a butterfly to approximate the Breeden and Litzenberger (1978) approach, by buying the put at $k - \epsilon$ and the put at $k + \epsilon$, and selling two of the puts at k. The value of this butterfly is $\mathcal{B}_{i,t}(k) = put_{i,t}(k-\epsilon) + put_{i,t}(k+\epsilon) - 2 \times put_{i,t}(k)$, which can be simplified to

$$\mathcal{B}_{i,t}(k) = p^M B^{-\gamma} \Lambda\left(k, R_i^M\right) \epsilon + \left(1 - p^M\right) bs_i^{''}(k) \epsilon^2 \tag{2}$$

where $bs_i^{''}(k)$ is the second derivative of $bs_i(k)$ with respect to k. $\Lambda\left(k, R_i^M\right)$ equals $1 - \frac{\left|k - R_i^M\right|}{\epsilon}$ for $\left|k - R_i^M\right| \leq \epsilon$, and 0 otherwise. The function Λ resembles the payoff of a butterfly around

k, as "max $\left(k - R_i^M, 0\right)$" describes the payoff of a put option with strike k. In other words, the function Λ is Λ-shaped, reaching the maximal value of 1 when $k = R_i^M$.

From this point on, we deviate from the recovery literature by focusing on the change, rather than the level, of $\mathcal{B}_{i,t}(k)$. The reason is simple: suppose we find a proxy for p^M, $\mathcal{V}\left(p^M\right)$, we would have

$$\frac{d\mathcal{B}_{i,t}(k)}{d\mathcal{V}} = \frac{1}{\mathcal{V}'}\left[B^{-\gamma}\Lambda\left(k, R_i^M\right)\epsilon - bs_i^{''}(k)\,\epsilon^2\right]. \tag{3}$$

For small ϵ such that $\epsilon \gg \epsilon^2$, $\frac{d\mathcal{B}_{i,t}(k)}{d\mathcal{V}}$ is maximized approximately when $\Lambda\left(k, R_i^M\right)$ is, i.e., when $k = R_i^M$. Therefore, we can recover R_i^M via

$$R_i^M \approx \arg\max_k \frac{d\mathcal{B}_{i,t}(k)}{d\mathcal{V}}. \tag{4}$$

The initial step in our implementation process involves selecting a variable, $\mathcal{V}$, that is positively related to the probability of a market crash. Then, for each stock, we look for the moneyness of a butterfly spread whose price change is maximally sensitive to the changes in $\mathcal{V}$. The limitation is that this procedure requires high-quality options data, as the recovery is biased if ϵ is large. Therefore, we focus on the S&P 500 constituents for the main analysis. There are several additional decisions that need to be made, which we will outline in detail below.

2.1 Implementation of the procedure

2.1.1 Choosing the variable $\mathcal{V}$

The condition $\mathcal{V}' > 0$ indicates that we look for an index that rises when a market crash becomes more likely. Ideally, this index can be observed at high frequency to facilitate the measurement of $\frac{d\mathcal{B}_{i,t}(k)}{d\mathcal{V}}$. Therefore, we focus on indices constructed using equity or derivatives prices rather than those based on surveys or valuation ratios. A direct measure of the crash probability would satisfy this condition. However, any measure of the varying probability of rare events is inherently difficult to validate. Our approach is far less demanding and, as a result, circumvents this issue.

Our choice is the VIX index, which is famously dubbed the "fear index." Of course, VIX contains compensation for diffusion, spike, and crash. Formally, VIX is calculated with a basket of out-of-money options that are weighted by their strikes, or

$$\mathrm{VIX}_{t\to T} = 2R_{f,t\to T}\left\{\int_0^{F_{t,T}} \frac{\mathrm{put}_{t,T}(K)}{K^2}dK + \int_{F_{t,T}}^{\infty} \frac{\mathrm{call}_{t,T}(K)}{K^2}dK\right\} \tag{5}$$

where $R_{f,t\to T}$ is the risk-free interest rate, $F_{t,T}$ is the forward price, and K is the strike. As the formula suggests, VIX overweights out-of-money put options with low strikes. A decomposition by Bollerslev and Todorov (2011) finds that the compensation for the crash in the left tail is several times that for the right tail. Therefore, although VIX is not a direct measure of the crash probability, it rises as a market crash becomes more likely, which satisfies the requirement of $\mathcal{V}' > 0$.

We have considered several other candidates. Assuming that the marginal utility of the investor is inversely proportional to the gross return of the market R_T, Martin (2017) show that the probability of a crash over the period from t to T can be calculated as

$$\tilde{\mathbb{P}}(R_T < \alpha) = \alpha\left[\mathrm{put}'_{t,T}(\alpha S_t) - \frac{\mathrm{put}_{t,T}(\alpha S_t)}{\alpha S_t}\right] \tag{6}$$

where $\mathrm{put}'_{t,T}$ is the first derivative of put option price with respect to the strike, which equals to the price of a claim to \$1 contingent on $R_T < \alpha$. Given that our theme is to recover without a restriction on the stochastic discount factor, we do not adopt this measure. In addition, as $\tilde{\mathbb{P}}(R_T < \alpha)$ relies on the out-of-money put option prices, it is empirically correlated with other measures based on the same prices. We do not have the series at a daily frequency. However, using the monthly data, we find that the correlations amongst the probabilities with different thresholds (i.e., α) and tenors (i.e., T) are well above 0.8. The correlation between the computed crash probability and VIX of the same maturity is typically around 0.9. Therefore, despite its theoretical appeal as a measure of the level of the crash probability, the variation of this measure does not buy us much.

Lu and Murray (2019) propose an "AD Bear" portfolio by taking a long position in a put with a strike price K_1 and a short position in a put with a strike price K_2 and scaling both positions by $K_1 - K_2$. It is easy to see that, when K_1 and K_2 are sufficiently close around the point αS_t, the price of the AD Bear portfolio is the $\mathrm{put}'_{t,T}(\alpha S_t)$ term in Equation (6). In other words, the price of this portfolio approximately measures $\tilde{\mathbb{P}}(R_T < \alpha)$, except for the bias from ignoring the $-\frac{\mathrm{put}_{t,T}(\alpha S_t)}{\alpha S_t}$ term. Therefore, we do not adopt this measure.

Finally, as stock price reacts to the fear of a crash, we may use a negative stock return to indicate heightened crash risk. Empirically, stock return and VIX tend to move in the opposite direction. Indeed, in the earlier stage of this work, we experimented with several function forms of the stock return. However, the results are weaker than that obtained using VIX. Stock prices move for reasons other than the varying crash probability, exacerbating the measurement error issues.

Overall, the alternative measures we examined do not clearly dominate VIX. Following the advice of John Maynard Keynes, when attractiveness is subjective, we settle for the most popular choice.

2.1.2 Approximating the price of butterfly $\mathcal{B}_{i,t}(k)$

We use the volatility surface of OptionMetrics to calculate the prices of the butterflies. OptionMetrics calculates the implied volatilities for options on individual stocks with various maturities at each delta from -0.1 to -0.9 at 0.05 intervals each day. The OptionMetrics method interpolates using a kernel-smoothing technique to utilize all puts and calls with different deltas and maturities. Our results are similar with puts or calls of different maturities because every point on the volatility surface aggregates all the information in the options market of this stock. In this paper, we use the implied premium of put options with 30 days to maturity. Notice that the interpolation is not arbitrage-free and needs not to be, as we only extract a signal from the options market to trade the stock.

The volatility surface is indexed by deltas instead of strikes. To avoid the heavy burden of recreating the daily surface, we propose a workaround. We approximate the price of the butterfly at each delta Δ for stock i at time τ, $\tilde{\mathcal{B}}_{i,\tau}^{\Delta}$, by buying the put at $\Delta - 0.05$ and the put at $\Delta + 0.05$, and selling two of the puts at Δ, all with 30-day maturity. For this to be a butterfly, the implied strike K needs to satisfy $K_{i,\tau}(\Delta + 0.05) = 2K_{i,\tau}(\Delta) - K_{i,\tau}(\Delta - 0.05)$. In reality, this equality generally does not hold, which makes our butterfly "broken-winged". Fortunately, the approach in this paper does not rely on the price of the butterfly per se. Instead, it tracks the price changes. Specifically, we calculate the rolling six-month correlation between the changes in $\tilde{\mathcal{B}}_{i,\tau}^{\Delta}$ and VIX_τ, and pick the butterfly whose price is most likely to move in the same direction as VIX over the rolling six-month window, or

$$\Delta_{i,t}^{*} = \underset{\Delta \in \{-0.15,\ \ldots -0.85\}}{\arg\max} \left\{ \text{corr}_{t-6M \leq \tau \leq t} \left(\tilde{\mathcal{B}}_{i,\tau}^{\Delta} - \tilde{\mathcal{B}}_{i,\tau-1}^{\Delta},\ \text{VIX}_\tau - \text{VIX}_{\tau-1} \right) \right\}. \tag{7}$$

The implied strike at each delta changes daily. Therefore, having found the corresponding $\Delta_{i,t}^{*}$, we use the average implied strike at this delta relative to the stock price over the six-month estimation window, which is denoted as $\overline{\frac{\mathrm{K}_{i,\tau}\left(\Delta_{i,t}^{*}\right)}{\mathrm{P}_{i,\tau}}}$. This workaround is unnecessary if one starts with a volatility surface indexed by equal-interval strikes instead of deltas.

2.1.3 Calculating butterfly implied return $\text{BIR}_{i,t}$

Finally, we define the butterfly implied return (BIR) for stock i at time t as

$$\text{BIR}_{i,t} = \overline{\frac{K_{i,\tau}\left(\Delta_{i,t}^{*}\right)}{P_{i,\tau}}} - 1. \tag{8}$$

BIR measures the return expectation of stock i in 30 days if a market crash occurs, i.e., $R_i^M - 1$. In principle, we could have one BIR each day for each stock. However, as the calculation is made on a rolling basis, the resulting BIR only changes gradually. Therefore, we use BIR at the end of each month. Further, we require that the prices of the butterfly are positive and that there are at least 60 valid daily observations in the six-month estimation window. The result is a panel of BIR at a monthly frequency from June 1996 to December 2019, with on average 492 of the S&P 500 constituents appearing each month.

The measurement error exists due to (i) the ϵ^2 term in Equation (3), which can be alleviated with a more "dense" volatility surface, (ii) our choice of the proxy for the probability of a market crash in Equation (4), which can be improved with new financial instruments, e.g., event contacts offered by CME Group, (iii) the use of a rolling window in Equation (7), which can be improved with a shorter window of tick-by-tick data, and (iv) the workaround to transform delta to strike in Equation (8), which can be improved if one is willing to expend resources to generate a volatility surface indexed by strikes.

2.1.4 What kind of crash does BIR measure?

One limitation of this procedure is that the crash's nature and severity are ambiguous. As VIX utilizes all the out-of-money put options, it responds to the possibility of crashes of any type. Theoretically, the procedure can be improved by using a variable $\mathcal{V}$ related to a specific kind of crash (for example, by changing the threshold α in Equation (6)). In practice, we are constrained by the data quality. At daily or lower frequency, these measures tend to have a high correlation among themselves, and with VIX.

It turns out that the procedure already allows us to identify the severity of the crash. However, one more step is needed, as we explain in Section 3. The severity varies over time and exhibits clear pro-cyclicality, with the median being approximately an annualized drop of 30% in the index level, as seen from Figure 4. Therefore, BIR measures the performance of individual stocks in relatively severe market crashes.

2.2 Basic Statistics in Cross-Section

We start by examining what kind of stocks are picked by our BIR. Panel A of Table 1 reports the time series statistics of the cross-sectional mean and dispersion of BIR. BIR is negative on average, which is unsurprising as it is the supposed return during a crash. The monthly first-order autocorrelation is 0.97, corresponding to a half-life of 22 months. Based on BIR, we divide stocks into ten groups at the end of each month. The lowest-BIR group (Group 1) of "vulnerable" stocks has a typical moneyness of -12%, while the highest-BIR group (Group 10) of "resilient" stocks each month has a typical moneyness of $+8\%$. Although certain types of stock hedge against market crashes, we think the positive moneyness is mainly attributable to the measurement error in BIR, and the identified "vulnerable" stocks primarily drive the cross-sectional results. Later, we have several pieces of evidence that support this view. From now on, we will focus on these "vulnerable" stocks.

As these are constituents of the S&P 500 index, the difference in size across the groups is limited. However, smaller firms are more likely to appear in the extreme portfolios. This occurs because for smaller and thus more volatile firms, the implied strike at the end of the volatility surface ($\Delta = -0.1$ or -0.9) tends to be far from the current stock price. A similar logic applies to the average beta, which is larger for extreme portfolios. However, the characteristics do not change smoothly between groups. Comparing Group 1 with the neighboring Group 2, we see a significant gap in the CAPM beta. In addition, the firms in Group 1 invest heavily with much lower profitability, suggesting they are likely to be the hardest hit in a recession.[2] This characterization is reminiscent of Fama and French (2015), who mention the difficulty in capturing "the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability." Indeed, a strategy based on BIR produces a positive alpha under the five-factor model of Fama and

[2] We use operating profit divided by book equity to measure profitability, as in Fama and French (2015). The cash-based measure of profitability (Ball et al., 2015) actually declines smoothly from Group 1 (average 18.8%) to Group 10 (average 17.3%). While operating profitability is not significantly priced in the cross-section of the S&P500 stock returns (Table 2), cash-based profitability is. Nevertheless, $\mathrm{BIR}_{i,t-1}$ is significant with either one as the control.

French (2015), as we show in Section 2.3.2.

Table 1: **Basic Statistics**

In each month t, we calculate the cross-sectional number of observations n_t, mean μ_t, and standard deviation σ_t of $\text{BIR}_{i,t-1}$ across stocks. Panel A reports the time series statistics of n_t, μ_t, and σ_t. Based on $\text{BIR}_{i,t-1}$, we divide the stocks into ten groups at the end of month t and calculate the group means of the common stock characteristics. Panel B reports the time series average of the group means for the following variables: CAPM beta (beta), market equity (size), book-to-market equity (BMdec), investment (AssetGrowth), operating profitability (OperProf), and momentum (Mom12m), with the parentheses showing the corresponding variable names used in Chen and Zimmermann (2022).

Panel A: Time Series Properties of the Cross-Sectional Moments

Avg $n_t\,(\text{BIR}_{i,t-1})$	$\mu_t\,(\text{BIR}_{i,t-1})$ in %			$\sigma_t\,(\text{BIR}_{i,t-1})$ in %		
	Avg	Std	ac(1)	Avg	Std	ac(1)
492	-2.401	2.289	0.969	5.854	1.936	0.978

Panel B: Time Series Average of Characteristics by Group

Group	$\text{BIR}_{i,t-1}$ in %	CAPM Beta	ME in \$Bn	B/M	INV	OP	MOM in %
1	-11.85	1.11	20.89	0.49	0.21	0.24	12.10
2	-8.55	0.81	26.52	0.49	0.12	0.44	10.91
3	-6.55	0.73	30.73	0.47	0.11	0.34	10.86
4	-4.74	0.71	27.91	0.49	0.12	0.44	10.91
5	-3.12	0.67	29.23	0.49	0.11	0.42	11.10
6	-1.65	0.66	30.31	0.49	0.12	0.39	12.18
7	-0.21	0.65	28.09	0.47	0.11	0.43	11.66
8	1.28	0.68	27.27	0.47	0.10	0.42	12.93
9	3.12	0.73	26.08	0.48	0.13	0.27	12.27
10	8.23	0.92	19.85	0.53	0.13	0.26	12.75

The most important question is whether BIR is informative about individual stock returns in a market crash. We formally answer this question in Section 2.3. Here, we illustrate by forming ten value-weighted portfolios at the end of month t, based on $\text{BIR}_{i,t-1}$, calculated using daily data from months $t-6$ to $t-1$. The realized return of each portfolio $r_{t+1}^{(p)}$ is for month $t+1$. Each portfolio's value-weighted average $\text{BIR}_{i,t-1}$ is coded $\text{BIR}_t^{(p)}$ with a subscript t as the portfolio weight (the market capitalization at the end of month t) is known in month t.

Figure 1 plots the time series average of realized return $r_{t+1}^{(p)}$ of each portfolio p against the time series average of the implied return $\text{BIR}_t^{(p)}$. In the full sample, the realized return broadly increases in the implied return, even though the relationship is not strictly monotonic. When a market crash occurs, the realized returns align well with the implied return amongst the vulnerable groups (i.e., $p = 1,\ 2,\ 3$). In contrast, the realized return is indistinguishable from the market return for the resilient groups. This is consistent with our view that the positive BIR, as shown in last three rows of Table 1, primarily reflects measurement errors. The strength of our procedure is that it identifies the worst performers in a market downturn.

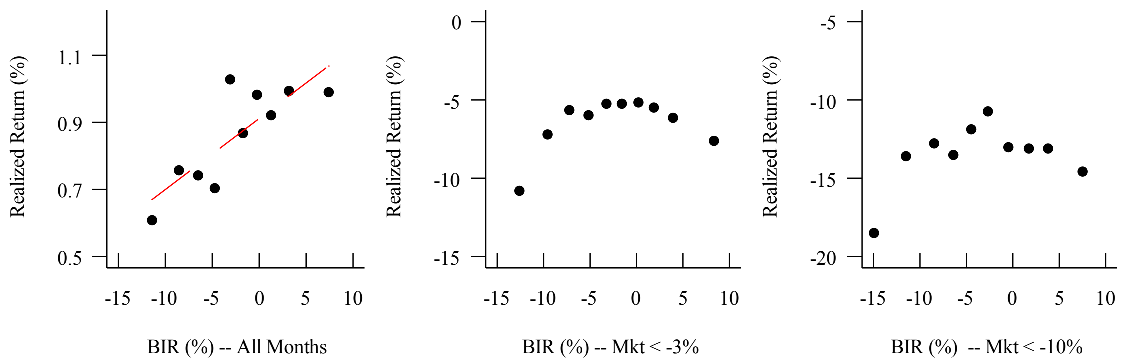

Figure 1: **The relationship between time series average $\mathbf{BIR}_t^{(p)}$ and return $r_{t+1}^{(p)}$.** The horizontal axis plot $\mathrm{BIR}_t^{(p)}$, which is the average $\mathrm{BIR}_{i,t-1}$ of each portfolio in Panel B of Table 1 weighted by market capitalization in month t. The vertical axis plots $r_{t+1}^{(p)}$, which is the realized return of each portfolio in month $t+1$. Both are value-weighted with month t market capitalization. The right panel uses all months; the middle panel uses months with a market (S&P 500) realized return below -3%; the right panel uses months with a market realized return below -10%.

2.3 Results of the Standard Asset Pricing Tests

2.3.1 Fama and MacBeth (1973) regression

We now formally study the pricing of BIR, starting with the Fama and MacBeth (1973) regression. In each month $t+1$, we regress the realized return $r_{i,t+1}$ on $\mathrm{BIR}_{i,t-1}$. As the subscript indicates, we skip one month after obtaining $\mathrm{BIR}_{i,t-1}$, which is estimated with a six-month rolling window. Table 10 in the Internet Appendix shows the robustness of the results with different rolling windows and skipping periods. The main analysis uses the data from 1996 to 2019, which was the data available to us when we finish the first version of this paper. In Tables 11 and 12 in the Internet Appendix, we extend the sample with data from 2020 and 2021, and conclusions remain unchanged.

Table 2 reports the Fama and MacBeth (1973) regression results, where most of the control variables are based on Chen and Zimmermann (2022) except for the risk-neutral moments we calculate ourselves. Using the results in Column 1, the t-stat for β_{t+1}^{BIR} over the full sample is 3.26, which approximately translates to an annual Sharpe ratio of $\frac{3.26}{\sqrt{2019-1996}} = 0.68$. Although the Sharpe ratio is meaningful, a t-value of around three may seem statistically low compared to figures reported in recent research on the cross-section of returns (Harvey, Liu, and Zhu, 2016). However, our test uses only the most liquid stocks (constituents of the S&P 500), with a much shorter data period (from 1996). A fairer comparison is with the performance of the established characteristics using the same sample, which shows weak cross-sectional predictability among the S&P 500 stocks (as found, for exmaple, by Nagel, 2005; Martin and Wagner, 2019). In all of the specifications of Table 2, the t-statistics for β^{CAPM}, $SIZE$, and BE/ME never exceed two and are not reported for brevity. Controlling for investment INV and profitability OP does lower the significance level of BIR. However, this is simply because the OP values are missing for one-third of the observations, which are consequentially dropped from the regression. In Columns 4 - 7, we create a new variable OP^* by replacing the missing OP values with zero, which restores the significance level of BIR. In Columns 5 and 6, the result of BIR is consistent over time, albeit slightly better in the past ten years. We think this is partly attributable to the low liquidity in the options market in the early years of the sample. As we see later in Section 2.3.2, a strategy based on BIR performs poorly in the first three years of the sample period, i.e., from 1996 to 1999.

In Column 7, we control for related covariance and characteristics identified in the literature, namely the volatility beta β^{VIX} from Ang, Hodrick, Xing, and Zhang (2006b), and the second and third risk-neutral moments RNV and RNS from Bakshi, Kapadia, and Madan (2003). For these three variables, we use the values at the end of month t (rather than

$t-1$ as BIR) to give them the best chance of predicting future returns. We also control for volatility spread and skew, the results of which are incorporated into Table 4. Finally, given that BIR measures one kind of downside risk, we control for downside betas of Ang, Chen, and Xing (2006a) and Kelly and Jiang (2014). For brevity, we do not report this set of results as the downside betas are insignificant in the sample and period considered here, and the coefficient of BIR is essentially unchanged.

The statistical significance remains stable across various specifications, indicating that BIR contains information independent of the established predictors. Although bits and pieces have appeared before, such as treating VIX as a state variable (Ang, Hodrick, Xing, and Zhang, 2006b) or utilizing the moneyness dimension from the option prices (Xing, Zhang, and Zhao, 2010), the procedure in this paper combines different elements and recovers the return in a market downturn in a nearly "model-free" way, which is unique in the literature.

As for the economic significance, the coefficient for each month can be seen as a trading strategy with a weight proportional to the normalized value of $\mathrm{BIR}_{i,t-1}$:

$$\beta_{t+1}^{BIR} = \frac{\mathrm{cov}\left(\mathrm{BIR}_{i,t-1}, r_{i,t+1}\right)}{\mathrm{var}\left(\mathrm{BIR}_{i,t-1}\right)} = L_t^{BIR} \times \sum_i w_{i,t}^{BIR} r_{i,t+1}, \tag{9}$$

where $L_t^{BIR} = \frac{1}{\sigma(\mathrm{BIR}_{i,t-1})}$ is the leverage and $w_{i,t}^{BIR} = \frac{\mathrm{BIR}_{i,t-1} - \mu(\mathrm{BIR}_{i,t-1})}{\sigma(\mathrm{BIR}_{i,t-1})}$ is the normalized signal for stock i. This strategy trades almost all S&P500 stocks each month, and is leveraged by a factor of L_t^{BIR}, which is time-varying and scales down in bad times when the dispersion of BIR is larger. Although the leverage is high (around 17, based on the average number from Table 1), the t-stat of β_{t+1}^{BIR}, which is related to the Sharpe ratio of the strategy (around 0.68), is not affected by the leverage. The Sharpe ratio is comparable to the information ratio of a tradable strategy based on BIR, as shown below.

Table 2: **Fama-MacBeth Regression**

This table reports the time series average coefficients and standard errors of the Fama-MacBeth regression of return $r_{i,t+1}$ on a list of variables including $\text{BIR}_{i,t-1}$, β, ME, B/M, INV, OP, and MOM as in Table 1; OP^*, which sets the missing values of OP to zero; REV, which is the return in month t; β^{VIX}, which is the volatility beta in Ang et al. (2006b) estimated using daily data in month t; RNV and RNS, which are, respectively, the second (not annualized, not in percentage form) and third risk-neutral moments in Bakshi et al. (2003), constructed using 30-day maturity data from the volatility surface at the end of month t. The standard errors in parenthesis are adjusted according to Newey and West (1987), with six lags.

	The dependent variable is $r_{i,t+1}$ in percentage points						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full	Full	Full	Full	<=2009	>2009	Full
$\text{BIR}_{i,t-1}$	2.35***	1.87***	1.33**	1.82***	1.63*	2.08***	1.51***
	(0.72)	(0.54)	(0.64)	(0.58)	(0.87)	(0.67)	(0.56)
β, ME, B/M	N	Y	Y	Y	Y	Y	Y
INV			-0.09	-0.18*	-0.16	-0.21	-0.22**
			(0.10)	(0.10)	(0.11)	(0.16)	(0.09)
OP			0.02				
			(0.02)				
OP^*				0.00	-0.01	0.01	0.00
				(0.02)	(0.04)	(0.01)	(0.02)
MOM			-0.08	0.02	-0.13	0.22	0.09
			(0.41)	(0.42)	(0.67)	(0.38)	(0.38)
REV			-1.26*	-1.37*	-1.33	-1.43	-1.56**
			(0.67)	(0.70)	(0.98)	(0.95)	(0.68)
β^{VIX}							-8.89
							(11.32)
RNV							-10.09
							(11.12)
RNS							0.28***
							(0.10)

2.3.2 Alpha from betting with butterfly (BwB)

We now construct an implementable strategy called "Betting with Butterfly" (BwB). Following the convention of Fama and French (1993), we trade the top and bottom 30% of stocks.

The return from BwB is

$$r_{t+1}^{\mathrm{BwB}} = \frac{r_{t+1}^{(10)} + r_{t+1}^{(9)} + r_{t+1}^{(8)}}{3} - \frac{r_{t+1}^{(3)} + r_{t+1}^{(2)} + r_{t+1}^{(1)}}{3}. \tag{10}$$

The subscript indicates that the portfolio formation takes place at the end of month t, which is one month after determining $\mathrm{BIR}_{i,t-1}$. The return is realized in the following month, $t+1$. From August 1996 to December 2019, the average monthly return of BwB is 0.26%. BwB is a relatively low-risk investment, with a monthly volatility of 2.1%. BwB trades only the liquid S&P 500 stocks, and it has low turnover due to the signal being estimated over a rolling window. As a result, BwB is highly practical and easy to implement.

Table 3 reports the result from the time series regressions of BwB return on the common risk factors. The relatively short period puts BwB at a disadvantage from a statistical point of view. Nevertheless, the alpha ranges from 0.28% to 0.39% monthly, and all are significant at the conventional levels. The annualized information ratio with respect to the three-factor model is $\frac{\frac{0.39}{0.12}}{\sqrt{2019-1996}} = 0.66$, which is comparable to the Sharpe ratio reported in Section 2.3.1.

BwB is a long-short strategy, and the long and short legs of BwB share many characteristics, as presented in Table 10 in the Internet Appendix. For example, under the five-factor model of Hou, Mo, Xue, and Zhang (2020), both legs have a significant negative loading on the size factor and a positive loading on the investment factor. However, the alpha primarily comes from the short leg (i.e., Groups 1, 2, and 3). In unreported results, we find the alpha with just the short leg is more significant than the alpha of the long-short strategy. As previously noted, the procedure in this paper is more successful in identifying the worst performers, which BwB shorts to hedge the crash. This allows BwB to benefit from a market downturn, leading to a negative loading on the market factor. However, BwB suffers a heavy loss when (and if) the market rebounds as the short leg consists of stocks that suffer most from a market crash and thus potentially benefit most from a market rebound.

BwB negatively correlates with the value factor HML, which performs poorly in extreme

Table 3: **Regressing $r_t^{\mathbf{BwB}}$ on the Common Risk Factors**

The dependent variable is the return from BwB in percentage points. The common risk factors comprise (1) FF3: the three stock factors of Fama and French (1993), (2) FF5: the five factors of Fama and French (2015), (3) M4: the four factors of Stambaugh and Yuan (2017), and (4) Q5: the five factors of Hou, Mo, Xue, and Zhang (2020).

	FF3	FF5		M4		Q5
Alpha	0.39***	0.28**	Alpha	0.32**	Alpha	0.34**
	(0.12)	(0.12)		(0.14)		(0.13)
MktRf	-0.16***	-0.10***	MktRf	-0.08**	R_MKT	-0.13***
	(0.03)	(0.03)		(0.04)		(0.03)
SMB	-0.03	-0.02	SMB	-0.04	R_ME	-0.02
	(0.04)	(0.04)		(0.04)		(0.04)
HML	-0.10**	-0.24***	MGMT	-0.02	R_IA	0.02
	(0.04)	(0.05)		(0.05)		(0.06)
CMA		0.28***	PERF	0.14***	R_ROE	0.06
		(0.08)		(0.03)		(0.06)
RMW		0.08			R_EG	-0.02
		(0.06)				(0.07)
#Obs	281	281		245		281

events. For example, Koijen, Lustig, and van Nieuwerburgh (2017) document that value stocks substantially reduce cash dividend payments during a recession. The loadings on the rest of the factors are low, except in Column 2, where both HML and CMA appear on the right-hand side. If HML is omitted, the loading on the Fama-French investment factor CMA drops essentially to zero, similar to the loading on the Hou et al. (2020) investment factor R_IA in Column 4. With the five-factor model of Hou et al. (2020), the only two significant coefficients are the negative loading on the market factor and the positive intercept.

The loss from a market rebound is similar to that of the momentum strategy (Daniel and Moskowitz, 2016). However, the momentum strategy gains when market conditions go from bad to worse, while BwB profits at the beginning of the crash. Although the above models do not include momentum due to the high turnover, we show in Table 2 that MOM has essentially no effect on the pricing of BIR. In unreported results, we find that adding the momentum factor to FF3 and FF5 slightly (but non-significantly) increases the alpha of BwB.

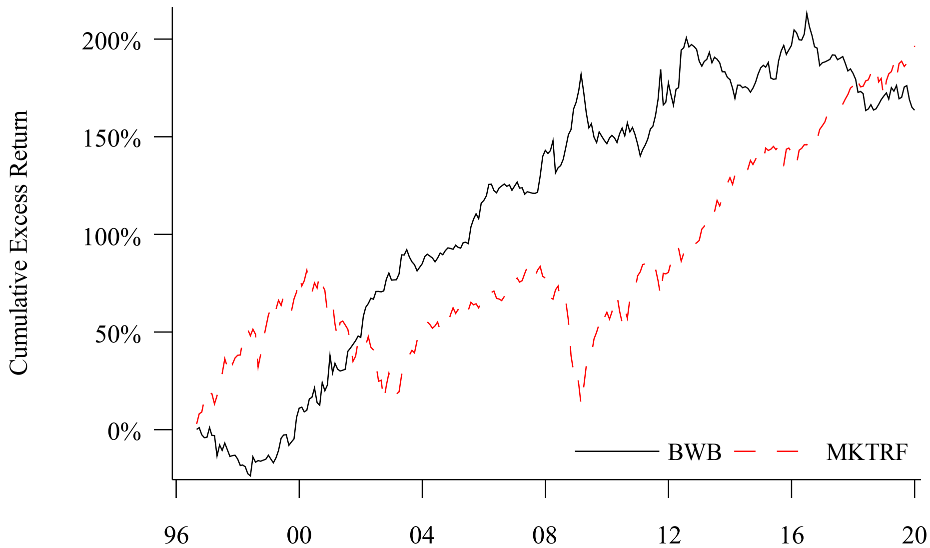

Figure 2: **Cumulative excess returns of BwB and the market**, both of which are scaled to have an ex post annual volatility of 16%.

We also experiment with alternative methods to construct tradable strategies. Among different strategies, the alpha of BwB in Table 3 is an approximate representation of the lower bound. In our experiments, the highest alpha is achieved, perhaps unsurprisingly, by longing the equal-weighted bottom decile while shorting the equal-weighted top decile, which generates a positive alpha between 0.47% (relative to FF5, t=2.8) and 0.61% (relative to Q5, t=3.1). Given the academic focus of this paper, we do not report the full results. Instead, we turn to a more intriguing question: if BwB hedges market downturns, why does it have a positive alpha?

2.4 Channels for the Positive Alpha

2.4.1 A mispricing explanation

A possible explanation for the positive alpha is that the participants in the option market have superior information that becomes slowly incorporated into the stock prices. This explanation dates back to Black (1975) that options embed leverage. It also assumes that the consequences of trading options (implied volatility and volume) are ignored by stock traders, which is plausible over short windows before earnings announcements (Jin, Livnat, and Zhang, 2012), takeover announcements (Augustin, Brenner, and Subrahmanyam, 2019), or analyst recommendation revisions (Hayunga and Lung, 2014). However, there is also an argument that the evidence reflects the carry cost of shorting (Muravyev, Pearson, and Pollet, 2022) or short-lived price pressure (Goncalves-Pinto, Grundy, Hameed, van der Heijden, and Zhu, 2020).

The information in the context of this paper, which is the relative performance in the event of a market crash, differs from the type of insider information in the abovementioned literature. Nevertheless, we perform tests using variables that relate to the relative informativeness of options and the stock market: (1) the option-to-stock volume ratio, as we expect that relatively active trading reflects an information advantage on the options side (Roll, Schwartz, and Subrahmanyam, 2010); (2) change in the option-to-stock volume ratio; (3) percentage change in options volume; (4) negative percentage change in stock volume; (5-6) call minus put volatility spread (Bali and Hovakimian, 2009); and (7-8) at-the-money call minus out-of-the-money put volatility skew (Xing, Zhang, and Zhao, 2010). Table 4 reports the results after interacting the six variables with BIR. Under the mispricing explanation, the interaction terms between BIR and the measures of option information advantage should be positive. However, we fail to obtain a significant positive coefficient for any of the interactions, suggesting that the predictability of BIR is similar for all stocks, regardless of the relative informativeness between option and stock.[3]

[3]For volatility spread and skew, we use the month t value. Both the coefficients on the variables and

Table 4: **Interacting BIR with Measures of Relative Informativeness**

In this Fama-MacBeth regression, the variable RI_i corresponds to (1) the option-to-stock volume ratio, (2) change in the option-to-stock volume ratio, (3) percentage change in options volume, (4) (negative) percentage change in stock volume, (5-6) volatility spread from call minus put, and (7-8) volatility skew from at-the-money call minus out-of-the-money put. The control variables comprise β^{CAPM}, ME, B/M, INV, OP, MOM, and REV. The standard errors in parentheses are adjusted according to Newey and West (1987), with six lags.

	The dependent variable is $r_{i,t+1}$ in percentage points							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RI_i=	O/S	$\Delta O/S$	ΔO	$-\Delta S$	C-P Spread		A-O Skew	
$\mathrm{BIR}_{i,t-1}$	2.16***	1.90***	1.71***	1.30*	1.80***	1.61***	1.59***	1.15
	(0.70)	(0.59)	(0.62)	(0.73)	(0.57)	(0.59)	(0.55)	(0.89)
$\mathrm{BIR}_{i,t-1} \times \mathrm{RI}_i$	-3.02	-7.82	0.04	1.92		-8.23		-2.36
	(4.28)	(8.90)	(0.48)	(1.51)		(27.85)		(8.70)
RI_i	0.28	0.70	-0.03	0.21*	4.46***	3.26*	2.29***	1.84**
	(0.49)	(0.60)	(0.02)	(0.11)	(1.27)	(1.68)	(0.85)	(0.78)
Controls	β^{CAPM}, ME, B/M, INV, OP^*, MOM, and REV							

We have two additional reasons to downplay the possibility of a mispricing channel. First, the sample in this paper comprises the most liquid and carefully analyzed stocks, i.e., the constituents of the S&P 500 composite index. Second, although the alpha might be consistent with mispricing, the loss incurred by BwB when the market rebounds is difficult to explain. Therefore, mispricing is unlikely the main driver of the return of BwB.

2.4.2 A rational explanation

The alpha of BwB has a rational explanation consistent with its loading on the market factor. From the point of view of a representative household, BwB loses money at the most inopportune time: the bottom of the crash when investors' marginal utility is at its highest.

the interaction terms are non-significant with the month $t-1$ value. The coefficient on BIR becomes non-significant only if we include the interaction term with the month t volatility skew, as shown in the last column of Table 4. We do not believe that the pricing of BIR is due to its correlation with volatility skew, because if we omit the interaction term, they are both significant, as we see in Column 7 of Table 4.

As a result, BwB requires compensation in the form of a positive alpha over the full sample. For investors whose consumption process differs from that of the representative household, it is possible to benefit from investing in BwB.

To investigate this rational explanation, we incorporate BIR into the list of characteristics of the demand system proposed by Koijen and Yogo (2019). The demand system models investor j's portfolio weight in asset i, $\omega_{jt}(i)$, as an exponentially affine function of characteristics $\{x_{kt}(n)\}$,

$$\log \omega_{jt}(i) = a_{it} + \sum_{k=0}^{K} \beta_{k,jt} x_{kt}(i) + \beta_{BIR,jt} \mathrm{BIR}_{i,t} + \log \epsilon_{jt}(i)$$

where the constant a_{it} absorbs the choice of outside assets. We use the same list of characteristics $\{x_{kt}(n)\}$ as Koijen and Yogo (2019), which includes market and book equities (both in log), profitability, investment, dividend, and market beta. For the investors, we focus on three types of investors who are relatively flexible: investment advisors, mutual funds, and households. For the first two types, we calculate the AUM-weighted average loading $\beta_{BIR,jt}$ across the same type of investor, which is $\beta_{BIR,t}$. The time series of $\beta_{BIR,t}$ thus reflects the average tendency of each type of investor to pursue the BwB strategy.

We make three changes to the estimation procedure. First, to ensure that the characteristics are available to investors when they are making decisions, Koijen and Yogo (2019) match the stockholding data at the end of a quarter (let us call it month t) with the CAPM beta at the end of month $t-1$, and other accounting-related characteristics at the end of month $t-6$, assuming a 6-month gap for accounting information release. For the newly added BIR, to be conservative about the timing, we use the end-of-month $t-1$ value, which is the same as the timing of the CAPM beta. Second, the definition of outside assets is different. In Koijen and Yogo (2019), the inside assets comprise common stocks with valid stock characteristics $\{x_{kt}(n)\}$ and returns. When BIR is added, the inside assets need to have non-missing values of BIR, which restricts the inside assets to S&P 500 stocks. All other assets become outside assets, which are absorbed by the constant a_{it}. Koijen and

Yogo (2019) individually estimate the coefficients for investors with enough number of inside assets, and assign the other investors into groups. We adopt the same grouping as Koijen and Yogo (2019). Luckily, we do not encounter convergence issues even though the number of assets is at most 500 in our case. Finally, in Koijen and Yogo (2019), the stock characteristics are winsorized at 2.5% and 97.5% by month (for dividend to book equity, they impose non-negativity and winsorize only at the 97.5% level). However, as BIR is a return without extreme values, we winsorize BIR at 0.5% and 99.5% levels by month.

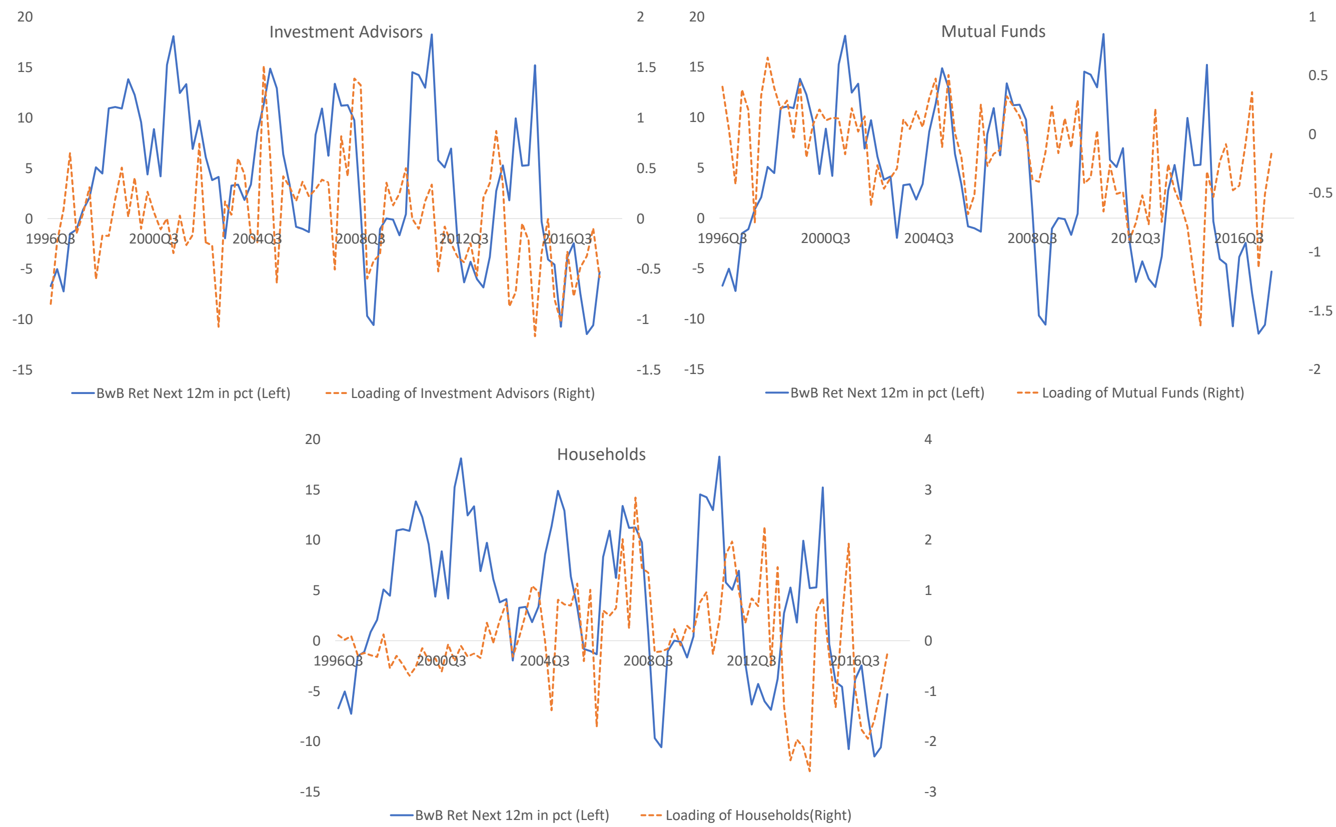

Figure 3: **Time series of $\beta_{BIR,t}$ and future BwB return**. The blue lines show the future 12-month returns of BwB, which use the primary axis. The dashed orange lines are $\beta_{BIR,t}$, which is the average portfolio weight's loading on BIR shown separately for investment advisors, mutual funds, and households.

As recommended by Koijen and Yogo (2019), we use the nonlinear GMM estimation method. The time series of estimated loading $\beta_{BIR,t}$ of each investor type is plotted in Figure 3 alongside the future 12-month return of BwB. We find that investment advisors (which include hedge fund managers) are good at timing BwB. Throughout the sample period, investment advisors skillfully time the market downturns by increasing their weights on the resilient stocks (or exiting stocks whose prices are about to fall dramatically), thus benefiting from the high BwB return when the market crashes, consistent with their behavior in the technology bubble (Brunnermeier and Nagel, 2004). In comparison, the loading of mutual funds is negative on average. We can see that the change in mutual fund loadings predicts future BwB returns prior to 2008 but not afterward. Finally, households appear to be on the other side of BwB for most of the sample period, in the sense that the sharp changes in households' loadings typically predict BwB returns in the opposite direction. However, as emphasized, we cannot label such portfolio choices simply as "smart" vs. "dumb" money. As investors differ in their consumption processes, the same strategy might appear safe to one group of investors and very risky to another group, even though the alpha of the strategy is positive relative to academic risk factors.

2.4.3 Further implications of the rational explanation

Due to data limitations, our analysis focuses on the S&P 500 stocks. However, the BwB time series allows us to expand our attention to all stocks. Following the rational explanation that shows BwB is risky, one immediate implication is that a stock that co-moves positively with BwB is also risky and thus requires compensation in the form of a higher return. We test this implication on all common stocks listed on the NYSE, NASDAQ, and AMEX from July 1997 (one year after the first BwB return) to December 2019. With a slight alteration to the notation, we estimate $\beta_{i,t}^{BwB}$ with the daily returns of stock i and BwB, using a one-year rolling window ending at month t that contains at least 200 valid observations. When estimating $\beta_{i,t}^{BwB}$, we experiment with three sets of control variables: (1) MktRf only; (2) MktRf, SMB, and HML; and (3) MktRf, SMB, HML, CMA, and RMW. We refer to the resulting $\beta_{i,t}^{BwB}$ as the CAPM-version, FF3-version, and FF5-version, respectively. We then test the pricing of $\beta_{i,t}^{BwB}$ in the cross-section of all common stock returns and report the results in Table 5.

Table 5: **Pricing of $\beta_{i,t}^{BwB}$ in the Cross-Section of All Common Stock Returns**

We estimate $\beta_{i,t}^{BwB}$ from the daily returns of stock i and BwB using a one year rolling window ending at month t that contains at least 200 valid observations. When estimating $\beta_{i,t}^{BwB}$, we use three sets of control variables: (1) MktRf only; (2) MktRf, SMB, and HML; and (3) MktRf, SMB, HML, CMA, and RMW. We refer to the resulting $\beta_{i,t}^{BwB}$ as the CAPM-version, FF3-version, and FF5-version, respectively. The control variables are defined in Table 2. The standard errors in parentheses are adjusted according to Newey and West (1987), with six lags. The data ranges from July 1997 to December 2019 due to the availability of $\beta_{i,t}^{BwB}$, and includes all common stocks listed on the NYSE, NASDAQ, and AMEX.

Panel A	Time-series average of cross-sectional statistics of $\beta_{i,t}^{BwB}$								
	Mean	Std	Min	25%	50%	75%	Max	Skew	N
CAPM-version	-0.10	0.79	-7.22	-0.45	-0.05	0.31	6.19	-0.64	4460
FF3-version	-0.04	0.78	-7.24	-0.39	0.01	0.36	6.35	-0.63	4460
FF5-version	-0.02	0.79	-7.50	-0.38	0.01	0.37	7.02	-0.42	4460

Panel B	Fama-Macbeth regression coefficient on $\beta_{i,t}^{BwB}$								
	The dependent variable is $r_{i,t+1}$ in percentage points								
	CAPM-version			FF3-version			FF5-version		
$\beta_{i,t}^{BwB}$	0.17*	0.17*	0.20**	0.19**	0.18**	0.21***	0.12*	0.12*	0.15**
	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
β, ME, B/M	Y	Y	Y	Y	Y	Y	Y	Y	Y
INV, OP^*		Y	Y		Y	Y		Y	Y
MOM, REV			Y			Y			Y

The results across all specifications confirm that a positive covariance with BwB is associated with a higher expected return. Of the three versions, the FF3-version performs slightly better. The coefficients from the FF3-version suggest that a 2-standard-deviation difference in $\beta_{i,t}^{BwB}$ creates a return spread of $2 \times 0.78 \times 0.21\% = 0.33\%$ per month, or 3.9 percent per annum, which is economically meaningful and plausible. Of course, the inclusivity of the sample makes it difficult to implement the result as a trading strategy. In unreported results, we also run a horserace regression of BIR and $\beta_{i,t}^{BwB}$ using the subsample of S&P 500 stocks. We find that the coefficient of BIR is essentially unchanged by $\beta_{i,t}^{BwB}$, which itself is not significant in this subsample.

3 Aggregated BIRM in the Time Series

It is meaningful to aggregate BIR in each cross-section, as BIR refers to the future returns in the same state; that is, in a market crash. The aggregated series measures the return of the S&P 500 index when it crashes. Unlike the cross-sectional analysis in Section 2, which is mostly monthly, the time series tests in this section use annualized returns. As a result, we use an annualized series, BIRM (BIR of the Market), which is defined as

$$\mathrm{BIRM}_t = A(\sum_i \mathrm{MktCap}_{i,t}\mathrm{BIR}_{i,t} / \sum_i \mathrm{MktCap}_{i,t}) \tag{11}$$

where $\mathrm{MktCap}_{i,t}$ is the market capitalization of stock i at the end of month t, and $A(x) = (1+x)^{12} - 1$ annualizes the series. As plotted in Figure 4, BIRM is significantly lower during recession periods. It is also responsive to other crises that unfold slowly, such as the worsening of the European debt crisis around 2011.

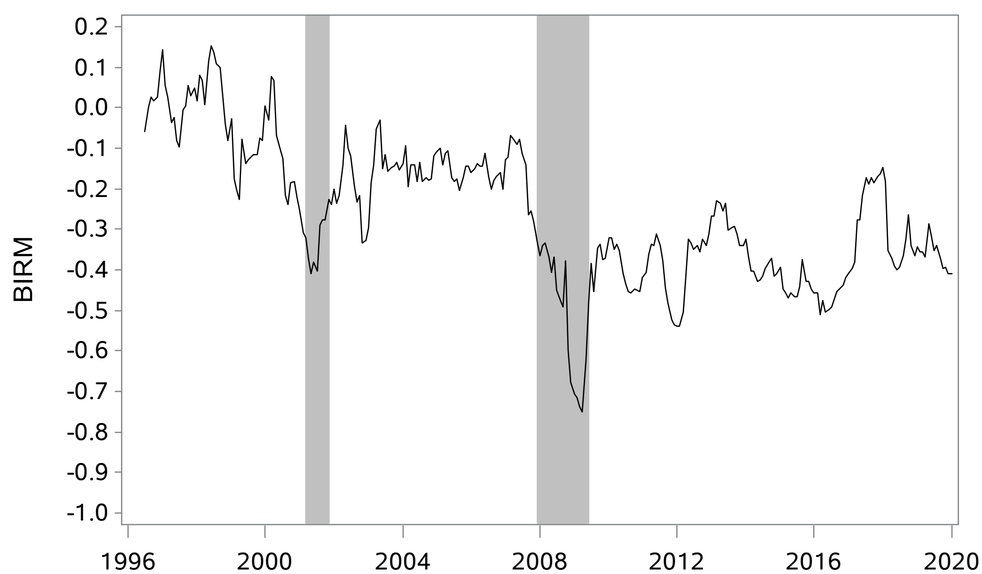

Figure 4: **Time series of BIRM$_t$**, which is the average of $\mathrm{BIR}_{i,t}$ across all stocks, weighted by the market cap at the end of month t and then annualized. The shaded periods are NBER recessions.

3.1 BIRM and Expected Return

The severity of a crash is an important determination of equity risk premium. Quantitatively, if we regress a measure of equity risk premium $\mathrm{EP}_t = \frac{1}{n}\mathbb{E}_t\left(R_{M,t\to t+n} - R_{f,t}\right)$ on BIRM_t, i.e.,

$$\mathrm{EP}_t = \underbrace{\beta_0^{EP}}_{>0} + \underbrace{\beta_B^{EP}}_{<0} \times \mathrm{BIRM}_t + \epsilon_t^{EP} \tag{12}$$

the resulting coefficient β_B^{EP} equals the difference between the physical (p^M of Eq. 1) and risk-neutral probability ($p^M B^{-\gamma}$, with $B^{-\gamma} > 1$) of a market crash. This difference is negative, as the pricing kernel is higher in a market crash. The constant term β_0^{EP} captures the remaining equity risk premium due to the varying crash probability or other risks, such as rare disasters.

We rely on the argument of Martin (2017) that risk-neutral variance $R_{f,t} \cdot \mathrm{SVIX}^2_{t\to t+n}$ is an approximate measure of the equity risk premium EP_t. In Table 6, we regress $R_{f,t} \cdot \mathrm{SVIX}^2_{t\to t+n}$ on BIRM_t, with the log dividend yield dp_t as the control.[4] Dividend yield is countercyclical and thus one of the most commonly used proxies for the equity risk premium. However, BIRM subsumes the effect of the dividend yield. The key coefficient is β_B^{EP}, which is negative as expected. The intercept β_0^{EP} is 3%, which is around half of the equity risk premium, suggesting that BIRM explains the other half of the average equity risk premium. In terms of time variation, the R^2 of the univariate regression ranges from 18% to 31%, suggesting that BIRM is also an important driver of the varying equity risk premium.

It is also unsurprising that BIRM is informative about the realized return, which we verify in Table 13 and 14 the Internet Appendix with predictive regressions. However, unlike Equation (12), which holds in every period, the coefficients obtained from the predictive

[4] To make the intercept interpretable, we use $dp_t - \overline{dp}$, where $\overline{dp} = -4.2$ is the approximately in-sample mean (the natural logarithm of 1.5%). This adjustment turns out to be unnecessary in this regression, as the coefficient on $dp_t - \overline{dp}$ is essentially zero when controlling for BIRM_t. The SVIX series from Ian Martin's website ends in January 2012. The result with the extended series that we calculate ourselves is similar. The Internet Appendix shows that the results are stable throughout the sub-periods by repeating the regression in consecutive five-year windows.

Table 6: **BIRM_t and Equity Risk Premium**

This table reports the results of Regression (12). The proxy for equity risk premium EP_t is the end-of-month number of $R_{f,t} \cdot \text{SVIX}^2_{t \to t+n}$ from January 1996 to January 2012. dp_t is the natural logarithm of the annual dividend to price ratio for the S&P 500. We set $\overline{dp} = -4.2$, which is the natural logarithm of 1.5%. In the parentheses are the Hansen and Hodrick (1980) standard errors with a bandwidth of 18.

Dependent	$\text{EP}_t = R_{f,t} \cdot \text{SVIX}^2_{t \to t+n}$					
	(1)	(2)	(3)	(4)	(5)	(6)
		$n = \frac{1}{12}$			$n = 1$	
Intercept	0.03***	0.04***	0.03***	0.03***	0.04***	0.03***
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
BIRM_t	-0.09**		-0.07**	-0.07**		-0.07***
	(0.04)		(0.03)	(0.03)		(0.02)
$dp_t - \overline{dp}$		0.07**	0.03		0.04	0.01
		(0.03)	(0.02)		(0.03)	(0.02)
R^2	0.18	0.14	0.21	0.31	0.12	0.31
N	188	188	188	188	188	188

regressions change significantly over time, depending on whether a market crash occurs. Therefore, we do not find evidence for out-of-sample predictability.

3.2 BIRM as the expectation of return

We establish that BIRM_t is negatively correlated with the equity risk premium, which is known professionally as the expected excess return. On the other hand, BIRM_t represents the expectation of return in one of the likely future states; that is, a market crash. Several prominent surveys feature questions about the expectation of return. Could the answer be influenced by the expectation of crash severity?

We find that this is the case. Following Greenwood and Shleifer (2014), we focus on the

Gallup sentiment series, which calculates the difference between the percentage of bullish and bearish investors. Although this series does not directly measure the expectation of return, it is highly correlated (84% in levels and 65% in one-month changes) with the answer to another question, "what is the estimate of the percentage return you *expect* on the market over the next 12 months?", which is only available for a shorter period. Therefore, the Gallup sentiment series, shown by the dashed red line in Figure 5, is used to describe the variation in the expectation of return among the survey participants.

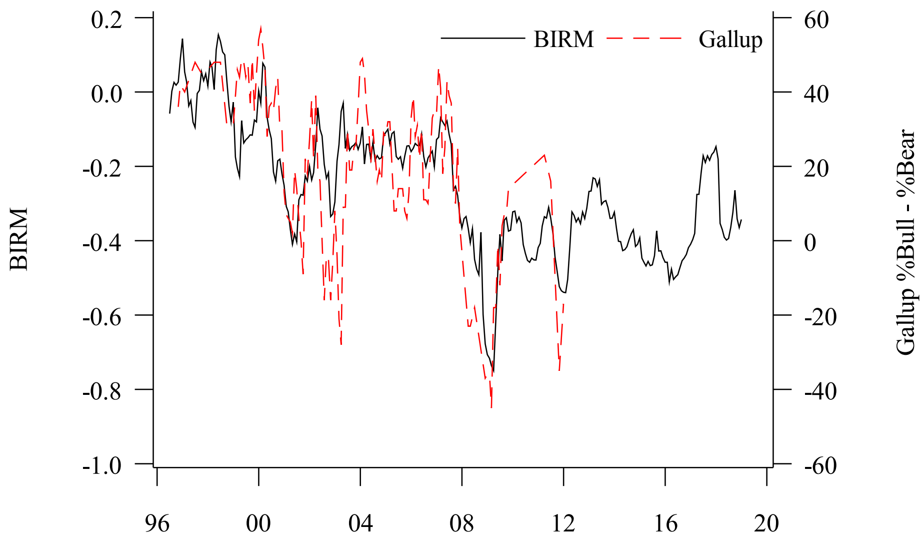

Figure 5: **BIRM$_t$ and the Gallup sentiment series**, the latter being calculated as the difference between the percentage of bullish and bearish investors.

The survey expectation has a notable characteristic of being pro-cyclical, which differs from most theoretical measures of expected return. This difference is one of the most fundamental puzzles, and it appears that the only way to resolve it is through behavioral explanations (Barberis et al., 2015; Hirshleifer et al., 2015; Cassella and Gulen, 2018; Nagel and Xu, 2022). BIRM is positively correlated with the survey expectation, as demonstrated in Figure 5, and on the other hand, based on a pure rational consideration, negatively correlated with the expected return, as shown in Section 3.1. The question remains as to why BIRM, which measures the return in a market crash, plays a central role in the expectations of the survey participants.

We think the answer lies in the interpretation of the word "expect," as pointed out in Cochrane (2017). The time series of BIRM allows us to give content to this argument. After all, the colloquial meaning of "expect" is to "regard (something) as likely to happen." For example, if survey participants are fixated on a market crash that they consider to be the most likely outcome in their own probability measure, then the return they expect, or *regard as likely to happen*, would be the return in that state, or BIRM. Of course, this answer differs from the textbook definition of "expect," which is the average value under the objective probability measure. This textbook answer is theoretically and empirically negatively correlated with BIRM, as we explain in Section 3.1.

3.3 Pricing of BIRM in the Cross-section of Stock Returns

As clear from Figure 4, a negative shock to BIRM is concurrent with a decrease in consumption. By Merton (1973), BIRM should carry a positive price of risk as all risk-averse utility maximizers choose to underweight stocks with positive exposure to BIRM. To test this implication, we use an APT specification with the only systematic risks being the market return and news about BIRM. In this specification, the pricing error of the test assets is

$$g_T\left(b\right)=\frac{1}{T}\sum_{t=1}^{T}u_t\left(b\right)\equiv\frac{1}{T}\sum_{t=1}^{T}R_t^e-\gamma^{\mathrm{M}}R_t^e\left(R_t^{\mathrm{M,e}}-\mu^{\mathrm{M}}\right)-\gamma^{\mathrm{B}}R_t^e\left(\Delta\mathrm{BIRM}_t-\mu^{\mathrm{B}}\right) \tag{13}$$

where R_t^e is a row vector of the excess returns of the test assets and $R_t^{\mathrm{M,e}}$ is the excess return of the market. $\Delta\mathrm{BIRM}_t=\mathrm{BIRM}_t-\mathrm{BIRM}_{t-1}$ is the (annualized) news about the severity of a crash. μ^{M} and μ^{B} are the in-sample means of $R_t^{\mathrm{M,e}}$ and $\Delta\mathrm{BIRM}_t$, respectively.[5] In addition to 25 size/BM portfolios, the test assets also include portfolios with weaker factor structures such as 10 beta portfolios and 10 iVol portfolios, following the advice of Lewellen, Nagel,

[5]Prior to estimation, we calculate μ^{M} and μ^{B} in-sample. In Table 15 of the Internet Appendix, we use the specification in which the means are estimated rather than calculated, similar to Maio and Santa-Clara (2012), and the point estimates for γ^{M} and γ^{B} are very similar.

and Shanken (2010).[6] The range of the data is from July 1996 to December 2019.

We are interested in the prices of risk γ^{M} and γ^{B}. It is common to interpret $R_t^{\mathrm{M},e}$ as the excess return to the wealth portfolio of the marginal investor. In this case, the γ^{M} parameter is the coefficient of relative risk aversion, so it should be positive and not too large. We also expect γ^{B} to be positive, as assets that perform poorly when BIRM decreases are considered risky. We estimate γ^{M} and γ^{B} in two stages. The first stage produces the parameter estimate

$$\hat{b}_1 = \mathrm{argmin}_{\{b_1\}} g_T(b_1)' \, g_T(b_1)$$

which minimizes the sum of the squared pricing error. The first stage estimate of γ^{B} is positive, though only marginally significant. We also perform the second stage, which minimizes the (squared) Sharpe ratio attainable with the unspanned residuals, or

$$\hat{b}_2 = \mathrm{argmin}_{\{b_2\}} g_T(b_2)' \, \hat{S}^{-1} g_T(b_2)$$

where $\hat{S}$ is the estimate of the spectral density matrix from the first stage. As we can see in Table 7, the second stage estimate of γ^{B} is positive and significant, suggesting that the severity of the market crash carries a positive price of risk in the cross-section of stock returns.

Table 7: **The Price of Risk via GMM Estimation**

We estimate parameters $b = \left\{\gamma^{\mathrm{M}}, \gamma^{\mathrm{B}}\right\}$ with the moment condition (13). The standard errors use the adjustment of Newey and West (1987), with six lags in each stage.

	First Stage		Second Stage	
	γ^{M}	γ^{B}	γ^{M}	γ^{B}
$\hat{b}$	2.15	5.09*	3.51**	7.60***
s.e.	(1.49)	(2.90)	(1.49)	(2.65)

[6] In unreported results, we also include the industry portfolios in test assets. However, due to the large cross-section and the short sample period, we do not find a significant price of risk for $\Delta\mathrm{BIRM}_t$.

The GMM approach uses the full sample to estimate the price of risk. In unreported results, we use another approach akin to the one in Section 2.4.3. We directly calculate the ΔBIRM_t beta of an individual stock with a rolling window and then estimate the price of risk using the Fama-MacBeth regression. However, since the frequency of BIRM_t is monthly, the resulting beta is very noisy and does not show a significant price of risk. Additionally, in the subsample of S&P 500 stocks, the ΔBIRM_t beta has essentially no effect on the pricing of BIR.

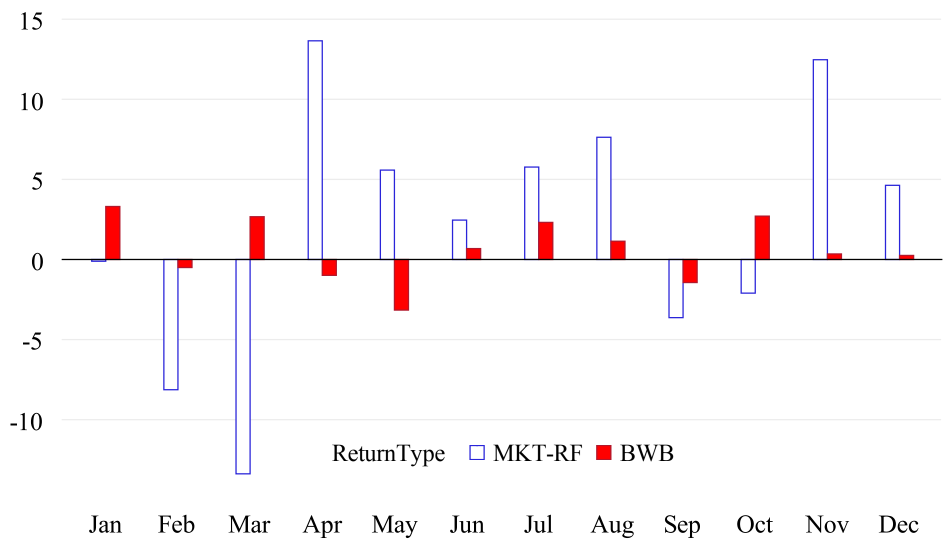

Figure 6: **Monthly excess returns of BwB and the market in 2020** measured in percentage points.

4 Concluding Remarks

This paper proposes an innovative approach to uncover the time-varying expectation of individual stock returns during a market crash, which can then be aggregated into a measure of the market return in a future crash. The key inputs are the prices of the butterfly options, which contain information about a narrow range of stock prices, and the VIX index, which contains information about a future market crash. The expectation of return is useful in both the cross-section and the time series.

After the circulation of this paper and the availability of data for 2020 from OptionMetrics, we find that the original conclusions in the paper remain unchanged by the new data. Figure 6 plots the excess return of the market and the long-short return of BwB in each month of 2020. First, BwB remains a low-risk bet, far less volatile than the market. Second, between February and March, there is a market crash of more than 20%, while BwB realizes a positive 2% return in this period by shorting the worst performers. When the market recoups its previous loss in April and May, BwB starts to lose, consistent with what we find in Section 2.3.2. The market realized an annualized Sharpe ratio of 0.9 in 2020, more than double the historical average. However, the Sharpe ratio of BwB in the same period is 1.1, which is even higher. Overall, the results for 2020 provide out-of-sample validation of the findings of this paper.

References

An, Byeong-Je, Andrew Ang, Turan G Bali, and Nusret Cakici, 2014, The joint cross section of stocks and options, *Journal of Finance* 69, 2279–2337.

Ang, Andrew, Joseph Chen, and Yuhang Xing, 2006a, Downside risk, *Review of Financial Studies* 19, 1191–1239.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006b, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Augustin, Patrick, Menachem Brenner, and Marti G Subrahmanyam, 2019, Informed options trading prior to takeover announcements: Insider trading?, *Management Science* 65, 5697–5720.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Bali, Turan G, and Armen Hovakimian, 2009, Volatility spreads and expected stock returns, *Management Science* 55, 1797–1812.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. V. Nikolaev, 2015, Deflating profitability, *Journal of Financial Economics* 117, 225–248.

Baqaee, David Rezza, 2020, Asymmetric inflation expectations, downward rigidity of wages, and asymmetric business cycles, *Journal of Monetary Economics* 114, 174–193.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-capm: An extrapolative capital asset pricing model, *Journal of Financial Economics* 115, 1–24.

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho, 2022, Survey data and subjective beliefs in business cycle models, Working paper.

Black, Fischer, 1975, Fact and fantasy in the use of options, *Financial Analysts Journal* 31, 36–41.

Bollerslev, Tim, and Viktor Todorov, 2011, Tails, fears, and risk premia, *The Journal of Finance* 66, 2165–2211.

Borovička, Jaroslav, Lars Peter Hansen, and José A Scheinkman, 2016, Misspecified recovery, *Journal of Finance* 71, 2493–2544.

Breeden, Douglas T, and Robert H Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621–651.

Brunnermeier, Markus K, and Stefan Nagel, 2004, Hedge funds and the technology bubble, *Journal of Finance* 59, 2013–2040.

Cassella, Stefano, and Huseyin Gulen, 2018, Extrapolation bias and the predictability of stock returns by price-scaled variables, *Review of Financial Studies* 31, 4345–4397.

Chang, Bo Young, Peter Christoffersen, and Kris Jacobs, 2013, Market skewness risk and the cross section of stock returns, *Journal of Financial Economics* 107, 46–68.

Chen, Andrew Y, and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 11, 207–264.

Cochrane, John H, 2017, Macro-finance, *Review of Finance* 21, 945–985.

Cremers, Martijn, Michael Halling, and David Weinbaum, 2015, Aggregate jump and volatility risk in the cross-section of stock returns, *Journal of Finance* 70, 577–614.

Cremers, Martijn, and David Weinbaum, 2010, Deviations from put-call parity and stock return predictability, *Journal of Financial and Quantitative Analysis* 45, 335–367.

Daniel, Kent, and Tobias J Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

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

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.

Goncalves-Pinto, Luis, Bruce D Grundy, Allaudeen Hameed, Thijs van der Heijden, and Yichao Zhu, 2020, Why do option prices predict stock returns? the role of price pressure in the stock market, *Management Science* 66, 3903–3926.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Hansen, Lars Peter, and Robert J Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hayunga, Darren K, and Peter P Lung, 2014, Trading in the options market around financial analysts' consensus revisions, *Journal of Financial and Quantitative Analysis* 49, 725–747.

Hirshleifer, David, Jun Li, and Jianfeng Yu, 2015, Asset pricing in production economies with extrapolative expectations, *Journal of Monetary Economics* 76, 87–106.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2020, An augmented q-factor model with expected growth, *Review of Finance* .

Jackwerth, Jens Carsten, and Marco Menner, 2020, Does the ross recovery theorem work empirically?, *Journal of Financial Economics* 137, 723–739.

Jensen, Christian Skov, David Lando, and Lasse Heje Pedersen, 2019, Generalized recovery, *Journal of Financial Economics* 133, 154–174.

Jin, Wen, Joshua Livnat, and Yuan Zhang, 2012, Option prices leading equity prices: Do option traders have an information advantage?, *Journal of Accounting Research* 50, 401–432.

Kadan, Ohad, and Xiaoxiao Tang, 2020, A bound on expected stock returns, *Review of Financial Studies* 33, 1565–1617.

Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *Review of Financial Studies* 27, 2841–2871.

Koijen, Ralph SJ, Hanno Lustig, and Stijn van Nieuwerburgh, 2017, The cross-section and time series of stock and bond returns, *Journal of Monetary Economics* 88, 50–69.

Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

Lu, Zhongjin, and Scott Murray, 2019, Bear beta, *Journal of Financial Economics* 131, 736–760.

Maio, Paulo, and Pedro Santa-Clara, 2012, Multifactor models and their consistency with the icapm, *Journal of Financial Economics* 106, 586–613.

Martin, Ian, 2017, What is the expected return on the market?, *Quarterly Journal of Economics* 132, 367–433.

Martin, Ian, and Christian Wagner, 2019, What is the expected return on a stock?, *Journal of Finance* 74, 1887–1929.

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Muravyev, Dmitriy, Neil D. Pearson, and Joshua Matthew Pollet, 2022, Why does options market information predict stock returns?, Working paper.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Nagel, Stefan, and Zhengyang Xu, 2022, Asset pricing with fading memory, *Review of Financial Studies* 35, 2190—2245.

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

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam, 2010, O/s: The relative trading activity in options and stock, *Journal of Financial Economics* 96, 1–17.

Ross, Stephen A, 2015, The recovery theorem, *Journal of Finance* 70, 615–648.

Schneider, Paul, and Fabio Trojani, 2019, (almost) model-free recovery, *Journal of Finance* 74, 323–370.

Stambaugh, Robert F, and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

Stilger, Przemysław S, Alexandros Kostakis, and Ser-Huang Poon, 2017, What does risk-neutral skewness tell us about future stock returns?, *Management Science* 63, 1814–1834.

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao, 2010, What does the individual option volatility smirk tell us about future equity returns?, *Journal of Financial and Quantitative Analysis* 45, 641–662.

Internet Appendix of "Butterfly Implied Returns"

5 Results Complementing the Cross-sectional Analysis of Section 2

5.1 BwB generates robust alpha with different rolling windows and skipping periods

To estimate BIR, we use a rolling six-month window in our main analysis. To make the strategy implementable, we skip a month before forming the BwB strategy. In this section, we experiment with different specifications of rolling windows and skipping periods. A longer rolling window produces a more precise estimation of the correlation. However, if the correlation is time-varying, information from earlier periods might be stale. Table 8 summarizes the alphas from the alternative specifications. The six-month window offers a good balance between the two considerations. On the other hand, the length of the skipping periods plays a minor role, and the results are similar across the different choices (0, 1, or 2 months).

5.2 Equal-weighted BwB performs even better

In the main analysis, we focus on the results based on the value-weighted approach, which is in line with the common practice in cross-sectional asset pricing studies. However, since we only use the constituents of the S&P 500 index, an equal-weighted portfolio is also relevant and practical. Table 9 presents the returns of the equal-weighted portfolios and the alpha of the equal-weighted Betting-with-Butterfly (BwB_ew) strategy. Although the magnitudes are similar, BwB_ew outperforms the original BwB in terms of statistical significance, resulting in a higher information ratio.

Table 8: **BwB with Different Rolling Windows and Skipping Periods**

This table examines the robustness of BwB strategy under alternative rolling windows (in BIR estimation) and alternative skipping period (before forming portfolios). In the main analysis, we estimate BIR using a six-month rolling window which ends at the end of month $t-1$, then form portfolios at the end of month t, which corresponds to the (Rolling, Skipping) pair of (6, 1). For each pair of parameters, we report the BwB strategy alphas with respect to different factor models used in Table 3. The data ends in December 2019.

Rolling	Skipping	Alpha			
		FF3	FF5	M4	Q5
3	0	0.30**	0.25**	0.21	0.27**
		(0.12)	(0.12)	(0.14)	(0.13)
	1	0.19	0.13	0.17	0.24*
		(0.12)	(0.12)	(0.14)	(0.13)
	2	0.24**	0.22*	0.17	0.29**
		(0.11)	(0.12)	(0.13)	(0.12)
6	0	0.36***	0.31**	0.29**	0.32**
		(0.12)	(0.12)	(0.14)	(0.13)
6	**1**	**0.39*** **	**0.28** **	**0.32** **	**0.34** **
		(0.12)	**(0.12)**	**(0.14)**	**(0.13)**
	2	0.41***	0.34**	0.30**	0.36**
		(0.13)	(0.13)	(0.15)	(0.14)
12	0	0.30**	0.19	0.24	0.28*
		(0.13)	(0.13)	(0.15)	(0.14)
	1	0.37***	0.26**	0.31**	0.32**
		(0.12)	(0.13)	(0.14)	(0.14)
	2	0.28**	0.20	0.24	0.28*
		(0.13)	(0.13)	(0.15)	(0.14)

5.3 Alpha of BwB primarily comes from the short leg of "vulnerable" stocks

Table 10 reports alphas separately for the long and short legs of the original BwB strategy. The findings indicate that the alpha of BwB mainly originates from the short leg. This means that the approach used in this paper identifies the stocks that are most susceptible to market crashes and have the worst performance.

5.4 The result is robust in the extended sample period (1996-2021)

In the main analysis, we use the options data from 1996 to 2019, which was the data available to us when we finish the first version of this paper. In Tables 11 and 12, we extend the sample with data from 2020 and 2021. The conclusion of the paper is unchanged. In Section 4, we highlight the performance of BwB during the COVID period: using monthly returns, the market realizes an annualized Sharpe ratio of 0.9 in 2020, which is more than double the historical average. However, the Sharpe ratio of BwB is even higher at 1.1 during the same period.

Table 9: **Equal-weighted Portfolios and BwB**

Based on $\text{BIR}_{i,t-1}$, we divide the stocks into 10 groups at the end of month t and form 10 *equal-weighted* portfolios. Panel A reports the time-series average excess returns of each portfolio and the 10–minus–1 returns (H–L) in month $t+1$, as well as alphas adjusted for the factor models used in Table 3. Panel B reports the alphas of an equal-weighted Betting-with-Butterfly (BwB_ew) strategy, which trades the top and bottom three equal-weighted portfolios. The standard errors in parenthesis are adjusted according to Newey and West (1987) with 6 lags. The data ends in December 2019.

Panel A: Excess Return and Alpha of the Equal-weighted Portfolios

	1	2	3	4	5	6	7	8	9	10	H-L
ExRet	0.57	0.57	0.86	0.86	0.90	0.85	0.80	0.84	0.96	1.06	0.49***
	(0.45)	(0.36)	(0.29)	(0.26)	(0.28)	(0.28)	(0.26)	(0.28)	(0.29)	(0.39)	(0.18)
CAPM	-0.41	-0.17	0.23	0.24	0.31	0.26	0.20	0.24	0.29	0.20	0.61***
	(0.16)	(0.16)	(0.17)	(0.16)	(0.17)	(0.15)	(0.16)	(0.18)	(0.13)	(0.19)	(0.17)
FF3	-0.47	-0.25	0.15	0.18	0.24	0.21	0.14	0.16	0.22	0.15	0.61***
	(0.14)	(0.11)	(0.12)	(0.12)	(0.12)	(0.11)	(0.12)	(0.12)	(0.10)	(0.17)	(0.17)
FF5	-0.28	-0.34	-0.00	-0.00	0.04	0.05	-0.03	-0.04	0.10	0.19	0.47***
	(0.14)	(0.13)	(0.11)	(0.12)	(0.10)	(0.11)	(0.12)	(0.11)	(0.11)	(0.19)	(0.17)
M4	0.04	-0.18	0.18	0.11	0.21	0.16	0.12	0.12	0.23	0.58	0.54***
	(0.18)	(0.15)	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.14)	(0.14)	(0.23)	(0.20)
Q5	-0.19	-0.27	0.02	0.10	0.07	0.13	0.02	-0.02	0.15	0.42	0.61***
	(0.13)	(0.12)	(0.12)	(0.14)	(0.12)	(0.13)	(0.12)	(0.13)	(0.10)	(0.19)	(0.20)

Panel B: Alpha of BwB_ew

	FF3	FF5		M4		Q5
Alpha	0.36***	0.29***	Alpha	0.30**	Alpha	0.33***
	(0.10)	(0.10)		(0.12)		(0.11)
MktRf	-0.10***	-0.06**	MktRf	-0.05*	R_MKT	-0.09***
	(0.02)	(0.03)		(0.03)		(0.03)
SMB	-0.04	-0.03	SMB	-0.05	R_ME	-0.05
	(0.03)	(0.04)		(0.04)		(0.03)
HML	-0.00	-0.10**	MGMT	0.06	R_IA	0.08
	(0.03)	(0.04)		(0.04)		(0.05)
CMA		0.18***	PERF	0.04*	R_ROE	-0.02
		(0.06)		(0.03)		(0.05)
RMW		0.06			R_EG	0.03
		(0.05)				(0.06)

Table 10: **Regressing the Long and Short Legs of BwB on the Common Risk Factors**

This table repeats the analysis in Table 3 for the long and short legs of BwB. The short leg comprise of stocks with lowest BIR and the excess return is $\frac{r^{(3)}_{t+1}+r^{(2)}_{t+1}+r^{(1)}_{t+1}}{3} - r_{f,t}$. The excess return of the long leg is $\frac{r^{(10)}_{t+1}+r^{(9)}_{t+1}+r^{(8)}_{t+1}}{3} - r_{f,t}$. The data ends in December 2019.

Panel A: Dependent variable is $\frac{r^{(3)}_{t+1}+r^{(2)}_{t+1}+r^{(1)}_{t+1}}{3} - r_{f,t}$ (the short leg)						
	FF3	FF5		M4		Q5
Alpha	-0.27***	-0.20**	Alpha	-0.14	Alpha	-0.25***
	(0.08)	(0.09)		(0.10)		(0.10)
MktRf	1.18***	1.14***	MktRf	1.12***	R_MKT	1.19***
	(0.02)	(0.02)		(0.03)		(0.02)
SMB	-0.11***	-0.11***	SMB	-0.11***	R_ME	-0.12***
	(0.03)	(0.03)		(0.03)		(0.03)
HML	0.18***	0.29***	MGMT	0.08**	R_IA	0.14***
	(0.03)	(0.04)		(0.03)		(0.04)
CMA		-0.21***	PERF	-0.14***	R_ROE	-0.04
		(0.05)		(0.02)		(0.04)
RMW		-0.05			R_EG	0.06
		(0.04)				(0.05)

Panel B: Dependent variable is $\frac{r^{(10)}_{t+1}+r^{(9)}_{t+1}+r^{(8)}_{t+1}}{3} - r_{f,t}$ (the long leg)						
	FF3	FF5		M4		Q5
Alpha	0.11	0.08	Alpha	0.18**	Alpha	0.09
	(0.07)	(0.07)		(0.08)		(0.08)
MktRf	1.03***	1.04***	MktRf	1.04***	R_MKT	1.05***
	(0.02)	(0.02)		(0.02)		(0.02)
SMB	-0.14***	-0.13***	SMB	-0.15***	R_ME	-0.13***
	(0.02)	(0.03)		(0.03)		(0.02)
HML	0.08***	0.04	MGMT	0.06**	R_IA	0.16***
	(0.02)	(0.03)		(0.03)		(0.04)
CMA		0.07	PERF	-0.00	R_ROE	0.02
		(0.05)		(0.02)		(0.03)
RMW		0.03			R_EG	0.03
		(0.03)				(0.04)

Table 11: **Fama-MacBeth Regression with an Extended Sample (1996-2021) and a Recent Sample (2002-2021)**

This table reports the time series average coefficients and standard errors of the Fama-MacBeth regression of return $r_{i,t+1}$ on a list of variables as defined in Table 2, with the data extended to December 2021, as well as the recent data from the last 20 years (January 2002 to December 2021). The standard errors in parenthesis are adjusted according to Newey and West (1987), with six lags.

	The dependent variable is $r_{i,t+1}$ in percentage points									
	Extended Sample (1996-2021)					Recent Sample (2002-2021)				
$\mathrm{BIR}_{i,t-1}$	2.13***	1.65***	1.19*	1.62***	1.35**	2.47***	2.09***	1.62**	2.09***	1.73***
	(0.69)	(0.53)	(0.61)	(0.56)	(0.55)	(0.79)	(0.59)	(0.67)	(0.63)	(0.62)
β, ME, B/M	N	Y	Y	Y	Y	N	Y	Y	Y	Y
INV			-0.07	-0.15	-0.18*			-0.06	-0.12	-0.13
			(0.10)	(0.10)	(0.09)			(0.11)	(0.11)	(0.10)
OP			0.02					0.02		
			(0.02)					(0.02)		
OP^{*}				0.00	0.00				0.02	0.02
				(0.02)	(0.02)				(0.02)	(0.02)
MOM			-7.22	0.68	9.54			-21.15	-14.89	-4.82
			(38.07)	(38.92)	(35.50)			(45.44)	(46.89)	(42.15)
REV			-1.14*	-1.29*	-1.43**			-1.02	-1.05	-1.29
			(0.65)	(0.68)	(0.67)			(0.76)	(0.80)	(0.79)
β^{VIX}					-8.96					-15.01
					(10.84)					(12.97)
RNV					-8.05					-11.54
					(10.44)					(11.65)
RNS					0.30***					0.20**
					(0.09)					(0.09)

Table 12: **Alpha of BwB with an Extended Sample (1996-2021) and a Recent Sample (2002-2021)**

This table repeats the analysis of Table 3 with the data extended to December 2021 as well as the recent data from the last 20 years (January 2002 to December 2021).. The result in Column "M4" is unchanged with the extended sample as we do not have the updated factor returns.

	Extended		Recent			Extended	Recent		Extended	Recent
	FF3	FF5	FF3	FF5		M4	M4		Q5	Q5
Alpha	0.34***	0.25**	0.34***	0.25*	Alpha	0.32**	0.32**	Alpha	0.31**	0.33**
	(0.12)	(0.12)	(0.13)	(0.13)		(0.14)	(0.15)		(0.13)	(0.13)
MktRf	-0.14***	-0.10***	-0.17***	-0.14***	MktRf	-0.08**	-0.10**	R_MKT	-0.12***	-0.16***
	(0.03)	(0.03)	(0.03)	(0.03)		(0.04)	(0.04)		(0.03)	(0.04)
SMB	-0.03	-0.02	-0.06	-0.06	SMB	-0.04	-0.07	R_ME	-0.03	-0.13**
	(0.04)	(0.04)	(0.05)	(0.05)		(0.04)	(0.06)		(0.04)	(0.06)
HML	-0.08**	-0.18***	-0.06	-0.16***	MGMT	-0.02	0.17**	R_IA	0.05	0.16**
	(0.03)	(0.05)	(0.05)	(0.05)		(0.05)	(0.07)		(0.06)	(0.07)
CMA		0.23***		0.35***	PERF	0.14***	0.16***	R_ROE	0.04	0.00
		(0.07)		(0.08)		(0.03)	(0.04)		(0.05)	(0.06)
RMW		0.05		0.06				R_EG	-0.01	0.00
		(0.05)		(0.06)					(0.07)	(0.08)
#Obs	305	305	240	240		245	180		305	240

6 Results Complementing the Time Series Analysis of Section 3

6.1 BIRM$_t$ and the realized market return

In equilibrium, asset prices adjust such that BIRM is positively correlated with returns in the crash state and negatively correlated with returns in non-crash states. Therefore, when regressing realized returns on BIRM in a long sample with crashes, we obtain a positive coefficient equal to the physical probability of a market crash (p^M) only if we control for the expectation of returns in a non-crash state. We use the following specification

$$\frac{1}{n}\left(R_{M,t\to t+n} - R_{f,t}\right) = \underbrace{\beta_0^R}_{=0} + \underbrace{\beta_B^R}_{>0} \times \mathrm{BIRM}_t + \underbrace{\beta_{NB}^R}_{>0} \times \left(dp_t - \overline{dp}\right) + \epsilon_{t+n}^R \qquad (14)$$

where $dp_t - \overline{dp}$ captures the return if a crash does not occur, in which case dp_t mean-reverts to its mean $\overline{dp}$. In this regression, β_B^R directly measures the (ex post) probability of a crash, p^M, which should be positive. To operationalize this regression, it is necessary to introduce some forward-looking bias and set $\overline{dp}$ to its in-sample mean. We use $\overline{dp} = -4.2$, which is the natural logarithm of 1.5%. Table 13 reports the results over the next month and the next year, which are similar in all aspects except for R^2. This is not surprising, as the result using the monthly return reflects transitory shocks that are not captured by the two variables.

As expected, the coefficient β_B^R is positive. The combination of β_B^{EP} and β_B^R allows us to quantitatively evaluate the importance of a market crash. With the results from annual data in the last columns of Table 6 and 13, we know that the probability of a crash is $p^M = \beta_B^R = 0.41$, which corresponds to two crashes every five years. When a crash occurs, the marginal utility is multiplied by $B^{-\gamma} = 1 - \frac{\beta_B^{EP}}{\beta_B^R} = 1.17$, or increases by 17%. If we further take a stance on the risk aversion by setting $\gamma = 4$, we can infer that the consumption of the representative investor falls by $1 - B = 4\%$ in a market crash.

Interpreting the coefficients β_{NB}^R is more subtle, as the variable $dp_t - \overline{dp}$ is a mere proxy

Table 13: **Time Series Predictability of BIRM$_t$**

This table reports the results of Regression (14). The realized excess market return $\frac{12}{n}\left(R_{M,t\to t+n}-R_{f,t}\right)$ ends in December 2019. On the right-hand side of the regression, BIRM is for the period March 1996 to December 2019. dp_t is the natural logarithm of the annual dividend to price ratio for the S&P 500. We set $\overline{dp}=-4.2$, which is the natural logarithm of 1.5%. In the parentheses are the Hansen and Hodrick (1980) standard errors with a bandwidth of $12\times n$.

Dependent	$\frac{1}{n}\left(R_{M,t\to t+n}-R_{f,t}\right)$					
	$n=\frac{1}{12}$			$n=1$		
Intercept	0.10	0.02	0.10	0.07	-0.00	0.06*
	(0.06)	(0.04)	(0.06)	(0.06)	(0.04)	(0.04)
BIRM$_t$	0.08		0.49**	-0.02		0.41***
	(0.24)		(0.20)			(0.09)
$dp_t-\overline{dp}$		0.33	0.59***		0.41***	0.63***
		(0.23)	(0.21)		(0.14)	(0.09)
R^2	0.00	0.01	0.03	0.00	0.24	0.35
N	282	282	282	271	271	271

for the future return. Over short horizons, if the log dividend-to-price ratio dp_t mean-reverts to $\overline{dp}$, the return can be approximated by $\phi^{dp}\left(dp_t-\overline{dp}\right)$, where ϕ^{dp} is the speed of mean-reversion. To make quantitative inferences, one needs to take a stand on the magnitude of ϕ^{dp}, which is not easy. For example, Lettau and van Nieuwerburgh (2007) challenge the conventional wisdom that ϕ^{dp} is low. In their Table 2, the AR(1) coefficient is 0.61 with two breaks, implying that $\phi^{dp}=0.39$ annually. We observe an even more speedy reversion after the Great Recession, which leads to strong return predictability. More recently, the stock market exhibits a V-shaped rebound from the low in March 2020, reflecting dissipation of the initial fear over COVID-19. The debate about the persistence of the valuation ratio is likely to continue for years to come and is beyond the scope of this paper.

Finally, we highlight another difference between β_B^R and the coefficient β_B^{EP} in the main text. The relationship in Regression (12) holds in every period. Therefore, β_B^{EP} is always the difference between the physical and risk neutral probability of a crash, which is a small negative number. In contrast, the relationship in Regression (14) only holds in a long sample in which the probability of crash converges to the mean. In a shorter sample, β_B^R is heavily influenced by the crash frequency. Figure 7 examines how the two coefficientsβ_B^{EP} and β_B^R evolve over time with a rolling 60-month window. As expected, the resulting β_B^{EP} (solid line) oscillates within a tight and mostly negative range, whereas β_B^R (dashed line) is far more volatile, shooting up if when a market crash occurs in the rolling window.

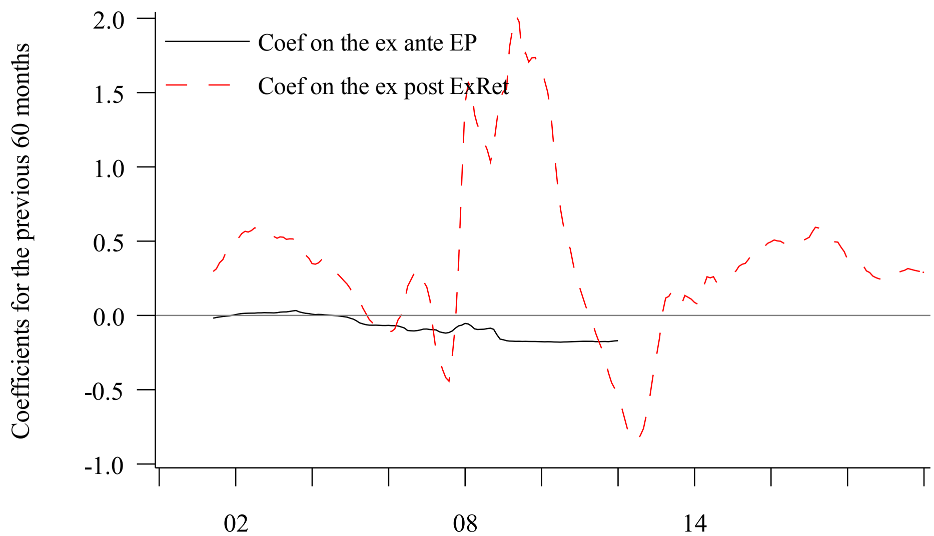

Figure 7: **Coefficients of the rolling estimation of the regressions** (12) **and** (14)**.** We plot the time series of β_B^{EP} (solid line) in Regression (12) and β_B^R (dashed line) in Regression (14), estimated with rolling data over 60 months.

6.2 BIRM$_t$ significantly predicts future 12-month return with Amihud and Hurvich (2004) correction

Regression (14) is subject to the Stambaugh (1999) bias. The bias arises due to the measurement error in the persistence of the predictors BIRM$_t$ and $dp_t - \overline{dp}$, whose shocks in the future are correlated with the shock of $R_{M,t\to t+n} - R_{f,t}$. We follow the procedure proposed in Amihud and Hurvich (2004) to reduce the bias. First, we estimate a univariate AR(1) model for $x_t = \text{BIRM}_t$ or $dp_t - \overline{dp}$:

$$x_{t+1} = \theta_x + \rho_x x_t + \nu_{t+1}^x. \tag{15}$$

The estimated ρ_x is biased towards zero in a small sample. Therefore, we obtain the (doubly) bias-adjusted coefficient $\hat{\rho}_x^c$ as

$$\hat{\rho}_x^c = \hat{\rho}_x + \frac{1+3\hat{\rho}_x}{N} + \frac{3\left(1+3\hat{\rho}_x\right)}{N^2} \tag{16}$$

with $N = 271$. The adjusted error terms, which will be added to the predictive regression, are calculated as $\nu_{t+1}^{x,c} = x_{t+1} - \left(\hat{\theta}_x + \hat{\rho}_x^c x_t\right)$.

Finally, we run the augmented predictive regression

$$\frac{1}{n}\left(R_{M,t\to t+n} - R_{f,t}\right) = \beta_0^R + \beta_{BIRM}^R \times \text{BIRM}_t + \beta_{dp}^R \times \left(dp_t - \overline{dp}\right) + \phi_{BIRM}\nu_{t+n}^{BIRM,c} + \phi_{dp}\nu_{t+n}^{dp,c} + \epsilon_{t+n}^R, \tag{17}$$

with the adjusted standard error for the predictive coefficient as

$$s.e.\left(\widehat{\beta_x^R}\right)^c = \sqrt{s.e.\left(\widehat{\beta_x^R}\right)^2 + \left[\widehat{\phi_x}\left(1 + \frac{3}{N} + \frac{9}{N^2}\right)s.e.\left(\hat{\rho}_x\right)\right]^2} \tag{18}$$

for $x = \text{BIRM}$ or dp. Table 14 compare the result with and without the bias reduction. We focus on 12-month return by setting $n = 1$. The result clearly shows a bias in the estimated coefficient. For example, the coefficient of BIRM_t is adjusted from 0.41 down to 0.15. However, our conclusion remains that BIRM_t significantly predicts future return, conditional on dividend yield.

6.3 The price of risk of ΔBIRM_t is robust to the specification

Maio and Santa-Clara (2012) use a different GMM specification which allows for estimation

Table 14: **Predictive Regression with Amihud and Hurvich (2004) Correction**

Panel A: AR(1) Model in Equations (15) and (16)

$x_t =$	$\hat{\theta}_x$	$\hat{\rho}_x$	$s.e.\left(\hat{\rho}_x\right)$	$\hat{\rho}_x^c$
BIRM_t	-0.1306	0.5465	0.048	0.5563
$dp_t - \overline{dp}$	0.0606	0.6477	0.047	0.6586

Panel B: Predictive Regression in Equation (17)

	Amihud and Hurvich (2004) corrected				Table 6	
$x_t =$	$\widehat{\beta_x^R}$	$s.e.\left(\widehat{\beta_x^R}\right)$	$\widehat{\phi_x}$	$s.e.\left(\widehat{\beta_x^R}\right)^c$	$\widehat{\beta_x^R}$	$s.e.\left(\widehat{\beta_x^R}\right)$
BIRM_t	**0.15*** **	0.040	0.1182	*(0.04)*	**0.41*** **	*(0.09)*
$dp_t - \overline{dp}$	**0.48*** **	0.032	-0.7024	*(0.05)*	**0.63*** **	*(0.09)*

error of the factor means

$$g_T(b) = \frac{1}{T}\sum_{t=1}^{T} u_t(b) \equiv \frac{1}{T}\sum_{t=1}^{T} \begin{cases} R_t^e - \gamma^{\mathrm{M}} R_t^e \left(R_t^{\mathrm{M}} - \mu^{\mathrm{M}}\right) - \gamma^{\mathrm{B}} R_t^e \left(\Delta\mathrm{BIRM}_t - \mu^{\mathrm{B}}\right) \\ R_t^{\mathrm{M}} - \mu^{\mathrm{M}} \\ \Delta\mathrm{BIRM}_t - \mu^{\mathrm{B}} \end{cases} . \tag{19}$$

In this specification, the Jacobian matrix is

$$d_T = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} -R_t^e\left(R_t^{\mathrm{M}} - \mu^{\mathrm{M}}\right), & -R_t^e\left(\Delta\mathrm{BIRM}_t - \mu^{\mathrm{B}}\right), & \gamma^{\mathrm{M}} R_t^e, & \gamma^{\mathrm{B}} R_t^e \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

As in the main text, R_t^e is a row vector of the excess returns of test assets, and R_t^{M} is the excess return of market. $\Delta\mathrm{BIRM}_t = \mathrm{BIRM}_t - \mathrm{BIRM}_{t-1}$ is the news about the annualized expectation of return, whose volatility is comparable to that of the realized returns. The four parameters $b = \left\{\gamma^{\mathrm{M}}, \gamma^{\mathrm{B}}, \mu^{\mathrm{M}}, \mu^{\mathrm{B}}\right\}$ are to be estimated.

We differ from Maio and Santa-Clara (2012) in two ways. First, to address the concerns raised by Lewellen, Nagel, and Shanken (2010), we add 10 beta-sorted portfolios and 10 iVol-sorted portfolios into the test assets. Second, we use the following weighting matrix in the second stage

$$W = \begin{bmatrix} \hat{S}_{(n)}^{-1} & \mathbf{0} \\ \mathbf{0} & 100^2 \times I_{(2)} \end{bmatrix}$$

where $\hat{S}_{(n)}^{-1}$ inverts the top-left $n \times n$ sub-matrix of $\hat{S}$ ($n = 45$ in this case), which is the estimate of spectral density matrix from the first stage. The difference here is that we multiply the identity matrix $I_{(2)}$ by 100^2 to give the two parts comparable weights, since $\hat{S}$ measures the second moments of monthly returns.

Table 15: **Result of an Alternative GMM Specification**

We use the moment condition in Equation (19). The standard errors are $\text{var}\left(\hat{b}\right) = \frac{1}{T}\left(d_T'\hat{S}^{-1}d_T\right)^{-1}$ with Newey and West (1987) adjustment $\hat{S} = \sum_{j=-L}^{L}\left(1 - \frac{|j|}{L+1}\right)\frac{1}{T}\sum_{t=1}^{T}\left[u_t\left(\hat{b}\right)u_{t-j}\left(\hat{b}\right)'\right]$ where $u_t = 0$ if $t < 1$ or $t > T$. The number of lags is $L = 6$ in both stages.

	First Stage			
	γ^{M}	γ^{B}	$100 \times \mu^{\text{M}}$	$100 \times \mu^{\text{B}}$
$\hat{b}_1$	1.95	5.82	0.59	-0.29
$s.e.\left(\hat{b}_1\right)$	(1.39)	(2.84)	(0.24)	(0.25)
	Second Stage			
	γ^{M}	γ^{B}	$100 \times \mu^{\text{M}}$	$100 \times \mu^{\text{B}}$
$\hat{b}_2$	3.00	8.05	0.64	-0.13
$s.e.\left(\hat{b}_2\right)$	(1.13)	(2.62)	(0.22)	(0.23)

The GMM procedure is equivalent to a two-pass estimation, where we first use the time series of return of each portfolio to estimate the betas on the two risk, $\hat{\beta}^M$ and $\hat{\beta}^B$, then run a cross-section regression

$$\mu = \text{constant} + \lambda_M \hat{\beta}^M + \lambda_B \hat{\beta}^B + \epsilon \tag{20}$$

to obtain the risk premium λ_M and λ_B. Here $\mu = \frac{1}{T}\sum_{t=1}^{T} R_t^e$ is the average excess return of the test assets. In Table 16, we report results from OLS and GLS (μ on $V^{-1/2}\left[\iota\ \hat{\beta}^M\ \hat{\beta}^B\right]$, where $V^{-1/2} = Q\Lambda^{-1/2}Q'$ with Q and Λ from an eigen-decomposition of $var\left(R_t^e\right)$), with and without the constant term. The estimated monthly risk premium λ_B (around 1%) approximately equals γ^{B} (Table 15, e.g., 5.82 from the first stage) times the monthly variance of $\Delta\mathrm{BIRM}_t$ (approximately 0.002).

Table 16: **Result of the Two-pass Regression**
This table reports the slopes and standard errors from the cross-section regression of Equation (20).

	Constant	$\lambda_M \times 100$	$\lambda_B \times 100$	R^2
OLS		0.61	1.28	
		(0.04)	(0.79)	
	0.91	-0.18	1.00	14%
	(0.09)	(0.08)	(0.44)	
GLS		0.71	1.26	
		(0.44)	(1.09)	
	1.19	-0.48	0.56	32%
	(0.24)	(0.43)	(0.89)	

References

Amihud, Yakov, and Clifford M Hurvich, 2004, Predictive regressions: A reduced-bias estimation method, *Journal of Financial and Quantitative Analysis* 39, 813–841.

Hansen, Lars Peter, and Robert J Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

Lettau, Martin, and Stijn van Nieuwerburgh, 2007, Reconciling the return predictability evidence, *Review of Financial Studies* 21, 1607–1652.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

Maio, Paulo, and Pedro Santa-Clara, 2012, Multifactor models and their consistency with the icapm, *Journal of Financial Economics* 106, 586–613.

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

Stambaugh, Robert F, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.