L'Institut bénéficie du soutien financier de l'Autorité des marchés financiers ainsi que du ministère des Finances du Québec

Document de recherche

DR 20-08

Common Factors in Equity Option Returns

Publié Septembre 2020

Ce document de recherche a été rédigée par :

Alex Horenstein, University of Miami
Aurelio Vasquez, ITAM School of Business
Xiao Xiao, Erasmus University Rotterdam

Common Factors in Equity Option Returns

Alex Horenstein[1], Aurelio Vasquez*[2], and Xiao Xiao[3,4]

[1]Department of Economics, University of Miami

[2]ITAM, School of Business

[3]Erasmus School of Economics, Erasmus University Rotterdam

[4]Erasmus Research Institute of Management - ERIM

September 23, 2020

Abstract

This paper studies the factor structure of the cross-section of delta-hedged equity option returns. We find that a four-factor model explains the cross-section and time-series of equity option returns. Out of the four factors, three are characteristic-based factors from the long-short option portfolios based on firm size, idiosyncratic volatility, and the difference between implied and historical volatilities. The fourth factor is the market volatility risk factor proxied by the delta-hedged option return of the the S&P 500 index. Traditional stock return factors cannot price the cross-section of equity option returns.

JEL Classification: C14, G13, G17

Keywords: Cross Section of Option Returns, Latent Factors, Rank Estimation

*Corresponding to: Aurelio Vasquez, ITAM, Rio Hondo 1, Alvaro Obregón, Mexico City, Mexico. Tel: (52) 55 5628 4000 x.6518; Email: aurelio.vasquez@itam.mx.

1 Introduction

The identification of factors that drive the comovements of asset returns is a central question in empirical asset pricing. Existing papers on multi-factor asset pricing models mainly focus on common factors in stock returns.[1] However the factor structure of the cross-section and time-series of equity option returns is less understood. Options are often viewed as merely leveraged positions in the underlying stocks. However Bakshi and Kapadia (2003) show that delta-hedged option returns contain risk premiums beyond the equity premium such as the variance risk premium. To study the factor structure of option returns provides information on the factors that drive the cross-section of variance risk premiums. This is the main goal of our study.

Factor identification for the cross-section of option returns cannot be directly done at the firm level. Even by working with the most liquid options—ATM options with one-month to maturity—firms come in and out of the sample and a balanced panel of option returns is not possible. To circumvent this problem we work with option portfolios constructed from firm characteristics that predict option returns. We include eleven characteristics that predict option returns: size, reversal, momentum, profitability, cash holding and analyst forecast dispersion in Cao et al. (2017), credit rating in Vasquez and Xiao (2018), the deviation of realized volatility from implied volatility in Goyal and Saretto (2009), idiosyncratic volatility in Cao and Han (2013), and the volatility term structure in Vasquez (2017). Empirically, we construct 105 option portfolios sorted by the eleven characteristics using monthly portfolios of delta-hedged options from January 1996 to December 2015. The eleven characteristic-based factors are constructed based on long-short strategies of decile or quintile returns. Additionally, we include two option market factors: the delta-hedged option return of the S&P 500 index and the value-weighted delta-hedged option return of the S&P 500 index components. Overall we include 13 candidate factors in total.

We find that a four-factor model explains the cross-sectional and time-series variation of delta-hedged option returns. These factors include the long-short option portfolios formed on size, idiosyncratic volatility, and volatility deviation (the difference between historical and

[1] Recent studies include Barillas and Shanken (2018), Ahn et al. (2018), Hou et al. (2018), Feng et al. (2017) among others.

implied volatilities), and the delta-hedged option return of the S&P 500 index.[2] The first three factors capture the cross-sectional comovements in option returns while the market volatility risk factor mainly explains their time-series variation.

Using the identification protocol by Pukthuanthong et al. (2018), we determine that there is one strong factor and possibly up to five weak factors that can explain the cross-section of option returns. We first employ a latent variable analysis of the covariance matrix of option returns to understand its factor structure. Latent variables are not directly observed but are econometrically inferred from observed variables. We estimate the number of latent factors using six identification methods and then estimate the common latent factors using the asymptotic principal component analysis (PCA) suggested in Connor and Korajczyk (1986). We conclude that up to six latent factors are needed to explain the covariance matrix of option returns. We report a 96% correlation between the observed option returns and the six-latent factor model's predicted returns.

To provide an economic interpretation of the information captured by the six latent factors, we explore which of the 13 (observed) factors—11 long-short option portfolios and two market option factors—better explain the covariation of option returns. First, we use the rank-estimation method suggested by Ahn et al. (2018) on the 13 candidate factors and find that out of the 462 different combinations of 6 factors, only one set can generate a full-rank beta matrix. The unique set contains the long-short factors constructed from option returns on size, cash holding, analyst dispersion, idiosyncratic volatility, volatility deviation, and credit rating. Canonical correlation analysis between these six factors and the six latent factors estimated from PCA confirm the results. The six characteristic-based factor candidates contain most of the relevant information on the cross-section of delta-hedged option returns.

Ultimately we find that only three out of the six candidate factors suffice to explain the cross-section of option returns: size, idiosyncratic volatility, and volatility deviation. The correlation between the average option portfolio returns and the three factor's predicted returns is 95%. The three remaining factors add negligible explanatory power to the risk-

[2] Size and idiosyncratic volatility factors are constructed with delta-hedged option returns and are not to be confused with similar factors constructed from stock returns.

expected return relationship. To explain the time-series (but not the cross-section) of option returns a fourth factor is added to the model: the delta-hedged option return of the S&P 500 index. This approach is analogous to the one in the stock market where the stock returns' market portfolio is used as a factor even if it has little to no variability in factor loadings and can only explain the time-series dimension of stock returns (Fama and French (1992) and Ahn and Horenstein (2018)). Using the four-factor model on characteristic based option portfolios as test assets, we find that the adjusted R^2 in Fama-MacBeth regressions ranges from 84% to 90% for different setups. To address the critique by Lewellen et al. (2010), we also test the proposed model on a set of delta-hedge option portfolios constructed based on industrial classification and find that the correlation between observed returns and model returns is 84%.

This paper is related to the literature that explores the factor structure of options. Most of the research explores the factor structure of the S&P 500 index options: Jones (2006) study index option prices and Büchner and Kelly (2019) study index option returns. On the cross-section of options, Christoffersen et al. (2017a) study the factor structure of the option prices, not option returns. More related to our paper is Brooks et al. (2018) and Karakaya (2014) who study the cross-section of option returns. Brooks et al. (2018) use the LASSSO estimator to document that 15 characteristics out of 99 provide independent information on the cross-section of option returns. One main difference is that we relate the candidate factors to the covariance matrix of option returns to ensure that these factors capture different information. Karakaya (2014) studies the factors structure of 30 option portfolios constructed based on maturity and moneyness of option returns of individual firms. In our paper we study the factor structure inside one of these 30 portfolios: the one of ATM options with 1-month to maturity. The literature documents cross-sectional differences in option returns based on at least 11 characteristics within this portfolio. Therefore the factors captured by our study and Karakaya (2014) are different in spirit. Like in the stock return literature, multiple predictive characteristics are proposed but not necessarily all of them capture different information. We are the first to address this issue in the option market and find that four factors summarize all the information about the cross-section and time-series among the 13 candidate factors proposed so far in the option return literature.

In Section 2 we present our main analytical results motivating the factor structure in delta-hedged option returns. Section 3 explains the data used for our empirical analysis. In Section 4 we perform our quantitative studies: in Section 4.1 we analyze the factor structure in option returns using latent variable techniques, in Section 4.2 we study which option returns predictors better explain the factor structure in option returns, and in Section 4.3 we conduct a robustness test for different definitions of delta-hedged option returns. We conclude in Section 5.

2 Theoretical motivation: Delta-hedged equity option gains in a multi-factor framework

In this section, we derive expected delta-hedged equity option gains in a multi-factor framework in which stock returns and variance are driven by multiple factors. The results show that delta-hedged option gains have no sensitivity to standard factors in stock returns, just to those related to volatility risk.

We denote the stock price and the variance of stock return for firm i as S_t^i and V_t^i. The variance of stock i is driven by a factor structure: $V_{f,t}^j$, $j = 1, ..., n$, where the factors are independent of each other. The stock price evolves according to the process:

$$\frac{dS_t^i}{S_t^i} = \mu_t^i(S_t^i, V_t^i)dt + \sqrt{V_t^i}dW_{1t}^i,$$
$$V_t^i = \sum_{j=1}^{n} \beta^j V_{f,t}^j + Z_t^i,$$
$$dV_{f,t}^j = \theta^j(V_{f,t}^j)dt + \eta^j(V_{f,t}^j)dW_{2t}^{i,j}.$$

To simplify the analysis, we assume that the correlations among the standard Brownian motions W_{1t}^i and $W_{2t}^{i,j}$ are all 0. Relaxing this assumption and allowing leverage effect does not change the main result of the model. Note that if the stock variance is only driven by the market index variance, the factor structure of the stock variance is based on CAPM where stock returns have a market component and an idiosyncratic component: $R_t^i = \hat{\alpha}^i + \hat{\beta}^i R_t^m + \hat{\epsilon}_t^i$, where $\hat{\epsilon}_t^i$ is uncorrelated with R_t^m. $\beta_i = (\hat{\beta}^i)^2$ is the sensitivity of individual variance with respect to the variance of the market index and $V_{f,t}^1$ corresponds to the variance of the market index. Similarly, if the stock return is driven by the Fama-French three factor model, $V_{f,t}^1$, $V_{f,t}^2$ and $V_{f,t}^3$ corresponds to the variance of the market index, the variance of the SMB factor, and the variance of the HML factor.

By Ito's lemma, we can write the call option price as

$$C_{t+\tau}^i = C_t^i + \int_t^{t+\tau} \Delta_u^i dS_u^i + \int_t^{t+\tau} \frac{\partial C^i}{\partial V^i} dV_u^i + \int_t^{t+\tau} b_u^i du \tag{1}$$

where $\Delta_u^i = \frac{\partial C_u^i}{\partial S_u^i}$ is the delta of the call option and

$$b_u^i = \frac{\partial C^i}{\partial u} + \frac{1}{2} V^i (S^i)^2 \frac{\partial^2 C^i}{\partial (S^i)^2} + \frac{1}{2} \sum_{j=1}^n (\beta^j)^2 (\eta^j)^2 \frac{\partial^2 C^i}{\partial (V^i)^2}.$$

The no-arbitrage assumption implies that the valuation equation that determines the call option price is:

$$\begin{aligned} &\frac{1}{2} V^i (S^i)^2 \frac{\partial^2 C^i}{\partial (S^i)^2} + \frac{1}{2} \sum_{j=1}^n (\beta^j)^2 (\eta^j)^2 \frac{\partial^2 C^i}{\partial (V^i)^2} + r S_i \frac{\partial C^i}{\partial S_i} + \\ &[\sum_{j=1}^n \beta^j (\theta^j (V_{f,t}^j) - \lambda^j (V_{f,t}^j))] \frac{\partial C^i}{\partial V^i} + \frac{\partial C^i}{\partial t} - r C^i = 0, \end{aligned} \tag{2}$$

where $\lambda^j(V_{f,t}^j) = -cov_t(\frac{dm_t}{m_t}, dV_{f,t}^j)$ is the variance risk premium for factor j given a pricing kernel m_t.

Combining Equation (1) and (2), we have:

$$\begin{aligned} C_{t+\tau}^i - C_t^i = &\int_t^{t+\tau} \Delta_u^i dS_u^i + \int_t^{t+\tau} r(C^i - S_i \frac{\partial C^i}{\partial S_i}) du + \\ \int_t^{t+\tau} [\sum_{j=1}^n \beta^j \lambda^j (V_{f,t}^j)] \frac{\partial C^i}{\partial V^i} du &+ \int_t^{t+\tau} [\sum_{j=1}^n \beta^j \theta^j (V_{f,t}^j) \frac{\partial C^i}{\partial V^i} dW_2^{i,j}]. \end{aligned} \tag{3}$$

With a delta-hedged portfolio, we buy the call option and dynamically delta-hedge the option position with time-varying delta Δ_u^i. The delta-hedged gain $\Pi_{t,t+\tau}^i$ is the gain or loss on a delta-hedged option portfolio in excess of the risk-free rate earned by this portfolio and is defined as

$$\Pi^i_{t,t+\tau} = C^i_{t+\tau} - C^i_t - \int_t^{t+\tau} \Delta^i_u dS^i_u - \int_t^{t+\tau} r(C^i - S_i \frac{\partial C^i}{\partial S_i})du.$$

From the definition of delta-hedged gain and Equation (3), we obtain the expectation of the delta-hedged gain for stock option i:

$$\begin{aligned} E[\Pi^i_{t,t+\tau}] &= E[\int_t^{t+\tau} [\sum_{j=1}^{n} \beta^j \lambda^j (V^j_{f,t})] \frac{\partial C^i}{\partial V^i} du + \int_t^{t+\tau} [\sum_{j=1}^{n} \beta^j \theta^j (V^j_{f,t}) \frac{\partial C^i}{\partial V^i} dW_2^{i,j}]] \\ &= \sum_{j=1}^{n} \beta^j E[\int_t^{t+\tau} \lambda^j (V^j_{f,t})] \frac{\partial C^i}{\partial V^i} du] \end{aligned} \tag{4}$$

The result shows that if variance risks of the factors $V^1_{f,t}$, $V^2_{f,t}$, ..., $V^n_{f,t}$ are priced, the expected delta-hedged gain of a equity option is driven by the exposures to each variance risk β_j, the price of variance risk λ^j, and the vega of the option $\frac{\partial C^i}{\partial V^i}$.

Bakshi and Kapadia (2003) show that the expected delta-hedged gain of the index option is closely related to the price of volatility risk. If we consider the following price process with stochastic volatility for the market index S^m

$$\begin{aligned} \frac{dS^m_t}{S^i_t} &= \mu^m_t(S^m_t, V^m_t)dt + \sqrt{V^m_t} dW^m_{1t}, \\ dV^m_t &= \theta^m(V^m_t)dt + \eta^m(V^m_t)dW^m_{2t}, \end{aligned}$$

and a call option C^m_t written on the market index, Bakshi and Kapadia (2003) derive the expected delta-hedged gain for the index option as

$$E[\Pi^m_{t,t+\tau}] = E[\int_t^{t+\tau} \lambda^m(V^m_t) \frac{\partial C^m}{\partial V^m} du].$$

It follows that if we use options with similar maturity and moneyness for the index and the stocks, their vega sensitivities will also be of similar magnitude. Hence, we can use the delta-hedged gain of index options to replicate the price of volatility risk of the index return. However, there are no traded options on stock return factors other than the market factor. We follow the literature on stock return factors and work with the long-short characteristic-based option portfolios as potential candidate of factors. The details of the characteristics and factors are provided in Section 3.

3 Data and variables description

3.1 Data and sample coverage

We obtain option data on individual stocks from the OptionMetrics Ivy DB database. Sample period is from January 1996 to December 2015. Implied volatility and Greeks are calculated by OptionMetrics using the binomial tree from Cox et al. (1979). We obtain stock returns, prices and credit ratings from the Center for Research on Security Prices (CRSP); balance sheet data from Compustat and analyst coverage and forecast data from I/B/E/S.

We apply several filters to select the options in our sample. First, to avoid illiquid options, we exclude options if the trading volume is zero, the bid quote is zero, the bid quote is smaller than the ask quote, or the average of the bid and ask price is lower than 0.125 dollars. Second, to remove the effect of early exercise premium in American options, we discard options whose underlying stock pays a dividend during the remaining life of the option. Therefore, options in our sample are very close to European style options. Third, we exclude all options that violate no-arbitrage restrictions. Fourth, we only keep options with moneyness between 0.8 and 1.2. At the end of each month and for each stock with options, we select a call option that is the closest to being at-the-money with the shortest maturity among those options with more than one month to maturity. We drop options whose maturity is different from the majority of options. Our final sample contains 327,016 option-month observations for calls. The time to maturity ranges from 47 to 50 days.

3.2 Construction of the delta-hedged option returns

Since an option is a derivative written on a stock, option returns are highly sensitive to stock returns. In this paper, following the literature, we study the gain of delta-hedged options, such that the portfolio gain is not sensitive to the movement of the underlying stock. Empirical studies find that the average gain of the delta-hedged option portfolios is negative for both indexes and individual stocks (Bakshi and Kapadia (2003), Carr and Wu

(2009), and Cao and Han (2013)). Bakshi and Kapadia (2003) show that the sign and the magnitude of delta-hedged gain are related to the variance risk premium and the jump risk premium. The delta-hedged option position is constructed by holding a long position in an option, hedged by a short position of delta shares on the underlying stock. The definition of delta-hedged option gain follows Bakshi and Kapadia (2003) and is given by

$$\Pi_{t,t+\tau} = O_{t+\tau} - O_t - \int_t^{t+\tau} \Delta_u dS_u - \int_t^{t+\tau} r_u(O_u - \Delta_u S_u)du,$$

where C_t represents the price of an European option at time t, $\Delta_u = \frac{\partial C_u}{\partial S_t}$ is the option delta at time u, and r_u is the annualized risk-free rate at time u. We consider a portfolio of an option that is hedged discretely N times over the period $[t, t+\tau]$, where the hedge is rebalanced at each date t_n, $n = 0, 1, ...N-1$. As shown by Bakshi and Kapadia (2003) in a simulation setting, the use of the Black-Scholes hedge ratio has a negligible bias on delta-hedged gains. The discrete delta-hedged option gain up to maturity $t + \tau$ is defined as

$$\Pi_{t,t+\tau} = O_{t+\tau} - O_t - \sum_{n=0}^{N-1} \Delta_{t_n}[S_{t_{n+1}} - S_{t_n}] - \sum_{n=0}^{N-1} \frac{a_n r_{t_n}}{365}(O_{t_n} - \Delta_{t_n} S_{t_n}), \tag{5}$$

where O_t is the price of the option, Δ_{t_n} is the delta of the option at time t_n, r_{t_n} is the annualized risk free rate, and a_n is the number of calendar days between t_n and $t_n + 1$. This definition is used to compute the delta-hedged gain for call and put options by using the corresponding price and delta. To make the delta-hedged gains comparable across stocks we use delta-hedged option returns defined as the delta-hedged option gain $\Pi_{t,t+\tau}$ scaled by the absolute value of the securities involved, i.e. $\Delta_t S_t - O_t$ for call options. We start the position at the beginning of each month and close the position at the end of each month. We work with monthly returns through the empirical analysis.

3.3 Test portfolios and factor candidates in the equity option market

In the literature on the cross-section of stock returns, long-short portfolios are commonly used as stock return factors. These factors are constructed with portfolios composed by ranked stocks by certain characteristic, such as the size factor, the value factor or the momentum factor. We follow the same procedure for the equity option market and consider the predictors of option returns documented in the literature. These predictors are then used to sort portfolios and construct the list of candidate factors. The characteristics that predict option returns along with the long-short factors in the brackets are:

(1) Size (LS_{size}): The natural logarithm of the market value of the firm's equity (in Cao et al. (2017)).

(2) Stock return reversal ($LS_{reversal}$): The lagged one-month return (Cao et al. (2017)).

(3) Stock return momentum (LS_{mom}): The cumulative return on the stock over the 11 months ending at the beginning of the previous month (in Cao et al. (2017))

(4)CH (LS_{ch}): Cash-to-assets ratio, as in in Cao et al. (2017), defined as the value of corporate cash holdings over the value of the firm's total assets.

(5) Profit (LS_{profit}): Profitability, calculated as earnings divided by book equity in which earnings are defined as income before extraordinary items, as in in Cao et al. (2017).

(6) Disp (LS_{disp}): Analyst earnings forecast dispersion, computed as the standard deviation of annual earnings-per-share forecasts scaled by the absolute value of the average outstanding forecast (in Cao et al. (2017)).

Cao et al. (2017) find that delta-hedged option gains increase with size, momentum, reversal, and profitability and decrease with cash holding and analyst forecast dispersion. We also consider other option predictors in the literature related to volatility.

(7) Ivol (LS_{ivol}): Stock return idiosyncratic volatility, as in Ang et al. (2006). Cao and Han (2013) find that delta-hedged equity option return decreases monotonically with an increase in the idiosyncratic volatility of the underlying stock.

(8) Voldev (LS_{voldev}): The log difference between the realized volatility and the Black-Scholes implied volatility for at-the-money options. Goyal and Saretto (2009) find that the higher the difference, the higher the future straddle return of the equity option.

(9) Vts slope (LS_{vts}): the difference between long-term and short-term implied volatility. Vasquez (2017) finds that straddle portfolios with high slopes of the volatility term structure outperform straddle portfolios with low slopes by a significant amount.

(10) BidAsk (LS_{bidask}): The ratio of the difference between the bid and ask quotes of option to the midpoint of the bid and ask quotes at the end of previous month. Christoffersen

et al. (2017b) find that option illiquidity has strong risk premia in the equity option market. We use bid-ask spread as a proxy of option illiquidity due to data availability.

(11) Rating (LS_{rating}): Credit ratings are provided by Standard & Poor's and are mapped to 22 numerical values, where 1 corresponds to the highest rating (AAA) and 22 corresponds to the lowest rating (D). Vasquez and Xiao (2018) find that credit rating is a strong predictor of future option returns. Options with lower credit rating have more negative delta-hedged returns in the future.

At the end of each month, we sort all stock options into 10 portfolios based on the first 10 characteristics described above We sort stock options into 5 quintiles by credit rating because there are less than 10 different ratings in some months, which leads to missing data in the portfolio returns. We then start the position at the end of the month and hold it until the end of next month. Their corresponding delta-hedged option returns are calculated according to Section 3.2. We consider the 105 portfolios sorted by 11 different characteristics as test assets, such that they have enough heterogeneity and the underlying risk premium associated factors can be identified. The 11 candidate factors are the 10-1 (5-1 for credit rating) return spreads based on the 11 characteristics. We also consider two candidate factors related to common volatility risk:

(12) Delta-hedged return of the S&P 500 index option (DH_{idx}): DH_{idx} is a proxy for the market volatility risk in Coval and Shumway (2001) and Carr and Wu (2009).

(13) Delta-hedged return of stock options (DH_{stk}): DH_{stk} is the value-weighted delta-hedged returns on the individual stocks that are components of the S&P 500 index. It is used as a measure of common individual stock variance risk.

Table 1 reports summary statistics for the average returns of delta-hedged option portfolios sorted by the 11 predictors. The table shows that the long-short returns constructed by buying the top decile (quintile) and selling the bottom decile (quintile) are all significantly different from zero. The average return spreads range from -1.48% to 2.31% with t-statistics ranging from -10.49 to 16.05. The delta-hedged equity option returns increase with size, reversal, momentum, profitability, volatility deviation, and the slope of volatility term structure, while they decrease with cash holding, analyst dispersion, idiosyncratic volatility, bid ask spread, and credit rating.

[Table 1 around here]

Since the delta-hedged return of the S&P 500 index options is on average negative, which represents the negative price of variance risk, we construct the long-short factors based on the return spreads such that they are all on average negative. The long portfolio is the one with the highest payoff in bad states of nature as a hedge and the short portfolio is the one with the lowest payoff in bad states of nature. The summary statistics of the long-short factors including mean, standard deviation, skewness, kurtosis, 10th, 25th, 50th, 75th and 90th percentiles are reported in Table 2.

[Table 2 around here]

Table 3 shows the correlations among the long-short portfolios of the option return predictors. We observe that the correlation coefficients among the strategies are mostly below 0.5. Only the correlations between LS_{disp} and LS_{ivol} is 0.57, LS_{disp} and LS_{profit} is 0.53, and LS_{voldev} and LS_{credit} is 0.53. The low correlation among the long-short candidate factors suggests that these variables might capture distinctive information on the cross-section of delta-hedged option returns. But, how many different factors do these 13 candidate factors capture? How many of them are relevant for explaining covariance matrix of option returns? Do they capture similar information than stock returns factors? We address these questions in the next section.

[Table 3 around here]

4 Empirical Procedure and Results

Following the factor identification protocol from Pukthuanthong et al. (2018), we perform a latent variable analysis of the covariance matrix of delta-hedged option returns to uncover its factor structure. First we estimate the number of common factors that explain the covariation in option returns using six different methodologies. Then we compute these factors using principle component analysis (PCA) as suggested by Connor and Korajczyk (1986). Finally, we check their explanatory power on the cross-section of option returns.

Factors estimated by PCA are difficult to interpret economically. Therefore, we test if some factor candidates with economic interpretation are related to the covariance matrix of delta-hedged option returns using the rank-estimation method as suggested by Ahn et al. (2018). Third, we perform several tests to assess how much of the common variation captured by the PCA factors is captured by the selected factor candidates. For this purpose, we use standard regression analysis as well as a canonical correlation analysis as suggested in Pukthuanthong et al. (2018). Finally, we test whether the factors selected from the previous steps command risk premiums and propose a four factor model to explain the cross-section of delta-hedged option returns.

4.1 Number of latent factors that drive the comovement of delta-hedged option returns

As stated in Pukthuanthong et al. (2018), a necessary condition for any empirical factor candidate is to be related to the principal components of the covariance matrix. In this section, we aim to identify the factors that drive option returns.

Section 2 shows that under a stochastic volatility model the delta-hedged option return is driven by a linear factor model when the stock variance follows a multi-factor structure. In this section we estimate the number of factors under an approximate linear factor structure as defined in Chamberlain and Rothschild (1983).[3] More precisely, let x_{it} be the response variable for the ith cross-section unit at time t ($i = 1, 2, ..., N$, and $t = 1, 2, ..., T$). Explicitly, x_{it} can be the return on a delta-hedged option portfolio i at time t. The response variables x_{it} depend on K empirical factors $f_t = (f_{1t}, ..., f_{Kt})'$. That is,

$$x_t = \alpha + Bf_t + \epsilon_t,$$

[3]The advantage of working with approximate factor models as opposed to the classic exact factor models (e.g. Ross (1976)) is that the former allows for a certain degree of correlation across idiosyncratic terms while the later impose an orthogonality condition on the covariance matrix of the idiosyncratic component

where $x_t = (x_{1t}, ..., x_{Nt})'$ is the N-vector of response variables at time t; $\alpha = (\alpha_1, ..., \alpha_N)$ is the N-vector of individual intercepts; B is the $N \times K$ matrix of factor loadings (beta matrix); and $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{Nt})'$ is the N-vector of idiosyncratic components at time t. The idiosyncratic components ϵ_{it} can be weakly cross-sectionally and time-series correlated.

To estimate the number of factors K in delta-hedged option returns, we use as response variables data on all the delta-hedged portfolios defined in Section 3.3 (N = 105 portfolios) during the entire sample period from January 1996 to December 2015 (T = 240 months). As a preliminary step, we plot in Figure 1 the largest fifteen eigenvalues from the sample second-moment matrix of the "doubly demeaned" delta-hedged portfolio returns.[4]

The figure, known as a "scree plot", indicates that there are about six common factors, and one of them has much stronger explanatory power than the other five factors.

[Figure 1 around here]

Pukthuanthong et al. (2018) suggest that the number of factors should be designated in advance; for example, the number of factors could be chosen such that the cumulative variance explained by the principal components is at least ninety percent. However, the "ninety-percent" is an arbitrary cut-off point and a scree plot is not a formal statistical tool to estimate the number of factors either.

Instead, we use six different methodologies to estimate the number of factors necessary to explain the comovement of option returns: the Eigenvalue Ratio (ER) and Growth Ratio (GR) estimators of Ahn and Horenstein (2013), the Edge Distribution (ED) estimator of Onatski (2010), the BIC3 and IC1 estimators of Bai and Ng (2002), and the Modified Information Criterion estimator (ABC) of Alessi et al. (2010). Intuitively, these methods separate relevant information from noise exploiting the differential convergence rate of the eigenvalues that correspont to the common and idiosyncratic components of the covariance matrix from a factor model. A brief explanation about these methods is provided in Appendix A.1.

[4] Let x_{it} be the observed value of response variable i. Then, the "doubly demeaned" data is $x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}$, where $\bar{x}_i = \sum_{t=1}^{T} x_{it}/T$, $\bar{x}_t = \sum_{i=1}^{N} x_{it}/N$, and $\bar{x} = \sum_{i=1}^{N} \bar{x}_i/N$. Ahn and Horenstein (2013) recommends using doubly demeaned data when estimating the number of factors with eigenvalue-based methods to reduce the one-factor bias problem arising in finite samples when means are different than zero.

[Table 4 around here]

Table 4 reports the number of factors necessary to explain the comovement of option returns according to the six methodologies. The first column confirms our preliminary results from the scree test that there are between 1 and 6 common factors in doubly-demeaned delta-hedged option returns. We work with "doubly-demeaned" delta-hedged option returns so that none of the factors has constant loadings and that the factors explain cross-sectional differences in option returns. While ER, GR, and ED capture 1 common factor, BIC3 captures 3 common factors, and IC1 and ABC capture 6 common factors.[5] The factor structure is consistent with having 1 strong factor and possibly up to 5 weak factors.

When we use raw option returns, all estimators capture an additional factor as reported in the second column of Table 4. This additional factor is highly correlated with the equally-weighted portfolio (EWP) of option returns (99% correlation), it has unitary loadings, and explains the time-series of returns but not the cross-section as detailed in Appendix A.3. This result leads some researchers to argue that the most important factor is the market portfolio (e.g. Brown (1989), Ferson and Korajczyk (1995)), and that this factor is captured by the first PC from raw returns (or excess-returns over the risk free rate). This factor with constant betas explains the time-series variation in returns but cannot explain the cross-section. As Ahn and Horenstein (2018) show, to better capture the factors with variation in betas and the relevant factors to explain the cross-section of option returns, we can extract common factors from excess returns over the equally-weighted portfolio (or any well-diversified portfolio) and use them as response variables for testing assets pricing models.[6] Appendix A.3 shows that

[5]For all these estimators we set the parameter kmax, the maximum number of factors to test for, equal to 15.

[6]The intuition behind this suggestion is very simple. Let's assume that asset returns are driven by Black (1972) CAPM, i.e., $E(r_i) = E(r_0) + \beta_i \times E(r_{Mkt} - r_0)$ where r_0 and r_{Mkt} are the returns on the zero-beta and market portfolios, respectively. Note that this is a two factor model in which one of the factors, r_0, has unitary loadings. The common factors can be estimated by PCA using a large cross-section of data as shown by Bai (2003). However, the estimated factors are normalized eigenvectors that are unique only up to rotation, which could render impossible to disentangle r_0 from r_{Mkt}. Now, let's define the equally-weighted portfolio (EWP) as $\bar{r} = \sum_{i=1}^{N} r_i/N$ and assume for now that $\beta_{EWP} = 1$. If we use as response variables excess returns over $\bar{r}$ we have the following pricing equation: $E(r_i) - E(\bar{r}) = (\beta_i - 1) \times E(r_{Mkt} - r_0)$. Now, applying PCA to the modified data will lead to the estimation of the only factor necessary to explain the cross-section of expected returns. There are other advantages of using this simple modification to the data for testing asset

EWP contains unitary loadings. Therefore the latent model used for estimating the common factors explaining the cross-section of delta-hedged option returns consists of the following equation:

$$r_{it} - r_{EWP,t} = \alpha_i + \sum_{k=1}^{6} \tilde{\beta}_{ik} f_{kt} + \epsilon_{it},$$

where r_i is the return on the ith delta-hedged portfolio, r_{EWP} is the return on the equally-weighted portfolio (EWP) constructed using the 105 characteristic-based delta-hedged option portfolios, f_k corresponds to the kth PC factor[7] (where $k = 1, \ldots, 6$), $\tilde{\beta}_{ik} = (\beta_{ik} - 1)$ is the sensitivity of portfolio i to factor k, α_i is the pricing error of the model with respect to portfolio i, and ϵ_i is the idiosyncratic component of portfolio i. The PC factors have been extracted using the same rotation as Bai and Ng (2002).

Panel (a) of Table 5 shows the performance of the proposed model with 6 latent variables as factors. For comparison purposes, we present on Panel (b) the performance of the model using 13 candidate factors. The performance metrics we analyze are the percentage of non-zero alphas statistically different from 5%, the average adjusted R^2 across option portfolios, the correlation between the portfolios expected (excess) returns and the betas of each factor, the correlation between the portfolios expected (excess) returns and the predicted returns from the model (betas times factor premiums), and the annualized average absolute pricing error (AAAPE) generated by each model.

[Table 5 around here]

Table 5 shows that the model with 6 latent variables outperforms the model using the 13 candidate factors according to the performance metrics. The latent model produces less non-zero alphas statistically different than zero, slightly higher adjusted R^2, and lower annualized average absolute pricing errors. This result is not surprising in principle, since the latent factors are the most correlated factors with the second-moment matrix of our sample

pricing models that are discussed in Ahn and Horenstein (2018).

[7] The kth PC factor is the normalized eigenvector corresponding to the kth largest eigenvalue.

of delta-hedged option returns. The correlation between expected returns and predicted returns is quite high for both models: 96.7% for the model with 6 latent variables and 96.3% for the model with 13 candidate factors. A striking results is the 96% correlation between expected returns and the beta of the first latent factor. Figure 2 plots the scatter diagram between average returns and the betas of the first latent factor. There is a clear linear relationship between the two variables. Higher first-latent factor betas translate into higher average option returns.

[Figure 2 around here]

Consistent with the estimation for the number of factors in the previous section, the first factor seems to suffice for explaining the co-movement in delta-hedged returns. However, the other 5 weak factors might contain some relevant information. To decide how many, if any, of these weak factors are necessary to improve the model's performance, we evaluate six different latent factor models; on each model we sequentially increase the number of factors used as independent variables from 1 to 6. We use the following performance metrics for the 6 factor models to conclude how many factor are needed: (i) the number of non-zero alphas statistically significant at the 5% or less, (ii) the average adjusted R^2 generated by the model, and (iii) the annualized average absolute pricing error (AAAPE). We do not use the correlation between expected returns and realized returns because we already know that the first factor generates a 96% correlation of the total of 97% generated by all six latent factors.

Panel (c) of Table 5 reports the performance metrics for the 6 latent factor models. A model with one latent factor generates 54% of alphas different from zero. After adding a second factor to the model, the percentage decreases to 22% and the minimum non-zero alphas is obtained with 3 factors at 18%. According to this metric, additional factors beyond the third one do not add relevant pricing information. The average adjusted R^2 metric shows that the first factor explains an average of 14% of the return variation, the second factor explains 6%, and after that each additional factor only explains 3%. Finally, the third metric (AAAPE) decreases with the number of factors. However, the change is quite small after the third factor. Overall, a model with three latent factors provides the best fit to price delta-hedged option returns.

In the next section we explore the economic interpretation of the factors by studying which of the 13 candidate factors explain the 6 latent factors. Given that there are at most 6 common factors in total and at most 3 common factors relevant for pricing, many of the 13 candidate factors proposed as predictors might contain redundant information. Hence it is crucial to identify the best candidate factors for pricing.

4.2 Relevant candidate factors of the cross-section of delta-hedged option returns

In the previous section we find that there are at most 6 factors in option returns capable to explain their cross-sectional variation. However, we do not know how many of the 6 factors are captured by the 13 variables that predict option returns. This can be solved by estimating the rank of the beta matrix produced by the 13 variables when regressed onto the 105 delta-hedged option portfolios. As Ahn et al. (2018) point out, "the rank of the beta matrix corresponding to a set of factors equals the number of factors whose prices are identifiable." In other words, the rank of the beta matrix tells us the number of different sources of stock returns' comovement captured by a set of factors. When we apply the RBIC estimator developed in Ahn et al. (2018), we find that the rank of the beta matrix generated by the 13 variables proposed as predictors of option returns equals 6. Therefore, the 13 variables are capturing all the relevant factors containing information about the co-movement of delta-hedged option returns. This result is consistent with our analysis in Section 4.1.

Table 6 shows the correlation coefficients between the 13 candidate factors and the six latent factors. Since the EWP can be considered a factor, although it has unitary loadings, we add it to the table.

[Table 6 around here]

Many characteristic-based factors are relatively highly correlated with the latent factors, suggesting that some candidate factors might be capturing similar information. Therefore, we now estimate the minimum set of variables necessary to capture the 6 common latent factors using rank estimation methods. More precisely, we generate all combinations of different sets containing 6 factors from the 13 candidate factors and check which sets generate a full-rank beta matrix. From the 13 candidate factors, we can create 462 different sets of 6 factors. We find that only 1 of these 462 sets generates a full-rank beta matrix using the delta-hedged portfolio returns over the EWP as response variables. The unique set generating a full rank beta matrix contains the following factors: LS_{size}, LS_{ch}, LS_{disp}, LS_{ivol}, LS_{voldev}, and LS_{rating}. These six factors suffice to capture most (if not all) the relevant information on the cross-section of delta-hedged option returns.

Figure 3 shows the relationship between average option portfolio returns and predicted returns by a model containing the six relevant factors. The two variables have a correlation coefficient of 95%. As shown in Panel (B) of Table 5, this correlation increases to 96% if we use all of the 13 candidate factors. Overall, these six variables contain all relevant information about the cross-section of option returns.

[Figure 3 around here]

But how well do the six candidate factors explain the six latent factors? To answer this question, we use canonical correlation analysis as suggested in the factor identification protocol by Pukthuanthong et al. (2018). Canonical correlations assesses the relationship between the two factor matrices by finding linear combinations of the variables that maximally correlate. A brief explanation on this procedure is on Appendix A.2.

[Table 7 around here]

Table 7 reports the canonical correlations for different pairs of variables. The first column shows the canonical correlation of the set containing 6 latent factors and the one with 13 factor candidates. The first five canonical correlations are quite high. This result indicates that five linear combinations of the 13 factor candidates are highly correlated with five linear combinations of the 6 latent factors. It appears that the candidate factors almost perfectly capture five dimensions spanned by the 6 latent factors. However one dimension is missed; if it were the one spanned by the first latent factor, the implication would be that the candidate factors are missing the most important dimension. Therefore, in the second column we compare the 3 most relevant latent factors with the 13 factor candidates. We find that the 13 candidate factors capture almost perfectly the three most important latent factors for pricing. The 13 factor candidates contain the most relevant information for pricing delta-hedged option returns. Columns three and four compare the latent factors with the subset of the 6 factor candidates that are most relevant with our rank estimation exercise. We find that the 6 most relevant factor candidates capture almost perfectly 3 of the latent factors, and that these latent factors are the 3 most relevant ones. This further confirms our rank estimation exercise.

Next we select which of the six relevant factor candidates are able to capture most of the information in the three main latent factors. We rank by importance the six variables which are sufficient to explain the cross-section of delta-hedged option returns. The most relevant variable is the one that produces by itself the largest correlation coefficient between beta and average returns. The second variable in our rank is the one that most increases the correlation between average returns and predicted returns in a model already containing the first variable as regressor.

We find that we only need 3 out of the 6 candidate variables to get a correlation of 95% between average observed and predicted returns. LS_{size} is the variable that produces the highest correlation of 83%, LS_{ivol} increases that correlation to 89%, and LS_{voldev} further increases it to 95%. This does not mean that the other three variables—LS_{ch}, LS_{disp1}, and LS_{rating}—are uninformative. However, their added explanatory power to the cross-section of option returns is negligible once we control for the selected three factors. In fact, the three variables chosen generate quite high canonical correlation with respect to the first two latent factors of 93% and 89%, arguably the most important ones. The third latent factor is explained with lesser accuracy as its canonical correlation coefficient is only 58%.

Overall, all our results are consistent with our previous estimation on the number of factors showing that there is one strong factors and 5 weak factors, where the weakest 3 factors are much weaker than the other two (see Figure 1). Given these results, we now propose an asset pricing model that can capture well the time-series and cross-sectional variation in delta-hedged returns. In the previous analysis, we find that the three most important factors to capture the cross-sectional variation in delta-hedged option returns are LS_{size}, LS_{ivol}, and LS_{voldev}. We also show and further investigate in Appendix A.3 that a variable that captures well the factor with unitary loadings is the delta-hedged return of S&P 500 index options DH_{idx}, which is commonly considered the market factor in the option returns literature (Goyal and Saretto (2009) and Cao and Han (2013)). Thus, we propose the following four-factor model for pricing delta-hedged equity option returns:[8]

$$r_{it} = \alpha_i + \beta_{idx,i} DH_{idx,t} + \beta_{size,i} LS_{size,t} + \beta_{ivol,i} LS_{ivol,t} + \beta_{voldev,i} LS_{voldev,t}$$

Table 8 below shows the performance metrics for the four-factor model in Panel (a) and the four-factor model augmented with LS_{ch}, LS_{disp1}, and LS_{rating} in Panel (b).

[Table 8 around here]

We observe in Table 8 that the four-factor and the seven-factor models perform almost identically according to the performance metrics. The non-zero alphas, the adjusted r-square, the correlation between average returns and returns predicted by the model, and the average absolute annualized alphas are quite similar for both models. We conclude that the four-factor model captures all the necessary information for pricing purposes.

Since the factor structure might change over time, we now evaluate the performance of the proposed factor model for different sub-samples. We run monthly rolling regressions using 60 months of data in each iteration. Since our data comprise the period January 1996 to December 2015, we have 181 regressions in total. For each regression we calculate the correlation between the average returns of the delta-hedged portfolios and the predicted returns by our four factor model. The average correlation is 86%, with a standard deviation of 4%, a maximum value of 93% and a minimum of 72%. Figure 4 shows the evolution of the 5-year

[8]Note that the delta-hedged returns are in excess of the risk free rate as in Equation 5.

correlation over time. Overall the predictive power of the four factor model is quite stable in our sample period.

[Figure 4 around here]

We further check if the factors proposed in our model are priced in the cross-section by running Fama-MacBeth regressions for the whole sample and two 10-year sub-samples. Regression results are reported in Table 9. To avoid multicollinearity problems arising from having a factor with near constant loadings (see Ahn et al. (2013)), we only test the three factors that show variability in their loadings: LS_{size}, LS_{ivol}, and LS_{voldev}. Table 9 shows that the proposed factors are priced since all the coefficients are statistically significant in the full sample and the two subsamples. As such, the factors in our model pass all the criteria in the protocol established by Pukthuanthong et al. (2018) to find relevant factors.

[Table 9 around here]

4.3 Robustness checks: Pricing other portfolios

In the previous section, we search for the best model – based on tradable factors – to price the returns of characteristic-based delta-hedged call option portfolios. In this section, we evaluate the performance of the four-factor model developed in the previous section on three additional set of portfolios: 26 delta-hedged call option industry portfolios, 105 characteristic-sorted delta-hedged put option portfolios, and 105 variance risk premium portfolios.

We construct industry portfolios by categorizing each firm into the industry group with two-digit code provided by OptionMetrics. After removing portfolios with missing data, we have 26 industry portfolios in total. The 105 characteristic-sorted delta-hedged put option portfolios are constructed in the same way as for the call option portfolios. The only difference is that we use at-the-money put options to construct the portfolios. Finally, since the expected return of the delta-hedged option portfolio is closely related to the variance risk premium of the underlying asset, we also consider the variance risk premium (VRP) portfolios. The returns of the VRP portfolios are constructed as the returns of the synthetic variance swaps. Variance risk premiums are inferred as the difference between the synthetic variance swap rate and the realized return variance as in Carr and Wu (2009).

We start by showing the relationship between the predicted expected returns by the four-factor model and the realized expected return by the test portfolios. Figure 5, Panels (a), (b), (c) and (d) show the results for the 105 characteristic-sorted delta-hedged call option portfolios, 26 industry delta-hedged call option portfolios, 105 characteristic-sorted delta-hedged put option portfolios, and 105 VRP portfolios, respectively.

[Insert Figure 5 around here]

Figure 5 shows that predicted returns by the model are inline with the realized average returns of the option portfolios. To further study the performance of the four-factor model for other portfolios, we report in Table 10 several performance metrics. For comparison purposes, we report the previous results for the 105 characteristic-sorted delta-hedged call option in Panel (a), while panels (b), (c), and (d) show the results for the 26 delta-hedged call option industry portfolios, 105 characteristic-sorted delta-hedged put option portfolios, and 105 VRP portfolios, respectively.

[Insert Table 10 around here]

The performance metrics for the four-factor model are quite outstanding and do not differ much across portfolios. For example, in most cases, very few portfolios produce abnormal returns statistically different from zero. Only 2 out of the 26 industry portfolios and 7 out of the 105 characteristic-sorted delta-hedged put portfolios have significant alphas. The correlation between realized portfolio returns and the predicted returns by the model is quite high: 0.825 for the 26 delta-hedged industry portfolios, 0.866 for the 105 characteristic-sorted delta-hedged put portfolios, and 0.874 for the 105 VRP portfolios. Overall, the four-factor model we proposed can capture the variance risk premium in option returns and the model fits quite well the risk-return relationship in the space of delta-hedge option returns.

5 Conclusion

Despite the very large and still growing literature on common factors in stock returns, there is limited understanding about the factor structure in delta-hedged equity option returns. In this paper, we motivate our empirical analysis by showing that in a stochastic volatility model, the expected delta-hedged option returns are driven by factors related to volatility risks and not by stock returns' factors.

In the empirical analysis, we construct 105 option portfolios sorted by 11 characteristics using monthly portfolios of delta-hedged options from January 1996 to December 2015. We choose these characteristics because the literature has documented that they predict future option returns. The characteristic-based factors are constructed based on long-short strategies of decile portfolios. We also consider delta-hedged return of the S&P 500 and average delta-hedged return of the equity options as two additional candidate factors. Using latent factor techniques on the 105 portfolios, we find strong evidence for the existence of at most six factors in equity options returns, where three suffice to explain its cross-sectional variation. Using the 11 characteristic-based option factors as candidate factors, we find that three of them suffice to capture the relevant latent factors and to explain the time series and cross-section of equity option returns. The three factors are long-short factors constructed based on size, idiosyncratic volatility, and volatility deviation.

Fama-MacBeth regressions show that the exposures to the three factors are statistically significant when explaining average delta-hedged returns of the 105 portfolios in the full sample and in two subsamples with adjusted R^2 above 80%.

Finally, we propose a four-factor pricing model for delta-hedged option returns that adds the delta-hedged return of the S&P 500 index options to the three aforementioned characteristic-based factors. This factor improves the explanatory power of the model over in the time-series dimension. In the main analysis we derive the proposed factor model using delta-hedged call option portfolio returns. The explanatory power of our proposed factor model extends to delta-hedged industry call option portfolios, and delta-hedged put option portfolios and variance risk premium portfolios sorted by the 11 characteristics previously discussed.

References

Seung C Ahn and Alex R Horenstein. Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227, 2013.

Seung C Ahn and Alex R Horenstein. Asset pricing and excess returns over the market return. *Working Paper*, 2018.

Seung C Ahn, M Fabricio Perez, and Christopher Gadarowski. Two-pass estimation of risk premiums with multicollinear and near-invariant betas. *Journal of Empirical Finance*, 20: 1–17, 2013.

Seung C Ahn, Alex R Horenstein, and Na Wang. Beta matrix and common factors in stock returns. *Journal of Financial and Quantitative Analysis*, 53(3):1417–1440, 2018.

Lucia Alessi, Matteo Barigozzi, and Marco Capasso. Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, 80 (23-24):1806–1813, 2010.

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1): 135–171, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

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

Francisco Barillas and Jay Shanken. Comparing asset pricing models. *The Journal of Finance*, 73(2):715–754, 2018.

Fischer Black. Capital market equilibrium with restricted borrowing. *The Journal of business*, 45(3):444–455, 1972.

Robert Brooks, Don M Chance, and Mobina Shafaati. The cross-section of individual equity option returns. *Working Paper*, 2018.

Stephen J Brown. The number of factors in security returns. *the Journal of Finance*, 44(5): 1247–1262, 1989.

Matthias Büchner and Bryan Kelly. A factor model for option returns. *Working Paper*, 2019.

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

Jie Cao, Bing Han, Qing Tong, and Xintong Zhan. Option return predictability. *Working Paper*, 2017.

Peter Carr and Liuren Wu. Variance risk premiums. *Review of Financial Studies*, 22(3): 1311–1341, 2009.

Raymond B Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966.

Gary Chamberlain and Michael Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281, 1983.

Peter Christoffersen, Mathieu Fournier, and Kris Jacobs. The factor structure in equity options. *The Review of Financial Studies*, 31(2):595–637, 2017a.

Peter Christoffersen, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui. Illiquidity premia in the equity options market. *The Review of Financial Studies*, 31(3):811–851, 2017b.

Gregory Connor and Robert A Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3): 373–394, 1986.

Joshua D Coval and Tyler Shumway. Expected option returns. *The Journal of Finance*, 56 (3):983–1009, 2001.

John C Cox, Stephen A Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

Karl B Diether, Christopher J Malloy, and Anna Scherbina. Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5):2113–2141, 2002.

Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

Eugene F Fama and Kenneth R French. Profitability, investment and average returns. *Journal of Financial Economics*, 82(3):491–518, 2006.

Guanhao Feng, Stefano Giglio, and Dacheng Xiu. Taming the factor zoo. *Working Paper*, 2017.

Wayne E Ferson and Robert A Korajczyk. Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business*, pages 309–349, 1995.

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

Kewei Hou, Haitao Mo, Chen Xue, and Lu Zhang. Which factors? *Working Paper, The Ohio State University*, 2018.

Christopher S Jones. A nonlinear factor analysis of s&p 500 index option returns. *The Journal of Finance*, 61(5):2325–2363, 2006.

Mustafa Mete Karakaya. Characteristics and expected returns in individual equity options. *Working Paper*, 2014.

Jonathan Lewellen, Stefan Nagel, and Jay Shanken. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2):175–194, 2010.

Alexei Onatski. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016, 2010.

Berardino Palazzo. Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1):162–185, 2012.

Kuntara Pukthuanthong, Richard Roll, and Avanidhar Subrahmanyam. A protocol for factor identification. *The Review of Financial Studies*, Forthcoming, 2018.

Stephen Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

Jay Shanken. On the estimation of beta-pricing models. *The Review of Financial Studies*, 5 (1):1–33, 1992.

Aurelio Vasquez. Equity volatility term structures and the cross section of option returns. *Journal of Financial and Quantitative Analysis*, 52(6):2727–2754, 2017.

Aurelio Vasquez and Xiao Xiao. Default risk and option returns. *Working Paper*, 2018.

A Appendix

A.1 Summary of the number of factors tests used in the paper

One of the most popular methods used to determine the number of common factors is the Scree Test developed by Cattell (1966). The test is a visual way to find out the number of factors from the eigenvalues of XX', where X is the $T \times N$ matrix of response variables. Cattell describes his method as follows. “If I plotted the principal components in their sizes, as a diminishing series, and then joined up the points all through the number of variables concerned, a relatively sharp break appeared where the true number of factors ended and the ‘detritus’, presumably due to error factors, appeared. From the analogy of the steep descent of a mountain till one comes to the scree of rubble at the foot of it, I decided to call this the Scree Test.”

The Scree Test is an eye-ball test (not a formal estimator with known asymptotic properties). Bai and Ng (2002) (BN, 2002) propose the first consistent estimator for the number of factors. Their method can be interpreted as finding a consistent threshold to divide the eigenvalues corresponding to common factors from those corresponding to noise. To be specific, denote $\psi_k(A)$ as the kth largest eigenvalue of a positive semi-definite matrix A. Define

$$\tilde{\mu}_{NT,k} = \psi_k(\frac{1}{NT}XX') = \psi_k(\frac{1}{NT}XX'),$$

where $k = 1, ..., m$, and $m = min(T, N)$. Then, $T\tilde{\mu}_{NT,k}$ ($N\tilde{\mu}_{NT,k}$) is the kth largest eigenvalue of the sample covariance matrix of $x_{i\cdot}$ ($x_{\cdot i}$) if the means of x_{it} are all zeros. Ahn and Horenstein (2013) show that if r is the true number of factors, then $\tilde{\mu}_{NT,k} = O_p(1)$ for $k \leq r$ while $\tilde{\mu}_{NT,k} = O_p(m^{-1})$ for $k > r$.

Next we describe all the estimators used in the paper. Let $\tilde{F}^k$ be the $T \times k$ matrix of the eigenvectors corresponding to the largest k eigenvalues of $XX'/(NT)$, which, following Bai and Ng (2002), is normalized to either $(\tilde{F}^k)'(\tilde{F}^k)/T = I_k$ or $(\tilde{\Lambda}^k)'(\tilde{\Lambda}^k)/N = I_k$. Let $V(k)$ be the mean of squared residuals from the regressions of $x_{i\cdot}$ on $\tilde{F}^k$ and with $\hat{\sigma}^2 = V(kmax)$, where $kmax$ is the maximum number of factors to be tested. BN proposed minimizing two types of criteria to consistently estimate the number of factors:

$$PC(k) = V(k) + \hat{\sigma}^2 k g(N,T),$$

$$IC(k) = \ln(V(k)) + kg(N,T),$$

where $g(N,T)$ is a penalty function such that $g(N,T) \to 0$ and $g(N,T)m \to \infty$ as $N,T \to \infty$. A BN estimator is obtained by minimizing these functions. BIC3 is a PC criterion where $g(N,T) = \frac{(N+T-k)\ln(NT)}{NT}$, while IC1 is an IC criterion where $g(N,T) = \frac{N+T}{NT}\ln(\frac{NT}{N+T})$. Note that the threshold values used for the estimators are not unique. Specifically, any finite multiple of a valid threshold value is also asymptotically valid. To show the relationship between BN and the Scree Test, note that

$$V(k) = \frac{1}{NT}\sum_{i=1}^{N}(x'_{i\cdot}x_{i\cdot} - x'_{i\cdot}P(\tilde{F}^k)x_{i\cdot}) = \sum_{j=k+1}^{T}\tilde{\mu}_{NT,j} \tag{6}$$

See Ahn and Horenstein (2013) for more details. Then, $\hat{\sigma}^2 = V(kmax) = \sum_{j=kmax+1}^{T}\tilde{\mu}_{NT,j}$ and $PC(k)$ can be written as

$$PC(k) = \sum_{j=k+1}^{T}\tilde{\mu}_{NT,j} + (\sum_{j=kmax+1}^{T}\tilde{\mu}_{NT,j}) \times kg(N,T),$$

Let $\tilde{k}_{PC} = min_{k \leq kmax} PC(k)$, using Equation 6 and the monotonicity of the eigenvalues, we can show that $\tilde{k}_{PC} = min_{k \leq kmax}\{k|\tilde{\mu}_{NT,j} \geq \hat{\sigma}^2 g(N,T)\}$. Thus, the PC estimation can be viewed as a Scree Test using $\hat{\sigma}^2 g(N,T)$ as a threshold value.

Overall, the BN estimators can be roughly understood as a formalization of the Scree Test. The ER (eigenvalue ratio) estimator of Ahn and Horenstein (2013) can be viewed as a modified version of the PC estimator that does not require the use of threshold values. The ER estimator is defined by maximizing the following criterion function:

$$ER(k) = \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} = \frac{V(k-1) - V(k)}{V(k) - V(k+1)}$$

Thus, the ER estimator is the value of k that maximizes the ratio of the changes in $V(k)$

between $k-1$ and k. A similar interpretation can be given to the GR (growth ratio) developed also in Ahn and Horenstein (2013), which is defined by maximizing the following criterion function:

$$GR(k) = \frac{ln(1 + \tilde{\mu}_{NT,k})}{ln(1 + \tilde{\mu}_{NT,k+1})} = \frac{ln(V(k-1)/V(k))}{ln(V(k)/V(k+1))}$$

where $k \leq kmax$. An advantage of Ahn and Horenstein (2013) estimators is that they do not depend on a pre-specified threshold function. This is important because the threshold function depends mostly on the values of N and T and not too much on the particular data generating process of the observations. In addition, there are infinitely many consistent threshold functions that could be used that might lead to different estimation results in finite samples.

Alessi et al. (2010) (ABC, 2010) propose to estimate the number of factors using different subsamples and different multiples of the BN penalty functions. The final estimators is the one that is invariant to the subsamples used and the change in the multiplicative constant of the penalty function within a certain range. Overall, the ABC estimator can be considered as a refinement of BN. The last estimator we use is the Edge Distribution (ED) estimator of Onatski (2010) that estimates the number of factors using differenced eigenvalues. A novelty of his approach is that the threshold value is estimated, not pre-specified like in BN or ABC.

To sum up, we use different estimators but all of them can be linked to the behavior of the eigenvalues from the second moment matrix of the data. The rationale to all of them is to separate information from noise, with the known property that eigenvalues corresponding to eigenvectors having common information do not vanish as the dimension of the panel (N, T) increases while those corresponding to eigenvectors related to the idiosyncratic component do vanish. BN separates information from noise using a pre-specified penalty function. ABC refines BN but still uses a pre-specified penalty function. ED estimates the penalty function from the data. ER and GR do not use penalty functions.

A.2 Canonical Correlation

In this subsection, we explain the details of the canonical correlation. Suppose there are K latent factors selected from Section 4.1: $F = (f_1, f_2, ..., f_L)'$ and L candidate factors as return spreads constructed by sorting portfolios based on characteristics: $X = (x_1, x_2, ..., x_L)$. Their mean factors are μ_F and μ_X and their covariance matrices are Σ_F and Σ_X. The covariance matrix between F and X is $\Sigma_{FX} = E[(X - \mu_X)(f - \mu_F)']$.

Define linear combinations of X and F as U and V: $U = a'F$ and $V = b'X$, where a and b are constant vectors. Note that the variance of the two vectors and the covariance between the two vectors are:

$$Var(U) = a'\Sigma_F a, \;\; Var(V) = b'\Sigma_X b, \;\text{ and }\; Cov(U, V) = a'\Sigma_{FX} b.$$

The first pair of canonical variate (U_1, V_1) is defined via the pair of linear combination vectors $\{a_1, b_1\}$ that maximize the following correlation, subject to the condition that U_1 and V_1 have unit variance:

$$Corr(U, V) = \frac{Cov(U, V)}{\sqrt{Var(U)}\sqrt{Var(V)}} = \frac{a'\Sigma_{FX} b}{\sqrt{a'\Sigma_F a}\sqrt{b'\Sigma_X b}}.$$

The remaining canonical variates (U_p, V_p) maximize the above correlation subject to having unit variance and being uncorrelated with $((U_q, V_q))$ for all $q < p$. The p-th pair of canonical variates is given by,

$$U_p = u_p'\Sigma_F^{-1/2} F, \;\text{ and }\; V_p = v_p'\Sigma_X^{-1/2} X$$

where u_p is the p-th eigenvector of $\Sigma_F^{-1/2}\Sigma_{FX}\Sigma_X^{-1}\Sigma_{XF}\Sigma_F^{-1/2}$ and v_p is the p-th eigenvector of $\Sigma_X^{-1/2}\Sigma_{XF}\Sigma_F^{-1}\Sigma_{FX}\Sigma_X^{-1/2}$. The p-th canonical correlation is given by $Corr(U_p, V_p) = \rho_p$, where $'\rho_p^2$ is the p-th eigenvalue of $\Sigma_F^{-1/2}\Sigma_{FX}\Sigma_X^{-1}\Sigma_{XF}\Sigma_F^{-1/2}$ and $\Sigma_X^{-1/2}\Sigma_{XF}\Sigma_F^{-1}\Sigma_{FX}\Sigma_X^{-1/2}$.

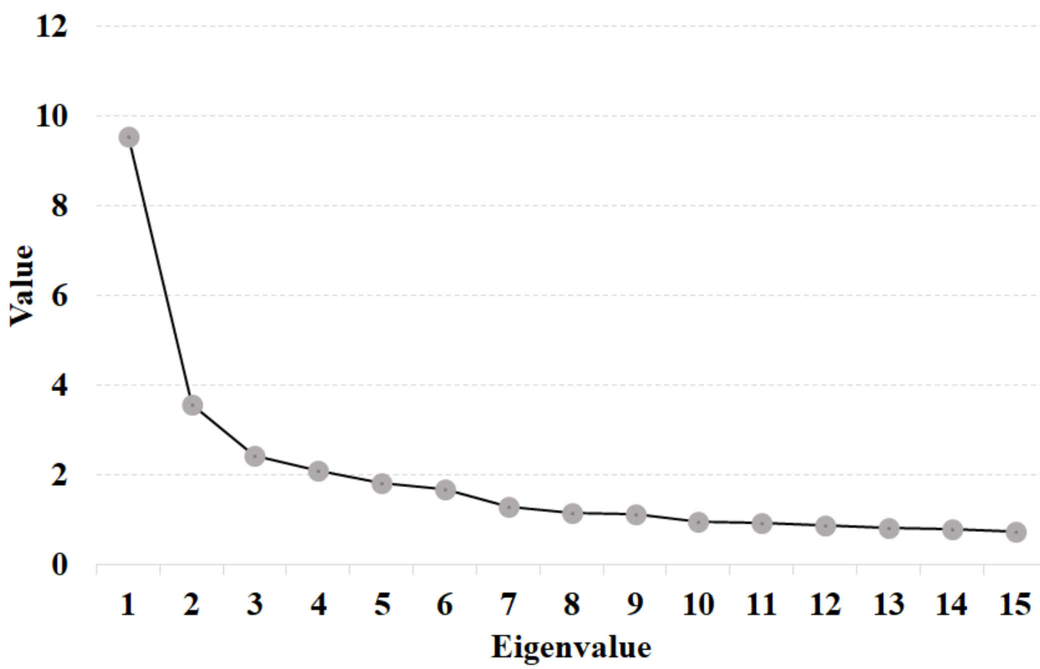
A.3 Equal weighted portfolio (EWP) contains unitary loadings

In the main body of the paper, we argue that EWP is a factor with constant loadings. Ahn et al. (2013) show that these factors lead to specious conclusions of statistical significance when using them as explanatory variables in typical asset pricing tests like the Fama-MacBeth two pass regression method. In addition, Ahn et al. (2018) show that subtracting a factor with constant loadings from the response variables in asset pricing tests helps to better identify the relevant pricing factors (see footnote 5). Therefore, in the main body of the paper we use as response variable excess delta-hedged portfolio returns over the EWP for estimating the principal component factors. In this Appendix we show that not only EWP is a factor with unitary loadings but also the $Straddle_{index}$ factor, which is usually used as a market factor in the option returns literature (see Goyal and Saretto (2009) and Cao and Han (2013) for example).

To assess the variability of the different factor loadings we use the invariant beta (IB) estimator proposed in Ahn and Horenstein (2013). More precisely, if we have N response variables and K factors, then $IB_k = N\bar{\beta}_k^2/\sum_{i=1}^{N}\hat{\beta}_{ik}^2$ where $\hat{\beta}_{ik}$ is the estimated beta for asset i corresponding to factor k and $\bar{\beta}_k = \sum_{i=1}^{N}\hat{\beta}_{ik}/N$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$. The IB_k estimator is between 0 and 1 and is equivalent to the uncentered R-square from regressing a vector of ones onto the vector of factor k's loadings. The closer to 1 the value of IB the closer to a constant the vector of factor loadings. Table A1 below shows the values of the IB estimator for two factor models. Panel (a) shows the results from regressing the 105 delta-hedged option returns on a model consisting of the EWP and the three characteristic-based option factors from our four-factor model (Size, ivol, and voldev) and Panel (b) show the results fusing the proposed four-factor model with the $Straddle_{index}$ as factor market factor. The table also reports the average value of the factor loading as well as the R^2 of the models and the correlation between the predicted returns and the average returns of the models.

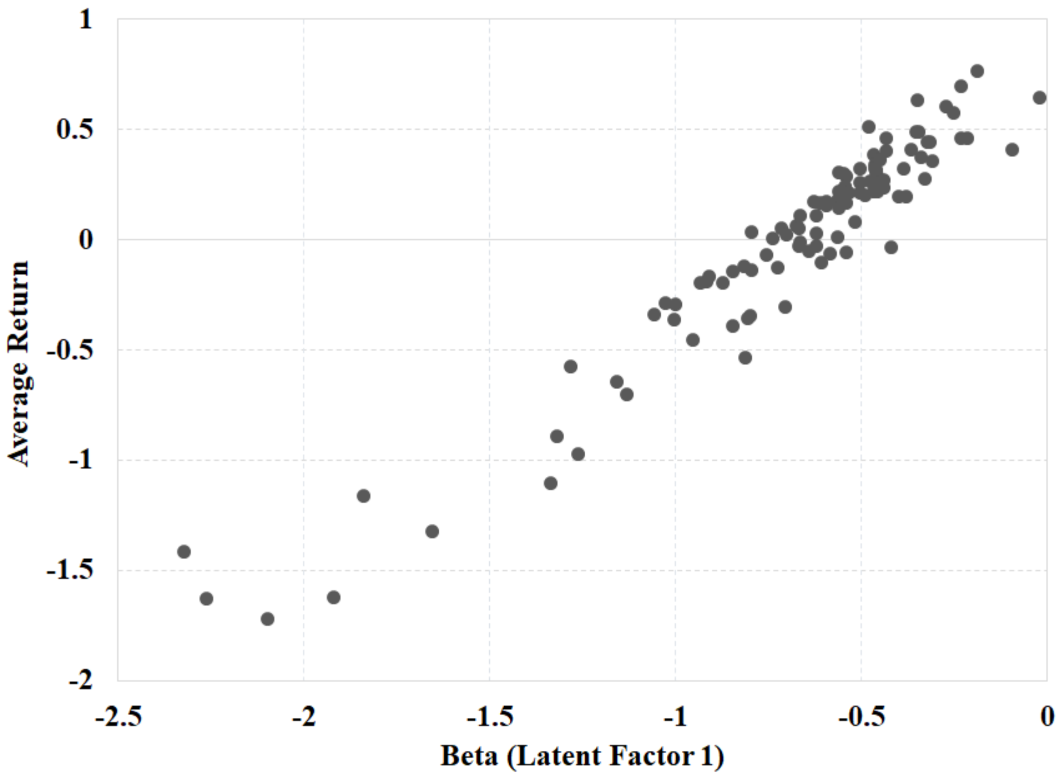
The Table shows that the IB estimator is almost equal to 1 for both EWP and the $Straddle_{Index}$ factor. As such, these factors might have good explanatory power for the time series regressions but fail to explain the cross-section of delta-hedged option returns, as shown by the low correlation between their estimated betas and the average returns of the delta-hedged portfolios. Both models produce the same value of the correlation coefficient between expected and predicted returns. This is expected since factors with constant betas add no information for the cross-section of expected returns. Finally, the R^2 of the model using EWP is much higher. This is expected given that EWP is an "in-sample" market portfolio.

Figure 1: Eigenvalues from the Second-moment Matrix of the "Doubly-demeaned" Delta-hedged Call Option Portfolio Returns



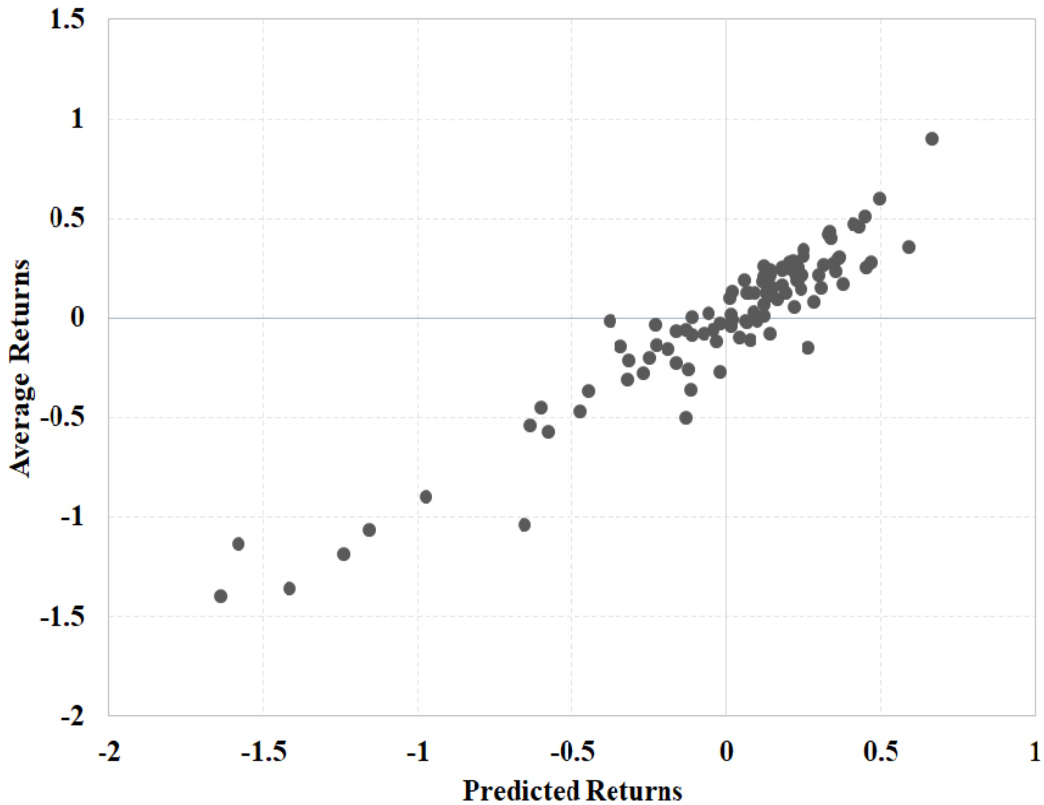
This figure shows the largest fifteen eigenvalues from the sample second-moment matrix of the "doubly demeaned" returns of the 105 delta-hedged portfolios. The sample period is from January 1996 to December 2015.

Figure 2: Beta-return Relationship of the First Latent Factor



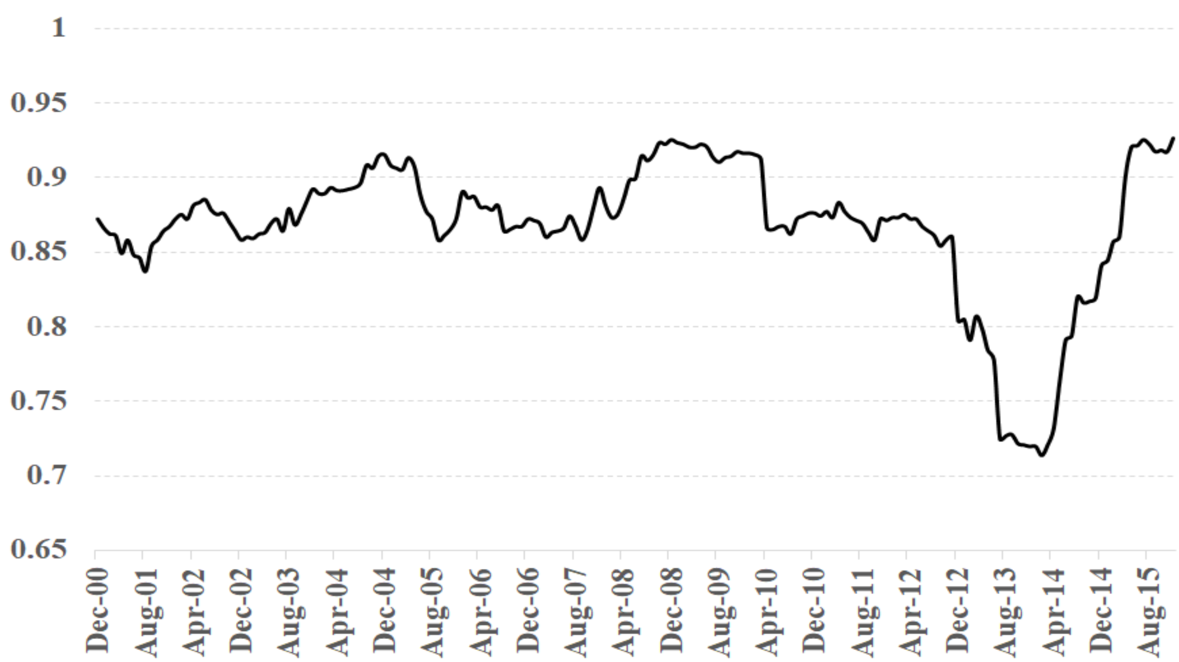
This figure shows the scatter diagram between the average returns of the 105 delta-hedged call option portfolios and their corresponding betas of the first latent factor. The sample period is from January 1996 to December 2015.

Figure 3: Average Returns and Predicted Returns by the Model with Six Latent Factors



This figure shows the relationship between the average returns of the 105 delta-hedged call option portfolios and the predicted returns by the model containing the six latent factors. The sample period is from January 1996 to December 2015.

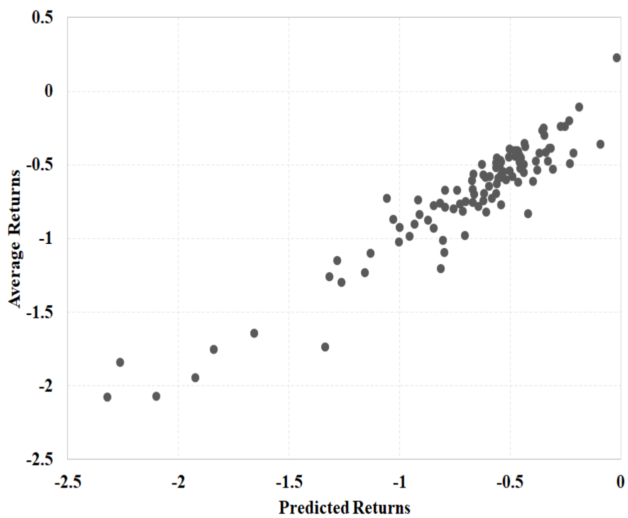
Figure 4: Correlation of Average Return and Predicted Return by the Four-factor Model Over Time



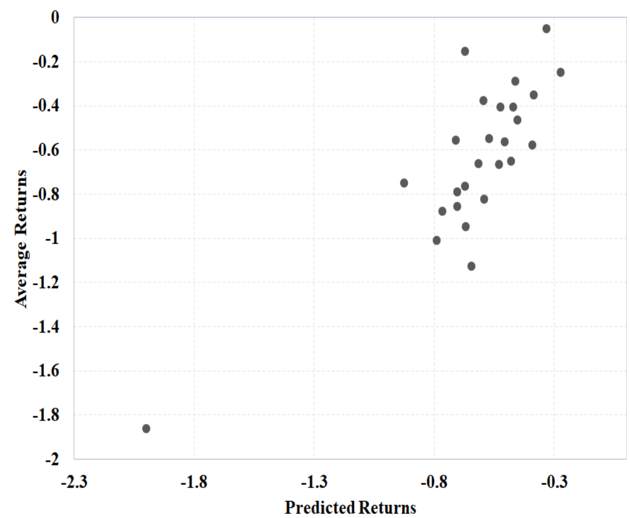
This figure shows the correlation of average return and predicted returns by the four-factor model over time. We run monthly rolling regressions using 60-month of data in each regression. For each regression we calculate the correlation between the average returns of the delta-hedged portfolios and the predicted returns by our three-factor model.

Figure 5: Average Returns and Predicted Returns by the Four Factor Model

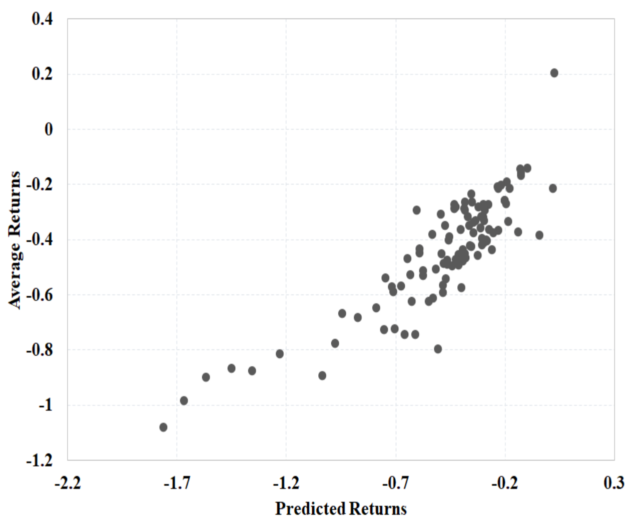
(a) 105 Call Portfolios



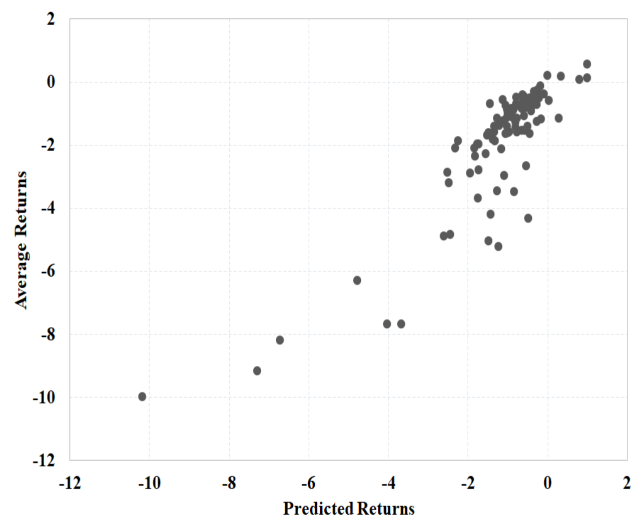
(b) 26 Industry Portfolios



(c) 105 Put Portfolios



(d) 105 VRP Portfolios



This figure shows the scatter diagram between the realized average return of the test portfolios and the predicted returns by the four factor model. Panels (a), (b), (c) and (d) show the results for the 105 characteristic-sorted delta-hedged call option portfolios, 26 industry delta-hedged call option portfolios, 105 characteristic-sorted delta-hedged put option portfolios, and 105 variance risk premium portfolios, respectively. The sample period is from January 1996 to December 2015.

Table 1: Delta-hedged option return sorted by 11 characteristics

	1	2	3	4	5	6	7	8	9	10	10 − 1
Size	-2.10	-1.26	-0.78	-0.63	-0.55	-0.44	-0.42	-0.31	-0.27	-0.23	1.87***
	(-14.50)	(-9.22)	(-6.06)	(-5.66)	(-4.87)	(-4.17)	(-3.97)	(-2.96)	(-2.44)	(-2.26)	(16.05)
Reversal	-1.31	-0.82	-0.65	-0.67	-0.58	-0.58	-0.53	-0.54	-0.59	-0.72	0.60***
	(-7.84)	(-6.28)	(-5.89)	(-6.41)	(-5.14)	(-5.71)	(-5.71)	(-5.45)	(-5.77)	(-5.13)	(5.45)
Mom	-1.31	-0.82	-0.73	-0.62	-0.56	-0.54	-0.50	-0.54	-0.54	-0.80	0.51***
	(-9.17)	(-6.71)	(-6.50)	(-6.02)	(-5.36)	(-5.30)	(-4.85)	(-5.13)	(-4.50)	(-4.90)	(3.53)
Ch	-0.49	-0.45	-0.46	-0.48	-0.55	-0.60	-0.65	-0.69	-0.81	-1.66	-1.18***
	(-4.55)	(-4.19)	(-4.30)	(-4.27)	(-4.84)	(-5.31)	(-5.98)	(-5.55)	(-6.65)	(-10.36)	(-8.65)
Profit	-1.91	-0.95	-0.66	-0.55	-0.48	-0.46	-0.46	-0.36	-0.46	-0.54	1.37***
	(-12.04)	(-7.95)	(-5.57)	(-5.06)	(-4.43)	(-4.42)	(-4.45)	(-3.42)	(-4.46)	(-4.60)	(14.29)
Disp	-0.46	-0.45	-0.45	-0.42	-0.47	-0.62	-0.61	-0.75	-0.91	-1.13	-0.67***
	(-4.76)	(-4.52)	(-4.32)	(-3.69)	(-4.32)	(-5.61)	(-5.03)	(-6.20)	(-7.10)	(-8.34)	(-7.77)
Ivol	-0.35	-0.34	-0.37	-0.43	-0.51	-0.53	-0.72	-0.80	-1.12	-1.83	-1.48***
	(-4.33)	(-3.75)	(-3.79)	(-4.05)	(-4.83)	(-4.64)	(-6.22)	(-5.79)	(-7.57)	(-11.69)	(-13.04)
Voldev	-2.32	-1.27	-0.93	-0.81	-0.60	-0.44	-0.32	-0.22	-0.08	-0.00	2.31***
	(-20.32)	(-12.17)	(-8.77)	(-8.27)	(-5.37)	(-3.75)	(-2.61)	(-1.58)	(-0.61)	(-0.02)	(14.85)
Vts	-2.26	-1.00	-0.69	-0.61	-0.44	-0.38	-0.34	-0.30	-0.34	-0.53	1.72***
	(-14.67)	(-7.18)	(-5.84)	(-5.87)	(-3.96)	(-3.47)	(-3.27)	(-2.77)	(-3.27)	(-4.78)	(16.80)
BidAsk	-0.21	-0.39	-0.41	-0.54	-0.65	-0.79	-0.86	-0.91	-1.02	-1.06	-0.85***
	(-1.88)	(-3.81)	(-3.57)	(-5.18)	(-5.75)	(-7.02)	(-7.93)	(-8.02)	(-8.33)	(-7.50)	(-10.36)
Credit	-0.18	-0.25	-0.31	-0.49	-0.97						-0.78***
	(-1.79)	(-2.47)	(-3.22)	(-4.38)	(-7.91)						(-10.49)

This table reports summary statistics of delta-hedged option returns (in percentage) sorted by various characteristics. The sample period is January 1996 to December 2015. Size is the natural logarithm of the market value of the firm's equity. Reversal is the lagged one-month return. Mom is the cumulative return on the stock over the 11 months ending at the beginning of the previous month. CH is the Cash-to-assets ratio, defined as the value of corporate cash holdings over the value of the firm's total assets in Palazzo (2012). Profit is calculated as earnings divided by book equity in Fama and French (2006). Disp is the analyst earnings forecast dispersion, computed as the standard deviation of annual earnings-per-share forecasts scaled by the absolute value of the average outstanding forecast in Diether et al. (2002)). Ivol is the annualized stock return idiosyncratic volatility in Ang et al. (2006). Voldev is the log difference between the realized volatility and the Black-Scholes implied volatility for at-the-money options in Goyal and Saretto (2009). Vts is the volatility term structure, defined as the difference between long-term and short-term implied volatility in Vasquez (2017). BidAsk is the difference between the bid and ask quotes of option divided by the midpoint of the bid and ask quotes at the end of the previous month. Credit is the credit ratings provided by Standard & Poor's. The ratings are mapped to 22 numerical values, where 1 corresponds to the highest rating (AAA) and 22 corresponds to the lowest rating (D). We report equal-weighted returns and 10-1 return spread in the table. The Newey-West t-statistics are reported in Parenthesis.

Table 2: Summary Statistics of the Long-short Factors

	Mean	Std. Dev.	10th. Pctl.	25th. Pctl.	50th. Pctl.	75th. Pctl.	90th. Pctl.	Skew	Kurt
LS_{size}	-1.87	1.65	-3.99	-2.85	-1.82	-1.03	0.07	0.01	0.80
$LS_{reversal}$	-0.60	1.57	-2.33	-1.29	-0.68	0.23	1.09	0.33	3.37
LS_{mom}	-0.51	1.95	-2.60	-1.36	-0.35	0.48	1.58	-0.63	3.27
LS_{ch}	-1.18	1.66	-3.00	-2.25	-1.11	-0.29	0.76	0.06	1.04
LS_{profit}	-1.37	1.46	-3.19	-2.23	-1.31	-0.50	0.17	-0.20	1.27
LS_{disp}	-0.67	1.27	-2.00	-1.32	-0.73	-0.10	0.60	-0.27	5.59
LS_{ivol}	-1.48	1.69	-3.33	-2.42	-1.62	-0.40	0.47	0.30	2.76
LS_{voldev}	-2.31	1.64	-4.70	-3.26	-1.96	-1.10	-0.48	-0.69	0.57
LS_{vts}	-1.72	1.50	-3.57	-2.56	-1.71	-0.91	-0.21	0.29	2.54
LS_{bidask}	-0.85	1.26	-2.37	-1.58	-0.91	-0.21	0.68	-0.18	2.93
LS_{credit}	-1.03	1.47	-2.81	-1.90	-1.07	-0.23	0.63	-0.08	2.77
DH_{idx}	-0.16	0.36	-0.47	-0.38	-0.22	0.02	0.32	1.08	3.84
DH_{stk}	-0.06	0.19	-0.22	-0.16	-0.08	0.03	0.13	-0.90	16.19

This table reports summary statistics of the returns on long-short portfolios (in percentage) that go long in stock options with high (low) values for a certain characteristic and short stock options with low (high) values. To make the average sign of the factors returns consistent with the negative market straddle return, which represents variance risk premium, we go long in options with low values and short with high values for size, reversal, momentum, profitability, volatility deviation and slope of the volatility term structure, such that all factor returns are on average negative. The sample period is from January 1996 to December 2015.

Table 3: Correlation Matrix of the Candidate Factors in the Equity Option Market

	LS_{size}	$LS_{reversal}$	LS_{mom}	LS_{ch}	LS_{profit}	LS_{disp}	LS_{ivol}	LS_{voldev}	LS_{vts}	LS_{bidask}	LS_{credit}	DH_{idx}	DH_{stk}
LS_{size}	1.00												
$LS_{reversal}$	0.17	1.00											
LS_{mom}	0.33	0.07	1.00										
LS_{ch}	0.28	0.03	-0.22	1.00									
LS_{profit}	0.51	0.06	0.10	0.50	1.00								
LS_{disp}	0.26	0.04	0.12	0.31	0.43	1.00							
LS_{ivol}	0.49	0.20	-0.02	0.53	0.57	0.50	1.00						
LS_{voldev}	0.13	-0.01	0.14	0.05	0.08	-0.01	-0.21	1.00					
LS_{vts}	0.15	0.33	0.03	0.20	0.22	0.30	0.40	0.16	1.00				
LS_{bidask}	0.50	0.13	0.18	-0.06	0.19	0.00	0.03	-0.01	0.00	1.00			
LS_{credit}	0.47	0.23	0.16	0.23	0.36	0.37	0.53	0.06	0.35	0.16	1.00		
DH_{idx}	0.04	0.08	-0.07	0.15	0.11	0.16	0.19	-0.07	0.09	0.01	0.09	1.00	
DH_{stk}	0.03	0.11	-0.15	0.24	0.13	0.22	0.27	-0.03	0.20	-0.09	0.12	0.55	1.00

This table reports correlation matrix of the 13 candidate factors in the equity option market. The first 10 factors are 10-1 return spread sorted by size, reversal, momentum, cash holding, profitability, analyst forecast dispersion, idiosyncratic volatility, deviation of log realized volatility from log implied volatility, volatility term structure and bid-ask spread. The 11th factor is the 5-1 return spread sorted by credit rating. The last two factors are the delta-hedged return of the S&P 500 index options (DH_{idx}) and average delta-hedged return from the stock options that are components in S&P500 DH_{stk}. Sample period is from January 1996 to December 2015.

Table 4: Estimation for the Number of Factors in the Delta-hedged Option Portfolios

	Demeaned Data	Raw Data
Eigenvalue Ratio (ER) estimator in Ahn and Horenstein (2013)	1	NA
Growth Ratio (GR) estimator in Ahn and Horenstein (2013)	1	NA
Edge Distribution (ED) estimator in Onatski (2010)	1	2
Modified Bayesian information criterion (BIC3) estimator in Bai and Ng (2002)	3	4
Information Criterion (IC1) estimator in Bai and Ng (2002)	6	7
Modified Information Criterion estimator (ABC) in Alessi et al. (2010)	6	7

This table presents results obtained from estimating the number of factors using the Eigenvalue Ratio (ER) and Growth Ratio (GR) estimators of Ahn and Horenstein (2013), the Edge Distribution (ED) estimator of Onatski (2010), the BIC3 and IC1 estimators of Bai and Ng (2002), and the Modified Information Criterion estimator (ABC) of Alessi et al. (2010). The test assets are 105 characteristic-sorted delta-hedged option portfolios reported in Table 1. The sample period is January 1996 to December 2015. We apply the estimators to doubly-demeaned data (Column 1) and to the raw data (Column 2). ER and GR are applied only to doubly-demeaned data.

Table 5: Performance of the Model with 6 Latent Factors and the Model with 13 Candidate Factors

Panel A: Model with 6 Latent Factors		Panel B: Model with 13 Candidate Factors	
Non-zero Alphas	0.238	Non-zero Alphas	0.295
Average Adj. R^2	0.318	Average Adj. R^2	0.303
Corr E(R)-Predicted Return	0.967	Corr E(R)-Predicted Return	0.963
Average Abs. Alpha (%)	1.010	Average Abs. Alpha (%)	1.070
Corr E(R) and Beta of Factor 1	-0.96	Corr E(R) and Beta of LS_{size}	-0.475
Corr E(R) and Beta of Factor 2	0.143	Corr E(R) and Beta of $LS_{reversal}$	-0.154
Corr E(R) and Beta of Factor 3	0.048	Corr E(R) and Beta of LS_{mom}	-0.046
Corr E(R) and Beta of Factor 4	-0.008	Corr E(R) and Beta of LS_{ch}	-0.344
Corr E(R) and Beta of Factor 5	0.009	Corr E(R) and Beta of LS_{profit}	-0.295
Corr E(R) and Beta of Factor 6	-0.005	Corr E(R) and Beta of LS_{disp}	-0.23
		Corr E(R) and Beta of LS_{ivol}	-0.373
		Corr E(R) and Beta of LS_{voldev}	-0.429
		Corr E(R) and Beta of LS_{vts}	-0.38
		Corr E(R) and Beta of LS_{bidask}	-0.077
		Corr E(R) and Beta of LS_{credit}	-0.216
		Corr E(R) and Beta of STR_{idx}	-0.01
		Corr E(R) and Beta of STR_{stk}	0.19

Panel C: Increase the number of factors from 1 to 6

	1 Factor	2 Factors	3 Factors	4 Factors	5 Factors	6 Factors
Non-zero alphas (5%)	0.543	0.219	0.181	0.21	0.257	0.238
Average Adj R^2	0.138	0.195	0.232	0.26	0.288	0.318
Average Abs. Alpha (%) 2.038	1.254	1.048	1.033	1.008	1.010	

This table shows the performance of the proposed model with 6 latent factors in Panel A and the performance of a model with 13 candidate factors in Panel B. In Panel A and B, we report the percentage of pricing errors statistically different than 5% generated by each model, the average adjusted R^2 across option portfolios each model generates, the annualized average absolute alpha in percentage, the correlation between the portfolios expected (excess) returns and the betas of each factor, the correlation between the portfolios expected (excess) returns and the models' predicted returns (betas times factor premiums). Panel C presents the performance metrics of 6 factor models, in which we increase the number of factors used as independent variables sequentially from 1 to 6. We report the number of pricing errors statistically significant at the 5% or less, the average adjusted R^2 generated by the model, and the annualized average absolute alpha in percentage.

Table 6: Correlation Coefficients between the 13 Candidate Factors and Six Latent Factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	EWP
LS_{size}	0.71	0.02	0.48	-0.25	-0.24	-0.15	0.25
$LS_{reversal}$	0.28	0.11	0.13	0.64	-0.25	-0.09	0.24
LS_{mom}	0.13	-0.43	0.72	0.01	0.35	-0.01	-0.22
LS_{ch}	0.64	0.60	-0.23	-0.36	-0.06	0.02	0.33
LS_{profit}	0.71	0.31	0.16	-0.31	0.03	-0.23	0.30
LS_{disp}	0.56	0.39	0.23	-0.02	0.40	0.01	0.37
LS_{ivol}	0.80	0.72	0.25	-0.04	0.10	-0.15	0.56
LS_{voldev}	0.20	-0.50	-0.34	0.00	0.23	0.27	-0.18
LS_{vts}	0.53	0.27	-0.01	0.54	0.13	0.13	0.34
LS_{bidask}	0.20	-0.23	0.33	-0.06	-0.56	-0.21	0.10
LS_{credit}	0.63	0.28	0.33	0.13	-0.01	0.19	0.36
DH_{idx}	0.14	0.19	0.04	0.04	-0.02	0.00	0.38
DH_{stk}	0.21	0.30	-0.04	0.08	0.03	0.00	0.50

This table shows the Pearson correlation coefficients between the 13 candidate factors and the six latent factors. The 13 candidate factors are described in Table 1. The sample period is from January 1996 to December 2015.

Table 7: Canonical Correlation

	6 Latent Factors vs 13 Candidate Factors	3 Latent Factors vs 13 Candidate Factors	6 Latent Factors vs 6 Candidate Factors	3 Latent Factors vs 6 Candidate Factors
1	0.99	0.98	0.98	0.96
2	0.98	0.96	0.93	0.90
3	0.95	0.89	0.85	0.75
4	0.91		0.69	
5	0.91		0.42	
6	0.42		0.22	

This table shows the canonical correlations for different pairs of factors. The first column shows the canonical correlation corresponding to the 6 latent factor and the 13 candidate factors. In the second column we compare the 3 most relevant latent factors with the 13 candidate factors. Columns three and four compare the latent factors with the subset of the 6 candidate factors we find most relevant with our rank estimation exercise, which are LS_{size}, LS_{ch}, LS_{disp}, LS_{ivol}, LS_{voldev}, and LS_{rating} A brief explanation on the canonical correlation is in Appendix A.2.

Table 8: Performance of the Models with 4 Factors and 7 Factors

(105 Characteristic-sorted Portfolios)

Panel A: Model with 4 Candidate Factors		Panel B: Model with 7 Candidate Factors	
Non-zero Alphas	0.05	Non-zero Alphas	0.07
Average Adj. R^2	0.34	Average Adj. R^2	0.35
Corr E(R)-Predicted Return	0.95	Corr E(R)-Predicted Return	0.95
Average Abs. Alpha (%)	1.220	Average Abs. Alpha (%)	1.210
Corr E(R) and Beta of DH_{idx}	-0.10	Corr E(R) and Beta of DH_{idx}	0.06
Corr E(R) and Beta of LS_{size}	-0.50	Corr E(R) and Beta of LS_{size}	-0.47
Corr E(R) and Beta of LS_{ivol}	-0.61	Corr E(R) and Beta of LS_{ivol}	-0.47
Corr E(R) and Beta of LS_{voldev}	-0.63	Corr E(R) and Beta of LS_{voldev}	-0.52
		Corr E(R) and Beta of LS_{ch}	-0.32
		Corr E(R) and Beta of LS_{disp}	-0.28
		Corr E(R) and Beta of LS_{credit}	-0.30

This table shows the performance of the model with 4 option factors DH_{idx}, LS_{size}, LS_{ivol}, LS_{voldev} in Panel A and the performance of the model augmented with LS_{ch}, LS_{disp}, and LS_{credit} in Panel B. The test assets are the 105 characteristic-sorted portfolios. We report the percentage of pricing errors statistically different than 5% generated by each model, the average Adjusted R^2 across option portfolios each model generates, the correlation between the portfolios expected (excess) returns and the betas of each factor, the correlation between the portfolios expected (excess) returns and the models' predicted returns (betas times factor premiums).

Table 9: Fama-Macbeth Regressions for the 105 Delta-hedged Call Option Portfolios

	Intercept	LS_{size}	LS_{ivol}	LS_{voldev}	Adjusted R^2
Full sample	0.49	-2.11	-1.35	-2.31	0.90
	(9.19)	(-13.47)	(-10.66)	(-9.16)	
Jan. 1996 - Dec. 2005	-0.00	-2.35	-1.38	-2.39	0.84
	(-0.00)	(-18.63)	(-9.12)	(-9.73)	
Jan. 2006 - Dec. 2015	-0.00	-1.59	-1.23	-2.10	0.84
	(-0.01)	(-14.25)	(-10.28)	(-13.14)	

This table reports Fama-MacBeth regression results for the 105 portfolios for the full sample from January 1996 to December 2015 and for the two sub-samples from January 1996 to December 2005 and from January 2006 to December 2015. We first run the time-series regression of the return of the 105 portfolios on the three factors and get the betas. We then run the cross-section regression of average return of the 105 portfolios on the estimated betas from the first step. The table reports regression coefficients, adjusted R^2 and the t-statistics from the second step of regression. The standard errors are corrected by Shanken (1992).

Table 10: Pricing Performance of the Models with four Factors

(Four Sets of Test Assets)

	105 Call Portfolios	26 Industry Call Portfolios	105 Put Portfolios	105 VRP Portfolios
Non-zero alphas	4.80%	7.69%	6.67%	10.47%
Average Adj. R^2	0.338	0.222	0.301	0.256
Corr E(R)-Predicted Return	0.946	0.825	0.866	0.874
Average Abs. Alpha (%)	1.220	1.951	1.458	8.697
Corr E(R) and Beta of STR_{idx}	-0.100	0.290	-0.287	-0.318
Corr E(R) and Beta of LS_{size}	-0.503	0.097	-0.317	-0.525
Corr E(R) and Beta of LS_{ivol}	-0.605	-0.700	-0.594	-0.532
Corr E(R) and Beta of LS_{voldev}	-0.627	-0.737	-0.244	0.065

This table shows the performance of the model with four option factors DH_{idx}, LS_{size}, LS_{ivol}, and LS_{voldev}. The test assets are the 105 characteristic-sorted delta-hedged call option portfolios, 26 industry delta-hedged call option portfolios, 105 characteristic-sorted delta-hedged put option portfolios, and 105 characteristic-sorted VRP portfolios. We report the percentage of pricing errors statistically different than 5% generated by each model, the average adjusted R^2 across option portfolios, the correlation between the portfolios expected (excess) returns and the models' predicted returns (betas times factor premiums), the annualized average absolute alpha in percentage, and the correlation between the portfolios expected (excess) returns and the betas of each factor.

Table A1: Values of the IB Estimator for Two Factor Models

Panel (a)					
		EWP	LS_{size}	LS_{ivol}	LS_{voldev}
Average Adj R^2	0.87				
Corr E(R)-Beta		-0.10	0.50	0.61	0.63
IB		0.995	0.00	0.00	0.00
Corr E(R)-Predicted Return	0.95				
Panel (b)					
		DH_{idx}	LS_{size}	LS_{ivol}	LS_{voldev}
Average Adj R^2	0.34				
Corr E(R)-Beta		0.10	0.50	0.61	0.63
IB		0.991	0.00	0.87	0.19
Corr E(R)-Predicted Return	0.95				

This table shows the values of the IB estimator proposed in Ahn and Horenstein (2013) for two factor models. Panel (a) shows the results from regressing the 105 delta-hedged option returns on a model consisting of the EWP (equal-weighted portfolio) and the three option factors proposed in the main body of the paper (Size, ivol, and voldev) and Panel (b) show the results from a similar factor model in which EWP has been replaced by the DH_{idx} factor (delta-hedged return of the S&P 500 index options). The table also reports the average value of the factor loadings as well as the R^2 of the models and the correlation between the predicted returns and the average returns of the models.