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

February 1, 2019

Abstract

This paper studies the factor structure of the cross-section of delta-hedged equity option returns. Using latent factor techniques, we find strong evidence supporting the existence of a factor structure in equity options returns. We find that a four-factor model captures relevant latent factors and explains the time series and the cross-section of equity option returns. The factors are the market volatility risk factor and three characteristic-based factors related to firm size, idiosyncratic volatility, and the difference between implied and historical volatilities. 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 factor that drives the cross-section of variance risk premiums. This is the main goal of our study.

Using a multifactor stochastic volatility model, we show that the expected delta-hedged equity option returns are driven by the volatility risk of the priced factors. Since no options are traded on the stock return factors, 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. We also include two additional market factors: the delta-hedged return of the S&P 500 index option and the value-weighted delta-hedged return on the individual stocks that are components of the S&P 500 index. We consider 13 candidate factors in total.

Our identification procedure follows the factor identification protocol suggested by Pukthuanthong et al. (2018). We first employ a latent variable analysis of the covariance matrix of the delta-hedged option returns to understand the factor structure for the cross-section of option returns. Latent variables are not directly observed but are econometrically inferred

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

from observed variables. We estimate the number of latent factors using six identification methods: 1) the Eigenvalue Ratio (ER) and 2) the Growth Ratio (GR) estimators by Ahn and Horenstein (2013), 3) the Edge Distribution (ED) estimator of Onatski (2010), 4) the BIC3 and 5) IC1 estimators of Bai and Ng (2002), and 6) the Modified Information Criterion estimator (ABC) of Alessi et al. (2010). We find that there is one strong factor and possibly up to five weak factors that can explain the cross-section of delta-hedged option returns.

Next we estimate the common latent factors using principal component analysis (PCA) as suggested in Connor and Korajczyk (1986) and study how well they explain the cross-section of delta-hedged option returns. We find that the six latent factors explain on average 31.8% of the time series variation of the 105 portfolios. The correlation between the average return of the 105 portfolios and the model's predicted return is 96%.

Latent factors estimated with PCA are difficult to interpret economically since they are not directly observable. To circumvent this problem, we test if the 13 (observed) candidate factors are related to the covariance matrix of delta-hedged option returns. Using the rank-estimation method suggested by Ahn et al. (2018) on the 13 candidate factors, we 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 size, cash holding, analyst dispersion, idiosyncratic volatility, volatility deviation, and credit rating. These six factors contain most of the relevant information on the cross-section of delta-hedged option returns. Canonical correlation analysis further confirms this result. Linear combinations of the six characteristic-based factor candidates are highly correlated with linear combinations of the six latent factors estimated from PCA.

Next, we find that three out of the six factor candidates suffice to explain the cross-section of delta-hedged option returns: size, idiosyncratic volatility, and volatility deviation. The correlation between the average return of the 105 portfolios and the three factor's predicted returns is 0.95. The added explanatory power of the three remaining factors: cash holding, analyst dispersion, and credit rating is negligible. The fourth factor we propose is the market volatility risk factor. This factor's loadings are almost constant and as such it is useful to explain the time-series of option returns but not the cross-section.[2]

[2] Fama and French (1992) and more recently Ahn and Horenstein (2018) show that the stock returns'

Given the empirical results, we propose a four-factor model that captures the time-series and cross-sectional variation in delta-hedged option returns. The factors include the delta-hedged return of the S&P 500 index options, size, idiosyncratic volatility, and volatility deviation. The market volatility risk factor mainly explains the time-series variation of the delta-hedged option returns. The latter three factors capture the cross-sectional comovements in option returns. Using the 105 option portfolios as test assets in Fama-MacBeth regressions, we find that the exposures to the three factors are statistically significant in explaining average delta-hedged returns of the 105 portfolios. The adjusted R^2 ranges from 84% to 90% when using the full sample or subsamples. To address the critique in Lewellen et al. (2010), we also test the proposed model on a set of delta-hedge option portfolios constructed based on industrial classification instead of firm characteristics and find that the model explains their cross-section quite well.

Lastly, we investigate how much information about the cross-section of delta-hedged option returns are contained in commonly used factors to price the cross-section of stock returns. We find that stock factors have little to no explanatory power over the 6 latent factors necessary to explain the cross-section of delta-hedged option returns. The only stock factor that is sometimes significant at the 1% or less when regressed onto option latent factors is the BAB factor (for the case of the 2nd and 3rd most important factors). As expected, the information contained in stock factors seems not relevant for pricing the cross-section of delta-hedged option returns.

This paper differs from the existing literature in option factors in two important ways. First, different from Christoffersen et al. (2017a) who study the factor structure of equity volatility levels, skews, and term structures, we focus on the factor structure of delta-hedged equity option returns, which reflect the risk premium required by bearing the unfavorable variance risk. While the implications of their paper are mainly on equity option valuation, our analysis uncovers the factor structure of equity option returns and the embedded variance risk premium. Second, in the spirit of Ahn et al. (2018) and Feng et al. (2017), this paper

Market portfolio shows little to no variability in factor loadings and is mostly useful to explain the time-series dimension of stock returns. We find an analogous relationship between the market volatility risk factor and option returns. The market volatility risk factor explains the time-series but not the cross-section of option returns attempts to avoid the proliferation of option returns' factors capturing similar information. 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 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 that best explain the factor structure in option returns, and in Section 4.3 we study the common information between factors in stock returns and factor in option returns. We conclude in Section 5.

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

In this section, we show expected delta-hedged equity option gains in a multi-factor framework, in which stock return and variance are driven by multiple factors. The results show that the 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 with 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 that 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, variance of the SMB factor and 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:

$$C^i_{t+\tau} - C^i_t = \int_t^{t+\tau} \Delta^i_u dS^i_u + \int_t^{t+\tau} r(C^i - S_i \frac{\partial C^i}{\partial S_i})du +$$
$$\int_t^{t+\tau} [\sum_{j=1}^n \beta^j \lambda^j (V^j_{f,t})] \frac{\partial C^i}{\partial V^i} du + \int_t^{t+\tau} [\sum_{j=1}^n \beta^j \theta^j (V^j_{f,t}) \frac{\partial C^i}{\partial V^i} dW_2^{i,j}]. \quad (3)$$

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

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

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

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

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 exposure to the 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 for the market index S^m with stochastic volatility,

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

and a call option written on the market index is C^m_t, according to Bakshi and Kapadia

(2003), the expected delta-hedged gain for the index option is:

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

It follows that, if $\frac{\partial C^m}{\partial V^m}$ and $\frac{\partial C^i}{\partial V^i}$ are at the similar level, we can use the delta-hedged gain of the index options to replicate the price of volatility risk of the index return. However, there are no traded options on other stock return factors. In this paper, following the literature on stock return factors, we consider some 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 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 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 in calculating 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 initial investment: $O_t - \Delta_t S_t$. 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 cross-section of stock returns, long-short factors are commonly used to describe stock returns. These factors are constructed with portfolios composed by ranked stocks by certain characteristics, such as the size factor, the value factor and the momentum factor. In the equity option market, following the similar logic, we consider the following predictors which has been shown to have strong predictability in the literature. The predictors are then used to sort portfolios and construct factors.

We consider six equity characteristics in Cao et al. (2017), which have been found to be significant predictors of delta-hedged equity option return in the next month. The equity characteristics are as follows, with notation of the long-short factors in the brackets.

(1) Size (LS_{size}): The natural logarithm of the market value of the firm's equity (e.g.Banz (1981) and Fama and French (1992)).

(2) Stock return reversal ($LS_{reversal}$): The lagged one-month return. (Jegadeesh (1990))

(3) Stock return momentum (LS_{mom}): The cumulative return on the stock over the 11 months ending at the beginning of the previous month (Jegadeesh and Titman (1993))

(4)CH (LS_{ch}): Cash-to-assets ratio, as in Palazzo (2012), 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 Fama and French (2006).

(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 (Diether et al. (2002)). 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 deciles 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 beginning of the next month and hold the position until the end of that 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 detected. 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 the 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 below shows the summary statistics for the average returns of the decile (quintile) delta-hedged option portfolio 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]

To conclude this section, Table 3 shows the correlation coefficients between the option returns predictors. The table shows that the correlation coefficients between the strategies are mostly below 0.5 with the exception that the correlation between LS_{disp} and LS_{ivol} is 0.57, the correlation between LS_{disp} and LS_{profit} is 0.53 and the correlation between LS_{voldev} and LS_{credit} is 0.53. The low correlation among the long-short factors suggests that these variables might capture different information about the cross-section of delta-hedged option returns. How many different factors do these 13 factors capture? How many of them are relevant for explaining cross-section of option returns? Do they capture similar information to stock returns factors? We answer these questions in the next section.

[Table 3 around here]

4 Empirical Procedure and Results

Following the suggestions in Pukthuanthong et al. (2018), we first perform a latent variable analysis of the covariance matrix of delta-hedged option returns to uncover its factor structure. For that purpose, we test the number of common factors in delta-hedged option returns. Then we estimate those factors using principle component analysis (PCA) as suggested in Connor and Korajczyk (1986) and study how well they explain the cross-section of delta-hedged 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 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 premium 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 to be a factor is that it should be related to the principal components of the covariance matrix. In this section, we aim to identify factors that drive option return systematically.

In Section 2, we show that, under a stochastic volatility model, the delta-hedged return of an equity option portfolio is consistent with a linear factor model if the stock variance follows a multi-factor structure. In this section we use several techniques developed to estimate the number of factors in approximate linear factor models 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,

$$x_{it} = \alpha + Bf_t + \epsilon_t,$$

where $x_{it} = (x_{1t}, ..., x_{Kt})'$; α is the N-vector of individual intercepts; B is the $N \times K$ matrix of factor loadings (beta matrix); and ϵ_t is the N-vector of idiosyncratic components of individual returns at time t. The entries in ϵ_t can be cross-sectionally correlated. We denote the ith row of α and B by α_i and $\beta_t = (\beta_{1i}, ..., \beta_{Ki})'$, respectively.

To estimate the number of factors K in delta-hedged option returns, we use as response variables data on all the delta-hedged portfolio 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. Therefore, we also estimate the number of factors

[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

[4] Let x_{it} be the VW-excess return on asset i. Then, the "doubly demeaned" excess return equals $x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}$, where $\bar{x}_i = \sum_{t=1}^{T} x_{it}/T$, $\bar{x}_t = \sum_{i=1}^{N} x_{it}/N$, and $\bar{x} = \sum_{i=1}^{N} x_i/N$. See Ahn and Horenstein (2013) for the justification of using these doubly demeaned excess returns for estimation of the number of factors.

using several consistent methods. More precisely, in Table 4 we present the 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). A brief explanation about these methods is provided in Appendix A.1. We apply the estimators to doubly-demeaned data (Column 1) and to the raw data (Column 2) in Table 4. ER and GR have been applied only to doubly-demeaned data.

[Table 4 around here]

First, 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. While ER, GR, and ED capture 1 common factor, BIC3 captures 3 common factors, IC1 captures 6 common factors as well as ABC[5]. The factor structure is consistent with having 1 strong factor and possibly up to 5 weak factors. Second, all estimators capture an additional factor when we use raw data. This is consistent with the existence of an additional factor having constant factor loadings. More precisely, this factor with constant factor loadings, corresponding to the first principal component (PC) extracted from raw data, is the equally-weighted portfolio (EWP) of the response variables (correlation between the first PC extracted from raw data and EWP is 0.99). This result has led to many 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). While this factor has good power to explain the average time series variation in returns, given that it produces constant betas, it can barely explain the cross-section of returns. 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 and use as response variables

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

for testing assets pricing models the excess returns over EWP or any other well-diversified portfolios. In the Appendix A.3, we show that EWP contains unitary loadings. Therefore, our base latent model consists of the following equation:

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

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-portfolios, f_k corresponds to the kth PC factor[6] (where $k = 1, \ldots, 6$), β_{ik} 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 purpose, we present on Panel (b) the performance of a model using the 13 candidate factors. The performance metrics we analyze are the percentage of pricing errors statistically different from 5% generated by each model, the average adjusted R^2 across option portfolios generated by each model, 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).

[Table 5 around here]

The performance metrics of the model with latent variables improve over those of a model using the 13 candidate factors used to predict option returns. It produces less pricing errors 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 of delta-hedged option returns. The correlation between expected returns and predicted returns is quite high

[6]The kth PC factor is the one corresponding to the kth largest eigenvalue.

for both models: 0.97 for the model with 6 latent variables and 0.96 for the model with 13 candidate factors. What is striking from these results is the correlation between expected returns and the beta of the first latent factor, which amounts to 0.96. Figure 2 plots the scatter diagram between average returns and the betas of the first latent factor, showing a clear linear relationship between the two variables.

[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. We now analyze how many, if any, of those weak factors are necessary to improve the model performance.

To answer these question, Panel (c) Table 5 shows the performance metrics of 6 factor models, in which we increase the number of factors used as independent variables sequentially from 1 to 6. This panel contains 3 metrics: (i) the number of pricing errors 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 show the correlation between expected returns and realized returns because we already know that in this case the first factor generates a correlation 0.96 of the total of 0.97 generated by all the six latent factors.

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 is obtained with 3 factors at 18% of pricing errors different than zero. According to this metric, additional factors beyond the third do not add relevant pricing information. The average adjusted R^2 metric shows that the first factor explains an average of 14% of the returns 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, this table suggests that at most three of the latent factors suffice to have the best fit to price our set of delta-hedged option returns.

It is important to note that we work with "doubly-demeaned" delta-hedged returns. This implies that the factors we analyze are the ones that do not have constant loadings and are useful for explaining the cross-section of delta-hedged returns. As previously discussed, and further developed in Appendix A.3, EWP, which corresponds to the first PC estimated from raw data, is an additional factor that has explanatory power for the time series of returns but not over their cross-section since it has unitary loadings. We use this factor later on to add the market portfolio to the proposed benchmark pricing model.

Finally, the problem with statistical and latent factors is that it is hard to understand them with economic interpretation. As such, it is important to study which factors in the 13 candidate factors are important to explain the latent factors, especially the first one. At the same time, given that there are at most 6 common factors in total, and at most 3 common factors relevant for pricing, many of the 13 variables proposed as predictors of option returns might contain redundant information. Hence, it is important to separate the most relevant variables for pricing, which is the focus of the next section.

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 used to 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 below 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. As in the stock returns case, it seems that the variable used to capture the market factor has unitary loadings (e.g. Ahn et al. (2018)). This means that the delta-hedged index option returns might be good for capturing the average time-series variation in delta-hedged option returns but not the cross-sectional variation. We further study this result in Appendix A.3.

Since many variables seem to be capturing similar information, we would like to know the minimum set of variables necessary to capture the 6 common latent factors. We use rank estimation methods to answer this question. More precisely, we generate all combinations of different sets containing 6 factors from the 13 candidate factors and check if any set generates a full-rank beta matrix. From the set of 13 candidate factors, we can create 462 different sets of 6 factors. It turns out that only 1 of these 462 sets generates a full rank beta matrix when using the returns of the delta-hedged portfolios 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}. This means that these six factors suffice to capture most (if not all) the relevant information about the cross-section of delta-hedged option returns. Figure 3 shows the relationship between the average returns of the option portfolios and the predicted returns by a model containing the six relevant factors. The two variables have a correlation coefficient of 0.95. As shown in Panel (B) of Table 5, this correlation increases to 0.96 if we use all of the 13 candidate factors. Overall, these six variables contain all relevant information about the cross-section of delta-hedged option returns.

[Figure 3 around here]

Previous results show that the six latent factors are captured by the six factor candidates. However, we do not know how well the candidate factors capture the latent ones. To study this, following Pukthuanthong et al. (2018) we use canonical correlation analysis. A brief explanation on this procedure is on Appendix A.2. Table 7 below shows the canonical correlations for different pairs of variables.

[Table 7 around here]

The first column shows the canonical correlation corresponding using the 6 latent factors and the 13 factor candidates. It shows that 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, we do not know which dimension is missed by just looking at the canonical correlation coefficients. 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, meaning that the 13 factor candidates contain the most important 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 also capture almost perfectly 3 of the latent factors, and that those latent factors are the 3 most relevant ones. This further confirms our rank estimation exercise.

We next select the most relevant factors among the six relevant factor candidates, which are able to capture the most relevant three latent factors. For this purpose, we rank by importance the six variables which are sufficient to explain the cross-section of delta-hedged option returns. We pick the most relevant variable as the one that produces by itself the largest correlation coefficient between beta and average return. The second variable in our rank is the one that increases the correlation between average return and predicted returns the most in a model already containing the first variable as regressor. We find that we only need 3 out of the 6 variables to have a correlation between average returns and predicted returns of 0.95. LS_{size} is the variable that produces the highest correlation, 0.83. LS_{ivol} increases that correlation further to 0.89 and LS_{voldev} increases it to 0.95. This does not mean that the other variables are uninformative, however, the added explanatory power to the cross-section of option returns by LS_{ch}, LS_{disp1}, and LS_{rating} 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 (0.93 and 0.89), arguably the most important ones. The third latent factor is explained with lesser accuracy (canonical correlation is 0.58).

Overall, all our results are consistent with our previous estimation of 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 construct 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 comovement 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 such as in Goyal and Saretto (2009) and Cao and Han (2013). Thus, we propose the following four-factor model for pricing the delta-hedged equity option returns[7]:

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

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 performance of the four-factor model does not increase when

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

augmented by LS_{ch}, LS_{disp}, and LS_{rating}. We conclude that the four-factor model captures all the necessary information to price the delta-hedged option returns. However, the factor structure might change through time and it is important to evaluate the performance of the proposed factor model in different sub-samples. For this purpose, we run monthly rolling regressions using 60-month 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 0.86, with a standard deviation of 0.04, a maximum value of 0.93 and a minimum of 0.72. Figure 4 below shows the value of this parameter through time. This figure shows that the predictive power of the four factor model we propose is quite stable throughout the time span analyzed.

[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 present 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 two subsamples. As such, the factors in our model passes all the criteria in the protocol established by Pukthuanthong et al. (2018) to find relevant factors.

[Table 9 around here]

As a further robustness check, we evaluate the performance of the four-factor and the seven-factor models for industry portfolios. We construct the industry portfolios by categorizing each firm into the industry groups with two-digit code provided by OptionMetrics. After removing portfolios with missing data, we have 26 industry portfolios in total. We report performance matrix for the industry portfolios the same as for the 105 characteristic-sorted portfolios in Table 10.

[Table 10 around here]

Note that only 2 of the 26 industry portfolios have non-zero alphas in the four-factor model and the correlation between expected returns and realized returns is 0.83. Again, the model with four factors works as well as the model with seven factors. The main difference between using industry portfolios and characteristic-based portfolios is that LS_{size} does not seem to be relevant for industry portfolios while LS_{ch} does. However, LS_{ch} does not add additional explanation power once we use the four factor model with LS_{size}, LS_{ivol}, and LS_{voldev}.

4.3 Are stock factors important in explaining the cross section of option returns?

We now analyze how much information about the cross-section of delta-hedged option returns are contained in commonly used factors to price the cross-section of stock returns. We include the following stock market factors as potential factors to explain the variation in equity option returns: the five factors in Fama and French (2016) (MKT, SMB, HML, RMW, CMA), momentum factor in Carhart (1997) (MOM), stock market liquidity risk factors in Pástor and Stambaugh (2003) (PS_{innov}, PS_{level}, PS_{vwf}), betting-against-beta factor in Frazzini and Pedersen (2014) (BAB), two mispricing factors in Stambaugh and Yuan (2016) (MGMT and PERF).

Table 11 reports the correlation coefficients between the 12 stock factors and the estimated 6 latent option factors plus EWP. The table shows that stock factors are not highly correlated with the latent option factors. The first PC factor, which is the most important one, reports correlations below (absolute) 19% with the stock factors.

[Table 11 around here]

To further study the relationship between stock return factors and option return factors, we regress each latent option factor onto all the stock factors. Table 12 reports the regression results. The table shows that the market factor from stock returns is highly significant for explaining the variation of EWP. This suggests that the factor with unitary loading in stocks is related to that of option returns, which possibly comes from the leverage effect that stock return and stock volatility are correlated.

[Table 12 around here]

Stock factors explain around 32% of the variation of EWP, but very little of that of the 6 latent factors necessary to explain the cross-section of delta-hedged option returns. The only stock factor that is sometimes significant at the 1% or less when regressed onto option PC factors is the BAB factor (for the case of the 2nd and 3rd factor). As expected, the information contained in stock factors seems negligible when pricing the cross-section of delta-hedged 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. 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-section variation. Using the 11 characteristic-based option factors as candidate factors, we find 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.

Using these three factors in the Fama-MacBeth regression, we find that the exposures to the three factors are statistically significant in explaining average delta-hedged return of the 105 portfolios in full sample and two subsamples with adjusted R^2 higher than 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 does not add explanatory power to the cross-section of option returns but it improves the explanatory power of the model over the time-series dimension. As expected from the theoretical model, stock factors seems negligible to price the cross-section of delta-hedged option returns.

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 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.

Rolf W Banz. The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18, 1981.

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

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

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.

Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52 (1):57–82, 1997.

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.

Eugene F Fama and Kenneth R French. Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29(1):69–103, 2016.

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.

Andrea Frazzini and Lasse Heje Pedersen. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.

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.

Narasimhan Jegadeesh. Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3):881–898, 1990.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993.

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.

L'uboš Pástor and Robert F Stambaugh. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685, 2003.

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.

Robert F Stambaugh and Yu Yuan. Mispricing factors. *The Review of Financial Studies*, 30 (4):1270–1315, 2016.

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 $k < r$ 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 is normalized so that $(\tilde{F}^k)'(\tilde{F}^k)/T = I_k$. 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:

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

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

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 the two BN criteria can be written as

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

$$IC(k) = ln(\sum_{j=k+1}^{T}\tilde{\mu}_{NT,j}) + kg(N,T).$$

Let $\tilde{k}_{PC} = min_{k\leq kmax}PC(k)$ and $\tilde{k}_{IC} = min_{k\leq kmax}IC(k)$, 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 estimator 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 estimators 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 $V(k)$ at 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 $k < kmax$. 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 intrinsic characteristics of the data. In addition, there are infinitely many consistent threshold functions that could be used in Bai and Ng (2002).

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 in 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 divide information from noise, with the known property that eigenvalues corresponding to eigenvectors having common information do not vanish as the dimension of the panel increases (N,T) 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 $\Sigma_X^{-1/2}\Sigma_{XF}\Sigma_F^{-1}\Sigma_{FX}\Sigma_X^{-1/2}$.

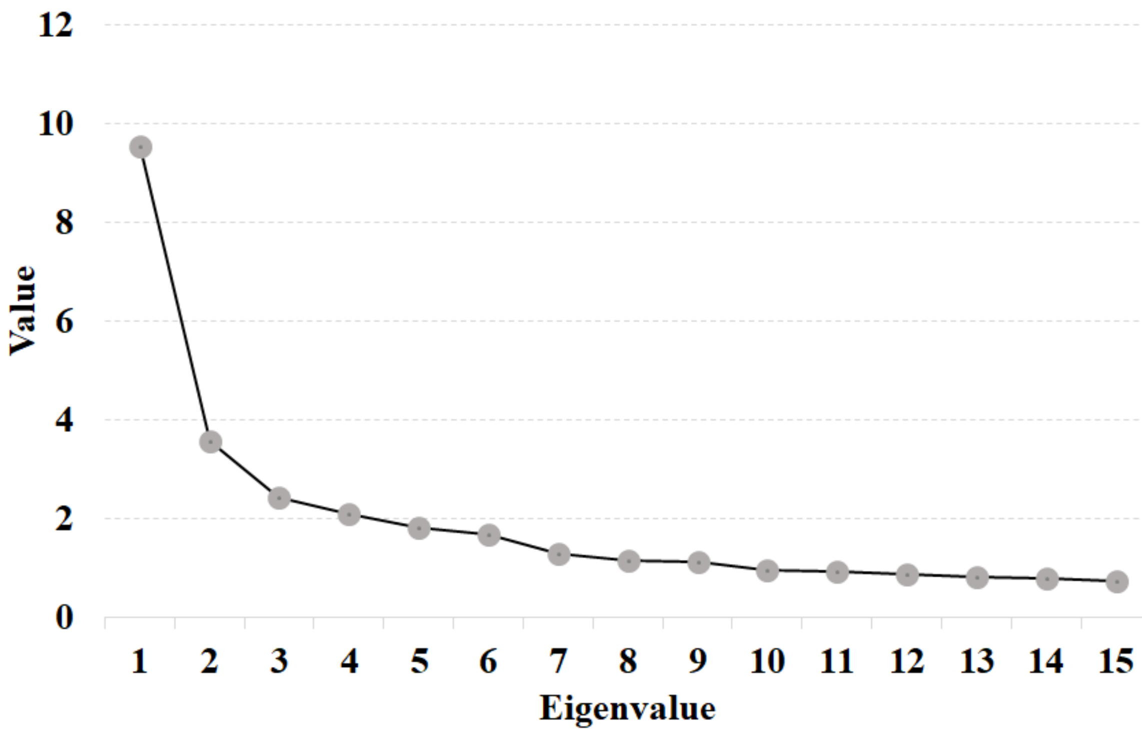
A.3 Equal weighted portfolio (EWP) contains unitary loadings

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 help better identify the relevant factors. Therefore, in the main body of the paper we use as response variable in our test excess delta-hedged portfolio returns over the EWP. 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 (Goyal and Saretto (2009) and Cao and Han (2013) for example).

To assess the variability of the different factor loadings we use the 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 AX 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 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 $Straddle_{index}$ factor. 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.

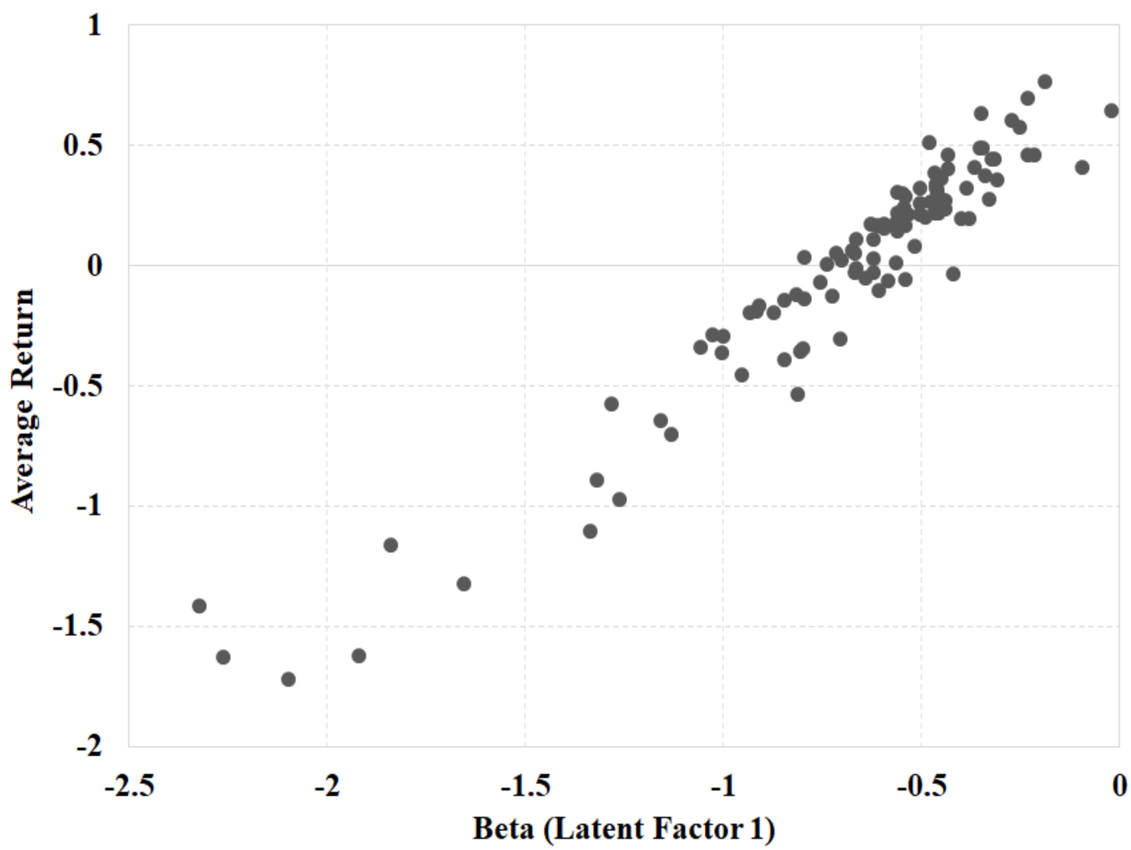
The Table shows that the IB estimator is almost equal to 1 for both EWP and the $Straddle_{Index}$ factor. As such, these factors might have good explanatory power for the time series regressions but fail to explain the cross-section of delta-hedged option returns, as shown by the low correlation between their estimated betas and the average returns of the delta-hedged portfolios.

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



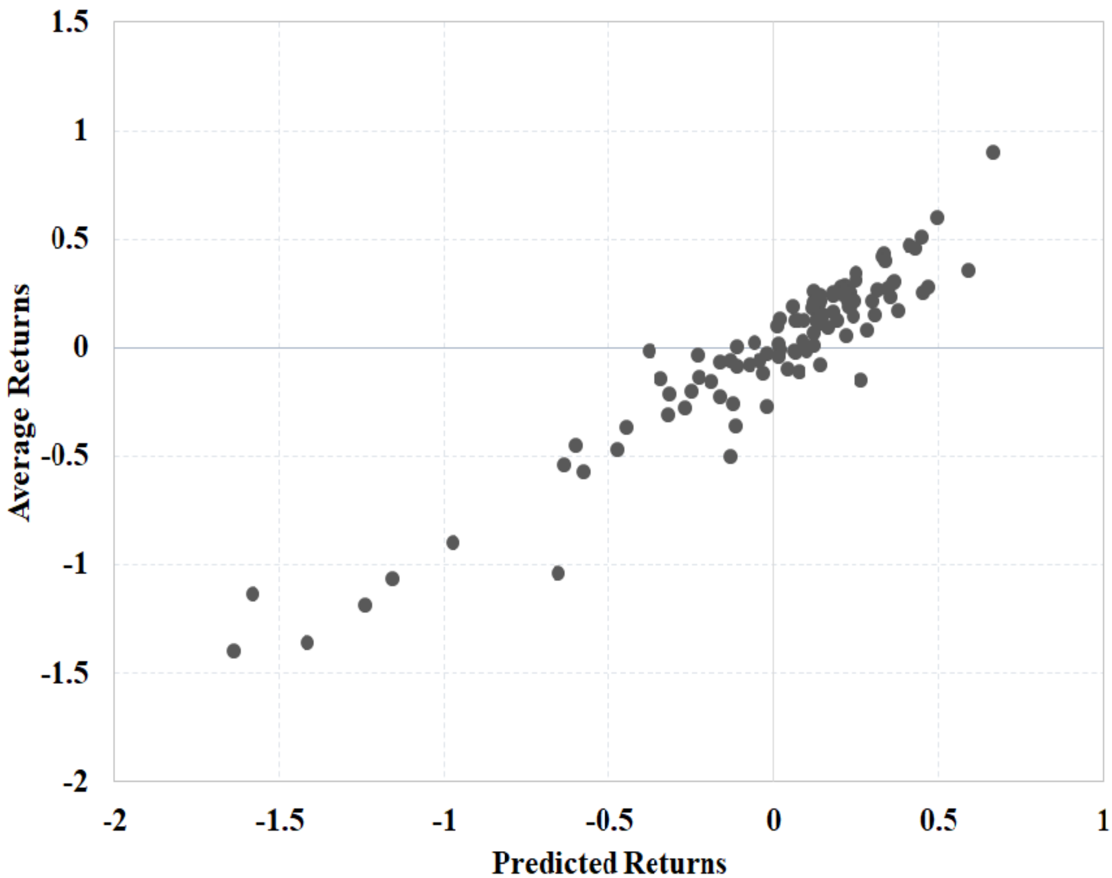
This figure shows the largest fifteen eigenvalues from the sample second-moment matrix of the "doubly demeaned" returns of the 105 delta-hedged portfolios. The sample period is from January 1996 to December 2015.

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



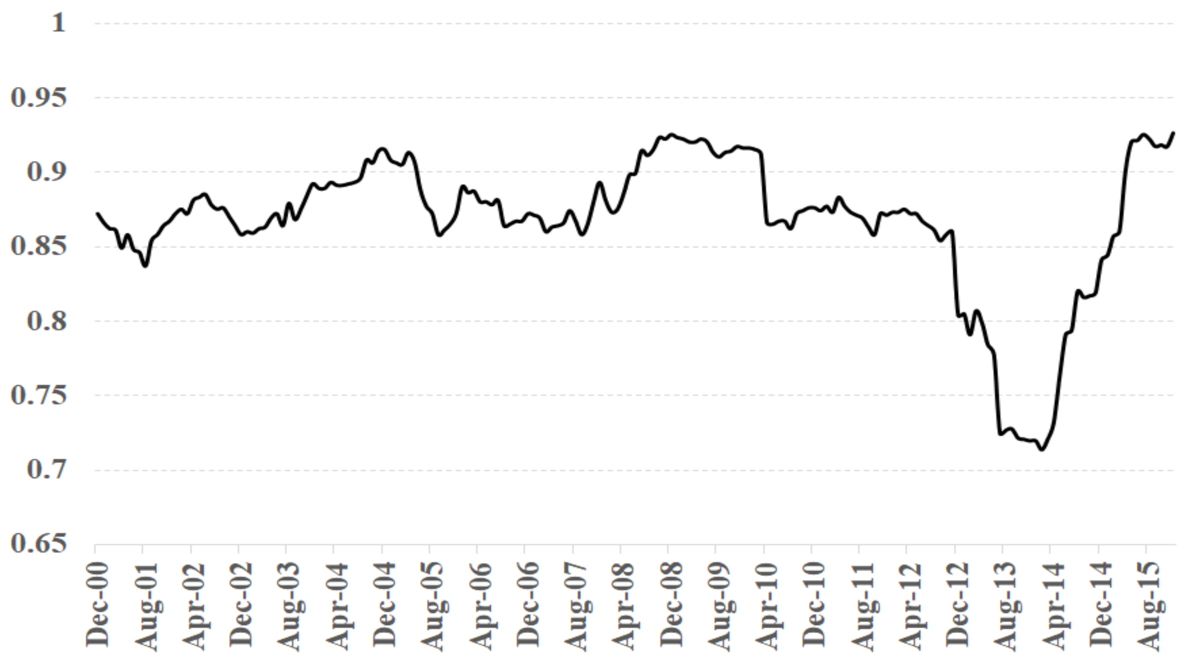
This figure shows the scatter diagram between the average returns of the 105 delta-hedged option portfolios and their corresponding betas of the first latent factor. The sample period is from January 1996 to December 2015.

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 option portfolios and the predicted returns by the model containing the six latent factors.

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



We run monthly rolling regressions using 60-month of data in each iteration. Since our data comprise the period January 1996 – December 2015, we have 181 regressions. 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. This figure shows the correlation over time.

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 (annualized)	1.625	Average Abs. Alpha (annualized)	1.785
Corr E(R) and Beta of Factor 1	0.96	LS_{size}	0.475
Corr E(R) and Beta of Factor 2	-0.143	$LS_{reversal}$	0.154
Corr E(R) and Beta of Factor 3	-0.048	LS_{mom}	0.046
Corr E(R) and Beta of Factor 4	0.008	LS_{ch}	0.344
Corr E(R) and Beta of Factor 5	-0.009	LS_{profit}	0.295
Corr E(R) and Beta of Factor 6	0.005	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
AAAPE	5.461	2.275	1.719	1.681	1.628	1.625

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 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 pricing error (AAAPE).

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 (annualized)	2.18	Average Abs. Alpha (annualized)	2.14
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 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	2.07	1.45	1.99	0.84
	(0.00)	(12.74)	(9.77)	(5.96)	
Jan. 2006 - Dec. 2015	0.00	1.68	1.22	2.14	0.84
	(-0.01)	(14.25)	(8.59)	(10.81)	

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: Performance of the Models with 4 Factors and 7 Factors

(26 Industry Portfolios)

Panel A: Model with 4 candidate factors		Panel B: Model with 7 candidate factors	
Non-zero alphas	0.08	Non-zero alphas	0.08
Average Adj R2	0.22	Average Adj R2	0.25
Corr E(R)-Predicted Return	0.83	Corr E(R)-Predicted Return	0.81
Average Abs. Alpha (annualized)	5.01	Average Abs. Alpha (annualized)	5.05
Corr E(R) and Beta of DH_{idx}	-0.29	Corr E(R) and Beta of DH_{idx}	-0.43
Corr E(R) and Beta of LS_{size}	-0.10	Corr E(R) and Beta of LS_{size}	-0.09
Corr E(R) and Beta of LS_{ivol}	0.70	Corr E(R) and Beta of LS_{ivol}	0.44
Corr E(R) and Beta of LS_{voldev}	0.74	Corr E(R) and Beta of LS_{voldev}	0.54
		Corr E(R) and Beta of LS_{ch}	0.71
		Corr E(R) and Beta of LS_{disp}	-0.14
		Corr E(R) and Beta of LS_{credit}	0.03

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 26 industry characteristic-sorted portfolios. We report the percentage of pricing errors statistically different than 5% generated by each model, the average Adjusted R2 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 11: Correlation Coefficients of the Stock Factors and the Six Latent Option Factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	EWP
MKT	0.10	0.23	-0.12	-0.01	-0.08	-0.10	0.36
SMB	0.19	0.16	-0.08	-0.02	-0.05	-0.08	0.29
HML	-0.13	-0.14	0.09	-0.07	-0.15	-0.14	-0.05
RMW	-0.12	-0.21	0.15	-0.07	0.06	-0.07	-0.20
CMA	-0.14	-0.18	0.04	0.05	-0.06	-0.04	-0.11
MOM	0.18	0.12	0.21	0.06	0.12	0.00	0.02
BAB	0.09	0.10	0.31	-0.04	-0.12	-0.10	0.16
MGMT	-0.10	-0.19	0.09	0.00	0.02	0.00	-0.20
PERF	0.03	-0.10	0.22	0.03	0.14	0.09	-0.11
$LIQD_{LEVEL}$	0.05	0.14	0.09	0.04	-0.33	0.09	-0.25
$LIQD_{INNOV}$	0.03	0.09	0.11	-0.02	-0.21	-0.11	-0.21
$LIQD_{VWF}$	0.06	-0.05	0.10	0.03	-0.12	0.00	-0.18

This table shows the correlation coefficients between the stock factors and the estimated 6 latent option factors plus EWP. Stock factors include the five factors in Fama and French (2016) (MKT, SMB, HML, RMW, CMA), momentum factor in Carhart (1997) (MOM), stock market liquidity risk factors in Pástor and Stambaugh (2003) (PS_{innov}, PS_{level}, PS_{VWF}), betting-against-beta factor in Frazzini and Pedersen (2014) (BAB) and the two mispricing factors in Stambaugh and Yuan (2016) (MGMT and PERF).

Table 12: Regression of the Latent Option Factors and EWP on Stock Factors

	Latent Factor 1	Latent Factor 2	Latent Factor 3	Latent Factor 4	Latent Factor 5	Latent Factor 6	EWP
MKT	1.20	3.83**	-0.80	-0.13	1.19	-2.73	11.58***
SMB	3.47***	2.58	-2.36	-2.74	0.93	-3.89	10.50***
HML	-2.20	-4.02	7.06*	-4.14	-8.39*	-1.94	-1.56
RMW	0.05	-2.44	-7.01*	-1.83	8.51*	-10.22**	-2.42
CMA	-5.23*	-5.79	-7.75	8.24	0.77	-3.95	-3.78
MOM	2.47**	4.53**	0.37	1.37	3.80*	-4.40**	-0.28
BAB	2.72**	7.66***	6.73	-1.13	-3.48*	-1.01	12.66***
MGMT	3.15	0.68	0.49	-0.78	3.23	5.46	0.50
PERF	-2.09	-5.33**	5.42**	-0.94	-2.90	5.98**	0.74
$LIQD_{level}$	0.00	1.36	-1.08	1.67	-5.06***	0.00	2.65
$LIQD_{innov}$	-0.22	-1.25	2.91**	-1.44	1.55	-0.83	0.63
$LIQD_{vwf}$	0.29	-2.91*	1.59	0.68	-2.13	0.74	-2.61
Adjusted R^2	0.13	0.21	0.16	0.04	0.19	0.09	0.32

This table reports the regression result of each option latent factor and EWP (equal-weighted portfolio of all 105 portfolios) on the stock factors. Stock factors include the five factors in Fama and French (2016) (MKT, SMB, HML, RMW, CMA), momentum factor in Carhart (1997) (MOM), stock market liquidity risk factors in Pástor and Stambaugh (2003) (PS_{innov}, PS_{level}, PS_{VWF}), betting-against-beta factor in Frazzini and Pedersen (2014) (BAB) and the two mispricing factors in Stambaugh and Yuan (2016) (MGMT and PERF).

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.