Option-Based Estimation of Co-Skewness and Co-Kurtosis Risk Premia*

Peter Christoffersen
University of Toronto,
CBS and CREATES

Mathieu Fournier
HEC Montreal

Kris Jacobs
University of Houston

Mehdi Karoui
OMERS

December 26, 2014

Abstract

We show that the price of risk for equity factors that are nonlinear in the market return are readily obtained using index option prices. We apply this insight to the price of co-skewness and co-kurtosis risk. The price of co-skewness risk corresponds to the spread between the physical and the risk-neutral second moments, and the price of co-kurtosis risk corresponds to the spread between the physical and the risk-neutral third moments. Our option-based estimates of the prices of risk lead to reasonable values of the associated risk premia. An out-of-sample analysis of factor models with co-skewness and co-kurtosis risk indicates that the new estimates of the price of risk improve the models' performance. Models with higher-order market moments also robustly outperform standard competitors such as the CAPM and the Fama-French model.

JEL Classification: G12, G13, G17

Keywords: Co-skewness; co-kurtosis; risk premia; options; cross-section; out-of-sample.

*We would like to thank the Bank of Canada, the Global Risk Institute, and SSHRC for financial support. For helpful comments we thank Tim Bollerslev, Bryan Kelly, Eric Renault, Michael Rockinger, Bas Werker, and seminar participants at the SoFiE meeting in Toronto. Correspondence to: Kris Jacobs, C. T. Bauer College of Business, 334 Melcher Hall, University of Houston, Houston, TX 77204-6021; Tel: 713-743-2826; Fax: 713-743-4622; E-mail: kjacobs@bauer.uh.edu.

Contents

1 Introduction

The specification and performance of factor models are of paramount importance for financial research and practice, and have been the subject of intense debate for a long time. The Capital Asset Pricing Model (CAPM) has been criticized from different angles, and although its performance improves substantially when evaluating the model conditionally rather than unconditionally, there is widespread consensus that models with better explanatory power are badly needed.

Many alternative models have been proposed over the past four decades, with limited success. One class of models attempts to find new factors using economic intuition or more formal economic modeling. The performance of these models in cross-sectional pricing has been rather disappointing. For instance, aggregate consumption, which is a state variable suggested by theory, has been shown to have limited explanatory power for the cross-section of stock returns. Another class of models constructs factors using a more reduced-form approach, partly based on well-documented stylized facts. The standard examples in this literature are the three-factor model of Fama and French (1993), which includes market, book-to-market and size factors, and the four-factor model suggested by Carhart (1997), which additionally includes a momentum factor. The cross-sectional explanatory power of these models is often judged as satisfactory, but the lack of economic and theoretical foundations is cause for concern.[1]

In view of the state of the literature, further evidence on the pricing of the cross-section of stock returns is therefore a priority. This paper contributes to a literature that goes back to Kraus and Litzenberger (1976), who argue that if investors care about portfolio skewness, co-skewness enters as a second pricing factor in addition to the market portfolio.[2] This argument has later been applied to investor preferences over portfolio kurtosis, leading to co-kurtosis as an additional factor (see, for instance, Ang, Chen, and Xing (2006), Dittmar (2002), Guidolin and Timmermann (2008), and Scott and Horvath (1980)).[3] Despite several important contributions by among others Bansal and Viswanathan (1993), Leland (1997), Lim (1989), Harvey and Siddique (2000), and Dittmar (2002), and despite the theory's obvious intuitive appeal, there seems to be no widespread consensus on the importance of

[1] An extensive literature has sprung up that attempts to provide economic underpinnings for the Fama-French and Carhart factors. See for example Liew and Vassalou (2000) for a risk-based explanation, and Chan, Karceski, and Lakonishok (2003) for a behavioral explanation.

[2] In a related literature, Ang, Hodrick, Xing, and Zhang (2006) analyze the performance of volatility as a pricing factor.

[3] See also Arditti (1967), Rubinstein (1976), and Golec and Tamarkin (1998) for related work.

this literature for cross-sectional asset pricing.

One possible drawback of co-skewness and co-kurtosis as cross-sectional pricing factors is measurement. Measurement is especially difficult when analyzing conditional co-skewness and co-kurtosis.[4] Most existing papers estimate and test the importance of co-skewness and co-kurtosis using two-stage cross-sectional regressions. For a classical example of this type of conditional analysis, see for instance Harvey and Siddique (2000). This approach necessitates the estimation of co-skewness betas in a first stage. These betas are subsequently used in the second-stage cross-sectional regression. It is well-known that the estimation of betas in the first-stage regression is noisy, and these errors carry over in the second-stage cross-sectional regression. While these problems apply to virtually all implementations of cross-sectional models, including the CAPM, they may be especially serious in the case of co-skewness and co-kurtosis. The simple basic intuition is that the higher the moment, the more difficult it is to estimate precisely. This argument applies a fortiori to the estimation of co-measures of higher moments, such as co-skewness and co-kurtosis, and the betas for these factors. Therefore, errors in estimated betas may be large for these models, leading to biases in the cross-sectional estimation of the price of risk that are potentially much larger than in the competing case of the CAPM or the Fama-French three-factor model.

We propose a new strategy to estimating the price of co-skewness and co-kurtosis risk, which avoids the problems inherent in the second-stage cross-sectional regression. Our approach can also be used to estimate the price of other risks, provided that they are nonlinear functions of the market return. We derive our result based on the well-known representation of cross-sectional asset pricing models that relies on the stochastic discount factor or SDF (see Cochrane (2005)). The CAPM corresponds to the assumption of linearity of the SDF with respect to the market return. A quadratic SDF implies that investors require compensation not only for the exposure to market returns but also for the exposure to squared market returns, which leads to co-skewness risk aversion.[5] SDFs that are higher-order functions of the market return lead to progressively more complex co-movements with market returns as pricing factors.

The key difference between our approach and existing studies is that we explicitly impose restrictions on the pricing of both stocks and contingent claims. This allows us to derive explicit formulas for the time-varying price of risk for the exposure to any nonlinear function

[4] Kraus and Litzenberger (1976) provide an unconditional empirical analysis of co-skewness.

[5] See Dittmar (2002) for an investigation of higher moments in cross-sectional pricing using this approach. See Bakshi, Madan, and Panayotov (2010) for evidence that pricing kernels are U-shaped as a function of market returns.

of the market return. For instance, for the case of co-skewness risk we show that the price of co-skewness risk corresponds to the spread between the physical and the risk-neutral second moment. Similarly, the price of co-kurtosis risk is given by the spread between the physical and the risk-neutral third moment. To provide intuition for this result, consider the special case where the SDF is a linear function of the market return, which corresponds to the CAPM. In this case, our general result shows that the price of risk can be estimated as the difference between the spread between the physical and risk-neutral first moment. This equals the market return minus the risk-free rate, which is of course the classical CAPM result.

We empirically investigate the performance of our approach for the pricing of co-skewness and co-kurtosis risk. Using monthly data for the period 1996-2012, we find that the price of co-skewness risk has the expected negative sign in every month in our sample, and the price of co-kurtosis risk has the expected positive sign in most months. On average, both estimated prices of risk are larger in absolute value than the traditional estimates obtained using a two-stage Fama-MacBeth approach. More importantly, while the average prices of risk obtained using the Fama-MacBeth approach have the theoretically anticipated signs on average, they are often estimated with the opposite sign. We evaluate the cross-sectional performance of our newly proposed estimates out-of-sample, and find that they outperform implementations of the CAPM and the Fama-French three factor model that use cross-sectional regressions to estimate the price of risk.

The paper proceeds as follows. Section 2 describes our alternative approach to the measurement of (nonlinear) market risk. Section 3 presents an empirical investigation of co-skewness risk. Section 4 investigates co-kurtosis risk. Section 5 concludes.

2 Measuring Market Risks: An Option-Based Approach

In this section we provide an overview of multifactor asset pricing models in which cross-sectional differences in expected returns between assets are determined by their exposure to risk factors that are nonlinear functions of the market return. This setting corresponds to assuming SDFs that are nonlinear in the market return. We proceed to propose an option-based approach to measuring the price of risk for these types of exposures. We investigate two special cases that are of significant empirical interest: exposure to the squared market return R_m^2, which captures co-skewness risk; and exposure to the third power of the market return R_m^3, which captures co-kurtosis risk.

2.1 Measuring Co-Skewness Risk

Before we introduce the general case, we first discuss two specific examples to provide more intuition for our approach. We begin with co-skewness risk. Let m_{t+1} denote the stochastic discount factor

$$m_{t+1} = a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right), \tag{1}$$

where $R_{m,}$ denotes the stock market return, and $E_t^P(.)$ denotes the expectation under the physical probability measure. Similar to Harvey and Siddique (2000, henceforth HS), our setup is based on the assumption of a quadratic SDF. As explained by HS (2000), a quadratic SDF is consistent with several utility-based asset pricing models. The performance of quadratic pricing kernels is studied in Bansal and Viswanathan (1993) and Chabi-Yo (2008).

Given this SDF, we can establish pricing restrictions on any asset return. The key feature of our approach is that we jointly consider theoretical restrictions on stocks and contingent claims, whereas the existing cross-sectional asset pricing literature focuses exclusively on the underlying assets. Our approach enables the specification of new estimators for the price of co-skewness risk which can be easily implemented using short data windows.

Denote the return on a stock j by R_j and the return on a contingent claim on the stock by R_i. The existing literature contains several measures of co-skewness risk, which all capture covariation between the stock return and the squared market return. Kraus and Litzenberger (1976, henceforth KL) define co-skewness risk by $\frac{E^P\left[(R_j-\overline{R}_j)(R_m-\overline{R}_m)^2\right]}{E^P\left[(R_m-\overline{R}_m)^3\right]}$. HS (2000) mainly focus on $cov(R_j, R_m^2)$ in their theoretical analysis but consider four different co-skewness measures in their empirical analysis. Our measure of co-skewness risk is the beta with respect to R_m^2 in a multivariate regression. This measure allows for mathematical tractability in the derivation of the price of risk as shown in the following proposition. The proposition presents the pricing implications of the SDF defined in equation (1).

Proposition 1 *If the stochastic discount factor (SDF) has the following form:*

$$m_{t+1} = a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right),$$

then the cross-sectional pricing restrictions are

$$E_t^P(R_{j,t+1}) - R_f = \lambda_t^{MKT}\beta_{j,t}^{MKT} + \lambda_t^{COSK}\beta_{j,t}^{COSK}, \tag{2}$$

and

$$E_t^P(R_{i,t+1}) - R_f = \lambda_t^{MKT}\beta_{i,t}^{MKT} + \lambda_t^{COSK}\beta_{i,t}^{COSK}, \tag{3}$$

where β_t^{MKT} *and* β_t^{COSK} *are the loadings from the projection of the asset returns on* $R_{m,t+1}$ *and* $R_{m,t+1}^2$. *The price of covariance risk,* λ_t^{MKT}*, is*

$$\lambda_t^{MKT} = E_t^P(R_{m,t+1}) - R_f, \tag{4}$$

and the price of co-skewness risk, λ_t^{COSK}*, is*

$$\lambda_t^{COSK} = E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2). \tag{5}$$

where $E_t^P(.)$ *and* $E_t^Q(.)$ *denote the expectation under the physical and risk-neutral probability measures, respectively.*

Proof. Linear factor models, in which the stochastic discount factor is $m_{t+1} = a_t + \mathbf{b}_t'\left(\widetilde{\mathbf{f}}_{t+1} - E_t^P(\widetilde{\mathbf{f}}_{t+1})\right) = a_t + \mathbf{b}_t'\mathbf{f}_{t+1}$, are equivalent to beta-representation models with the vector of risk factors $\mathbf{f}$

$$E_t^P(R_{j,t+1}) - R_{f,t} = \boldsymbol{\lambda}_t'\boldsymbol{\beta}_{j,t}, \tag{6}$$

where $\boldsymbol{\lambda}_t' = \frac{-1}{a_t}\mathbf{b}_t'E_t^P(\mathbf{f}_{t+1}\mathbf{f}_{t+1}')$, $(1+R_{f,t}) = \frac{1}{a_t} = \frac{1}{E_t^P(m_{t+1})}$, $\boldsymbol{\beta}_{j,t} = \left[E_t^P(\mathbf{f}_{t+1}\mathbf{f}_{t+1}')\right]^{-1} E_t^P(\mathbf{f}_{t+1}R_{j,t+1})$, see for instance Cochrane (2005). Since the pricing kernel prices all the assets including contingent claims, the above equation also holds for any claim i whose price is contingent on the stock j and has a payoff function $\Psi(R_{j,t+1})$, for any function $\Psi(.)$. From equation (6) we have

$$E_t^P\left(\frac{\Psi(R_{j,t+1}) - P_{i,t}}{P_{i,t}}\right) - R_{f,t} = \boldsymbol{\lambda}_t'\boldsymbol{\beta}_{i,t}, \tag{7}$$

where $P_{i,t}$ is the price of the contingent claim i. Using the definition of $\boldsymbol{\beta}_{i,t}$ we have

$$E_t^P\left(\frac{\Psi(R_{j,t+1}) - P_{i,t}}{P_{i,t}}\right) - R_{f,t} = \boldsymbol{\lambda}_t'\left[E_t^P((\mathbf{f}_{t+1}\mathbf{f}_{t+1}')\right]^{-1} E_t^P\left((\mathbf{f}_{t+1}\frac{\Psi(R_{j,t+1}) - P_{i,t}}{P_{i,t}}\right). \tag{8}$$

Rearranging and using $E_t^P(\mathbf{f}_{t+1}) = 0$ gives

$$E_t^P\left[\Psi(R_{j,t+1})\right] - P_{i,t}\left(1 + R_{f,t}\right) = \boldsymbol{\lambda}_t'\left[E_t^P((\mathbf{f}_{t+1}\mathbf{f}_{t+1}')\right]^{-1} E_t^P\left[(\mathbf{f}_{t+1}\Psi(R_{j,t+1})\right] \tag{9}$$

The no-arbitrage condition ensures the existence of at least one risk-neutral measure Q such that $P_{i,t} = \frac{1}{(1+R_{f,t})} E_t^Q [\Psi(R_{j,t+1})]$. Therefore, we obtain

$$E_t^P [\Psi(R_{j,t+1})] - E_t^Q [\Psi(R_{j,t+1})] = \boldsymbol{\lambda}_t' \boldsymbol{\beta}_{\Psi,t} \tag{10}$$

where $\boldsymbol{\beta}_{\Psi,t}$ is from the projection of $\Psi(R_{j,t+1})$ on $\mathbf{f}_{t+1}$.

For $m_{t+1} = a_t + b_{1,t} \left(R_{m,t+1} - E_t^P(R_{m,t+1}) \right) + b_{2,t} \left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2) \right)$ and $\Psi = R_m$, equation (10) reduces to equation (4). Applying equation (10) for $\Psi = R_m^2$, we recover equation (5). ■

Proposition 1 shows that the price of co-skewness risk corresponds to the spread between the physical and the risk-neutral second moments for the market return. Unlike other moments, the second moment is fairly easy to estimate under both the physical and risk-neutral probability measures. The literature contains a wealth of robust approaches for modeling the physical volatility of stock returns. The risk-neutral moment can be estimated from option market data either by the implied volatility of option pricing models, or alternatively using a model-free approach based as in Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003).

A number of existing studies relate the volatility spread to risk aversion (see Bakshi and Madan (2006)) or the price of correlation risk (see Driessen, Maenhout and Vilkov (2009)). Proposition 1 shows that if the pricing kernel is quadratic, then the volatility spread is equal to the price of co-skewness risk.

Proposition 1 allows for separate identification of the price of covariance (λ_t^{MKT}) and co-skewness (λ_t^{COSK}) risk. Note that this result is simply an application of the general result that if the factor is a portfolio, then the expected return on the factor is equal to the factor risk premium. Importantly, the result holds regardless of assumptions on other risk factors. This is in stark contrast with risk premia estimated from two-pass cross-sectional regressions for which the empirical results depend on the other risk factors considered in the regression. Our approach also has the advantage of easily capturing time variation in risk premia.

The existing empirical evidence clearly indicates that risk-neutral variance is larger than physical variance, therefore suggesting a negative price of co-skewness risk. See for instance Bakshi and Madan (2006), Bollerslev, Tauchen, and Zhou (2009), Carr and Wu (2009), and Jackwerth and Rubinstein (1996). A negative price of risk is consistent with theory. Assets with lower (more negative) co-skewness decrease the total skewness of the portfolio and increase the likelihood of extreme losses. Assets with lower co-skewness are thus perceived by investors to be riskier and should command higher risk premiums.

While our approach to estimating the price of co-skewness risk is different from the existing literature and the betas are defined (and/or scaled) differently, the implications for the risk premia on the assets are of course the same. Using the fact that $E_t^P(R_{m,t+1}) - R_f = \lambda_t^{MKT}$ and $E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2) = \lambda_t^{COSK}$, we can re-write equation (2) of proposition 1 as follows

$$E_t^P(R_{j,t+1}) - R_f = \beta_{j,t}^{MKT}\left[E_t^P(R_{m,t+1}) - R_f\right] + \beta_{j,t}^{COSK}\left[E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2)\right], \quad (11)$$

which can also be written as

$$E_t^P(R_{j,t+1}) - R_f = c_t + \beta_{j,t}^{MKT} E_t^P\left(R_{m,t+1}\right) + \beta_{j,t}^{COSK} E_t^P\left(R_{m,t+1}^2\right), \quad (12)$$

where $c_t = -\beta_{j,t}^{MKT} R_f - \beta_{j,t}^{COSK} E_t^Q(R_{m,t+1}^2)$. Equation (12) shows the link between our method and the approaches in KL (1976) and HS (2000). It is equivalent to equation (6) of KL (1976) and equation (8) of HS (2000).

The crucial difference between our approach and the one in KL (1976) and HS (2000) is that we explicitly impose the pricing restrictions on contingent claims. This additional restriction leads to a very simple estimator of the price of risk.

2.2 Measuring Co-Kurtosis Risk

A natural extension of the quadratic pricing kernel discussed in the previous section is the cubic pricing kernel studied in Dittmar (2002), given by

$$\begin{aligned} m_{t+1} &= a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right) \\ &\quad + b_{3,t}\left(R_{m,t+1}^3 - E_t^P(R_{m,t+1}^3)\right). \end{aligned} \quad (13)$$

A cubic pricing kernel is consistent with investors' preferences for higher order moments, specifically skewness and kurtosis. See Dittmar (2002) and HS (2000) for more details. As before, we first make an assumption on the shape of the SDF and then derive pricing restrictions. In this case, the expected excess return on any asset will be related to co-kurtosis risk, in addition to covariance risk and co-skewness risk. As explained by Dittmar (2002), kurtosis measures the likelihood of extreme values and co-kurtosis captures the sensitivity of asset returns to extreme market return realizations. If investors are averse to extreme values, they require higher compensation for assets with higher co-kurtosis risk, meaning that the price of co-kurtosis risk should be positive. See Guidolin and Timmermann (2008) and Scott and Horvath (1980) for a more detailed discussion. Similar to co-skewness risk, co-kurtosis risk has been defined in various ways in previous studies. For instance, Ang, Chen and Xing (2006) measure co-kurtosis risk using $\frac{E^P\left[(R_j-\overline{R}_j)(R_m-\overline{R}_m)^3\right]}{\sqrt{E^P\left[(R_j-\overline{R}_j)^2\right]}\left(E^P\left[(R_m-\overline{R}_m)^2\right]\right)^{3/2}}$, and Guidolin and Timmermann (2008) use $cov(R_j, R_m^3)$. In this paper, we measure co-kurtosis risk by the return's beta with respect to the cubic market return R_m^3. We denote the co-kurtosis beta of a stock j (contingent claim i) by $\beta_{j,t}^{COKU}$ ($\beta_{i,t}^{COKU}$).

The following proposition presents the estimator for the co-kurtosis price of risk and the cross-sectional pricing restrictions.

Proposition 2 *If the stochastic discount factor (SDF) has the following form:*

$$\begin{aligned} m_{t+1} &= a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right) \\ &\quad + b_{3,t}\left(R_{m,t+1}^3 - E_t^P(R_{m,t+1}^3)\right), \end{aligned}$$

then the cross-sectional restrictions are

$$E_t^P(R_{j,t+1}) - R_f = \lambda_t^{MKT}\beta_{j,t}^{MKT} + \lambda_t^{COSK}\beta_{j,t}^{COSK} + \lambda_t^{COKU}\beta_{j,t}^{COKU}, \tag{14}$$

and

$$E_t^P(R_{i,t+1}) - R_f = \lambda_t^{MKT}\beta_{i,t}^{MKT} + \lambda_t^{COSK}\beta_{i,t}^{COSK} + \lambda_t^{COKU}\beta_{i,t}^{COKU}, \tag{15}$$

where β_t^{MKT}, β_t^{COSK}, and β_t^{COKU} are from the projection of asset returns on $R_{m,t+1}$, $R_{m,t+1}^2$ and $R_{m,t+1}^3$, respectively. The prices of covariance, λ_t^{MKT}, and co-skewness risk λ_t^{COSK} are

$$\lambda_t^{MKT} = E_t^P(R_{m,t+1}) - R_f, \tag{16}$$

$$\lambda_t^{COSK} = E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2), \tag{17}$$

and the price of co-kurtosis risk, λ_t^{COKU}, is

$$\lambda_t^{COKU} = E_t^P(R_{m,t+1}^3) - E_t^Q(R_{m,t+1}^3), \tag{18}$$

where $E_t^P(.)$ and $E_t^Q(.)$ denote the expectation under the physical respectively risk-neutral probability measure.

Proof. The structure of the proof largely follows the proof of Proposition 1. Assuming that $m_{t+1} = a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right) + b_{3,t}\left(R_{m,t+1}^3 - E_t^P(R_{m,t+1}^3)\right)$, then, as in Proposition 1, applying equation (10) for $\Psi = R_m$, we recover equation (16), and applying equation (10) for $\Psi = R_m^2$, we recover equation (17). In addition, applying equation (10) for $\Psi = R_m^3$, we obtain equation (18). ■

Proposition 2 shows that the price of co-kurtosis risk is equal to the spread between the market physical and risk-neutral third moments. Clearly third moments are harder to estimate than second moments. Nevertheless, existing evidence (see for instance Bakshi, Kapadia, and Madan (2003)) indicates that the risk-neutral distribution for the market return is more left skewed than the physical distribution, therefore suggesting a positive price of co-kurtosis risk. This is entirely consistent with theory, as explained earlier in this section.

2.3 The General Case

We now examine more general nonlinearities in the SDF. Preference theory is relatively silent about the sign of terms in the SDF higher than the third order, and therefore we do not extend our empirical analysis beyond the cubic SDF. However, while the empirical focus of this paper is on co-skewness and co-kurtosis risk, our approach can be used for any source of risk that is an arbitrary nonlinear (including linear) function of the market return. This does not just include powers of the market return, it includes more complex nonlinear relationships, such as for instance measures of downside risk as in Ang, Chen, and Xing (2006). We now present the general result which nests among many other results the results for co-skewness risk in Section 2.1 and co-kurtosis risk in Section 2.2.

Proposition 3 *If the stochastic discount factor (SDF) has the following form:*

$$m_{t+1} = a_t + \sum_k b_{k,t}\left(G_k(R_{m,t+1}) - E_t^P\left[G_k(R_{m,t+1})\right]\right) + \sum_l c_{l,t}\left(f_{l,t+1} - E_t^P(f_{l,t+1})\right),$$

then the cross-sectional pricing restrictions are

$$E_t^P(R_{j,t+1}) - R_f = \sum_k \lambda_t^k \beta_{j,t}^k + \sum_l \gamma_t^l \beta_{j,t}^l, \tag{19}$$

and

$$E_t^P(R_{i,t+1}) - R_f = \sum_k \lambda_t^k \beta_{i,t}^k + \sum_l \gamma_t^l \beta_{i,t}^l, \tag{20}$$

where the β_t^k and β_t^l are from the projection of asset returns on $G_k(R_{m,t+1})$ and $f_{l,t+1}$ respectively, and γ^l is the price of risk associated with the factor f_l. The price of risk associated with the exposure to a nonlinear function, G_k, of the market return, λ_t^k, is

$$\lambda_t^k = E_t^P(G_k(R_{m,t+1})) - E_t^Q(G_k(R_{m,t+1})), \tag{21}$$

where $E_t^P(.)$ and $E_t^Q(.)$ denote the expectation under the physical respectively the risk-neutral probability measure.

Proof. The structure of the proof is again similar to the proof of Proposition 1. If $m_{t+1} = a_t + \sum_k b_{k,t}\left(G_k(R_{m,t+1}) - E_t^P\left[G_k(R_{m,t+1})\right]\right) + \sum_l c_{l,t}\left(f_{l,t+1} - E_t^P(f_{l,t+1})\right)$, then applying equation (10) for $\Psi = G_k(R_{m,t+1})$ we obtain equation (21). ■

Proposition 3 shows that the reward for exposure to any nonlinear function G of the market return is determined by the spread between the physical and the risk-neutral expectations of this function. The proposition also demonstrates that we can easily incorporate factors that are not necessarily functions of the market return.

3 Estimating the Price of Co-Skewness Risk

We begin the empirical investigation by documenting the price of co-skewness risk using the estimators presented in Proposition 1. The implementation of our approach requires the estimation of physical and risk-neutral conditional expectations. For the price of co-skewness risk, we need to estimate the second conditional moment under the risk-neutral measure, $E_t^Q(R_{m,t+1}^2)$, and under the physical measure, $E_t^P(R_{m,t+1}^2)$. We first discuss the estimation of these moments. Subsequently we estimate the price of co-skewness risk and compare our estimate with more conventional regression-based estimates.

3.1 Estimating the Risk-Neutral Variance

We estimate the risk-neutral variance in two ways. In our benchmark analysis, we use the square of the VIX index as our estimate for the risk-neutral variance. The VIX provides a very simple benchmark because the data are readily available from the Chicago Board of Options Exchange (CBOE). Using the VIX has a number of advantages. The construction of the VIX is exogenous to our experiment, and so it is not possible to design it to maximize performance. Even more importantly, the VIX is available for a longer sample period than the available alternatives. We obtain data for the period January 1986 to December 2012. For existing studies that use the VIX as a proxy for the risk-neutral second moment see for instance Bollerslev, Tauchen, and Zhou (2009). In the robustness analysis in Section 3.6, we use an alternative approach to compute the risk-neutral variance, following Bakshi and Madan (2000).

3.2 Estimating the Physical Variance

The literature contains a large number of models for estimating physical variance. In our benchmark analysis, we use a simple and robust implementation of the heterogeneous autoregressive model (HAR) of Corsi (2009), defined as follows

$$RP_{t+1,t+K} = \phi_0 + \phi_1 RP_{t-1,t} + \phi_2 RP_{t-4,t} + \phi_3 RP_{t-20,t+K} + \varepsilon_{RP,t}, \tag{22}$$

where

$$RP_{s,s+\tau} = RP_s + RP_{s+1} + ... + RP_{s+\tau}, \tag{23}$$

and

$$\begin{aligned} RP_t &= \ln(S_t^{High}/S_t^{Open})\left[\ln(S_t^{High}/S_t^{Open}) - \ln(S_t^{close}/S_t^{Open})\right] \qquad (24) \\ &\quad + \ln(S_t^{Low}/S_t^{Open})\left[\ln(S_t^{Low}/S_t^{Open}) - \ln(S_t^{close}/S_t^{Open})\right], \qquad (25) \end{aligned}$$

where S_t^{Close} (S_t^{Open}) denotes the close (open) stock price and S_t^{High} (S_t^{Low}) denotes the highest (lowest) price on day t. We estimate the HAR model using OLS and a recursive ten-year window. To ensure consistency with our measure of the risk-neutral variance, we generate one-month forecasts of the physical variance at the end of every month.

In the robustness analysis in Section 3.6, we use several alternative approaches to estimate the physical variance. We use a simple autoregressive model on realized variances, the NGARCH model of Engle and Ng (1993), and the Heston (1993) stochastic volatility model. For each of these models, we also use a recursive ten-year window.

3.3 The Price of Co-Skewness Risk

Using the estimates of the physical and risk-neutral second moments, the estimated price of co-skewness risk for month t is now simply

$$\widehat{\lambda}_t^{COSK} = \widehat{E}_t^P(R_{m,t+1}^2) - \widehat{E}_t^Q(R_{m,t+1}^2)$$

Table 1 reports descriptive statistics for the estimates of the moments and the price of risk. Figure 1 depicts the time series of the price of co-skewness with the corresponding estimated physical and risk-neutral moments required to compute these prices. The figures show some spikes surrounding the 1987 stock market crash, the 1998 LTCM collapse, the WorldCom bankruptcy in 2002, and the subprime crisis. These spikes occur for both the risk-neutral as well as physical moments, but the spikes in the physical variance are relatively smaller than the risk-neutral spikes except for the one during the subprime crisis. This is to some extent due to the choice of the model for the physical variance. Other approaches for modeling the physical variance in some cases yield larger spikes, but they do not affect our results for cross-sectional pricing. We discuss these results in more detail in Section 3.6 below.

The co-skewness risk premium is negative for almost all months. On average the co-skewness price of risk is equal to -0.271. These findings are consistent with theory, and with existing empirical studies that document a negative price of co-skewness risk, see for instance KL (1976) and HS (2000). However, it is critical to emphasize that these existing estimates are typically averages of the price of risk over several years. Most studies estimate prices of risk using a two-pass Fama-MacBeth (1973) setup and report the average estimates of the month-by-month cross-sectional regressions. Often the estimates of the price of risk have the opposite sign over shorter time periods, as we will demonstrate below. What is remarkable about the results reported in Figure 1 is that we have genuinely conditional month-by-month estimates of the price of risk that have the theoretically expected sign in almost every month. Note also that while there is no guarantee that these results for co-skewness will continue to hold in the future, we know that usually implied variances exceed historical variances have exceeded implied variances. Because of this stylized fact, when using this method we can expect theoretically plausible estimates of co-skewness risk most of the time.

3.4 Regression-Based Estimates of the Price of Co-Skewness Risk

The studies referenced in Section 3.3 use different sample periods and implementations of the cross-sectional regressions. To provide more insight into our new estimates of co-skewness risk, we now compare our estimates with estimates obtained using regression methods, using samples for the same period 1986-2012. We report results from Fama-MacBeth regressions using the classical setup. We first obtain betas using sixty monthly returns, and subsequently we run a cross-sectional regression for the next month.

Table 2 reports results for two factor models. The first model incorporates co-skewness exposure but also exposure to the market factor. The second model also includes the Fama-French (1993) size and book-to-market factors, and the momentum factor. For each regression, following Fama and MacBeth (1973), we report the average of the cross-sectional regression estimates as well as the t-statistics on these averages. We report on four cross-sectional datasets that are commonly used in the existing literature. We use portfolios formed on size and book-to-market ratio, portfolios formed on size and momentum, portfolios formed on size and short-term reversal, and portfolios formed on size and long-term reversal. The data on these portfolios, as well as the data on the Fama-French and momentum factors we use to analyze competing models, are collected from Kenneth French's online data library. We report on the 1986-2012 period but also on the longer 1966-2012 period to provide additional perspective.

Figures 2 and 3 report more detailed results for two specifications. Figure 2 reports on the univariate model that exclusively contains co-skewness exposure. We depict the average returns as well as the average co-skewness betas for both sample periods used in Table 2. Figure 3 reports on the model that includes the co-skewness and market factors using the same time period used for our estimates in Table 1, 1986 through 2012. We report the time-series of the cross-sectional regression estimates of the price of risk. The estimates for the price of co-skewness risk reported in the first model in Panel A of Table 2 are the averages of the time series in Figure 2.

Consider first the results for 1986-2012 in Panel A of Table 2 and Figure 2. For our purpose, the most important conclusion is that the estimates of the price of co-skewness risk critically depend on the assets used in estimation. For the univariate models displayed in Figure 2, the estimate of the price of co-skewness risk is -0.084 when using the twenty-five size and book-to-market portfolios. When using the twenty-five size and momentum portfolios, the estimate is -0.182. However, when using the size and short-term reversal portfolios and the size and long-term reversal portfolios, the estimates are positive. The only estimate that is statistically significant is the one obtained using the twenty-five size and momentum portfolios. Panel A of Table 2 indicates that when including the market factor in the regressions, the results do not change much. The estimates for the size and short-term reversal portfolios and the size and long-term reversal portfolios are now negative but they are not statistically significant. Our first conclusion is that the choice of test assets is critical for the estimate of the price of co-skewness risk.

Our second conclusion is that our newly proposed estimate of the price of co-skewness risk in Table 1, which is equal to -0.271, is much larger (in absolute value) than any of the estimates obtained using the regression approach. This of course does not necessarily mean that our estimate is superior; in order to demonstrate that we have to show that the larger estimate leads to improved fit. We address this in Section 3.5 below.

Including additional factors in the cross-sectional model does not change this conclusion. Table 2 also reports results for the price of co-skewness risk when the Fama-French factors as well as the momentum factor are included in the regressions. The resulting estimates are smaller in absolute value and are always statistically insignificant.

Finally, it could be argued that our 1986-2012 sample period is relatively short to reliable estimate the price of co-skewness risk using a regression approach. We use the 1986-2012 period to compare the results to our newly proposed estimates, which are limited to this sample period because of the availability of risk-neutral second moments. Panel B of Table 2 therefore also reports results for the longer 1966-2012 period. The resulting estimates of the price of co-skewness risk are very similar to those obtained for the 1986-2012 period, and also strongly differ across test assets.

The time-series of the cross-sectional estimates of the price of co-skewness risk in Figure 3 yields another important conclusion. It is clear that the cross-sectional estimates vary a lot over time, and that they are often positive, even when the averages reported in Panel A of Table 2 are negative. We have to interpret the evidence in Figure 2 with caution, because the essence of the Fama-MacBeth cross-sectional procedure is of course to estimate the price of risk by averaging the time series of cross-sectional estimates. In other words, the fact that the estimates in Figure 2 are positive for some months may not in itself constitute a problem. Nevertheless, the contrast with the results for our newly proposed method in Figure 1 is stark. In Figure 1, we also report estimates for every month. It is striking that the monthly estimates are almost all negative. This of course also explains why the negative average estimate of -0.271 for our approach is so much larger (in absolute value), because the negative peaks are not cancelled out by positive estimates in other months. At a minimum, we can conclude that our newly proposed estimator provides us with a genuinely conditional month-by-month estimate of the price of risk that almost always has the sign suggested by theory.

Figure 3 provides additional insight into the properties of the regression estimates. Based on the results in Figure 2 and Table 2, we concluded that there were substantial differences between different test assets. But Figure 3 instead indicates substantial commonality between test assets in the month-by-month estimates of the price of risk. In other words, the four time series in Figure 3 are highly correlated. Table 2 indicates that the only test assets that yield a significantly negative price of co-skewness risk are the twenty-five size and momentum portfolios. Figure 3 indicates that this can be explained by the fact that the regression estimates for these test assets vary less over time compared to the estimates for other test assets, even though the monthly estimates are also often positive.

In summary, a comparison of our newly proposed estimates of the price of co-skewness risk with regression-based estimates yields three important conclusions. First, regression-based estimates critically depend on the test assets used in estimation, whereas our approach is by design independent of the test assets. Second, our estimate -0.271 indicates a role for co-skewness that is much larger in magnitude. Third, when looking a month-by-month estimates we obtain a consistently negative sign of the price of co-skewness risk in our approach. While the regression approach is of course mainly focused on the overall average of the cross-sectional coefficients, the estimates are positive for many months, and this has implications for the statistical significance of the estimates. Moreover, the averaging needed to obtain reliable results with the regression approach makes the estimates less genuinely conditional.

We therefore conclude that our approach is economically appealing. To show that it improves on regression-based estimates, we have to demonstrate that it leads to a better fit. This is the subject to which we now turn.

3.5 Comparing Model Fit: Out-of-Sample Tests

When using regression-based methods, the cross-sectional or Fama-MacBeth regressions which provide estimates of the prices of risk are also used to evaluate cross-sectional fit and assess the model's performance. For instance, Table 2 reports on model performance using the R-square. Even though there are many other related evaluation criteria, in the overwhelming majority of cases these evaluation criteria are similar to the R-square in Table

2 in the sense that they are in-sample. Table 2 also highlights a common drawback of such in-sample comparisons, in the sense that models with more factors often lead to a better fit.

It is important to note that in our approach, we construct betas or loadings in exactly the same way as in the traditional Fama-MacBeth setup, but the price of risk is not estimated from a cross-sectional regression. Instead it is estimated as a historical risk premium, and subsequently it is used to assess cross-sectional fit. This difference can best be understood by referring to the well-known case of the CAPM. The CAPM is often evaluated using the Fama-MacBeth approach, by first estimating betas and then running cross-sectional regressions. But alternatively the price of risk for the CAPM could be estimated using the historical market risk premium, and the cross-sectional fit of the CAPM could be evaluated using this price of risk and (the same) estimated betas. It does not make sense to compare the in-sample cross-sectional R-square of the CAPM when the price of risk is estimated in the regression with an R-square obtained by inserting the historical risk premium in the same sample. This amounts to comparing an in-sample fit with an out-of-sample fit. We therefore implement tests of our models using a genuinely out-of-sample approach for all models. Out-of-sample testing of cross-sectional models is becoming increasingly popular, see for instance Simin (2008) and Ferson, Nallareddy, and Xie (2012).

We therefore present out-of-sample results, and we use two evaluation criteria. Denote the one step-ahead forecast provided by the model for security j by $\widehat{R}_{j,t+1}^{Model}$. In our implementation, which is recursive, this forecast uses information available up to time t. The first evaluation criterion is the mean of the squared forecast error, also used by Simin (2008), which is given by

$$RMSFE_{j,OS} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(R_{j,t+1} - \widehat{R}_{j,t+1}^{Model}\right)^2} \tag{26}$$

where T is the number of time periods in the sample. We can compute this measure for each individual portfolio j, but because of space constraints we report the average over the test portfolios. Our second evaluation criteria is adapted from the time-series literature. We use the out-of-sample R-square suggested by Campbell and Thompson (2008), which has become the standard in the time-series literature, see for instance Rapach and Zhou (2013). The out-of-sample $R_{j,OS}^2$ for a security j is defined by

$$R_{j,OS}^2 = 1 - \frac{\sum_t \left(R_{j,t} - \widehat{R}_{j,t+1}^{Model}\right)^2}{\sum_t \left(R_{j,t} - \overline{R}_{j,t-59:t}\right)^2} \tag{27}$$

where $\overline{R}_{j,t-59:t} = \frac{1}{60}\sum_{s=t-59}^{t} R_{j,s}$. This R-square can again be computed for every portfolio, but because of space constraints we report the average across portfolios for each model.

Note that this out-of-sample R-square uses the historical return on the test portfolio as a benchmark. If a candidate model performs as well as the historical return on the test portfolio, the resulting R-square will be zero. R-squares will be negative for models that do not perform well in forecasting. Consequently, the values of this out-of-sample R-square should not be confused with the R-squares one typically obtains from a cross-sectional or time-series regression, for example. In fact, R-squares can be expected to be very small, and a small positive R-square is an indicator of success. See Campbell and Thompson (2008), Rapach, Strauss, and Zhou (2010), and Rapach and Zhou (2013) for a detailed discussion.

We compare the cross-sectional performance of our newly proposed estimates of the price of co-skewness and co-kurtosis risk to a number of other specifications based on these two evaluation criteria. One set of specifications is based on historical risk premia, in the other one the risk premia are estimated using cross-sectional regressions. The models that use cross-sectional regressions to estimate the risk premia are the model with market covariance risk (the CAPM), the model with market covariance and co-skewness risk (CAPM+COSK), and the Fama-French three-factor model (FF). The specifications based on historical risk premia are: CAPM, COSK, and CAPM+COSK. We also include a hybrid approach CSCAPM+COKU, where the market risk premium is estimated using a cross-sectional regression.

To provide more intuition, consider the implementation of the two types of specifications using the CAPM as an example.

For the CAPM, the one step-ahead forecast of $\widehat{R}_{j,t+1}^{CAPM}$ using information available up to time t is

$$\widehat{R}_{j,t+1}^{CAPM} = \widehat{\lambda}_t^{mkt}\widehat{\beta}_{j,t}^{mkt} \tag{28}$$

The betas for both implementations are the same, and are obtained by regressing R_j on R_m, using a rolling window of 60 months from $t-59$ to t. However, estimates of the covariance price of risk, $\widehat{\lambda}_t^{mkt}$, are obtained in two ways. The first approach uses the sample mean of the market excess return over the past 60 months. The second approach is to estimate the price of risk using a cross-sectional regression:

$$R_{j,t} = \boldsymbol{\lambda}_t^{mkt}\,\widehat{\boldsymbol{\beta}}_j^{mkt} + u_{j,t}, \quad j = 1, ...N \tag{29}$$

Note that in principle we can at each time t use this price of risk $\boldsymbol{\lambda}_t$ to construct the forecast of $\widehat{R}_{j,t+1}^{CAPM}$. However, we found that this leads to extremely poor forecasts, which is due to the time variation in these cross-sectional estimates, as evidenced by the estimates for co-skewness in Figure 2. To provide better out-of-sample competitors for our estimators of co-skewness that use historical risk premia, we therefore use averages of the cross-sectional averages of $\boldsymbol{\lambda}_t$ for the past 60 months, which provided better forecasts, and which is more in line with the conventional (in-sample) implementation of Fama-MacBeth regressions.

Table 3 presents results for the same four sets of test portfolios used in Table 2. Panel A presents the out-of-sample R-square R_{OS}^2, and Panel B presents the out-of-sample RMSFEs. Consider the out-of-sample RMSFEs in Panel B, which we have multiplied by 100, following the convention adopted by Simin (2008). To interpret these numbers, note that if the forecast is the historical average, the magnitude should be similar to a monthly volatility. For a stock with 30% annual volatility, the monthly volatility is 8.66%. The second and third columns present results that are obtained using our newly proposed estimates of the price of co-skewness and co-kurtosis risk. The co-skewness based forecasts, in column 2, provide the lowest forecast errors for all four sets of test portfolios.

The out-of-sample RMSFEs provide a useful ranking of the models, but it takes some effort to interpret the magnitudes. The out-of-sample R-square R_{OS}^2 evaluation criterion is perhaps easier to understand intuitively. Panel A of Table 3 presents the results. Recall that a positive out-of-sample R-square means that the model forecasts better than the historical average return on the asset.

The performance of our newly proposed co-skewness measure $COSK$ in the second column of the top four rows is impressive. It yields a positive R-square for all four sets of test portfolios. The out-of-sample performance of the other models is mixed. Arguably the best competitor is the regression-based implementation of the CAPM, but this model does poorly for the twenty-five size and book-to-market portfolios. The out-of-sample performance of the Fama-French model is disappointing. It may seem surprising that the FF model performs so poorly for the case of the 25 size and book-to-market portfolios, but note that the FF model is not typically evaluated in a genuine out-of-sample setting.

It is important to keep in mind that in a genuine out-of-sample setting, these very small positive R-squares are economically meaningful. This criterion is typically used in the time-series literature, and even there R-squares of 1-2% are the exception rather than the rule, with many candidate forecasts yielding negative R-squares, see Campbell and Thompson (2008), Rapach, Strauss, and Zhou (2010), and Rapach and Zhou (2013). The performance of the newly proposed estimate of the price of co-skewness risk is therefore impressive, especially because forecasting with a cross-sectional model is even harder than time series forecasting.

3.6 Robustness

We now report on several robustness exercises, using alternative measures of conditional physical and risk-neutral second moments.

We used the VIX as our measure of the risk-neutral second moment in our benchmark results. In the robustness analysis we use an alternative approach to compute the risk-neutral variance, following Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003). This approach requires a continuum of out-of-the money call and put options which is approximated using cubic spline interpolation techniques. See the appendix for more details. We implement this approach using data on S&P500 index options from OptionMetrics for the period January 1996 to December 2012. We use the implied volatility estimates reported in OptionMetrics to approximate a continuum of implied volatilities which are in turn converted to a continuum of prices. For strike prices outside the range available, we simply use the implied volatility of the lowest or highest available strike price.

Following standard practice, we filter out options that (i) violate no-arbitrage conditions; (ii) have missing or extreme implied volatility (larger than 200% or lower than 0.01%); (iii) with open-interest or bid price equal to zero; and (iv) have a bid-ask spread lower than the minimum tick size, i.e., bid-ask spread below \$0.05 for options with prices lower than \$3 and bid-ask spread below \$0.10 for option with prices equal or higher than \$3.

We investigate three alternative approaches for modeling the conditional physical variance. We first consider a simple autoregressive model on realized variances. The one-step ahead forecast of the physical second moment is estimated from the following monthly regression

$$\kappa_t^2 = a_0 + a_1 \kappa_{t-1}^2 + u_t^\sigma, \tag{30}$$

where $\kappa_t^2 = \sum_{d \in t} R_{d,t}^2$, and $R_{d,t}$ is the daily return in day d of month t.

In addition to the autoregressive model we also use an NGARCH model (Engle and Ng, 1993) to estimate the physical variance

$$\begin{aligned} R_t &= h_t z_t \qquad\qquad z_t \sim N(0,1) && (31) \\ h_t^2 &= a_0 + b_0 h_{t-1}^2 \left(z_{t-1} - d_0\right)^2 + c_0 h_{t-1}^2. && (32) \end{aligned}$$

The T-day ahead forecast can be computed as follows

$$E_t[R^2_{t+1:t+T}] = Th_0^2 + (h^2_{t+1} - h_0^2)\frac{1-(b_0+c_0+b_0d_0^2)^T}{1-b_0-c_0-b_0d_0^2}, \tag{33}$$

where $h_0^2 = \frac{a_0}{1-b_0-c_0-b_0d_0^2}$. Finally we also use the Heston (1993) stochastic volatility model in which the underlying stock price S_t is given by

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t}dW_{S,t}, \tag{34}$$

and the instantaneous variance is

$$dv_t = \kappa(\theta - v_t)dt + \eta\sqrt{v_t}dW_{v,t}, \tag{35}$$

where $W_{S,t}$ and $W_{v,t}$ are two correlated Brownian motion processes with $dW_{v,t}dW_{S,t} = \rho dt$. We estimate this model using the particle filter.

Table 4 presents the results. Panel A contains the estimates of the price of risk obtained using the different approaches. Panels B and C contain the out-of-sample results. To save space, we limit ourselves to the out-of-sample R-squares R^2_{OS}.

The estimates of the price of risk in Panel A vary between -0.123 and -0.316. Recall that our benchmark estimate in Table 1 was -0.271. These estimates are quite similar and they are all larger (in absolute value) than the cross-sectional estimates in Table 2. The out-of-sample R-squares in Panel B are positive in twenty-six out of twenty-eight cases, which is quite impressive. We conclude that our newly proposed estimates of the price of risk are rather robust across different empirical implementations, and that the resulting out-of-sample performance is much better than that of regression-based implementations of models with co-skewness risk as well as competing models.

4 Estimating the Price of Co-Skewness and Co-Kurtosis Risk

We now provide estimates of the price of co-skewness and co-kurtosis risk using the estimators presented in Propositions 1 and 2. For the price of co-skewness risk, we need to estimate the second conditional moment under the risk-neutral measure, $E_t^Q(R^2_{m,t+1})$, and under the physical measure, $E_t^P(R^2_{m,t+1})$,just as in Section 3. For the price of co-kurtosis risk, we need to estimate the third conditional moment under the risk-neutral measure $E_t^Q(R^3_{m,t+1})$ and under the physical measure $E_t^P(R_{m,t+1}^3)$.

It is important to realize the main differences between this empirical exercise and the one in Section 3. Most importantly, estimating third moments is harder than estimating second moments. This is the main reason that we first provide estimates of co-skewness risk using methods that do not require us to model the third moments. When considering the estimation of risk-neutral and physical conditional third moments, the question then arises which of these tasks is most challenging. Perhaps somewhat surprisingly, the modeling of the physical third moment is relatively more difficult.

4.1 Modeling Risk-Neutral and Physical Skewness

We estimate risk-neutral variance and skewness using the method of Bakshi and Madan (2000), as explained in the appendix. We use data on S&P500 index options from OptionMetrics for the period from January 1996 to December 2012. We use the data filters discussed in Section 3.6.

For the physical moments, we want to impose internal consistency and obtain estimates of the physical conditional second and third moment using the same model. The estimation of conditional higher moments is notoriously difficult. We use a version of the Jondeau and Rockinger (2003) model. Our implementation is close to the model they refer to as Model 2, which is among the more parsimonious models they consider. We found this model converged well in estimation and for our purposes it is sufficiently richly parameterized. We implement this model using monthly data. The model is given by

$$R_{m,t} = h_t z_t \qquad\qquad z_t \sim GT(z_t|\eta_t, \lambda_t),$$

where $R_{m,t}$ is the return on the market in month m, GT denotes the generalized student-t distribution, and where the higher-moment dynamics are modeled via

$$\begin{aligned}
h_t^2 &= a_0 + b_0^+ \left(R_{m,t-1}^+\right)^2 + b_0^- \left(R_{m,t-1}^-\right)^2 + c_0 h_{t-1}^2, \\
\widetilde{\eta}_t &= a_1 + b_1^+ R_{m,t-1}^+ + b_1^- R_{m,t-1}^-, \\
\widetilde{\lambda}_t &= a_2 + b_2^+ R_{m,t-1}^2, \\
\eta_t &= g_{]2,+30]}\left(\widetilde{\eta}_t\right), \text{ and } \lambda_t = g_{]-1,1]}\left(\widetilde{\lambda}_t\right)
\end{aligned}$$

where $R_m^+ = \max(R_m, 0)$ and $R_m^- = \max(-R_m, 0)$. The logistic map

$$g_{]x_L,x_U]}(x) = x_L + \frac{x_U - x_L}{1+\exp(-x)}$$

ensures that $2 < \eta_t < \infty$ and $-1 < \lambda_t < 1$, which are necessary conditions for the existence of the GT distribution. Note that we have set the conditional mean return to zero here because it is difficult to model and unlikely to matter much for the dynamics of higher moments.

The density of Hansen's (1994) GT distribution is defined by

$$GT(z_t|\eta_t, \lambda_t) = \begin{cases} b_t c_t \left(1 + \frac{1}{\eta_t - 2}\left(\frac{b_t z_t + a_t}{1-\lambda_t}\right)^2\right)^{-(\eta_t+1)/2} & \text{if } z_t < -a_t/b_t, \\ b_t c_t \left(1 + \frac{1}{\eta_t - 2}\left(\frac{b_t z_t + a_t}{1+\lambda_t}\right)^2\right)^{-(\eta_t+1)/2} & \text{if } z_t \geq -a_t/b_t, \end{cases}$$

where

$$a_t \equiv 4\lambda_t c_t \frac{\eta_t - 2}{\eta_t - 1}, \quad b_t \equiv 1 + 3\lambda_t^2 - a_t^2, \quad c_t \equiv \frac{\Gamma\left(\left(\eta_t + 1\right)/2\right)}{\sqrt{\pi\left(\eta_t - 2\right)}\Gamma\left(\eta_t/2\right)}.$$

We need the non-centered second and third conditional moments, which can be computed as follows

$$E_t^P\left[R_{m,t+1}^2\right] = h_{t+1}^2,$$

and

$$E_t^P\left[R_{m,t+1}^3\right] = h_{t+1}^3\left[m_{3,t+1} - 3a_{t+1}m_{2,t+1} + 2a_{t+1}^3\right]/b_{t+1}^3.$$

where

$$m_{2,t} = 1 + 3\lambda_t^2, \quad m_{3,t} = 16c_t\lambda_t\left(1+\lambda_t^2\right)\frac{\left(\eta_t - 2\right)^2}{\left(\eta_t - 1\right)\left(\eta_t - 3\right)},$$

Note that the third moment exists in the model so long as $\eta_t > 3$.

Because estimation of conditional higher moments is difficult, we conducted an extensive robustness analysis. We computed the physical conditional second and third moments from the alternative model in Leon, Rubio, and Serna (2005), and the resulting estimates are very similar. For our purpose, both the Jondeau and Rockinger (2003) and Leon, Rubio, and Serna (2005) approaches have the advantage that estimates of the conditional third and second moment are internally consistent. If, as in Section 3, we limit ourselves to co-skewness, we only require the physical second moment, for which more straightforward estimation techniques are available.

4.2 The Price of Co-Kurtosis Risk

The price of co-skewness and co-kurtosis risk for month t can now simply be computed as

$$\begin{aligned}\widehat{\lambda}_t^{COSK} &= \widehat{E}_t^P(R_{m,t+1}^2) - \widehat{E}_t^Q(R_{m,t+1}^2)\\ \widehat{\lambda}_t^{COKU} &= \widehat{E}_t^P(R_{m,t+1}^3) - \widehat{E}_t^Q(R_{m,t+1}^3)\end{aligned}$$

Table 5 reports descriptive statistics for the relevant moments and the estimated prices of risk. Figure 4 depicts the time series of the prices of co-skewness and co-kurtosis risk with the corresponding estimated physical and risk-neutral moments required to compute these prices. As in Figure 1, the figures show spikes surrounding the 1998 LTCM collapse, the WorldCom bankruptcy in 2002, and the subprime crisis. These spikes occur for both the second and third moments, and for risk-neutral as well as physical moments.

Just as in Figure 1, the co-skewness risk premium in Figure 4 is negative for almost all months. Also consistent with theory, the co-kurtosis risk premium in Figure 4 is positive for almost all months, but following a sharp peak during the recent financial crisis, it briefly turns negative. Table 5 reports that on average the co-skewness price of risk is equal to -0.274, remarkably close to the average in Table 1. The price of co-kurtosis risk is equal to 0.0116 on average. Existing empirical studies have also documented positive prices of co-kurtosis risk, see for instance Ang, Chen, and Xing (2006) who find that stocks with higher co-kurtosis earn higher returns.

Figures 5 and 6 report on estimates of co-kurtosis risk obtained using Fama-MacBeth regressions. Figure 5 indicates that the month-by-month estimates of the rice of co-kurtosis risk vary significantly over time, and that they are often negative. Compared to the time series of the prices of co-skewness risk in Figure 3, the time series of the prices of co-kurtosis risk are less correlated across test assets. When averaging over time, the estimate is significantly negative for the twenty-five size and momentum portfolios. This is also the case for the univariate regressions in Panel A of Figure 6 However, Panel B of Figure 6 indicates that this may be due to the relatively short time period. When using the longer 1966-2012 time period, all four estimates of co-skewness risk are positive, although not always significant.

For the 1996-2012 sample period in Panel A of Figure 6, only one set of test portfolios yields a statistically significant positive result, the twenty-five size and short term reversal portfolios. The resulting estimate of the price of co-skewness risk is 0.020, of the same order of magnitude as our new estimate of 0.0116 in Table 5. The estimates obtained for the

1966-2012 sample period in Panel B of Figure 6 are also of the same order of magnitude, but somewhat larger. We conclude that our new estimates of the price of co-kurtosis risk are more similar to regression-based estimates compared to the estimates of co-skewness risk.

Table 6 presents out-of-sample R-squares and root mean squared forecast errors using these estimates of the prices of co-skewness and co-kurtosis risk, and compares the resulting fit with the fit of regression-based approaches.

Because the estimates of the price of co-skewness risk are similar to the ones obtained in Section 3.3, it is not surprising that the resulting R-squares and RMSFEs are similar to the ones in Table 3. The model with co-kurtosis risk does not do as well as the model with co-skewness risk only, but it performs better than the CAPM implemented with historical risk premia.

Table 7 reports on several robustness exercises. Unlike in the case of the physical second moment which is used to compute co-skewness risk, the literature does not contain a wealth of alternatives to model the physical third moment. As previously mentioned, we implemented the model of Leon, Rubio, and Serna (2005) and obtained similar results. Because of the inherent difficulty in modeling the physical third moment, we also implemented some alternatives that are much simpler. Table 7 reports on three alternatives for physical skewness: a constant skewness computed using daily data, a constant skewness computed using monthly data, and finally the case of zero skewness, which serves as a sanity check.

The results in Table 7 indicate that the resulting price of co-kurtosis risk is similar to the one in Table 5. The out-of-sample performance of the co-kurtosis model is substantially better than the one in Table 6. This suggests that our estimates of the conditional physical third moment may be too noisy, thus affecting the out-of-sample fit.

5 Conclusion

We propose an alternative strategy to estimate the price of possibly nonlinear exposures to market risk, which avoids the errors inherent in the cross-sectional regression approach. The key difference between our approach and existing studies is that we explicitly impose the resulting pricing restrictions on both stocks and contingent claims. We study two important applications of our general approach. The price of co-skewness risk in our framework corresponds to the spread between the physical and the risk-neutral second moment. The price of co-kurtosis risk is similarly given by the spread between the physical and the risk-neutral third moment.

Using monthly data for the period 1996-2012, we find that the price of co-skewness risk has the theoretically expected negative sign in each month, and the price of co-kurtosis risk has the theoretically expected positive sign in most months. In contrast, the prices of risk obtained using regression-based approaches do not always have the theoretically anticipated signs on average. Our approach also provides genuinely conditional estimates of the price of risk at monthly or even higher frequencies. When using a regression-based approach, monthly estimates are available, but they are very imprecise, and they are therefore usually averaged over a large number of months. An out-of-sample analysis of factor models with co-skewness and co-kurtosis risk indicates that the new estimates of the price of risk improve the models' performance. The models also robustly outperform competitors such as the CAPM and the Fama-French model.

Some questions remain, and a number of extensions could prove interesting. First, while the estimated price of co-skewness risk leads to a more than satisfactory out-of-sample cross-sectional fit when used by itself, its performance is worse when combined with the CAPM risk factor. It may prove useful to further investigate the resulting biases. Second, the focus of this paper is on improving measurement. While we believe that our measure of the price of co-skewness risk improves on existing techniques, we worry that the estimated betas we use in the analysis may be noisy. Improved estimation of betas may be worth exploring, and may lead to better out-of-sample performance. The estimation approach proposed by Bali and Engle (2010) may be especially promising in this regard.

Appendix: Extracting Option Implied Moments

Bakshi and Madan (2000) show that any twice-continuously differentiable payoff function, $H[S]$, can be spanned by a portfolio of risk-free bonds, the underlying asset and out-of-the-money calls and puts as follows

$$H[S] = H\left[\overline{S}\right] + \left(S - \overline{S}\right) H_S\left[\overline{S}\right] + \int_{\overline{S}}^{\infty} H_{SS}[K](S-K)^+ dK$$
$$+ \int_0^{\overline{S}} H_{SS}[K](K-S)^+ dK. \tag{36}$$

The prices of these contracts are

$$E_t^Q\left\{e^{-r\tau}H[S]\right\} = \left(H\left[\overline{S}\right] - \overline{S}H_S\left[\overline{S}\right]\right)e^{-r\tau} + H_S\left[\overline{S}\right]S(t)$$
$$+ \int_{\overline{S}}^{\infty} H_{SS}[K]\,C(t,\tau;K)\,dK + \int_0^{\overline{S}} H_{SS}[K]\,P(t,\tau;K)\,dK. \tag{37}$$

where $C_t(\tau, K)$ and $P_t(\tau, K)$ are prices of the European call and put options with time-to-maturity τ and strike price K. As a result, we can calculate the prices of derivatives whose payoffs only depend on the future S, given the prices of (i) the risk free zero coupon bond, r, (ii) the current value of the underlying stock, $\overline{S}$, and (iii) a series of OTM calls and puts.

For our purposes, let $R(t,\tau) = \ln S(t+\tau) - \ln S(t)$, and first consider the function

$$H[S_{t+\tau}] = R_{t+\tau}^2 = (\ln S_{t+\tau} - \ln S_t)^2 \tag{38}$$

Using this, we can get the risk-neutral raw second moment via

$$\begin{aligned} E_t^Q\left[R_{t+\tau}^2\right] &= e^{r\tau}\int_{S_t}^{\infty}\frac{2\left(1-\ln\left[K/S_t\right]\right)}{K^2}C_t(\tau,K)\,dK \\ &\quad + e^{r\tau}\int_0^{S_t}\frac{2\left(1+\ln\left[S_t/K\right]\right)}{K^2}P_t(\tau,K)\,dK. \end{aligned}$$

Now, let

$$H[S_{t+\tau}] = R_{t+\tau}^3 = (\ln S_{t+\tau} - \ln S_t)^3 \tag{39}$$

then we get the option-implied raw third moment via

$$
\begin{aligned}
E_t^Q\left[R_{t+\tau}^3\right] = & \; e^{r\tau}\int_{S_t}^{\infty}\frac{6\ln\left[K/S_t\right]-3\left(\ln\left[K/S_t\right]\right)^2}{K^2}C_t\left(\tau,K\right)dK \\
& -e^{r\tau}\int_{0}^{S_t}\frac{6\ln\left[S_t/K\right]+3\left(\ln\left[S_t/K\right]\right)^2}{K^2}P_t\left(\tau,K\right)dK.
\end{aligned}
$$

When computing these moments, we eliminate put options with strike prices of more than 105% of the underlying asset price ($K/S > 1.05$) and call options with strike prices of less than 95% of the underlying asset price ($K/S < 0.95$). We only estimate the moments for days that have at least two OTM call prices and two OTM put prices available.

Since we do not have a continuity of strike prices, we calculate the integrals using cubic splines. For each maturity, we interpolate implied volatilities using a cubic spline across moneyness levels (K/S) to obtain a continuum of implied volatilities. For moneyness levels below or above the available moneyness level in the market, we use the implied volatility of the lowest or highest available strike price. After implementing this interpolation-extrapolation technique, we obtain a fine grid of implied volatilities for moneyness levels between 1% and 300%. We then convert these implied volatilities into call and put prices using the following rule: moneyness levels smaller than 100% ($K/S < 1$) are used to generate put prices and moneyness levels larger than 100% ($K/S > 1$) are used to generate call prices using trapezoidal numerical integration. Linear interpolation between maturities is used to calculate the moments for a fixed 30-day horizon.

References

[1] Ang, A., J. Chen, and Y. Xing (2006), Downside Risk, *Review of Financial Studies*, 19, 1191–1239.

[2] Ang, A., R. Hodrick, Y. Xing, and X. Zhang (2006), The Cross-Section of Volatility and Expected Returns, *Journal of Finance*, 61, 259–299.

[3] Arditti, F. (1967), Risk and the Required Returns on Equity, *Journal of Finance*, 22, 19–36.

[4] Bakshi, G., N. Kapadia, and D. Madan (2003), Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *Review of Financial Studies*, 16 (1), 101–143.

[5] Bakshi, G., and D. Madan (2000), Spanning and Derivative Security Valuation, *Journal of Financial Economics*, 55, 205–238.

[6] Bakshi, G., and D. Madan (2006), A Theory of Volatility Spreads, *Management Science*, 52, 1945–1956.

[7] Bakshi, G., D. Madan, and G. Panayotov, (2010), Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels, *Journal of Financial Economics*, 97, 130–154.

[8] Bali, T. G., O. Demirtas, and H. Levy (2009), Is There an Intertemporal Relation between Downside Risk and Expected Returns? *Journal of Financial and Quantitative Analysis*, 44, 883–909.

[9] Bali, T. G., and R. F. Engle (2010), The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations, *Journal of Monetary Economics*, 57, 377–390.

[10] Bansal, R., and S. Viswanathan (1993), No-Arbitrage and Arbitrage Pricing: A New Approach, *Journal of Finance*, 48, 1231–1262.

[11] Black, F., and M. Scholes (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637–654.

[12] Bondarenko, O. (2003), Why Are Put Options So Expensive? Working Paper, University of Illinois, Chicago.

[13] Bollerslev, T., G. Tauchen, and H. Zhou (2009), Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies*, 22, 4463–4492.

[14] Broadie, M., M. Chernov, and M. Johannes (2009), Understanding Index Option Returns, *Review of Financial Studies*, 22 (11), 4493–4529.

[15] Campbell, J. Y., and S. B. Thompson (2008), Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial Studies*, 21, 1509–31.

[16] Carhart, M. (1997), On Persistence of Mutual Fund Performance, *Journal of Finance*, 52, 57–82.

[17] Carr, P., and L. Wu (2009), Variance Risk Premia, *Review of Financial Studies*, 22, 1311–1341.

[18] Chabi-Yo, F. (2008), Conditioning Information and Variance Bounds on Pricing Kernels with Higher-Order Moments: Theory and Evidence, *Review of Financial Studies*, 21, 181–231.

[19] Chan, L. K.C., J. Karceski, and J. Lakonishok (2003), The Level and Persistence of Growth Rates, *Journal of Finance*, 58, 643–684.

[20] Cochrane, J. (2005), *Asset Pricing*, Princeton University Press, Princeton, NJ.

[21] Corsi, F. (2009), A Simple Approximate Long Memory Model of Realized Volatility, *Journal of Financial Econometrics*, 7, 174–196.

[22] Coval, J. D., and T. Shumway (2001), Expected Option Returns, *Journal of Finance*, 56, 983–1009.

[23] Dittmar, R. F. (2002), Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns, *Journal of Finance*, 57, 369–403.

[24] Driessen, J. , P. J. Maenhout, and G. Vilkov (2009), The Price of Correlation Risk: Evidence from Equity Options, *Journal of Finance*, 64, 1377–1406.

[25] Engle, R., and V. Ng (1993), Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, 48, 1749–1778.

[26] Fama, E. F., and K. R. French (1993), Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33, 3–56.

[27] Fama, E. F., and J. D. MacBeth (1973), Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy*, 81, 607–636.

[28] Ferson, W., S. Nallareddy, and B. Xie (2012), The "Out-of-Sample" Performance of Long Run Risk Models, *Journal of Financial Economics*, 107, 537–556.

[29] Golec, J., and M. Tamarkin (1998), Bettors Love Skewness, Not Risk, at the Horse Track, *Journal of Political Economy*, 106, 205–225.

[30] Guidolin, M., and A. Timmermann (2008), International Asset Allocation under Regime Switching, Skew, and Kurtosis Preferences, *Review of Financial Studies*, 21, 889–935.

[31] Hansen, B. (1994), Autoregressive Conditional Density Estimation, *International Economic Review*, 35, 705-730.

[32] Harvey, C. R., and A. Siddique (2000), Conditional Skewness in Asset Pricing, *Journal of Finance*, 55 (3), 1263–1295.

[33] Jackwerth, J., and M. Rubinstein (1996), Recovering Probability Distributions from Option Prices, *Journal of Finance* 51, 1611–1631.

[34] Jondeau, E., and M. Rockinger (2003), Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements, *Journal of Economic Dynamics and Control*, 27, 1699–1737.

[35] Kraus, A., and R. H. Litzenberger (1976), Skewness Preference and the Valuation of Risk Assets, *Journal of Finance*, 31, 1085–1100.

[36] Leland, H. (1997), Beyond Mean-Variance: Performance Measurement of Portfolios Using Options or Dynamic Strategies, Working paper, University of California, Berkeley.

[37] León, A., G. Rubio, and G. Serna (2005), Autoregressive Conditional Volatility, Skewness and Kurtosis, *The Quarterly Review of Economics and Finance*, 45, 599–618.

[38] Lewellen, J., S. Nagel, and J. Shanken (2010), A Skeptical Appraisal of Asset Pricing Tests, *Journal of Financial Economics*, 96, 175–194.

[39] Liew, J., and M. Vassalou (2000), Can Book-to-Market, Size, and Momentum Be Risk Factors that Predict Economic Growth?, *Journal of Financial Economics*, 57, 221–245.

[40] Lim, K.-G. (1989), A New Test of the Three-Moment Capital Asset Pricing Model, *Journal of Financial and Quantitative Analysis,* 24, 205-216.

[41] Rapach, D. E., J. K. Strauss, and G. Zhou, (2010), Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, *Review of Financial Studies*, 23, 821–862.

[42] Rapach, D. E., and G. Zhou (2013), Forecasting Stock Returns, in the *Handbook of Economic Forecasting*, Volume 2, edited by G. Elliott and A. Timmermann, Elsevier: San-Diego, pp. 328–383.

[43] Rubinstein, M. (1973), The Fundamental Theory of Parameter-Preference Security Valuation, *Journal of Financial and Quantitative Analysis*, 8, 61–69.

[44] Santa-Clara, P., and A. Saretto (2009), Option Strategies: Good Deals and Margin Calls, *Journal of Financial Markets*, 12, 391–417.

[45] Simin, T. (2008), The Poor Predictive Performance of Asset Pricing Models, *Journal of Financial and Quantitative Analysis*, 43, 355-380.

[46] Scott, R. C., and P. A. Horvath (1980), On the Direction of Preference for Moments of Higher Order than the Variance, *Journal of Finance*, 35, 915–919.

Figure 1: The Option-Based Price of Co-Skewness Risk

We plot the times series for the conditional physical and risk-neutral second moments (monthly in percentage) and the price of co-skewness risk. The physical second moment is estimated using an HAR model and the risk-neutral second moment is proxied by the VIX. The time-varying price of co-skewness risk is equal to the spread between the physical and risk-neutral moments. The sample period is from January 1986 to December 2012.

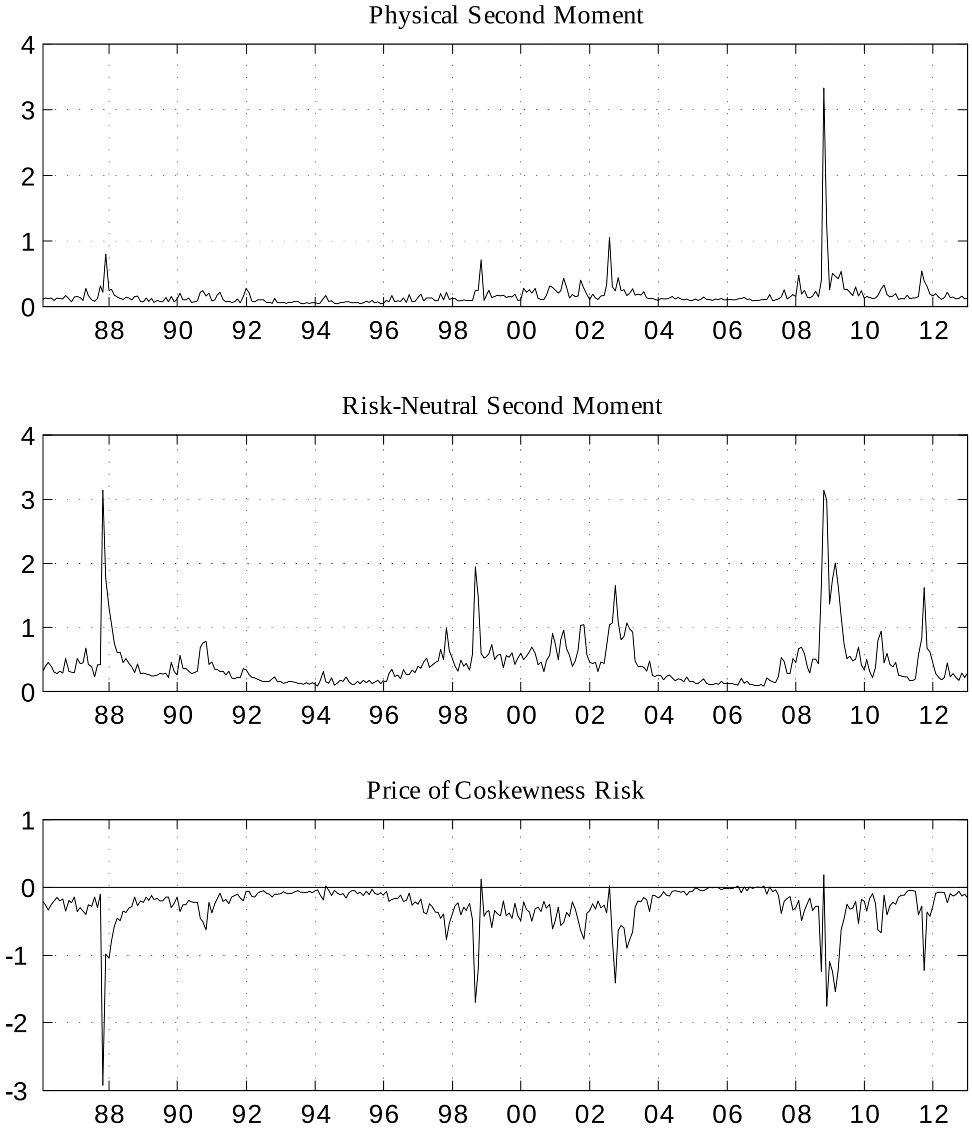
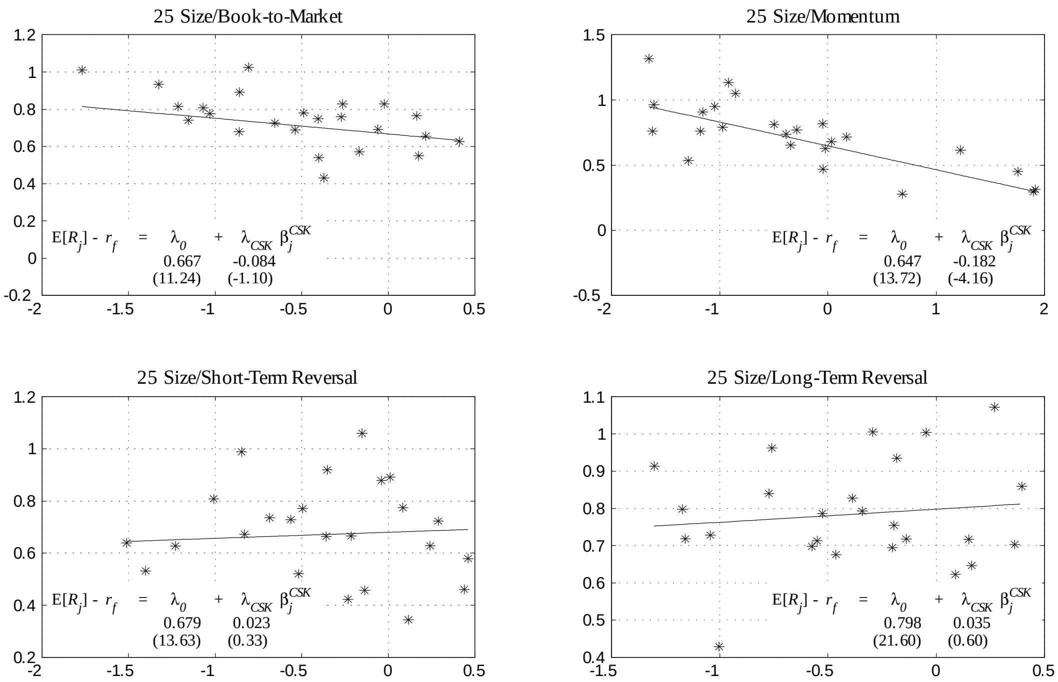


Figure 2: The Cross-Section of Returns and Co-Skewness Betas

We plot average excess returns (monthly, in percentages), $E[R_{j,t}] - r_f$, against co-skewness betas, β_j^{COKU}, for four sets of portfolios. The co-skewness beta, β_j^{COSK}, is computed from the regression of monthly excess returns on market returns and squared market returns. We consider two periods, 1986-2012 and 1966-2012, and four sets of test portfolios.

Panel A. 1986 - 2012



Panel B. 1966 - 2012

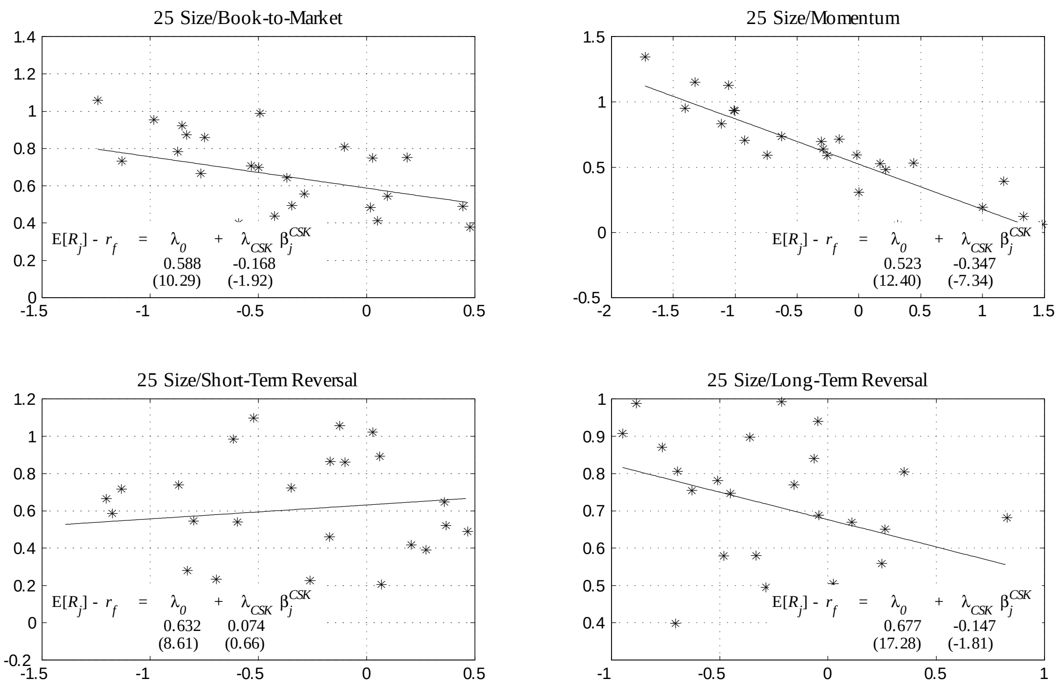


Figure 3: Regression-Based Estimates of the Price of Co-Skewness Risk

We plot the time series for the cross-sectional prices of co-skewness risk. Each month, we estimate the co-skewness beta using a 60-month rolling window of monthly returns from the following time series regression

$$R_{j,t} - r_f = \alpha_j + \beta_{j,t}^{MKT} R_{MKT,t} + \beta_{j,t}^{COSK} R_{MKT,t}^2 + \varepsilon_{j,t}.$$

We then run the following cross-sectional regression using the estimated betas and returns for the next month

$$R_{j,t+1} - r_f = \lambda_0 + \beta_{j,t}^{MKT} \lambda_{MKT} + \beta_{j,t}^{COSK} \lambda_{COSK} + e_{j,t}.$$

We consider four sets of test portfolios: 25 Size/Book-to-Market, 25 Size/Momentum, 25 Size/Short-Term Reversal and 25 Size/Long-Term Reversal. The sample period is from January 1986 through December 2012.

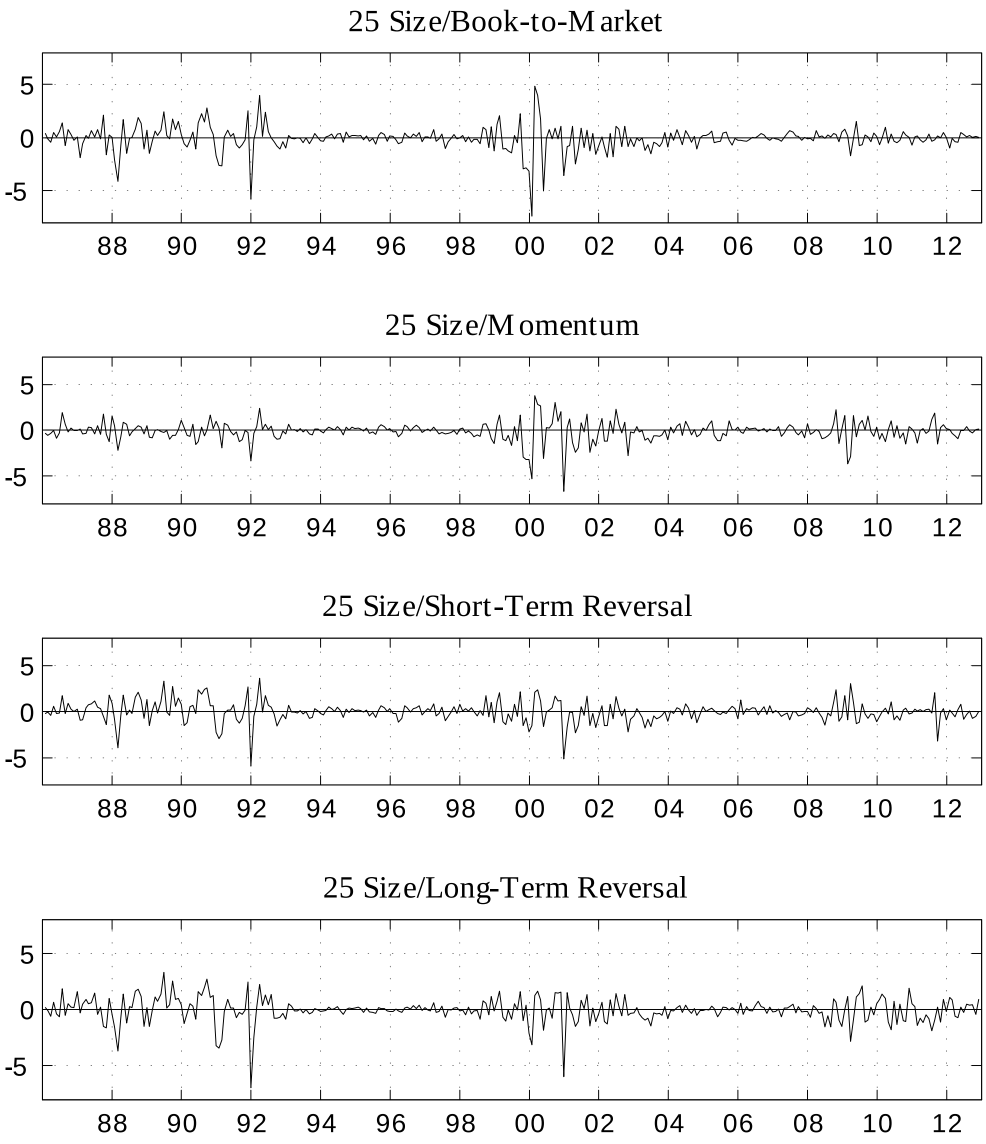


Figure 4: The Option-Based Price of Co-Skewness and Co-Kurtosis Risk

We plot the time series for the conditional physical and risk-neutral moments (in percentage per month). We also plot the price of co-skewness and co-kurtosis risk. The physical moments are estimated using the autoregressive conditional volatility, skewness, and kurtosis model of Jondeau and Rockinger (2003). The risk-neutral moments are estimated using the model-free approach in Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003). The sample period is from January 1996 to December 2012.

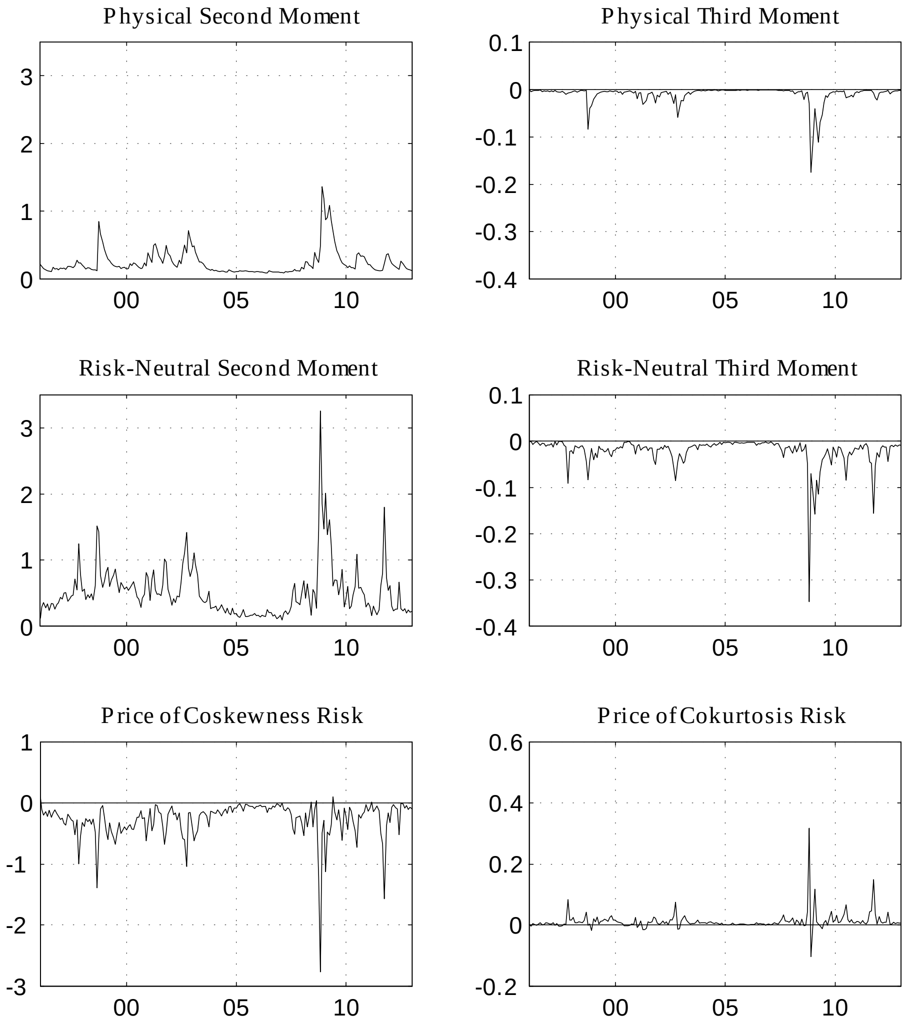


Figure 5: Regression-Based Estimates of the Price of Co-Kurtosis Risk

We plot the time series for the cross-sectional prices of co-kurtosis risk. Each month, we estimate the co-kurtosis beta using a 60-month rolling window of monthly returns from the following time series regression

$$R_{j,t} - r_f = \alpha_j + \beta_{j,t}^{MKT} R_{MKT,t} + \beta_{j,t}^{COSK} R_{MKT,t}^2 + \beta_{j,t}^{COKU} R_{MKT,t}^3 + \varepsilon_{j,t}.$$

We then run the following cross-sectional regression using the estimated betas and returns for the next month

$$R_{j,t+1} - r_f = \lambda_0 + \beta_{j,t}^{MKT} \lambda_{MKT} + \beta_{j,t}^{COSK} \lambda_{COSK} + \beta_{j,t}^{COKU} \lambda_{COKU} + e_{j,t}.$$

We consider four sets of portfolios: 25 Size/Book-to-Market, 25 Size/Momentum, 25 Size/Short-Term Reversal and 25 Size/Long-Term Reversal. The sample period is from January 1996 to December 2012.

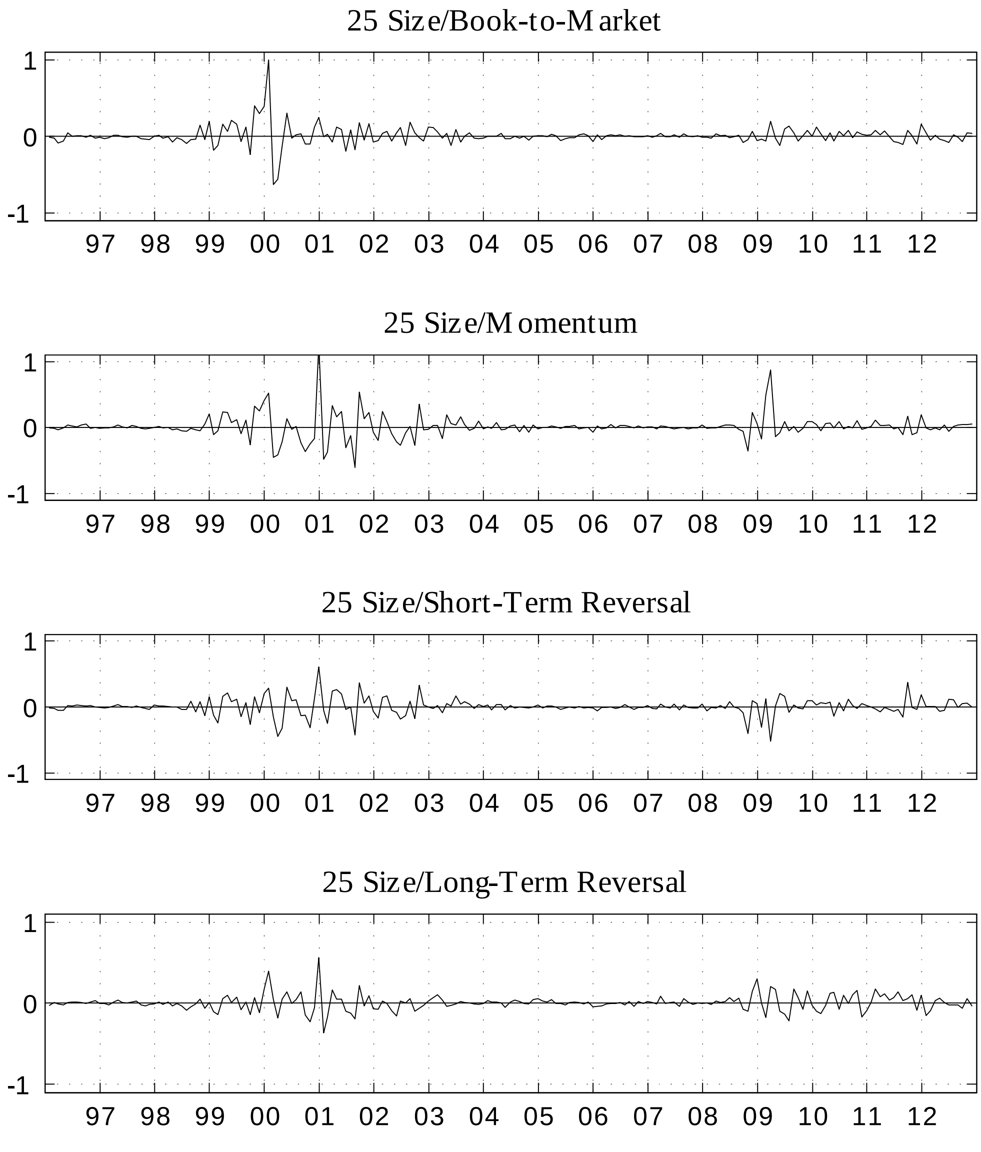
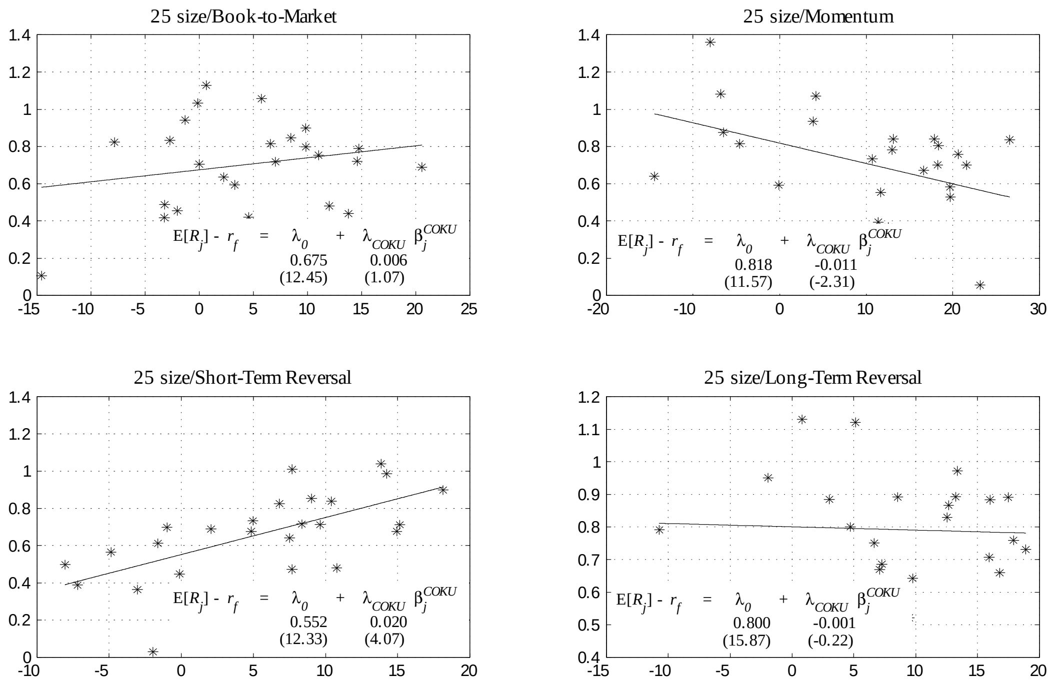


Figure 6: The Cross-Section of Returns and Co-Kurtosis Betas

We plot average excess returns (in percentages per month), $E[R_{j,t}] - r_f$, against co-skewness betas, β_j^{COKU}, for four sets of portfolios. The co-skewness beta, β_j^{COKU}, is computed from the regression of monthly excess returns on market returns, squared market returns and cubic market returns. We consider two periods, 1996-2012 and 1966-2012, and four sets of test portfolios.

Panel A. 1996 - 2012



Panel B. 1966 - 2012

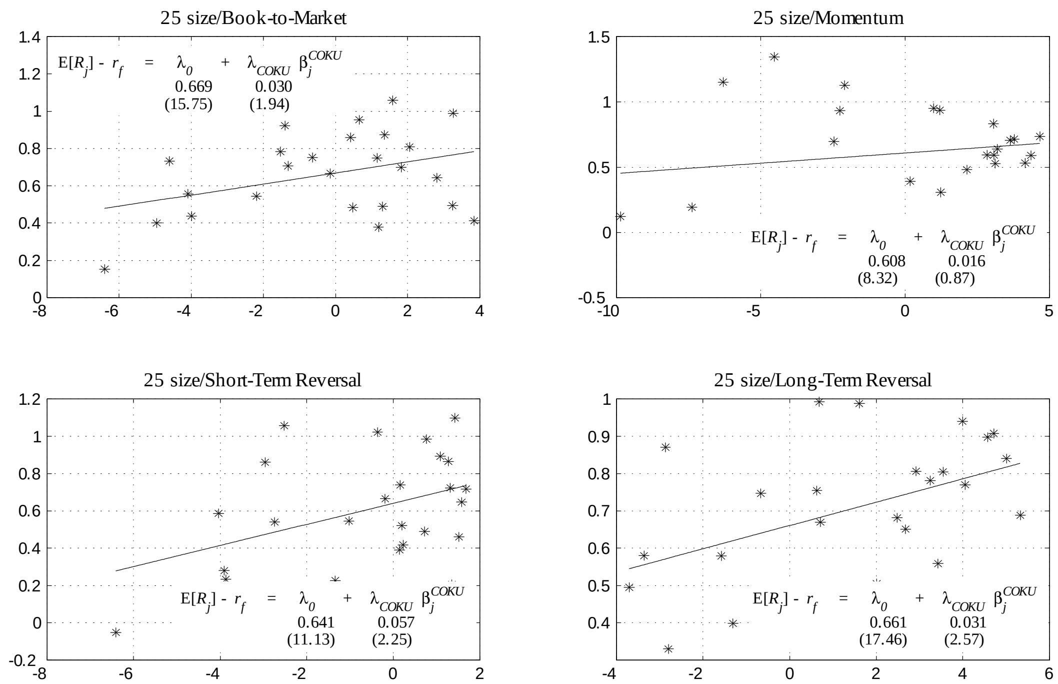


Table 1: The Option-Based Price of Co-Skewness Risk

The table provides descriptive statistics for the physical and risk-neutral expectations and the price of co-skewness risk. The data are monthly.. The physical second moment is estimated using an HAR model and the risk-neutral second moment is proxied by the VIX. The time-varying price of co-skewness risk is equal to the spread between the physical and risk-neutral moments. The data are monthly and the sample period is from January 1986 to December 2012.

	$E_t^P[R_{t+1}^2]$	$E_t^Q[R_{t+1}^2]$	$E_t^P[R_{t+1}^2] - E_t^Q[R_{t+1}^2]$
mean	0.1675	0.4381	-0.2707
std	0.2158	0.4133	0.3105
skew	10.5092	3.4318	-3.5761
kurt	145.8792	19.0463	23.1180
$\rho(1)$	0.4013	0.7525	0.4949

Table 2: Regression-Based Estimates of the Price of Co-Skewness Risk

The table shows the results of cross-sectional Fama-MacBeth regressions using monthly returns. Each month, we estimate the betas using a 60-month rolling window of monthly returns from a time series regression of the following general form

$$\begin{aligned} R_{j,t} - r_f &= \alpha_j + \beta_{j,t}^{MKT} R_{MKT,t} + \beta_{j,t}^{HML} R_{HML,t} + \beta_{j,t}^{SMB} R_{SMB,t} + \beta_{j,t}^{MOM} R_{MOM,t} \\ &\quad + \beta_{j,t}^{COSK} R_{MKT,t}^2 + \varepsilon_{j,t}. \end{aligned}$$

We then run the following cross-sectional regression using the estimated betas and returns for the next month

$$\begin{aligned} R_{j,t+1} - r_f &= \lambda_0 + \beta_{j,t}^{MKT} \lambda_{MKT} + \beta_{j,t}^{HML} \lambda_{HML} + \beta_{j,t}^{SMB} \lambda_{SMB} + \beta_{j,t}^{MOM} \lambda_{MOM} \\ &\quad + \beta_{j,t}^{COSK} \lambda_{COSK} + e_{j,t}. \end{aligned}$$

We report the mean (in percentage) of the estimates and the Fama-MacBeth t-statistics with Newey-West correction for serial correlation, using 1 lag. We consider two periods, 1986-2012 and 1966-2012, and four sets of test assets.

Panel A: 1986 - 2012

	25 Size/BM		25 size/Mom		25 size/STR		25 size/LTR	
λ_0	0.968	1.197	0.069	0.871	0.215	0.111	0.133	0.274
	(2.28)	(3.33)	(0.17)	(2.81)	(0.59)	(0.26)	(0.45)	(0.85)
λ^{MKT}	-0.396	-0.600	0.486	-0.154	0.372	0.477	0.569	0.417
	(-0.85)	(-1.58)	(1.07)	(-0.42)	(0.83)	(1.05)	(1.67)	(1.14)
λ^{HML}		0.043		0.103		-0.090		0.125
		(0.24)		(0.53)		(-0.50)		(0.60)
λ^{SMB}		0.274		-0.177		0.116		0.101
		(1.45)		(-0.68)		(0.39)		(0.43)
λ^{MOM}		0.737		0.532		-0.507		0.113
		(1.82)		(1.90)		(-1.18)		(0.35)
λ^{COSK}	-0.080	-0.059	-0.148	-0.037	-0.008	0.023	-0.058	-0.035
	(-1.21)	(-1.16)	(-2.45)	(-0.87)	(-0.13)	(0.42)	(-0.92)	(-0.60)
Adj R^2	26.75	46.30	25.05	54.54	29.71	48.70	21.58	43.38

Panel B: 1966 - 2012

	25 Size/BM		25 size/Mom		25 size/STR		25 size/LTR	
λ_0	0.755	0.892	0.107	0.807	-0.538	-0.152	0.252	0.647
	(2.40)	(3.48)	(0.35)	(3.42)	(-1.68)	(-0.49)	(1.11)	(2.67)
λ^{MKT}	-0.251	-0.440	0.397	-0.258	1.017	0.555	0.364	-0.123
	(-0.70)	(-1.60)	(1.17)	(-0.96)	(2.70)	(1.67)	(1.32)	(-0.44)
λ^{HML}		0.193		0.227		0.138		0.229
		(1.39)		(1.56)		(0.96)		(1.50)
λ^{SMB}		0.372		-0.068		0.199		0.231
		(2.72)		(-0.39)		(0.92)		(1.42)
λ^{MOM}		0.433		0.680		-1.116		0.042
		(1.58)		(3.56)		(-3.71)		(0.17)
λ^{COSK}	-0.046	-0.045	-0.146	-0.060	-0.027	0.034	-0.040	-0.049
	(-1.05)	(-1.32)	(-3.56)	(-2.12)	(-0.58)	(0.91)	(-0.95)	(-1.35)
Adj R^2	30.16	48.95	26.37	54.40	30.26	49.61	23.07	43.47

Table 3: Out-of-Sample Tests

We test the out-of-sample performance of several competing models. The price of risk is estimated using cross-sectional regressions or historical risk premia. Using historical risk premia, we provide out-of-sample predictions for the CAPM, the model with a co-skewness premium COSK, and the model with market and co-skewness factors CAPM + COSK. Using cross-sectional regressions, we provide predictions for the CAPM, the model with market and co-skewness factors CAPM + COSK, and the Fama-French three-factor model FF. We also use a hybrid model with regression-based co-variance premium and historical co-skewness premium. This model is referred to as CSCAPM + COSK. We consider four sets of portfolios. We compute out-of-sample $R-$squares for each portfolio and report the average in Panel A. Panel B reports the average of the square root of the mean forecasting errors (RMSFE). The sample period is from January 1986 to December 2012.

Panel A: Out of Sample R-squares

Prices of Risk Estimated from Historical Risk Premia				
	CAPM	COSK	CAPM +COSK	CSCAPM + COSK
25 Size/Book-to-Market	-0.252	1.690	-0.005	-2.061
25 Size/Momentum	-0.553	0.377	-1.623	0.353
25 Size/Short-Term Reversal	-0.301	1.362	-0.676	0.304
25 Size/Long-Term Reversal	-0.420	0.888	-0.533	0.720
Prices of Risk Estimated from Cross-Sectional Regressions				
	CAPM	CAPM +COSK	FF	
25 Size/Book-to-Market	-3.905	-4.534	-7.141	
25 Size/Momentum	0.879	-0.214	-6.237	
25 Size/Short-Term Reversal	0.469	0.524	-0.691	
25 Size/Long-Term Reversal	0.440	0.509	-0.750	

Panel B: RMSFEs

Prices of Risk Estimated from Historical Risk Premia				
	CAPM	COSK	CAPM +COSK	CSCAPM +COSK
25 Size/Book-to-Market	5.709	5.651	5.704	5.75
25 Size/Momentum	6.088	6.061	6.124	6.065
25 Size/Short-Term Reversal	5.977	5.923	5.987	5.957
25 Size/Long-Term Reversal	5.624	5.587	5.629	5.593
Prices of Risk Estimated from Cross-Sectional Regressions				
	CAPM	CAPM +COSK	FF	
25 Size/Book-to-Market	5.805	5.822	5.891	
25 Size/Momentum	6.044	6.077	6.247	
25 Size/Short-Term Reversal	5.953	5.951	5.985	
25 Size/Long-Term Reversal	5.599	5.598	5.631	

Table 4: Robustness. Price of Risk and Out-of-Sample Performance

In Panel A, we provide estimates of the price of co-skewness risk using alternative estimators of the physical and risk-neutral second moments. In Panel B, we document the out-of-sample performance of the model with co-skewness risk using these different moment estimators. In Panel B we consider four sets of portfolios. The sample periods differ dependent on the availability of data.

Price of Coskewness Risk

	$E_t^P[R_{t+1}^2]$	$E_t^Q[R_{t+1}^2]$	$E_t^P[R_{t+1}^2] - E_t^Q[R_{t+1}^2]$
NGARCH + VIX (1986 - 2012)	0.2960	0.4381	-0.1421
NGARCH + BKM (1996 - 2012)	0.3445	0.5153	-0.1708
Heston + VIX (1986 - 2012)	0.1539	0.4386	-0.2847
Heston + BKM (1996 - 2012)	0.2132	0.5167	-0.3035
AR + VIX (1986 - 2012)	0.3152	0.4381	-0.1229
AR + BKM (1996 - 2012)	0.3462	0.5153	-0.1691
HAR + BKM (1996 - 2012)	0.1990	0.5153	-0.3163

Panel B: Out-of-Sample R-squares

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR
NGARCH + VIX (1986 - 2012)	1.976	0.762	1.679	1.236
NGARCH + BKM (1996 - 2012)	2.165	0.345	1.709	1.056
Heston + VIX (1986 - 2012)	1.521	0.492	1.243	0.750
Heston + BKM (1996 - 2012)	1.437	-0.070	1.076	0.120
AR + VIX (1986 - 2012)	0.974	0.618	1.352	0.333
AR + BKM (1996 - 2012)	0.925	0.278	1.442	-0.091
HAR + BKM (1996 - 2012)	1.999	0.203	1.602	0.765

Table 5: The Option-Based Price of Co-Skewness and Co-Kurtosis Risk

The table provides descriptive statistics for the physical and risk-neutral expectations and the price of co-skewness and co-kurtosis risk. The data are monthly and the mean and the standard deviation are reported in percentages. The second and third conditional physical moments are estimated using the autoregressive conditional volatility, skewness, and kurtosis model of Jondeau and Rockinger (2003). The risk-neutral moments are estimated using the model-free approach in Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003). The sample period is from January 1996 to December 2012.

Panel A: Coskewness Risk

	$E_t^P[R_{t+1}^2]$	$E_t^Q[R_{t+1}^2]$	$E_t^P[R_{t+1}^2] - E_t^Q[R_{t+1}^2]$
mean	0.2408	0.5153	-0.2745
std	0.1939	0.3961	0.3024
skew	2.8889	2.7374	-3.8285
kurt	13.2043	15.0935	27.0616
$\rho(1)$	0.8421	0.7278	0.4257

Panel B: Cokurtosis Risk

	$E_t^P[R_{t+1}^3]$	$E_t^Q[R_{t+1}^3]$	$E_t^P[R_{t+1}^3] - E_t^Q[R_{t+1}^3]$
mean	-0.0104	-0.0220	0.0116
std	-4.9370	-5.7427	6.1949
skew	-4.9370	-5.4637	7.8020
kurt	33.3414	49.7137	63.3882
$\rho(1)$	0.6639	0.4515	-0.0446

Table 6: Out-of-Sample Tests with Co-Kurtosis Risk

We test the out-of-sample performance of several competing models. The price of risk is estimated using cross-sectional regressions or historical risk premia. Using historical risk premia, we provide out-of-sample predictions for the CAPM, the model with a co-skewness premium COSK, the model with a co-kurtosis premium COKU, and the model with market and co-skewness factors and market and co-kurtosis factors, CAPM + COSK and CAPM + COKU respectively. Using cross-sectional regressions, we provide predictions for the CAPM, CAPM + COSK, CAPM + COKU, and the Fama-French three-factor model FF. We also use hybrid models with regression-based co-variance premium and historical co-skewness and co-kurtosis premiums, CSCAPM + COSK and CSCAPM + COKU. We consider four sets of portfolios. Panel A reports out-of-sample $R-$squares. Panel B reports root mean squared forecasting errors (RMSFE). The sample period is from January 1986 to December 2012.

Panel A: Out-of-Sample R-squares

Prices of Risk Estimated from Historical Risk Premia							
	CAPM	COSK	COKU	CAPM +COSK	CAPM +COKU	CSCAPM +COSK	CSCAPM +COKU
25 Size/Book-to-Market	-0.402	2.074	0.396	0.088	-0.702	-0.342	-2.927
25 Size/Momentum	-0.731	0.249	-0.432	-2.033	-1.729	0.818	0.270
25 Size/Short-Term Reversal	-0.418	1.306	0.397	-1.015	-1.113	0.473	0.335
25 Size/Long-Term Reversal	-0.644	1.001	-0.331	-0.719	-1.410	1.485	0.023

Prices of Risk Estimated from Cross-Sectional Regressions				
	CAPM	CAPM +COSK	CAPM +COKU	FF
25 Size/Book-to-Market	-1.943	-2.471	-3.931	-7.966
25 Size/Momentum	1.595	0.960	0.615	-4.681
25 Size/Short-Term Reversal	1.048	1.736	0.007	-0.392
25 Size/Long-Term Reversal	0.998	1.046	0.178	0.055

Panel B: RMSFEs

Prices of Risk Estimated from Historical Risk Premia							
	CAPM	COSK	COKU	CAPM +COSK	CAPM +COKU	CSCAPM +COSK	CSCAPM +COKU
25 Size/Book-to-Market	6.083	6.005	6.056	6.068	6.092	6.074	6.154
25 Size/Momentum	6.557	6.525	6.548	6.600	6.595	6.510	6.531
25 Size/Short-Term Reversal	6.424	6.365	6.399	6.441	6.449	6.393	6.401
25 Size/Long-Term Reversal	5.934	5.883	5.924	5.936	5.956	5.871	5.916

Prices of Risk Estimated from Cross-Sectional Regressions				
	CAPM	CAPM +COSK	CAPM +COKU	FF
25 Size/Book-to-Market	6.125	6.140	6.192	6.294
25 Size/Momentum	6.482	6.504	6.519	6.678
25 Size/Short-Term Reversal	6.376	6.354	6.412	6.419
25 Size/Long-Term Reversal	5.886	5.885	5.912	5.913

Table 7: Robustness. Co-Kurtosis Price of Risk and Out-of-Sample Performance

Panel A provides estimates of the co-kurtosis risk premium using alternative measures of the physical third moment. Panel B reports out-of-sample R-squares using these alternative estimates. Panel B considers four sets of portfolios. We compute out-of-sample $R-$squares for each portfolio and report the average. The sample period is from January 1996 to December 2012.

Panel A: The Price of Co-Kurtosis Risk

	$E_t^P[R_{t+1}^2]$	$E_t^Q[R_{t+1}^2]$	$E_t^P[R_{t+1}^2] - E_t^Q[R_{t+1}^2]$
const skew (daily)	-0.0006	-0.0220	0.0214
const skew (monthly)	-0.0015	-0.0220	0.0205
zero skew	0.0000	-0.0220	0.0220

Panel B: Out-of-Sample R-squares

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR
const skew (daily)	1.087	0.510	1.098	0.360
const skew (monthly)	1.038	0.385	0.974	0.172
zero skew	0.934	0.399	1.007	0.214