Option-Based Estimation of the Price of Co-Skewness and Co-Kurtosis Risk*

Peter Christoffersen
University of Toronto

Mathieu Fournier
HEC Montréal

Kris Jacobs
University of Houston

Mehdi Karoui
OMERS

October 13, 2017

Abstract

We show that the prices of risk for factors that are nonlinear in the market return are readily obtained using index option prices. The price of co-skewness risk corresponds to the market variance risk premium, and the price of co-kurtosis risk corresponds to the market skewness risk premium. 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 compared to regression-based estimates.
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, Adlai Fisher, Bryan Kelly, Eric Renault, Michael Rockinger, Sang Byung Seo, Bas Werker, Xiaoyan Zhang, and seminar participants at the AFA meeting in Boston and 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.

1 Introduction

Kraus and Litzenberger (1976) show that if investors care about portfolio skewness, co-skewness is relevant for pricing assets in addition to co-variation with the market portfolio. If investors care about portfolio kurtosis, co-kurtosis is also relevant.[1] We propose a new strategy to estimate the price of co-skewness and co-kurtosis risk, which avoids problems inherent in the use of two-stage cross-sectional or Fama-MacBeth regressions. Our approach can also be used to estimate the price of other risks, provided that they are nonlinear functions of the market return, and it can easily accommodate the presence of other 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 expressions for the time-varying price of risk of the exposure to any nonlinear function of the market return. Remarkably, we can show that the price of co-skewness risk corresponds to the spread between the physical and the risk-neutral second moment, the market variance risk premium. Similarly, the price of co-kurtosis risk is given by the spread between the physical and the risk-neutral third moment, the market skewness risk premium.

To provide intuition for this result, consider the special case where the stochastic discount factor (SDF) is a linear function of the market return, which corresponds to the Capital Asset Pricing Model (CAPM). In this case, our general result implies that the price of risk corresponds to 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. While information from index options is not required in the linear SDF case, we show that whenever the SDF is nonlinear, information from index option prices can be used to pin down the price of risk of the nonlinear factor. Fortunately, we have particularly rich option information on the market index, as index options are among the most heavily traded contracts on the market. This makes our theoretical results very practical to implement.

Despite the obvious intuitive appeal of co-skewness and co-kurtosis, there seems to be no widespread consensus on their empirical relevance for cross-sectional asset pricing. We address one of the most important drawbacks of the approach, which is the measurement of the risk premia. Measurement is especially difficult when analyzing *conditional* co-skewness

[1] For important contributions to the literature on co-skewness and co-kurtosis, see for example Ang, Chen, and Xing (2006), Bansal and Viswanathan (1993), Dittmar (2002), Guidolin and Timmermann (2008), Harvey and Siddique (2000), Leland (1997), Lim (1989), Schneider (2015), and Scott and Horvath (1980). See also Rubinstein (1973) and Golec and Tamarkin (1998) for related work.

and co-kurtosis.[2] Most existing papers estimate and test the importance of co-skewness and co-kurtosis using two-stage cross-sectional or Fama-MacBeth regressions. 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.[3] 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 reason is that the higher the moment, the more difficult it is to precisely estimate betas. 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 estimates of the price of risk that are potentially much larger than in the case of the CAPM or the Fama-French three-factor model.

We empirically investigate the performance of our approach for the pricing of co-skewness and co-kurtosis risk. Using monthly data for the periods 1986-2012 and 1996-2012 respectively, we find that our conditional prices of co-skewness and co-kurtosis risk are stable and have the expected sign in almost every month in our sample. On average, both estimated prices of risk are larger in absolute value than the traditional estimates obtained using a two-stage Fama-MacBeth approach. While the average prices of risk obtained using the Fama-MacBeth approach most often have the theoretically anticipated signs, the estimates are noisy and reliable conditional estimates are not available. We evaluate the cross-sectional performance of our newly proposed estimates out-of-sample, and find that they outperform regression-based estimates of the price of risk.

In addition to the existing work on co-skewness and co-kurtosis discussed above, our work is also closely related to other strands of literature. Several studies investigate the importance of conditional skewness for the cross-section of stock returns and asset allocation. Recent work includes Conrad, Dittmar, and Ghysels (2013), Ghysels, Plazzi, and Valkanov (2015), and Schneider, Wagner, and Zechner (2015). Another literature uses higher moments of index returns extracted from option data as pricing factors. This literature follows Ang, Hodrick, Xing, and Zhang (2006), who show that the VIX is a priced factor in the cross-section of

[2] Kraus and Litzenberger (1976) provide an unconditional empirical analysis of co-skewness. For a classical example of a conditional analysis, see for instance Harvey and Siddique (2000). For a comparison of the performance of conditional and unconditional approaches, see Ghysels (1998).

[3] See, e.g. Jagannathan and Wang (1998), Shanken (1992), Kan and Zhang (1999), Kleibergen (2009), and Gospodinov, Kan, and Robotti (2014).

stock returns. Bali and Zhou (2014) investigate the performance of market variance risk as a proxy for economic uncertainty. Our finding that the price of co-skewness risk corresponds to the variance risk premium therefore provides another motivation for these results. Chang, Christoffersen, and Jacobs (2013) show that option-implied skewness is priced in the cross-section of stock returns, and this finding can be explained by our results on co-kurtosis. Cremers, Halling, and Weinbaum (2015) show that aggregate jump risk as well as aggregate volatility risk are priced in the cross-section of stock returns.

Several recent studies investigate whether the SDF is monotonic in market returns. Bakshi, Madan, and Panayotov (2010), Patton and Timmermann (2010), and Chaudhuri and Schroder (2015) reject a monotonic relationship and provide strong evidence that the pricing kernel is U-shaped. Non-linear SDFs can be motivated by the existence of investors' preference for higher order moments. For instance, a quadratic SDF is obtained in equilibrium when the third derivative of the marginal agent's utility function is non-zero, which is closely related to skewness preferences. Similarly, a non-zero fourth derivative of the utility function is related to kurtosis preferences.

Finally, our work is also related to the recent literature on disaster risk and heavy-tailed shocks in equilibrium models. Extending the approaches of Rietz (1988) and Barro (2006), Gabaix (2012), Gourio (2012) and Wachter (2013) argue that such heavy-tailed shocks to endowments or productivity can help explain long-standing puzzles in asset pricing.[4] Wachter (2013) notes that these studies are related to the co-skewness and co-kurtosis literature which examines the effects of nonnormalities on risk premia and the cross-section of returns. Drechsler and Yaron (2011) study the impact of heavy-tailed shocks in a long-run risk model and emphasize the importance of the variance risk premium in capturing attitudes towards uncertainty. They explicitly discuss the importance of combining derivatives information and information on the underlying to capture empirically relevant risk measures. Our approach theoretically justifies this approach and demonstrates its relevance for pricing the cross-section of returns.

The paper proceeds as follows. Section 2 discusses our alternative approach as applied to the measurement of the price of co-skewness risk. Section 3 presents empirical results for co-skewness risk and compares the pricing performance of the new approach with regression-based estimates. Section 4 discusses the predictive performance of the new estimates. Section

[4]See Hansen and Singleton (1982) and Mehra and Prescott (1985) for seminal examples of asset pricing puzzles in consumption-based models. For additional work on equilibrium models of disaster risk, see Chen, Joslin, and Tran (2012), Julliard and Ghosh (2012), and Liu, Pan, and Wang (2005). For recent work on tail risk see Bollerslev and Todorov (2011) and Kelly and Jiang (2014).

5 applies the new approach to the measurement of co-kurtosis risk as well as more general nonlinear market risk. Section 6 concludes.

2 Measuring Co-Skewness Risk: An Option-Based Approach

In this section we investigate an asset pricing model in which cross-sectional differences in expected returns between assets are determined by their exposure to the squared market return in addition to the market return itself. We proceed to propose an option-based approach to measuring the price of risk for the nonlinear (co-skewness) exposure.

2.1 A Quadratic SDF

Absence of arbitrage implies the existence of a stochastic discount factor, m_{t+1}, that prices any asset with risky return, $R_{j,t+1}$, using the moment condition

$$E_t^P\left[\left(1+R_{j,t+1}\right)m_{t+1}\right]=1, \tag{1}$$

where $E_t^P(.)$ denotes the expectation under the physical probability measure. We assume that the SDF can be written as a representative investor's marginal rate of substitution between current and future wealth,

$$m_{t+1}=\frac{U'(W_{t+1})}{U'(W_t)}, \tag{2}$$

where $U'(.)$ is marginal utility and W is aggregate wealth. We do not make a specific assumption regarding the form of the representative investor's utility function; instead, as in Kraus and Litzenberger (1976), Barone-Adesi (1985), Harvey and Siddique (2000) and Dittmar (2002) among others, we take a second-order Taylor approximation for $U'(W_{t+1})$ around W_t to write

$$m_{t+1}\approx 1+\left(\frac{U''(W_t)W_t}{U'(W_t)}\right)R_{m,t+1}+\left(\frac{U'''(W_t)W_t^2}{2U'(W_t)}\right)R_{m,t+1}^2, \tag{3}$$

where we have used $W_{t+1}=W_t\left(1+R_{m,t+1}\right)$ and $R_{m,t+1}$ denotes the stock market return, which we will use as a proxy for the return on the wealth portfolio. From equation (1) we have $E_t^P(m_{t+1}) = \frac{1}{\left(1+R_{f,t}\right)}$ for the risk-free rate, which gives

$$\frac{1}{(1+R_{f,t})} \approx 1 + \left(\frac{U''(W_t)W_t}{U'(W_t)}\right) E_t^P(R_{m,t+1}) + \left(\frac{U'''(W_t)W_t^2}{2U'(W_t)}\right) E_t^P(R_{m,t+1}^2). \tag{4}$$

Combining equations (3) and (4), we obtain the following form for the SDF

$$m_{t+1} \approx a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right), \tag{5}$$

where $a_t = 1/\left(1+R_{f,t}\right)$, $b_{1,t} = \left(U''(W_t)W_t\right)/U'(W_t)$, and $b_{2,t} = \left(U'''(W_t)W_t^2\right)/\left(2U'(W_t)\right)$. Similar to Harvey and Siddique (2000), henceforth, HS, our setup is based on the assumption of $U'''(.) \neq 0$ which implies a quadratic SDF. Dittmar (2002) assumes declining absolute risk aversion which implies that $U'''(.) > 0$. The performance of quadratic pricing kernels is studied further in Bansal and Viswanathan (1993) and Chabi-Yo (2008) among others.[5]

2.2 The Price of Co-Skewness Risk

Given the SDF in equation (5), 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, harvesting the rich information in index option prices.

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, which presents the pricing implications of the SDF defined in equation (5).

[5] Chabi-Yo, Leisen and Renault (2014) study a general equilibrium setting for skewness risk. They find that the quadratic SDF in (5) obtains as a special case in a single-period setting with a representative investor. Allowing for heterogenous agents and/or multiperiod models results in more general prices of risk than the ones derived in our complete-market setup. Investor heterogeneity is studied in, among others, Constantinides and Duffie (1996) and Mitton and Vorkink (2007).

Proposition 1 *In the absence of arbitrage opportunities, if the stochastic discount factor (SDF) has the following form*

$$m_{t+1} = a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right), \tag{6}$$

then the cross-sectional pricing restriction on stock returns is

$$E_t^P(R_{j,t+1}) - R_{f,t} = \lambda_t^{MKT}\beta_{j,t}^{MKT} + \lambda_t^{COSK}\beta_{j,t}^{COSK}, \tag{7}$$

where $\beta_{j,t}^{MKT}$ and $\beta_{j,t}^{COSK}$ 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

$$\lambda_t^{MKT} = E_t^P(R_{m,t+1}) - R_{f,t}, \tag{8}$$

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

$$\lambda_t^{COSK} = E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2). \tag{9}$$

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 mean zero risk factors $\mathbf{f}$ satisfying

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

where $\boldsymbol{\lambda}_t' = \frac{-1}{a_t}\mathbf{b}_t'E_t^P(\mathbf{f}_{t+1}\mathbf{f}_{t+1}')$, $(1+R_{f,t}) = \frac{1}{a_t} = \frac{1}{E_t^P(m_{t+1})}$, and $\boldsymbol{\beta}_{j,t} = \left[E_t^P(\mathbf{f}_{t+1}\mathbf{f}_{t+1}')\right]^{-1}E_t^P(\mathbf{f}_{t+1}R_{j,t+1})$, see for instance Cochrane (2005). Since the pricing kernel prices all the assets, the above equation also holds for any contingent claim with payoff Ψ, which can be a function of the market index return or of the stock return. Consequently, applying equation (10) to Ψ gives

$$E_t^P\left(R_{\Psi,t+1}\right) - R_{f,t} = E_t^P\left(\frac{\Psi_{t+1} - P_t}{P_t}\right) - R_{f,t} = \boldsymbol{\lambda}_t'\boldsymbol{\beta}_{\Psi,t}, \tag{11}$$

where P_t is the price of the contingent claim Ψ and R_Ψ is the return on the contingent claim.

Using the definition of $\boldsymbol{\beta}_{\Psi,t}$ we have

$$E_t^P\left(\frac{\Psi_{t+1}-P_t}{P_t}\right)-R_{f,t}=\boldsymbol{\lambda}_t'\left[E_t^P(\mathbf{f}_{t+1}\mathbf{f}_{t+1}')\right]^{-1}E_t^P\left(\mathbf{f}_{t+1}\left(\frac{\Psi_{t+1}-P_t}{P_t}\right)\right). \tag{12}$$

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

$$\begin{aligned}E_t^P(\Psi_{t+1})-P_t\left(1+R_{f,t}\right) &= \boldsymbol{\lambda}_t'\left[E_t^P(\mathbf{f}_{t+1}\mathbf{f}_{t+1}')\right]^{-1}E_t^P\left(\mathbf{f}_{t+1}\Psi_{t+1}\right)\\ &= \boldsymbol{\lambda}_t'\tilde{\boldsymbol{\beta}}_{\Psi,t},\end{aligned} \tag{13}$$

where $\tilde{\boldsymbol{\beta}}_{\Psi,t}$ is from the projection of Ψ on $\mathbf{f}$. The no-arbitrage condition ensures the existence of at least one risk-neutral measure Q such that $P_t=\frac{E_t^Q(\Psi_{t+1})}{\left(1+R_{f,t}\right)}$. Therefore, we get

$$E_t^P\left(\Psi_{t+1}\right)-E_t^Q\left(\Psi_{t+1}\right)=\boldsymbol{\lambda}_t'\tilde{\boldsymbol{\beta}}_{\Psi,t}. \tag{14}$$

To obtain the result in equation (8) from equation (14), we now consider the contingent claim $\Psi_{t+1}\equiv R_{m,t+1}$. If a return is also a factor, it has a loading of one onto itself and zero onto the other factors. Given the SDF (6), this gives $\tilde{\boldsymbol{\beta}}_{\Psi,t}=[1\ 0]'$ and equation (14) reduces to equation (8). Similarly, using $\Psi_{t+1}\equiv R_{m,t+1}^2$, we obtain $\tilde{\boldsymbol{\beta}}_{\Psi,t}=[0\ 1]'$ which applied to equation (14) gives the result in equation (9). ■

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. 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.

The spread between the physical and risk neutral market variance is often termed the variance risk premium and it has been found to be one of the best predictors of market returns. See, for example, Bollerslev, Tauchen, and Zhou (2009) and Bollerslev, Marrone, Xu and Zhou (2014). Our analysis above suggests that the variance risk premium is a predictor of market returns because it provides information about 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. In our empirical implementation, we show that an additional advantage of our approach is that the period-by-period estimates of the conditional price of risk are rather reliable and precise, in contrast with the estimates obtained using the regression-based approach.

The existing empirical evidence clearly indicates that risk-neutral variance is larger than physical variance, therefore suggesting a negative price of co-skewness risk.[6] 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.

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 as in Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003).

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 $E_t^P(R_{m,t+1}) - R_{f,t} = \lambda_t^{MKT}$ 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 (7) of Proposition 1 as follows

$$E_t^P(R_{j,t+1}) - R_{f,t} = \beta_{j,t}^{MKT}\left[E_t^P(R_{m,t+1}) - R_{f,t}\right] + \beta_{j,t}^{COSK}\left[E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2)\right], \quad (15)$$

which can also be written as

$$E_t^P(R_{j,t+1}) - R_{f,t} = c_{j,t} + \beta_{j,t}^{MKT}\left[E_t^P(R_{m,t+1}) - R_{f,t}\right] + \beta_{j,t}^{COSK} E_t^P\left[(R_{m,t+1} - E_t^P(R_{m,t+1}))^2\right], \quad (16)$$

where $c_{j,t} = \beta_{j,t}^{COSK}\left[(E_t^P(R_{m,t+1}))^2 - E_t^Q(R_{m,t+1}^2)\right]$. Equation (16) directly establishes the link between our approach and existing work. For instance, it is equivalent to equation (6) in KL (1976) and it can be re-written as equation (8) in HS (2000).

[6] See for instance Bakshi and Madan (2006), Bollerslev, Tauchen, and Zhou (2009), Carr and Wu (2009), and Jackwerth and Rubinstein (1996).

The crucial difference between our approach and the one in KL (1976) and HS (2000) is that we explicitly impose no-arbitrage restrictions on contingent claims prices, so that the pricing kernel prices all assets in the economy. This additional assumption leads to a very simple estimator of the price of risk.[7]

2.3 A First Look at the Option-Implied Price of Co-Skewness Risk

Figure 1 plots our option-based estimate of $\lambda_{OI,t}^{COSK} = E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2)$, where $E_t^P(R_{m,t+1}^2)$ in the top panel is obtained using the heterogeneous autoregressive (HAR) model of Corsi (2009), and $E_t^Q(R_{m,t+1}^2)$ in the middle panel is estimated by the squared value of the VIX. The details of the estimation will be provided in Section 3 below. For now, note that the price of co-skewness risk, $\lambda_{OI,t}^{COSK}$, in the bottom panel of Figure 1 exhibits interesting spikes surrounding the 1998 LTCM collapse, the WorldCom bankruptcy in 2002, the credit crisis in 2008, and bad economic news in Europe and the US at the end of 2011. Both the risk-neutral and physical variance display spikes around those events, but the spikes in the physical variance are typically relatively smaller than the risk-neutral spikes, leading to the negative spikes in the price of co-skewness risk. Note also that the new genuinely conditional estimates of the price of co-skewness risk in the bottom panel of Figure 1 have the theoretically expected negative sign in almost every month.

3 Estimating the Price of Co-Skewness Risk

We now present estimates of 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 discuss the differences between our new estimates and conventional regression-based estimates.

[7] Barras and Malkhozov (2016) find that the market variance risk premium (VRP) measured in the equity market is similar to the VRP estimated from options but reject the null hypothesis that they are identical. Our approach assumes that these risks have the same prices in different markets.

3.1 Estimating the Risk-Neutral Second Moment

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.[8] Using the VIX has a number of advantages. The VIX provides a very simple benchmark because the data are readily available from the Chicago Board of Options Exchange (CBOE). 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 use data for the ticker VXO throughout and obtain data for the period January 1986 to December 2012. For existing studies that use the VIX squared as a proxy for the expected risk-neutral second moment with one month horizon, see for instance Bollerslev, Tauchen, and Zhou (2009). In the robustness analysis in Section 4.2, we use an alternative approach to compute the risk-neutral variance, following Bakshi and Madan (2000).

3.2 Estimating the Physical Second Moment

Our benchmark implementation uses the so-called heterogeneous autoregressive (HAR) model of Corsi (2009) estimated on daily, weekly, and monthly realized variances via

$$V^m_{t+1,t+K} = \phi_0 + \phi_1 V^m_{t-1,t} + \phi_2 V^m_{t-4,t} + \phi_3 V^m_{t-20,t} + \varepsilon^m_{t+1,t+K}, \tag{17}$$

where $V^m_{t+1,t+K}$ denotes the market index K-days ahead integrated variance. In the previous equation, the variance terms are defined as

$$V^m_{s,s+\tau} = V^m_s + V^m_{s+1} + ... + V^m_{s+\tau}, \tag{18}$$

with the daily variance given as in Rogers and Satchell (1991) by

$$\begin{aligned} V^m_t &= \ln(S^{High}_t/S^{Open}_t)\left[\ln(S^{High}_t/S^{Open}_t) - \ln(S^{Close}_t/S^{Open}_t)\right] \\ &\quad + \ln(S^{Low}_t/S^{Open}_t)\left[\ln(S^{Low}_t/S^{Open}_t) - \ln(S^{Close}_t/S^{Open}_t)\right], \end{aligned} \tag{19}$$

[8] The theoretical results in Section 2 are based on the uncentered moments. Throughout our empirical work we use both centered and uncentered moments, and the results are very similar.

where S_t^{Close} (S_t^{Open}) is the close (open) price of the market index, measured by the S&P 500, and S_t^{High} (S_t^{Low}) denotes the market index highest (lowest) price on day t.[9] The HAR model in (17)-(19) parsimoniously allows for a highly persistent dynamic in volatility and employs the intraday information available in our relatively long historical sample. We estimate the model using OLS and a recursive ten-year (120-month) window. For related applications of high-low information in dynamic volatility models, see Azalideh, Brandt and Diebold (2002), Chou (2005), and Brandt and Jones (2006).[10]

3.3 Option-Based Estimates of the Price of Co-Skewness Risk

Using our benchmark HAR estimate of the physical second moment, and our benchmark VIX risk-neutral second moment, the estimated price of co-skewness risk for month t is now simply

$$\widehat{\lambda}_{OI,t}^{COSK} = \widehat{E}_t^P(R_{m,t+1}^2) - \widehat{E}_t^Q(R_{m,t+1}^2). \tag{20}$$

Table 1 reports descriptive statistics for the estimates of the moments and the price of risk. We discuss the robustness of our results to the choice of physical and risk-neutral variance model in Section 4.2 below. Table 1 and Figure 1 indicate that the co-skewness price of risk is negative for almost all months. On average it is equal to -0.146. This negative sign is consistent with theory, and several existing empirical studies document a negative price of co-skewness risk as well, see for instance KL (1976) and HS (2000).

3.4 Regression-Based Estimates

Table 2 reports results for three models based on cross-sectional regressions. The first model incorporates exposure to the market factor and the second combines co-skewness and market factors. The third model also includes the Fama-French (1993) size and book-to-market factors, and the momentum factor. We estimate betas using the same window we use for the HAR model, ten years of monthly returns, and subsequently run a cross-sectional regression for the next month.

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

[9] We re-scale V_t^m to ensure its average equals the sample variance from daily close-to-close returns.

[10] Corsi (2009) and subsequent HAR papers typically rely on high-frequency intraday returns to compute daily variance proxies. However, high-frequency returns are not readily available in the beginning of our sample period.

on four cross-sectional datasets that are commonly used in the existing literature. We use portfolios formed on size and book-to-market, on size and momentum, on size and short-term reversal, and 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 also report on the dataset that combines these four sets of test assets. This test uses 100 test assets.

Table 2 contains results for two sample periods: Our benchmark sample (1986-2012) in Panel A and a somewhat shorter sample (1996-2012) in Panel B. We consider the shorter sample to enable comparisons with other measures of the option-based risk premium, which are not available for the 1986-2012 sample, in the robustness section below.

Although the adjusted R^2s in Table 2 increase substantially when the co-skew factor is included in the model, an important conclusion is that the estimate of the price of co-skewness risk depends critically on the model and portfolios used in estimation. Focusing on the 1986-2012 sample in Panel A, the price of co-skewness risk ranges from a significant -0.155 in the model with the Fama-French and momentum factors estimated on size and long-term reversal portfolios to an insignificant 0.008 in the bivariate model estimated on size and short-term reversal portfolios.[11]

Figure 2 reports on the bivariate model that includes the co-skewness and market factors. We report the time-series of the month-by-month cross-sectional regression estimates of the price of co-skewness risk. The estimates for the price of co-skewness risk in the bivariate models in Table 2 are thus the averages of the time series in Figure 2. 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 2 indicates that this can be explained by the fact that the regression estimates are noisy, and the 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.

The essence of the Fama-MacBeth procedure is of course to estimate the price of risk by averaging the time series of cross-sectional estimates. The fact that the estimates in Figure 2 are positive for some months therefore does not constitute a problem in itself. But 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 Table 2 are negative. Figure 2 therefore suggests that noise in the month-by-month estimates is an important problem with regression-based estimation

[11] For expositional convenience we report the estimated prices of co-skewness and co-kurtosis risk, as well as the moments used to construct these estimates, in percentage terms, i.e. multiplied by 100.

of the price of co-skewness risk.

3.5 Comparing the Estimates

The above analysis reveals that there are some very important differences between our empirical results and regression-based estimates of the price of co-skewness risk.

First, the average of our newly proposed estimate of the price of co-skewness risk in Table 1, which is equal to -0.146, is larger (in absolute value) than all but one of the estimates obtained using the regression approach in Section 3.4. 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 4 below.

Second, it is interesting to compare the time-series of conditional estimates in Figure 1 with the time-series for the regression-based estimates in Figure 2. The monthly estimates in Figure 1 are almost all negative, and the difference with Figure 2 is striking. This of course also explains why the negative average estimate of -0.146 for our approach is larger (in absolute value).

This comparison between the time series in Figures 1 and 2 must of course be interpreted with some caution. Existing studies report averages of the price of risk over several years. They 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, rather than the time series in Figure 2. Indeed, it can be argued that the focus of the Fama-MacBeth approach is to obtain estimates of the price of risk by averaging the time-series in Figure 2, and therefore the time series itself is not meaningful. The month-by-month estimates of the price of risk may not have the theoretically expected negative sign, but this does not invalidate the unconditional estimate.

From this perspective, what is truly remarkable about our option-implied estimates 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. The regression-based approach obviously does not provide us with such results. Moreover, while there is no guarantee that these negative estimates for the price of co-skewness risk will continue to obtain in the future, we know that implied variances usually exceed historical variances. Because of this stylized fact, our approach is more likely to yield plausible estimates of co-skewness risk.

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 unconditional estimate of the price of co-skewness risk is -0.146 and indicates a role for co-skewness that is large in magnitude compared to regression-based approaches. Third, we consistently obtain negative estimates of the price of conditional co-skewness risk in our approach, which is not the case with regression-based methods.

We therefore conclude that our approach appears to be economically appealing. But the important question is of course how our new conditional, option-implied price of co-skewness risk can help in real-time investor decision marking. This is the topic to which we now turn.

4 Comparing Predictive Performance

In this section we compare the predictive performance of the option-implied and regression-based estimates of the co-skewness price of risk.

4.1 Comparing Cross-Sectional Predictive Performance

In this section our goal is to compare the predictive performance of the option-implied and the regression-based estimates of the price of co-skewness risk from a cross-sectional perspective. We proceed as follows. For each portfolio p in a given set of test assets, the one-month ahead expected return from co-skewness risk is defined by

$$COSK_{t+1|t}^{p,k} = \lambda_{k,t}^{COSK} \beta_{p,t}^{COSK}, \tag{21}$$

where $\lambda_{k,t}^{COSK}$ is either the option-implied price of co-skewness risk, which we denote $\lambda_{OI,t}^{COSK}$, using the HAR & VIX^2 specification, or the regression-based price of risk, which we denote $\lambda_{RB,t}^{COSK}$. In each case the estimate is constructed using information through the current month t.

Portfolio p's exposures $\beta_{p,t}^{COSK}$ to the market return squared and $\beta_{p,t}^{MKT}$ are estimated via OLS regression including month t returns. Based on the model predictions, we can define the monthly forecast error for each model by

$$\varepsilon_{t+1}^{p,k} = R_{p,t+1} - \lambda_{RB,t}^{MKT} \beta_{p,t}^{MKT} - COSK_{t+1|t}^{p,k}, \tag{22}$$

where $R_{p,t+1}$ is the ex-post realized portfolio and market excess returns next month. Note that because we are focused on comparing predictions from different estimates of the co-skewness price of risk, we subtract portfolio p's expected return due to its market exposure. When evaluating the regression-based price of co-skewness risk, $\lambda_{RB,t}^{MKT}$ is estimated using a bivariate model that includes both market return and market return squared. When evaluating the option-implied price of co-skewness risk, $\lambda_{RB,t}^{MKT}$ is estimated independently of $\lambda_{RB,t}^{COSK}$. This approach prevents measurement errors in $\lambda_{RB,t}^{COSK}$ to spill over into $\lambda_{RB,t}^{MKT}$. To construct the forecast errors (22), we need the exposures $\beta_{p,t}^{MKT}$ and $\beta_{p,t}^{COSK}$. We estimate the MKT and $COSK$ betas of each portfolio jointly by bivariate regressions where the factors are the market excess return and the market excess return squared. This ensures consistency of our methodology with the SDF specified in Proposition 1.

When implementing equations (21) and (22), we have to decide on the rolling window to estimate the portfolio exposures $\beta_{p,t}^{COSK}$ and $\beta_{p,t}^{MKT}$, but also the window used for the prices of risk. Given the genuinely conditional nature of the option-based price of risk, we expect it to do better with shorter windows compared to the regression based approach. In our benchmark results, we report on a twelve-month moving average for the price of risk and a 120 month-window for the factor exposures. We also report robustness results for alternative windows.

Armed with the monthly forecast errors of each model, we compute two measures of relative performance across risk price estimation methods. The first is a simple difference in mean squared error that exploits the information embedded in the cross-section of portfolios in a given set of test assets. We refer to it as ΔMSE defined by

$$\Delta MSE = \frac{1}{T}\sum_{t=1}^{T} \Delta MSE_t, \tag{23}$$

where the monthly difference in mean squared error satisfies

$$\Delta MSE_t = \left(\frac{1}{25}\sum_{p=1}^{25}\left(\varepsilon_t^{p,RB}\right)^2 - \left(\varepsilon_t^{p,OI}\right)^2\right)\cdot 12\cdot 100. \tag{24}$$

Note that we multiply the cross-sectional average of the monthly difference in mean squared errors by 12 to annualize and by 100 to express it in percentage. Clearly, a positive ΔMSE indicates a superior forecast performance of the option-implied estimate versus the regression-based estimate.

We also consider a relative R^2 measure to evaluate the forecast performance across the risk price estimates. We compute the average R^2 across portfolios (in percent) using[12]

$$R^2 = \frac{1}{25} \sum_{p=1}^{25} R_p^2, \tag{25}$$

where the R-Squared of a given portfolio p in a given set of test assets is given by

$$R_p^2 = \left(1 - \frac{\sum_{t=1}^{T} \left(\varepsilon_t^{p,OI} \right)^2}{\sum_{t=1}^{T} \left(\varepsilon_t^{p,RB} \right)^2} \right) \cdot 100. \tag{26}$$

Similarly to the ΔMSE measure above, the R^2 is signed and measures the forecast performance of the option model relative to the regression model. A large and positive R^2 again indicates the superior forecast performance of the option-implied forecast versus the regression-based forecast. In contrast, a negative R^2 indicates superior forecast performance of the regression-based price of risk relative to the option-implied price of risk.

Based on these two measures of performance, our goal is to conduct statistical inference about the superiority of one price of risk calculation relative to the other. While statistical inference about ΔMSE can easily be conducted using standard Fama-MacBeth methodology by exploiting the time-series of the monthly ΔMSE_t, inference about the R^2 faces two important challenges. First, the two models we consider are non-nested while asymptotic theories and critical values for out-of-sample R-Squared are mostly available for the case of nested models. Second, our benchmark sample period of 1986-2012 is relatively short, which may cast doubts about the validity of standard asymptotic arguments. As a result, several of our empirical tests below rely on bootstrap methods for inference purposes.

Table 3 contains the results for ΔMSE and R^2. Consider first Panel A. For each of the four sets of 25 test assets, the ΔMSE is positive. The differences in mean squared error are large and range from 0.325 to 0.081. They are significantly different from zero at the 5% level in two of the four cases according to the Newey-West (NW) t-statistic when using the Student's t distribution. To benchmark the one-sided student-t p-values, we also report bootstrapped p-values. For a given set of test assets, the p-value corresponds to the probability of observing a ΔMSE under the null (when the forecast performance of both

[12] We also report tests results that combine the four sets of test assets. In this case, equations (24) and (25) must be adjusted by substituting 100 instead of 25.

models is equal) that is at least as large as its sample value. To this end, we bootstrap 100,000 paths of option-implied and regression-based squared forecast errors under the null hypothesis to obtain the distribution of ΔMSE. Based on that distribution we then infer the proportion of paths having higher ΔMSE than the sample value reported in Panel A. Table 3 shows that while the bootstrap and Student-t p-values are similar, the bootstrap indicates statistical significance in three of the four cases.[13] Panel A of Table 3 also contains a bootstrapped 90% confidence interval for the empirical distribution of ΔMSE. Note that three of the four 5^{th} percentile bounds of the empirical distribution of ΔMSE are larger than zero. In these cases the forecasting ability of the option-implied price of risk dominates the regression based price of risk more than 95% of the time.

Panel B of Table 3 reports various statistics on the R^2 computed across the 25 portfolios for each set of test assets. R^2s are economically large across test assets and range from 1.71% for size/LTR to 5.80% for Size/BM. Most importantly, the one-sided p-values of the R^2 are very small in three out of four cases, showing that the option-implied price of co-skewness risk significantly outperforms the regression-based estimate when using R^2.

The last column in Panels A and B reports results for all 100 test assets simultaneously. Both the ΔMSE and R^2 results are economically large and statistically significant.

The results in Table 3 use a twelve-month moving average for the price of risk and a 120 month-window for the factor exposures. Table 4 reports on alternative rolling windows, both for the price of risk and the factor exposure. We report on the relative R^2 for the 100 portfolios, but conclusions based on ΔMSE are similar. Panel A reports the relative R^2 as a function of the rolling window and Panel B reports the corresponding bootstrapped p-values. As expected, the evidence in favor of the option-based price of risk is strongest when the estimate of the price of risk is based on a single month. This is not surprising because as the length of the moving average used for the prices of risk increases, both prices of risk converge toward their unconditional means, and the conditional nature of the test disappears. Perhaps most importantly, the R^2 are always positive regardless of the estimation windows and are often highly statistically significant.

We conclude from Tables 3-4 that the cross-sectional predictive evidence in favor of our new option-based estimates of the price of co-skewness risk is strong, economically large, and statistically significant.

[13] Details on the bootstrap procedures can be found in Appendix A.1 and A.2.

4.2 Robustness

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

In our benchmark results we simply use the square of the VIX index as our estimate for the risk-neutral variance. In the robustness analysis, we use the Bakshi, Kapadia, and Madan (BKM, 2003) replication approach to obtain an estimate of the risk-neutral second moment. The BKM approach is based on a continuum of out-of-the money call and put options which is approximated using cubic spline interpolation techniques.

Let S_t denote the value of the market index and $R_{m,t+\tau} = \ln S_{t+\tau} - \ln S_t$ its return over the horizon τ. We can get the risk-neutral second moment via

$$\begin{aligned} E_t^Q\left[R_{m,t+\tau}^2\right] &= e^{r\tau}\int_{S_t}^{\infty}\frac{2\left(1-\ln\left[K/S_t\right]\right)}{K^2}C_t\left(\tau,K\right)dK \\ &\quad + e^{r\tau}\int_{0}^{S_t}\frac{2\left(1+\ln\left[S_t/K\right]\right)}{K^2}P_t\left(\tau,K\right)dK. \end{aligned} \tag{27}$$

where $C_t\left(\tau,K\right)$ and $P_t\left(\tau,K\right)$ are call and put options quoted at time t with maturity τ and strike price K. See Appendix B for more details on the implementation and the option data used to construct the BKM measure.

We investigate two alternative approaches to the HAR approach for modeling the conditional physical variance. The first is a simple autoregressive (AR) model on realized variances. Like the HAR model above, it is estimated using OLS and a ten-year recursive window. The one-step ahead forecast of the physical second moment is estimated from the following monthly regression

$$V_t^m = a_0 + a_1 V_{t-1}^m + u_{V,t}, \tag{28}$$

where $V_t^m = \sum_{d\in t} R_{m,d,t}^2$, $R_{m,d,t}$ denotes the daily market index return in day d of month t, and $u_{V,t}$ is the variance innovation.

We also consider the Heston (1993) stochastic volatility model in which the market index return follows

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{V_t^m}dW_{S,t}, \tag{29}$$

and the instantaneous variance dynamic is

$$dV_t^m = \kappa(\theta - V_t^m)dt + \eta\sqrt{V_t^m}dW_{V,t}, \tag{30}$$

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 the Euler discretization of this model using maximum likelihood and the particle filter on a ten-year moving window of daily returns. To ensure consistency with our measure of the risk-neutral variance, we generate the forecasts of the physical one-month total variance

$$\hat{V}^m_{t+1,t+21} = \theta + (V_t - \theta) \frac{1 - \exp(-(21/252)\,\kappa)}{\kappa\,(21/252)}, \tag{31}$$

from the model at the end of every month.

Figure 3 plots the six different option-implied prices of co-skewness risk constructed by combining one of the two risk-neutral variances (VIX or BKM) with one of the three physical variances (HAR, Heston, and AR). The top-left panel contains our benchmark HAR/VIX estimate from Figure 1 for reference. The six panels in Figure 3 demonstrate that the estimated prices of co-skewness risk are quite robust across methods and are highly correlated.

Table 5 presents cross-sectional prediction results using the different co-skew estimates in Figure 3. Panel A of Table 5 first reports estimates of the average price of risk obtained using the different approaches which can be compared with the benchmark case in Table 1. Note that the Heston model tends to deliver relatively low average physical variance estimates leading to an average price of co-skewness risk that is larger in magnitude. However, the differences are small and of course what ultimately matters is the dynamics of the estimated price of risk.

Panel B of Table 5 contains the ΔMSE and R^2 metrics for comparing the option-implied and regression-based estimates of the price of co-skew risk. A comparison with Panels A and B of Table 3 shows that the results in Table 3 are very robust. The p-values are small in Table 5 for three of the four sets of test assets, indicating that the option-implied estimates of co-skew risk are significantly better than the regression-based estimates in a ΔMSE-sense. The one-sided p-values for the significance of R^2 are very small everywhere in Table 5 leading to the same conclusion: The option-implied co-skew risk estimates deliver a superior fit to the cross-section of equity portfolio ex-post co-skewness returns regardless of which risk-neutral and physical variance estimate is used.

In Table 6 we consider the 1986-2007 sample period. This provides insights into the impact of the turbulent 2008-2009 financial crisis period on the performance of the option-implied price of risk. Panel A of Table 6 shows the ΔMSE results and should be compared with Panel A in Table 3. Note that our results are robust and if anything even stronger in favor of the option-implied estimates than the benchmark results in Table 3. Using all 100 test assets, the ΔMSE is 0.241 which is higher than the average of 0.193 obtained during the 1986-2012 sample period. We conclude that the superior performance of the new estimate of the price of Co-Skewness risk is not driven by the recent financial crisis.

Panel B of Table 5 shows that the R^2 are again large. Using all test assets, the R^2 is equal to 5.09%, compared to 3.70% in Table 3. We conclude that the improvements offered by the option-implied estimates of co-skewness risk over the regression-based benchmark are pervasive across different time periods and are not driven by the financial crisis.

5 The Price of Co-Kurtosis Risk

In this section we provide a framework for estimating the price of co-kurtosis risk using index option prices and we provide an initial estimate of this conditional price of risk. We also briefly discuss how our approach can be generalized beyond co-skewness and co-kurtosis to more general risks that are nonlinear in the market return.

5.1 Measuring Co-Kurtosis Risk

A natural extension of the quadratic pricing kernel in equation (5) is the cubic pricing kernel studied in Dittmar (2002), given by

$$\begin{aligned} m_{t+1} &= a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right) \\ &\quad + b_{3,t}\left(R_{m,t+1}^3 - E_t^P(R_{m,t+1}^3)\right). \end{aligned} \tag{32}$$

A cubic pricing kernel is consistent with investors' preferences for higher order moments, specifically skewness and kurtosis (i.e. $U'''(.) \neq 0$ and $U''''(.) \neq 0$). 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 by $\beta_{j,t}^{COKU}$.

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

Proposition 2 *In the absence of arbitrage opportunities, if the stochastic discount factor (SDF) has the following form:*

$$\begin{aligned} m_{t+1} &= a_t + b_{1,t}\left(R_{m,t+1} - E_t^P(R_{m,t+1})\right) + b_{2,t}\left(R_{m,t+1}^2 - E_t^P(R_{m,t+1}^2)\right) \\ &\quad + b_{3,t}\left(R_{m,t+1}^3 - E_t^P(R_{m,t+1}^3)\right), \end{aligned} \tag{33}$$

then the cross-sectional restriction on stock returns is

$$E_t^P(R_{j,t+1}) - R_{f,t} = \lambda_t^{MKT}\beta_{j,t}^{MKT} + \lambda_t^{COSK}\beta_{j,t}^{COSK} + \lambda_t^{COKU}\beta_{j,t}^{COKU}, \tag{34}$$

where $\beta_{j,t}^{MKT}$, $\beta_{j,t}^{COSK}$, and $\beta_{j,t}^{COKU}$ 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

$$\lambda_t^{MKT} = E_t^P(R_{m,t+1}) - R_{f,t}, \tag{35}$$

$$\lambda_t^{COSK} = E_t^P(R_{m,t+1}^2) - E_t^Q(R_{m,t+1}^2), \tag{36}$$

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

$$\lambda_t^{COKU} = E_t^P(R_{m,t+1}^3) - E_t^Q(R_{m,t+1}^3), \tag{37}$$

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. Given equation (33), applying equation (14) for $\Psi_{t+1} \equiv R_{m,t+1}$ as in Proposition 1, we recover equation (35), and applying equation (14) for $\Psi_{t+1} \equiv R_{m,t+1}^2$, we recover equation (36). In addition, applying equation (14) for $\Psi_{t+1} \equiv R_{m,t+1}^3$, we obtain equation (37). This again uses the results that a return which is also a factor has a loading of one onto itself and zero on the other factors. ■

Proposition 2 shows that the price of co-kurtosis risk is equal to the spread between the market physical and risk-neutral third moments. Existing empirical 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 consistent with theory, as explained earlier in this section and in Dittmar (2002).

5.2 Estimating Risk-Neutral and Physical Third Moments

It is well known that capturing the time-variation in the physical third moment is extremely difficult, see for instance Jondeau and Rockinger (2003). It is also well known that for index returns, the risk-neutral third moment is on average much larger (in absolute value) than the physical third moment, see for instance Bakshi and Kapadia (2003). Given these stylized facts, we proceed as follows.

We estimate the risk-neutral third moment using the method of Bakshi and Madan (2000). We implement the Bakshi and Madan (2000) approach using data on S&P500 index options from OptionMetrics for the period January 1996 to December 2012. As in equation (27) above let S_t denote the value of the market index and $R_{m,t+\tau} = \ln S_{t+\tau} - \ln S_t$ its return over the horizon τ. We can get the option-implied third moment via

$$\begin{aligned} E_t^Q \left[R_{m,t+\tau}^3\right] &= e^{r\tau} \int_{S_t}^{\infty} \frac{6 \ln\left[K/S_t\right] - 3\left(\ln\left[K/S_t\right]\right)^2}{K^2} C_t\left(\tau, K\right) dK \\ &\quad - e^{r\tau} \int_{0}^{S_t} \frac{6 \ln\left[S_t/K\right] + 3\left(\ln\left[S_t/K\right]\right)^2}{K^2} P_t\left(\tau, K\right) dK. \end{aligned} \tag{38}$$

where $C_t(\tau, K)$ and $P_t(\tau, K)$ are call and put options quoted at time t with maturity τ and strike price K. See Appendix B for more details on the implementation and the data.

It is well known that modeling the third moment dynamics under the physical measure is challenging, partly because it is much less persistent than the second moment–particularly so at the monthly frequency. Our own empirical implementation confirmed these challenges. The unconditional third moment estimate for monthly S&P500 returns during 1986-2012 is not statistically different from zero at conventional confidence levels, and moreover it is very small compared to the estimates of risk-neutral moments in our sample. These stylized facts motivated us to set the physical third moment equal to zero in our benchmark implementation. Setting the physical third moment to zero is preferable to using noisy estimates. In this case the price of co-kurtosis risk for month t can simply be computed as

$$\widehat{\lambda}_{OI,t}^{COKU} = \widehat{E}_t^P(R_{m,t+1}^3) - \widehat{E}_t^Q(R_{m,t+1}^3) = -\widehat{E}_t^Q(R_{m,t+1}^3). \tag{39}$$

Below we also report on alternative modeling assumptions for the physical third moment.

5.3 Empirical Results

Figure 4 depicts the time series of the price of co-kurtosis risk for our benchmark implementation, which is simply the negative of the risk-neutral third moment. Consistent with theory, the price of co-kurtosis risk in Figure 4 is positive throughout the period. Panel A of Table 7 indicates that it is equal to 0.022 on average.[14] 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. See for instance Ang, Chen, and Xing (2006), who find that stocks with higher co-kurtosis earn higher returns.

Panel A of Table 7 also reports on the price of co-kurtosis risk using two alternative estimates for the physical third moment: first, a constant third moment computed using daily data; second, a fully dynamic physical third moment, estimated using a version of the dynamic moment model in Jondeau and Rockinger (2003) described in Appendix C. Our implementation is close to the model Jondeau and Rockinger (2003) refer to as Model 2, which is among the more parsimonious models they consider and which is sufficiently richly parameterized for our purposes. Panel A of Table 7 indicates that one of the resulting estimates of the price of co-kurtosis risk is very similar to the estimate of 0.022 obtained using a zero physical third moment; the third estimate is positive but smaller.

Figure 5 reports on estimates of co-kurtosis risk obtained using Fama-MacBeth regressions from a bivariate model that includes the co-kurtosis and market factors. Figure 5 indicates that the month-by-month estimates of the price of co-kurtosis risk vary considerably over time, and that they are often negative. However, the average of the time-series estimates are all positive for the four sets of test assets and range from 0.0069 for Size/Short-term reversal to 0.0210 for Size/Book-to-market. In line with the conclusions drawn for Figure 2, the statistical significance of these estimates is weak and strongly depends on the set of test

[14]Recall that for expositional convenience these estimates of the moments and the prices of co-skewness and co-kurtosis risk are reported in percentage terms, i.e. multiplied by 100.

assets considered. Compared to the regression-based time series of the prices of co-skewness risk in Figure 2, the time series of the prices of co-kurtosis risk are less correlated across test assets.

Tables 7 and 8 report on the point estimates as well as the statistical significance for the out-of-sample ΔMSE and R^2 metrics applied to the co-kurtosis price of risk, constructed as in Section 4. Panel B of Table 7 reports on a bivariate model with the market price of risk and the co-kurtosis price of risk. Table 8 reports on the model that also includes the co-skew price of risk. In Table 7 we report on all three methods used to compute the physical third moment. In Table 8 we only report on two implementations with zero and constant physical skew because of space constraints and the excessive number of permutations, but we get similar conclusions when using the other estimate of the physical third moment.

All of the ΔMSE and R^2 in Table 7 are positive. Regardless of the model used for the physical third moment, the improvements are smallest for the 25 size/LTR portfolios, and the statistical evidence in favor of the co-skew price of risk is weakest for these test portfolios. For the other three sets of test portfolios, the p-values are very small both for the ΔMSE and R^2 measures. This is also the case when using all 100 test assets in the last column.

When using both the co-skew and the co-kurtosis price of risk in Table 8, the statistical evidence in favor of the model is even stronger, in the sense that the p-values are now also small for the 25 size/LTR portfolios. Note that these are out-of-sample tests, and that the performance of the model therefore cannot simply be attributed to the fact that more pricing factors are included.

We conclude that the model with co-kurtosis risk performs well in predicting the cross-section of portfolio returns when using price of risk estimated from option-based risk premia.

5.4 More General SDFs

Proposition 3 in Appendix D examines 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. While the empirical focus of this paper is on co-skewness and co-kurtosis risk, our approach can be used for virtually any source of risk that is an integrable function of the market return. This does not just include expectations of 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 refer to Appendix D for the details.

6 Conclusion

We propose an alternative strategy for estimating the price of possibly nonlinear exposures to market risk that avoids some shortcomings inherent in the cross-sectional regression approach. The key difference between our approach and existing methods is that we explicitly impose the 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.

We find that the price of co-skewness risk has the theoretically expected negative sign in almost every month, and the price of co-kurtosis risk has the theoretically expected positive sign in almost every month. In contrast, the prices of risk obtained using regression-based approaches do not always have the theoretically anticipated signs, even on average. Our approach thus 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 compared to regression-based estimates.

Several extensions of our work are possible. For computing the price of co-kurtosis risk, it may prove interesting to use the CBOE SKEW index, which, like the VIX, is readily available. Improved modeling of the physical third moment may also lead to improvements in the estimates of the price of co-kurtosis risk. While the focus of our new approach is on improving measurement of the price of risk, we worry that the estimated betas we use in the analysis may be noisy and impact the performance of the newly proposed prices of risk. 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. Finally, it would be useful to reliably assess the statistical significance of the price of co-skewness and co-kurtosis risk that takes into account the uncertainty in the various steps involved in the computation.

Appendix A.1: Bootstrapping ΔMSE and R^2

To assess the statistical significance of ΔMSE and R^2, we rely on bootstrapping. Our null hypothesis is that the option and regression predictions perform equally well, that is $\Delta MSE = 0$ and $R^2 = 0$, when forecasting the one-month ahead portfolio returns for a given set of test assets. This will be the case when the sample average of $\left(\varepsilon_t^{p,OI}\right)^2$ is equal to that of $\left(\varepsilon_t^{p,RB}\right)^2$ for each portfolio in the set considered. To see this, let us consider ΔMSE first. Given equations (23) and (24), we can write

$$\begin{aligned} \Delta MSE &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{25}\sum_{p=1}^{25}\left(\varepsilon_t^{p,RB}\right)^2 - \left(\varepsilon_t^{p,OI}\right)^2\right)\cdot 12\cdot 100 \\ &= \frac{12\cdot 100}{25}\sum_{p=1}^{25}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2\right). \end{aligned} \tag{40}$$

Thus, $\frac{1}{T}\sum\limits_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2 = \frac{1}{T}\sum\limits_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2$ implies $\Delta MSE = 0$. A similar argument applies for the R^2 since from equations (25) and (26), we have

$$\begin{aligned} R^2 &= \frac{100}{25}\sum_{p=1}^{25}\left(1 - \frac{\sum_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2}{\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2}\right) \\ &= \frac{100}{25}\sum_{p=1}^{25}\left(\frac{\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2}\right). \end{aligned} \tag{41}$$

Therefore $\frac{1}{T}\sum\limits_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2 = \frac{1}{T}\sum\limits_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2$ also results in $R^2 = 0$.

In order to draw statistical inference on the significance level of ΔMSE and R^2, we must obtain an estimate of their respective distribution under H_0. To this end, we first adjust the sample average of $\left(\varepsilon_t^{p,OI}\right)^2$ to match that of $\left(\varepsilon_t^{p,RB}\right)^2$ for each portfolio in a given set of test assets. This is done by recentering $\left(\tilde{\varepsilon}_t^{p,OI}\right)^2$ according to

$$\left(\tilde{\varepsilon}_t^{p,OI}\right)^2 = \left(\varepsilon_t^{p,OI}\right)^2 - \frac{1}{T}\sum_{u=1}^{T}\left(\varepsilon_u^{p,OI}\right)^2 + \frac{1}{T}\sum_{u=1}^{T}\left(\varepsilon_u^{p,RB}\right)^2 \text{ for all } t \text{ and } p. \tag{42}$$

Note that based on the time-series of adjusted squared errors $\left\{\left(\tilde{\varepsilon}_t^{p,OI}\right)^2, \left(\varepsilon_t^{p,RB}\right)^2\right\}_{t=1}^T$, we have $\Delta MSE = 0$ and $R^2 = 0$ by construction.[15]

The bootstrap simulations for ΔMSE and R^2 now proceed as follows:

Step 1: We draw with replacement from $\left\{\left(\tilde{\varepsilon}_u^{p,OI}\right)^2, \left(\varepsilon_u^{p,RB}\right)^2\right\}_{u=1}^T$ to obtain a bootstrapped sample of the squared errors under H_0 of size T. We denote it $\left\{\left(\tilde{\varepsilon}_{u,b}^{p,OI}\right)^2, \left(\varepsilon_{u,b}^{p,RB}\right)^2\right\}_{u=1}^T$ where b refers to a particular bootstrap sample.

Step 2: Using the bootstrapped sample $\left\{\left(\tilde{\varepsilon}_{u,b}^{p,OI}\right)^2, \left(\varepsilon_{u,b}^{p,RB}\right)^2\right\}_{u=1}^T$, we then compute ΔMSE_b and R_b^2 according to equations (23), (24), (25), and (26).

Step 3: We repeat Steps 1 and 2 B times to obtain $\{\Delta MSE_b, R_b^2\}_{b=1}^B$ which can now be used to estimate the distributions of ΔMSE and R^2 under the null hypothesis.

Step 4: Finally, we calculate the one-sided p-value of the ΔMSE and R^2 obtained for a given set of test assets by computing

$$p\text{-}value\,(\Delta MSE) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{\Delta MSE_b > \Delta MSE\right\} \tag{43}$$

$$p\text{-}value\left(R^2\right) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{R_b^2 > R^2\right\} \tag{44}$$

where $\mathbf{1}\{*\}$ is the indicator function.

We are also interested in estimating the significance level of ΔMSE and R^2 when the 25 portfolios of the four test assets are combined together. In this case, we have a total of 100 portfolios instead of 25 when calculating ΔMSE and R^2. We use a similar methodology as the one described above. We first bootstrap $\left\{\left(\tilde{\varepsilon}_u^{p,OI}\right)^2, \left(\varepsilon_u^{p,RB}\right)^2\right\}_{u=1}^T$ for each of the 100 portfolios, compute the corresponding ΔMSE_b and R_b^2, and then infer the one-sided p-values of these statistics.

Appendix A.2: Bootstrapping Confidence Intervals

Given the time-series of the estimators' (unadjusted) squared errors, $\left\{\left(\varepsilon_t^{p,OI}\right)^2, \left(\varepsilon_t^{p,RB}\right)^2\right\}_{t=1}^T$, the bootstrap simulations for the confidence intervals proceed as follows:

[15] Adjusting the regression-based squared errors instead of the option-implied errors does not change their respective variance and thus has no impact on the results.

Step 1: We draw with replacement from $\left\{\left(\varepsilon_t^{p,OI}\right)^2, \left(\varepsilon_t^{p,RB}\right)^2\right\}_{t=1}^{T}$ to obtain a bootstrapped sample of size T of the squared errors. We denote it $\left\{\left(\varepsilon_{u,b}^{p,OI}\right)^2, \left(\varepsilon_{u,b}^{p,RB}\right)^2\right\}_{u=1}^{T}$ where b denotes a particular bootstrap sample.

Step 2: Using the bootstrapped sample $\left\{\left(\varepsilon_{u,b}^{p,OI}\right)^2, \left(\varepsilon_{u,b}^{p,RB}\right)^2\right\}_{u=1}^{T}$, we then compute ΔMSE_b and R_b^2 according to equations (23), (24), (25) and (26).

Step 3: We repeat Steps 1 and 2 B times to obtain $\{\Delta MSE_b, R_b^2\}_{b=1}^{B}$ which can now be used to estimate the empirical distributions of ΔMSE and R^2 including their .05 and .95 percentiles.

Appendix B: Extracting Option-Implied Moments

We implement estimation of the risk-neutral second and third moments using the method of Bakshi and Madan (2000). We use 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 available range, 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.

Let S_t denote the value of the market index and $R_{m,t+\tau} = \ln S_{t+\tau} - \ln S_t$ its return over the horizon τ. We can get the risk-neutral second moment via

$$\begin{aligned} E_t^Q\left[R_{m,t+\tau}^2\right] &= e^{r\tau}\int_{S_t}^{\infty}\frac{2\left(1-\ln\left[K/S_t\right]\right)}{K^2}C_t\left(\tau,K\right)dK \\ &\quad +e^{r\tau}\int_{0}^{S_t}\frac{2\left(1+\ln\left[S_t/K\right]\right)}{K^2}P_t\left(\tau,K\right)dK. \end{aligned}$$

where $C_t\left(\tau,K\right)$ and $P_t\left(\tau,K\right)$ are call and put options quoted at time t with maturity τ and strike price K. We can get the option-implied third moment via

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

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 continuum 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.

Appendix C: Dynamic Conditional Skewness

Our implementation of the Jondeau and Rockinger (2003) model is close to the model they refer to as Model 2. 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 t, 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

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

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 $\eta_t > 3$. We estimate the model monthly by maximum likelihood using 10-year rolling windows of returns.

Appendix D: General Nonlinear SDFs

We now present the general case which nests the results in Propositions 1 and 2.

Proposition 3 *In the absence of arbitrage opportunities, if the stochastic discount factor (SDF) has the following form:*

$$m_{t+1} = a_t + \sum_i b_{i,t} \left(G_i(R_{m,t+1}) - E_t^P \left(G_i(R_{m,t+1}) \right) \right) + \sum_l c_{l,t} \left(f_{l,t+1} - E_t^P (f_{l,t+1}) \right), \tag{45}$$

where $G_i(R_{m,t+1})$ is a nonlinear function of the market return and $f_{l,t+1}$ is the realization of risk factor l, then the cross-sectional pricing restriction for stock returns is

$$E_t^P(R_{j,t+1}) - R_{f,t} = \sum_i \lambda_t^i \beta_{j,t}^i + \sum_l \gamma_t^l \beta_{j,t}^l, \tag{46}$$

where $\beta_{j,t}^i$ and $\beta_{j,t}^l$ are from the projection of asset returns on $G_i(R_{m,t+1})$ 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 G_i, of the market return, λ_t^i, is

$$\lambda_t^i = E_t^P(G_i(R_{m,t+1})) - E_t^Q(G_i(R_{m,t+1})), \tag{47}$$

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. Given equation (45), then applying equation (14) for $\Psi_{t+1} \equiv G_i(R_{m,t+1})$, we obtain equation (47). ■

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 allow for factors that are not necessarily functions of the market return.

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] Azalideh, S., M. Brandt, and F. Diebold (2002), Range-Based Estimation of Stochastic Volatility Models, *Journal of Finance*, 57, 1047–1091.

[4] Bakshi, G., and N. Kapadia (2003), Delta-Hedged Gains and the Negative Market Volatility Risk Premium, *Review of Financial Studies*, 16, 527–566.

[5] 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.

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

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

[8] 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.

[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] Bali, T. G., and H. Zhou (2014), Risk, Uncertainty, and Expected Returns, Forthcoming in the *Journal of Financial and Quantitative Analysis.*

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

[12] Barone-Adesi, G. (1985), Arbitrage Equilibrium with Skewed Asset Returns, *Journal of Financial and Quantitative Analysis*, 20, 299–313.

[13] Barras, L., and A. Malkhozov (2016), Does Variance Risk Have Two Prices? Evidence from the Equity and Option Markets, *Journal of Financial Economics*, 121, 79–92.

[14] Barro, R. (2006), Rare Disasters and Asset Markets in the Twentieth Century, *Quarterly Journal of Economics*, 121, 823–866.

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

[16] Bollerslev, T., J. Marrone, L. Xu, and H. Zhou (2014), Predicting Stock Returns with Variance Risk Premia: Statistical Inference and International Evidence, *Journal of Financial and Quantitative Analysis*, 49, 633–661.

[17] Bollerslev, T., and V. Todorov (2011), Tails, Fears, and Risk Premia, *Journal of Finance*, 66, 2165–2211.

[18] Brandt, M., and C. Jones (2006), Volatility Forecasting with Range-Based EGARCH Models, *Journal of Business and Economic Statistics*, 24, 470–486.

[19] 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.

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

[21] 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.

[22] Chabi-Yo, F., D. Leisen, and E. Renault (2014), Aggregation of Preferences for Skewed Asset Returns, *Journal of Economic Theory*, 154, 453–489.

[23] Chang, B.-Y., P. Christoffersen, and K. Jacobs (2013), Market Skewness Risk and the Cross-Section of Stock Returns, *Journal of Financial Economics*, 107, 46-68.

[24] Chaudhuri, R., and M. Schroder (2015), Monotonicity of the Stochastic Discount Factor and Expected Option Returns, *Review of Financial Studies*, Forthcoming.

[25] Chen, H., S. Joslin, and N.-K. Tran (2012), Rare Disasters and Risk Sharing with Heterogeneous Beliefs, *Review of Financial Studies*, 25, 2189–2224.

[26] Chou, R. (2005), Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range, *Journal of Money, Credit and Banking*, 37, 561–582.

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

[28] Conrad, J., R. Dittmar, and E. Ghysels (2013), Ex Ante Skewness and Expected Stock Returns, *Journal of Finance*, 68, 85–124.

[29] Constantinides, G., and D. Duffie (1996), Asset Pricing with Heterogeneous Consumers, *Journal of Political Economy*, 104, 219–240.

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

[31] Cremers, M., M. Halling, and D. Weinbaum (2015), Aggregate Jump and Volatility Risk in the Cross-Section of Stock Returns, *Journal of Finance*, 70, 577–614.

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

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

[34] Drechsler, I., and A. Yaron (2011), What's Vol Got to Do with It, *Review of Financial Studies*, 24, 1–45.

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

[36] 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.

[37] Fama, E.F., and K. R. French (1997), Industry Costs of Equity, *Journal of Financial Economics*, 43, 153–193.

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

[39] 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.

[40] Gabaix, X. (2012), Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance, *Quarterly Journal of Economics*, 127, 645–700.

[41] Ghysels, E. (1998), On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?, *Journal of Finance*, 53, 549–573.

[42] Ghysels, E., A. Plazzi, and R. Valkanov (2015), Why Invest in Emerging Markets? The Role of Conditional Return Asymmetry, *Journal of Finance*, Forthcoming.

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

[44] Gospodinov, N., R. Kan, and C. Robotti (2014), Misspecication-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors, *Review of Financial Studies*, 27, 2139–2170.

[45] Gourio, F. (2012), Disaster Risk and Business Cycles, *American Economic Review*, 102, 2734–2766.

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

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

[48] Hansen, L., and K. Singleton (1982), Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models, *Econometrica*, 50, 1269–1286.

[49] Heston, S. (1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies,* 6, 327–343.

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

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

[52] Jagannathan, R., and Z. Wang (1998), An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression, *Journal of Finance*, 53, 1285–1309.

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

[54] Julliard, C., and A. Ghosh (2012), Can Rare Events Explain the Equity Premium Puzzle?, *Review of Financial Studies*, 25, 3037–3076.

[55] Kan, R., and C. Zhang (1999), Two-Pass Tests of Asset Pricing Models with Useless Factors, *Journal of Finance*, 54, 204–235.

[56] Kelly, B., and H. Jiang (2014), Tail Risk and Asset Prices, *Review of Financial Studies*, 27, 2841–2871.

[57] Kleibergen, F. (2009), Tests of Risk Premia in Linear Factor Models, *Journal of Econometrics*, 149, 149–173.

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

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

[60] Levi, Y., and I. Welch (2014), Long-Term Capital Budgeting, Working Paper, University of California, Los Angeles.

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

[62] Liu, J., J. Pan, and T. Wang, 2005, An equilibrium model of rare-event premia and its implication for option smirks, *Review of Financial Studies*, 18, 131–164.

[63] Mehra, R., and E. Prescott (1985), The Equity Premium: A Puzzle, *Journal of Monetary Economics*, 15, 145–161.

[64] Mitton, T., and K. Vorkink (2007), Equilibrium Underdiversification and the Preference for Skewness, *Review of Financial Studies*, 20, 1255–1288.

[65] Patton, A., and A. Timmermann (2010), Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM, and Portfolio Sorts, *Journal of Financial Economics*, 98, 605–625.

[66] 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.

[67] 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.

[68] Rietz, T. (1988), The Equity Risk Premium: A Solution, *Journal of Monetary Economics*, 22, 117–131.

[69] Rogers, L., and S. Satchell (1991), Estimating Variance from High, Low and Closing Prices, *Annals of Applied Probability*, 1, 504–512.

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

[71] Schneider, P. (2015), Generalized Risk Premia, *Journal of Financial Economics*, 116, 487–504.

[72] Schneider, P., C. Wagner, and J. Zechner (2015), Low-Risk Anomalies? Working Paper, Copenhagen Business School.

[73] 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.

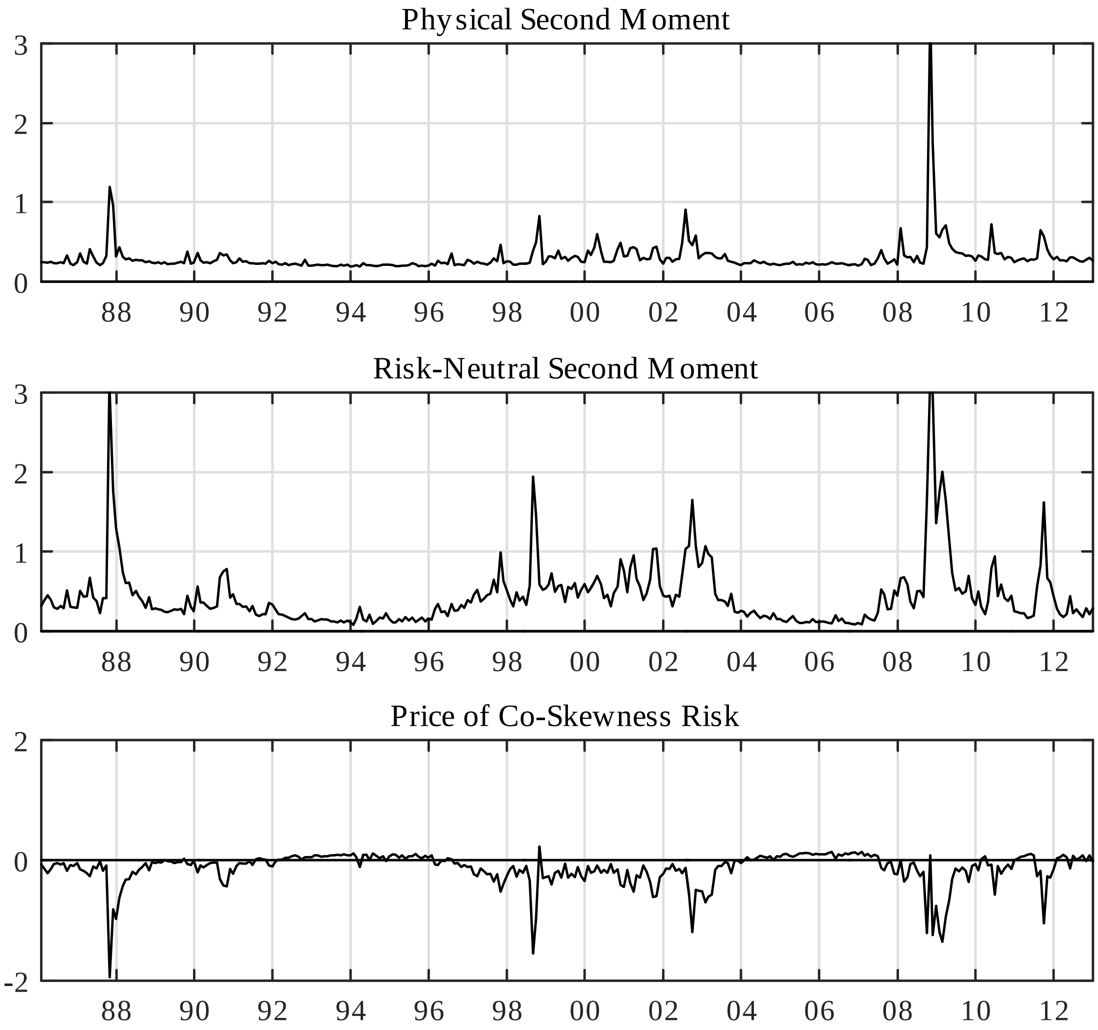
[74] Shanken, J. (1992), On the Estimation of Beta Pricing Models, *Review of Financial Studies*, 5, 1–33.

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

[76] Wachter, J. (2013), Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?, *Journal of Finance*, 68, 987–1035.

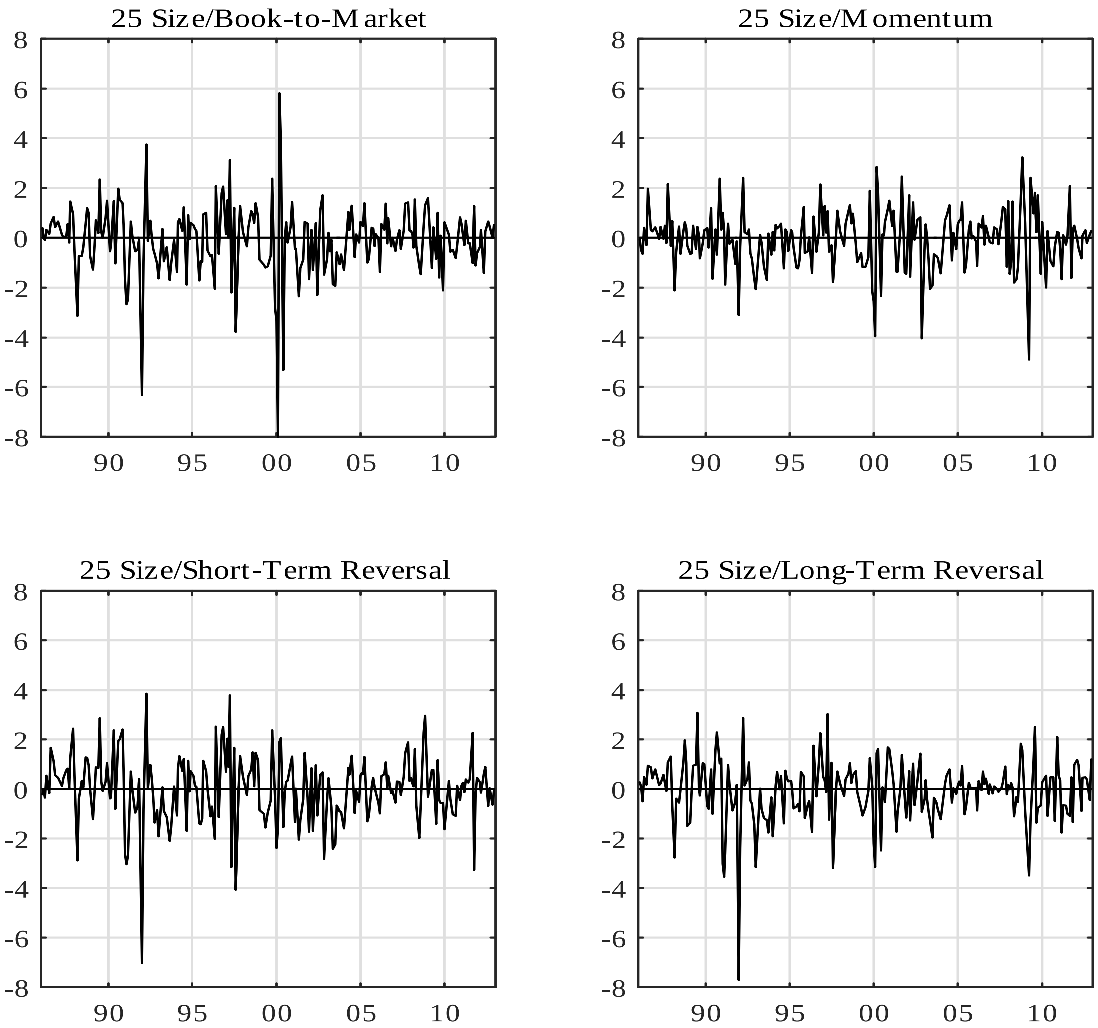
[77] Welch, I., and A. Goyal (2008), A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *Review of Financial Studies*, 21, 1455–1508.

Figure 1: Option-Based Estimates of the Price of Co-Skewness Risk.



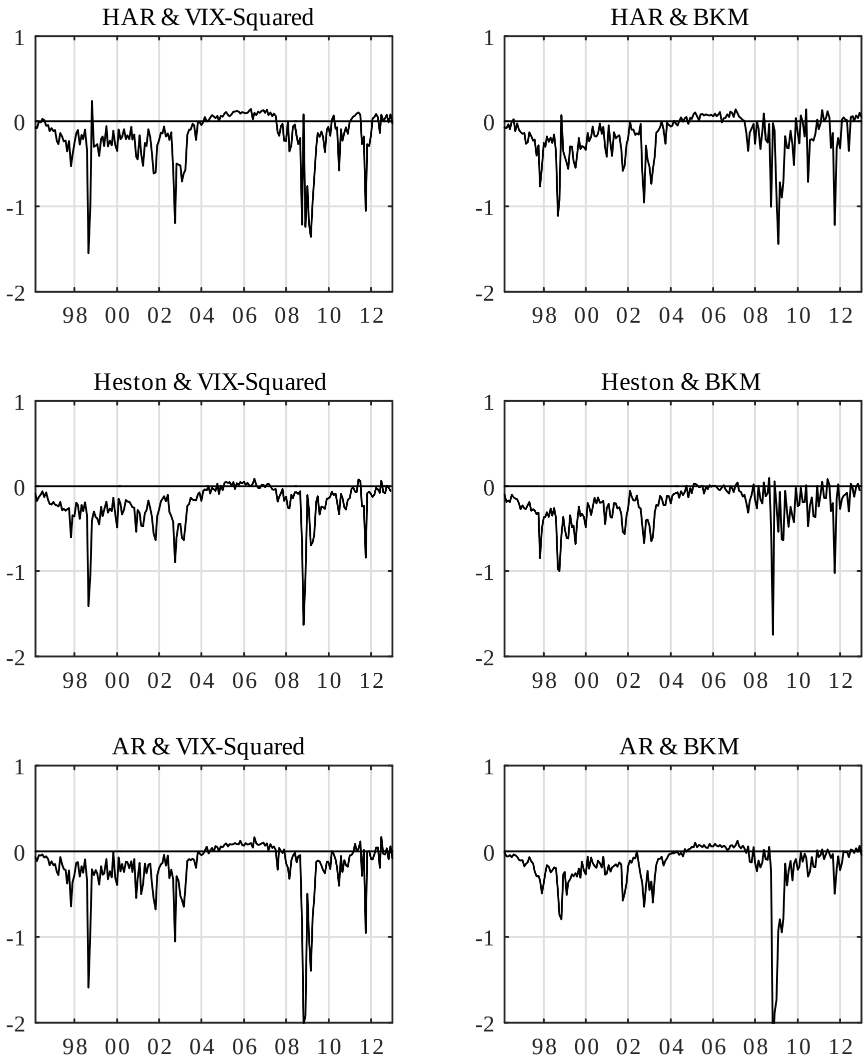
Notes to Figure: We plot the time series for the conditional physical and risk-neutral second moments (monthly in percentages) and the price of co-skewness risk, multiplied by 100 for expositional convenience. The physical second moment is estimated using the HAR model and the risk-neutral second moment is constructed using VIX^2. 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: Regression-Based Estimates of the Price of Co-Skewness Risk.



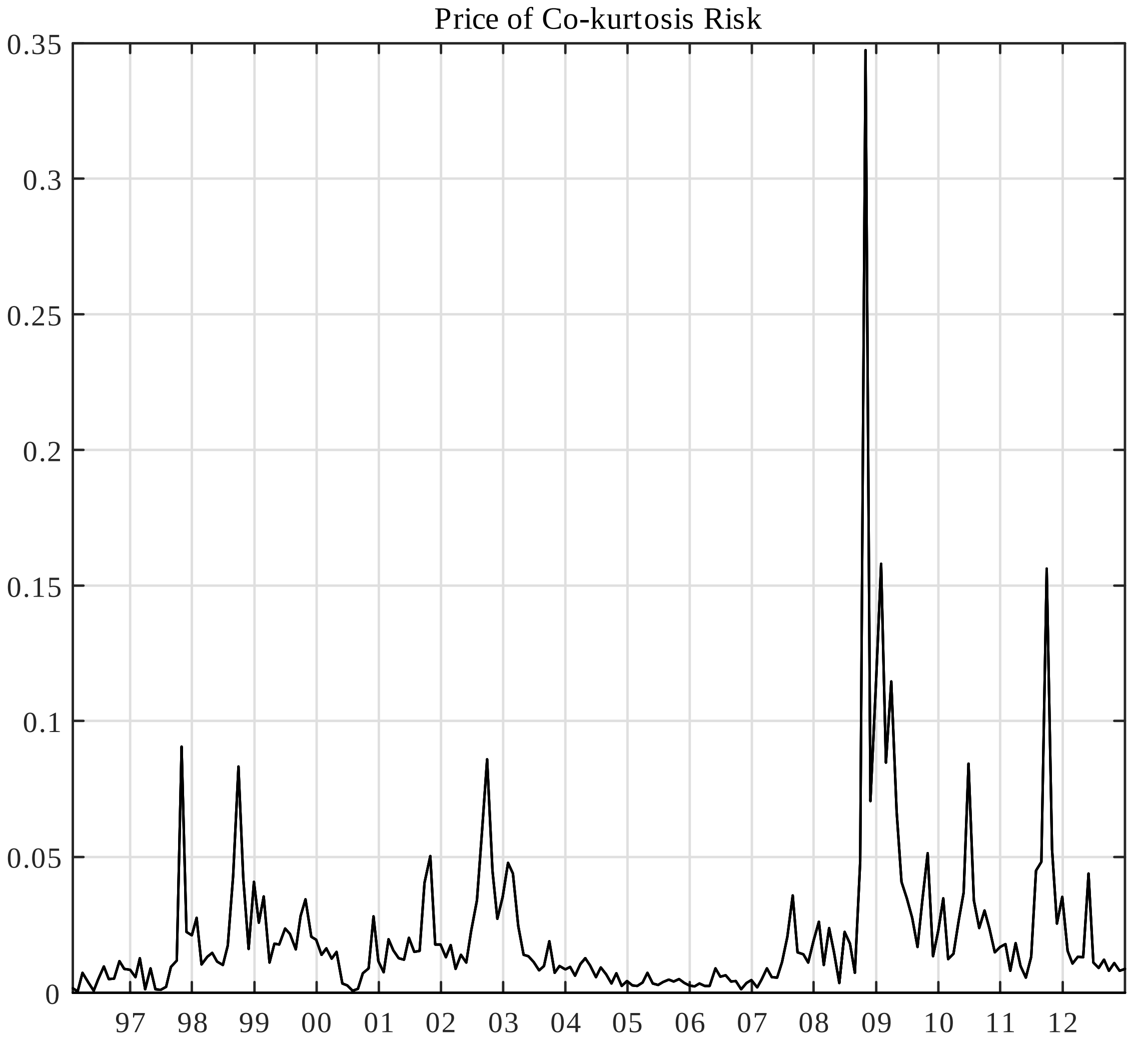
Notes to Figure: We plot time series of the cross-sectional prices of co-skewness risk, multiplied by 100 for expositional convenience. Each month, we estimate factor exposures using a 60-month rolling window of monthly returns from a time series regression of excess portfolio returns on the factors. We then run a cross-sectional regression of returns on estimated exposures to obtain the risk premiums. 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 3: Various Option-Implied Estimates of the Price of Co-Skewness Risk Using Alternative Models for Physical and Risk Neutral Variance. 1996-2012.



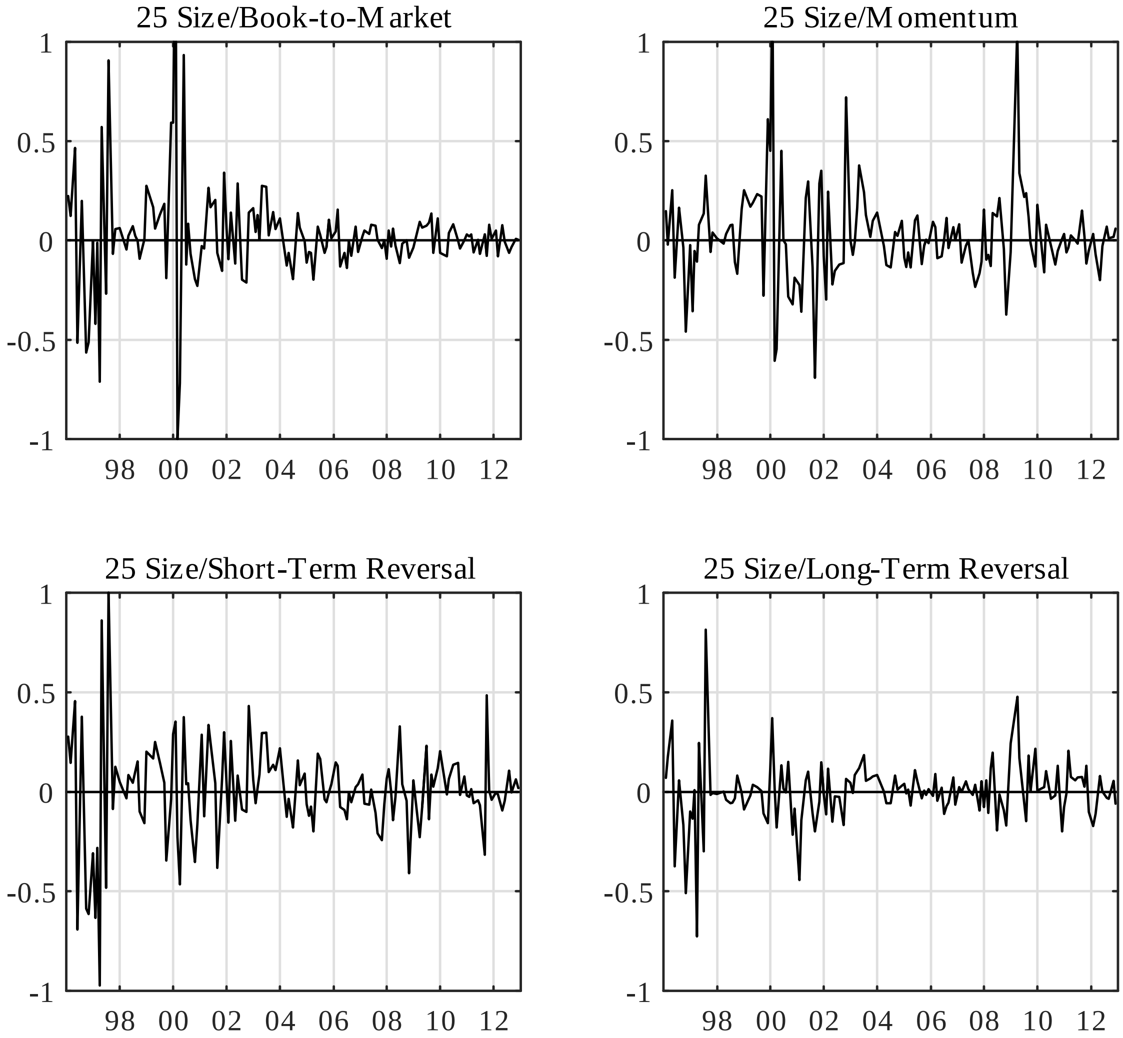
Notes to Figure: We plot the price of co-skewness risk defined as the physical variance minus the risk-neutral variance. The physical variance is either from the Corsi (2009) HAR model estimated on daily high-low returns, from the Heston (1993) model estimated on returns, or from an AR model estimated on realized monthly variances from daily squared returns. The risk-neutral variance is either the square of the VIX index from the CBOE or based on Bakshi, Kapadia and Madan (BKM, 2003).

Figure 4: Option-Based Estimates of the Price of Co-Kurtosis Risk



Notes to Figure: We plot the price of co-kurtosis risk, multiplied by 100 for expositional convenience. We report on the benchmark case where the physical third moment is set equal to zero. The risk-neutral moment is 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.



Notes to Figure: We plot time series of the cross-sectional prices of co-kurtosis risk, multiplied by 100 for expositional convenience. Each month, we estimate factor exposures using a 60-month rolling window of monthly returns from a time series regression of excess portfolio returns on the factors. We then run a cross-sectional regression of returns on estimated exposures to obtain the risk premiums. 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 1996 through December 2012.

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

We provide descriptive statistics for the physical and risk-neutral second moments and the price of co-skewness risk. The physical second moment is estimated using the HAR model and the risk-neutral second moment is proxied by the squared VIX. The time-varying price of co-skewness risk is equal to the spread between the physical and risk-neutral moments. The moments and prices of risk are multiplied by 100 for expositional convenience. The data are monthly and the sample period is from January 1986 to December 2012.

	Physical Second Moment	Risk Neutral Second Moment	Price of Co-Skewness Risk
mean	0.3034	0.4499	-0.1464
std	0.2168	0.4133	0.2289
skew	8.9863	3.4318	-2.2645
kurt	109.6172	19.1077	12.8347
$\rho(1)$	0.4996	0.7525	0.4352

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

We show the results of cross-sectional Fama-MacBeth regressions using monthly returns. Each month, we estimate factor exposures using a 120-month rolling window of monthly returns from a time series regression of excess portfolio returns on the factors. We then run a cross-sectional regression of returns on estimated exposures to obtain the price of risk, λ. We report the mean (in percentage) of the price of risk estimates and the Fama-MacBeth t-statistics with Newey-West correction for serial correlation, using 1 lag. We consider two periods, 1986-2012 and 1996-2012, and four sets of test assets, obtained using sorts on size and respectively book-to-market (BM), momentum (Mom), short-term reversal (STR), and long-term reversal (LTR).

Panel A: 1986-2012

	25 Size/BM			25 Size/Mom			25 Size/STR			25 Size/LTR		
λ^0	1.535	1.489	1.013	0.025	-0.190	0.388	0.653	0.353	0.698	0.511	0.470	-0.353
	2.74	*2.30*	*2.25*	*0.06*	*-0.43*	*0.84*	*1.64*	*0.84*	*1.25*	*1.58*	*1.28*	*-0.78*
λ^{MKT}	-0.617	-0.634	-0.240	0.813	1.002	0.441	0.165	0.494	0.034	0.440	0.442	1.141
	-1.04	*-0.98*	*-0.46*	*1.59*	*1.91*	*0.90*	*0.33*	*0.94*	*0.06*	*1.11*	*1.03*	*2.29*
λ^{HML}			0.104			0.151			0.136			0.131
			0.48			*0.67*			*0.58*			*0.51*
λ^{SMB}			0.443			0.283			0.232			0.652
			1.96			*0.77*			*0.51*			*2.22*
λ^{MOM}			0.395			0.504			-0.355			0.218
			0.63			*1.48*			*-0.64*			*0.46*
λ^{COSK}		-0.098	-0.044		-0.070	0.010		0.008	0.123		-0.094	-0.155
		-1.02	*-0.59*		*-0.90*	*0.15*		*0.08*	*1.67*		*-1.05*	*-1.96*
Adj R^2	15.44	30.42	44.29	14.01	27.40	54.36	13.78	31.86	50.17	7.39	21.28	43.10

Panel B: 1996-2012

	25 Size/BM			25 Size/Mom			25 Size/STR			25 Size/LTR		
λ^0	1.536	1.518	1.352	0.577	0.541	0.675	0.447	0.558	0.845	0.830	0.921	-0.021
	1.94	*1.60*	*2.33*	*1.12*	*0.92*	*1.13*	*0.92*	*1.07*	*1.17*	*2.17*	*2.20*	*-0.04*
λ^{MKT}	-0.702	-0.796	-0.744	0.191	0.186	-0.025	0.256	0.155	-0.292	0.050	-0.085	0.652
	-0.83	*-0.85*	*-1.06*	*0.28*	*0.26*	*-0.04*	*0.40*	*0.23*	*-0.40*	*0.09*	*-0.15*	*1.10*
λ^{HML}			0.210			0.296			0.268			0.214
			0.71			*0.98*			*0.82*			*0.59*
λ^{SMB}			0.610			0.457			0.321			0.864
			1.87			*0.93*			*0.50*			*2.10*
λ^{MOM}			0.917			0.415			-0.311			0.385
			0.97			*0.80*			*-0.42*			*0.63*
λ^{COSK}		-0.108	-0.011		-0.046	0.040		-0.007	0.133		-0.045	-0.100
		-0.89	*-0.13*		*-0.42*	*0.50*		*-0.07*	*1.78*		*-0.50*	*-1.63*
Adj R^2	15.49	28.84	45.22	16.82	28.35	57.74	14.95	31.58	52.34	10.11	22.89	46.37

Table 3: Co-Skewness Out-of-Sample Tests, 1986-2012

For each portfolio p in a given set of test assets, we estimate model k's monthly forecast error as

$$\varepsilon_{t+1}^{p,\mathrm{k}} = R_{p,t+1} - \lambda_{RB,t}^{MKT}\beta_{p,t}^{MKT} - \lambda_{k,t}^{COSK}\beta_{p,t}^{COSK},$$

where $\lambda_{RB,t}^{MKT}$ is the 12-month moving average of the regression based price of market risk, and $\beta_{p,t}^{F}$ is portfolio p's exposure to factor F estimated by multivariate OLS regression based on the most recent 120 months, including month t. $\varepsilon_{t+1}^{p,\mathrm{OI}}$ is then constructed using $\lambda_{OI,t}^{COSK}$ which is the 12-month moving average of the option-implied price of co-skewness risk (estimated using HAR & VIX²) and $\varepsilon_{t+1}^{p,\mathrm{RB}}$ is constructed using $\lambda_{RB,t}^{COSK}$ which is the 12-month moving average of the regression based price of risk. Each month, we calculate from these the cross-sectional difference in mean squared error of the models' forecasts using all 25 portfolios p in a given set of test assets according to

$$\Delta MSE_t = \left(\frac{1}{25}\sum_{p=1}^{25}\left(\varepsilon_t^{p,RB}\right)^2 - \left(\varepsilon_t^{p,OI}\right)^2\right) \times 12 \times 100,$$

where we multiply by 12 to annualize and by 100 to express the difference in percentages. We then compute the sample average of this measure for each test asset, $\Delta MSE = \frac{\sum_{t=1}^{T}\Delta MSE_t}{T}$. We also compute the percentage relative R^2 across portfolios according to

$$R^2 = \frac{1}{25}\sum_{p=1}^{25}\left(1 - \frac{\sum_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2}{\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2}\right) \times 100.$$

Panel A reports the ΔMSE for each set of test assets, the Newey-West p-value, the bootstrapped (B.S.) one-sided p-value of ΔMSE, and the 90% confidence bounds. Panel B reports the R^2 for each set of test assets, its bootstrapped p-value and 90% confidence bounds. In the last column of each panel, we report results obtained by using all 100 portfolios. We refer to Appendix A.1 and A.2 for further details on the estimation of p-values and confidence bounds. All bootstrap results are based on 100,000 draws. The data are monthly and the sample period is from January 1986 to December 2012.

Panel A: Difference in Mean Squared Error

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
ΔMSE	0.325	0.134	0.233	0.081	0.193
NW p-value	*0.30%*	*5.35%*	*0.30%*	*15.61%*	*0.40%*
B.S. p-value	*0.15%*	*3.37%*	*0.14%*	*13.58%*	*0.16%*
B.S. 5 Percentile Bound	0.171	0.015	0.112	-0.040	0.091
B.S. 95 Percentile Bound	0.492	0.255	0.359	0.202	0.298

Panel B: Relative R-Squared

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
R^2	5.80	2.61	4.69	1.71	3.70
B.S. p-value	*0.03%*	*2.62%*	*0.05%*	*17.33%*	*0.13%*
B.S. 5 Percentile Bound	3.18	0.32	2.32	-1.29	1.65
B.S. 95 Percentile Bound	8.45	4.82	6.98	4.77	5.74

Table 4: Robustness Analysis: Estimation Window for Beta and the Price of Risk

Panel A reports the relative R-Squared for the 100 test portfolios. Panel B reports the corresponding *p*-values. The estimation of betas is based on rolling windows ranging from 36 to 120 months. The estimation of the price of risk is based on rolling windows ranging from 1 to 120 months. We refer to Table 3 and Appendices A.1 and A.2 for details on the computation of the relative R-Squared and the bootstrapped p-values. The sample periods vary depending on the estimation windows used for the price of risk. For example, the sample period is 1986-2012 when using unsmoothed monthly estimates (the first row in both panels). It is 1991-2012 when using ten-year averages of the monthly estimates (the next-to-last row in both panels).

Panel A: Relative R-Squared

# of Months used to Compute Price of Risk.	# of Months Used to Estimate Beta: 36	48	60	72	84	96	108	120	Average
1	10.19	10.17	11.87	12.49	14.77	16.94	17.50	16.45	13.80
12	2.83	2.30	2.17	2.81	2.49	3.99	3.89	3.70	3.02
24	2.10	1.37	0.53	1.17	0.85	1.46	1.13	1.24	1.23
36	1.88	1.15	0.56	0.98	0.90	1.42	1.23	1.46	1.20
48	1.29	0.99	0.41	0.88	0.72	0.93	0.72	0.85	0.85
60	0.76	0.72	0.35	0.65	0.65	0.63	0.41	0.71	0.61
72	0.33	0.46	0.24	0.59	0.67	0.57	0.47	0.67	0.50
84	0.04	0.44	0.26	0.53	0.74	0.66	0.60	0.82	0.51
96	0.05	0.48	0.24	0.44	0.68	0.61	0.57	0.68	0.47
108	0.23	0.63	0.64	0.78	0.96	0.82	0.73	0.78	0.70
120	0.20	0.61	0.52	0.58	0.67	0.61	0.56	0.62	0.55
Average	1.81	1.76	1.62	1.99	2.19	2.60	2.53	2.54	2.13

Panel B: One-Sided Bootstrapped P-Values

# of Months used to Compute Price of Risk.	# of Months Used to Estimate Beta: 36	48	60	72	84	96	108	120	Average
1	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
12	0.2%	0.6%	1.2%	1.0%	1.5%	0.0%	0.1%	0.1%	0.6%
24	0.1%	0.5%	20.6%	6.6%	15.7%	4.5%	11.5%	9.6%	8.6%
36	0.0%	0.4%	11.1%	5.0%	8.1%	1.0%	3.7%	2.6%	4.0%
48	0.0%	0.1%	14.0%	3.6%	8.7%	2.6%	9.4%	9.1%	5.9%
60	1.5%	0.8%	12.5%	4.3%	8.5%	7.3%	20.5%	12.3%	8.5%
72	14.7%	5.2%	15.2%	2.7%	3.8%	4.8%	13.1%	10.0%	8.7%
84	43.5%	5.6%	9.6%	2.6%	1.3%	1.2%	4.9%	3.9%	9.1%
96	41.9%	4.2%	10.1%	4.0%	1.2%	0.6%	3.3%	4.9%	8.8%
108	17.2%	1.2%	0.1%	0.1%	0.1%	0.1%	1.0%	2.3%	2.8%
120	18.6%	0.5%	0.6%	1.3%	1.8%	1.0%	3.0%	3.8%	3.8%
Average	12.5%	1.7%	8.6%	2.8%	4.6%	2.1%	6.4%	5.3%	5.5%

Table 5: Robustness Analysis for the Price of Co-Skewness Risk

In Panel A, we provide estimates of the average price of co-skewness risk using alternative estimators of the monthly physical and risk-neutral moments. The moments and prices of risk are multiplied by 100 for expositional convenience. In Panel B, we document the ΔMSE and R^2 for each set of test assets and for the 100 portfolios (last column), using different moment estimators for the option-implied price of co-skewness risk. We refer to Table 3 and Appendices A.1 and A.2 for details on the computation of these statistics and their bootstrapped (B.S.) p-values. The sample periods are 1986-2012 and 1996-2012.

Panel A: Average Price of Co-Skewness Risk

Sample Period	Physical 2nd Moment	Risk-Neutral 2nd Moment	Physical Estimate	Risk-Neutral Estimate	Average Price of Co-Skewness Risk
1986-2012	Heston	VIX^2	0.2519	0.4499	-0.1979
1986-2012	AR	VIX^2	0.2921	0.4499	-0.1578
1996-2012	HAR	BKM	0.3405	0.5153	-0.1748
1996-2012	Heston	BKM	0.2944	0.5153	-0.2209
1996-2012	AR	BKM	0.3367	0.5153	-0.1786

Panel B: Difference in Mean Squared Error and Relative R-Squared

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
		Heston & VIX^2 (1986-2012)			
ΔMSE	0.330	0.128	0.226	0.074	0.190
B.S. p-value	*0.11%*	*4.23%*	*0.18%*	*15.98%*	*0.19%*
R^2	5.87	2.48	4.55	1.53	3.61
B.S. p-value	*0.02%*	*3.44%*	*0.07%*	*20.45%*	*0.21%*
		AR & VIX^2 (1986-2012)			
ΔMSE	0.325	0.129	0.231	0.080	0.191
B.S. p-value	*0.16%*	*3.89%*	*0.14%*	*14.00%*	*0.18%*
R^2	5.79	2.54	4.67	1.69	3.67
B.S. p-value	*0.03%*	*2.90%*	*0.06%*	*17.81%*	*0.14%*
		HAR & BKM (1996-2012)			
ΔMSE	0.492	0.223	0.207	0.107	0.257
B.S. p-value	*0.14%*	*1.92%*	*1.45%*	*8.91%*	*0.26%*
R^2	7.42	3.74	3.50	1.97	4.16
B.S. p-value	*0.01%*	*1.36%*	*1.27%*	*12.60%*	*0.23%*
		Heston & BKM (1996-2012)			
ΔMSE	0.489	0.215	0.191	0.097	0.248
B.S. p-value	*0.14%*	*2.73%*	*2.51%*	*11.87%*	*0.48%*
R^2	7.33	3.61	3.21	1.72	3.97
B.S. p-value	*0.02%*	*2.07%*	*2.34%*	*16.57%*	*0.54%*
		AR & BKM (1996-2012)			
ΔMSE	0.493	0.229	0.218	0.109	0.262
B.S. p-value	*0.17%*	*1.62%*	*1.12%*	*8.15%*	*0.24%*
R^2	7.44	3.88	3.70	2.03	4.26
B.S. p-value	*0.01%*	*0.99%*	*0.93%*	*11.65%*	*0.19%*

Table 6: Robustness Analysis: 1986-2007 Sample

Panel A reports the ΔMSE for each set of test assets. The option-implied price of co-skewness risk is measured using HAR and VIX². Below the ΔMSE estimates, we report the bootstrapped one-sided *p*-values, allowing for one autocorrelation lag of the monthly differences in mean squared errors, and the bootstrapped 90% confidence interval bounds. We refer to Table 3 and Appendices A.1 and A.2 for details on the computation of ΔMSE and the bootstrapped (B.S.) confidence bounds. Panel B reports for each set of test assets the relative R^2, the bootstrapped p-values of the R^2 and the lower and upper 90% confidence interval bounds. In the last column of both panels, we report on the pricing performance when all 100 portfolios are considered. We refer to Appendices A.1 and A.2 for further details about the methodology used to estimate the p-values and confidence bounds of the R^2. All bootstrap results are based on 100,000 sample draws. The data are monthly and the sample period is from January 1986 to December 2007.

Panel A: Difference in Mean Squared Error

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
ΔMSE	0.379	0.188	0.267	0.129	0.241
B.S. p-value	*0.20%*	*1.27%*	*0.19%*	*6.12%*	*0.11%*
B.S. 5 Percentile Bound	0.191	0.051	0.125	-0.006	0.118
B.S. 95 Percentile Bound	0.582	0.326	0.417	0.267	0.367

Panel B: Relative R-Squared

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
R^2	7.01	3.93	5.99	3.43	5.09
B.S. p-value	*0.04%*	*1.00%*	*0.04%*	*7.78%*	*0.09%*
B.S. 5 Percentile Bound	3.69	1.00	2.88	-0.54	2.37
B.S. 95 Percentile Bound	10.29	6.76	8.89	7.53	7.78

Table 7: Co-Kurtosis Out-of-Sample Tests, 1996-2012

For each portfolio p in a given set of test assets, we estimate model k's monthly forecast error as

$$\varepsilon_{t+1}^{p,k} = R_{p,t+1} - \lambda_{RB,t}^{MKT}\beta_{p,t}^{CAPM} - \lambda_{k,t}^{COKU}\beta_{p,t}^{COKU},$$

where $\lambda_{RB,t}^{MKT}$ is the 12-month moving average of the regression based price of market risk, and $\beta_{p,t}^{F}$ is portfolio p's exposure to factor F estimated by OLS regression based on the most recent 120 months, including month t. $\lambda_{RB,t}^{COKU}$ is the 12-month moving average of the regression based price of co-kurtosis risk and $\lambda_{OI,t}^{COKU}$ is the 12-month moving average of the option-implied price of co-kurtosis risk. Each month, we calculate the cross-sectional difference in mean squared error of the models' forecasts using all portfolios p in a given set of test assets according to

$$\Delta MSE_t = \left(\frac{1}{25}\sum_{p=1}^{25}\left(\varepsilon_t^{p,RB}\right)^2 - \left(\varepsilon_t^{p,OI}\right)^2\right) \times 12 \times 100,$$

where we multiply by 12 to annualize and by 100 to express the difference in percentages. We then compute the sample average of this measure for each test asset, $\Delta MSE = \frac{\sum_{t=1}^{T}\Delta MSE_t}{T}$. We compute the percentage relative R^2 across portfolios according to

$$R^2 = \frac{1}{25}\sum_{p=1}^{25}\left(1 - \frac{\sum_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2}{\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2}\right) \times 100.$$

In Panel A, we provide estimates of the average price of co-kurtosis risk using alternative estimators of the monthly physical and risk-neutral moments. Panel B reports the ΔMSE for each set of test assets, the bootstrapped (B.S.) one-sided p-value of ΔMSE, as well as the 90% confidence interval bounds. We also report the R^2 for each set of test assets, its bootstrapped p-value and 90% confidence bounds. In the last column of Panel B, we report on results when considering all 100 portfolios. We refer to Appendices A.1 and A.2 for further details about the methodology developed to estimate p-values and confidence bounds. All bootstrap results are based on 100,000 draws. The data are monthly and the sample period is from January 1996 to December 2012.

Panel A: The Average Price of Co-Kurtosis Risk

Physical Third Moment Specification	Average Physical Third Moment	Average Risk-Neutral Third Moment	Average Price of Co-Kurtosis Risk
Zero Skewness	0	-0.0220	0.0220
Constant Skewness from Daily Returns	-0.0006	-0.0220	0.0214
Jondeau and Rockinger (2003)	-0.0104	-0.0220	0.0116

Panel B: Difference in Mean Squared Error and Relative R-Squared

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
		Zero Skewness & BKM			
ΔMSE	0.199	0.488	0.296	0.016	0.250
B.S. p-value	*0.93%*	*0.04%*	*0.25%*	*37.83%*	*0.01%*
R^2	3.45	8.66	5.63	0.44	4.54
B.S. p-value	*0.39%*	*0.00%*	*0.05%*	*35.37%*	*0.00%*
		Constant Skewness & BKM			
ΔMSE	0.199	0.487	0.294	0.017	0.249
B.S. p-value	*0.95%*	*0.04%*	*0.26%*	*36.95%*	*0.01%*
R^2	3.44	8.64	5.61	0.46	4.54
B.S. p-value	*0.41%*	*0.00%*	*0.06%*	*34.51%*	*0.00%*
		Jondeau and Rockinger (2003) & BKM			
ΔMSE	0.182	0.464	0.278	0.001	0.231
B.S. p-value	*2.04%*	*0.07%*	*0.40%*	*50.55%*	*0.03%*
R^2	3.16	8.29	5.33	0.12	4.22
B.S. p-value	*1.06%*	*0.00%*	*0.10%*	*46.63%*	*0.00%*

Table 8: Co-Skewness and Co-Kurtosis Out-of-Sample Tests, 1996-2012

For each portfolio *p* in a given set of test assets, we estimate model *k*'s monthly forcast error as

$$\varepsilon_{t+1}^{p,\mathrm{k}} = R_{p,t+1} - \lambda_{RB,t}^{MKT}\beta_{p,t}^{CAPM} - \lambda_{k,t}^{COSK}\beta_{p,t}^{COSK} - \lambda_{k,t}^{COKU}\beta_{p,t}^{COKU},$$

where $\lambda_{RB,t}^{MKT}$ is the 12-month moving average of the regression based price of market risk, and $\beta_{p,t}^{F}$ is portfolio *p*'s exposure to factor *F* estimated by OLS regression based on the most recent 120 months, including month *t*. $\lambda_{RB,t}^{COSK}$ and $\lambda_{RB,t}^{COKU}$ are the 12-month moving average of the regression based prices of risk. $\lambda_{OI,t}^{COSK}$ and $\lambda_{OI,t}^{COKU}$ are the 12-month moving average of the option-implied prices of risk. Each month, we calculate the cross-sectional difference in mean squared error of the models' forecast using all portfolios p in a given set of test assets according to

$$\Delta MSE_t = \left(\frac{1}{25}\sum_{p=1}^{25}\left(\varepsilon_t^{p,RB}\right)^2 - \left(\varepsilon_t^{p,OI}\right)^2\right) \times 12 \times 100,$$

where we multiply by 12 to annualize and by 100 to express the difference in percentages. We then compute the sample average of this measure for each test asset, $\Delta MSE = \frac{\sum_{t=1}^{T}\Delta MSE_t}{T}$. We compute the percentage relative R^2 across portfolios according to

$$R^2 = \frac{1}{25}\sum_{p=1}^{25}\left(1 - \frac{\sum_{t=1}^{T}\left(\varepsilon_t^{p,OI}\right)^2}{\sum_{t=1}^{T}\left(\varepsilon_t^{p,RB}\right)^2}\right) \times 100.$$

Panel A reports the ΔMSE for each set of test assets, the bootstrapped (B.S.) one-sided *p*-value of ΔMSE and its 90% confidence interval bounds. Panel B reports the average R^2 for each set of test assets, its bootstrapped *p*-value and 90% confidence bounds. In the last column of both panels, we report results when considering all 100 portfolios. We refer to Appendix A.1 and A.2 for further details about the methodology developed to estimate *p*-values and confidence bounds. All bootstrap results are based on 100,000 draws. The data are monthly and the sample period is from January 1996 to December 2012.

Panel A: Difference in Mean Squared Error

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
	COSK: HAR& BKM, COKU: Zero Skew & BKM				
ΔMSE	0.483	0.542	0.360	0.258	0.411
B.S. p-value	*0.00%*	*0.00%*	*0.00%*	*1.30%*	*0.00%*
B.S. 5 Percentile Bound	0.265	0.307	0.169	0.049	0.235
B.S. 95 Percentile Bound	0.710	0.775	0.560	0.464	0.585
	COSK: HAR& BKM, COKU: Constant Skew & BKM				
ΔMSE	0.486	0.544	0.360	0.260	0.413
B.S. p-value	*0.03%*	*0.01%*	*0.22%*	*2.00%*	*0.01%*
B.S. 5 Percentile Bound	0.267	0.313	0.167	0.048	0.236
B.S. 95 Percentile Bound	0.710	0.781	0.562	0.469	0.591

Panel B: Average Relative R-Squared

	25 Size/BM	25 Size/Mom	25 Size/STR	25 Size/LTR	All
	COSK: HAR& BKM, COKU: Zero Skew & BKM				
R^2	7.80	9.40	6.37	5.49	7.27
B.S. p-value	*0.00%*	*0.00%*	*0.00%*	*1.30%*	*0.00%*
B.S. 5 Percentile Bound	4.46	6.05	3.13	1.10	4.24
B.S. 95 Percentile Bound	10.92	12.81	9.72	10.10	10.03
	COSK: HAR& BKM, COKU: Constant Skew & BKM				
R^2	7.84	9.41	6.38	5.54	7.29
B.S. p-value	*0.00%*	*0.00%*	*0.09%*	*1.73%*	*0.00%*
B.S. 5 Percentile Bound	4.61	5.97	3.10	1.08	4.29
B.S. 95 Percentile Bound	10.94	12.68	9.69	10.00	10.20