Robust Conditional Kurtosis and the Cross-Section of International Stock Returns*

Ruifeng Liu
Anheuser-Busch InBev

Alex Maynard
University of Guelph

Ilias Tsiakas
University of Guelph

June 2024

Abstract

We introduce robust kurtosis, which is a new quantile-based measure for the kurtosis of stock returns. For approximately normal distributions, robust kurtosis is equivalent to the traditional moment-based kurtosis. For fat-tailed distributions, when kurtosis matters the most, robust kurtosis provides a distinct and reliable measure. Using a cross-section of international stock index returns, we find that robust kurtosis carries a significant negative premium: higher robust kurtosis is related to lower future stock returns, especially for developed markets. This contrasts with the significant positive premium associated with robust skewness, especially for emerging markets.

JEL classification: C58, G11, G15.

Keywords: Robust Kurtosis, Robust Skewness, Conditional Quantiles, MIDAS, International Stock Returns.

*__Acknowledgements:__ The authors are grateful for useful comments and discussions to Brian Ferguson, Chuan Goh, Nikola Gradojevic, Michael Hoy, Joann Jasiak, Fred Liu, Mengke Luo, Dongmeng Ren, Thanasis Stengos, Lars Stentoft, Yiguo Sun, Jianhan Zhang, and seminar participants at St. Mary's University, the University of Guelph, the 2019 CEA conference, the 2019 RCEA conference, the 2020 IAES conference, the 2020 DSE Winter School, the 2023 AiE conference and the 2023 CESG conference. This research was supported by the Social Sciences and Humanities Research Council of Canada (grant number 430417 for Alex Maynard and 430418 for Ilias Tsiakas).

Corresponding author: Ilias Tsiakas, Department of Economics and Finance, Lang School of Business and Economics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Email: itsiakas@uoguelph.ca.

1 Introduction

Outside of special cases, such as quadratic utility or normally distributed returns, skewness and kurtosis both matter to investors. Skewness measures the asymmetry of the underlying return distribution. Kurtosis captures the extent to which the probability mass is concentrated on the center and tails as opposed to the "shoulders" of the return distribution.[1] In the context of financial returns, negative skewness indicates downside risk, high kurtosis reflects tail risk, and the interaction of the two may be important for investment decisions.[2]

In this paper, we introduce robust kurtosis, which is a new quantile-based measure for the kurtosis of financial returns. The new measure of robust kurtosis complements the existing measure of robust skewness (Ghysels, Plazzi and Valkanov, 2016). Since robust kurtosis is based on quantiles, it is by design insensitive to extreme outliers. More importantly, it has the following crucial advantages over existing measures: (1) for the normal distribution, robust kurtosis is perfectly equivalent to the traditional moment-based kurtosis; (2) for moderately fat-tailed distributions, robust kurtosis approximates the moment-based kurtosis and essentially provides the same information; and (3) for distributions with tails, which are sufficiently fat so that the moment-based measure is poorly estimated or even does not exist, robust kurtosis provides a distinct and more reliable alternative to the traditional measure. When kurtosis matters the most, as in the case of non-normal distributions typically observed in financial returns, the performance of robust kurtosis is excellent. In short, robust kurtosis is a reliable measure for all distributions since it approximates well the traditional kurtosis measure when traditional kurtosis can be accurately estimated, but replaces it by a robust alternative when it becomes unreliable. To the best of our knowledge, the robust kurtosis measure and its properties are novel contributions to the literature.

---

1 Balanda and Macgillivray (1988) define kurtosis as the location-free and scale-free movement of probability mass from the shoulders of the distribution to its center and tails.

2 For example, Kraus and Litzenberger (1976), Harvey and Siddique (2000), Neuberger (2012), Jondeau, Zhang and Zhu (2019), and Langlois (2020), among others, assess the role of skewness in the investment decision. The importance of kurtosis is demonstrated by Scott and Horvath (1980), Dittmar (2002), Guidolin and Timmermann (2008), and Jondeau and Rockinger (2012).

We also make an empirical contribution to the literature by evaluating the ability of *conditional* robust kurtosis, together with conditional robust skewness, to predict the cross-section of international stock index returns. The robust kurtosis estimator is based on quantiles, which are straightforward to estimate conditionally, thus leading to a robust measure of conditional kurtosis.[3] Specifically, we assess whether robust skewness and robust kurtosis can consistently predict the stock market performance of a large cross-section of countries. Our data sample comprises the monthly returns of 39 US dollar-denominated country stock indices for the sample period of January 1996 to June 2019. The cross-section includes 23 developed markets and 16 emerging markets.

Our empirical approach involves the following steps. First, we account for the dynamics of robust skewness and robust kurtosis using the conditional mixed-data sampling (MIDAS) quantile model proposed by Ghysels, Plazzi and Valkanov (2016). A crucial advantage of the MIDAS quantile model is that it conditions on daily returns to estimate the quantiles of monthly returns, which is especially suitable given the length of our sample period. Second, we perform standard portfolio sorts based on either robust skewness or robust kurtosis to rank international stock index returns, allocate countries to portfolios and examine the portfolios' future performance. Finally, we estimate Fama and MacBeth (1973) cross-sectional regressions, which formally assess the predictive ability of robust skewness and robust kurtosis, while controlling for volatility and standard economic fundamentals.[4]

Our main empirical finding is that there is a significant *negative* premium associated with robust kurtosis, especially across *developed markets*: the higher the robust kurtosis the lower the future international stock returns. This finding contrasts with the significant *positive* premium associated with robust skewness, especially across *emerging markets*. Overall,

[3] While conditional kurtosis can be modelled parametrically in fat-tailed extensions of GARCH models (see, e.g., Hansen, 1994; Jondeau and Rockinger, 2003; Brooks *et al.*, 2005; León, Rubio and Serna, 2005; Bekaert, Engstrom and Ermolov, 2015), the moment-based estimator cannot be easily generalized to include a conditional version.

[4] Our main analysis is performed ex-ante (i.e., conditioning on $t-1$ information) but in-sample (i.e., using the full sample MIDAS estimates). In the Online Appendix, we also report out-of-sample results.

across all countries, the Fama and MacBeth (1973) regressions indicate that robust kurtosis is dominant over robust skewness.[5]

These findings become stronger when we restrict our analysis to countries with positive excess robust kurtosis, i.e., robust kurtosis higher than the value of 3 associated with the normal distribution. It is sensible that the fourth moment is informative primarily for non-normal distributions indicated by positive excess kurtosis. We consider this restriction to be economically motivated since investors are unlikely to care about kurtosis if the distribution tails are thinner than the normal.[6] Using the restricted cross-section, a zero-investment portfolio that every month goes long on the top 30% of countries with the highest robust kurtosis and short on the bottom 30% of countries with the lowest robust kurtosis delivers an annualized return of −5.3%. In contrast, a similar portfolio based on robust skewness delivers an annualized return of 3.6%.

Consistent with a long line of research, our results confirm the importance of higher moments in predicting expected stock returns. However, the direction of this predictive relation appears to be inconsistent with some previous theoretical contributions. Specifically, the positive cross-sectional relation between skewness and expected stock returns is inconsistent with the theoretical findings of Arditti (1967, 1971), Kraus and Litzenberger (1976), Harvey and Siddique (2000) and Mitton and Vorkink (2007), which predict a negative relation. The negative cross-sectional relation between kurtosis and expected stock returns is inconsistent with the theoretical findings of Scott and Horvath (1980), Kimball (1993) and Dittmar (2002), which predict a positive relation. Our analysis, therefore, sheds new light on this literature by providing evidence of a negative premium associated with conditional robust kurtosis in predicting the cross-section of international stock returns.[7]

---

5 We also find that volatility has an insignificant relation to future stock returns. Therefore, the predictive ability of the third and the fourth moment is distinct from that of the second moment.

6 Imposing economic constraints on kurtosis is roughly analogous to the economic constraints imposed on equity premium prediction by Campbell and Thompson (2008).

7 For additional contributions on the predictive ability of skewness and kurtosis, see Kumar (2009), Boyer, Mitton and Vorkink (2010), Bali, Cakici and Whitelaw (2011), and Amaya *et al.* (2015).

The main implication of our findings is that high conditional skewness or low conditional excess kurtosis is likely to improve the benefits of international diversification. However, in the opposite case of low conditional skewness or high conditional excess kurtosis, a US investor is better off investing domestically by lowering the asset allocation to international stock investments. In other words, our main results provide indirect support for the home bias of a US investor who faces international stock investments with low (negative) robust skewness or high excess robust kurtosis. This is consistent with the analysis of Guidolin and Timmermann (2008), who find that a combination of investor preferences that put weight on the skewness and kurtosis of portfolio returns along with time-variations in international investment opportunities is important in explaining the home bias of US investors.

The new robust kurtosis measure is linked to the original quantile-based kurtosis measure of Moors (1988) but the two measures are distinct. Robust kurtosis has a different interpretation since it can be understood as a transformation of Moors' measure that is approximately comparable to the traditional kurtosis measure, when the latter exists and can be reasonably well estimated. It can therefore be interpreted as the value of the moment-based kurtosis implied by Moors' measure. By contrast, Moors' measure is not easily interpretable and cannot be compared to the traditional kurtosis measure. In addition, the empirical performance (e.g., in portfolio sorts) of Moors' measure is poor compared to robust kurtosis. For these reasons, we believe that both conceptually and empirically robust kurtosis is distinct to Moors' kurtosis.

The paper is organized as follows. In the next section, we set the stage by reviewing standard measures of skewness and kurtosis. Then, we introduce and derive the new robust kurtosis measure. The performance of robust kurtosis for alternative distributions is evaluated through Monte Carlo simulations in Section 3. Section 4 implements conditional measures of robust skewness and kurtosis on the cross-section of international stock returns. In Section 5, we assess the predictive ability of the robust measures. Finally, Section 6 concludes. Detailed econometric derivations and additional material are included in our Online Appendix.

2 New Robust Kurtosis Measure

Among several measures of skewness and kurtosis proposed in the literature, the most popular continue to be the moment-based measures of Pearson (1905). Despite their widespread use, these measures have two shortcomings. First, they require the existence of the third and fourth moments. Otherwise, they are undefined. Second, their estimators are based on sample moments, which are highly sensitive to outliers. The effect of outliers is magnified by being raised to the third and fourth powers (Kim and White, 2004), resulting in estimator variances that depend on the existence of the sixth and eighth moments, respectively. These shortcomings limit the usefulness of moment-based estimators for financial data, where asymmetry, fat tails and large outliers are standard elements of the return distribution.

In this context, it would be useful to derive robust measures for the skewness and kurtosis of the underlying return distribution using information from quantile-based measures so that three conditions hold: (1) the robust measures are insensitive to outliers, (2) they remain well-defined even in extreme cases, when the moments do not exist, and (3) they are equivalent to the moment-based measures for normal and approximately normal distributions.[8] In recent work, Ghysels, Plazzi and Valkanov (2016) do so for robust skewness. In this paper, we do so for robust kurtosis.

2.1 Moment-Based and Quantile-Based Measures

In this section, we briefly define the moment-based and quantile-based measures of skewness and kurtosis for the h-period continuously compounded returns $r_{t,h}$ defined as follows:

$$r_{t,h} = \sum_{j=1}^{h} r_{t+1-j}, \; h \geq 1, \tag{1}$$

8 The term "approximately normal" refers to normal distributions and moderately fat-tailed distributions for short-horizon returns, which are aggregated over long horizons to become approximately normal.

where r_t is the one-period return. We define the mean and variance of $r_{t,h}$ as $\mu_h = E[r_{t,h}]$ and $\sigma_h^2 = E[(r_{t,h} - \mu_h)^2]$. The quantile based measures discussed below depend on $Q_\alpha(r_{t,h})$, which is defined as the α-th quantile of $r_{t,h}$, for the following seven quantile levels: $\boldsymbol{\alpha} = (\alpha_1, \ldots \alpha_7)' = (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875)'$. We collect these quantiles in the vector $\boldsymbol{Q}(r_{t,h}) = (Q_{\alpha_1}(r_{t,h}), Q_{\alpha_2}(r_{t,h}) \ldots Q_{\alpha_7}(r_{t,h}))'$.[9] For notational simplicity, we suppress the subscript h for $h = 1$ and suppress dependence on $r_{t,h}$ when possible.

The conventional moment-based measure of skewness for $r_{t,h}$ is defined as follows:

$$S = E\left[\left(\frac{r_{t,h} - \mu_h}{\sigma_h}\right)^3\right]. \tag{2}$$

As an alternative to $S = S(r_{t,h})$, Bowley (1920) introduces the following quantile-based measure:

$$QS(\boldsymbol{Q}) = \frac{\Big(Q_{0.75}(r_{t,h}) - Q_{0.5}(r_{t,h})\Big) - \Big(Q_{0.5}(r_{t,h}) - Q_{0.25}(r_{t,h})\Big)}{Q_{0.75}(r_{t,h}) - Q_{0.25}(r_{t,h})}. \tag{3}$$

Bowley's skewness measure captures the asymmetry of the interquartile range with respect to the median. QS is equal to zero for all symmetric distributions. A positive (negative) value of QS indicates skewness in the right (left) tail of the distribution. Unlike the moment-based measure, the value of QS is restricted to lie between -1 and 1, inclusive.

The conventional moment-based measure of kurtosis for $r_{t,h}$ is defined as follows:

$$K = E\left[\left(\frac{r_{t,h} - \mu_h}{\sigma_h}\right)^4\right]. \tag{4}$$

As an alternative to $K = K(r_{t,h})$, Moors (1988) proposes the following quantile-based measure:

$$QK(\boldsymbol{Q}) = \frac{\Big(Q_{0.875}(r_{t,h}) - Q_{0.625}(r_{t,h})\Big) + \Big(Q_{0.375}(r_{t,h}) - Q_{0.125}(r_{t,h})\Big)}{Q_{0.75}(r_{t,h}) - Q_{0.25}(r_{t,h})}. \tag{5}$$

In words, Moors' kurtosis measure is defined as the ratio of the sum of the two shoulders of

[9] We use the convention of bold-facing variables to denote vectors across multiple quantiles.

the distribution to the interquartile range. This is based on the intuition that a distribution with fat tails and a peaked center has relatively wide shoulders. The standard normal distribution has a value of $QK = 1.23$, with higher values representing excess kurtosis.

Since they depend on quantiles rather than moments, QS and QK are more robust to outliers than S and K. This is analogous to the greater robustness of the median relative to the mean. However, QS and QK are not directly equivalent to S and K, respectively. For this reason, our analysis focuses on the new robust kurtosis measure, which is directly equivalent to the moment-based estimator for normal or approximately normal distributions.[10]

2.2 Derivation of Robust Kurtosis

In this section, we derive our new measure of robust kurtosis. We begin by noting that QS and QK are location- and scale-invariant. Therefore, we can express QS and QK in terms of the standardized h-period return:

$$z_{t,h} = \frac{r_{t,h} - \mu_h}{\sigma_h}, \tag{6}$$

so that $QS\left(\boldsymbol{Q}(r_{t,h})\right) = QS\left(\boldsymbol{Q}(z_{t,h})\right)$ and $QK\left(\boldsymbol{Q}(r_{t,h})\right) = QK\left(\boldsymbol{Q}(z_{t,h})\right)$. For simplicity, we assume that the one-period return r_t follows an i.i.d. process with a finite third and fourth moment and an absolutely continuous cumulative distribution function (CDF).

The Cornish-Fisher expansion (Cornish and Fisher, 1938), which is an inversion of the Edgeworth expansion (Edgeworth, 1905, 1907), can be used to obtain an approximation to the finite-sample distribution of $z_{t,h}$ when h is of moderate length. The *first order* Cornish-Fisher expansion for the α-th quantile of the standardized h-period return $z_{t,h}$ is given by:

$$Q_\alpha(z_{t,h}) = \varphi_\alpha + \frac{(\varphi_\alpha^2 - 1)S(z_{t,h})}{6} + o\big(h^{-1/2}\big), \tag{7}$$

[10] See Section C of the Online Appendix for a detailed discussion of why rescaling QK by $\frac{3}{1.23}$ does not make it equivalent to K except when returns are exactly normally distributed.

where φ_α is the α-th quantile of the standard normal distribution. The first term on the right-hand-side (RHS) of Equation (7) approximates $Q_\alpha(z_{t,h})$ by the quantile of the standard normal distribution, φ_α. The second term incorporates the effect of skewness on $z_{t,h}$ in order to improve the quality of the approximation.[11]

The *second order* Cornish-Fisher expansion is given by:

$$Q_\alpha(z_{t,h}) = \varphi_\alpha + \frac{(\varphi_\alpha^2 - 1)S(z_{t,h})}{6} + \frac{(\varphi_\alpha^3 - 3\varphi_\alpha)\big(K(z_{t,h}) - 3\big)}{24} + \frac{(5\varphi_\alpha - 2\varphi_\alpha^3)S^2(z_{t,h})}{36} + o\big(h^{-1}\big). \tag{8}$$

The first two terms of Equation (8) above are the same as in Equation (7). The third term improves the approximation by further taking into account the effect of kurtosis on the standardized return quantile. The last term corrects for the secondary effect of skewness.[12]

Based on Equation (7), Ghysels, Plazzi and Valkanov (2016) propose a measure of robust skewness RS for the h-period return (see Sections A.1 and A.2 of the Online Appendix for the proof):

$$RS(\boldsymbol{Q}) = \frac{6QS(\boldsymbol{Q})}{\varphi_{0.75}}. \tag{9}$$

Based on Equations (8) and (9), we derive an approximate relation between the moment-based kurtosis K and the quantile-based kurtosis QK as follows (see Sections A.3 and A.4 of the Online Appendix for details):

$$K = 3 + \frac{c_4 QK(\boldsymbol{Q}) - c_1}{c_5 QK(\boldsymbol{Q}) - c_2} + \left[\frac{c_6 QK(\boldsymbol{Q}) - c_3}{c_5 QK(\boldsymbol{Q}) - c_2}\right] \frac{36QS^2(\boldsymbol{Q})}{\varphi_{0.75}^2} + o(h^{-1/2}), \tag{10}$$

where $K = K(r_{t,h})$, $Q = Q(r_{t,h})$, and c_1 to c_6 are positive constants, which are functions of the standard normal quantiles.[13] Note that in the third term of the RHS of Equation (10):

[11] The second term is of order $O(h^{-1/2})$ since $S(z_{t,h})$ is $O(h^{-1/2})$.

[12] The third and fourth terms are both of order $O(h^{-1})$ since $K(z_{t,h})$ is $O(h^{-1})$ and $S(z_{t,h})$ is $O(h^{-1/2})$.

[13] The exact formulas for c_1 to c_6 are: (1) $c_1 = 2(\varphi_{0.875} - \varphi_{0.625}) \approx 1.6634$; $c_2 = \frac{-1}{12}\left(\varphi_{0.875}^3 - 3\varphi_{0.875} - \varphi_{0.625}^3 + 3\varphi_{0.625}\right) \approx 0.0838$; $c_3 = \frac{1}{18}\left(5\varphi_{0.875} - 2\varphi_{0.875}^3 - 5\varphi_{0.625} + 2\varphi_{0.625}^3\right) \approx 0.0655$; $c_4 = 2\varphi_{0.75} \approx 1.3490$; $c_5 = \frac{-1}{12}\left(\varphi_{0.75}^3 - 3\varphi_{0.75}\right) \approx 0.1431$; $c_6 = \frac{1}{18}\left(5\varphi_{0.75} - 2\varphi_{0.75}^3\right) \approx 0.1533$.

$\frac{36QS^2(\boldsymbol{Q})}{\varphi_{0.75}^2} = RS^2(\boldsymbol{Q})$.[14]

Equation (10) affords several insights. First, it provides a direct link between the moment-based kurtosis measure (K) and the quantile-based kurtosis measure (QK) that strengthens as the return horizon increases. Second, it makes explicit the influence of skewness (or asymmetry) on kurtosis. This influence is captured succinctly by the third term, which allows us to quantify the contribution of skewness to kurtosis. Finally, it motivates our new robust kurtosis measure as a novel quantile-based measure that approximates the moment-based measure at long horizons.

Before we introduce our new robust kurtosis measure, we need to address two shortcomings of Equation (10), which are especially relevant for thin-tailed distributions. The first shortcoming relates to the fact that the second and third terms of Equation (10) are not defined if the denominator is equal to zero, which will be the case when $QK = c2/c5 \approx 0.5856$. The second shortcoming relates to the fact that the sum of the first three terms of K can be negative for $QK < 1.08$. To address both shortcomings, and to ensure that K remains a continuous function of $Q_\alpha(z_{t,h})$, we smoothly winsorize QK at the rounded-up value of 1.1. Following Engle and Manganelli (2004), the smoothly winsorized QK is defined as:

$$QK^* = 1.1G(QK) + QK\big(1 - G(QK)\big), \tag{11}$$

where $G(x) = 1/\big(1 + e^{b(x-a)}\big)$ is a logistic function, $a = 1.1$ is the winsorization point of QK, and b is a positive finite number that controls the rate of change of the logistic function. The higher the value of b, the less smooth the winsorization. We set $b = 1000$. It is important to note that QK^* differs from QK only for extremely thin-tailed distributions, which are rarely applicable to financial data, such as the inverse triangular distribution with a domain

14 Note that in Equations (7) and (8), the quantiles are expressed in terms of standardized returns, $z_{t,h}$ because the Cornish-Fisher expansions are based on standardized returns. However, in Equations (9) and (10), skewness and kurtosis are expressed in terms of non-standardized returns, $r_{t,h}$. This is because skewness and kurtosis in all their forms (moment-based, quantile-based and robust) are scale-invariant.

on $[-1, 1]$. Otherwise, QK and QK^* are identical.[15]

Based on this analysis, we propose the following new robust measure of kurtosis:

$$RK(\boldsymbol{Q}) = 3 + \frac{c_4 QK^*(\boldsymbol{Q}) - c_1}{c_5 QK^*(\boldsymbol{Q}) - c_2} + \left[\frac{c_6 QK^*(\boldsymbol{Q}) - c_3}{c_5 QK^*(\boldsymbol{Q}) - c_2}\right] \frac{36 QS^2(\boldsymbol{Q})}{\varphi_{0.75}^2}. \tag{12}$$

RK is always positive and is always increasing in QK^* except for extremely skewed distributions. Again, note that RS and RK are scale-invariant.[16]

The main advantage of the new robust kurtosis measure is that it shares desirable properties with both the moment-based and the quantile-based measures. For long horizons, it corresponds to the moment-based measure when it exists. For short horizons, it is a distinct measure that exploits information from the quantiles of the distribution. In doing so, it is robust to outliers and is well defined, even when the tails are too fat for the moments of the distribution to exist. In short, therefore, this is a new and versatile measure that provides a novel view of the kurtosis of financial returns.

2.3 Estimation and Inference

The RK measure defined in Equation (12) is a function of QS and QK^*, which are in turn functions of a finite set of quantiles. We use the sample quantiles $\widehat{Q}_\alpha$ to estimate the population quantiles $Q_\alpha(r_{t,h})$. Then we estimate $RK(\boldsymbol{Q})$ as follows:

$$\widehat{RK} = RK(\widehat{\boldsymbol{Q}}), \tag{13}$$

where $\widehat{\boldsymbol{Q}} = (\hat{Q}_{\alpha_1}, \widehat{Q}_{\alpha_2}, \dots \widehat{Q}_{\alpha_7})'$ is the vector of sample quantiles.

In Section B.1 of the Online Appendix, we show the consistency and asymptotic normality of

15 For further details on the smooth winsorization, see Section A.5 of the Online Appendix.

16 Note that the first order derivative of RK with respect to QK^* is positive if $QS(r_{t,h}) \in (-0.93, 0.93)$. Recall that the range for $QS(r_{t,h})$ is $[-1, 1]$.

$\widehat{RK}$ as an estimator of RK, as $T \to \infty$ under weak conditions on the existence of moments and allowing for weakly dependent data. This accommodates many models commonly used in finance, including GARCH processes, continuous-time diffusion and stochastic volatility models. It is important to note that the estimated robust kurtosis $\widehat{RK}$ converges to the robust kurtosis estimator RK, not to the moment-based estimator K. As a result, the asymptotic results do *not* rely on the Cornish-Fisher expansion or the assumption that h diverges. Instead, they hold for any fixed h, making them suitable for empirical applications in which h is typically small relative to T.[17]

3 Monte Carlo Simulation

In this section, we compare the performance of the robust kurtosis estimator ($\widehat{RK}$) to the traditional moment-based estimator ($\widehat{K}$) by designing Monte Carlo simulations under different distributional assumptions and alternative return horizons. We assess performance with the Root Mean Squared Error (RMSE) and the Root Mean Squared Proportional Error (RMSPE). The RMSE is equal to the square root of the average squared estimation error. The RMSPE is calculated in the same way except that the estimation error is scaled by the population value. By design, the RMSPE accounts for the fact that RK and K may differ in their population values. The formulas for computing $RMSPE$ and $RMSE$ are provided in Section E of the Online Appendix.

The simulations have the following specification. In terms of return horizons, we specify $h = \{1, 5, 22, 66, 250\}$, where $h = 1$ corresponds to the one-day horizon and, therefore, $h = 5/22/66/250$ correspond to a weekly/monthly/quarterly/yearly horizon. We denote N as the number of non-overlapping h-period returns and $T = Nh$ as the sample size of one-

[17] As pointed out by a reviewer, robust kurtosis could alternatively be considered as an estimator of Pearson kurtosis if both h and T diverge. Combining our consistency result ($\widehat{RK} \to RK$ as $T \to \infty$ for fixed h) with the Cornish Fisher expansion ($RK \to K$ as $h \to \infty$) implies $\lim_{h\to\infty} \lim_{T\to\infty} \widehat{RK} - K \to 0$. A full analysis could be a direction for future theoretical research. However, we do not adopt this interpretation here since in practice h is typically small relative to T.

period returns. We set $N = 300$, $N = 1000$ and $N = 6000$ non-overlapping observations. The values for N are chosen to match realistic data sets for the most practically relevant horizons. The lowest value ($N = 300$) roughly corresponds to the monthly sample size used in our empirical analysis. The highest value ($N = 6000$) roughly matches the daily sample size used in our empirical analysis. The intermediate value ($N = 1000$) can be thought of as corresponding to 83 years of monthly data or 4 years of daily data. Each experiment is repeated for $10,000$ simulations.[18]

In terms of return distributions, we consider eight data generating processes (DGPs) for the one-period return r_t: the benchmark standard normal distribution $N(0,1)$; the standard normal distribution with ten artificially generated outliers;[19] the symmetric Student-$t(5)$ distribution to account for fat tails; the Skewed-t $skt(5,-0.3)$ distribution to account for asymmetry and fat tails; and the $GARCH(1,1)$ model with $N(0,1)$, $t(10)$, $t(7)$ or $t(5)$ innovations to account for volatility clustering with varying degrees of tail thickness. The $GARCH(1,1)$ parameters are set to be $\alpha = 0.08$ (on the lagged squared residuals) and $\beta = 0.90$ (on the lagged variance).[20]

3.1 Performance of the K Estimator

We begin by assessing the performance of the traditional moment-based kurtosis estimator, $\widehat{K}$ in terms of its RMSPE. The RMSPEs from the Monte Carlo simulations are reported in Table 1. Our main findings can be summarized as follows.

First, as expected, $\widehat{K}$ performs very well for the normal distribution across all horizons. This implies that $\widehat{K}$ is accurately estimated when it matters the least, i.e., when the distri-

18 Since N is fixed when varying h, a given value of N cannot provide a realistic sample size at all horizons. We set the value of N to be realistic for one of the horizons that we view as most practically relevant.

19 The outliers are generated as in Kim and White (2004) and Ghysels, Plazzi and Valkanov (2016). See Section D of the Online Appendix for details.

20 These parameters are motivated empirically. When estimating the $GARCH$-t model across all countries, the median estimates are $\hat{\alpha} = 0.08$ and $\hat{\beta} = 0.90$ with 6.64 degrees of freedom.

bution is normal. Second, $\widehat{K}$ displays a high RMSPE in the presence of outliers or when the distribution is fat-tailed, especially for short horizons. Therefore, $\widehat{K}$ is not accurately estimated when it matters the most, i.e., for fat-tailed distributions. Third, the performance of $\widehat{K}$ worsens substantially when GARCH volatility is combined with a Student-t distribution. Therefore, $\widehat{K}$ performs the worst for the distribution (GARCH-t) that best represents financial returns data.[21]

Fourth, for sufficiently fat-tailed distributions (such as the Student-$t(5)$) and fat-tailed distributions with volatility dependence (such as the GARCH-$t(7)$), the performance of $\widehat{K}$ worsens as the sample size increases from $N = 300$ to $N = 6000$. For fat-tailed distributions without GARCH volatility, this is true for short horizons. For fat-tailed distributions with GARCH volatility, this is true for all but the yearly horizon. Therefore, the performance of $\widehat{K}$ is poor for realistic distributions and it actually worsens with higher sample size.

While the previous finding may run contrary to the usual finding of improved estimation accuracy with sample size, it results from the strict variance condition of the fourth sample moment. In an i.i.d. setting, it is straightforward to show that the variance of $\widehat{K}$ depends on the eighth moment of returns, which is infinite under many realistic return distributions. Even if the eighth moment exists, the convergence of its variance may be slow (see, e.g., Eberl and Klar, 2020).[22] Intuitively, a larger sample size involves a higher probability of an extreme outlier. When the tails are sufficiently thick, this effect swamps the additional averaging afforded by the larger sample size. When both fat tails and GARCH effects are present, large outliers tend to arrive in clusters, making their impact difficult to average out.

In conclusion, the Monte Carlo simulation results show that the traditional moment-based estimator $\widehat{K}$ performs well for only two cases: (1) normally distributed returns, and (2) fat-tailed distributions when the horizon is long and in the absence of volatility clustering. In

[21] Note that we do not report the RMPSE for $\widehat{K}$ for the GARCH-$t(5)$ because for this model the fourth moment does not exist (see, Ling and McAleer, 2002).

[22] Eberl and Klar (2020) show that, in an i.i.d. setting, the existence of variance of sample skewness depends on the sixth moment and the variance converges slowly even if the sixth moment exists. Based on a nearly identical argument, the variance of sample kurtosis depends on the eighth moment.

other words, $\widehat{K}$ performs well only when returns are normal or approximately normal, i.e., when kurtosis is of little or no importance in modelling returns. For the type of distributions typically observed in financial returns characterized by GARCH volatility and fat tails, the traditional moment-based estimator is not a reliable estimator of kurtosis.

3.2 Performance of the RK Estimator

Next, we turn to the performance of the robust kurtosis estimator, $\widehat{RK}$. The results reported in Table 1 are striking: the RMSPE of $\widehat{RK}$ is low and remarkably stable across all horizons and return distributions. Notably, $\widehat{RK}$ often performs better for returns with fat tails, asymmetry and volatility clustering than for normally distributed returns. For example, for $N = 300$ across all distributions and horizons, the highest RMSPE is equal to 0.60 (N(0,1) with outliers for $h = 1$) and the lowest is equal to 0.32 (GARCH-$t(5)$ for $h = 1$). In other words, the RMSPE of $\widehat{RK}$ for perhaps the most realistic distribution (GARCH-$t(5)$) is lower than for all other distributions, including the normal.

Notably, the RMSPE of $\widehat{RK}$ becomes consistently lower as we move to a higher sample size, i.e., from $N = 300$ to $N = 6000$. Hence, not only is $\widehat{RK}$ a reliable estimator with a low RMSPE across all cases, but its performance also improves with sample size.

Additionally, $\widehat{RK}$ performs better than $\widehat{K}$ at short horizons (e.g., $h = 1$) for almost all distributions other than the normal. Importantly, for the GARCH-t distributions, which are the most relevant for financial returns data, $\widehat{RK}$ perfoms better than $\widehat{K}$ both for short and for long horizons. Consequently, $\widehat{RK}$ is a far more reliable estimator than $\widehat{K}$ for financial returns data characterized by non-normal distributions.

To conclude, based on the RMSPE, our Monte Carlo evidence indicates that the performance of the robust kurtosis estimator is consistently excellent in all cases regardless of horizon and distribution. More importantly, the robust kurtosis estimator performs substantially better than the traditional moment-based estimator when the distribution deviates meaningfully from normality. This makes $\widehat{RK}$ an ideal kurtosis estimator for short- and long-horizon financial returns data characterized by fat tails and volatility clustering.

3.3 Comparison of K and RK based on RMSE

We have so far compared $\widehat{K}$ and $\widehat{RK}$ using the RMSPE, which rescales the estimation error by the population value. In Table 2, we present the more commonly reported RMSE without rescaling. Overall, the RMSE results are similar to the RMSPE results but now the dominance of $\widehat{RK}$ over $\widehat{K}$ is more pronounced. Specifically, for short horizons ($h = 1$), $\widehat{RK}$ outperforms $\widehat{K}$ across almost all non-normal distributions. For realistic return distributions, such as GARCH-t, $\widehat{RK}$ outperforms $\widehat{K}$ across all horizons and sample sizes. For all other cases, $\widehat{RK}$ outperforms $\widehat{K}$ for low horizons and high sample sizes. Overall, these results confirm that $\widehat{RK}$ is a superior estimator for financial asset returns.

3.4 The Population Values of K and RK

It is important to highlight that by design the population values of RK and K may differ. The population values reported in Table 3 indicate that K and RK have similar values for two cases: (1) the benchmark normal distribution across all horizons, and (2) for approximately normal distributions. For all other cases, the two estimators deviate in population value. The deviation becomes more pronounced at short-horizons and in the presence of fat tails combined with volatility clustering. By design, therefore, K and RK provide distinct measures of kurtosis for the most relevant distributions associated with financial returns.[23]

It is important not to confuse the population values with performance. When returns are (approximately) normal, K and RK have similar values and both are estimated reliably. In all other cases, when kurtosis matters the most, K has a high value that cannot be estimated

[23] The population values are calculated analytically when feasible. This is the case only for the normal distribution across all horizons and for the $t(5)$ distribution when $h = 1$. For all other cases, we draw $10,000$ artificial samples for r_t with length $T = 20,000h$ and then compute the non-overlapping h-period return for each sample. The average across these simulations is used as the population value.

reliably, whereas RK has a lower value than K that can be estimated reliably. These findings are especially true for the one-month horizon used in our empirical analysis.

4 Empirical Application: International Stock Markets

4.1 Data Description

In this section, we implement robust skewness and robust kurtosis on the cross-section of international stock index returns. Our empirical analysis focuses on the monthly returns of 39 US dollar-denominated country stock indices for the sample period of January 1, 1996 to June 30, 2019. The data sample comprises 23 developed markets (DM) from FTSE and 16 emerging markets (EM) from the S&P/IFCI indices. We refer to the union of DM and EM as ALL. The returns data are obtained from the *Refinitiv Eikon* database. These data are similar to the data used by Ghysels, Plazzi and Valkanov (2016) extended to June 2019.

Our analysis is based on daily simple returns, R_t, to calculate the monthly returns $R_{t,22} = \prod_{j=0}^{21}(1 + R_{t-j}) - 1$ using the most recent 22 daily observations, hence $h = 22$.[24] We simplify our notation by adopting a monthly frequency and referring to monthly returns $R_{t,22}$ as R_τ, where the monthly index τ is related to the daily index t as follows: for $t = 22, 44, 66, \ldots$ days, $\tau = t/22 = 1, 2, 3, \ldots$ months. Going forward, using our monthly notation $\tau - 1$ refers to the previous month and $\tau - 1/22$ refers to the previous day. To be consistent with the literature, our main analysis uses non-overlapping monthly returns.

[24] Although the formal derivation of RS and RK is based on log returns, our empirical application follows Ghysels, Plazzi and Valkanov (2016) by employing simple returns. As a robustness check, in Table A7 of the Online Appendix, we confirm that our main findings hold using log returns.

4.2 Conditional Robust Skewness and Kurtosis

A large body of research has documented that the skewness and kurtosis of stock returns are time-varying (see, e.g., Bekaert and Harvey, 1997; León, Rubio and Serna, 2005; Guidolin and Timmermann, 2008; Jondeau and Rockinger, 2003, 2012). To capture the dynamics of robust skewness and robust kurtosis, we employ the conditional MIDAS quantile model proposed by Ghysels (2014) and Ghysels, Plazzi and Valkanov (2016). A crucial aspect of the MIDAS quantile model is that it conditions on higher frequency (daily) returns to estimate the quantiles of lower frequency (monthly) returns. This is in contrast to other conditional quantile models, such as the CAViaR model of Engle and Manganelli (2004) and White, Kim and Manganelli (2010), which use same-frequency returns to estimate conditional quantiles.[25]

Accordingly, our choice of the MIDAS model as a centerpiece of the analysis is due to: (1) its ability to exploit information from daily returns in estimating monthly conditional quantiles, (2) its prominence in the literature (see, e.g., Ghysels, Santa-Clara and Valkanov, 2005), and (3) its well-established good performance (see, e.g., Ghysels, Santa-Clara and Valkanov, 2006; Andreou, Ghysels and Kourtellos, 2010). In this context, our empirical results not only depend on the definition of robust skewness and robust kurtosis but they also depend on the ability of the MIDAS model to deliver reliable estimates for the conditional quantiles.

Following the mixed frequency approach, we model the conditional quantile of the monthly returns R_τ using lagged daily regressors. We condition on $\mathcal{F}_{\tau-1} = \sigma(x_{\tau-1}, x_{\tau-1-1/22}, x_{\tau-1-2/22}, \ldots)$, the information set including the daily regressors $x_{\tau-1-d/22}$ at a lag of one month plus $d = 0, 1, 2, \ldots$ days. The conditional quantile, denoted by $Q_{\alpha,\tau-1}(R_\tau)$, is defined as the quantile of R_τ conditional on $\mathcal{F}_{\tau-1}$. The relevant conditional quantiles are collected in the vector $\boldsymbol{Q}_{\tau-1} = (Q_{\alpha_1,\tau-1}(R_\tau), Q_{\alpha_2,\tau-1}(R_\tau), \ldots, Q_{\alpha_7,\tau-1}(R_\tau))'$. Finally, we define the population measures of conditional robust skewness and kurtosis as $RS_{\tau-1} = RS(\boldsymbol{Q}_{\tau-1})$ and $RK_{\tau-1} = RK(\boldsymbol{Q}_{\tau-1})$, respectively.

[25] Aggregating the conditioning variables in order to match the frequency of the dependent variable tends to produce a bias in the results (see, e.g., Breitung and Swanson, 2002.)

To avoid parameter proliferation, the MIDAS quantile model specifies each conditional quantile nonlinearly using the D daily lags in $X_{\tau-1} = (1, x_{\tau-1}, x_{\tau-1-1/22}, x_{\tau-1-2/22}, \ldots x_{\tau-1-(D-1)/22})'$, employing just the three parameters $\theta_\alpha = (a_{0,\alpha}, a_{1,\alpha}, \kappa_\alpha)'$ as follows:

$$Q_{\alpha,\tau-1}(R_\tau) = X'_{\tau-1}\beta(\theta_\alpha) = a_{0,\alpha} + a_{1,\alpha}\sum_{d=0}^{D-1} B(d+1;\kappa_\alpha)x_{\tau-1-d/22}. \tag{14}$$

We define:

$$\beta(\theta_\alpha) = [a_{0,\alpha}, a_{1,\alpha}B(1,\kappa_\alpha), a_{1,\alpha}B(2,\kappa_\alpha), \ldots, a_{1,\alpha}B(22,\kappa_\alpha)]', \tag{15}$$

$$B(d;\kappa) = \frac{f\left(\frac{d}{D};\kappa\right)}{\sum_{d=1}^{D} f\left(\frac{d}{D};\kappa\right)}, \tag{16}$$

$$f\left(\frac{d}{D};\kappa\right) = \frac{\left(1-\frac{d}{D}\right)^{\kappa-1}\Gamma(1+\kappa)}{\Gamma(\kappa)}, \tag{17}$$

where $\Gamma(\cdot)$ is the gamma function. We estimate the parameters θ_α with nonlinear quantile regression using the same objective function as Ghysels, Plazzi and Valkanov (2016).[26] The estimation algorithm sets $\kappa > 0$. For the vast majority of cases, we find that $\widehat{\kappa} > 1$, which implies slowly decaying weights. We also set the maximum lag to $D = 250$ and $x_{\tau-1-d/22} = |R_{\tau-1-d/22}|$ as in Engle and Manganelli (2004) and Ghysels, Plazzi and Valkanov (2016). Hence the conditioning information consists of lagged daily absolute returns over the past one year. In our main empirical analysis, the parameters of the MIDAS quantile model are estimated using the full sample. Therefore, our analysis is performed ex-ante (i.e., conditioning on $\tau - 1$ information) but in-sample (i.e., using the full sample MIDAS estimates). In Section K of the Online Appendix, we also provide out-of-sample results.[27]

Armed with the MIDAS parameter estimates, we generate the conditional quantiles (e.g., $\widehat{Q}_{0.75,\tau-1}(R_\tau)$) using the fitted values of the MIDAS model and collect them in the vector $\widehat{\boldsymbol{Q}}_{\tau-1} = (\widehat{Q}_{\alpha_1}(R_\tau), \widehat{Q}_{\alpha_2}(R_\tau), \ldots \widehat{Q}_{\alpha_7}(R_\tau))'$. Then, we plug $\widehat{\boldsymbol{Q}}_{\tau-1}$ into Equations (9) and (12)

---

[26] See Equation (IA2) on page 8 of the Ghysels, Plazzi and Valkanov (2016) Online Appendix. See also the MIDAS Matlab toolbox written by Eric Ghysels and collaborators for the code used in our estimation.

[27] Following Ghysels, Plazzi and Valkanov (2016), the estimation of the MIDAS parameters involves overlapping returns but the conditional $\widehat{RS}$ and $\widehat{RK}$ are constructed with non-overlapping returns.

to define fitted conditional robust skewness and kurtosis by:

$$\widehat{RS}_{\tau-1} = RS(\widehat{\boldsymbol{Q}}_{\tau-1}) \quad \text{and} \tag{18}$$

$$\widehat{RK}_{\tau-1} = RK(\widehat{\boldsymbol{Q}}_{\tau-1}). \tag{19}$$

In short, Equations (18) and (19) provide the conditional robust skewness and kurtosis measures that we use throughout our empirical analysis.

Next, we establish that $\widehat{RK}_{\tau-1}$ is a consistent estimator of its population counterpart, $RK_{\tau-1}$, and then derive its asymptotic normality. The asymptotic properties of $\widehat{RK}_{\tau-1}$ depend on those of $\widehat{\boldsymbol{Q}}_{\tau-1}$, which in turn depend on the quantile MIDAS estimators $\widehat{\theta}_{\alpha}$. Since Ghysels (2014) and Ghysels, Plazzi and Valkanov (2016) do not provide detailed asymptotic results for their quantile MIDAS estimator, we show the consistency and asymptotic normality of $\hat{\theta}_{\alpha}$ in Propositions 10 and 11 of Section B.2 of the Online Appendix. Then, the propositions below show that $\widehat{RK}_{\tau-1}$ is a consistent and asymptotically normal estimator.

Proposition 1. *(Consistency) Assume that a_1 is bounded away from zero and that Assumptions 5,6 and 8 of Online Appendix B.2 are satisfied for $\boldsymbol{\alpha} = \{0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875\}'$. Then as $N \to \infty$, we have:*

(i) Conditional on $\mathcal{F}_{t-1}$, $\widehat{RK}_{\tau-1} \to_p RK_{\tau-1}$ for any given $X_{\tau-1}$ and

(ii) $P\left(\widehat{RK}_{\tau-1} - RK_{\tau-1} \to_p 0\right) = 1$, where P denotes probability.

Proof. See Online Appendix B.3. □

Proposition 2. *(Asymptotic Normality) Assume that a_1 is bounded away from zero and parts (a),(b) and (d) of Assumption 9 of Online Appendix B.2 are satisfied for $\boldsymbol{\alpha} = \{0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875\}'$. Then, conditional on $\mathcal{F}_{\tau-1}$,*

$$\sqrt{N}\left(\widehat{RK}_{\tau-1} - RK_{\tau-1}\right) \xrightarrow{d} N\left(0, \boldsymbol{R}'_{Q,\tau-1}\boldsymbol{\Omega}_{\tau-1}\boldsymbol{R}_{\boldsymbol{Q},\tau-1}\right)$$

for any given $X_{\tau-1}$ as $N \to \infty$, where $\boldsymbol{R}_{\boldsymbol{Q},\tau-1} = \frac{\partial RK_{\tau-1}}{\partial \boldsymbol{Q}_{\tau-1}}$ is the 7×1 vector of partial *derivatives and is defined analogously to $R_{\boldsymbol{Q}}$ in Proposition 9 of Online Appendix B.1,* $\boldsymbol{\Omega}_{\tau-1} = \left(I_7 \otimes X'_{\tau-1}\right) \boldsymbol{G_\theta} \boldsymbol{\Sigma} \boldsymbol{G'_\theta} \left(I_7 \otimes X_{\tau-1}\right)$, *$I_7$ is a 7×7 identity matrix, $\otimes$ denotes the Kronecker product, and $\boldsymbol{G_\theta}$ is a $7q \times 7q$ block diagonal partitioned matrix with the ith diagonal block given by the $q \times q$ matrix of partial derivatives, $\frac{\partial \beta(\theta_{\alpha_i})}{\partial \theta'_{\alpha_i}}$, for $i = 1, 2, \ldots, 7$.*

Proof. See Online Appendix B.3. □

In the MIDAS quantile model, κ is unidentified if a_1 is exactly zero, which explains the restriction on a_1. The remaining assumptions, intermediate results and proofs, together with additional discussion, can be found in Sections B.2 and B.3 of the Online Appendix.

5 The Cross-Section of International Stock Returns

In this section, we evaluate the predictive ability of conditional robust skewness and conditional robust kurtosis for the cross-section of international stock returns. Our empirical analysis is based on portfolio sorts using either robust skewness or robust kurtosis to rank international stock returns and formal Fama and MacBeth (1973) cross-sectional regressions.

5.1 Portfolio Sorts Based on Positive Excess Kurtosis

We begin with generating portfolios by sorting international stock index returns on their individual RS or RK.[28] At the end of each month, the stock returns are ranked according to their RS or RK for that month and are then allocated into three portfolios: Low, Medium and High. Consistent with standard practice (see, e.g., Fama and French, 1993), the allocation breakpoints are the 30th and 70th percentile of the monthly RS/RK. Allocating countries to three portfolios ensures that there is a sufficient number of countries in each portfolio even for the smaller EM cross-section. For each portfolio, we compute the

[28] To simplify notation, we henceforth denote $\widehat{RS}_{\tau-1}$ and $\widehat{RK}_{\tau-1}$ simply as RS and RK.

one-month ahead equally-weighted return and the Newey and West (1987) t-statistic.[29] This procedure is repeated every month. In short, the portfolio sorts allow us to assess whether RS/RK have predictive information for portfolio returns over the next month.

In our empirical analysis, it is sensible to discern the case when returns display fatter tails than the normal distribution and, as a result, kurtosis exceeds the value of 3 (i.e., $RK > 3$). Intuitively, kurtosis is a measure of the importance of outliers and, if outliers matter, kurtosis must be higher than 3. Therefore, it is natural to expect that the fourth moment is informative primarily for non-normal distributions indicated by positive *excess* kurtosis.

In this context, we perform the portfolio sorts based on RS and RK using only countries for which on a given month $RK > 3$. We denote $RS^{+}_{i,\tau-1} = RS_{i,\tau-1}$ and $RK^{+}_{i,\tau-1} = RK_{i,\tau-1}$ for $RK_{i,\tau-1} > 3$, otherwise we drop country i in month τ. This restricts our analysis to the countries for which excess robust kurtosis is positive. We consider this restriction to be economically motivated since investors are unlikely to care about kurtosis if it indicates that the distribution tails are thinner than the normal. Additionally, this restriction accounts for the cases when robust kurtosis has a low value due to estimation error. Imposing an economic constraint on kurtosis is conceptually similar to the economic constraints imposed on equity premium prediction by Campbell and Thompson (2008).[30]

The results on the portfolio sorts based on RS^{+} (Panel A) and RK^{+} (Panel B) are reported in Table 4. Specifically, across all cases (DM, EM and ALL), mean returns are monotonically increasing for RS^{+} and monotonically decreasing for RK^{+}. This is an important finding: higher RS^{+} consistently leads to *higher* returns, whereas higher RK^{+} consistently leads to *lower* returns. RS^{+} has a stronger effect on EM: the High-minus-Low portfolio delivers a statistically significant return of 0.76% per month (or 9.12% per year). RK^{+} is highly significant for ALL countries: the High-minus-Low portfolio delivers a statistically significant

[29] In computing the Newey and West (1987) t-statistic, we use the Bartlett kernel with the data-driven bandwidth parameter selected by the AR(1) model (Andrews, 1991).

[30] On average, 32 of the 39 countries display positive excess kurtosis, see Table A6 and Figure 5 of the Online Appendix. For portfolio results without this restriction, see Table A5 of the Online Appendix.

return of −0.44% per month (or −5.28% per year).

In conclusion, our evidence on portfolio sorts indicates the following: (1) robust skewness has strong positive predictive power for future EM stock returns; (2) robust kurtosis has strong negative predictive power for future stock returns; and (3) these findings become more pronounced and more significant when robust kurtosis deviates from normality ($RK > 3$). The latter two results comprise novel contributions to the literature.

5.2 Fama-MacBeth Regressions

In this section, we formalize our analysis of the predictive relation between stock returns and the conditional RS and RK by estimating Fama and MacBeth (1973) cross-sectional regressions. At each month τ, we estimate the following cross-sectional regression across $i = 1, ..., 39$ countries:

$$R_{i,\tau} = \beta_{0,\tau} + \beta_{1,\tau} VOL_{i,\tau-1} + \beta_{2,\tau} RS_{i,\tau-1} + \beta_{3,\tau} RK_{i,\tau-1} + \epsilon_{i,\tau}, \tag{20}$$

where $R_{i,\tau}$ is the monthly return of the ith country stock index for month τ, and $RS_{i,\tau-1}$ and $RK_{i,\tau-1}$ are the conditional robust skewness and kurtosis of the ith country for month $\tau - 1$, respectively. $VOL_{i,\tau-1}$ controls for the realized volatility of each stock, which is computed as the square root of the sum of the squared daily returns over the last month. Following Fama and MacBeth (1973), we collect the estimated coefficients every month and report the mean coefficients as well as Newey and West (1987) t-statistics. In some specifications, we replace RS and RK by RS^+ and RK^+, which restrict the analysis to countries which on a given month exhibit positive excess robust kurtosis.

In addition to the Fama-MacBeth regression in Equation (20), we also estimate an augmented specification by adding two important economic fundamentals: the lagged dividend yield ($DY_{\tau-1}$) and the lagged short interest rate ($IR_{\tau-1}$). The former is defined as the log of the 12-month rolling dividend ending at time $\tau - 1$ divided by the price at $\tau - 2$ (see., e.g., Welch and Goyal, 2008). The latter is defined as the lagged 1-month Eurodeposit rate for each country where available, otherwise we choose a similar 1-month interest rate based on data availability. The data on DY and IR are obtained from the *Refinitiv Eikon* database.[31]

The results from the Fama-MacBeth regressions are presented in Table 5. The main findings are as follows. First, for DM countries, only RK^+ is statistically significant with a negative coefficient (t-stat=-1.98) for the full specification. Therefore, for developed markets, the higher the deviation from normality in terms of robust kurtosis the lower the future stock returns. This result confirms the negative predictive ability of RK^+ that we first established with the portfolio sorts. The stronger results for RK^+ relative to RK indicate that it is positive excess robust kurtosis that matters most to investors. This makes intuitive sense, since it is under positive excess kurtosis that outliers become relevant.

Second, for EM countries, only RS (or RS^+) is statistically significant with a positive coefficient. For example, in regression (5) that conditions on VOL, RS^+ and RK^+, the coefficient on RS^+ is positive and significant (t-stat=2.40). This result confirms the positive predictive ability of RS for EM that we first established with the portfolio sorts. However, the significance of RS^+ disappears when we also condition on the two economic fundamentals, $ln(DY)$ and IR. Therefore, although RS^+ is significant in the absence of economic fundamentals, it is not significant in the presence of economic fundamentals.

Third, for ALL countries, both RS^+ and RK^+ are individually significant with RS^+ being positive and RK^+ being negative. However, when they are together in the same regression, only RK^+ retains its statistical significance (t-stat=-2.00). Indeed, RK^+ remains significant when we also add the economic fundamentals (t-stat=-2.32). Therefore, across all countries, the negative effect of RK^+ is stronger than the positive effect of RS^+. This is our strongest result since for ALL countries RK^+ is the only predictor that retains its statistical significance across all specifications.[32]

[31] We select DY and IR as the main economic fundamentals used in the Fama-MacBeth regression because of data availability, their importance in equity premium prediction (see., e.g., Tsiakas, Li and Zhang, 2020), and to avoid high dimensionality in estimation, especially for the EM group of only 16 countries.

[32] We consider our evidence of statistical significance (when present) as particularly strong because it is based on a cross-section of 39 countries over a sample period of 24 years. This is a small sample

Fourth, VOL is consistently insignificant in all Fama-MacBeth regressions. This is strong evidence that the second moment does not provide any predictive ability for international stock returns over and above the third and fourth moments. This finding further motivates our focus on skewness and kurtosis.

Finally, fifth, the two economic fundamentals are statistically insignificant in all cases. Although they do affect the significance of other variables (notably RS^{+}) they do not individually exhibit statistical significance.

In summary, the Fama-MacBeth regressions indicate that robust kurtosis carries a negative premium primarily for developed markets, whereas robust skewness carries a positive premium primarily for emerging markets. Hence each measure is dominant for a different group of countries. Across all regressions, volatility is consistently insignificant. Similarly, the dividend yield and the short interest rate are also insignificant. For the full cross-section of countries, the predictive ability of robust kurtosis is stronger and more significant than that of robust skewness. Robust kurtosis remains significant in the presence of volatility, skewness and economic fundamentals. In short, for the full cross-section of countries, robust kurtosis is dominant over robust skewness and is a powerful variable in predicting the cross-section of international stock returns.[33]

6 Conclusion

In this paper, we introduce robust kurtosis, which is a new quantile-based measure for the kurtosis of stock returns. The new robust kurtosis measure complements the existing measure of robust skewness. We provide both an econometric and an empirical analysis of robust kurtosis. Our econometric analysis derives robust kurtosis using a second-order Cornish-

compared to standard asset pricing studies focusing on individual US equities over long periods.

33 In Table A7 of the Online Appendix, we use log returns to form VOL, RS and RK, and find that the results remain qualitatively the same. In Section K and Table A9 of the Online Appendix, we also perform an out-of-sample analysis, which indicates that the effect of RK is a robust finding.

Fisher expansion and assesses the finite-sample properties of robust kurtosis using Monte Carlo simulations under different distributional specifications. In the Online Appendix, we also establish the consistency and asymptotic normality of robust kurtosis. For the normal distribution, we find that robust kurtosis is equivalent to the moment-based kurtosis. For moderately fat-tailed distributions, robust kurtosis essentially provides the same information as the moment-based measure. However, for severely fat-tailed distributions, when kurtosis matters the most, robust kurtosis provides a distinct and more reliable alternative to the moment-based kurtosis for measuring the thickness of the distribution tails (and center) relative to the shoulder.

Our empirical analysis provides a comprehensive evaluation of the predictive information of robust skewness and robust kurtosis for the cross-section of international stock index returns. Our methodology is based on portfolio sorts and Fama-MacBeth cross-sectional regressions. Our findings confirm the previously established result regarding robust skewness: higher robust skewness in emerging markets is related to higher future stock returns. More importantly, we establish a new empirical result regarding robust kurtosis: higher robust kurtosis is related to lower future stock returns. This result holds for all countries but is particularly strong for developed markets. Furthermore, robust kurtosis tends to have higher predictive power when restricting the analysis to countries with positive excess robust kurtosis. In other words, when the return distribution deviates meaningfully from normality, robust kurtosis tends to be more informative.

In short, our empirical findings indicate that robust skewness carries a positive premium primarily for emerging markets. In contrast, robust kurtosis carries a negative premium primarily for developed markets. Taken together, for the full cross-section of countries, robust kurtosis dominates the effect of robust skewness. These results are robust to controlling for volatility and standard economic fundamentals. In conclusion, robust skewness together with robust kurtosis provide powerful predictors for future international stock returns across developed and emerging markets and highlight the role of higher moments in justifying international diversification.

References

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez. 2015. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118:135-167.

Andreou, E., E. Ghysels, and A. Kourtellos. 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics* 158:246-261.

Andrews, D. W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59:817-858.

Arditti, F. D. 1967. Risk and the required return on equity. *Journal of Finance* 22:19-36.

Arditti, F. D. 1971. Another look at mutual fund performance. *Journal of Financial and Quantitative Analysis* 6:909-912.

Balanda, K. P. and H. L. Macgillivray. 1988. Kurtosis: A critical review. *American Statistician* 42:111-119.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–446.

Bekaert, G., E. Engstrom, and A. Ermolov. 2015. Bad environments, good environments: A non-Gaussian asymmetric volatility model. *Journal of Econometrics* 186:258-275.

Bekaert, G. and C. R. Harvey. 1997. Emerging equity market volatility. *Journal of Financial Economics* 43:29-77.

Bowley, A. L. 1920. *Elements of Statistics*. New York: Scribner.

Boyer, B., T. Mitton, and K. Vorkink. 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23:169–202.

Breitung, J. and N. R. Swanson. 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis* 23:651-665.

Brooks, C., S. P. Burke, S. Heravi and G. Persand. 2005. Autoregressive conditional kurtosis. *Journal of Financial Econometrics* 3:399-421.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:1509-1531.

Cornish, E. and R. Fisher. 1938. Moments and cumulants in the specification of distributions. *Review of International Statistical Institute* 5:307-320.

Dittmar, R. F. 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance* 57:369-403.

Eberl, A. and B. Klar. 2020. Asymptotic distributions and performance of empirical skewness measures. *Computational Statistics & Data Analysis* 146:106939.

Edgeworth, F. Y. 1905. The law of error. In *Transactions of the Cambridge Philosophical Society Vol.20.* Cambridge: Cambridge University Press.

Edgeworth, F. Y. 1907. On the representation of statistical frequency by a series. *Journal of the Royal Statistical Society* 70:102-106.

Engle, R. F. and S. Manganelli. 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22:367-381.

Fama, E. and K. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3-56.

Fama, E. and J. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–636.

Ghysels, E. 2014. Conditional skewness with quantile regression models: SoFiE presidential address and a tribute to Hal White. *Journal of Financial Econometrics* 12:620-644.

Ghysels, E., A. Plazzi, and R. Valkanov. 2016. Why invest in emerging markets? The role of conditional return asymmetry. *Journal of Finance* 71:2145-2192.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2005. There is a risk-return tradeoff after all. *Journal of Financial Economics* 76:509-548.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2006. Predicting volatility: How to get the most out of returns data sampled at different frequencies. *Journal of Econometrics* 131:59-95.

Guidolin, M. and A. Timmermann. 2008. International asset allocation under regime switching, skew and kurtosis preferences. *Review of Financial Studies* 21:889-935.

Hansen, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review* 35:705-730.

Harvey, C. R. and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 52:1263-1295.

Jondeau, E. and M. Rockinger. 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27:1699-1737.

Jondeau, E. and M. Rockinger. 2012. On the importance of time variability in higher moments for asset allocation. *Journal of Financial Econometrics* 10:84-123.

Jondeau, E., Q. Zhang, and X. Zhu. 2019. Average skewness matters. *Journal of Financial Economics* 134:29-47.

Kim, T. H. and H. White. 2004. On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1:56-73.

Kimball, M. S. 1993. Standard risk aversion. *Econometrica* 61:589-611.

Kraus, A. and R. H. Litzenberger. 1976. Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085-1100.

Kumar, A. 2009. Who gambles in the stock market? *Journal of Finance* 64:1889-1933.

Langlois, H. 2020. Measuring skewness premia. *Journal of Financial Economics* 135:399-424.

León, Á, G. Rubio, and G. Serna. 2005. Autoregressive conditional volatility, skewness and kurtosis. *Quarterly Review of Economics and Finance* 45:599-618.

Ling, S. and M. McAleer. 2002. Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* 106:109-117.

Mitton, T. and K. Vorkink. 2007. Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies* 20:1255-1288.

Moors, J. 1988. A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)* 37:25-32.

Neuberger, A. 2012. Realized skewness. *Review of Financial Studies* 25:3424-3255.

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

Pearson, K. 1905. Skew variation, a rejoinder. *Biometrika* 4:169-212.

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.

Tsiakas, I., J. Li, and H. Zhang. 2020. Equity premium prediction and the state of the economy. *Journal of Empirical Finance* 58:75-95.

Welch, I. and A. Goyal. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21:1455-1508.

White, H., T. H. Kim, and S. Manganelli. 2010. Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR. In *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*. ed. T. Bollerslev, J. Russell, and M. Watson, Oxford: Oxford University Press.

Table 1: Performance of Kurtosis Estimators based on RMSPE

This table reports the root mean squared proportional error (RMSPE) of the estimators for robust kurtosis (RK) and moment-based kurtosis (K) using Monte Carlo simulation. For each distribution, we draw 10,000 artificial samples of the one-period return r_t with sample size Nh, where $N = \{300, 1000, 6000\}$ and $h = \{1, 5, 22, 66, 250\}$ is the horizon of non-overlapping returns. The RMSPE is defined as the average across simulations of the square root of the squared ratio of the difference between the estimator and its population value divided by its population value, where the population value is taken from Table 3.

		h =1	h =5	h =22	h =66	h =250	h =1	h =5	h =22	h =66	h =250
		Panel A: $N(0,1)$					Panel B: $N(0,1)$ with outliers				
N = 300	$\widehat{RK}$	0.56	0.56	0.57	0.58	0.57	0.60	0.58	0.59	0.58	0.57
	$\widehat{K}$	0.09	0.09	0.09	0.09	0.09	7.20	1.32	0.21	0.19	0.10
N = 1000	$\widehat{RK}$	0.30	0.29	0.29	0.29	0.29	0.30	0.29	0.29	0.29	0.29
	$\widehat{K}$	0.05	0.05	0.05	0.05	0.05	11.79	1.39	0.12	0.06	0.05
N = 6000	$\widehat{RK}$	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
	$\widehat{K}$	0.02	0.02	0.02	0.02	0.02	9.06	0.52	0.04	0.02	0.02
		Panel C: $t(5)$					Panel D: $skt(5, -0.3)$				
N = 300	$\widehat{RK}$	0.39	0.49	0.54	0.56	0.55	0.47	0.50	0.54	0.54	0.50
	$\widehat{K}$	0.73	0.49	0.20	0.15	0.09	0.74	0.51	0.24	0.16	0.10
N = 1000	$\widehat{RK}$	0.18	0.24	0.28	0.28	0.29	0.24	0.25	0.28	0.27	0.26
	$\widehat{K}$	1.02	0.69	0.22	0.08	0.05	0.90	0.68	0.35	0.13	0.07
N = 6000	$\widehat{RK}$	0.07	0.09	0.10	0.11	0.11	0.10	0.10	0.11	0.11	0.12
	$\widehat{K}$	1.24	3.73	0.19	0.05	0.02	1.31	0.87	0.54	0.11	0.03
		Panel E: $GARCH(0.9, 0.08)$ with $N(0,1)$ innovation					Panel F: $GARCH(0.9, 0.08)$ with $t(10)$ innovation				
N = 300	$\widehat{RK}$	0.47	0.42	0.44	0.46	0.51	0.38	0.40	0.41	0.43	0.49
	$\widehat{K}$	0.27	0.33	0.41	0.40	0.27	0.44	0.44	0.55	0.57	0.46
N = 1000	$\widehat{RK}$	0.24	0.21	0.21	0.22	0.25	0.19	0.19	0.19	0.21	0.24
	$\widehat{K}$	0.24	0.31	0.45	0.33	0.20	0.39	0.50	0.63	0.51	0.67
N = 6000	$\widehat{RK}$	0.09	0.08	0.08	0.09	0.10	0.07	0.07	0.07	0.08	0.09
	$\widehat{K}$	0.28	0.29	0.20	0.15	0.10	0.49	0.69	1.12	0.91	0.37
		Panel G: $GARCH(0.9, 0.08)$ with $t(7)$ innovation					Panel H: $GARCH(0.9, 0.08)$ with $t(5)$ innovation				
N = 300	$\widehat{RK}$	0.36	0.38	0.40	0.42	0.48	0.32	0.36	0.38	0.41	0.46
	$\widehat{K}$	0.55	0.53	0.61	0.71	0.71	N/A	N/A	N/A	N/A	N/A
N = 1000	$\widehat{RK}$	0.17	0.18	0.19	0.20	0.23	0.16	0.17	0.18	0.19	0.23
	$\widehat{K}$	0.49	0.66	0.90	1.05	1.35	N/A	N/A	N/A	N/A	N/A
N = 6000	$\widehat{RK}$	0.07	0.07	0.07	0.08	0.09	0.06	0.06	0.07	0.07	0.09
	$\widehat{K}$	0.73	1.31	1.34	1.21	1.08	N/A	N/A	N/A	N/A	N/A

Table 2: Performance of Kurtosis Estimators based on RMSE

This table reports the root mean squared error (RMSE) of the estimators for robust kurtosis (RK) and moment-based kurtosis (K) using Monte Carlo simulation. For each distribution, we draw 10,000 artificial samples of the one-period return r_t with sample size Nh, where $N = \{300, 1000, 6000\}$ and $h = \{1, 5, 22, 66, 250\}$ is the horizon of non-overlapping returns. The RMSE is defined as the average across simulations of the square root of the squared difference between the estimator and its population value, where the latter is taken from Table 3.

		h =1	h =5	h =22	h =66	h =250	h =1	h =5	h =22	h =66	h =250
		Panel A: $N(0,1)$					Panel B: $N(0,1)$ with outliers				
N = 300	$\widehat{RK}$	1.69	1.69	1.71	1.73	1.71	1.79	1.75	1.76	1.75	1.71
	$\widehat{K}$	0.28	0.28	0.28	0.28	0.28	21.60	3.95	0.64	0.56	0.29
N = 1000	$\widehat{RK}$	0.89	0.87	0.87	0.88	0.87	0.90	0.88	0.88	0.88	0.87
	$\widehat{K}$	0.15	0.15	0.15	0.16	0.16	35.36	4.17	0.37	0.17	0.16
N = 6000	$\widehat{RK}$	0.34	0.34	0.33	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	$\widehat{K}$	0.06	0.06	0.06	0.06	0.06	27.17	1.57	0.12	0.07	0.06
		Panel C: $t(5)$					Panel D: $skt(5,-0.3)$				
N = 300	$\widehat{RK}$	1.64	1.68	1.69	1.71	1.68	2.35	1.82	1.73	1.71	1.64
	$\widehat{K}$	6.58	2.01	0.66	0.47	0.29	8.30	2.36	0.80	0.51	0.31
N = 1000	$\widehat{RK}$	0.76	0.83	0.87	0.87	0.87	1.21	0.92	0.89	0.87	0.86
	$\widehat{K}$	9.17	2.87	0.72	0.25	0.16	10.13	3.19	1.17	0.39	0.20
N = 6000	$\widehat{RK}$	0.28	0.32	0.33	0.34	0.33	0.49	0.37	0.35	0.34	0.40
	$\widehat{K}$	11.18	15.42	0.63	0.16	0.07	14.75	4.05	1.82	0.35	0.09
		Panel E: $GARCH(0.9, 0.08)$ with $N(0,1)$ innovation					Panel F: $GARCH(0.9, 0.08)$ with $t(10)$ innovation				
N = 300	$\widehat{RK}$	1.68	1.64	1.68	1.69	1.70	1.64	1.67	1.64	1.63	1.69
	$\widehat{K}$	1.18	1.63	1.97	1.79	1.00	3.08	3.08	3.65	3.41	1.99
N = 1000	$\widehat{RK}$	0.86	0.80	0.79	0.81	0.85	0.82	0.78	0.78	0.79	0.82
	$\widehat{K}$	1.03	1.51	2.18	1.47	0.73	2.72	3.54	4.13	3.06	2.88
N = 6000	$\widehat{RK}$	0.33	0.30	0.31	0.31	0.32	0.32	0.29	0.29	0.31	0.32
	$\widehat{K}$	1.21	1.43	0.98	0.68	0.36	3.44	4.85	7.37	5.47	1.60
		Panel G: $GARCH(0.9, 0.08)$ with $t(7)$ innovation					Panel H: $GARCH(0.9, 0.08)$ with $t(5)$ innovation				
N = 300	$\widehat{RK}$	1.63	1.64	1.64	1.65	1.69	1.58	1.62	1.61	1.63	1.66
	$\widehat{K}$	5.57	4.98	5.76	5.51	3.75	N/A	N/A	N/A	N/A	N/A
N = 1000	$\widehat{RK}$	0.79	0.77	0.78	0.79	0.82	0.78	0.76	0.75	0.78	0.82
	$\widehat{K}$	5.04	6.19	8.42	8.16	7.18	N/A	N/A	N/A	N/A	N/A
N = 6000	$\widehat{RK}$	0.31	0.29	0.29	0.30	0.31	0.30	0.29	0.29	0.29	0.31
	$\widehat{K}$	7.41	12.28	12.57	9.36	5.73	N/A	N/A	N/A	N/A	N/A

Table 3: The Population Value of Kurtosis Measures

This table reports the population value for the robust kurtosis (RK) and the moment-based kurtosis (K). The population values are computed for non-overlapping returns across five return horizons $h = \{1, 5, 22, 66, 250\}$ and seven return distributions. The table reports analytical results for the normal distribution across all horizons and for the Student-$t(5)$ distribution when $h = 1$. For all other cases, the population values are simulated: we draw 10,000 artificial samples for the one-period return r_t with length $T = 20,000h$, and use these to compute the non-overlapping h-period return for each sample. The population value is closely approximated by taking the average across the simulations.

	h = 1	h = 5	h = 22	h = 66	h = 250
Panel A: $N(0,1)$					
RK	3	3	3	3	3
K	3	3	3	3	3
Panel B: $t(5)$					
RK	4.20	3.45	3.16	3.07	3.02
K	9	4.14	3.27	3.09	3.02
Panel C: $skt(5, -0.3)$					
RK	5.05	3.64	3.23	3.16	3.27
K	11.26	4.65	3.37	3.11	2.98
Panel D: $GARCH(0.9, 0.08)$ with $N(0,1)$ innovation					
RK	3.61	3.89	3.82	3.65	3.36
K	4.35	4.91	4.82	4.43	3.68
Panel E: $GARCH(0.9, 0.08)$ with $t(10)$ innovation					
RK	4.27	4.15	4.01	3.80	3.47
K	7.07	7.01	6.58	6.02	4.31
Panel F: $GARCH(0.9, 0.08)$ with $t(7)$ innovation					
RK	4.56	4.28	4.11	3.88	3.54
K	10.20	9.38	9.38	7.76	5.30
Panel G: $GARCH(0.9, 0.08)$ with $t(5)$ innovation					
RK	4.94	4.49	4.26	4.01	3.65
K	N/A	N/A	N/A	N/A	N/A

Table 4: Portfolio Sorts Based on RS^+ or RK^+

This table displays the performance of portfolios sorted on either conditional robust skewness (RS^+) or conditional robust kurtosis (RK^+). In this table, the portfolio sorts use only countries for which RK deviates from normality: $RS^+_{i,\tau-1} = RS_{i,\tau-1}$ and $RK^+_{i,\tau-1} = RK_{i,\tau-1}$ for $RK_{i,\tau-1} > 3$, otherwise we drop country i in month τ. The portfolios are rebalanced at the end of every month using the 30th percentile for the Low portfolio and the 70th percentile for the High portfolio. The table reports the mean of the one-month ahead return of equally-weighted portfolios. Returns are reported in monthly percent. We also report Newey-West t-statistics (in parenthesis) and the average RS^+ or RK^+ for each portfolio. The sample period ranges from January 1997 to June 2019. For notational simplicity, we suppress the circumflex on RS^+ and RK^+ and their dependence on $\tau - 1$.

Panel A: Portfolio Sort on RS^+

	Low	Medium	High	High−Low
DM				
Mean Return	0.66	0.82	0.88	0.21
(t-stat)	(1.86)	(2.37)	(2.36)	(1.33)
Average RS^+	-1.37	-0.58	0.28	
EM				
Mean Return	0.54	1.10	1.30	0.76
(t-stat)	(1.08)	(2.42)	(2.37)	(2.05)
Average RS^+	-1.19	-0.17	0.96	
ALL				
Mean Return	0.70	0.88	1.01	0.30
(t-stat)	(1.88)	(2.42)	(2.33)	(1.45)
Average RS^+	-1.37	-0.43	0.65	

Panel B: Portfolio Sort on RK^+

	Low	Medium	High	High−Low
DM				
Mean Return	0.90	0.77	0.65	-0.25
(t-stat)	(2.37)	(2.24)	(1.84)	(-1.52)
Average RK^+	3.67	4.83	7.18	
EM				
Mean Return	1.28	0.95	0.69	-0.59
(t-stat)	(2.64)	(2.23)	(1.21)	(-1.91)
Average RK^+	3.61	4.70	8.34	
ALL				
Mean Return	1.10	0.83	0.66	-0.44
(t-stat)	(2.79)	(2.30)	(1.61)	(-2.82)
Average RK^+	3.61	4.72	7.74	

Table 5: Fama-MacBeth Regressions

This table reports the results from Fama and MacBeth (1973) cross-sectional regressions for 23 developed markets (DM) and 16 emerging markets (EM). Each month, we regress monthly returns on the lagged conditioning variables. The table reports the time-series mean of each coefficient scaled by 10^3 except for the short interest rate IR, which is unscaled. Newey-West t-statistics are reported in parenthesis. ***,** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The sample period ranges from January 1997 to June 2019. For notational simplicity, we suppress all circumflexes and the dependence of the lagged regressors on $\tau - 1$.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: DM Countries						
Constant	5.40*	5.71	5.48	8.64 **	9.12**	11.10
	(1.67)	(1.61)	(1.59)	(2.11)	(2.14)	(1.06)
VOL	36.88	40.46	41.67	39.41	39.25	22.45
	(0.66)	(0.72)	(0.67)	(0.62)	(0.61)	(0.34)
RS	1.07					
	(1.35)					
RK		-0.28				
		(-0.81)				
RS^+			1.47		0.72	-0.45
			(1.63)		(0.58)	(-0.35)
RK^+				-0.84**	-0.77	-1.27**
				(-1.96)	(-1.35)	(-1.98)
$ln(DY)$						-0.25
						(-0.10)
IR						-0.27
						(-0.26)
Panel B: EM Countries						
Constant	7.87*	7.24*	8.90	10.87*	12.43*	13.52
	(1.70)	(1.67)	(1.55)	(1.69)	(1.89)	(0.92)
VOL	-4.16	54.78	-9.83	56.84	7.33	68.11
	(-0.08)	(0.95)	(-0.13)	(0.74)	(0.09)	(-0.85)
RS	3.44**					
	(2.26)					
RK		-0.69				
		(-1.02)				
RS^+			3.27**		4.44**	1.67
			(1.99)		(2.40)	(0.70)
RK^+				-1.49	-1.11	-1.09
				(-1.50)	(-1.04)	(-0.73)
$ln(DY)$						-0.94
						(-0.26)
IR						0.20
						(0.36)
Panel C: ALL Countries						
Constant	5.69*	5.51*	6.27*	8.03**	9.33**	7.23
	(1.71)	(1.68)	(1.82)	(2.09)	(2.46)	(1.07)
VOL	37.70	52.82	33.95	46.90	41.12	36.02
	(0.80)	(1.11)	(0.65)	(0.89)	(0.78)	(0.76)
RS	1.72*					
	(1.92)					
RK		-0.34				
		(-1.25)				
RS^+			1.57*		1.04	0.58
			(1.81)		(1.02)	(0.49)
RK^+				-0.79**	-0.83**	-1.13**
				(-2.06)	(-2.00)	(-2.32)
$ln(DY)$						-0.86
						(-0.54)
IR						-0.04
						(-0.09)