Option-Implied Asymmetry and Market Returns

Conall O'Sullivan[a] and Yan Wang[a]

[a]Michael Smurfit Graduate Business School, University College Dublin

June 2022

In particular, we would like to thank Michael Brennan and Peter Carr for their valuable feedback on this paper. We would also like to thank Mike Lipkin, Scott Murray, Thierry Post, Jean-Paul Renne, David Shimko, Viktor Todorov, Grigory Vilkov, Yajun Xiao, Geoffery Zheng, and Xin Zhou, as well as seminar participants of the BQE Lecture Series at NYU Tandon in October 2020, and participants of the ICDCM 2020, the SWFA 2021 Annual Meeting, and the DMC 2021 for helpful comments and valuable feedback. Conall O'Sullivan (conall.osullivan@ucd.ie) would like to acknowledge the support of Science Foundation Ireland under grant number 16/SPP/3347. Yan Wang (yan.wang@ucdconnect.ie) would like to acknowledge the funding received from the UCD Michael Smurfit Graduate Business School Doctoral Scholarship. The authors have no declarations of interest to make. All errors are our own.

Option-Implied Asymmetry and Market Returns

Abstract

We propose a novel method to estimate risk-neutral quantiles that uses sorting to minimize an objective function given by a convex combination of call and put option prices over the range of available strike prices. We demonstrate that this new method significantly improves the accuracy of quantile estimates relative to existing approaches. We use the method to estimate a novel risk-neutral quantile-based asymmetry measure (RNA-Q) from S&P 500 index options. In contrast to existing risk-neutral skewness measures, we find that RNA-Q is significantly negatively linked to future market excess returns at horizons ranging from one to twelve weeks. Our findings suggest that ex-ante systematic asymmetry does matter when predicting excess market returns.

Keywords: Model-free quantiles, asymmetry, skewness, forecasting, equity risk premium

JEL: C14, G12, G17

1 Introduction

We propose a novel model-free approach to extract risk-neutral quantiles from option prices that improves upon the most commonly used approach in the literature pioneered by Breeden and Litzenberger (1978) and Banz and Miller (1978) resulting in more accurate quantile estimates. We use risk-neutral quantiles from the tails of the risk-neutral distribution to construct a quantile based measure of risk-neutral asymmetry (hereafter denoted as RNA-Q). The leading term in RNA-Q is a quantile-based risk-neutral skewness term. However, RNA-Q is also related to the higher order cumulants thus goes beyond skewness in measuring asymmetry. We estimate RNA-Q using daily data on S&P 500 index options and take RNA-Q to be a proxy for ex-ante market asymmetry. We find that RNA-Q negatively predicts S&P 500 weekly excess returns at horizons ranging from 1 to 12 weeks[1]. This result is robust to the inclusion of risk-neutral skewness[2] as a control along with other commonly used market return predictors such as the variance risk premium. In particular, we find that RNA-Q contains information on subsequent market returns when estimated with tail quantiles but not when RNA-Q is estimated using the most extreme tail quantiles with probability levels in the range of 1% to 3% in the left tail and 97% to 99% in the right tail.

Return distribution asymmetry is typically summarised by skewness, the third moment of return, a natural extension to distributions based on mean and variance alone. The three-moment CAPM of Kraus and Litzenberger (1976) concludes that systematic skewness (also referred to as coskewness), and not idiosyncratic skewness, should be priced in the cross-section of stock returns. Indeed, conditional coskewness is shown

[1] This result holds regardless of whether we average the RNA-Q value over each trading day in the week or evaluate RNA-Q on the last trading day in the week

[2] We use the Bakshi, Kapadia, and Madan (2003) measure of risk-neutral skewness (denoted as RNS-BKM). This can be used as an alternative proxy for ex-ante asymmetry. RNS-BKM, formulated on the the model-free methodology of Bakshi and Madan (2000) and Carr and Madan (2001), is frequently used to to proxy for ex-ante skewness in the literature. See, for example, Rehman and Vilkov (2012), Chang, Christoffersen, and Jacobs (2013), Conrad, Dittmar, and Ghysels (2013), DeMiguel, Plyakha, Uppal, and Vilkov (2013), Stilger, Kostakis, and Poon (2017), Chordia, Lin, and Xiang (2021), for a small sample of papers that use BKM skewness.

to be priced in the cross-section of stock returns in Harvey and Siddique (2000). Furthermore, both the static three-moment CAPM of Kraus and Litzenberger (1976) and the dynamic quadratic pricing kernel in Harvey and Siddique (2000) imply that ex-ante market skewness negatively predicts subsequent market excess returns. However, as far as the authors are aware, this result has so far eluded the literature or results have been very weak[3]. Jondeau, Zhang, and Zhu (2019) find that average cross-sectional skewness negatively predicts subsequent market returns but that market skewness contains no predictive power for market returns. Chang, Christoffersen, and Jacobs (2013) extract risk-neutral skewness (RNS) and risk-neutral kurtosis estimates using the approach of Bakshi, Kapadia, and Madan (2003) (hereafter we abbreviate this RNS estimate as RNS-BKM) from S&P 500 index options. Taking RNS-BKM as a proxy for ex-ante market skewness, they find that stocks with higher exposure to innovations in RNS-BKM generate lower returns on average. This further emphasises the negative link between systematic skewness and the cross-section of stock returns.

A number of models that deviate from the respresentative agent/expected utility framework conclude that idiosyncratic along with systematic skewness may be priced (see, e.g., Mitton and Vorkink 2007, Brunnermeier, Gollier, and Parker 2007 and Barberis and Huang 2008). A negative link between ex-ante total skewness, proxied by RNS-BKM extacted from individual stock options, and subsequent stock returns is found in Conrad, Dittmar, and Ghysels (2013) agreeing with the aforementioned theoretical models. Boyer and Vorkink (2014) show that ex-ante total skewness in option returns is strongly linked with subsequent negative average options returns. However, more recent results demonstrate a positive link between total RNS and subsequent stock returns. In particular, Chordia, Lin, and Xiang (2021), Stilger, Kostakis, and Poon (2017) and

[3] Chang, Zhang, and Zhao (2011) show that market skewness estimated with daily return data is marginally significantly associated with subsequent monthly market returns. The skewness coefficient is significantly negative for monthly and quarterly returns and (in)significantly positive for semi-annual (annual) returns. However, their data sample from January 1996 to December 2005 is very short making it difficult to draw robust conclusions, in particular, when using longer horizon overlapping holding periods.

Rehman and Vilkov (2012) (Bali, Hu, and Murray 2014) show that there is a positive link between ex-ante stock skewness and subsequent stock returns (ex-ante expected stock returns derived from analyst expectations) using RNS-BKM as a proxy for ex-ante skewness. Informed trading (Chordia, Lin, and Xiang 2021), low RNS stocks being overpriced combined with limits to arbitrage (Stilger, Kostakis, and Poon 2017) along with mispricing in the stock market (Rehman and Vilkov 2012) are shown to be the likely reasons for the positive causal link between RNS and stock returns. The contradictory results regarding the sign of the link between RNS and stock returns can be partially explained by the formation period used to estimate RNS (Stilger, Kostakis, and Poon 2017). A positive link between RNS and stock returns ensues when the most recently available end-of-month data for RNS is used, whereas a negative relation is found when the RNS is averaged over a longer formation period[4]. Borochin, Chang, and Wu (2020) also show that RNS extracted from short maturity options positively predicts stock returns due to informed trading whereas, RNS extracted from longer maturity options negatively predicts stock returns consistent with skewness preference.

Along with the mixed empirical results discussed above there are a number of arguments against the use of skewness as a risk factor. Brockett and Kahane (1992) show that skewness preference is not necessarily a consequence of expected utility maximization for investors with utility functions that have a positive third derivative. The coskewness risk premium reported in Harvey and Siddique (2000) weakens considerably when cubic utility functions are restricted to display risk aversion over the full wealth domain (Post and Levy 2005, Post, Van Vliet, and Levy 2008). Martin (2013) shows that higher order cumulants/moments (not just the third moment) of the consumption growth process make a significant contribution to the equity risk premium in a consumption-based asset pricing model that allows for infrequent disasters. Jiang, Wu, Zhou, and Zhu (2020)

[4]For instance, in Conrad, Dittmar, and Ghysels (2013) stocks are sorted based on RNS-BKM averaged over the proceeding quarter with stock returns evaluated at a quarterly holding period. This result is also consistent with mean reversion in RNS.

show that skewness is significantly negatively associated with subsequent stock returns but only in periods of high volatility or high sentiment. Jiang, Wu, Zhou, and Zhu (2020) show that their measure of asymmetry, which focuses on the asymmetry between the probability of upside versus downside returns in excess of one standard deviation, is more consistently negatively associated with subsequent stock returns than skewness. This motivates us to assess whether a risk-neutral measure of systematic asymmetry, that goes beyond skewness, is significantly linked to future market excess returns. Our measure of asymmetry focuses on asymmetry in the risk-neutral quantiles located in the tails of the distribution. Using the Cornish-Fisher (CF) expansion (Cornish and Fisher 1938, Abramowitz and Stegun 1972), we show that the leading term in our risk-neutral asymmetry measure is equal to a quantile-based version of risk-neutral skewness but that the asymmetry measure also depends on higher order standard cumulants.

We offer a number of contributions to the literature. First, we propose a new method to estimate risk-neutral quantiles. Our method involves sorting an objective function of option prices to the desired probability level α, whereas traditional methods (see, e.g., Breeden and Litzenberger 1978) estimate risk-neutral quantiles by sorting an objective function of first derivatives of option prices. Sorting option prices, rather than first derivatives of option prices, turns out to be crucial in reducing standard errors of risk-neutral quantile estimates. Breeden and Litzenberger (1978) risk-neutral quantiles estimated with nonparametric methods will have larger asymptotic variance compared to quantiles estimated with our approach. This is because our approach sidesteps the need for differentiation and thus avoids the "curse of differentiation"[5]. Second, we demonstrate the superiority of our method to extract risk-neutral quantiles relative to the existing method with Monte Carlo experiments. We mimic S&P 500 index option price quote conditions at different dates in the sample and show that risk-neutral quantiles

[5]The curse of differentiation refers to the fact that functions are estimated more accurately than function derivatives when using nonparametric methods (see, e.g., Fan and Gijbels 1996, Aït-Sahalia and Duarte 2003, and Bondarenko 2003).

have lower standard errors and do not suffer from quantile crossing (when extracted quantiles are nonmonotonic in the probability level α) unlike the existing approach. Third, we construct a novel measure of risk-neutral asymmetry that we refer to as a standardised symmetric quantile sum (SSQS) and estimate SSQS using S&P 500 index options. SSQS measures the relative asymmetry in the distance between the right strike price to the forward price and the left strike price to the forward price in a zero cost risk reversal consisting of a long position in a binary call option and a short position in binary put option. Using the CF expansion, we show that SSQS is related to a quantile-based risk-neutral asymmetry measure (RNA-Q) that has a leading skewness term along with higher order terms related to higher order cumulants/moments. Fourth, we provide theoretical motivation that excess market returns are decreasing in RNA-Q in a two-period representative investor skewness aware CARA economy. Fifth, we show that RNA-Q negatively predicts future excess market returns. This is in contrast to existing measures of risk-neutral skewness, such as the measure proposed in Bakshi, Kapadia, and Madan (2003), where there is no significant relation between RNS-BKM and future excess market returns. As a result we show that ex-ante systematic asymmetry does matter in the prediction of market excess returns. This result is robust to a number of alternative option implied predictors of market returns popular in the literature.

The remainder of this article proceeds as follows. Section 2 provides more background on option implied measures of asymmetry and tail risk. Section 3 introduces our novel method to estimate risk-neutral quantiles and outlines, with Monte Carlo experiments, how this novel method results in more accurate quantiles relative to existing methods. Section 4 introduces our novel measure of risk-neutral asymmetry and links this risk-neutral asymmetry to RNA-Q. Section 5 provides theoretical motivation. Section 6 outlines the S&P 500 index options data and the method used to augment the strike price in order to extract risk-neutral quantiles. Section 7 reports empirical results and the final section concludes.

2 Background

Option markets facilitate the pricing of elementary contingent claims (Arrow 1951; Debreu 1952) along the state dimension, spanning and augmenting the payoff space of financial markets (Ross 1976). Option prices at different strikes contain refined information on state variables that define the opportunity set of investors (Merton 1973). Exploiting different characteristics of risk-neutral and physical return processes, under certain circumstances, options can be used for direct inference on risk premiums, i.e., the wedge between two conditional expectations, for risk factors that are traditionally considered unspanned. For instance, Bollerslev, Tauchen, and Zhou (2009) show that the difference between the risk-neutral and physical conditional variances of equity returns, the variance risk premium (VRP), is a proxy for the volatility of a time-varying volatility in the consumption growth process, which is an important risk factor that is priced in the ERP. Bollerslev and Todorov (2011) find that short-term out-of-the-money (OTM) options contain information on the "crash-o-phobia" that is effectively purged from the compensation for a time-varying jump risk. The "crash-o-phobia" accounts for a large fraction of the ERP and the VRP (as defined in Carr and Wu 2009). Furthermore, under the assumption that the physical large jump intensity process is symmetric, combined with the empirical evidence that the risk-neutral right jump tail is negligible, Bollerslev, Todorov, and Xu (2015) show that the risk-neutral negative jump variation approximates the "crash-o-phobia" and helps explain the ERP predictability by the VRP.

Whilst investors may pay a premium for stocks with positive asymmetrical return distributions, investors have an aversion to negative asymmetry as more negatively skewed stocks are more likely to be subject to left tail disaster risk. The quantile-based RNA measure we propose is more closely related to option implied measures of disaster risk than RNS-BKM as we use tail quantiles to extract RNA-Q. Furthermore, RNA-Q encapsulates higher order standard cumulants that are important in capturing the shape of the return distribution. Options provide an alternative means of estimating time-varying economic disaster risk without the need to consider a long time series to circumvent the Peso problem or, as considered in Kelly and Jiang (2014), a large cross section of asset returns. Furthermore, option-implied measures are real-time reflections of the current states of the economy incorporating information on the conditional expectation of future states. As a result, risk measures based on option prices are forward-looking and potentially more informative than those estimated from historical data. However, option-implied measures are risk-neutral therefore embed both the representative agent's preferences and subjective expectations with additional assumptions needed to disentangle these components. Nonetheless, significant progress has been made in the literature that allows for the recovery of either of these components from a risk-neutral distribution in a semi-model-free way (see, e.g., Ross 2015; Carr and Yu 2012; Martin 2017; Jensen, Lando, and Pedersen 2019; Schneider and Trojani 2019; Kadan and Tang 2020; Jackwerth and Menner 2020).

Certain issues arise when extracting option-implied information using model-free measures or through, more technically, spanning (see Bakshi and Madan 2000). Option prices are quoted in the market only for a finite strike range and over a discrete set of strike prices. This renders inferences on risk-neutral distributions problematic, especially at the tails (see, e.g., Jiang and Tian 2007; Figlewski 2009; Andersen, Bondarenko, and Gonzalez-Perez 2015). Despite these limitations, the estimation of option-implied distribution tails and disaster risk measures has been a topic of significant research activity (see, e.g., Carr, Ellis, and Gupta 1998; Bollerslev and Todorov 2011; Du and Kapadia 2012; Vilkov and Xiao 2013; Siriwardane 2015; Hao 2017; Lu and Murray 2019). The quantile-based RNA measure we propose is particularly sensitive to deep out-of-the-money option prices. As a result, in the next section, we introduce a novel method to estimate risk-neutral quantiles that is more robust to measurement error in the tails than the existing quantile estimation approach.

3 Model-Free Quantiles

In this section a novel method to estimate risk-neutral quantiles is introduced. Provided the strike range is sufficient to encompass the quantiles to be estimated, this method leads to risk-neutral quantile estimates that have smaller estimation errors relative to existing procedures, particularly for quantiles in the tails of the distribution.

3.1 Definition

The method we propose to estimate risk-neutral quantiles is related to the objective function introduced in quantile regression (see Koenker and Bassett 1978). We apply the objective function under the context of option pricing.

Proposition 3.1. *The α quantile for a risk-neutral distribution of horizon T, $Q_{\alpha,T}$, is the solution to the following minimization problem*

$$Q_{\alpha,T} = \arg\min_{K \in \mathbb{K}} \; (1-\alpha)\, Put\,(K,T) + \alpha\, Call\,(K,T) \;, \tag{1}$$

where K denotes the strike price and $\mathbb{K}$ denotes the set of strike prices.

Proof. The objective function is a nonnegative weighted sum of convex functions given the no-arbitrage condition, i.e., the second derivative of put or call prices with respect to strike price is positive (see, e.g., Carr and Madan 2005). Hence, the objective function is convex (see, e.g., Boyd and Vandenberghe 2004) and the first order condition with respect to K gives a global solution:

$$\begin{aligned}
\frac{\partial}{\partial K}\left[(1-\alpha)\, e^{-r_f T}\int_0^K (K-S)\, dF_T(S) + \alpha\, e^{-r_f T}\int_K^\infty (S-K)\, dF_T(S)\right] &= 0\\
(1-\alpha)\, F_T(K) + \alpha\, (F_T(K)-1) &= 0\\
F_T(K) &= \alpha
\end{aligned}$$

where r_f denotes the risk-free rate and $F_T(S)$ denotes the risk-neutral distribution function of the optioned asset price S for horizon T. □

In Proposition 3.1, no assumption is made regarding the return generating process, except for the implicit assumption of no-arbitrage and consequently the existence of a risk-neutral measure. Thus, risk-neutral quantile estimates that are the solution of Eq. (1) are model-free. In the following, we refer to risk-neutral quantiles estimated as solutions to Eq. (1) as model-free quantiles (MFQs).

3.2 Illustrative Example

Our novel procedure to estimate risk-neutral quantiles is as follows. We first augment the discrete set of available option price quotes to obtain a continuum of call and put option prices over a high resolution grid of strike prices. We then solve the optimization problem by finding the value on the x-axis (strike grid) that minimizes the objective function in Eq. (1). We refer to this approach of finding the minimum as sorting as it does not require information on the derivatives of the objective function. This optimization problem is convex if the set of augmented high resolution chain of option prices is convex in the strike price. This is equivalent to the principal of no-arbitrage holding for the augmented high resolution chain of option prices. In practise, this requires that the input market option price quotes are convex in strike and that the augmenting procedure preserves this convexity. We return to this issue in Section 3.4.

Figure 1 illustrates the method by depicting the estimate for the risk-neutral price quantile Q_α when $\alpha = 5\%$ under the Black-Scholes model. We set the underlying price to 100, the risk-free rate to 0, time to maturity to 1 year, and volatility to 20%. The top panel depicts the objective function of Eq. (1) and the solution that minimizes the objective function by sorting where, for illustration purposes only, 10^6 strikes are equally spaced between 50 and 150 (in a subsequent subsection we conduct Monte Carlo experiments to assess the performance of the method using a set of option price quotes across the strike range that replicate index option market conditions). The bottom panel depicts the cumulative distribution function of the log-normal distribution given these parameters. Figure 1 shows the procedure delivers (almost) the exact $Q_{5\%}$ which is equal to 70.54. However, it should be noted that Proposition 3.1 is model-free hence our procedure does not rely on the assumption that the underlying price follows geometric Brownian motion.

[Figure 1 about here.]

3.3 Discussion

In their seminal paper Breeden and Litzenberger (1978) (hereafter BL) show that the pricing of elementary contingent claims (or Arrow-Debreu securities) is equivalent to solving second partial derivatives of a European call option pricing function with respect to strike prices. Inherent in this result is the connection between a risk-neutral cumulative distribution function (RN-CDF) and first partial derivatives of a European call or put option pricing function with respect to strike prices (see, e.g., Figlewski 2009) given by

$$F(K) = e^{r_f T} \frac{\partial\, Call\,(K,T)}{\partial K} + 1\,, \text{ or} \tag{2}$$

$$F(K) = e^{r_f T} \frac{\partial\, Put\,(K,T)}{\partial K}\,, \tag{3}$$

where $F(K)$ denotes the RN-CDF of the optioned asset price and we have suppressed the dependence on maturity T to save on notation. To estimate quantiles with the BL approach (hereafter referred to as BLQs), the RN-CDF is first estimated by differentiation and then the quantile Q_α is given by the solution to:

$$Q_\alpha = \arg\min_{K \in \mathbb{K}} |F\,(K) - \alpha|.$$

As $F(K)$ can be estimated with a chain of call or put options using either Eq. (2) or Eq. (3), BLQs can be obtained using either:

$$Q_\alpha = \arg\min_{K\in\mathbb{K}} |e^{r_f T}\frac{\partial\, Call\,(K,T)}{\partial K} + 1 - \alpha|\,, \text{ or} \tag{4}$$

$$Q_\alpha = \arg\min_{K\in\mathbb{K}} |e^{r_f T}\frac{\partial\, Put\,(K,T)}{\partial K} - \alpha|\,. \tag{5}$$

The inputs to the model-free quantile estimator proposed in Eq. (1) are option prices, which renders Eq. (1) more flexible when used in practice compared to the BL approach. By applying put-call parity to Eq. (1), the quantile extraction method can be rewritten either as:

$$Q_\alpha = \arg\min_{K\in\mathbb{K}} \;\; Call\,(K,T) - (1-\alpha)\left(S - e^{-r_f T}K\right),\;\text{ or} \tag{6}$$

$$Q_\alpha = \arg\min_{K\in\mathbb{K}} \;\; Put\,(K,T) + \alpha\left(S - e^{-r_f T}K\right)\,. \tag{7}$$

Comparing Eq. (6) with Eq. (4) or Eq. (7) with Eq. (5) we see that the MFQ method requires minimising an objective function based on option prices whereas the BLQ method requires minimising an objective function based on the first derivative of option prices. In theory when an infinite number of market option price quotes are available over the strike range, the BLQ and MFQ solutions will coincide. Figure 2 depicts the two objective functions along with the corresponding quantiles when option prices are assumed to follow a Black-Scholes model. The spot price is $S = 100$, the interest rate is $r = 0$, the dividend yield is $q = 0$, the time-to-maturity is $T = 1$, and volatility is $\sigma = 0.20$. The set of strike prices ranges from 10 to 100 in steps of 0.01 index points. The true Black-Scholes quantiles at 1%, 5% and 10% are, respectively, 61.55, 70.54 and 75.85. The MFQ and BLQ estimates coincide in this case at, respectively, 61.55, 70.54, and 75.86.

[Figure 2 about here.]

In practise option price quotes are observed with measurement error due to discreteness of prices, asynchronous trading, bid-ask spreads and variation in liquidity across moneyness. Furthermore, option price quotes are available at a finite number of strike prices and over a finite strike range. As a result, to estimate quantiles the discrete set of option price quotes must be augmented to obtain a continuum of call and put option prices on a high resolution grid of strike prices. BLQs and MFQs will coincide if a parametric model is used to augment the set of available option price quotes to a high resolution grid, as shown above for the case of the Black-Scholes model. However, BLQs and MFQs will not necessarily coincide if nonparametric methods are used to augment the set of available option price quotes. BLQs require nonparametric derivative estimates whereas only level estimates are needed for MFQs. As nonparametric level estimates have lower asymptotic variance than nonparametric derivative estimates (see, e.g., Fan and Gijbels 1996), MFQs will be more accurate than BLQs.

Nonparametric estimators are commonly used for inference on risk-neutral distributions (see, e.g., Aït-Sahalia and Lo 1998; Aït-Sahalia and Duarte 2003; Fengler and Hin 2015; Ludwig 2015). Compared with parametric models, nonparametric estimators are not subject to misspecification errors thus, naturally complement model-free measures that require stringent conditions on the set of strike prices but (almost) no assumptions on underlying return distributions. Nevertheless, nonparametric estimates in general converge much slower than those of parametric models. The rate of convergence is even slower when estimating derivatives, which is anecdotally known as the "curse of differentiation" (as previously discussed in Section 1). It is also cumbersome to impose no-arbitrage conditions on nonparametric estimators. For these reasons, nonparametric inferences in the literature are mostly conducted on risk-neutral distributions of market indices, e.g., the S&P 500 index, where there is usually a large number of option price quotes available over different strikes at a given maturity. Moreover, when using the BLQ method (Eq. (4) or Eq. (5)) to estimate risk-neutral quantiles, a nonparametric estimator with an objective function that is designed to optimize the goodness-of-fit of levels (option market prices) as opposed to first derivatives is first applied to augment existing strike prices[6]. Proposition 3.1 accommodates the use of sorting or differentiation methods to estimate risk-neutral quantiles, whereas when differentiation followed by sorting is used to solve risk-neutral quantiles (based on Breeden and Litzenberger 1978), this inevitably exposes the quantile estimates to higher standard errors by lowering the rate of convergence. Moreover, MFQs are always monotonically increasing with α, even when the strike augmentation procedure does not guarantee convexity. To see this, we note that when α is increased in either Eq. (6) or Eq. (7) this rotates the objective function clock-wise. Given a solution to the objective function Q_{α_1} at a given value of α_1, a new solution Q_{α_2} at α_2, where $\alpha_2 > \alpha_1$, will either be larger or remain at the current solution, regardless of whether the strike augmentation procedure (smoother) is locally concave or convex. Quantile crossing can happen under BLQs when the strike augmentation procedure does not guarantee convexity.

3.4 Monte Carlo Experiments

To illustrate the improved convergence properties of MFQs we conduct a number of Monte Carlo experiments. We simulate a chain of noisy Black-Scholes out-of-the-money (OTM) call option prices and OTM put option prices using the approach in Bondarenko (2003). The noise in the option price is assumed to be a uniform random variable with a spread equal to half the maximal allowable spread according to CBOE rules (see Appendix A for more details). We calculate ITM option prices from OTM option prices using put-call parity. As the maximum allowable spread increases with the bid quote of the option price, options that are close to the money (further from the money)

[6] The goodness of fit of a nonparametric linear estimator amongst other linear estimators can be measured by the linear minimax efficiency (see, e.g., Fan and Gijbels 1996). In particular, local linear and quadratic regression estimators achieve the optimal linear minimax efficiency for the C_2 (twice continuously differentiable) class on estimating interior levels and slopes, respectively. The C_2 class is a bottom line requirement for modelling option prices, such that elementary contingent claim prices exist (see, e.g., Breeden and Litzenberger 1978; Härdle and Hlávka 2009).

will have higher (lower) measurement error. As we only simulate OTM options in this experiment, options that are closer to the money have the highest prices and hence the highest measurement error in absolute terms. Whereas, the measurement error in far OTM simulated option prices is larger in relative terms.

Table 1 depicts results from a Monte Carlo simulation where we compare model-free quantiles (MFQs) estimated using the approach in Eq. (1) with the Breeden and Litzenberger (1978) approach (BLQs). It should be noted that BLQs are also model-free estimates. In this table we deliberately do not preprocess the simulated noisy option prices to impose convexity with respect to strike prices. In the next table we do consider the impact of ensuring that the input option data is convex in the strike price. We replicate the conditions of the S&P 500 index near maturity chain of options on the most recently available data in our sample which corresponds to June 28, 2019. In this case there are a large number of option price quotes available over a wide range of strike prices. On this day, there were 160 different strike prices ranging from a minimum strike price of 2310 to a maximum strike price of 3155. Most option price quotes are spaced at uniform intervals of 5 index points, although the 2^{nd}, 5^{th}, 6^{th}, 7^{th}, and 8^{th} furthest OTM put option price quotes are 10 index points away from their nearest neighbours to the left. In the interests of replicating the experiments, we round down the inputs to the Black-Scholes model. As a result we use the following inputs to simulate Black-Scholes option prices: spot price $S = 2942$, interest rate $r = 0.02$, dividend yield $q = 0.02$, and time-to-maturity $T = 0.08$. We use a single implied volatility from the option quote with a strike price that is the closest to the forward price, i.e. an ATM forward implied volatility $\sigma = 0.13$.

[Table 1 about here]

Panel A of Table 1 compares the performance of MFQs to BLQs when the coarse grid option price chain is interpolated with cubic splines, with no smoothing applied, to obtain a fine resolution grid of option prices. We first convert the noisy OTM put (call) option prices to ITM call (put) option prices. We then combine the simulated OTM put (call) option prices with ITM put (call) options prices so that we have a full chain of put (call) option prices across the strike range. We use a cubic spline to interpolate the chain of simulated noisy call option prices at a very fine grid of strike prices with a uniform interval of 0.01 index points, where the original set of available option price quotes has a strike interval of 5 (or 10) index points. We separately interpolate the chain of simulated noisy put option prices. For MFQs, we form the objective function given in Eq. (1) for a given probability level α and find, by sorting, the point on the x-axis that minimizes this objective function. This point also determines the quantile corresponding to the probability level α. For BLQs, we numerically differentiate the fine grid of call option prices with respect to the strike price to estimate the CDF according to equation Eq. (2) and find, by sorting, the quantile on the CDF x-axis corresponding to the probability level α on the y-axis. We estimate MFQs and BLQs for $\alpha = 1\%, 2\%, \ldots, 99\%$. We also report the root mean square error, given by

$$RMSE = \sqrt{\sum_{\alpha} \left(\hat{Q}_{\alpha} - Q_{\alpha}^{BS} \right)^2} \tag{8}$$

for $\alpha = 1\%, 2\%, \ldots, 99\%$, where $\hat{Q}(\alpha)$ are the estimated quantiles using either MFQs or BLQs and $Q^{BS}(\alpha)$ are the true Black-Scholes quantiles. This measures, for one simulation of an option price chain, the deviation of estimated quantiles from Black-Scholes quantiles at 99 different quantile values. We then average the RMSE over all simulations.

As is clear from the results in Panel A of Table 1, MFQs have lower standard errors than BLQs when using cubic splines to interpolate the noisy option price chain. In particular, the standard errors are approximately 10 times lower at the 1% to 5% left tail quantiles and more than 3 times lower at the 95% to 99% right tail quantiles. The mean RMSE is 34.162 index points for the BLQ method and 7.835 index points for the MFQ method. The ratio of the mean RMSEs is 4.360 meaning that the BLQ method results in over 4 times higher RMSE than the MFQ method over the entire range of quantiles estimated. Figure 3(a) depicts the simulation with the highest RMSE value for the BLQ method and demonstrates the poor performance of BLQs relative to MFQs, particularly at the lower left quantiles. It is clear that BLQs result in quantile crossing whereas MFQs do not. Figure 3(b) depicts a typical simulation where it is still clear that BLQs are considerably more noisy than MFQs. We do not plot the simulation with the highest RMSE for the MFQ method but it should be noted that this simulation results in even higher RMSE for the BLQ method.

[Figure 3 about here.]

Panel B of Table 1 depicts results from a Monte Carlo simulation that compares MFQs and BLQs but where a nonparametric smoothing method is used to augment the noisy option price chain. We simulate market option prices with measurement error as before. However, we use a nonparametric local linear regression (LLR) method (see, e.g., Aït-Sahalia and Duarte 2003) to generate option prices at a high resolution grid of strike prices, with a strike interval of 0.01 index points, by smoothly approximating the chain of low resolution noisy option prices. We do this separately for the call and put option price chain. We use a Gaussian kernel and, for each simulation, the optimal bandwidth for LLR is chosen by leave-one-out least squares cross-validation (LSCV) using the chain of call option prices[7]. The use of LSCV in nonparametric physical quantile estimation is shown to be asymptotically optimal under certain conditions in Li, Lin, and Racine (2013). We estimate MFQs as previously outlined but using the two smoothed high resolution option price chains (separately smoothed call and put chains) as opposed to

[7] In further experiments we selected the optimal bandwidth using only OTM call option prices and OTM put option prices but the results are very similar to the results when using call option prices alone and are omitted to save space.

the cubic spline interpolated option price chains. To estimate BLQs with LLR we do not need to numerically differentiate the call option chain with respect to the strike price as we use the slope coefficient from LLR as the estimate of the first derivative for each grid point. However, the standard error for higher order derivatives is larger than for level estimates in local polynomial regression (see Eq. (3.7) Fan and Gijbels 1996 where it is shown that the asymptotic conditional variance of a LPR derivative estimate, $m^{(v)}$, increases with v, the order of the derivative, with a denominator proportional to h^{1+2v} where h is the grid spacing with $h \to 0$ in the asymptotic analysis, see also Eq. (3.16) Aït-Sahalia and Duarte 2003). Thus, even though we are using an estimate of the slope from the LLR output, we still expect MFQs to be more accurate than BLQs due to the "curse of differentiation"[8].

The results in Panel B of Table 1 demonstrate a marked improvement in BLQs compared to the previous case in Panel A when no smoothing is applied to the option price chains. This is to be expected as numerical differentiation of noisy data amplifies the noise (see, e.g., Ahnert and Abel 2007 and Ling 2006). However, MFQs still have lower standard errors, in particular, at the more extreme quantiles of 1% to 5% and 95% to 99%. The mean RMSE is 9.277 index points for the BLQ method and 6.578 index points for the MFQ method. The ratio of the mean RMSEs is 1.41 meaning that the BLQ method has an average RMSE that is approximately 41% higher than the average RMSE for the MFQ method across all the quantiles evaluated. Figure 3(c) depicts the simulation that results in the highest RMSE value for the BLQ method. This highlights the potential inaccuracies of BLQs, in particular, at the left most extreme strike prices even after smoothing is applied. In this simulation, where the leftmost OTM put option prices have larger measurement error by chance, MFQs are much more accurate than BLQs in capturing the left most quantiles. Figure 3(d) depicts a more typical simulation

[8] Unlike the curse of dimensionality that only has serious practical implications once the dimension of the problem is ≥ 4, the curse of differentiation applies even when comparing first derivative estimates to level estimates.

where BLQs and MFQs are so close that they are indistinguishable to the eye in the plot.

In Panel C of Table 1 we use local quadratic regression (LQR) to augment the noisy option price chain. The optimal order of the polynomial to use in LPR is $k+1$ when estimating a derivative of a function of order k, based on asymptotic analysis (see, e.g., Fan and Gijbels 1996). As a result we assess the performance of LQR given that we estimate a first order derivative in BLQs. As with LLR we use leave-one-out LSCV to select the bandwidth. The accuracy of the BLQ method deteriorates when LQR is used in place of LLR. The mean RMSE value rises to 19.045 index points for LQR relative to 9.277 index points under LLR and 34.162 index points under cubic splines. The accuracy of the the MFQ method also decreases when LLR is replaced with LQR but not nearly as drastically as that of the BLQ method. The mean RMSE value for the MFQ method is 7.375 index points under LQR which is higher than the mean RMSE value of 6.578 of the MFQ method under LLR but lower than the mean RMSE value of 7.835 when using the MFQ method with cubic splines. Furthermore, MFQs have lower standard errors than BLQs, in particular, at the more extreme quantiles of 1% to 5% and 95% to 99%. In agreement with Aït-Sahalia and Duarte (2003) we find that asymptotic theory is not necessarily a good guide to follow when dealing with sample sizes typically encountered in index options markets.

The "curse of differentiation" applies to any nonparametric method (see, e.g., Stone 1982). This is highlighted in Panel D of Table 1 where we use cubic B-splines to smooth option price chains. We use leave-one-out LSCV to choose the number of interior knots in the B-spline. The interior knots are placed in the strike domain to divide the strike price grid into uniform partitions. For example, when n_k interior knots are used, knots are placed to divide the strike domain into $1 + n_k$ uniform partitions. As with the cubic spline, LLR and LQR approaches, MFQs are more accurate than BLQs. MFQs have lower standard errors at the more extreme quantiles of 1% to 3% and 97% to 99%. The mean RMSE is 5.29 index points for the BLQ method and 5.22 index points for the MFQ method. The ratio of the mean RMSEs is 1.01 thus, the BLQ method is 1% less accurate than the MFQ method in terms of average RMSEs. Figure 3(e) depicts the simulation that results in the highest RMSE for BLQs. This figure highlights the problem that the BLQ method has at the left tail of the distribution as the B-spline fits the left boundary of the data. Not only do we observe quantile crossing, but the quantile estimates in the lower left tail are often 200 index points away from the true quantile value. This particular simulation causes problems for the BLQ method at the left tail but the MFQ method performs well and MFQs are more accurate than BLQs in capturing the left tail quantiles. Figure 3(f) depicts a typical simulation in terms of RMSEs and, even in this average simulation, MFQs are more accurate than BLQs at the left extreme strike prices.

In Table 2 we repeat the simulations in Table 1 but preprocess the simulated noisy option price data to ensure that the option prices are convex with respect to the strike price. We use the constrained least squares approach of Aït-Sahalia and Duarte (2003) to effectively clean (or "convexify") the noisy option price input data to ensure it is convex in strike price. Panel A of Table 2 reports results when cubic splines are used to augment the strike space. As cubic splines do not necessarily preserve convexity, we note that the resulting augmented high resolution option price chains will not necessarily be convex in strike, even though the coarse grid of input option prices are convex in strike. BLQs are much better behaved using preprocessed convex data with cubic splines. The mean RMSE value falls from 34.16 index points in Panel A of Table 1 to 8.27 index points in Panel A of Table 2 when the input data is cleaned. However, it should be noted that MFQs remain more accurate with a mean RMSE value that is 9% lower than the corresponding value for BLQs. Panel B and C of Table 2 report results when LLR and LQR are used, respectively, to augment the strike space. LLR preserves the convexity of the input option prices, as shown in Aït-Sahalia and Duarte (2003), so that the augmented high resolution option price chains remain convex in strike price. As a result we see a marked improvement in BLQ performance relative to their performance in Panel A. Under LLR, MFQs have mean RMSE values that are 0.4% lower than the corresponding values for BLQs when using leave-one-out LSCV to select the bandwidth. Preprocessing the input option data to ensure it is convex in strike along with the use of a convexity preserving smoothing process results in improved performance for BLQs however, MFQs are still marginally more accurate as evidenced by mean RMSE values.

LQR is not guaranteed to preserve convexity. However, according to asymptotic analysis, local regression with a polynomial of order 2 is optimal when estimating the first derivative required for BLQs. The former effect outweighs the latter in our experiments as LQR quantiles are less accurate than LLR quantiles under both the BLQ and MFQ method. Furthermore, the difference in performance between BLQs and MFQs widens with BLQs having mean RMSE values that are 7% higher than the corresponding values for MFQs compared to only 0.4% higher under LLR. Panel D of Table 2 report results when smoothing B-splines are used to augment the strike space. The number of knots are chosen by leave-one-out LSCV. Here we see that that the combination of preprocessing data along with smoothing the data with B-splines actually increases the mean RMSE values relative to the case where no preprocessing occurs in Table 1. For example, BLQs have a mean RMSE value equal to 5.29 index points in Panel D of Table 1. This increases to 7.22 index points in Panel D of Table 2. Similarly, MFQs have a mean RMSE value equal to 5.22 index points in Panel D of Table 1 which increases to 6.88 index points in Panel D of Table 2. In this case, the data is over smoothed when preprocessing along with smoothing B-splines are used to extract quantiles whether BLQs or MFQs are used. Despite the drop in performance induced by using convexity preprocessing before the application of smoothing B-splines, MFQs remain more accurate than BLQs.

[Table 2 about here]

We repeat the above experiments using a two-state mixture lognormal model (see, e.g., Bahra 1997 and Melick and Thomas 1997) to generate option prices that are a closer match to market prices. We do this both without and with preprocessing of input option prices to ensure convexity. We also conduct additional simulations replicating S&P 500 index option market conditions on a date approximately half way through our sample (July 30, 2007) and on an earlier date in our sample (July 30, 1999). In these cases there are, respectively, 52 and 30 unique strike option price quotes available. In the vast majority of cases the MFQ method outperforms the BLQ method with lower mean RMSEs and lower standard errors. Further discussion is deferred to Appendix A. In the next section we introduce a quantile-based measure of risk-neutral asymmetry that we link to skewness and higher order moments via the CF expansion.

4 Quantiles-Based Measures of Risk-Neutral Asymmetry

In this section we propose a novel measure of return asymmetry that we refer to as standardised symmetric quantile sums (SSQS). We then introduce a quantile-based asymmetry measure that is based on a CF expansion of SSQS.

Using the CF expansion up to third order (see, e.g., Cornish and Fisher 1938 and Abramowitz and Stegun 1972), we write return quantiles, $q_\alpha(r_\tau)$, in terms of standard normal quantiles, z_α, as follows:

$$q_\alpha(r_\tau) \approx \mu(r_\tau) + \sigma(r_\tau) z_\alpha + \frac{1}{6}\sigma(r_\tau)\left(z_\alpha^2 - 1\right)\gamma(r_\tau)$$

where r_τ is the return at horizon $\tau = T - t$, $\mu(r_\tau)$, $\sigma(r_\tau)$ and $\gamma(r_\tau)$ are, respectively, the mean, standard deviation and skewness of the return distribution at horizon τ. We then write risk-neutral price quantiles $Q_\alpha(S_\tau)$ in terms of risk-neutral return quantiles as follows:

$$
\begin{aligned}
\ln Q_\alpha(S_\tau) &= \ln S_0 + q_\alpha(r_\tau) \\
&\approx \ln S_0 + \mu(r_\tau) + \sigma(r_\tau) z_\alpha + \frac{1}{6}\sigma(r_\tau)\left(z_\alpha^2 - 1\right)\gamma(r_\tau) \\
&= \ln F_\tau + \sigma(r_\tau) z_\alpha + \frac{1}{6}\sigma(r_\tau)\left(z_\alpha^2 - 1\right)\gamma(r_\tau) \qquad (9)
\end{aligned}
$$

using the fact that the log forward price is given by $\ln F_\tau = \ln S_0 + \mu(r_\tau)$ with $\mu(r_\tau) = (r - q)\,\tau$. Let σ_{BKM} denote the risk-neutral standard deviation which we estimate using the method in Bakshi, Kapadia, and Madan (2003). We define SSQS as follows:

$$
\begin{aligned}
SSQS_\alpha(r_\tau) &= \frac{\ln\left(Q_\alpha(S_\tau)/F\right) + \ln\left(Q_{1-\alpha}(S_\tau)/F\right)}{\sigma_{BKM}(r_\tau)} \qquad (10) \\
&\approx \frac{1}{3}\left(z_\alpha^2 - 1\right)\gamma(r_\tau) \qquad (11)
\end{aligned}
$$

for $\alpha \in [0, 50\%)$, where Eq. (11) is derived by substituting Eq. (9) for both α and $1 - \alpha$ into Eq. (10). SSQS measures the asymmetry in the distance of the right strike price (1-α quantile) from the forward price and the left strike (α quantile) from the forward price in a zero cost binary option risk reversal[9]. The asymmetry measure is expressed in terms of returns relative to the forward price required to obtain (pay) a dollar payoff conditional on the right $1-\alpha$ (left α) quantile being reached at horizon τ by the underlying optioned asset. We further standardise to units of risk-neutral standard deviation to purge the impact that standard deviation has on quantiles (see Eq. (9)).

For a given quantile level α, the term multiplying skewness in Eq. (11) is a fixed scaling factor. To obtain a quantile-based risk-neutral asymmetry estimate (RNA-Q) with skewness as the leading term we then invert Eq. (11) to express skewness $\gamma(r_\tau)$ as

[9]This implicitly assumes that left quantile is negative and the right quantile is positive.

a function of $SSQS_\alpha(r_\tau)$:

$$RNA\text{-}Q_\alpha(r_\tau) := \gamma(r_\tau) \approx \frac{3}{(z_\alpha^2 - 1)} SSQS_\alpha(r_\tau). \tag{12}$$

The relation between RNA-Q and SSQS is not defined when $z_\alpha = \pm 1$, corresponding to $\alpha \approx 16\%$. The denominator term $z_\alpha^2 - 1 = 0$ when $z_\alpha = \pm 1$. Hence, RNA-Q should only be used for quantiles that are more (less) than one standard deviation away from the mean. In the appendix we show that RNA-Q also depends on higher order cumulants when higher order CF expansions are used.

RNA-Q deviates from the robust measure of asymmetry proposed in Hinkley (1975) (a generalisation of which is proposed in Groeneveld and Meeden 1984 and used in Ghysels, Plazzi, and Valkanov 2016) as RNA-Q uses the risk-neutral standard deviation as the denominator as opposed to the interquantile range. Both SSQS and RNA-Q can be viewed as alternative measures of asymmetry in a risk-neutral return distribution that are more robust to outliers than the usual moment based skewness measure. Related asymmetry measures of return distributions under the physical measure are studied in Ghysels, Plazzi, and Valkanov (2016) and Jiang, Wu, Zhou, and Zhu (2020). Ghysels, Plazzi, and Valkanov (2016) use a quantile-based measure of asymmetry and link this measure to physical skewness using the CF expansion. Using their quantile-based skewness measure, Ghysels, Plazzi, and Valkanov (2016) show that international portfolios with larger weights in emerging market indices have significantly positive skewness which increases the certainty equivalent gains of these portfolios. As mentioned previously, Jiang, Wu, Zhou, and Zhu (2020) propose a measure of asymmetry given by the difference in the upside and downside tail probabilities that stock returns exceed a one standard deviation move, and an entropy scaled version of this measure. They find that their tail based asymmetry measure (estimated using historical density functions on single stocks) results in stronger cross-sectional pricing effects than skewness. This suggests that tail focused measures of asymmetry can be useful in empirical studies. Furthermore, the theoretical argument used in Jiang, Wu, Zhou, and Zhu (2020) to motivate the pricing of asymmetry shows that market risk-neutral asymmetry, as opposed to single stock physical asymmetry measure, is priced. However, physical asymmetry measures are used in their cross-sectional pricing analysis. This provides motivation to assess whether our risk-neutral ex-ante measure of asymmetry, RNA-Q, predicts excess market returns. Before progressing to empirical results, we provide a different theoretical argument to that in Jiang, Wu, Zhou, and Zhu (2020) although both approaches amount to the same conclusion: that the equity risk premium is a declining function of risk-neutral asymmetry in a skewness aware economy.

5 Theoretical Motivation

To motivate our following empirical results we show that the ERP is a decreasing monotonic function of RNA-Q using the Smooth Half Normal (SHN)-CARA economy of de Roon and Karehnke (2017). Combining the SHN density function with a representative investor CARA economy, de Roon and Karehnke (2017) show that the ERP is a decreasing function of physical skewness and examine the variation of optimal portfolio weights as a function of skewness. We derive the risk-neutral density function and closed form expressions for RNA-Q in a SHN-CARA economy.

The SHN density function is given by:

$$g(x) = \begin{cases} \lambda_1 f(x; m, s_1) & \text{if } x \leq m, \\ \lambda_2 f(x; m, s_2) & \text{if } x > m. \end{cases}$$

where $f(x; \mu_x, \sigma_x)$ is the normal density function with mean μ_x and standard deviation σ_x, λ_1 and λ_2 are chosen to ensure the density function is continuous and integrates to one, and where s_1, s_2 and m are chosen to match the mean μ, variance σ and skewness

γ of the excess return distribution. Following de Roon and Karehnke (2017) we assume a two period economy with a representative investor that maximizes a CARA utility function $u(x) = -e^{-\theta w_0 (1+r_f+x)}$ where θ is the risk aversion coefficient, w_0 the initial wealth, r_f is the risk-free rate and x is the excess return with $x \sim SHN(\mu, \sigma, \gamma)$. The risk-neutral density $g^*(x)$ is related to the physical density function (subject to conditions such as complete and frictionless markets) as follows:

$$g^*(x) = \frac{u'((1+r_f+x))g(x)}{\int_{-\infty}^{\infty} u'((1+r_f+x))g(x)}$$

The resulting risk-neutral density function is given by:

$$g^*(x) = \begin{cases} \lambda_1^* f(x; m - \theta w_0 s_1^2, s_1) & \text{if } x \leq m, \\ \lambda_2^* f(x; m - \theta w_0 s_2^2, s_2) & \text{if } x > m. \end{cases}$$

where

$$\begin{aligned} \lambda_i^* &= \lambda_i \frac{c_i}{c}, \\ c_i &= e^{\frac{1}{2}\left(\theta^2 w_0^2 s_i^2\right)}, \\ c &= \lambda_1 c_1 \Phi\left(\theta w_0 s_1\right) + \lambda_2 c_2 \left(1 - \Phi\left(\theta w_0 s_2\right)\right) \end{aligned}$$

for $i = 1, 2$ and where $\Phi(x)$ denotes the standard normal CDF.

Left tail physical quantiles for the SHN density function are given in de Roon and Karehnke (2017). For completeness, we present left and right tail physical quantiles below along with risk-neutral quantiles, as derived in this paper, in a SHN-CARA economy:

$$\begin{aligned} q_\alpha &= \Phi\left(\tfrac{\alpha}{\lambda_1}\right) s_1 + m, & q_{1-\alpha} &= \Phi\left(1 - \tfrac{\alpha}{\lambda_2}\right) s_2 + m, \\ q_\alpha^* &= \Phi\left(\tfrac{\alpha}{\lambda_1^*}\right) s_1 + m - \theta w_0 s_1^2, & q_{1-\alpha}^* &= \Phi\left(1 - \tfrac{\alpha}{\lambda_2^*}\right) s_2 + m - \theta w_0 s_2^2, \end{aligned}$$

where q_α (q^*_α) denotes the physical (risk-neutral) quantile. Hence, the risk-neutral asymmetry measure SSQS is given by:

$$SSQS_\alpha = \frac{1}{\sigma^*}\left(\Phi\left(\frac{\alpha}{\lambda_1^*}\right)s_1 + \Phi\left(1-\frac{\alpha}{\lambda_2^*}\right)s_2 - \theta w_0(s_1^2+s_2^2) + 2\,(m-r_f)\right)$$

where α is the quantile probability level (e.g., $\alpha = 5\%$) and where σ^* is the risk-neutral standard deviation (the formula for σ^{*2} is given in Appendix B). In a SHN-CARA economy, the risk-neutral standard deviation σ^* is different to σ when physical skewness $\gamma \neq 0$. The VRP, defined as $VRP = \sigma^{*2} - \sigma^2$, is positive for negative skewness, negative for positive skewness and zero for zero skewness. However, the variation in σ^* with skewness does not alter the fact that the ERP is negatively related to the RNA-Q in a SHN-CARA economy using realistic parameter values as depicted below. Finally, RNA-Q is given by:

$$RNA\text{-}Q_\alpha = \frac{3}{(z_\alpha^2-1)}SSQS_\alpha.$$

We use the same parameters as de Roon and Karehnke (2017) for ease of comparison. Assume excess returns have a mean $\mu = 7.28\%$, a standard deviation $\sigma = 14.96\%$ and physical skewness is varied from -0.95 to 0.8. We use an initial wealth $w_0 = 1$ and a risk-free rate $r_f = 0$. When skewness is zero the risk aversion parameter that results in the investor being fully invested in the risky asset is given by $\theta = 3.2829$ and, following de Roon and Karehnke (2017), we round θ to 3.25. Figure 5 presents results for the SHN-CARA economy. Panel (a) depicts the physical and risk-neutral density functions for a single skewness value of -0.5. Panel (b) depicts the ERP versus physical skewness with a clear negative link from skewness to ERP replicating the result in de Roon and Karehnke (2017). Panel (c) shows the variation of RNA-Q and quantile-based physical asymmetry (PA-Q) with physical skewness. As expected both RNA-Q and PA-Q are increasing functions of physical skewness with RNA-Q more sensitive to skewness than

PA-Q. Finally, panel (d) illustrates the monotonic negative link from RNA-Q to ERP.

[Figure 5 about here.]

The above analysis can also be applied to the case of a representative investor with an S-shaped utility function (in place of a CARA utility function) along with a probability weighting scheme, that puts more weight on the tail probabilities, as in the Cumulative Prospect Theory (CPT) of Tversky and Kahneman (1992). Barberis and Huang (2008) use the CPT approach to motivate investor skewness preference. de Roon and Karehnke (2017) examine the relation between the ERP/optimal portfolio weights and physical skewness in such a SHN-CPT economy. Using the SHN-CPT economy with the same parameters as de Roon and Karehnke (2017), we find that the ERP is also a decreasing monotonic function of RNA-Q[10] with an even stronger negative link between ERP and RNA-Q relative to the link in the SHN-CARA economy. Thus, a monotonic negative association from RNA-Q to the ERP can be motivated using either a SHN-CARA or SHN-CPT economy.

6 Data and Strike Augmentation

In this section we outline the data and method used to estimate a time series of risk-neutral quantiles for the S&P 500 index. This panel of quantiles is then used to construct RNA-Qs along with control option implied predictors.

6.1 Data

The option data we use is from OptionMetrics. The sample period is from January 4, 1996 to June 28, 2019. We focus on constant 30-day risk-neutral quantiles which can be estimated using the most liquid options. We filter the S&P 500 index options

[10] Details are omitted to save space but are available upon request.

sample following the VIX methodology[11]. Prior to October 6, 2014, a near-term (with at least one week to maturity) and a next-term SPX option chains are included in the sample. Since then, the sample also includes Friday settled Weeklys[12] and contains two option chains with more than 23 days but no more than 37 days to maturity. We adjust maturities of options to distinguish between AM and PM settlements. The AM settled SPX options are counted one day less than the PM settled SPXW options. For SPX options, we use settlement dates as opposed to expiration dates to calculate maturities. The sample is further cleaned by eliminating any zero bid quotes, and by excluding any further OTM options once two zero bid quotes are encountered. In particular, we define an at-the-money (ATM) option by its strike price that equals to the futures price[13]. We obtain realized variance data from the Oxford-Man Institute of Quantitative Finance with daily data available from January 2000 to the end of our sample. Prior to 2000, we estimate S&P 500 realized variance using the daily sum of 5-minute squared returns of the SPDR S&P 500 ETF Trust (ticker SPY) extracted from the NYSE Trade and Quote (TAQ) database. We estimate 5-minute squared returns using mid-quote prices from 09:30 to 16:00 along with the previous overnight close-to-open squared return from 16:00 to 09:30. We obtain data on the left tail jump variation (LJV) and the risk-neutral probability of a 10% drop in the S&P 500 from tailindex.com (see Andersen and Todorov 2019).

[11] See the VIX white paper and Information Circular IC14-075 from CBOE for further information.

[12] A number of steps are taken to ensure a smooth transition to Weekly options including taking into account CBOE expiry adjustments due to holidays to identify Friday settled Weeklys using the SPXW symbol prefix in OptionMetrics IvyDB. Further detail is available from the authors upon request.

[13] We use the futures price provided by OptionMetrics directly which is different from the futures price calculated according to the CBOE methodology. Occasionally, the OptionMetrics futures price field has missing records for some option strikes. We backfill these missing records using valid futures price records that match the date and expiration fields of the missing records. Further detail on the difference between the OptionMetrics and CBOE futures prices can be found in the OptionMetrics manual and the VIX white paper.

6.2 Augmenting the Strike Space

We use smoothing cubic B-splines to obtain call and put option prices over a fine grid of strike prices. The number of knots in the B-spline is selected by LSCV and is applied separately for both the near and far maturity option chains in each date in the sample to obtain a constant maturity 30-day set of quantiles for each date. We have also tested the use of cubic-splines, LLR, LQR and the semi-parametric approach proposed in Figlewski (2009)[14] to augment the space of strike prices. However, we use cubic B-splines in the following empirical analysis given the robust performance of this method in the Monte Carlo simulations in Section 3.4. It must be noted that the subsequent empirical analysis of risk-neutral quantiles, quantile-based risk-neutral skewness measures, and market return predictive regression results are not overly sensitive to the choice of method used to augment the strike space provided quantiles are estimated with the model-free method proposed in this paper.

For each date within the sample, and for a given maturity, we first discard deep OTM options using a minimum bid cutoff of 50 cents. We combine put and call samples by retaining OTM and ATM options. We retain bid, ask, and mid quotes and use mid quotes to augment the strike price. Figure 4 illustrates the steps that are taken to estimate the 5% and 1% price quantiles on a single day. We demonstrate these steps using a sample of options that belong to the near-term chain of the S&P 500 index from two different dates: June 28, 2019 and January 17, 1996. The top left panel depicts mid-quote market prices for call and put options along with cubic B-spline smoothed option prices on June 28, 2019. The top right panel panel depicts the objective functions used to estimate model-free quantiles at $\alpha = 1\%, 5\%$ and 10%. The vertical line on both plots depicts the 1%, 5% and 10% quantiles, where the solution to the minimization

[14] This procedure uses a two-piece quartic polynomial that smoothly passes through bid-ask spreads of option IVs converted from option market prices using the Black-Scholes formula thus is not strictly nonparametric. Figlewski's procedure is used to estimate a time series of RN-PDFs in Birru and Figlewski (2012) and a time series of RN-CDFs in Linn, Shive, and Shumway (2018).

problem is found by sorting. The bottom left and right panels of Figure 4 depict the same plots for January 17, 1996.

[Figure 4 about here.]

We estimate risk-neutral price quantiles, $Q_\alpha(S_\tau)$, on each sample date, for near- and next-term option chains separately, and for $\alpha = 1\%$ to 99% in increments of 1% from January 1996 to June 2019[15]. We linearly interpolate two estimates from the near- and next-term option chains to obtain constant 30-day quantile estimates.

7 Empirical Results

In this section we report summary statistics, the correlation coefficient matrix, in-sample and out-of sample predictive regression results that assess the performance of RNA-Q in predicting market excess returns. To summarise results in a more succinct manner and to ensure stable estimates, we average the RNA-Q estimates over a range of α values, similar to the approaches of Groeneveld and Meeden (1984) and Ghysels, Plazzi, and Valkanov (2016) where quantile based measures of asymmetry are integrated over α. We focus on α values ranging from 1% to 10% to ensure we are far from the region of α where the CF approximation break downs ($\alpha \approx 16\%$). We first consider a range based tail RNA-Q measure by averaging RNA-Q$_\alpha$ over the range $\alpha \in [1\%, 2\%, \ldots, 10\%]$, denoted by RNA-Q$_{1-10}$. We further breakdown this RNA-Q measure by focusing on an extreme tail range, RNA-Q$_{1-3}$, an intermediate tail range, RNA-Q$_{4-7}$, and a moderate

[15] Our quantile estimates are based on a continuum of options augmented using cubic B-splines. We intentionally do not append risk-neutral tails that can be modelled using Generalised Extreme Value distributions, see for example Figlewski 2009. The reason is twofold. First, we would like to highlight another advantage of the proposed MFQ method, that is, model-free quantiles always converge to the smallest or largest option strikes when the solution for a fixed α is beyond the finite strike range quoted in the market. Second, there are a number of dates on which the quantile estimates (either near or far maturity or both) are equal to the smallest or largest strikes in the market (hence the true quantile exceeds these strikes). We restrict our attention to $\alpha \in [1\%, 99\%]$ in the empirical study as a trade-off between capturing the tails of a risk-neutral distribution and not using quantiles that regularly lie outside the strike range.

tail range, RNA-Q_{8-10}. In the remainder of the empirical section we use weekly data extracted every Wednesday, as is standard in the literature (see, for example, Figlewski 2018), by taking the average value from the previous Thursday to the current Wednesday.

7.1 Summary Statistics and Correlations

Panel A of Table 3 reports summary statistics for the weekly quantile based skewness measure RNA-Q_α estimated using a single quantile value for each tail with $\alpha = 1\%, \ldots, 10\%$ for the left tail and $1 - \alpha = 99\%, \ldots, 90\%$ for the right tail. RNA-Q_α have mean values that vary from -1.37 ($\alpha = 2\%$) to -1.15 ($\alpha = 9\%$) compared to the mean value of RNS-BKM of -1.69 (see Panel B). The standard deviation of the RNA-Q_α measures vary from 0.26 ($\alpha = 2\%$) to 0.44 ($\alpha = 10\%$) compared to a standard deviation of 0.67 for RNS-BKM. The $AR(1)$ coefficients for the RNA-Q_α measures vary from 0.80 ($\alpha = 1\%$) to 0.67 ($\alpha = 2\%$) with the $AR(1)$ coefficients approximately equal to 0.7 for the other RNA-Q_α values. This compares to a highly persistent $AR(1)$ coefficient of 0.94 for RNS-BKM. Also reported are the summary statistics for the different averaged RNA-Q estimates. Similar to single quantile pair estimates of RNA-Q_α, the averaged RNA-Q mean values have higher means and lower standard deviations than the corresponding mean and standard deviation of RNS-BKM. The $AR(1)$ coefficients for the range based RNA-Q measures are between 0.76 (for RNA-Q_{1-10}) and 0.71 (for RNA-Q_{8-10}).

[Table 3 about here.]

Panel B of Table 3 reports summary statistics for the controls used in predictive regressions. The use of σ^2_{BKM} as an ERP predictor is motivated by a Merton type ERP expression, where the instantaneous ERP is a linear function of instantaneous variance (see, e.g., Merton 1980 and Cochrane 2009)[16]. We also use Martin (2017)'s model-

[16]The VIX^2 is a forward-looking measure of integrated variance, which coincides with the risk-neutral BKM variance, σ^2_{BKM}, under a diffusion model (see, e.g., Du and Kapadia 2012). Summary statistics and predictive regressions that follow are very similar if we replace σ^2_{BKM} with VIX^2 thus VIX^2 summary statistics are not reported.

free measure of total variance times the gross risk-free rate of return, $R_f \cdot SVIX^2$, as a lower bound for the ERP that we denote as LB-SVIX. ID is the implied dividend yield of the S&P 500 index taken from OptionMetrics IvyDB. The implied dividend yield has been shown to predict market returns in Bilson, Kang, and Luo (2015) and Golez (2014)[17]. We also use volume and open interest weighted IV spreads, denoted respectively as VWVS and OWVS, put forth in Atilgan, Bali, and Demirtas (2015) as a measure of information flow from options markets to stock markets. These measures capture differences between OTM put option IVs and OTM call option IVs and are shown to be significant in predicting daily and weekly market excess returns. We consider ex-post realized variance (using the most recent 22 trading days) as a potential predictor of the ERP, as RV is shown to be a significant predictor of the ERP in Atilgan, Bali, and Demirtas (2015). We also use the lagged ERP as a control to account for possible serial correlation in weekly excess returns. We use the variance risk premium, VRP, as in Bollerslev, Tauchen, and Zhou (2009), that has been shown to be a robust predictor of market excess returns at three month horizons and longer. We define the VRP as: $VRP_t = IV_t - RV_t$ where we use implied variance $IV_t = \sigma^2_{BKM,t}$ and where $RV_t =$ ex-post 22-day realized variance. A related control we consider is the left risk-neutral jump variation (LJV) component of the VRP as proposed in Bollerslev and Todorov (2014). LJV is a component of the VRP that compensates investors for bearing jump risk. Bollerslev, Todorov, and Xu (2015) show that LJV extracted from short term deep-out-of-the money options (maturity between 8 and 45 days) is highly significant in predicting aggregate market returns. Using the same methodology, Andersen and Todorov (2019) construct a risk-neutral probability of a 10% stock market plunge over the next week using short term deep OTM options. Given the relationship of this variable

[17] Strictly speaking, Golez (2014) uses a corrected dividend price ratio to predict market returns that adjusts the dividend price ratio to account for time-varying dividend growth rates, where the latter is estimated as the difference between the implied dividend yield and the dividend price ratio. As a result, the corrected dividend price ratio is positively correlated with the implied dividend yield. Hence, we use the latter as a predictor.

to option implied asymmetry we use it as a control denoted by Probability Drop (PD). Finally, we use an option implied version of the extreme tail difference (ETD) measure as used in Jiang, Wu, Zhou, and Zhu (2020) on single stocks. We define ETD as: $ETD_t = P_{up,t} - P_{down,t}$, where P_{up} (P_{down}) is the risk-neutral probability of excess market returns being higher (less) than a one standard deviation move over the next month. All option implied controls are averaged over the week to be consistent with our use of RNA-Q[18]. This means that the VRP we use is based on an implied variance that is averaged over the week and a realized variance that is the sum of the previous 22-days. We also tested a VRP where the implied variance is extracted on the final day of the week and averaged over the previous 22-days, with the same 22-day rolling sum for RV as before, and results are unchanged.

The mean 30-day risk-neutral variance, RNV-BKM, is 0.39% (equivalent to an annualised risk-neutral variance/volatility of 4.67%/21.61%) and is very close to the mean of Martin's lower bound LB-SVIX of 0.35% (equivalent to an annualised mean lower bound of 4.18%). The mean volatility spreads (both volume and open interest weighted) are approximately 10%, meaning OTM put IV is on average 10% higher than ATM call IV. The mean ID is 1.79% with a standard deviation of 0.49%. The mean of the weekly frequency 1-month ERP (weekly ERP scaled to a one month horizon by multiplying by 4) is 0.48% (annualised mean of 5.68%) and has a standard deviation of 9.17%. Not reported are longer horizon ERP statistics whose one-month means remain relatively constant but where the one-month standard deviation falls with increasing horizons. For instance, the mean of the 4-week horizon ERP is 0.47% with a standard deviation of 4.32%. The mean RV is 0.25% (corresponding to an annualized realized variance/volatility of 3.04%/17.42%). The mean VRP is 0.136% (this corresponds to annualized difference in implied and realized standard deviation of 4.18%). The mean ETD is -4.37% meaning that the risk-neutral probability of a one standard deviation

[18] Results are very similar when controls are extracted on the final day of the week as opposed to using the estimate averaged over the week and are available upon request.

downward move is on average 4.37% higher than the corresponding risk-neutral probability of a one standard deviation upward move. The mean LJV value is 7.52 and the mean PD value is 72%. The $AR(1)$ coefficients for the volatility spread measures (0.74 for VWVS and 0.76 for OWVS), the VRP (0.80), ETD (0.70) and LJV (0.83) are similar to the averaged RNA-Q measures. These are considerably lower than the $AR(1)$ coefficients of the other controls. These $AR(1)$ coefficients are, respectively, 0.94 for RNS-BKM, 0.95 for both RNV-BKM and LB-SVIX, 0.99 for ID, 0.95 for RV and 0.98 for PD.

Table 4 reports correlations of the quantile based RNA-Q measures that are averaged over different ranges of quantiles, along with the controls used in predictive regressions. RNA-Q_{1-3} has a correlation of 0.64 (0.41) with RNA-Q_{4-7} (RNA-Q_{8-10}) which, although statistically significant at 1%, is sufficiently low to suggest that these RNA-Q measures may pick up different facets of the RN density tail behaviour. The correlation between RNA-Q_{4-7} and RNA-Q_{8-10} is quite high at 0.82 suggesting that the most extreme tail skewness, RNA-Q_{1-3}, behaves separately from the intermediate and moderate tail average range RNA-Q values, RNA-Q_{4-7} and RNA-Q_{8-10}. This is expected as the variation in the most extreme tail skewness measure, RNA-Q_{1-3}, depends on the variation in the most extreme strike prices for deep OTM options that make it past the VIX filtering rules.

[Table 4 about here.]

The correlations of RNS-BKM with quantile based RNA-Q values are generally far less than 1 with a value of 0.38 for RNA-Q_{1-3}, 0.12 for RNA-Q_{4-7}, 0 for RNA-Q_{8-10} and 0.15 for the overall tail average RNA-Q_{1-10}. Thus, the quantile based RNA-Q values are very distinct from the standard measure of RNS with the most extreme average RNA-Q value, RNA-Q_{1-3}, exhibiting the highest correlation with RNS-BKM, although this is still far from 1 at 0.38. This can be seen as evidence that RNS-BKM is sensitive to strike range variation.

An interesting point to note is that the correlation between RNS-BKM and RNV-BKM (RV) is statistically significantly positive at 0.23 (0.17). This is the opposite to what we would expect if we interpret RNS-BKM as a tail risk measure. Intuitively, one would expect a skewness based tail risk measure to become more negative (tail risk increases) as variance increases. This is exactly what we find with quantile based RNA-Q measures. The correlations between RNA-Q and RNV-BKM (RV) are statistically significantly negative for RNA-Q_{1-10}, RNA-Q_{4-7} and RNA-Q_{8-10}, although the correlation between RNV-BKM (RV) and RNA-Q_{1-3} is statistically insignificant at -0.04 (-0.03). Thus, the positive relation between the standard RN measures of skewness and variance (realized variance) is reversed when using quantile-based RN measures of asymmetry.

With the exception of the most extreme average RNA-Q_{1-3} measure, there is a significant positive (negative) correlation between the other RNA-Q measures and contemporaneous ERP (one-week ahead ERP) at a significance level of 5% (5%) for RNA-Q_{4-7} and 1% (10%) for RNA-Q_{1-10}. A similar (but reverse) pattern is observed for the VRP. The correlation between the VRP and contemporaneous ERP (one-week ahead ERP) is statistically significantly negative (positive) at the 1% level. The correlation of ID with ERP and one-week ahead ERP remains the same with a correlation value of 5% that is statistically significant at the 10% significance level. The $AR(1)$ coefficient of ID is particularly high at 0.99 and is the likely reason for this persistence in correlation values. Also worth noting is the high positive correlation of ETD with RNA-Q across all averaged ranges. This is to be expected as both RNA-Q and ETD are measures of risk-neutral asymmetry. LJV and PD are both significantly negatively correlated with the different RNA-Q measures. Downside risk-neutral asymmetry is high when LJV (PD) is high and this corresponds to low values for RNA-Q. Hence, these negative correlations are expected although are sufficiently far from one to suggest that LJV and PD contain different information to RNA-Q. Finally, we note that LJV and PD are highly correlated with the risk-neutral variance measures of RNV-BKM and LB.

7.2 Univariate Predictive Regressions

We run a sequence of univariate predictive regressions, using excess returns of the S&P 500 index as the dependent variable and RNA-Q as the independent variable as follows:

$$\frac{4}{h}\left(r_{t,t+h} - r^f_{t,t+h}\right) = \beta_0 + \beta_1 \times RNA\text{-}Q_{t,\alpha} + \epsilon_{t,t+h}, \tag{13}$$

where $r_{t,t+h}$ is the return on the S&P 500 index from t to $t+h$, h represents the forecast horizon that ranges from 1 week to 12 weeks in increments of one week and α represents the range of quantiles used to estimate RNA-Q. Weekly market excess returns are scaled to a one month horizon so that predictive regression coefficients are comparable across different return horizons. The risk-free rate, $r^f_{t,t+h}$, is observed at t and matures at $t+h$. We calculate risk-free rates using the zero curve provided by OptionMetrics which is derived from ICE IBA LIBOR rates and settlement prices of CME Eurodollar futures. The zero curve is discrete thus we use a piecewise cubic Hermite interpolating polynomial to back out 1-week to 12-week risk-free rates that match the return horizon $r_{t,t+h}$[19]. RNA-Q$_{t,\alpha}$ denotes the quantile-based RNA-Q measures observed at time t. These are categorised into the overall tail RNA-Q measure for $\alpha \in [1\%, \ldots, 10\%]$, the most extreme tail RNA-Q measure for $\alpha \in [1\%, 2\%, 3\%]$, the medium tail RNA-Q measure for $\alpha \in [4\%, 6\%, 7\%]$, and the moderate tail RNA-Q measure for $\alpha \in [8\%, 9\%, 10\%]$. The overlap in weekly observed variables generates a high degree of positive autocorrelation which biases the standard error estimates downwards. To account for this, we report Newey-West standard error estimates in the regression analysis using a lag length equal to two times the overlap in the excess returns (see Bollerslev, Tauchen, and Zhou 2009).

Table 5 presents results for in-sample univariate predictive regressions on market

[19] We also use daily risk-free rates from Kenneth French's data library and aggregate these daily risk-free rate to weekly values in the calculation of excess returns. Results using risk-free rates from Kenneth French's data library are very similar to the results reported here using LIBOR rates.

excess returns. Reported are beta coefficient estimates (β_1) for regressors, standardised beta coefficient estimates which are scaled by standard deviations of regressors, t-statistics based on Newey-West standard errors allowing for a lag equal to the two times the overlap in the dependent variable, corresponding p-values, and adjusted R^2s. Panel A reports results where the RNA-Q measures are averaged over each day of the week and Panel B reports results where the RNA-Q measures are extracted on the final day of the week.

[Table 5 about here.]

Focusing on Panel A we observe that RNA-Q_{1-10} (RNA-Q_{4-7}) significantly negatively predicts future market excess returns at all horizons bar the first week (all horizons) as expected from the skewness aware asset pricing model used in Section 5. There is no significant link between RNS-BKM and subsequent market returns confirming previous results that this standard RN measure of skewness does not predict market excess returns. This is not expected from the skewness aware asset pricing model we use. However, RNS-BKM is noisier than the RNA-Q measures with a full-sample standard deviation of 0.67 versus a standard deviation of 0.31 for RNA-Q_{4-7}. Furthermore, in unreported simulation results, similar to those conducted in Jiang and Tian (2005) and Jiang and Tian (2007) for risk-neutral variance, we find that RNS-BKM is very sensitive to tail truncation when a finite strike range is used to compute RNS-BKM[20]. However, quantile-based measures of risk-neutral asymmetry are not dependent on the strike range as long as the strike range covers the quantiles under consideration.

The statistical significance associated with RNA-Q_{1-10} is marginal at the 10% level for horizons 2W to 3W and 11W to 12W with significance peaking at the 5% level at

[20] We use the CGMY model of Carr, Geman, Madan, and Yor (2002) to simulate a chain of call and put options with model parameters taken from the paper. We simulate discrete strike prices with a truncated strike range that we vary from the minimum to the maximum observed strike range in our sample. We find that RNS-BKM is particularly sensitive to left strike truncation using the negatively skewed CGMY density function.

intermediate horizons of between 4W and 10W inclusive. RNA-Q_{4-7} is statistically significant at the 5% level for 1W to 4W and 9W to 11W horizons, 1% at horizons of 5W to 8W and is marginal at 10% at 12W. Andersen, Fusari, and Todorov (2020) find that a proxy for the time-varying left jump tail risk predicts future weekly returns of international equity market indices at longer horizons of more than 10 weeks. Interpreting RNA-Q as a tail risk measure, it is no surprise that the RNA-Q measures perform well at intermediate to longer horizon return forecasting. However, RNA-Q_{4-7} also significantly predicts market excess returns at shorter horizons, in particular, 1W ahead excess returns. Using RNA-Q_{4-7}, the t-statistic on 1W excess returns is -2.23. Since there is no overlap in 1W market returns, this t-statistic is not subject to the econometric issue that impact standard errors when forecasting overlapping returns as discussed in Hodrick (1992). The standardised beta coefficient shows that a one standard deviation increase in RNA-Q_{4-7} results in a 61 basis points decrease in market excess returns (7.32% annualised) at the 1-week horizon, which is extremely significant from an economic point of view. The standardised beta coefficients decrease gradually to approximately 50 basis points (bps) (6% annualised) at intermediate horizons and approximately 40 bps (4.8% annualised) at the longer horizons of 11-weeks but remain economically significant. The corresponding values for standardised betas of RNA-Q_{1-10} are 44 bps (5.3% annualised) at the 1-week horizon decreasing gradually to 30 bps (3.6% annualised) at the longest 12W horizon. Thus, RNA-Q_{1-10} is also economically significant in predicting market excess returns but the statistical and economic significance of RNA-Q_{1-10} is always weaker than that of RNA-Q_{4-7} at each forecast horizon considered. The extreme tail RNA-Q measure, RNA-Q_{1-3}, does not predict market excess returns at any horizon. The extreme left ($\alpha = 1\%$) and right ($1 - \alpha = 99\%$) tails are highly correlated with the minimum and maximum strike prices that satisfy the VIX filtering rules. As a result, RNA-Q_{1-3} often reflects the extreme strike price quotes as opposed to the ex-ante asymmetry in the underlying market return distribution. This is analogous to results in Andersen, Bondarenko, and Gonzalez-Perez (2015) where a high frequency measure of the VIX may change, not due to changes in underlying volatility, but as a result of the changes in the range of option price quotes used to calculate the VIX. Andersen, Bondarenko, and Gonzalez-Perez (2015) propose using a more consistent corridor volatility index that truncates the option price quotes used to calculate implied volatility at a range that is more coherent over time. Our RNA-Q measures automatically create such a consistent range by fixing the probability levels α and $1-\alpha$ at which we examine quantile variation over time. The moderate tail RNA-Q measure, RNA-Q_{8-10}, predicts market excess returns at horizons of 4W to 9W inclusive but the predictions are always weaker than the medium tail RNA-Q measure, RNA-Q_{4-7}, and the overall tail RNA-Q measure, RNA-Q_{1-10}. Thus, the 4%-7% regions of the left and right tails are sufficiently far away from the extreme strike price quotes to not be overly effected by changes in extreme strike price quotes but are still sufficiently far enough out in the tail to capture extreme asymmetric moves that investors are mindful of.

Panel B of Table 5 shows that results generally weaken slightly relative to Panel A when we use the most recently available RNA-Q values to predict subsequent weekly market returns as opposed to using RNA-Q values that have been averaged over the five days in the week. The relative strengths of the various predictors are the same as in Panel A with end of week RNA-Q_{4-7} remaining the strongest predictor of subsequent returns relative to the other skewness measures extracted at end of week.

Next we turn our attention to other potential ERP predictors frequently used in the literature. Table 6 presents results for in-sample univariate predictive regressions of market excess returns using the series of controls introduced in Section 7.2.

[Table 6 about here.]

The risk-neutral variance, RNV-BKM, and Martin's lower bound, LB-SVIX, are not significantly associated with the ERP at any horizon considered[21]. VWVS and OWVS

[21] This does not necessarily contradict the results in Martin 2017 where the null hypothesis that $\beta_1 = 1$

positively forecast S&P 500 index returns during the sample period we study. However, the beta coefficient estimates are only marginally statistically significant at 10% at longer horizons of 9W (8W) and higher for VWVS (OWVS)[22]. The positive association between VS and subsequent market excess return we find agrees with the interpretation of VS as a tail risk measure as opposed to an information based interpretation of VS. The implied dividend yield, ID, positively forecasts market excess returns, with the t-statistic rising above 2 at return horizons of 6 weeks and higher. The adjusted R^2s in these regressions are particularly high at longer horizons reaching a maximum of 5.50% at 12 weeks. It must be noted that the persistence of ID, as measured by the $AR(1)$ coefficient, is 0.99 and is much higher than the persistence of the RNA-Q measures. Therefore, the problems associated with highly persistent regressors in long horizon forecasts documented in Boudoukh, Richardson, and Whitelaw (2008) are less likely to be an issue with RNA-Qs than they are with implied dividend yields. Higher values of RV and lagged ERP are associated with lower values of subsequent ERP demonstrating, respectively, the leverage effect and negative serial correlation. However, there is no significant association between RV or lagged ERP and subsequent ERP. The VRP positively forecasts market excess returns but the t-statistics rise above 2 at horizons of 9 weeks to 11 weeks. In predictive regressions of Bollerslev, Tauchen, and Zhou (2009), the VRP is also positively linked to future excess returns and attains its highest significance at a 3-month return horizon using monthly frequency returns over a sample period of January 1990 to December 2007. In our sample of January 1996 to June 2019, using weekly frequency returns, we find the VRP attains its highest significance at intermediate to long horizons of 9 to 11 weeks. Thus, our results are similar, but not identical, to those in Bollerslev,

is not rejected in predictive regressions.

[22] These results are in contrast to those found in Atilgan, Bali, and Demirtas (2015), where both VWVS and OWVS negatively forecast S&P 500 index returns at daily and weekly horizons with informed trading found to be the main explanation. Atilgan, Bali, and Demirtas (2015) study a shorter sample period from January 4, 1996 to September 10, 2008. A potential explanation for the positive relation between VS and future market returns found in this article is that equity option markets have become much larger in notional value since 2009, therefore leading information priced into options by informed traders is no longer present in IV measures, such as VWVS and OWVS.

Tauchen, and Zhou (2009). This is to be expected given the differences in the frequency of returns, the different sample period used and the different data used to construct the realized variance. We find that ETD is significantly negatively linked to excess returns at horizons of 3W and higher with a significance level of 5% attained at horizons of 5W and higher. These results are similar to those we find for the RNA-Q measures which is to be expected given both ETD and RNA-Q are risk-neutral asymmetry measures that focus on the tails of the distribution. Finally, we find LJV and PD are not significant in predicting excess returns over the sample period considered.

7.3 Bivariate Predictive Regressions

In the following bivariate predictive regressions we focus on the RNA-Q measure that results in the most robust ERP predictor, RNA-Q_{4-7}. The results are quantitatively and qualitatively similar, although a little weaker, if we use RNA-Q_{1-10} in place of RNA-Q_{4-7}. Table 7 presents results on bivariate predictive regressions using RNA-Q_{4-7} along with a single control to jointly predict future market excess returns using regressions of the following form:

$$\frac{4}{h}\left(r_{t,t+h} - r^{f}_{t,t+h}\right) = \beta_0 + \beta_1 \times RNA\text{-}Q_{t,4\text{-}7} + \beta_2 \times X_t + \epsilon_{t,t+h}, \tag{14}$$

where X_t is the control variable observed at time t.

The intermediate tail RNA-Q measure, RNA-Q_{4-7}, is robust to the inclusion of the control predictors. The t-statistic on the RNA-Q_{4-7} coefficient remains below -2 at all horizons, with the exception of 12-week (11- and 12-week), when using RNS-BKM, RNV-BKM, LB, RV, or ERP-lag (LJV or PD) as a control. The t-statistic on the RNA-Q_{4-7} coefficient remains below -2 for horizons 1W to 8W when either of the IV spread measures is used as a control, 2W to 9W when implied dividend yield is used as a control and 2W to 10W when the VRP is used as a control. The t-statistic on

RNA-Q_{4-7} increases above -2 but remains close to -2 with the inclusion of ETD as a second predictor. However, the correlation between ETD and RNA_{4-7} is 62% as both variables are measures of risk-neutral asymmetry thus the inclusion of ETD in a bivariate regression is expected to reduce the impact of RNA-Q_{4-7} on the ERP.

The only control variables with beta coefficient t-statistics greater than 2 in the presence of RNA-Q_{4-7} are ID at horizons of 7W and up and the VRP at horizons of 7W to 9W. The implied dividend yield, ID, and the variance risk premium, VRP, have $AR(1)$ coefficients of 0.99 and 0.80, respectively, compared to 0.74 for RNA-Q_{4-7}. As previously noted the high persistence of ID, in particular, could lead to inference problems for this potential predictor (e.g., see Boudoukh, Richardson, and Whitelaw 2008). The control coefficient on ETD is not significant when ETD is used along with RNA-Q_{4-7} to predict the ERP. Thus, the information in RNA-Q_{4-7} is not captured by other commonly used predictors that are shown in the literature to be significant in predicting the ERP.

[Table 7 about here.]

7.4 Out-of-Sample Tests

Table 8 reports out-of-sample (OS) predictive regression statistics on subsequent market excess returns. The statistics reported are OS adjusted R^2s calculated as $1-(1-R^2_{OS}) \times \frac{T-1}{T-k-1}$, where R^2_{OS} is calculated as

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} (y_t - \bar{y}_t)^2}.$$

T is the number of OS forecasts, k is the number of predictive variables, y_t is the predicted variable, $\hat{y}_t$ is the forecast based on a predictive model that is estimated using predictive variables through the period $t-1$, and $\bar{y}_t$ is the forecast based on a nested model that restricts the coefficient on the assessed predictive variable to be 0. The nested model uses the historical average ERP to predict the next period ERP. We also report OS adjusted R^2 from restricted regressions, as in Campbell and Thompson (2008) (hereafter referred to as CT OS adjusted R^2), where the coefficient on the labelled predictor in the regression tables is restricted to be non-negative and, if the forecast is still negative, a second restriction is applied replacing the forecast with 0. Given the theoretical and empirical link between RNA-Q_{4-7} and the ERP is negative, we use *negative* RNA-Q_{4-7} in the CT OS tests to ensure the association between the regressor (-RNA-Q_{4-7}) and regressand (ERP) is positive. OS significance is assessed using a one-sided t-test on the mean squared prediction error (MSPE)-adjusted statistic from Clark and West (2007). We apply this test taking into account the autocorrelation in forecast errors by using Newey-West standard error estimates with lags set equal to 2 times the overlap in h-horizon returns in calculating t-statistics. Forecasts are made on a recursive basis with an initial estimation window set equal to 120 weeks, where the window expands with the full sample size equal to 1,225 weeks. One-sided t-statistics for the MSPE-adjusted statistic, with the null of equal MSPE, are reported in the table. For large enough sample sizes, as applies in our case, standard normal critical values can be used and we can reject the null at 10%/5%/1% if the t-statistic $>$ 1.282/1.645/2.33, respectively.

[Table 8 about here.]

In these OS tests, we consider RNA-Q_{4-7} and the three strongest controls from the in-sample tests: ID, VRP and ETD. Table 8 documents significant OS predictability of market excess returns using RNA-Q_{4-7} at all forecast horizons. The statistical significance is 5% at all horizons. The OS adjusted R^2 peaks at 2.6% at the 8W forecast horizon. The CT OS adjusted R^2s are generally lower than the unrestricted OS adjusted R^2s with the exception of the 1 week horizon. ID is insignificant for forecast horizons of 1 and 2 weeks, marginally significant at the 10% level for a forecast horizon of 3 weeks, and is significant at 5% for horizons of 4 weeks and higher. The OS R^2 and CT OS R^2 values for ID are higher than the respective values for RNA-Q_{4-7} at all forecast horizons however, it must be noted that this may partly stem from the highly persistent nature of ID and the impact this may have on longer horizon predictive regressions (as previously discussed in the in-sample results). When ID is used to predict the ERP, CT OS R^2s are higher (lower) than OS R^2s at shorter (longer) horizons with the crossover occurring at 5W. The variance risk premium, VRP, is insignificant in forecasting the ERP out-of-sample for forecast horizons of 1 to 6 weeks, is significant at the 5% level at horizons of 7W, 8W and 12W and is significant at the 1% level at 9W, 10W and 11W. The OS R^2 (CT OS R^2) values for the VRP are lower than the respective values for RNA-Q_{4-7} at all forecast horizons (forecast horizons from 1 week to 8 week). When the VRP is used to predict the ERP, CT OS R^2s are higher than OS R^2s at all forecast horizons. As with RNA-Q_{4-7}, we use the negative value of ETD to forecast the ERP out-of-sample to ensure a positive link between the regressor (-ETD) and the regressand (ERP) in the CT OS tests. ETD is not significant in ERP out-of-sample forecasts for horizons of 1W to 3W, is significant at the 5% level at horizons of 4W to 8W and 11W to 12W, and is significant at the 1% level at horizons of 9W and 10W. As with the VRP, the restricted OS CT forecasts result in higher adjusted OS R^2s than the unrestricted forecasts across all horizons. ETD has lower OS adjusted R^2 than RNA-Q_{4-7} at all horizons and lower (higher) CT OS adjusted R^2 than RNA-Q_{4-7} at horizons of 1W to 7W (8W to 12W). These results demonstrate that the in-sample results also hold out-of-sample with RNA-Q_{4-7} remaining a strong ERP predictor outperforming ETD and on-par with the performance of the VRP and ID but with better persistence properties than ID.

8 Conclusion

In this article we introduce a novel approach to estimate risk-neutral quantiles that results in more accurate risk-neutral quantile estimates relative to existing quantile estimation procedures. We apply the method to S&P 500 index options to estimate a time series of risk-neutral quantiles. We use these quantiles to construct a novel quantile based risk-neutral asymmetry measure RNA-Q. We show that RNA-Q constructed from quantiles in the left and right tails (with the exception of the most extreme quantiles) predicts market excess returns and is robust to a number of alternative option implied and other commonly used ERP predictors. Further research could focus on whether risk-neutral asymmetry predicts excess returns in an international setting and the construction of a dynamic economy in which conditional risk-neutral asymmetry predicts the ERP.

References

Abramowitz, M. and I. A. Stegun (1972, December). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55. Tenth Printing.* U.S. Government Printing Office.

Ahnert, K. and M. Abel (2007). Numerical differentiation of experimental data: local versus global methods. *Computer Physics Communications 177*(10), 764–774.

Aït-Sahalia, Y. and J. Duarte (2003, September). Nonparametric option pricing under shape restrictions. *Journal of Econometrics 116*(1), 9–47.

Andersen, T. and V. Todorov (2019). Tail index white paper. *Tail Index White Paper available at tailindex.com*.

Andersen, T. G., O. Bondarenko, and M. T. Gonzalez-Perez (2015, October). Exploring Return Dynamics via Corridor Implied Volatility. *The Review of Financial Studies 28*(10), 2902–2945.

Andersen, T. G., N. Fusari, and V. Todorov (2020, July). The Pricing of Tail Risk and the Equity Premium: Evidence From International Option Markets. *Journal of Business & Economic Statistics 38*(3), 662–678.

Arrow, K. J. (1951). An Extension of the Basic Theorems of Classical Welfare Economics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California.

Atilgan, Y., T. G. Bali, and K. O. Demirtas (2015, January). Implied Volatility Spreads and Expected Market Returns. *Journal of Business & Economic Statistics 33*(1), 87–101.

Aït-Sahalia, Y. and A. W. Lo (1998). Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *The Journal of Finance 53*(2), 499–547.

Bahra, B. (1997). Implied risk-neutral probability density functions from option prices: theory and application.

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

Bakshi, G. and D. Madan (2000, February). Spanning and derivative-security valuation. *Journal of Financial Economics 55*(2), 205–238.

Bali, T. G., J. Hu, and S. Murray (2014). Analyst price target expected returns and option implied risk. *Georgetown McDonough School of Business Research Paper* (2516937).

Banz, R. W. and M. H. Miller (1978). Prices for state-contingent claims: Some estimates and applications. *The Journal of Business 51*(4), 653–672.

Barberis, N. and M. Huang (2008, December). Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review 98*(5), 2066–2100.

Bilson, J. F., S. B. Kang, and H. Luo (2015). The term structure of implied dividend yields and expected returns. *Economics Letters 128*, 9–13.

Birru, J. and S. Figlewski (2012). Anatomy of a meltdown: The risk neutral density for the s&p 500 in the fall of 2008. *Journal of Financial Markets 15*(2), 151–180.

Bollerslev, T., G. Tauchen, and H. Zhou (2009, November). Expected Stock Returns and Variance Risk Premia. *The Review of Financial Studies 22*(11), 4463–4492.

Bollerslev, T. and V. Todorov (2011). Tails, fears, and risk premia. *The Journal of Finance 66*(6), 2165–2211.

Bollerslev, T. and V. Todorov (2014). Time-varying jump tails. *Journal of Econometrics 183*(2), 168–180.

Bollerslev, T., V. Todorov, and L. Xu (2015). Tail risk premia and return predictability. *Journal of Financial Economics 118*(1), 113–134.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics 116*(1-2), 85–112.

Borochin, P., H. Chang, and Y. Wu (2020). The information content of the term structure of risk-neutral skewness. *Journal of Empirical Finance 58*, 247–274.

Boudoukh, J., M. Richardson, and R. F. Whitelaw (2008). The myth of long-horizon predictability. *The Review of Financial Studies 21*(4), 1577–1605.

Boyd, S. P. and L. Vandenberghe (2004). *Convex optimization.* Cambridge, UK %3B New York: Cambridge.

Boyer, B. H. and K. Vorkink (2014). Stock options as lotteries. *The Journal of Finance 69*(4), 1485–1527.

Breeden, D. T. and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business 51*(4), 621–651.

Brockett, P. L. and Y. Kahane (1992). Risk, return, skewness and preference. *Management Science 38*(6), 851–866.

Brunnermeier, M. K., C. Gollier, and J. A. Parker (2007). Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review 97*(2), 159–165.

Campbell, J. Y. and S. B. Thompson (2008, July). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies 21*(4), 1509–1531.

Carr, P., K. Ellis, and V. Gupta (1998). Static hedging of exotic options. *The Journal of Finance 53*(3), 1165–1190.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business 75*(2), 305–332.

Carr, P. and D. Madan (2001). Optimal positioning in derivative securities. *Quantitative Finance 1*(1), 19–37.

Carr, P. and D. B. Madan (2005, September). A note on sufficient conditions for no arbitrage. *Finance Research Letters 2*(3), 125–130.

Carr, P. and L. Wu (2009). Variance risk premiums. *The Review of Financial Studies 22*(3), 1311–1341.

Carr, P. and J. Yu (2012, August). Risk, Return, and Ross Recovery. *The Journal of Derivatives 20*(1), 38–59.

Chang, B. Y., P. Christoffersen, and K. Jacobs (2013). Market skewness risk and the cross section of stock returns. *Journal of Financial Economics 107*(1), 46–68.

Chang, E. C., J. Zhang, and H. Zhao (2011). Expected stock returns and the conditional skewness. *The 2011 China International Conference in Finance, Unpublished working paper*.

Chordia, T., T.-C. Lin, and V. Xiang (2021). Risk-neutral skewness, informed trading, and the cross section of stock returns. *Journal of Financial and Quantitative*

Analysis 56(5), 1713–1737.

Clark, T. E. and K. D. West (2007, May). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291–311.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Conrad, J., R. F. Dittmar, and E. Ghysels (2013). Ex ante skewness and expected stock returns. *The Journal of Finance 68*(1), 85–124.

Cornish, E. A. and R. A. Fisher (1938). Moments and Cumulants in the Specification of Distributions. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute 5*(4), 307–320.

de Roon, F. and P. Karehnke (2017). A simple skewed distribution with asset pricing applications. *Review of Finance 21*(6), 2169–2197.

Debreu, G. (1952). A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences 38*(10), 886–893.

DeMiguel, V., Y. Plyakha, R. Uppal, and G. Vilkov (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis 48*(6), 1813–1845.

Du, J. and N. Kapadia (2012). *The tail in the volatility index*. Ph. D. thesis, University of Massachusetts, Amherst.

Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Volume 66. CRC Press.

Fengler, M. R. and L.-Y. Hin (2015, February). Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. *Journal of Econometrics 184*(2), 242–261.

Figlewski, S. (2009, February). Estimating the Implied Risk Neutral Density. SSRN Scholarly Paper ID 1354492, Social Science Research Network, Rochester, NY.

Figlewski, S. (2018). Risk neutral densities: A review. *Available at SSRN 3120028*.

Ghysels, E., A. Plazzi, and R. Valkanov (2016). Why invest in emerging markets? the role of conditional return asymmetry. *The Journal of Finance 71*(5), 2145–2192.

Golez, B. (2014). Expected returns and dividend growth rates implied by derivative markets. *The Review of Financial Studies 27*(3), 790–822.

Groeneveld, R. A. and G. Meeden (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician) 33*(4), 391–399.

Hao, J. (2017, September). A Model-Free Tail Risk Index and Its Return Predictability. SSRN Scholarly Paper ID 2826577, Social Science Research Network, Rochester, NY.

Harvey, C. R. and A. Siddique (2000). Conditional skewness in asset pricing tests. *The Journal of Finance 55*(3), 1263–1295.

Hinkley, D. V. (1975, April). On power transformations to symmetry. *Biometrika 62*(1), 101–111.

Hodrick, R. J. (1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *The Review of Financial Studies 5*(3), 357–386.

Härdle, W. and Z. Hlávka (2009, May). Dynamics of state price densities. *Journal of Econometrics 150*(1), 1–15.

Jackwerth, J. C. and M. Menner (2020, September). Does the Ross recovery theorem work empirically? *Journal of Financial Economics 137*(3), 723–739.

Jensen, C. S., D. Lando, and L. H. Pedersen (2019). Generalized recovery. *Journal of Financial Economics 133*(1), 154–174.

Jiang, G. J. and Y. S. Tian (2005, December). The Model-Free Implied Volatility and Its Information Content. *The Review of Financial Studies 18*(4), 1305–1342.

Jiang, G. J. and Y. S. Tian (2007, February). Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index. *The Journal of Derivatives 14*(3), 35–60.

Jiang, L., K. Wu, G. Zhou, and Y. Zhu (2020). Stock return asymmetry: Beyond skewness. *Journal of Financial and Quantitative Analysis 55*(2), 357–386.

Jondeau, E., Q. Zhang, and X. Zhu (2019). Average skewness matters. *Journal of Financial Economics 134*(1), 29–47.

Kadan, O. and X. Tang (2020). A bound on expected stock returns. *The Review of Financial Studies 33*(4), 1565–1617.

Kelly, B. and H. Jiang (2014, October). Tail Risk and Asset Prices. *Review of Financial Studies 27*(10), 2841–2871.

Koenker, R. and G. Bassett (1978). Regression Quantiles. *Econometrica 46*(1), 33–50. Publisher: [Wiley, Econometric Society].

Kraus, A. and R. H. Litzenberger (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance 31*(4), 1085–1100.

Li, Q., J. Lin, and J. S. Racine (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics 31*(1), 57–65.

Ling, L. (2006). Finding numerical derivatives for unstructured and noisy data by multiscale kernels. *SIAM Journal on Numerical Analysis 44*(4), 1780–1800.

Linn, M., S. Shive, and T. Shumway (2018, February). Pricing Kernel Monotonicity and Conditional Information. *The Review of Financial Studies 31*(2), 493–531.

Lu, Z. and S. Murray (2019, March). Bear beta. *Journal of Financial Economics 131*(3), 736–760.

Ludwig, M. (2015, February). Robust Estimation of Shape-Constrained State Price Density Surfaces. *The Journal of Derivatives 22*(3), 56–72.

Martin, I. (2017, February). What is the Expected Return on the Market?*. *The Quarterly Journal of Economics 132*(1), 367–433.

Martin, I. W. R. (2013, April). Consumption-Based Asset Pricing with Higher Cumulants. *The Review of Economic Studies 80*(2), 745–773.

Melick, W. R. and C. P. Thomas (1997). Recovering an asset's implied pdf from option prices: an application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis 32*(1), 91–115.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica 41*(5), 867–887.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*(4), 323–361.

Mitton, T. and K. Vorkink (2007). Equilibrium underdiversification and the preference for skewness. *The Review of Financial Studies 20*(4), 1255–1288.

Post, T. and H. Levy (2005). Does risk seeking drive stock prices? a stochastic dominance analysis of aggregate investor preferences and beliefs. *The Review of Financial Studies 18*(3), 925–953.

Post, T., P. Van Vliet, and H. Levy (2008). Risk aversion and skewness preference. *Journal of Banking & Finance 32*(7), 1178–1187.

Rehman, Z. and G. Vilkov (2012). Risk-neutral skewness: Return predictability and its sources. *Available at SSRN 1301648*.

Ross, S. (2015). The recovery theorem. *The Journal of Finance 70*(2), 615–648.

Ross, S. A. (1976). Options and efficiency. *The Quarterly Journal of Economics 90*(1), 75–89.

Schneider, P. and F. Trojani (2019). (Almost) Model-Free Recovery. *The Journal of Finance 74*(1), 323–370.

Siriwardane, E. (2015, October). The Probability of Rare Disasters: Estimation and Implications. SSRN Scholarly Paper ID 2584047, Social Science Research Network, Rochester, NY.

Stilger, P. S., A. Kostakis, and S.-H. Poon (2017). What does risk-neutral skewness tell us about future stock returns? *Management Science 63*(6), 1814–1834.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 1040–1053.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty 5*(4), 297–323.

Vilkov, G. and Y. Xiao (2013, January). Option-Implied Information and Predictability of Extreme Returns. SSRN Scholarly Paper ID 2209654, Social Science Research Network, Rochester, NY.

Table 1: Quantile estimation simulation: 2019 market conditions

Panel A: Interpolate Option Price Chain with Cubic Splines																			
Number Sims	1000																		
RMSE BLQ Mean	34.162																		
RMSE MFQ Mean	7.835																		
Ratio of RMSE	4.360																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2536.590	2616.220	2710.371	2747.587	2760.765	2805.180	2832.558	2852.110	2901.357	2941.054	2980.663	3030.350	3052.947	3080.483	3119.759	3126.184	3130.422	3135.205	3135.279
S.E.	3.993	3.793	2.497	1.459	1.590	0.583	0.567	0.556	0.626	0.605	0.627	0.642	0.639	0.690	0.691	0.643	0.587	0.483	0.508
Qnt MF Mean	2698.824	2726.617	2744.016	2757.999	2767.899	2805.229	2830.792	2850.920	2898.925	2940.209	2981.630	3031.662	3053.918	3081.400	3122.686	3134.789	3148.158	3152.802	3154.113
S.E.	0.258	0.219	0.206	0.201	0.184	0.178	0.175	0.176	0.185	0.189	0.186	0.176	0.186	0.191	0.204	0.220	0.182	0.090	0.043
Panel B: Smooth Option Price Chain with Local Linear Regression using CV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	9.277																		
RMSE MFQ Mean	6.578																		
Ratio of RMSE	1.410																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2674.447	2722.674	2742.498	2756.544	2767.398	2804.764	2830.072	2850.594	2898.972	2940.099	2981.495	3031.916	3054.173	3082.148	3123.938	3137.191	3145.513	3148.514	3148.995
S.E.	2.289	0.765	0.431	0.181	0.177	0.161	0.150	0.156	0.162	0.176	0.175	0.172	0.160	0.168	0.246	0.245	0.224	0.220	0.237
Qnt MF Mean	2697.982	2725.655	2743.090	2756.901	2767.356	2804.565	2830.262	2850.419	2898.747	2939.999	2981.712	3032.429	3054.273	3082.064	3123.456	3136.038	3150.619	3154.309	3154.921
S.E.	0.221	0.173	0.156	0.150	0.135	0.120	0.121	0.128	0.130	0.139	0.135	0.125	0.132	0.140	0.155	0.182	0.150	0.049	0.015
Panel C: Smooth Option Price Chain with Local Quadratic Regression using CV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	19.045																		
RMSE MFQ Mean	7.375																		
Ratio of RMSE	2.582																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2592.930	2701.649	2735.831	2754.489	2766.159	2804.858	2830.325	2851.065	2899.676	2940.002	2981.700	3031.358	3053.742	3082.146	3123.970	3135.110	3142.533	3145.761	3146.316
S.E.	4.241	2.549	1.520	0.763	0.643	0.280	0.257	0.266	0.295	0.291	0.298	0.292	0.271	0.328	0.404	0.359	0.308	0.295	0.355
Qnt MF Mean	2698.476	2725.950	2743.349	2757.332	2767.764	2804.805	2830.457	2850.604	2898.824	2940.054	2981.587	3032.040	3054.127	3081.898	3123.108	3135.377	3149.426	3153.440	3154.491
S.E.	0.253	0.208	0.191	0.186	0.167	0.154	0.159	0.161	0.164	0.176	0.166	0.161	0.165	0.174	0.188	0.210	0.169	0.073	0.033
Panel D: Smooth Option Price Chain with Cubic B-splines using CV to Select Number of Interior Knots																			
Number Sims	1000																		
RMSE BLQ Mean	5.290																		
RMSE MFQ Mean	5.224																		
Ratio of RMSE	1.013																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2698.669	2726.233	2743.335	2756.698	2767.648	2804.736	2830.061	2850.465	2898.611	2940.082	2981.857	3032.359	3054.253	3082.027	3122.958	3135.904	3150.934	3154.357	3154.525
S.E.	0.142	0.072	0.054	0.033	0.038	0.031	0.029	0.036	0.044	0.056	0.047	0.034	0.040	0.046	0.066	0.101	0.134	0.080	0.072
Qnt MF Mean	2698.752	2726.233	2743.319	2756.712	2767.648	2804.736	2830.061	2850.450	2898.611	2940.049	2981.900	3032.369	3054.286	3081.998	3122.931	3135.880	3151.305	3154.832	3154.980
S.E.	0.086	0.072	0.051	0.036	0.038	0.031	0.029	0.036	0.044	0.047	0.041	0.034	0.039	0.045	0.054	0.102	0.129	0.026	0.006

This table reports results from a Monte Carlo simulation to assess the accuracy of model-free quantiles (MFQs) versus quantiles estimated with the approach of Breeden and Litzenberger (1978) (BLQs) recovered from a chain of noisy Black-Scholes options prices with measurement error simulated using the approach in Bondarenko (2003). We replicate the conditions of the near maturity chain of options a date halfway through our data corresponding to June 28, 2019. This results in 160 different strike prices ranging from a minimum strike price of 2310 to a maximum strike price of 3155. The strike interval between deep OTM put option quotes varies between 5 and 10 index points, with option quotes spaced at uniform intervals of 5 index points from the strike price of 2440 upwards. We simulate Black-Scholes model prices using the following inputs: spot price $S = 2942$, interest rate $r = 0.02$, dividend yield $q = 0.02$, and time-to-maturity $T = 0.08$. We use an ATM forward implied volatility of $\sigma = 0.13$. RMSEs are calculated using quantiles for $\alpha \in \{1\%, 2\%, \ldots, 99\%\}$.

Table 2: Quantile estimation simulation: 2019 market conditions with preprocessed convex option prices

Panel A: Interpolate Option Price Chain with Cubic Splines																			
Number Sims	1000																		
RMSE BLQ Mean	8.266																		
RMSE MFQ Mean	7.560																		
Ratio of RMSE	1.093																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2699.471	2726.571	2743.927	2757.567	2767.896	2804.849	2830.538	2850.762	2898.647	2940.072	2981.812	3031.996	3053.999	3081.639	3122.805	3135.324	3145.095	3146.882	3147.461
S.E.	0.265	0.208	0.204	0.190	0.179	0.165	0.164	0.168	0.175	0.176	0.181	0.177	0.175	0.186	0.201	0.220	0.218	0.231	0.243
Qnt MF Mean	2698.877	2726.120	2743.396	2757.324	2767.602	2804.967	2830.527	2850.588	2898.794	2940.096	2981.578	3032.041	3054.091	3081.898	3123.115	3135.888	3149.527	3153.518	3154.483
S.E.	0.257	0.213	0.200	0.191	0.174	0.167	0.165	0.173	0.171	0.181	0.176	0.167	0.176	0.183	0.194	0.223	0.168	0.072	0.032

Panel B: Smooth Option Price Chain with Local Linear Regression using CV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	6.115																		
RMSE MFQ Mean	6.093																		
Ratio of RMSE	1.004																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2698.458	2725.724	2743.146	2756.803	2767.262	2804.582	2830.196	2850.312	2898.675	2940.036	2981.872	3032.402	3054.378	3082.095	3123.498	3136.247	3150.059	3154.132	3154.730
S.E.	0.176	0.141	0.125	0.118	0.107	0.098	0.096	0.102	0.108	0.113	0.110	0.101	0.109	0.112	0.127	0.143	0.139	0.048	0.022
Qnt MF Mean	2698.458	2725.724	2743.146	2756.803	2767.262	2804.582	2830.196	2850.312	2898.675	2940.036	2981.872	3032.402	3054.378	3082.095	3123.498	3136.251	3150.264	3154.355	3154.953
S.E.	0.176	0.141	0.125	0.118	0.107	0.098	0.096	0.102	0.108	0.113	0.110	0.101	0.109	0.112	0.127	0.144	0.144	0.048	0.012

Panel C: Smooth Option Price Chain with Local Quadratic Regression using CV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	7.352																		
RMSE MFQ Mean	6.849																		
Ratio of RMSE	1.073																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2699.009	2725.962	2743.466	2757.216	2767.524	2804.830	2830.242	2850.479	2898.709	2939.947	2981.627	3032.107	3054.215	3081.668	3123.002	3136.027	3145.903	3149.014	3149.927
S.E.	0.229	0.183	0.164	0.156	0.139	0.134	0.134	0.136	0.143	0.153	0.147	0.141	0.146	0.154	0.169	0.186	0.177	0.201	0.214
Qnt MF Mean	2698.780	2725.974	2743.373	2757.193	2767.456	2804.777	2830.448	2850.472	2898.727	2940.093	2981.714	3032.182	3054.176	3081.924	3123.191	3136.081	3149.590	3153.638	3154.632
S.E.	0.222	0.181	0.166	0.158	0.142	0.135	0.132	0.140	0.142	0.149	0.145	0.136	0.145	0.150	0.165	0.188	0.155	0.067	0.028

Panel D: Smooth Option Price Chain with Cubic B-splines using CV to Select Number of Interior Knots																			
Number Sims	1000																		
RMSE BLQ Mean	7.217																		
RMSE MFQ Mean	6.881																		
Ratio of RMSE	1.049																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	2698.983	2726.171	2743.564	2756.720	2767.469	2804.686	2830.078	2850.424	2898.651	2940.012	2981.963	3032.416	3054.216	3081.867	3123.313	3135.490	3150.526	3170.626	3202.566
Qnt BL Mean	2698.803	2726.310	2743.383	2757.250	2767.581	2804.849	2830.354	2850.462	2898.698	2939.827	2981.804	3032.203	3054.294	3081.770	3123.326	3135.960	3147.494	3149.886	3150.559
S.E.	0.205	0.175	0.159	0.150	0.138	0.132	0.129	0.134	0.143	0.146	0.142	0.135	0.137	0.146	0.167	0.202	0.179	0.175	0.181
Qnt MF Mean	2698.886	2726.180	2743.278	2757.203	2767.503	2804.829	2830.403	2850.436	2898.641	2940.113	2981.825	3032.107	3054.093	3081.884	3123.314	3136.173	3149.743	3153.509	3154.469
S.E.	0.213	0.178	0.161	0.150	0.141	0.132	0.129	0.137	0.145	0.149	0.148	0.139	0.141	0.147	0.170	0.204	0.153	0.070	0.033

This table reports results from a Monte Carlo simulation to assess the accuracy of model-free quantiles (MFQs) versus quantiles estimated with the approach of Breeden and Litzenberger (1978) (BLQs) recovered from a chain of noisy Black-Scholes options prices with measurement error simulated using the approach in Bondarenko (2003). As in Table 1 we replicate the conditions of the near maturity chain of options at the end of our data corresponding to June 28, 2019. In this table, option prices are first preprocessed using a constrained least squares method to ensure the simulated noisy option prices are convex in strike price.

Table 3: Summary Statistics

Panel A: RNA-Q

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	Q_{1-10}	Q_{1-3}	Q_{4-7}	Q_{8-10}
count	1226	1226	1226	1226	1226	1226	1226	1226	1226	1226	1226	1226	1226	1226
mean	-1.344	-1.372	-1.306	-1.240	-1.198	-1.175	-1.161	-1.154	-1.151	-1.156	-1.226	-1.341	-1.194	-1.154
std	0.300	0.256	0.284	0.307	0.319	0.325	0.338	0.360	0.394	0.444	0.278	0.244	0.309	0.396
min	-2.175	-2.406	-2.759	-2.996	-3.163	-3.224	-2.984	-3.109	-3.268	-3.358	-2.931	-2.401	-3.092	-3.245
25%	-1.550	-1.534	-1.483	-1.435	-1.404	-1.396	-1.387	-1.400	-1.420	-1.459	-1.417	-1.503	-1.398	-1.424
50%	-1.375	-1.382	-1.312	-1.248	-1.205	-1.172	-1.149	-1.140	-1.138	-1.141	-1.226	-1.360	-1.204	-1.144
75%	-1.131	-1.199	-1.114	-1.033	-0.986	-0.949	-0.927	-0.893	-0.872	-0.843	-1.030	-1.168	-0.983	-0.873
max	-0.101	-0.272	-0.449	-0.207	-0.118	-0.123	-0.147	-0.091	0.052	0.225	-0.314	-0.274	-0.149	0.062
AR1	0.797	0.676	0.684	0.705	0.714	0.723	0.712	0.705	0.706	0.700	0.756	0.752	0.739	0.709

Panel B: Controls

	RNS-BKM	RNV-BKM ×100	LB-SVIX ×100	VWVS ×100	OWVS ×100	ID ×100	ERP ×100	RV ×100	VRP ×100	ETD ×100	LJV ×100	PD
count	1226	1226	1226	1226	1226	1226	1225	1226	1226	1226	1226	1222
mean	-1.692	0.389	0.348	9.826	10.062	1.787	0.482	0.253	0.136	-4.365	7.524	0.728
std	0.674	0.411	0.340	2.108	2.098	0.488	9.166	0.403	0.233	2.364	3.268	0.496
min	-4.881	0.070	0.063	1.785	2.962	0.000	-65.624	0.012	-2.997	-13.228	1.498	0.132
25%	-1.925	0.166	0.153	8.437	8.593	1.564	-3.945	0.072	0.070	-6.050	5.618	0.406
50%	-1.550	0.282	0.260	9.949	10.138	1.899	1.197	0.139	0.119	-4.310	6.772	0.569
75%	-1.247	0.456	0.413	11.316	11.507	2.107	5.533	0.261	0.201	-2.764	8.671	0.879
max	-0.628	4.786	3.512	16.453	16.350	2.850	40.797	4.529	1.549	7.567	30.872	3.130
AR1	0.938	0.948	0.952	0.742	0.760	0.988	-0.067	0.952	0.801	0.698	0.826	0.975

Panel A reports the summary statistics for the RN quantile-based skewness estimates for quantiles $\alpha \in \{1\%, \ldots, 10\%\}$ and the averaged RNA-Q estimates over ranges $\alpha \in \{1\%, 2\%, 3\%\}$, $\alpha \in \{4\%, 5\%, 6\%, 7\%\}$, $\alpha \in \{8\%, 9\%, 10\%\}$ and $\alpha \in \{1\%, \ldots, 10\%\}$. Panel B reports summary statistics for the controls used in predictive regressions. RNS-BKM and RNV-BKM are, respectively, the RN skewness and variance measures of Bakshi, Kapadia, and Madan (2003). LB-SVIX is the lower bound on the annualised 30-day ERP of Martin (2017). VWVS and OWVS are, respectively, volume and open interest weighted implied volatility (IV) spreads used in Atilgan, Bali, and Demirtas (2015), where the IV spread is the difference between out-of-the-money put IVs and at-the-money call IVs. ID is the implied dividend yield of the S&P 500 index extracted from OptionMetrics. ERP is the realized excess return on the S&P 500 index that we use as a proxy for the equity risk premium. RV is 22-day realized variance. VRP is the difference between one-month implied variance, RNV-BKM, and realized variance, RV. LJV is the left tail jump variation component of the VRP as in Bollerslev, Todorov, and Xu (2015) and PD is the risk-neutral probability of a 10% stock market drop over the next week (Andersen and Todorov 2019). ETD is a risk-neutral version of the extreme tail difference measure of Jiang, Wu, Zhou, and Zhu (2020) estimated using SP 500 index options.

Table 4: Correlation Matrix

	RNA-Q_{1-3}	RNA-Q_{4-7}	RNA-Q_{8-10}	RNS-BKM	RNV-BKM	LB-SVIX	VWVS	OWVS	ID	ERP-1w	ERP-c	RV	VRP	ETD	LJV	PD
RNA-Q_{1-10}	0.72***	0.96***	0.9***	0.15***	-0.26***	-0.24***	-0.42***	-0.41***	-0.14***	-0.05*	0.08***	-0.21***	-0.08***	0.72***	-0.28***	-0.32***
RNA-Q_{1-3}		0.64***	0.41***	0.38***	-0.04	-0.02	-0.55***	-0.5***	-0.12***	-0.01	-0.05*	-0.03	-0.01	0.58***	-0.33***	-0.14***
RNA-Q_{4-7}			0.82***	0.12***	-0.23***	-0.22***	-0.35***	-0.33***	-0.17***	-0.07**	0.06**	-0.2***	-0.07**	0.62***	-0.21***	-0.3***
RNA-Q_{8-10}				-0.0	-0.33***	-0.32***	-0.29***	-0.3***	-0.07**	-0.04	0.15***	-0.27***	-0.12***	0.69***	-0.24***	-0.35***
RNS-BKM					0.23***	0.26***	-0.65***	-0.63***	-0.06**	-0.0	-0.03	0.17***	0.11***	0.40***	-0.19***	0.28***
RNV-BKM						1.0***	0.05*	0.06**	-0.16***	0.02	-0.15***	0.84***	0.32***	-0.16***	0.71***	0.76***
LB							0.02	0.03	-0.17***	0.02	-0.15***	0.83***	0.33***	-0.14***	0.69***	0.78***
VWVS								0.91***	0.15***	0.03	-0.05*	0.03	0.04	-0.61***	0.41***	0.05*
OWVS									0.08***	0.02	-0.11***	0.01	0.08***	-0.60***	0.4***	0.03
ID										0.05*	0.05*	-0.22***	0.09***	-0.13***	-0.02	-0.11***
ERP-1w											-0.07**	-0.02	0.08***	-0.04	0.01	0.01
ERP-c												-0.06**	-0.16***	0.09***	-0.11***	-0.01
RV													-0.25***	-0.11***	0.58***	0.6***
VRP														-0.08***	0.24***	0.32***
ETD															-0.1***	-0.09***
LJV																0.52***

This table reports the correlations and associated significance levels for the averaged RNA-Q estimates and the control variables used in predictive regressions. The controls are defined in Table 3. ERP-c is the contemporaneous equity risk premium and ERP-1w is the one-week ahead equity risk premium.

Table 5: Predictive Regression Results

	RNA-Q_{1-10}					RNA-Q_{1-3}					RNA-Q_{4-7}					RNA-Q_{8-10}					RNS-BKM				
	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2
Panel A: Skewness values averaged over week																									
1W	-1.576	-0.439	-1.579	0.114	0.001	-0.348	-0.085	-0.281	0.779	-0.001	-1.961	-0.607	-2.233	0.026	0.004	-0.871	-0.345	-1.246	0.213	0.001	-0.029	-0.020	-0.087	0.931	-0.001
2W	-1.253	-0.349	-1.674	0.094	0.002	-0.149	-0.036	-0.148	0.882	-0.001	-1.653	-0.512	-2.493	0.013	0.006	-0.664	-0.263	-1.283	0.199	0.001	-0.055	-0.037	-0.190	0.849	-0.001
3W	-1.224	-0.341	-1.799	0.072	0.004	0.045	0.011	0.049	0.961	-0.001	-1.604	-0.496	-2.514	0.012	0.009	-0.729	-0.289	-1.578	0.114	0.002	-0.019	-0.013	-0.069	0.945	-0.001
4W	-1.315	-0.366	-2.042	0.041	0.006	-0.197	-0.048	-0.228	0.820	-0.001	-1.592	-0.492	-2.555	0.011	0.012	-0.799	-0.316	-1.807	0.071	0.005	-0.049	-0.033	-0.171	0.864	-0.001
5W	-1.401	-0.390	-2.241	0.025	0.009	-0.445	-0.108	-0.517	0.605	-0.000	-1.630	-0.504	-2.621	0.009	0.016	-0.816	-0.323	-1.953	0.051	0.006	-0.057	-0.038	-0.193	0.847	-0.001
6W	-1.444	-0.402	-2.383	0.017	0.012	-0.572	-0.139	-0.666	0.505	0.001	-1.612	-0.499	-2.622	0.009	0.019	-0.853	-0.338	-2.161	0.031	0.008	-0.082	-0.055	-0.273	0.785	-0.001
7W	-1.471	-0.410	-2.443	0.015	0.015	-0.747	-0.182	-0.877	0.380	0.002	-1.631	-0.505	-2.688	0.007	0.022	-0.815	-0.323	-2.120	0.034	0.009	-0.122	-0.082	-0.402	0.688	-0.000
8W	-1.441	-0.401	-2.395	0.017	0.016	-0.942	-0.229	-1.127	0.260	0.005	-1.610	-0.498	-2.652	0.008	0.026	-0.709	-0.281	-1.864	0.062	0.008	-0.174	-0.117	-0.571	0.568	0.001
9W	-1.359	-0.378	-2.241	0.025	0.017	-1.038	-0.253	-1.232	0.218	0.007	-1.469	-0.454	-2.430	0.015	0.024	-0.652	-0.258	-1.725	0.085	0.007	-0.204	-0.137	-0.653	0.514	0.001
10W	-1.287	-0.359	-2.129	0.033	0.017	-1.080	-0.263	-1.283	0.199	0.009	-1.380	-0.427	-2.284	0.022	0.024	-0.591	-0.234	-1.552	0.121	0.007	-0.205	-0.138	-0.647	0.517	0.002
11W	-1.165	-0.324	-1.900	0.057	0.015	-0.966	-0.235	-1.168	0.243	0.007	-1.271	-0.393	-2.065	0.039	0.022	-0.521	-0.206	-1.319	0.187	0.006	-0.202	-0.137	-0.636	0.525	0.002
12W	-1.084	-0.302	-1.707	0.088	0.014	-0.922	-0.225	-1.119	0.263	0.007	-1.164	-0.360	-1.835	0.067	0.020	-0.491	-0.194	-1.189	0.234	0.005	-0.208	-0.141	-0.647	0.517	0.002
Panel B: Skewness values at week end																									
1W	-1.169	-0.385	-1.450	0.147	0.001	-0.511	-0.139	-0.445	0.657	-0.001	-1.036	-0.386	-1.468	0.142	0.001	-0.794	-0.392	-1.453	0.146	0.001	0.067	0.047	0.209	0.835	-0.001
2W	-1.017	-0.335	-1.696	0.090	0.002	-0.427	-0.116	-0.448	0.654	-0.000	-1.117	-0.416	-2.126	0.033	0.004	-0.532	-0.263	-1.313	0.189	0.001	-0.040	-0.028	-0.141	0.888	-0.001
3W	-0.549	-0.181	-0.960	0.337	0.000	0.096	0.026	0.109	0.913	-0.001	-0.775	-0.289	-1.541	0.123	0.002	-0.254	-0.126	-0.653	0.514	-0.000	-0.012	-0.008	-0.043	0.966	-0.001
4W	-0.594	-0.196	-1.154	0.248	0.001	0.040	0.011	0.049	0.961	-0.001	-0.763	-0.285	-1.644	0.100	0.003	-0.314	-0.155	-0.922	0.357	0.000	-0.004	-0.003	-0.015	0.988	-0.001
5W	-0.783	-0.258	-1.610	0.107	0.004	-0.235	-0.064	-0.298	0.766	-0.001	-0.897	-0.334	-1.960	0.050	0.007	-0.410	-0.203	-1.293	0.196	0.002	-0.032	-0.022	-0.117	0.907	-0.001
6W	-0.882	-0.291	-1.903	0.057	0.006	-0.348	-0.094	-0.462	0.644	-0.000	-0.942	-0.351	-2.053	0.040	0.009	-0.488	-0.241	-1.657	0.097	0.004	-0.058	-0.040	-0.204	0.839	-0.001
7W	-0.873	-0.288	-2.018	0.044	0.007	-0.437	-0.118	-0.591	0.555	0.000	-0.907	-0.338	-2.103	0.035	0.010	-0.475	-0.235	-1.771	0.076	0.004	-0.076	-0.053	-0.269	0.788	-0.001
8W	-0.979	-0.322	-2.195	0.028	0.010	-0.593	-0.161	-0.810	0.418	0.002	-1.027	-0.383	-2.297	0.022	0.015	-0.494	-0.244	-1.840	0.066	0.006	-0.130	-0.090	-0.453	0.651	0.000
9W	-0.974	-0.321	-2.213	0.027	0.012	-0.760	-0.206	-1.044	0.296	0.004	-0.975	-0.363	-2.217	0.027	0.015	-0.476	-0.235	-1.829	0.067	0.006	-0.166	-0.115	-0.572	0.568	0.001
10W	-0.997	-0.329	-2.232	0.026	0.014	-0.814	-0.220	-1.125	0.261	0.006	-0.978	-0.364	-2.215	0.027	0.017	-0.492	-0.243	-1.865	0.062	0.007	-0.180	-0.125	-0.608	0.543	0.001
11W	-0.934	-0.308	-2.148	0.032	0.013	-0.834	-0.226	-1.190	0.234	0.007	-0.914	-0.341	-2.119	0.034	0.017	-0.440	-0.217	-1.623	0.105	0.006	-0.197	-0.137	-0.663	0.508	0.002
12W	-0.819	-0.270	-1.765	0.078	0.011	-0.732	-0.198	-1.059	0.289	0.006	-0.789	-0.294	-1.737	0.082	0.013	-0.396	-0.196	-1.392	0.164	0.005	-0.188	-0.131	-0.629	0.529	0.002

Table 5 reports predictive regression results for h-week ahead index returns using RNA-Q_{1-10}, RNA-Q_{1-3}, RNA-Q_{4-7}, RNA-Q_{8-10}, and RNS-BKM as individual predictors. The t-statistics are estimated using Newey-West standard error estimates allowing for a lag equal to two times the overlap of the dependent variable. The period is from January 1996 through June 2019, and the frequency is weekly. Panel A reports results where the predictor has been averaged over the previous week. Panel B reports results where the predictor is the most recently available predictor at the week end prior to the h-week ahead return horizon. Smoothing cubic B-splines are used to create a high resolution grid of option prices over the strike price domain.

Table 6: Predictive Regression Results: Controls

	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2
										Panel A										
			RNV-BKM					LB-SVIX					VWVS					OWVS		
1W	0.466	0.192	0.365	0.715	-0.000	0.603	0.205	0.395	0.693	-0.000	0.115	0.243	0.755	0.450	-0.000	0.096	0.202	0.665	0.506	-0.000
2W	0.069	0.028	0.073	0.942	-0.001	0.187	0.064	0.166	0.868	-0.001	0.101	0.214	0.847	0.397	0.000	0.074	0.154	0.654	0.513	-0.000
3W	-0.003	-0.001	-0.003	0.997	-0.001	0.106	0.036	0.108	0.914	-0.001	0.084	0.177	0.791	0.429	0.000	0.080	0.168	0.788	0.431	0.000
4W	0.089	0.037	0.119	0.905	-0.001	0.168	0.057	0.181	0.857	-0.001	0.113	0.238	1.069	0.285	0.002	0.119	0.249	1.185	0.236	0.003
5W	0.093	0.038	0.132	0.895	-0.001	0.155	0.053	0.177	0.859	-0.001	0.128	0.270	1.232	0.218	0.004	0.125	0.262	1.303	0.192	0.004
6W	0.207	0.085	0.316	0.752	-0.000	0.284	0.097	0.347	0.729	-0.000	0.138	0.292	1.339	0.181	0.006	0.129	0.270	1.388	0.165	0.005
7W	0.252	0.104	0.393	0.694	0.000	0.322	0.109	0.403	0.687	0.000	0.143	0.302	1.388	0.165	0.008	0.142	0.297	1.490	0.136	0.007
8W	0.236	0.097	0.363	0.716	0.000	0.294	0.100	0.366	0.714	0.000	0.157	0.331	1.568	0.117	0.011	0.159	0.334	1.686	0.092	0.011
9W	0.255	0.105	0.386	0.699	0.001	0.303	0.103	0.375	0.707	0.000	0.170	0.359	1.687	0.092	0.015	0.168	0.352	1.745	0.081	0.014
10W	0.240	0.099	0.361	0.718	0.000	0.280	0.095	0.344	0.731	0.000	0.175	0.369	1.782	0.075	0.018	0.171	0.359	1.854	0.064	0.017
11W	0.175	0.072	0.259	0.796	-0.000	0.208	0.071	0.252	0.801	-0.000	0.179	0.377	1.883	0.060	0.020	0.171	0.359	1.910	0.056	0.018
12W	0.134	0.055	0.197	0.844	-0.000	0.159	0.054	0.191	0.848	-0.000	0.183	0.386	1.931	0.054	0.023	0.174	0.366	1.927	0.054	0.021
										Panel B										
			ID					RV					ERP-lag					VRP		
1W	0.922	0.450	1.344	0.179	0.002	-0.520	-0.210	-0.371	0.711	-0.000	-0.067	-0.613	-1.496	0.135	0.004	3.025	0.705	1.553	0.120	0.005
2W	0.963	0.470	1.701	0.089	0.005	-0.555	-0.224	-0.569	0.569	0.000	-0.040	-0.365	-1.531	0.126	0.003	1.899	0.443	1.367	0.172	0.004
3W	0.981	0.479	1.810	0.070	0.008	-0.492	-0.199	-0.688	0.492	0.001	-0.023	-0.212	-1.271	0.204	0.001	1.491	0.348	1.550	0.121	0.004
4W	0.981	0.479	1.867	0.062	0.011	-0.394	-0.159	-0.613	0.540	0.001	-0.014	-0.132	-0.822	0.411	0.000	1.485	0.346	1.551	0.121	0.006
5W	0.996	0.487	1.966	0.049	0.015	-0.282	-0.114	-0.502	0.616	0.000	-0.005	-0.042	-0.232	0.816	-0.001	1.157	0.270	1.374	0.169	0.004
6W	1.019	0.498	2.092	0.036	0.019	-0.175	-0.070	-0.355	0.722	-0.000	0.001	0.009	0.055	0.956	-0.001	1.185	0.276	1.478	0.140	0.005
7W	1.061	0.518	2.263	0.024	0.024	-0.184	-0.074	-0.392	0.695	-0.000	-0.016	-0.149	-1.305	0.192	0.001	1.350	0.315	1.769	0.077	0.008
8W	1.108	0.541	2.430	0.015	0.030	-0.188	-0.076	-0.404	0.686	-0.000	-0.011	-0.100	-0.970	0.332	0.000	1.311	0.306	1.974	0.048	0.009
9W	1.144	0.559	2.562	0.010	0.037	-0.209	-0.084	-0.450	0.653	0.000	-0.007	-0.065	-0.598	0.550	-0.000	1.427	0.333	2.280	0.023	0.013
10W	1.173	0.573	2.665	0.008	0.043	-0.255	-0.103	-0.533	0.594	0.001	-0.008	-0.075	-0.795	0.427	-0.000	1.519	0.354	2.351	0.019	0.016
11W	1.191	0.581	2.748	0.006	0.050	-0.270	-0.109	-0.547	0.584	0.001	-0.004	-0.035	-0.408	0.683	-0.001	1.364	0.318	2.039	0.041	0.014
12W	1.197	0.584	2.807	0.005	0.055	-0.259	-0.105	-0.507	0.612	0.001	-0.007	-0.068	-0.907	0.364	-0.000	1.204	0.281	1.716	0.086	0.012
										Panel B										
			ETD					LJV					PD							
1W	-0.174	-0.413	-1.558	0.119	0.001	0.022	0.073	0.176	0.860	-0.001	0.214	0.106	0.276	0.782	-0.001					
2W	-0.126	-0.299	-1.534	0.125	0.001	-0.020	-0.066	-0.208	0.835	-0.001	0.279	0.139	0.480	0.631	-0.000					
3W	-0.129	-0.306	-1.739	0.082	0.003	-0.027	-0.087	-0.316	0.752	-0.001	0.457	0.227	0.896	0.370	0.001					
4W	-0.142	-0.335	-1.944	0.052	0.005	-0.016	-0.053	-0.212	0.832	-0.001	0.531	0.264	1.094	0.274	0.003					
5W	-0.140	-0.332	-1.969	0.049	0.007	-0.008	-0.027	-0.116	0.908	-0.001	0.594	0.295	1.257	0.209	0.005					
6W	-0.149	-0.353	-2.172	0.030	0.009	0.008	0.026	0.116	0.907	-0.001	0.633	0.314	1.358	0.175	0.007					
7W	-0.142	-0.336	-2.096	0.036	0.009	0.028	0.092	0.428	0.669	-0.000	0.616	0.306	1.324	0.185	0.008					
8W	-0.141	-0.334	-2.212	0.027	0.011	0.045	0.148	0.709	0.478	0.001	0.602	0.299	1.294	0.196	0.009					
9W	-0.145	-0.343	-2.319	0.020	0.013	0.064	0.208	0.992	0.321	0.004	0.559	0.277	1.207	0.227	0.008					
10W	-0.137	-0.324	-2.241	0.025	0.013	0.070	0.227	1.079	0.281	0.006	0.495	0.246	1.064	0.288	0.007					
11W	-0.131	-0.309	-2.197	0.028	0.013	0.067	0.219	1.040	0.298	0.006	0.457	0.227	0.991	0.322	0.007					
12W	-0.133	-0.314	-2.221	0.026	0.015	0.066	0.216	1.050	0.294	0.007	0.439	0.218	0.986	0.324	0.007					

Table 6 reports predictive regression results for h-week ahead index returns using the controls as individual predictors. The t-statistics are estimated using Newey-West standard error estimates allowing for a lag equal to two times the overlap of the dependent variable. The period is from January 1996 through June 2019, and the frequency is weekly. The controls are defined in Table 3. Smoothing cubic B-splines are used to create a high resolution grid of option prices over the strike price domain to estimate option implied predictors such as RNV-BKM and LB-SVIX.

Table 7: Bivariate Predictive Regressions

	RNS-BKM	RNV-BKM	LB-SVIX	VWVS	OWVS	ID	RV	ERP-lag	VRP	ETD	LJV	PD
Panel A: t-statistic on RNA-Q_{4-7} coefficient												
1W	-2.151	-2.067	-2.052	-2.035	-2.063	-1.857	-2.400	-2.041	-1.968	-1.666	-2.254	-2.176
2W	-2.397	-2.417	-2.358	-2.293	-2.322	-2.045	-2.715	-2.213	-2.203	-2.046	-2.648	-2.324
3W	-2.447	-2.448	-2.390	-2.333	-2.250	-2.121	-2.717	-2.295	-2.272	-1.933	-2.651	-2.156
4W	-2.488	-2.432	-2.401	-2.185	-2.073	-2.192	-2.705	-2.395	-2.325	-1.725	-2.603	-2.155
5W	-2.567	-2.529	-2.504	-2.195	-2.134	-2.268	-2.750	-2.535	-2.453	-1.794	-2.644	-2.244
6W	-2.582	-2.494	-2.475	-2.187	-2.158	-2.285	-2.710	-2.572	-2.479	-1.721	-2.583	-2.256
7W	-2.660	-2.521	-2.510	-2.296	-2.218	-2.342	-2.753	-2.576	-2.543	-1.949	-2.576	-2.331
8W	-2.618	-2.503	-2.495	-2.282	-2.175	-2.296	-2.726	-2.574	-2.522	-2.002	-2.504	-2.328
9W	-2.383	-2.292	-2.290	-1.958	-1.903	-2.040	-2.530	-2.371	-2.295	-1.689	-2.235	-2.150
10W	-2.229	-2.176	-2.176	-1.780	-1.739	-1.876	-2.421	-2.225	-2.145	-1.589	-2.073	-2.066
11W	-2.002	-2.012	-2.008	-1.528	-1.516	-1.662	-2.233	-2.034	-1.941	-1.395	-1.871	-1.880
12W	-1.769	-1.806	-1.802	-1.277	-1.273	-1.447	-2.002	-1.791	-1.726	-1.118	-1.649	-1.655
Panel B: t-statistic on control coefficient												
	RNS-BKM	RNV-BKM	LB-SVIX	VWVS	OWVS	ID	RV	ERP-lag	VRP	ETD	LJV	PD
1W	0.184	0.137	0.181	0.177	0.131	0.932	-0.581	-1.463	1.486	-0.258	-0.133	-0.197
2W	0.069	-0.191	-0.079	0.236	0.084	1.310	-0.816	-1.468	1.282	0.044	-0.565	-0.037
3W	0.202	-0.311	-0.175	0.075	0.150	1.478	-1.028	-1.160	1.433	-0.065	-0.711	0.338
4W	0.107	-0.221	-0.127	0.353	0.525	1.572	-0.997	-0.684	1.432	-0.283	-0.632	0.518
5W	0.093	-0.252	-0.168	0.489	0.565	1.673	-0.957	-0.092	1.232	-0.184	-0.574	0.663
6W	0.010	-0.099	-0.028	0.605	0.617	1.817	-0.877	0.226	1.334	-0.358	-0.357	0.772
7W	-0.117	-0.038	0.011	0.653	0.743	1.983	-0.944	-1.069	1.644	-0.203	-0.073	0.720
8W	-0.299	-0.061	-0.023	0.868	0.975	2.164	-0.964	-0.706	1.883	-0.245	0.197	0.699
9W	-0.419	0.007	0.024	1.086	1.132	2.308	-0.970	-0.372	2.240	-0.621	0.534	0.668
10W	-0.429	0.005	0.013	1.219	1.258	2.441	-1.022	-0.540	2.332	-0.608	0.657	0.555
11W	-0.435	-0.069	-0.053	1.363	1.347	2.553	-0.986	-0.140	2.006	-0.690	0.655	0.521
12W	-0.468	-0.104	-0.089	1.484	1.416	2.643	-0.895	-0.626	1.670	-0.979	0.695	0.547

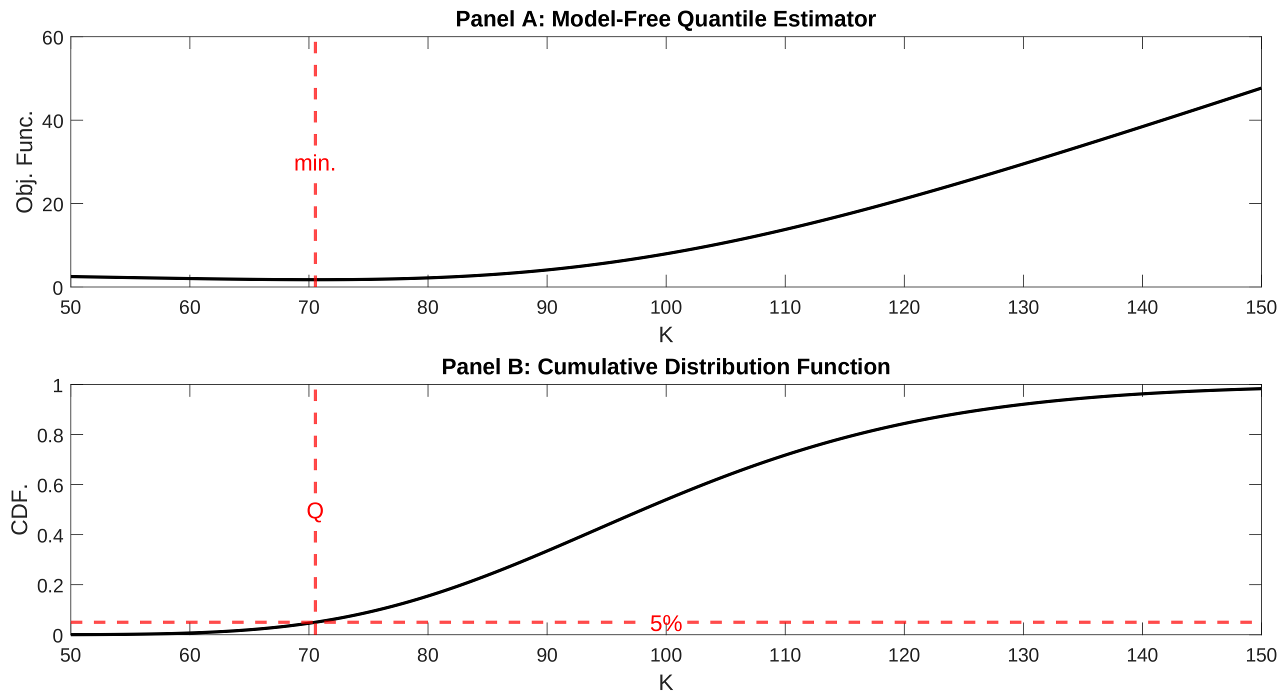
This table reports predictive regression results for h-week ahead excess index returns using RNA-Q and a single control in a bivariate regression. The period is from January 1996 through June 2019, and the frequency is weekly. The t-statistics are estimated using Newey-West standard error estimates allowing for a lag equal to two times the overlap of the dependent variable. Panel A reports the t-statistics for RNA-Q_{4-7} predictor when controls are used in bivariate predictive regressions. Panel B reports t-statistics for the controls. The controls are defined in Table 3.

Table 8: Out-Of-Sample Predictive Regressions

	RNA-Q			ID			VRP			ETD		
Week	OS $\bar{R}^2$	CT OS $\bar{R}^2$	CW t-stat	OS $\bar{R}^2$	CT OS $\bar{R}^2$	CW t-stat	OS $\bar{R}^2$	CT OS $\bar{R}^2$	CW t-stat	OS $\bar{R}^2$	CT OS $\bar{R}^2$	CW t-stat
1W	0.001	0.003	1.746**	-0.002	0.001	0.124	-0.001	0.002	0.620	-0.001	0.000	0.411
2W	0.003	0.003	2.068**	0.002	0.005	1.026	-0.002	0.002	0.464	-0.000	0.000	0.690
3W	0.007	0.005	2.045**	0.005	0.009	1.470*	0.000	0.001	1.033	0.001	0.003	1.199
4W	0.011	0.009	2.016**	0.009	0.011	1.670**	0.002	0.002	1.122	0.004	0.006	1.712**
5W	0.016	0.010	2.093**	0.013	0.013	1.830**	0.000	0.001	0.919	0.006	0.008	1.919**
6W	0.019	0.011	2.060**	0.017	0.016	1.947**	0.002	0.003	1.255	0.008	0.010	2.285**
7W	0.023	0.011	2.094**	0.021	0.020	2.075**	0.005	0.005	1.719**	0.009	0.011	2.271**
8W	0.026	0.011	2.081**	0.029	0.024	2.194**	0.007	0.008	2.074**	0.010	0.012	2.329**
9W	0.024	0.007	1.947**	0.035	0.027	2.268**	0.011	0.012	2.348***	0.012	0.014	2.518***
10W	0.024	0.006	1.879**	0.042	0.031	2.313**	0.015	0.015	2.522***	0.012	0.013	2.369***
11W	0.022	0.002	1.729**	0.048	0.033	2.321**	0.012	0.014	2.443***	0.012	0.014	2.259**
12W	0.020	-0.001	1.545*	0.053	0.035	2.310**	0.009	0.013	2.241**	0.013	0.015	2.268**

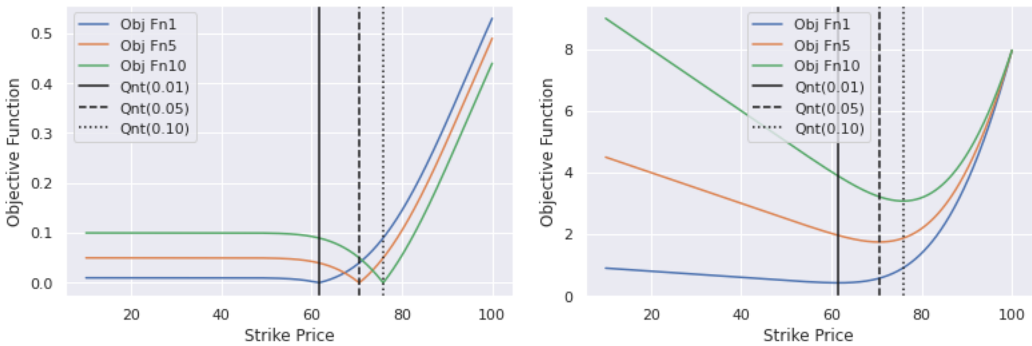
This Table reports OS adjusted R^2s (OS $\bar{R}^2$) and OS adjusted R^2s from restricted regressions, as in Campbell and Thompson (2008) (CT OS $\bar{R}^2$), where the coefficient on the predictor in the regression is restricted to be non-negative and, if the forecast is still negative, a second restriction is applied replacing the forecast with 0. Forecasts are made on a recursive basis with an initial estimation window set equal to 120 weeks, where the full sample size is 1,225 weeks. The period is from January 1996 through June 2019, and the frequency is weekly. The OS significance is assessed using a one-sided t-test on the mean squared prediction error (MSPE)-adjusted statistic in Clark and West (2007). We apply this test taking into account the autocorrelation in forecast errors by using Newey-West standard error estimates with lags set equal to 2 times the overlaps in calculating t-statistics. One-sided t-statistics for the MSPE-adjusted, for the null of equal MSPE, are reported in the table. For large enough sample sizes, as applies in this case, standard normal critical values can be used and we can reject the null at 10%/5%/1% if the t-statistic $> 1.282/1.645/2.33$, respectively.

Figure 1: Model-Free Quantiles: Black-Scholes Model



Panel A depicts the objective function of the model-free quantile at $\alpha = 5\%$ and the solution that minimizes the objective function in Eq. (1). In Panel B, the risk-neutral cumulative distribution function (RN-CDF) of the underlying price is depicted. To simulate option prices and depict the RN-CDF, we use the following parameters: $S = 100$, $r = 0$, $q = 0$, $T = 1$, $\sigma = 0.2$, assuming the underlying price follows geometric Brownian motion.

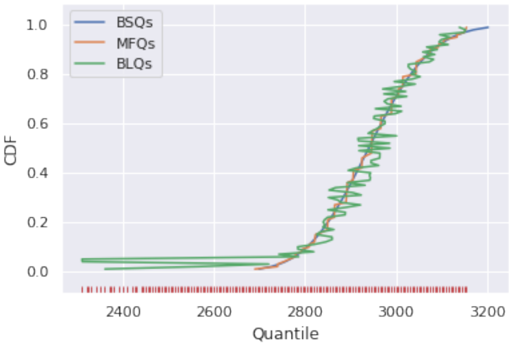
Figure 2: Objective Functions



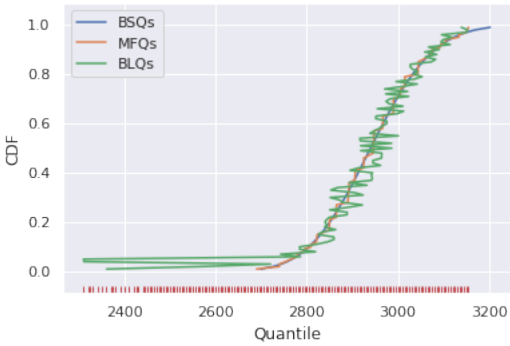
(a) BLQ (b) MFQ

This plot depicts the objective functions for the Breeden-Litzenberg quantile (BLQ) method and the model-free quantile (MFQ) method along with the estimated quantiles at 1%, 5% and 10%. BLQs and MFQs are estimated using Black-Scholes option prices. The spot price is $S = 100$, the interest rate is $r = 0$, the dividend yield is $q = 0$, the time-to-maturity is $T = 1$, and volatility is $\sigma = 0.10$. The strike prices range from 10 to 100 in steps of 0.01 index points.

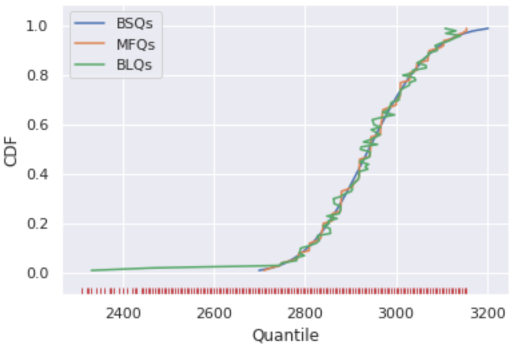
Figure 3: Quantiles with an Option Price Chain Interpolated by Cubic Splines



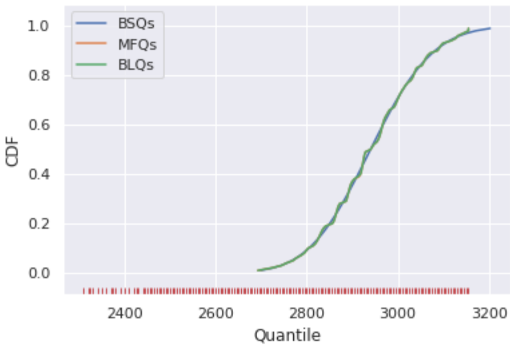
(a) BLQ Highest RMSE: Splines



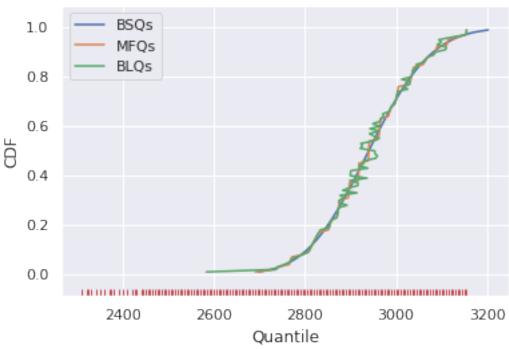
(b) BLQ Typical RMSE: Splines



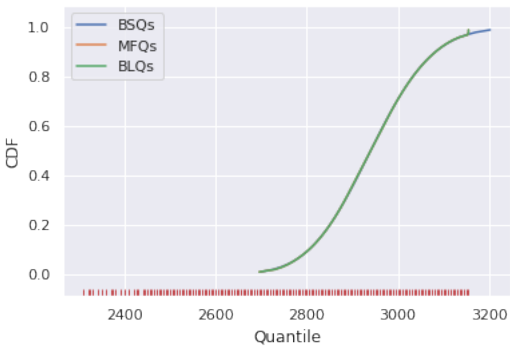
(c) BLQ Highest RMSE: Local Linear Regression



(d) BLQ Typical RMSE: Local Linear Regression



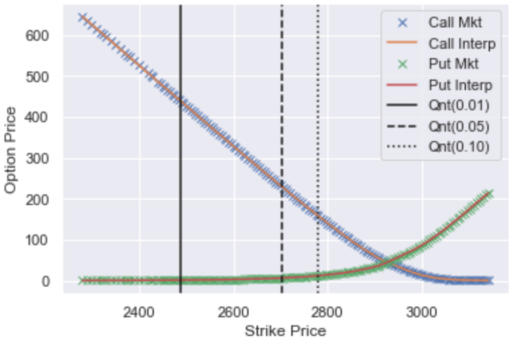
(e) BLQ Highest RMSE: Smoothing B-splines



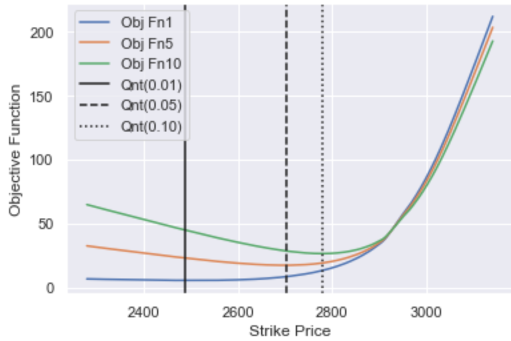
(f) BLQ Typical RMSE: Smoothing B-splines

This plot depicts the risk-neutral cumulative distribution function (RN-CDF) of the underlying optioned asset price where quantiles are estimated using the model-free method and the Breeden-Litzenberg method. BSQs are Black-Scholes quantiles calculated from the underlying log normal distribution with no noise. BLQs are quantiles estimated with the Breeden-Litzenberg method and MFQs are quantiles estimated with the model-free method. BLQs and MFQs use Black-Scholes option prices with added noise to simulate measurement error. The simulated option prices are available at discrete strike intervals depicted by the rug plot. The discrete option price chain is interpolated with cubic splines in panels (a) and (b), smoothed with local linear regression in panels (c) and (d) and smoothed with cubic B-splines in panels (e) and (f). Option prices are simulated to replicate the conditions in our data sample for S&P 500 index options on June 28, 2019. Option prices are available at 160 different strike prices ranging from a minimum strike price of 2310 to a maximum strike price of 3155 and spaced at uniform intervals of 5 index points, with the exception of some deep OTM options that are spaced at 10 index points. The spot price is $S = 2942$, the interest rate is $r = 0.02$, the dividend yield is $q = 0.02$, the time-to-maturity is $T = 0.08$, and volatility is $\sigma = 0.13$.

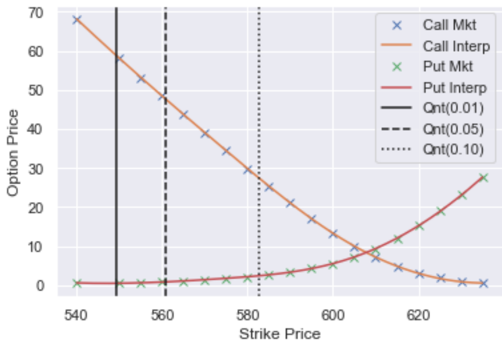
Figure 4: Risk-Neutral Quantile Estimation



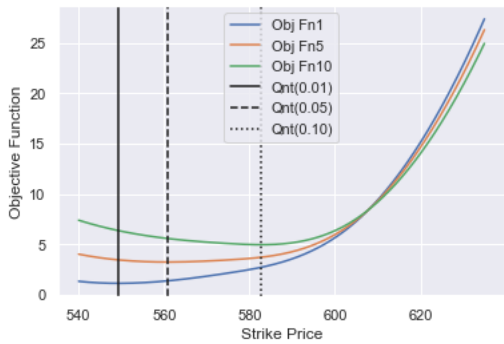
(a) Interpolated Option Prices 2019



(b) Objective Function $\alpha = 0.01, 0.05$ & 0.10



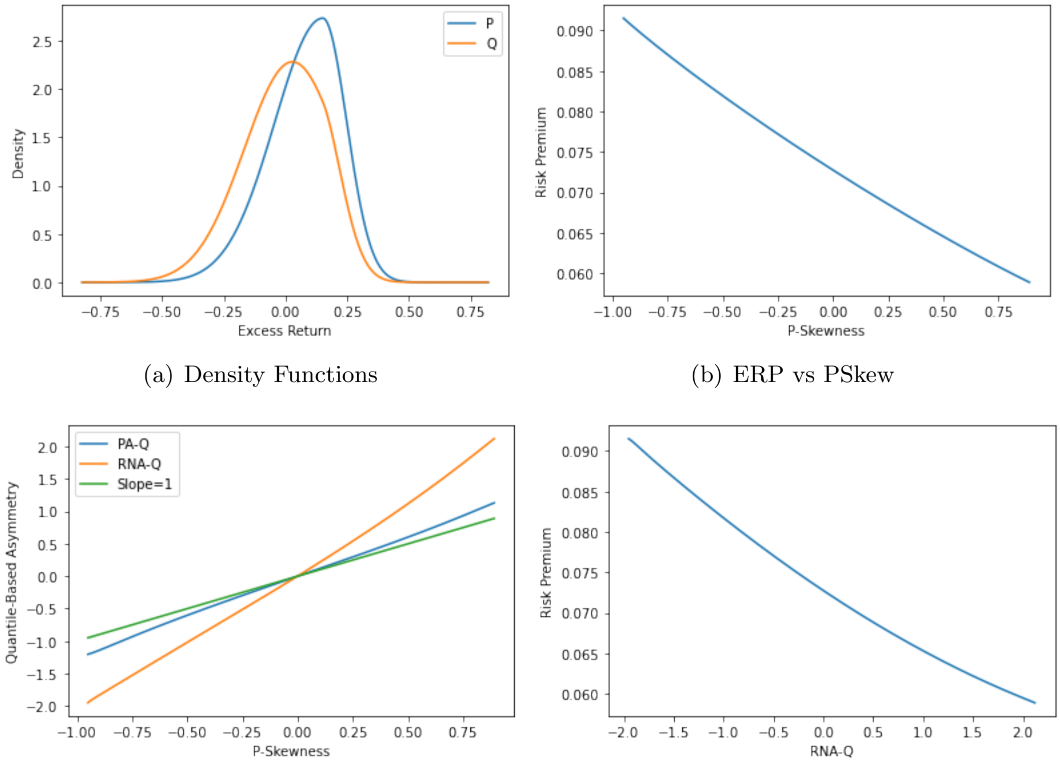
(c) Interpolated Option Prices 1996



(d) Objective Function $\alpha = 0.01, 0.05$ & 0.10

This plot depicts the process used to construct the objective function and estimate risk-neutral quantiles. The top left panel depicts mid-quote prices for call and put options along with the B-spline interpolated option prices using S&P 500 index options on June 28, 2019 with 28 days to maturity. The top right panel panel depicts the objective function used to estimate model-free quantiles at $\alpha = 1\%$, 5% and 10%, where the vertical lines on both the top left and right panel indicates these quantile. The bottom left and right panels depicts the same plots but using S&P 500 index options on January 17, 1996 with 29 days to maturity.

Figure 5: Smooth Half Normal CARA Economy



(a) Density Functions

(b) ERP vs PSkew

(c) RNA-Q vs PSkew

(d) ERP vs RNA-Q

These plots summarize various aspects of the Smooth Half Normal CARA economy. We set $\mu = 7.28\%$, $\sigma = 14.98\%$ and skewness varies from -0.95 to +0.8. Panel (a) depicts the physical and risk-neutral density functions for a single skewness value of -0.5. Panel (b) depicts the ERP versus physical skewness where skewness varies from -0.95 to 0.8. Panel (c) shows the variation of RNA-Q and quantile-based physical asymmetry (PA-Q) with physical skewness and panel (d) plots the ERP versus RNA-Q.

Appendix A Monte Carlo Experiments

A.1 Option Prices with Simulated Measurement Error

We provide further detail on the method used to simulate measurement error in option prices due to Bondarenko (2003). This approach sets the maximum strike dependent spreads, s, according to the CBOE rule book and generates uniformly distributed measurement error on $[-0.5s, 0.5s]$ where $s = q^a - q^b$ is the bid-ask spread and where (q^b, q^a) are the concurrent bid and ask-quotes for the option price q. The value of the spread depends on the moneyness of the option price quote q. Assuming s is proportional to the maximum spread permitted by the exchange, Bondarenko (2003) constructs a function $M(q)$ to represent the maximum spread for the quote q. Specifically,

$$
\begin{aligned}
&M(0) = \frac{1}{8}, \quad M(2) = \frac{1}{4}, \quad M(5) = \frac{3}{8}, \\
&M(10) = \frac{1}{2}, \quad M(20) = \frac{3}{4}, \\
&M(q) = 1, q \geq 50
\end{aligned}
$$

Furthermore, $M(q)$ is linearly interpolated for all $q \in [0, 50]$. As in Bondarenko (2003), we simulate a chain of call and put option prices and use put-call parity to convert ITM call (put) option prices to OTM put (call) option prices. Then we add measurement error to OTM option prices only. This results in measurement error that is smaller in absolute terms and larger in relative terms for far OTM option prices.

A.2 Further Monte Carlo Results

In this subsection of Appendix A we report further Monte Carlo simulation results on the performance of MFQs versus BLQs. Table A1 repeats the simulations in Table 1 using a two-state mixture lognormal model denoted as LN2 (see, e.g., Bahra 1997 and Melick and Thomas 1997) to generate option prices that are a closer match to market prices. The LN2 model consists of five parameters: p is the probability of being in state 1; μ_1 and μ_2 are the instantaneous drift of the asset price in states 1 and 2; and σ_1 and σ_2 are the instantaneous volatility of the asset price in states 1 and 2. Furthermore, the parameters should satisfy the following constraint so that the forward price from the LN2 model is equal to the market forward price: $pe^{\mu_1 T}+(1-p)e^{\mu_2 T} = e^{(r-q)T}$. We fit the LN2 model to the set of OTM call and put market prices available on June 28, 2019, imposing the above constraint. This results in the following set of parameters for the LN2 model: $p = 0.8702, \mu_1 = 0.1286, \mu_2 = -0.8601, \sigma_1 = 0.0929, \sigma_2 = 0.2278$. Furthermore, we simulate the model with the following set of inputs: $S = 2941.76, r = 0.0239, q = 0.0194$, and $T = 0.0767$[23]. The results in Table A1 are quantitatively and qualitatively similar to the results from Table 1. The MFQ method significantly outperforms the BLQ method when cubic splines are used to interpolate noisy option price data (Panel A) or when a LLR smoothing is applied as in Panel B where leave-one-out LSCV is used to select the bandwidth. The MFQ method also outperforms the BLQ method in Panels C when LQR smoothing is used and Panel D when cubic B-splines are used to smooth option prices.

We also run a Monte Carlo experiment using the LN2 where noisy option prices are preprocessed to be convex in the strike price with results presented in Table A2. MFQs are more accurate than BLQs in the case of cubic splines, LLR, and LQR. However, we find that BLQs are slightly more accurate than MFQs in terms of mean RMSE in the case of cubic B-spline smoothing. Comparing mean RMSE values with values in Panel D of Table A1, when cubic B-splines are used to smooth option prices that are not preprocessed to be convex, we find that preprocessing combined with cubic B-spline smoothing results in mean RMSE values that are approximately four times greater relative to the case when cubic B-spline smoothing is used without preprocessing. In the case of cubic B-splines, the experimental evidence suggests that preprocessing followed

[23]Unlike in the Black-Scholes case, we do not round down the inputs so that the constraint imposed on the fitted parameters ensures no-arbitrage.

by cubic B-spline smoothing results in less accurate quantiles when compared to simply using cubic B-splines to smooth option prices. In this latter case, the MFQ method is more accurate than the BLQ method.

In Table A3 we simulate S&P 500 index option market conditions on the July 30, 2007, for the near maturity option chain. We choose this date as it lies approximately half way through our data sample. In this case we replicate conditions where a medium number of option price quotes are available over a lower range of strike prices than in the previous simulations. On this day, there were 52 different strike prices ranging from a minimum strike price of 1290 to a maximum strike price of 1560. The left most four options are spaced at intervals of 10 index points whereas all other options are spaced at uniform intervals of 5 index points. The spot price is $S = 1474$, the interest rate is $r = 0.05$, the dividend yield is $q = 0.02$, and the time-to-maturity is $T = 0.05$. We use an ATM forward implied volatility of $\sigma = 0.200$ to simulate Black-Scholes model prices. In all cases, the MFQ method is more accurate than the BLQ method.

In Table A4 we simulate S&P 500 index option market conditions on the July 30, 1999, for the near maturity option chain. We choose this date as it lies near the beginning of our data sample. In this case we replicate conditions where a low number of option price quotes are available over a short range of strike prices than in the previous simulations. On this day, there were 30 different strike prices ranging from a minimum strike price of 1125 to a maximum strike price of 1425. The left most eight option quotes are spaced at intervals of 25 index points, then there is a gap of 20 index points to the 9th option quote with the remaining option quotes are spaced at uniform intervals of 5 index points. The spot price is $S = 1329$, the interest rate is $r = 0.05$, the dividend yield is $q = 0.01$, the time-to-maturity is $T = 0.05$, and we use a single ATM forward implied volatility of $\sigma = 0.22$ to simulate the Black-Scholes model prices. In all cases, the MFQ method is more accurate than the BLQ method.

Table A1: Quantile estimation simulation with LN2 model: 2019 market conditions

Panel A: Interpolate Option Price Chain with Cubic Splines																			
Number Sims	1000																		
RMSE BLQ Mean	35.728																		
RMSE MFQ Mean	6.385																		
Ratio of RMSE	5.595																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2492.023	2567.009	2608.737	2655.536	2685.203	2775.874	2834.571	2869.071	2923.423	2958.076	2989.738	3027.625	3045.045	3065.784	3097.817	3109.239	3118.434	3126.546	3133.398
S.E.	3.731	3.486	3.158	2.894	2.490	1.920	1.420	0.951	0.579	0.480	0.443	0.434	0.400	0.423	0.587	0.615	0.617	0.570	0.490
Qnt MF Mean	2513.784	2580.044	2625.899	2665.212	2699.717	2815.148	2856.269	2879.843	2925.502	2958.660	2991.407	3028.747	3045.119	3065.312	3095.693	3104.698	3115.404	3130.264	3149.479
S.E.	0.371	0.353	0.347	0.345	0.366	0.292	0.231	0.209	0.178	0.168	0.164	0.156	0.153	0.154	0.178	0.177	0.191	0.203	0.173

Panel B: Smooth Option Price Chain with Local Linear Regression using CV Bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	8.967																		
RMSE MFQ Mean	4.898																		
Ratio of RMSE	1.831																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2502.201	2577.368	2625.567	2666.212	2700.675	2811.217	2854.083	2878.982	2924.809	2959.097	2991.414	3029.387	3045.604	3066.116	3096.429	3105.383	3116.979	3133.517	3146.783
S.E.	1.330	0.915	0.828	0.889	0.874	0.525	0.281	0.213	0.154	0.121	0.110	0.097	0.105	0.099	0.122	0.140	0.179	0.254	0.191
Qnt MF Mean	2512.011	2577.762	2624.648	2663.193	2698.313	2814.358	2855.650	2879.099	2924.665	2959.274	2991.603	3029.391	3045.639	3066.018	3096.357	3105.124	3116.385	3131.266	3151.753
S.E.	0.357	0.327	0.316	0.338	0.340	0.255	0.190	0.163	0.130	0.108	0.101	0.090	0.091	0.097	0.110	0.118	0.141	0.155	0.135

Panel C: Smooth Option Price Chain with Local Quadratic Regression using CV Bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	2.267																		
RMSE MFQ Mean	2.109																		
Ratio of RMSE	1.075																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2512.045	2577.009	2623.933	2662.977	2697.775	2814.563	2855.545	2879.503	2924.952	2959.022	2991.670	3029.464	3045.536	3065.887	3096.712	3105.678	3116.339	3129.663	3147.269
S.E.	0.337	0.381	0.316	0.279	0.275	0.176	0.100	0.061	0.050	0.039	0.035	0.028	0.030	0.030	0.069	0.051	0.057	0.064	0.128
Qnt MF Mean	2511.909	2577.595	2623.931	2663.076	2698.085	2814.694	2855.707	2879.586	2925.002	2959.014	2991.689	3029.488	3045.545	3065.879	3096.658	3105.662	3116.299	3129.686	3147.621
S.E.	0.197	0.169	0.158	0.170	0.180	0.109	0.068	0.057	0.040	0.037	0.030	0.025	0.026	0.027	0.033	0.036	0.041	0.054	0.126

Panel D: Smooth Option Price Chain with Cubic B-splines using CV to Select Number of Interior Knots																			
Number Sims	1000																		
RMSE BLQ Mean	1.684																		
RMSE MFQ Mean	1.591																		
Ratio of RMSE	1.059																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2511.948	2577.413	2624.094	2663.125	2698.063	2814.456	2855.727	2879.572	2925.095	2958.991	2991.682	3029.319	3045.294	3065.510	3096.503	3105.410	3116.114	3130.950	3150.344
S.E.	0.143	0.148	0.144	0.194	0.164	0.108	0.069	0.054	0.042	0.036	0.030	0.023	0.024	0.028	0.042	0.050	0.055	0.122	0.149
Qnt MF Mean	2511.985	2577.382	2624.082	2662.929	2698.053	2814.459	2855.738	2879.598	2925.101	2959.009	2991.682	3029.319	3045.284	3065.510	3096.520	3105.419	3116.114	3130.687	3152.286
S.E.	0.138	0.120	0.114	0.130	0.135	0.110	0.068	0.050	0.042	0.037	0.030	0.023	0.025	0.028	0.041	0.050	0.055	0.111	0.125

This table reports results from a Monte Carlo simulation to assess the accuracy of model-free quantiles (MFQs) versus quantiles estimated with the approach of Breeden and Litzenberger (1978) (BLQs) recovered from a chain of noisy options prices where the measurement error noise is simulated using the approach in Bondarenko (2003). We replicate the conditions of the near maturity chain of options on a date at the end of our data corresponding to June 28, 2019. This results in 160 different strike prices ranging from a minimum strike price of 2310 to a maximum strike price of 3155. The strike interval between deep OTM put option quotes varies between 5 and 10 index points, with option quotes spaced at uniform intervals of 5 index points from the strike price of 2440 upwards. Option prices are simulated using a two-state lognormal mixture model (LN2) with the following inputs: spot price $S = 2941.76$, interest rate $r = 0.0239$, dividend yield $q = 0.0194$, and time-to-maturity $T = 0.0767$. We use the following parameters in the LN2 model: $p = 0.8702, \mu1 = 0.1286, \mu2 = -0.8601, \sigma1 = 0.0929, \sigma2 = 0.2278$. RMSEs are calculated using quantiles for $\alpha \in \{1\%, 2\%, \ldots, 99\%\}$.

Table A2: Quantile estimation simulation with LN2 model: 2019 market conditions with preprocessed convex option prices

Panel A: Interpolate Option Price Chain with Cubic Splines																			
Number Sims	1000																		
RMSE BLQ Mean	6.488																		
RMSE MFQ Mean	6.234																		
Ratio of RMSE	1.041																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2513.402	2578.515	2624.386	2662.755	2699.760	2814.840	2856.236	2879.857	2925.119	2958.953	2991.512	3029.182	3045.283	3065.447	3095.489	3104.554	3115.740	3130.562	3145.686
S.E.	0.425	0.392	0.394	0.399	0.423	0.311	0.235	0.201	0.172	0.160	0.145	0.135	0.139	0.150	0.164	0.166	0.187	0.205	0.225
Qnt MF Mean	2512.749	2578.220	2624.082	2662.700	2699.229	2814.570	2855.883	2879.741	2925.406	2959.140	2991.556	3029.256	3045.540	3065.607	3095.795	3104.949	3115.827	3131.012	3150.746
S.E.	0.394	0.356	0.363	0.371	0.377	0.296	0.229	0.200	0.172	0.160	0.149	0.140	0.143	0.149	0.164	0.163	0.179	0.205	0.152
Panel B: Smooth Option Price Chain with Local Linear Regression using CV Bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	5.753																		
RMSE MFQ Mean	5.749																		
Ratio of RMSE	1.001																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2512.660	2578.096	2624.064	2662.611	2699.115	2814.499	2855.701	2879.618	2925.328	2959.121	2991.540	3029.354	3045.576	3065.737	3095.853	3104.953	3115.968	3131.203	3150.793
S.E.	0.386	0.341	0.343	0.351	0.358	0.281	0.212	0.186	0.156	0.145	0.133	0.125	0.128	0.132	0.148	0.149	0.165	0.189	0.146
Qnt MF Mean	2512.733	2578.096	2624.063	2662.611	2699.116	2814.499	2855.701	2879.618	2925.328	2959.121	2991.540	3029.354	3045.576	3065.737	3095.853	3104.953	3115.968	3131.203	3150.809
S.E.	0.377	0.341	0.343	0.351	0.358	0.282	0.212	0.186	0.156	0.145	0.133	0.125	0.128	0.132	0.148	0.149	0.165	0.189	0.146
Panel C: Smooth Option Price Chain with Local Quadratic Regression using CV Bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	5.946																		
RMSE MFQ Mean	5.861																		
Ratio of RMSE	1.015																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2512.972	2578.105	2623.942	2662.983	2699.489	2814.800	2856.177	2879.673	2925.219	2959.103	2991.550	3029.269	3045.318	3065.629	3095.765	3104.715	3115.679	3130.826	3145.541
S.E.	0.398	0.352	0.353	0.365	0.372	0.288	0.220	0.188	0.158	0.148	0.133	0.127	0.130	0.132	0.150	0.151	0.167	0.195	0.201
Qnt MF Mean	2512.754	2578.161	2623.998	2662.699	2699.169	2814.517	2855.742	2879.703	2925.380	2959.149	2991.576	3029.272	3045.538	3065.642	3095.806	3104.974	3115.893	3131.127	3150.693
S.E.	0.380	0.345	0.348	0.359	0.363	0.286	0.215	0.189	0.160	0.148	0.137	0.127	0.131	0.138	0.152	0.152	0.168	0.194	0.149
Panel D: Smooth Option Price Chain with Cubic B-splines using CV to Select Number of Interior Knots																			
Number Sims	1000																		
RMSE BLQ Mean	5.146																		
RMSE MFQ Mean	5.229																		
Ratio of RMSE	0.984																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt LN2	2512.129	2577.335	2623.916	2662.922	2698.057	2814.635	2855.754	2879.610	2925.100	2959.090	2991.588	3029.358	3045.430	3065.701	3096.020	3104.954	3116.028	3130.952	3155.164
Qnt BL Mean	2512.958	2577.833	2623.847	2662.617	2698.784	2814.561	2855.823	2879.644	2925.272	2958.977	2991.615	3029.256	3045.524	3065.717	3095.844	3104.835	3115.863	3131.081	3148.747
S.E.	0.334	0.307	0.301	0.308	0.322	0.248	0.183	0.164	0.145	0.130	0.120	0.109	0.108	0.114	0.128	0.133	0.146	0.180	0.167
Qnt MF Mean	2512.751	2577.983	2624.187	2662.493	2699.068	2814.470	2855.745	2879.732	2925.421	2959.078	2991.689	3029.344	3045.508	3065.628	3095.886	3104.938	3115.871	3131.251	3150.884
S.E.	0.346	0.308	0.316	0.327	0.327	0.257	0.192	0.170	0.143	0.130	0.120	0.110	0.112	0.116	0.129	0.134	0.147	0.176	0.140

This table reports results from a Monte Carlo simulation to assess the accuracy of model-free quantiles (MFQs) versus quantiles estimated with the approach of Breeden and Litzenberger (1978) (BLQs) recovered from a chain of noisy options prices where the measurement error noise is simulated using the approach in Bondarenko (2003). We replicate the conditions of the near maturity chain of options on a date at the end of our data corresponding to June 28, 2019. This results in 160 different strike prices ranging from a minimum strike price of 2310 to a maximum strike price of 3155. The strike interval between deep OTM put option quotes varies between 5 and 10 index points, with option quotes spaced at uniform intervals of 5 index points from the strike price of 2440 upwards. Option prices are simulated using a two-state lognormal mixture model (LN2) with the following inputs: spot price $S = 2941.76$, interest rate $r = 0.0239$, dividend yield $q = 0.0194$, and time-to-maturity $T = 0.0767$. We use the following parameters in the LN2 model: $p = 0.8702, \mu1 = 0.1286, \mu2 = -0.8601, \sigma1 = 0.0929, \sigma2 = 0.2278$. RMSEs are calculated using quantiles for $\alpha \in \{1\%, 2\%, \ldots, 99\%\}$.

Table A3: Quantile estimation simulation: 2007 market conditions

Panel A: Interpolate Option Price Chain with Cubic Splines																			
Number Sims	1000																		
RMSE BLQ Mean	16.077																		
RMSE MFQ Mean	12.594																		
Ratio of RMSE	1.277																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1329.021	1345.322	1355.768	1363.680	1370.149	1392.592	1407.942	1420.262	1449.542	1474.737	1500.370	1531.302	1544.701	1561.727	1587.309	1594.839	1604.146	1616.602	1636.430
Qnt BL Mean	1324.187	1341.855	1351.778	1361.630	1368.677	1392.514	1408.802	1420.778	1451.291	1475.357	1499.306	1531.068	1543.573	1552.008	1555.047	1555.151	1555.219	1555.269	1555.309
S.E.	0.588	0.537	0.472	0.383	0.312	0.264	0.249	0.250	0.295	0.312	0.306	0.320	0.282	0.201	0.177	0.178	0.179	0.181	0.181
Qnt MF Mean	1328.968	1345.387	1355.960	1363.986	1370.467	1392.711	1408.338	1420.338	1449.867	1474.800	1500.048	1531.024	1544.545	1557.737	1559.642	1559.725	1559.789	1559.839	1559.879
S.E.	0.214	0.174	0.155	0.149	0.140	0.133	0.128	0.123	0.134	0.143	0.141	0.134	0.134	0.085	0.020	0.017	0.014	0.012	0.010
Panel B: Smooth Option Price Chain with Local Linear Regression using LSCV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	13.143																		
RMSE MFQ Mean	12.228																		
Ratio of RMSE	1.075																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1329.021	1345.322	1355.768	1363.680	1370.149	1392.592	1407.942	1420.262	1449.542	1474.737	1500.370	1531.302	1544.701	1561.727	1587.309	1594.839	1604.146	1616.602	1636.430
Qnt BL Mean	1326.525	1343.457	1355.320	1363.456	1370.018	1392.446	1407.970	1420.149	1449.737	1474.780	1499.888	1531.459	1545.374	1555.759	1556.767	1556.767	1556.767	1556.767	1556.767
S.E.	0.324	0.225	0.151	0.124	0.119	0.100	0.095	0.093	0.108	0.119	0.116	0.103	0.136	0.119	0.133	0.133	0.133	0.133	0.133
Qnt MF Mean	1327.771	1345.012	1355.540	1363.492	1369.871	1392.454	1408.090	1420.153	1449.539	1474.790	1500.240	1531.271	1544.709	1558.855	1560.000	1560.000	1560.000	1560.000	1560.000
S.E.	0.194	0.145	0.122	0.118	0.108	0.098	0.094	0.091	0.098	0.112	0.107	0.102	0.103	0.060	0.000	0.000	0.000	0.000	0.000
Panel C: Smooth Option Price Chain with Local Quadratic Regression using LSCV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	13.547																		
RMSE MFQ Mean	12.305																		
Ratio of RMSE	1.101																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1329.021	1345.322	1355.768	1363.680	1370.149	1392.592	1407.942	1420.262	1449.542	1474.737	1500.370	1531.302	1544.701	1561.727	1587.309	1594.839	1604.146	1616.602	1636.430
Qnt BL Mean	1321.864	1343.426	1354.921	1363.186	1369.973	1392.576	1408.336	1420.154	1449.800	1474.802	1499.739	1531.292	1545.290	1555.291	1556.595	1556.638	1556.655	1556.653	1556.646
S.E.	0.446	0.323	0.233	0.192	0.156	0.141	0.134	0.137	0.143	0.153	0.151	0.147	0.174	0.145	0.166	0.166	0.166	0.167	0.168
Qnt MF Mean	1328.324	1345.316	1355.781	1363.694	1370.152	1392.484	1408.171	1420.254	1449.602	1474.808	1500.097	1531.171	1544.664	1558.315	1559.935	1559.972	1559.987	1559.996	1559.999
S.E.	0.208	0.154	0.132	0.127	0.116	0.104	0.101	0.098	0.106	0.121	0.114	0.109	0.111	0.068	0.009	0.006	0.003	0.001	0.000
Panel D: Smooth Option Price Chain with Smoothing Cubic B-Splines																			
Number Sims	1000																		
RMSE BLQ Mean	11.973																		
RMSE MFQ Mean	11.861																		
Ratio of RMSE	1.009																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1329.021	1345.322	1355.768	1363.680	1370.149	1392.592	1407.942	1420.262	1449.542	1474.737	1500.370	1531.302	1544.701	1561.727	1587.309	1594.839	1604.146	1616.602	1636.430
Qnt BL Mean	1327.958	1344.766	1355.407	1363.567	1370.315	1392.669	1407.952	1420.269	1449.532	1474.806	1500.330	1531.225	1544.907	1558.715	1559.611	1559.618	1559.621	1559.624	1559.626
S.E.	0.213	0.056	0.047	0.038	0.029	0.027	0.023	0.023	0.032	0.037	0.036	0.040	0.056	0.067	0.052	0.052	0.052	0.052	0.052
Qnt MF Mean	1329.174	1344.766	1355.421	1363.597	1370.315	1392.677	1407.964	1420.269	1449.532	1474.785	1500.330	1531.226	1544.825	1559.087	1559.982	1559.989	1559.992	1559.995	1559.997
S.E.	0.091	0.056	0.047	0.038	0.029	0.029	0.023	0.023	0.032	0.037	0.036	0.039	0.048	0.050	0.006	0.004	0.003	0.002	0.001

This table reports results from a Monte Carlo simulation to assess the accuracy of model-free quantiles (MFQs) versus quantiles estimated with the approach of Breeden and Litzenberger (1978) and Banz and Miller (1978) (BLQs) recovered from a chain of noisy Black-Scholes options prices with measurement error simulated using the approach in Bondarenko (2003). We replicate the conditions of the near maturity chain of options a date halfway through our data corresponding to July 30, 2007. This results in 52 different strike prices ranging from a minimum strike price of 1290 to a maximum strike price of 1560. The left most four options are spaced at intervals of 10 index points, whereas all other options are spaced at uniform intervals of 5 index points. The spot price is $S = 1474$, the interest rate is $r = 0.05$, the dividend yield is $q = 0.02$, and the time-to-maturity is $T = 0.05$. We use an ATM forward implied volatility of $\sigma = 0.200$ to simulate Black-Scholes model prices. RMSEs are calculated using quantiles for $\alpha \in \{5\%, 6\%, \ldots, 95\%\}$.

Table A4: Quantile estimation simulation: 1999 market conditions

Panel A: Interpolate Option Price Chain with Cubic Splines																			
Number Sims	1000																		
RMSE BLQ Mean	10.936																		
RMSE MFQ Mean	10.063																		
Ratio of RMSE	1.087																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1186.225	1202.239	1212.512	1220.297	1226.667	1248.787	1263.936	1276.108	1305.076	1330.050	1355.502	1386.273	1399.623	1416.602	1442.147	1449.675	1458.983	1471.449	1491.315
Qnt BL Mean	1185.864	1203.007	1212.403	1220.050	1226.959	1248.836	1263.861	1276.601	1304.687	1329.992	1355.294	1386.221	1399.243	1416.613	1420.618	1420.755	1420.849	1420.911	1420.946
S.E.	0.168	0.112	0.100	0.072	0.072	0.054	0.064	0.065	0.117	0.127	0.130	0.121	0.120	0.148	0.149	0.152	0.154	0.155	0.156
Qnt MF Mean	1185.864	1203.007	1212.403	1220.050	1226.959	1248.836	1263.861	1276.601	1304.508	1329.877	1355.232	1386.258	1399.397	1416.914	1424.632	1424.770	1424.864	1424.926	1424.961
S.E.	0.168	0.112	0.100	0.072	0.072	0.054	0.064	0.065	0.108	0.133	0.131	0.124	0.128	0.139	0.024	0.017	0.012	0.008	0.005
Panel B: Smooth Option Price Chain with Local Linear Regression using LSCV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	14.423																		
RMSE MFQ Mean	11.24																		
Ratio of RMSE	1.283																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1186.225	1202.239	1212.512	1220.297	1226.667	1248.787	1263.936	1276.108	1305.076	1330.050	1355.502	1386.273	1399.623	1416.602	1442.147	1449.675	1458.983	1471.449	1491.315
Qnt BL Mean	1189.558	1204.489	1216.819	1227.632	1230.431	1255.518	1267.493	1290.177	1307.677	1328.644	1355.493	1386.401	1399.767	1417.636	1422.665	1422.681	1422.440	1417.301	1409.396
S.E.	0.592	0.679	0.525	0.562	0.673	0.575	0.468	0.409	0.233	0.133	0.089	0.081	0.089	0.117	0.114	0.114	0.212	0.853	1.284
Qnt MF Mean	1180.453	1198.799	1209.290	1223.364	1224.253	1249.153	1266.492	1274.292	1300.924	1325.639	1355.449	1386.546	1399.766	1418.584	1424.983	1424.998	1425.000	1425.000	1425.000
S.E.	0.457	0.150	0.393	0.166	0.069	0.068	0.367	0.069	0.203	0.331	0.083	0.073	0.078	0.129	0.007	0.002	0.000	0.000	0.000
Panel C: Smooth Option Price Chain with Local Quadratic Regression using LSCV bandwidth																			
Number Sims	1000																		
RMSE BLQ Mean	14.798																		
RMSE MFQ Mean	10.909																		
Ratio of RMSE	1.357																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1186.225	1202.239	1212.512	1220.297	1226.667	1248.787	1263.936	1276.108	1305.076	1330.050	1355.502	1386.273	1399.623	1416.602	1442.147	1449.675	1458.983	1471.449	1491.315
Qnt BL Mean	1171.354	1185.406	1211.522	1215.733	1223.228	1246.893	1263.173	1276.964	1306.916	1329.619	1355.282	1385.939	1399.662	1416.136	1416.754	1416.202	1413.464	1412.107	1410.459
S.E.	0.731	0.892	0.419	0.228	0.358	0.388	0.176	0.372	0.196	0.162	0.168	0.146	0.174	0.162	1.045	1.129	1.369	1.480	1.592
Qnt MF Mean	1182.865	1199.573	1211.582	1216.667	1224.553	1246.540	1263.284	1274.605	1303.613	1326.689	1355.286	1386.372	1399.496	1417.244	1424.735	1424.876	1424.952	1424.984	1424.995
S.E.	0.428	0.444	0.251	0.231	0.161	0.055	0.227	0.106	0.147	0.339	0.118	0.110	0.113	0.136	0.020	0.013	0.007	0.004	0.002
Panel D: Smooth Option Price Chain with Cubic B-splines using CV to Select Number of Interior Knots																			
Number Sims	1000																		
RMSE BLQ Mean	10.431																		
RMSE MFQ Mean	9.606																		
Ratio of RMSE	1.086																		
Quantile	1%	2%	3%	4%	5%	10%	15%	20%	35%	50%	65%	80%	85%	90%	95%	96%	97%	98%	99%
Qnt BS	1186.225	1202.239	1212.512	1220.297	1226.667	1248.787	1263.936	1276.108	1305.076	1330.050	1355.502	1386.273	1399.623	1416.602	1442.147	1449.675	1458.983	1471.449	1491.315
Qnt BL Mean	1179.142	1200.315	1211.803	1220.270	1226.668	1248.615	1263.803	1276.130	1305.192	1330.044	1355.450	1386.294	1399.807	1416.872	1422.073	1422.193	1422.255	1422.289	1422.306
S.E.	0.553	0.276	0.183	0.058	0.051	0.033	0.027	0.025	0.027	0.055	0.068	0.071	0.108	0.113	0.119	0.120	0.121	0.122	0.122
Qnt MF Mean	1185.488	1201.374	1212.151	1220.270	1226.668	1248.615	1263.803	1276.130	1305.192	1330.020	1355.500	1386.297	1399.612	1417.531	1424.758	1424.877	1424.939	1424.973	1424.990
S.E.	0.133	0.076	0.066	0.058	0.051	0.033	0.027	0.025	0.027	0.055	0.067	0.071	0.085	0.139	0.021	0.013	0.009	0.005	0.003

This table reports results from a Monte Carlo simulation to assess the accuracy of model-free quantiles (MFQs) versus quantiles estimated with the approach of Breeden and Litzenberger (1978) and Banz and Miller (1978) (BLQs) recovered from a chain of noisy Black-Scholes options prices with measurement error simulated using the approach in Bondarenko (2003). We replicate the conditions of the near maturity chain of options on a date at the beginning of the data sample which corresponds to July 30, 1999 but with option price parameters rounded down for ease of replication. This results in using 30 different strike prices ranging from a minimum strike price of 1125 to a maximum strike price of 1425. The left most eight options are spaced at intervals of 25 index points, then there is a gap of 20 index points to the 9th option with the remaining options spaced at uniform intervals of 5 index points. The spot price is $S = 1329$, the interest rate is $r = 0.05$, the dividend yield is $q = 0.01$, the time-to-maturity is $T = 0.05$, and we use a single ATM forward implied volatility of $\sigma = 0.22$ to simulate the Black-Scholes model prices. RMSEs are calculated using quantiles for $\alpha \in \{1\%, 2\%, 3\%, \ldots, 99\%\}$

Appendix B Risk-Neutral Mean and Variance in a Smooth Half Normal CARA Economy

The smooth half normal (SHN) density function (see de Roon and Karehnke 2017) is defined as follows:

$$g(x) = \begin{cases} \lambda_1 f(x; m, s_1) & \text{if } x \leq m, \\ \lambda_2 f(x; m, s_2) & \text{if } x > m. \end{cases}$$

where $f(x; \mu_x, \sigma_x)$ is the normal density function with mean μ_x and standard deviation σ_x, λ_1 and λ_2 are chosen to ensure the density function is continuous and integrates to one, and where s_1, s_2 and m are chosen to match the mean μ, variance σ and skewness γ of the excess return distribution. Following de Roon and Karehnke (2017) we assume a two period economy with a representative investor that maximizes a CARA utility function $u(x) = -e^{-\theta w_0 (1+r_f+x)}$ where θ is the risk aversion coefficient, w_0 the initial wealth, r_f is the risk-free rate and x is the excess return with $x \sim SHN(\mu, \sigma, \gamma)$. The risk-neutral density $g^*(x)$ is related to the physical density function (subject to conditions such as complete and frictionless markets) as follows:

$$g^*(x) = \frac{u'((1 + r_f + x))g(x)}{\int_{-\infty}^{\infty} u'((1 + r_f + x))g(x)}$$

The resulting risk-neutral density function is given by:

$$g^*(x) = \begin{cases} \lambda_1^* f(x; m - \theta w_0 s_1^2, s_1) & \text{if } x \leq m, \\ \lambda_2^* f(x; m - \theta w_0 s_2^2, s_2) & \text{if } x > m. \end{cases}$$

where

$$\lambda_i^* = \lambda_i \frac{c_i}{c},$$
$$c_i = e^{\frac{1}{2}\left(\theta^2 w_0^2 s_i^2\right)},$$
$$c = \lambda_1 c_1 \Phi\left(\theta w_0 s_1\right) + \lambda_2 c_2 \left(1 - \Phi\left(\theta w_0 s_2\right)\right),$$

for $i = 1, 2$ and where $\Phi(x)$ denotes the standard normal CDF. The risk-neutral conditional means are given by:

$$E^*[x|x \leq m] = m - \theta w_0 s_1^2 - s_1 \frac{\phi(\theta w_0 s_1)}{\Phi(\theta w_0 s_1)},$$
$$E^*[x|x > m] = m - \theta w_0 s_2^2 + s_2 \frac{\phi(\theta w_0 s_2)}{1 - \Phi(\theta w_0 s_2)},$$

where $\phi(x)$ denotes the standard normal PDF. The corresponding risk-neutral probabilities are given by:

$$\Pr{}^*[x \leq m] = \lambda_1^* \Phi(\theta w_0 s_1),$$
$$\Pr{}^*[x > m] = \lambda_2^* \left(1 - \Phi(\theta w_0 s_2)\right).$$

Hence, the risk-neutral mean is given by:

$$\begin{aligned} E^*[x] &= E^*[x|x \leq m] \Pr{}^*[x \leq m] + E^*[x|x > m] \Pr{}^*[x > m] \\ &= \lambda_1^* \left(m - \theta w_0 s_1^2\right) \Phi(\theta w_0 s_1) - \lambda_1^* s_1 \phi(\theta w_0 s_1) \\ &+ \lambda_2^* \left(m - \theta w_0 s_2^2\right) \left(1 - \Phi(\theta w_0 s_2)\right) + \lambda_2^* s_2 \phi(\theta w_0 s_2) \end{aligned}$$

The risk-neutral conditional expectations of the square of a SHN-CARA random variable are given by:

$$E^*[x^2|x \le m] = \left(m - \theta w_0 s_1^2\right)^2 + s_1^2 + s_1\left(\theta w_0 s_1^2 - 2m\right)\frac{\phi(\theta w_0 s_1)}{\Phi(\theta w_0 s_1)},$$
$$E^*[x^2|x > m] = \left(m - \theta w_0 s_2^2\right)^2 + s_2^2 + s_2\left(\theta w_0 s_2^2 + 2m\right)\frac{\phi(\theta w_0 s_2)}{1 - \Phi(\theta w_0 s_2)}.$$

The risk-neutral conditional variances are given by:

$$V^*[x|x \le m] = E^*[x^2|x \le m] - E^*[x|x \le m]^2,$$
$$V^*[x|x > m] = E^*[x^2|x > m] - E^*[x|x > m]^2.$$

Using the law of total variance, the risk-neutral variance in a SHN-CARA economy is given by:

$$\begin{aligned} V^*[x] &= V^*[x|x \le m]\operatorname{Pr}^*[x \le m] + V^*[x|x > m]\operatorname{Pr}^*[x > m] \\ &\quad + \left(E^*[x|x \le m] - E^*[x|x > m]\right)^2 \operatorname{Pr}^*[x \le m]\operatorname{Pr}^*[x > m]. \end{aligned}$$

It is interesting to note that in a SHN-CARA economy the variance risk premium defined as the difference between the risk-neutral and physical variance, $VRP = V^*[x] - V[x]$, is positive when the physical skewness γ is negative, zero when $\gamma = 0$ and negative when $\gamma > 0$. The fact that in a SHN-CARA economy the VRP is positive when skewness is negative is in agreement with the stylised empirical facts of negative skewness and positive VRP in major international equity markets.

Appendix C Further Predictive Regression Results

In this section of the Appendix we report further predictive regression results. Table C1 presents predictive regression results similar to those in Table 5 but where the quantile extraction method uses interpolating cubic-splines, as opposed to smoothing B-splines. As can be seen from the table, the results are quantitatively and qualitatively similar to the results in Table 5 highlighting that RNA-Qs are not overly sensitive to the procedure used to extract quantiles provided we use the MFQ method.

Table C2 reports predictive regression results where, following robustness tests in Table Martin (2017), we remove the period of the financial crisis from August 1, 2008 to July 31, 2009, in which variance spiked upwards, the stock market crashed and then subsequently recovered strongly. As can be seen in C2, the results for RNA-Q become even stronger when this period of the financial crisis is removed from the data sample.

Table C1: Predictive Regression Results: Cubic-splines

	RNA-Q_{1-10}					RNA-Q_{1-3}					RNA-Q_{4-7}					RNA-Q_{8-10}					RNS-BKM				
	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2
Panel A: Skew values averaged over week																									
1W	-1.542	-0.416	-1.596	0.110	0.001	-0.640	-0.156	-0.518	0.605	-0.001	-2.015	-0.569	-2.282	0.022	0.003	-0.832	-0.319	-1.162	0.245	0.000	-0.052	-0.035	-0.152	0.879	-0.001
2W	-1.161	-0.313	-1.491	0.136	0.001	-0.486	-0.118	-0.459	0.646	-0.000	-1.395	-0.394	-1.959	0.050	0.003	-0.712	-0.273	-1.279	0.201	0.001	-0.085	-0.058	-0.288	0.773	-0.001
3W	-1.140	-0.308	-1.604	0.109	0.003	-0.148	-0.036	-0.156	0.876	-0.001	-1.264	-0.358	-1.932	0.053	0.004	-0.909	-0.348	-1.776	0.076	0.004	-0.049	-0.033	-0.168	0.867	-0.001
4W	-1.325	-0.358	-1.949	0.051	0.005	-0.448	-0.109	-0.506	0.613	-0.000	-1.425	-0.403	-2.252	0.024	0.007	-0.978	-0.375	-1.929	0.054	0.006	-0.078	-0.053	-0.265	0.791	-0.001
5W	-1.360	-0.367	-2.053	0.040	0.007	-0.730	-0.178	-0.829	0.407	0.001	-1.460	-0.413	-2.337	0.019	0.010	-0.897	-0.344	-1.843	0.065	0.006	-0.091	-0.062	-0.300	0.764	-0.001
6W	-1.485	-0.401	-2.293	0.022	0.011	-0.866	-0.211	-0.982	0.326	0.002	-1.586	-0.448	-2.580	0.010	0.014	-0.959	-0.367	-2.063	0.039	0.009	-0.118	-0.080	-0.382	0.702	-0.000
7W	-1.523	-0.411	-2.338	0.019	0.013	-1.047	-0.255	-1.181	0.238	0.005	-1.652	-0.466	-2.646	0.008	0.017	-0.900	-0.344	-1.993	0.046	0.009	-0.159	-0.108	-0.513	0.608	0.000
8W	-1.465	-0.395	-2.236	0.025	0.014	-1.244	-0.303	-1.421	0.155	0.008	-1.558	-0.439	-2.468	0.014	0.018	-0.791	-0.302	-1.739	0.082	0.008	-0.210	-0.143	-0.671	0.502	0.001
9W	-1.401	-0.378	-2.090	0.037	0.015	-1.348	-0.328	-1.526	0.127	0.011	-1.537	-0.434	-2.387	0.017	0.020	-0.658	-0.251	-1.410	0.158	0.006	-0.239	-0.162	-0.743	0.458	0.002
10W	-1.348	-0.363	-1.996	0.046	0.015	-1.350	-0.329	-1.504	0.133	0.012	-1.445	-0.408	-2.219	0.026	0.019	-0.636	-0.243	-1.363	0.173	0.006	-0.240	-0.163	-0.733	0.463	0.002
11W	-1.278	-0.345	-1.862	0.063	0.015	-1.334	-0.325	-1.496	0.135	0.013	-1.355	-0.383	-2.046	0.041	0.019	-0.592	-0.227	-1.234	0.217	0.006	-0.239	-0.162	-0.725	0.468	0.003
12W	-1.196	-0.323	-1.716	0.086	0.014	-1.277	-0.311	-1.429	0.153	0.013	-1.250	-0.353	-1.856	0.063	0.017	-0.557	-0.213	-1.125	0.260	0.006	-0.248	-0.167	-0.740	0.459	0.003
Panel B: Skew values at week end																									
1W	-0.825	-0.252	-0.953	0.341	-0.000	-0.189	-0.053	-0.186	0.853	-0.001	-1.809	-0.592	-2.385	0.017	0.003	0.069	0.033	0.112	0.911	-0.001	0.082	0.056	0.247	0.805	-0.001
2W	-0.829	-0.254	-1.195	0.232	0.001	-0.099	-0.028	-0.111	0.912	-0.001	-1.287	-0.421	-2.136	0.033	0.003	-0.293	-0.140	-0.638	0.523	-0.000	-0.062	-0.043	-0.211	0.833	-0.001
3W	-0.684	-0.210	-1.061	0.289	0.001	0.308	0.087	0.382	0.703	-0.001	-1.067	-0.349	-1.903	0.057	0.004	-0.373	-0.179	-0.905	0.366	0.000	-0.032	-0.022	-0.114	0.909	-0.001
4W	-0.777	-0.238	-1.321	0.186	0.002	0.090	0.025	0.123	0.902	-0.001	-0.985	-0.323	-1.931	0.054	0.004	-0.477	-0.229	-1.236	0.216	0.002	-0.030	-0.020	-0.103	0.918	-0.001
5W	-0.852	-0.261	-1.509	0.131	0.003	-0.252	-0.071	-0.360	0.719	-0.001	-1.078	-0.353	-2.183	0.029	0.007	-0.403	-0.193	-1.080	0.280	0.001	-0.064	-0.044	-0.221	0.825	-0.001
6W	-1.009	-0.309	-1.882	0.060	0.006	-0.552	-0.155	-0.814	0.416	0.001	-1.122	-0.368	-2.356	0.018	0.009	-0.487	-0.233	-1.372	0.170	0.003	-0.089	-0.062	-0.303	0.762	-0.001
7W	-1.102	-0.338	-2.086	0.037	0.009	-0.691	-0.194	-1.040	0.298	0.002	-1.173	-0.384	-2.484	0.013	0.011	-0.534	-0.256	-1.585	0.113	0.005	-0.111	-0.077	-0.375	0.707	-0.000
8W	-1.148	-0.351	-2.137	0.033	0.011	-0.815	-0.229	-1.231	0.218	0.004	-1.155	-0.378	-2.368	0.018	0.013	-0.566	-0.271	-1.676	0.094	0.006	-0.168	-0.117	-0.561	0.575	0.000
9W	-1.140	-0.349	-2.106	0.035	0.012	-0.968	-0.272	-1.485	0.138	0.007	-1.161	-0.380	-2.394	0.017	0.015	-0.496	-0.238	-1.444	0.149	0.005	-0.204	-0.141	-0.670	0.503	0.001
10W	-1.121	-0.343	-2.041	0.041	0.013	-0.967	-0.272	-1.453	0.146	0.008	-1.137	-0.372	-2.245	0.025	0.016	-0.486	-0.233	-1.455	0.146	0.006	-0.211	-0.145	-0.679	0.497	0.002
11W	-1.074	-0.329	-1.997	0.046	0.014	-1.014	-0.285	-1.541	0.123	0.010	-1.033	-0.338	-2.103	0.035	0.015	-0.470	-0.225	-1.418	0.156	0.006	-0.224	-0.154	-0.717	0.473	0.002
12W	-0.999	-0.306	-1.793	0.073	0.013	-0.991	-0.278	-1.498	0.134	0.011	-0.930	-0.305	-1.826	0.068	0.013	-0.441	-0.211	-1.291	0.197	0.006	-0.218	-0.151	-0.693	0.488	0.003

Table C1 reports predictive regression results for h-week ahead index returns using RNA-Q_{1-10}, RNA-Q_{1-3}, RNA-Q_{4-7}, RNA-Q_{8-10}, and RNS-BKM as individual predictors. The t-statistics are estimated using Newey-West standard error estimates allowing for a lag equal to two times the overlap of the dependent variable. The period is from January 1996 through June 2019, and the frequency is weekly. Panel A reports results where the predictor has been averaged over the previous week. Panel B reports results where the predictor is the most recently available predictor at the end of the week prior to the h-week ahead return horizon.

Table C2: Predictive Regression Results: Removing Financial Crisis

	RNA-Q_{1-10}					RNA-Q_{1-3}					RNA-Q_{4-7}					RNA-Q_{8-10}					RNS-BKM				
	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2	Beta	Beta-Std	t-stat	p-val	Adj-R^2
Panel A: Skew values averaged over week																									
1W	-1.571	-0.434	-1.536	0.124	0.002	-0.240	-0.056	-0.179	0.858	-0.001	-2.146	-0.647	-2.281	0.023	0.005	-0.828	-0.324	-1.173	0.241	0.000	-0.093	-0.067	-0.286	0.775	-0.001
2W	-1.563	-0.432	-2.114	0.034	0.005	-0.028	-0.007	-0.028	0.978	-0.001	-2.018	-0.608	-2.951	0.003	0.011	-0.990	-0.388	-2.013	0.044	0.004	-0.118	-0.085	-0.442	0.658	-0.001
3W	-1.504	-0.416	-2.177	0.029	0.008	0.224	0.052	0.248	0.804	-0.001	-1.948	-0.587	-2.885	0.004	0.016	-1.037	-0.406	-2.267	0.023	0.007	-0.057	-0.041	-0.221	0.825	-0.001
4W	-1.484	-0.410	-2.194	0.028	0.010	0.044	0.010	0.053	0.958	-0.001	-1.851	-0.558	-2.694	0.007	0.020	-1.016	-0.398	-2.289	0.022	0.010	-0.059	-0.043	-0.231	0.817	-0.001
5W	-1.567	-0.433	-2.358	0.018	0.015	-0.235	-0.055	-0.305	0.760	-0.001	-1.888	-0.569	-2.728	0.006	0.027	-1.027	-0.402	-2.402	0.016	0.013	-0.060	-0.044	-0.234	0.815	-0.001
6W	-1.529	-0.422	-2.340	0.019	0.018	-0.343	-0.080	-0.466	0.641	-0.000	-1.775	-0.535	-2.562	0.010	0.029	-1.015	-0.397	-2.470	0.014	0.016	-0.087	-0.063	-0.333	0.739	-0.001
7W	-1.467	-0.405	-2.231	0.026	0.019	-0.470	-0.109	-0.660	0.509	0.000	-1.699	-0.512	-2.453	0.014	0.031	-0.927	-0.363	-2.277	0.023	0.015	-0.130	-0.094	-0.497	0.619	0.000
8W	-1.387	-0.383	-2.118	0.034	0.020	-0.630	-0.146	-0.909	0.363	0.002	-1.606	-0.484	-2.319	0.020	0.033	-0.811	-0.318	-2.072	0.038	0.014	-0.169	-0.123	-0.641	0.521	0.001
9W	-1.301	-0.359	-1.979	0.048	0.021	-0.742	-0.172	-1.067	0.286	0.004	-1.443	-0.435	-2.110	0.035	0.031	-0.758	-0.297	-1.964	0.049	0.014	-0.195	-0.142	-0.727	0.467	0.002
10W	-1.226	-0.339	-1.884	0.060	0.020	-0.847	-0.197	-1.190	0.234	0.006	-1.355	-0.408	-1.999	0.046	0.030	-0.666	-0.261	-1.786	0.074	0.012	-0.208	-0.151	-0.763	0.445	0.003
11W	-1.155	-0.319	-1.757	0.079	0.019	-0.838	-0.195	-1.186	0.236	0.007	-1.293	-0.390	-1.882	0.060	0.030	-0.600	-0.235	-1.581	0.114	0.010	-0.222	-0.161	-0.805	0.421	0.004
12W	-1.098	-0.304	-1.603	0.109	0.019	-0.838	-0.195	-1.160	0.246	0.007	-1.220	-0.368	-1.717	0.086	0.028	-0.565	-0.221	-1.416	0.157	0.009	-0.233	-0.169	-0.828	0.408	0.005
Panel B: Skew values at week end																									
1W	-1.313	-0.417	-1.493	0.135	0.001	-0.806	-0.205	-0.632	0.527	-0.000	-1.152	-0.416	-1.479	0.139	0.001	-0.862	-0.403	-1.474	0.141	0.001	0.012	0.009	0.039	0.969	-0.001
2W	-1.392	-0.442	-2.180	0.029	0.005	-0.651	-0.166	-0.684	0.494	-0.000	-1.452	-0.524	-2.544	0.011	0.008	-0.791	-0.369	-1.876	0.061	0.003	-0.104	-0.077	-0.414	0.679	-0.001
3W	-0.768	-0.244	-1.243	0.214	0.002	0.048	0.012	0.057	0.955	-0.001	-0.973	-0.351	-1.772	0.076	0.005	-0.421	-0.197	-0.992	0.321	0.001	-0.050	-0.038	-0.212	0.832	-0.001
4W	-0.752	-0.238	-1.375	0.169	0.003	0.084	0.021	0.107	0.915	-0.001	-0.892	-0.322	-1.755	0.079	0.006	-0.470	-0.220	-1.369	0.171	0.002	-0.008	-0.006	-0.034	0.973	-0.001
5W	-0.991	-0.314	-1.963	0.050	0.007	-0.115	-0.029	-0.162	0.871	-0.001	-1.114	-0.402	-2.223	0.026	0.013	-0.604	-0.282	-1.950	0.051	0.006	-0.028	-0.021	-0.117	0.907	-0.001
6W	-1.040	-0.330	-2.147	0.032	0.010	-0.221	-0.056	-0.337	0.736	-0.001	-1.117	-0.403	-2.217	0.027	0.016	-0.644	-0.301	-2.282	0.022	0.008	-0.050	-0.038	-0.209	0.834	-0.001
7W	-0.954	-0.303	-2.099	0.036	0.010	-0.298	-0.076	-0.481	0.630	-0.000	-0.992	-0.358	-2.089	0.037	0.015	-0.588	-0.275	-2.271	0.023	0.008	-0.067	-0.050	-0.281	0.778	-0.001
8W	-1.023	-0.325	-2.173	0.030	0.014	-0.377	-0.096	-0.635	0.526	0.000	-1.067	-0.385	-2.166	0.030	0.021	-0.611	-0.286	-2.350	0.019	0.011	-0.111	-0.083	-0.456	0.648	-0.000
9W	-1.042	-0.330	-2.224	0.026	0.017	-0.470	-0.120	-0.799	0.424	0.001	-1.027	-0.371	-2.096	0.036	0.022	-0.644	-0.301	-2.588	0.010	0.014	-0.142	-0.106	-0.578	0.563	0.001
10W	-1.028	-0.326	-2.170	0.030	0.019	-0.573	-0.146	-0.944	0.345	0.003	-1.007	-0.363	-2.065	0.039	0.024	-0.609	-0.285	-2.453	0.014	0.014	-0.170	-0.127	-0.676	0.499	0.002
11W	-1.021	-0.324	-2.230	0.026	0.020	-0.694	-0.176	-1.146	0.252	0.005	-1.007	-0.363	-2.117	0.034	0.026	-0.562	-0.263	-2.249	0.024	0.013	-0.206	-0.154	-0.812	0.417	0.004
12W	-0.927	-0.294	-1.881	0.060	0.018	-0.665	-0.169	-1.092	0.275	0.005	-0.900	-0.325	-1.802	0.072	0.022	-0.511	-0.239	-1.936	0.053	0.011	-0.210	-0.157	-0.812	0.417	0.004

Table C2 reports predictive regression results for h-week ahead index returns using RNA-Q_{1-10}, RNA-Q_{1-3}, RNA-Q_{4-7}, RNA-Q_{8-10}, and RNS-BKM as individual predictors. The t-statistics are estimated using Newey-West standard error estimates allowing for a lag equal to two times the overlap of the dependent variable. The period is from January 1996 through June 2019, and the frequency is weekly. The financial crisis period of August 1, 2008 to July 31, 2009 has been removed.

Appendix D Higher Order Cornish Fisher Expansion

The Cornish-Fisher (CF) expansion up to fifth order expresses a non-normal standard quantile, w, in terms of a standard normal quantile, x, and the the distribution cumulants κ_n as follows:

$$\begin{aligned} w \approx x &+ [\gamma_1 h_1(x)] \\ &+ \left[\gamma_2 h_2(x) + \gamma_1^2 h_{22}(x)\right] \\ &+ \left[\gamma_3 h_3(x) + \gamma_1 \gamma_2 h_{12}(x) + \gamma_1^3 h_{111}(x)\right] \\ &+ \left[\gamma_4 h_4(x) + \gamma_2^2 h_{22}(x) + \gamma_1 \gamma_3 h_{13}(x) + \gamma_1^2 \gamma_2 h_{112}(x) + \gamma_1^4 h_{1111}(x)\right] \end{aligned}$$

where

$$\begin{aligned} x &= \Phi^{-1}(\alpha) \\ \gamma_{r-2} &= \frac{\kappa_r}{\kappa_2^{r/2}};\ r \in \{3, 4, \ldots\} \\ h_1(x) &= \frac{1}{6}\mathrm{He}_2(x) \\ h_2(x) &= \frac{1}{24}\mathrm{He}_3(x) \\ h_{11}(x) &= -\frac{1}{36}\left[2\mathrm{He}_3(x) + \mathrm{He}_1(x)\right] \\ h_3(x) &= \frac{1}{120}\mathrm{He}_4(x) \\ h_{12}(x) &= -\frac{1}{24}\left[\mathrm{He}_4(x) + \mathrm{He}_2(x)\right] \\ h_{111}(x) &= \frac{1}{324}\left[12\mathrm{He}_4(x) + 19\mathrm{He}_2(x)\right] \\ h_4(x) &= \frac{1}{720}\mathrm{He}_5(x) \\ h_{22}(x) &= -\frac{1}{384}\left[3\mathrm{He}_5(x) + 6\mathrm{He}_3(x) + 2\mathrm{He}_1(x)\right] \\ h_{13}(x) &= -\frac{1}{180}\left[2\mathrm{He}_5(x) + 3\mathrm{He}_3(x)\right] \\ h_{112}(x) &= \frac{1}{288}\left[14\mathrm{He}_5(x) + 37\mathrm{He}_3(x) + 8\mathrm{He}_1(x)\right] \end{aligned}$$

$$h_{1111}(x) = -\frac{1}{7776}\left[252\mathrm{He}_5(x) + 832\mathrm{He}_3(x) + 227\mathrm{He}_1(x)\right]$$

where $He_n(x)$ are he probabilist's Hermite polynomials given by

$$\mathrm{He}_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}$$

Hermite polynomials $He_n(x)$ are even or odd functions depending on n with

$$\mathrm{He}_n(-x) = (-1)^n \mathrm{He}_n(-x)$$

The sum of two symmetric quantiles at, respectively, probability levels α and $1-\alpha$ is given by $w_\alpha + w_{1-\alpha}$. This sum is a function of the odd Hermite polynomials as the even Hermite polynomials cancel. Taking a fifth order expansion results in

$$w_\alpha + w_{1-\alpha} = 2\left(\gamma_1 h_1(x) + \gamma_3 h_3(x) + \gamma_1\gamma_2 h_{12}(x) + \gamma_1^3 h_{111}(x)\right)$$

Dividing across by $2h_1(x)$ yields the measure of asymmetry we use in the paper that we denote as RNA-Q

$$\begin{aligned}\text{RNA-Q} &= \frac{1}{2h_1(x)}\left(w_\alpha + w_{1-\alpha}\right) \\ &= \gamma_1 + \frac{1}{h_1(x)}\left(\gamma_3 h_3(x) + \gamma_1\gamma_2 h_{12}(x) + \gamma_1^3 h_{111}(x)\right)\end{aligned}$$

Along with skewness γ_1, we see that higher order cumulants γ_2 (kurtosis) and γ_3 (fifth order cumulant related to hyperskewness) also impact the asymmetry measure in the fifth order expansion.