1 Introduction to the Tail Indices

Volatility indices, and the VIX index in particular, have drawn an unprecedented amount of attention from investors and regulators in recent years. They serve as a benchmark of market risk and sentiment, a measure of portfolio risk, as well as an investment vehicle through the associated derivatives.

Formally, the VIX methodology is designed to capture the expected (risk-neutral) implied volatility of the underlying asset over a given future period. However, both academia and industry have noted that security price fluctuations–as captured by the volatility indices–are exposed to two distinct types of risk: diffusive and jump risk. The diffusive risk refers to continuous fluctuation–essentially the small price moves–while jump risk is associated with the return fluctuations stemming abrupt and discontinuous moves to the downside (crashes) or upside (spikes). There is growing recognition that investors demand very different compensation for the two types of risk exposure, leading to a desire to disentangle the implied volatility indices and measures into components that reflect the diffusive and jump risks, respectively.

The provision of an updated tail index is intended to provide the public with ready access to a robust jump risk measure. It is based on the procedure laid out originally in the article *Tail risk premia and return predictability* (Bollerslev, Todorov and Xu, Journal of Financial Economics, 2015). Using this methodology, we compute two separate measures of market jump risk. The **Left Jump Volatility** estimates the expected (risk-neutral) return volatility that stems from large negative price jumps, while the **Left Jump Probability** measure estimates the probability that the index drops by 10% or more over the coming week. Because the identification and separation of the diffusive versus the jump component is best achieved from short-dated options, our measures are generated from derivative securities that are close to expiry, although we avoid extremely short-lived options as they are prone to excessive measurement and microstructure biases. We also emphasize that our tail indices are derived from securities priced under the so-called risk-neutral measure, so they embed (large) risk premiums associated with downward market moves along with the expectations regarding the actual likelihood of such events. Consequently, these tail measures are more suitable as indicators of market risk pricing than objective measures of the likelihood that such adverse price moves actually will materialize. On the website, both indices are reported in percentage terms.

We provide tail index measures for a number of popular U.S. stock indices including the S&P 500 Index, the Dow Jones Industrial Average, the NASDAQ 100 Index, and the Russell 2000 Index, as well as a selection of prominent Exchange Traded Funds (ETFs). The same methodology may be applied to other assets or indices featuring a liquid option market.

2 Calculation

2.1 Data

The methodology relies on a nonparametric model for short-maturity options. We focus exclusively on options with a tenor between 6 and 31 trading days. Options with even shorter maturities are excluded to avoid complications stemming from measurement errors associated with frictions and low option prices.

To isolate the jump risk from the diffusion risk in the tail measure, we use only deep out-of-the-money (OTM) put options, which derive their value almost solely from the prospect of negative jumps. The threshold for the deep OTM puts is defined by the maturity-adjusted log-moneyness being less than or equal to -2.5,

$$\frac{ln(K/F_{t,\tau})}{\sigma_{ATM,\tau}\sqrt{\tau}} \leq -2.5$$

where,

$$K = \text{strike price}$$
$$F_{t,\tau} = \text{forward price of index at t, maturity } \tau$$
$$\tau = \frac{\text{number of trading days to maturity}}{252}$$
$$\sigma_{ATM,\tau} = \text{at-the-money (ATM) implied volatility}$$

Time-to-maturity τ is measured in terms of business days as a fraction of a year. By adjusting moneyness with implied (ATM) volatility and option maturity, the threshold is dynamic and robust, as it adapts suitably to shifting market conditions.

Deep OTM options are often fairly illiquid and have wide bid-ask spreads. We select only those with a positive bid price to exclude the most illiquid ones. As an indication of fair market value, we use the mid-quote, i.e., the average of the bid and ask quotes, as the price of the option. To avoid violations of the no-arbitrage principle and further robustify our procedure, we also enforce a monotone valuation rule: if a put option with a lower strike price has a higher or equal value to a preceding one with higher strike price, it is excluded from our computation. That is, moving from higher to lower strike prices, we ensure a strictly decreasing sequence of option valuations. Figure 1 provides an illustration of our approach.

Mid Price	Strike	Adj. Moneyness	
0.075	1525	-8.75	
0.075	1550	-8.05	
0.075	1560	-7.77	
0.075	1570	-7.49	
0.075	1575	-7.36	
0.1	1590	-6.95	
0.225	1595	-6.81	
0.1	1600	-6.68	
0.125	1610	-6.41	
0.1	1620	-6.14	
0.125	1625	-6.01	
0.1	1630	-5.88	
0.1	1635	-5.74	
0.1	1640	-5.61	Deep OTM
0.125	1645	-5.48	
0.15	1650	-5.35	
0.125	1655	-5.22	
0.15	1660	-5.09	
0.225	1665	-4.96	
0.275	1670	-4.83	
0.175	1675	-4.70	
0.175	1680	-4.57	
0.275	1685	-4.44	
0.175	1690	-4.32	
0.175	1695	-4.19	
0.25	1700	-4.06	
0.225	1705	-3.93	
0.3	1710	-3.81	
0.45	1715	-3.68	

Mid Price	Strike	Adj. Moneyness	
0.35	1720.00	-3.56	
0.35	1725.00	-3.43	
0.43	1730.00	-3.31	
0.45	1735.00	-3.18	Deep OTM
0.60	1740.00	-3.06	
0.68	1745.00	-2.93	
0.75	1750.00	-2.81	
0.60	1755.00	-2.69	
0.88	1760.00	-2.56	
0.90	1765.00	-2.44	
1.20	1770.00	-2.32	
1.08	1775.00	-2.20	
1.30	1780.00	-2.08	
1.25	1785.00	-1.95	
1.45	1790.00	-1.83	
1.75	1795.00	-1.71	
2.08	1800.00	-1.59	
2.30	1805.00	-1.47	
2.90	1810.00	-1.35	
3.13	1815.00	-1.24	
3.75	1820.00	-1.12	
4.30	1825.00	-1.00	
5.05	1830.00	-0.88	
5.85	1835.00	-0.76	
6.80	1840.00	-0.64	
7.85	1845.00	-0.53	
9.20	1850.00	-0.41	
10.60	1855.00	-0.29	
12.40	1860.00	-0.18	
14.40	1865.00	-0.06	

Figure 1: SPX option quotes on 2014/05/06 for 2014/05/17 expiry. The quotes highlighted in yellow are used in our calculation of the index. In practice, we pool data across multiple maturities for more robust estimation.

The ATM implied volatility, σ_{ATM}, is estimated by a weighted average of the implied volatility of the largest strike OTM put option and the lowest strike OTM call option. The formula is given by,

$$\sigma_{ATM,\tau} \ = \ \frac{F_{t,\tau} - K_P}{K_C - K_P} \sigma_{ATM,C} \ + \ \frac{K_C - F_{t,\tau}}{K_C - K_P} \sigma_{ATM,P}$$

where K_C and K_P denote the strike prices of the OTM call and put option, respectively, with strike prices closest to forward index. $\sigma_{ATM,C}$ and $\sigma_{ATM,P}$ indicate the corresponding Black-Scholes implied volatility of these close-to-the-money options.

2.2 Calculation

As mentioned, the tail indices are based on the procedure in Bollerslev, Todorov and Xu (2015). It exploits the fact that the density characterizing the return tails very generally take the form associated with the Fréchet Extreme Value Distribution (EVT); we refer to the paper for further details. To construct the measure, two parameters must be estimated period-per-period, namely the tail shape parameter α^- and the level shift parameter ϕ^-. They are obtained as follows,

$$\widehat{\alpha}_t^- = \operatorname*{median}_{2 \leq n \leq N} \left| 1 - \frac{\ln \frac{O_{t,\tau}(k_{t,i})}{O_{t,\tau}(k_{t,i-1})}}{k_{t,i} - k_{t,i-1}} \right|$$

$$\widehat{\phi}_t^- = \operatorname*{median}_{1 \leq n \leq N} \left| \ln \frac{e^{r_{t,\tau}} O_{t,\tau}(k_{t,i})}{\tau F_{t,\tau}} - (1 + \widehat{\alpha}_t^-)k_{t,i} + \ln(\widehat{\alpha}_t^- + 1) + \ln(\widehat{\alpha}_t^-) \right|$$

where,

$$N = \text{number of valid option quotes}$$
$$k_{t,i} = \ln(K_{t,i}/F_{t,\tau}) = \text{unadjusted log-moneyness of i'th option}$$
$$O_{t,\tau}(k_{t,i}) = \text{mid-quote for i'th option}$$
$$r_{t,\tau} = \text{risk-free interest rate for maturity } \tau$$

We use the 3-month U.S. treasury rate as our proxy for the risk-free rate. The market depth and the width of option range increased drastically over the last decade. To obtain a consistent and less noisy estimator for the early sample period, we pool the data to compute a single weekly α^- before 2013. From 2013 onward, we calculate α^- on a daily basis. For ETFs, due to lower liquidity, we pool the estimates across each week for the entire sample period. In either case, we only compute the measure if we have at least 4 pairs of $[O_{t,\tau}(k_{t,i}), O_{t,\tau}(k_{t,i-1})]$ available to ensure reasonable robustness. In contrast, ϕ^- is less sensitive to outliers, and we estimate it at a daily level throughout.

Given the point estimates $\widehat{\alpha}_t^-$ and $\widehat{\phi}_t^-$, our tail index measures **Left Jump Volatility** and **Left Jump Probability** are given by,

$$\text{LJV} = 100 \times \sqrt{\widehat{\phi}_t^- \cdot e^{-\widehat{\alpha}_t^- |\theta_t|} \frac{\widehat{\alpha}_t^- |\theta_t| (\widehat{\alpha}_t^- k_t + 2) + 2}{(\widehat{\alpha}_t^-)^3}}$$

$$\text{Left Jump Probability} = 100 \times \widehat{\phi}_t^- \cdot \frac{e^{-\widehat{\alpha}_t^- |\theta_t|}}{\widehat{\alpha}_t^-}$$

where θ_t denotes the threshold for a negative tail jump at the 1-week horizon. For LJV, we let $\theta_t = 10 \cdot \sigma_{ATM,30d} \cdot \sqrt{5/252}$, i.e., a ten standard deviation move. For the left jump probability, we fix it to 10%, so it indicates the probability of a price drop of more than (a constant) 10% over the coming week.

3 Technical Notes on Matlab Code

Two separate Matlab codes are used to compute the tail indices. The first takes raw data from OptionMetrics as input and creates a homogenous data file needed for the actual calculation. The second code then computes the actual indices from the standardized input from the first code. These Matlab codes are available for download on the website.

3.1 Parsing the Data: rawDataProcessing.m

The first program determines the available valid option quotes and organizes them into a convenient format for index computation in Matlab readable format. The data input consists of four separate csv files. We now describe these in turn.

The **Option data** file contains all option quotes in the date range. We require 14 data fields from the OptionMetrics database, as depicted in Figure 2

- Option Symbol
- Exercise Style
- Call/Put Flag
- Last Traded Date
- Open Interest
- Expiration Date
- Security ID
- Index Flag
- Issuer
- Highest Closing Bid
- Lowest Closing Ask
- Volume
- Strike Price Times 1000
- Implied Volatility

	2 date	3 symbol	4 exdate	5 last_date	6 cp_flag	7 strike_price	8 best_bid	9 best_offer	10 volume	11 open_interest	12 impl_volatility	13 optionid	14 index_flag	15 issuer	16 exercise_style
1	19960104	'098A3.1A'	19960316	19960104	'C'	600000	24.7500	25.7500	150	5633	0.1109	10003226	1	'CBOE S...	'E'
2	19960104	'09A49.0B'	19960316	19960104	'P'	615000	10.1250	10.8750	156	8302	0.1259	10111243	1	'CBOE S...	'E'
3	19960104	'098B6.4F'	19970621	19960104	'P'	550000	11.6250	13.2500	5	3282	0.1594	10008143	1	'CBOE S...	'E'
4	19960104	'099C9.98'	19960316	19960104	'P'	600000	6	6.5000	3917	5728	0.1369	10078616	1	'CBOE S...	'E'
5	19960104	'09C69.1E'	19960316	19960104	'P'	560000	1.7500	2	3	2971	0.1770	10250526	1	'CBOE S...	'E'
6	19960104	'0A02F.38'	19970621	19960104	'P'	500000	6	6.8750	10	1488	0.1759	10497848	1	'CBOE S...	'E'
7	19960104	'099D6.78'	19970621	19960104	'P'	700000	68	70	500	900	0.1227	10081912	1	'CBOE S...	'E'
8	19960104	'098D3.63'	19960217	19960104	'P'	620000	9.7500	10.2500	469	5621	0.1199	10015587	1	'CBOE S...	'E'
9	19960104	'0993A.AA'	19970621	NaN	'P'	525000	8.5000	9.3750	0	1522	0.1672	10042026	1	'CBOE S...	'E'
10	19960104	'09CFD.73'	19960622	NaN	'C'	475000	146.1250	147.1250	0	2700	NaN	10288499	1	'CBOE S...	'E'

Figure 2: Option data as read by Matlab

The second file is **Implied volatility**. It provides OptionMetrics extrapolated ATM implied volatility at the 30-day maturity. The data is used in calculating the benchmark to determine left jump events. The required data fields are,

- CUSIP

- Ticker

- SIC

- Index Flag

- Exchange Designator

- Class Designator

- Issue Type

- Industry Group

- Days to Expiration

- Delta

- Call/Put Flag

- Interpolated Implied Volatility

- Implied Strike Price

- Implied Premium

- Dispersion

The third file, **Underlying price**, provides prices for the relevant index/security. We request the following fields from OptionMetrics, see Figure 3:

- Ticker

- Bid/Low

- Ask/High

- Open Price

- Close Price

- Volume

- Return

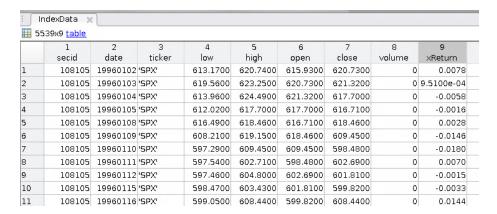|	1 secid	2 date	3 ticker	4 low	5 high	6 open	7 close	8 volume	9 xReturn
1	108105	19960102	'SPX'	613.1700	620.7400	615.9300	620.7300	0	0.0078
2	108105	19960103	'SPX'	619.5600	623.2500	620.7300	621.3200	0	9.5100e-04
3	108105	19960104	'SPX'	613.9600	624.4900	621.3200	617.7000	0	-0.0058
4	108105	19960105	'SPX'	612.0200	617.7000	617.7000	616.7100	0	-0.0016
5	108105	19960108	'SPX'	616.4900	618.4600	616.7100	618.4600	0	0.0028
6	108105	19960109	'SPX'	608.2100	619.1500	618.4600	609.4500	0	-0.0146
7	108105	19960110	'SPX'	597.2900	609.4500	609.4500	598.4800	0	-0.0180
8	108105	19960111	'SPX'	597.5400	602.7100	598.4800	602.6900	0	0.0070
9	108105	19960112	'SPX'	597.4600	604.8000	602.6900	601.8100	0	-0.0015
10	108105	19960115	'SPX'	598.4700	603.4300	601.8100	599.8200	0	-0.0033
11	108105	19960116	'SPX'	599.0500	608.4400	599.8200	608.4400	0	0.0144

Figure 3: Underlying price data as read by Matlab

The key information in this file is the "Close Price." Data from other sources are acceptable, but may necessitate minor adjustment in the code.

The fourth and last file, **Discount factor** is the interest rate series. We use the 30-day US Treasury bill series from WRDS-Federal Reserve Bank in our calculation. Any other discount factor may be substituted in its place. The csv file should contain two columns: column 1 is date in 'yyyymmdd' format, and column 2 is the interest rate series.

Overall, the Matlab program, exploiting these four files, generates a .mat data file with aligned option and ATM implied volatility data. The array 'alignedData' contains 13 columns:

- Date - in Matlab datenum format

- Call/Put flag - 1 for call and 0 for put

- Weekly option identifier

- Mid price

- Expiry date - in Matlab datenum format

- Strike price

- Spot(close) price of underlying

- Implied volatility

- Volume traded of the option contract on the specific date

- Bid-ask spread at close

- Open interest of option contract at close

- Last traded date of option contract in Matlab datenum format

- Discounting interest rate

7

3.2 Calculating Indices: tailMeasureCalculation.m

The second Matlab program takes the 'alignedData' as input and computes the indices. The calculation is performed in two steps - we first estimate the α_t^- series throughout the sample period, and then calculate ϕ_t^- and the associated tail indices.

The output consists of four Matlab arrays. The first one, 'result', summarizes the raw (un-smoothed) estimates for the parameters and indices:

- Date - in Matlab datenum format

- Put Count - number of valid put option quotes in the α estimation period, week or day

- Call Count - number of valid call option quotes in the α estimation period, week or day

- α^-: left tail shape parameter

- α^+: right tail shape parameter (not used in left tail indices)

- ϕ^-: left tail level parameter

- ϕ^+ : right tail level parameter (not used in left tail indices)

- Left jump intensity - See equation (5.1) in Bollerslev, Todorov and Xu (2015)

- Right jump intensity - See equation (5.1) in Bollerslev, Todorov and Xu (2015)

- Left jump variation: LJV^2 series without smoothing

- Right jump variation: RJV^2 series without smoothing (not used in left tail indices)

The other three arrays stores the indices exhibited for each security on the website: LJV series in 'LJVolatility', smoothed LJV series in 'LJVolatilityMA', and smoothed probability of a 10% weekly down move in 'leftDensityFixedMA'.