1 Introduction to Tail Index

Volatility indices, VIX in particular, have drawn unprecedented attention of 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 their derivatives. The methodology by which VIX and its analogs are derived captures implied volatility of the underlying asset as a whole. However, it has been noticed by both the academia and industry alike that security prices are exposed to two different types of risk: diffusion risk and jump risk. Diffusion refers to the continuous fluctuation, while jump risk realizes when a market event leads to a sudden crash or spike in underlying price. Traditional risk measures do not differentiate them.

Based on the paper *Tail risk premia and return predictability* (Bollerslev, Todorov and Xu, Journal of Financial Economics, 2015), we calculate two new measures of market risk, so as to provide a characterization of jump risk isolated from diffusion risk. **Left Jump Variation** estimates the variation of size of a possible negative jump, and **Left Jump Probability** estimates the possibility of it. Capture only the risk of a plunge in price, they have good indicative power on future market movement.

The tail index are calculated for the most popular US stock indices including S&P 500 Index, Dow Jones Industrial Average, NASDAQ 100 Index, Russell 2000 Index, as well as a number of top ETFs. The same methodology could be applied on other assets or indices with a liquid option market.

2 Calculation

2.1 Data

The methodology is based on a non-parametric model for short maturity options. We select those with maturities between 6 and 31 trading days, including both monthly and weekly options. Options with shorter maturities are excluded to avoid excessive noise in data.

In order to capture the jump risk only and eliminate diffusion risk in the tail measure, we select only deep out-of-the-money (OTM) put options, which bears its value almost solely on negative jump. The threshold of deep OTM is defined by maturity-adjusted log-moneyness less than or equal to -2.5,

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

where

$$K = \text{strike price}$$

$$F_{t,\tau} = \text{forward price of index at t, maturity } \tau$$

$$\tau = \frac{\text{number of trading days to maturity}}{252}$$

$$\sigma_{imp,\tau} = \text{at-the-money implied volatility}$$

Time to maturity τ is measured in terms of business days as fraction of a year. By adjusting the moneyness with implied volatility and option maturity, the threshold is more dynamic and robust under different market conditions.

Deep OTM options are often less liquid and have wider bid-ask quotes. We select only those with positive bid price to exclude the most illiquid ones. We then take the mid-price, i.e. average of bid and ask, as value of option. To avoid violation of no arbitrage principle, if a put option with lower strike price has higher or equal value than one with higher strike price, we would consider it as an outlier and exclude from our data set. Working from higher to lower strike price, we seek a strictly decreasing sequence of market value. See below for an example,

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
0.125	1645	-5.48	OTM
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
0.60	1740.00	-3.06	OTM
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 2014/05/06 option quotes, on 2014/05/17. Option quotes highlighted in yellow are valid to calculate the index. In practice, we pool data with multiple maturities for more robust estimation.

At-the-money implied volatility, σ_{imp}, is estimated by a weighted average of implied volatility of highest strike OTM put option and lowest strike OTM call option. The formula is given by,

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

where K_C and K_P are the strike price of the OTM call and put option respectively with strike price closest to forward index. $\sigma_{imp,C}$ and $\sigma_{imp,P}$ are Black-Scholes implied volatility of them.

2.2 Calculation

Two parameters are estimated in order to calculate the tail measures, tail shape parameter α^- and level shift parameter ϕ^-. They are estimated by;

$$\hat{\alpha_t}^- = \underset{2\leq n\leq N}{\text{median}}(|1 - \frac{\ln\frac{O_{t,\tau}(k_{t,i})}{O_{t,\tau}(k_{t,i-1})}}{k_{t,i} - k_{t,i-1}}|)$$

$$\hat{\phi_t}^- = \underset{1\leq n\leq N}{\text{median}}(|\ln\frac{e^{r_{t,\tau}}O_{t,\tau}(k_{t,i})}{\tau F_{t,\tau}} - (1+\hat{\alpha_t}^-)k_{t,i} + \ln(\hat{\alpha_t}^- + 1) + \ln(\hat{\alpha_t}^-)|)$$

where

$$N = \text{number of valid option quotes}$$

$$k_{t,i} = \ln(K_{t,i}/F_{t,\tau}) = \text{unadjusted log-moneyness of i-th option}$$

$$O_{t,\tau}(k_{t,i}) = \text{mid-price of i-th option}$$

$$r_{t,\tau} = \text{riskless interest rate with maturity } \tau$$

We use 3 month US treasury rate as our riskless rate in the calculation. The market depth and width of index options increased drastically in the last 10 years. To get more consistent estimation with less noise in earlier periods, we pool the data over each week to calculate α^- before 2013. From 2013, the market is liquid enough such that we could calculate α^- on a daily basis. For ETFs, due to less liquidity, we estimate over each week throughout the sample period. In both cases, we calculate only when we have at least 4 pairs of $[O_{t,\tau}(k_{t,i}), O_{t,\tau}(k_{t,i-1})]$ in the weekly or daily sample to achieve better robustness. ϕ^-, on the other hand, is less sensitive to extreme sample points. We estimate it daily throughout the sample.

With $\hat{\alpha_t}^-$ and $\hat{\phi_t}^-$, the tail measures Left Jump Variation and Left Jump Probability are given by

$$\text{LJV} = \hat{\phi_t}^- \cdot e^{-\hat{\alpha_t}^-|\theta_t|} \cdot \frac{\hat{\alpha_t}^-|\theta_t|(\hat{\alpha_t}^- k_t + 2) + 2}{(\hat{\alpha_t}^-)^3}$$

$$\text{Left Jump Probability} = \hat{\phi_t}^- \cdot \frac{e^{-\hat{\alpha_t}^-|\theta_t|}}{\hat{\alpha_t}^-}$$

Here, θ_t is a threshold of negative tail jump on 1 week horizon. For LJV, we define θ_t as $10 \cdot \sigma_{imp,30d} \cdot \sqrt{5/252}$, i.e. ten times standard deviation move. For left jump probability, we calculate two measures with $\theta_t = 10 \cdot \sigma_{imp,30d} \cdot \sqrt{5/252}$ and 10% respectively. The former is normalized to implied volatility and captures the tail event dynamically. The latter measures the probability of a plunge more than or equal to a constant 10% barrier over time.

3 Technical Notes on Matlab Code

A set of Matlab code implements the above model. It takes raw data from OptionMetrics as input. The first program converts the data into Matlab readable format, and a second program conducts the calculation.

The data input consists of four csv files. **Option data** contains all option quotes in the date range. We request 14 data fields from OptionMetrics:

- 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

OptionData_RAW

12366633x16 table

	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

Implied volatility is OptionMetrics extrapolated ATM implied volatility at 30 day maturity. The data is used in calculating the benchmark to determine left jump events. Request 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

Underlying price is the price series of the underlying index/security. We request the following fields from OptionMetrics:

- Ticker
- Bid/Low
- Ask/High
- Open Price
- Close Price
- Volume
- Return

IndexData

5539x9 table

	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 as well with minor adjustment in the data conversion program.

The last file, **discounting factor** is the interest rate series. We use 30-day US Treasury in our calculation. Our data source is WRDS-Federal Reserve Bank. Any other source works in the same manner. The csv file should contain two columns: column 1 is date in 'yyyymmdd' format, and column 2 is the interest rate series.

The first program results in a .mat data file with aligned option data and ATM implied volatility. 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

The second program implements the model and outputs in a 'result' table, with columns:

- 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^-
- α^+
- ϕ^-
- ϕ^+
- 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)
- Right jump variation (RJV)

Moving average series for LJV and left jump probability, as the final result, are stored in 'LJVMA' and 'leftDensityFixedMA'.