The Puzzle of Filtering Index Options

University of Chicago, Winter 2024*

Viren Desai, Harrison Holt, Ian Hammock

March 6, 2024

Abstract

In this report we summarize our efforts to replicate the data filtration process described in Appendix B of *The Puzzle of Index Option Returns* by Constantinides, Jackwerth, and Savov (2013). These filters shape the underlying distribution of implied volatility ("IV") and moneyness for a large cross-section of SPX index options (1 million+), and were used to build and price 54 option portfolios in the original paper. Due to the unavailability of SPX option data from 1985 to 1995, we focus our analysis on replicating the filtration results OptionMetrics data from 1996-01 to 2012-01. We then apply these filters to more recent data from 2012-02 to 2019-12. Through a sequence of data visualizations, we show that while the paper's intricately constructed data filters may yield elegant results when applied to one time period, these results do not necessarily port over to other time periods. The implications for option pricing models based on such time-fragile data filters would be an

*Final project for FINM 329; taught by Jeremy Bejarano.

interesting follow-up study. Our detailed analysis and code can be readily found on Github[1].

[1] https://tinyurl.com/3psws69d

1 Replicating Table B1

Appendix B of Constantinides, Jackwerth, and Savov (2013) outlines three levels of filters applied to millions of SPX call and put options with an intent to minimize quoting errors in the construction of the paper's 54 option portfolios. In this report we will summarize our implementation, and briefly discuss the challenges and differences encountered in our attempted replication and the subsequent reproduction for a later time period. Our results are summarized in Table 2.

1.1 Level 1 Filters

Methodology The Level 1 filters comprised of an "Identical" filter, to filter out duplicate options in the OptionMetrics data (measured by identical option type, strike, expiration date, and price), and an "Identical Except Price" filter (the "IEP filter"), which aimed to filter out options identical in all respects except price. In these cases, the paper's authors retained options whose quoted T-bill-based IV was closest to its moneyness neighbors, and deleted the rest.

Commentary on Results Two issues arose when applying this set of filters. First, there were quite a few options in the IEP filter that have no reported IV. While this was not explicitly addressed in the paper's description of the Level 1 filters, it created an issue with replication because the T-bill-based IV was required to identify the nearest moneyness neighbors. The missing IV option count here was limited to about 5 in the 1996-2012 dataset, but in the 2012-2019 dataset we observed 355,896 options with no reported IV. For the purpose of this filter, if an IV was not reported, it was not chosen as the option with IV closest to the at the money. Additionally, if an option group's "at the money" member could not have their IV calculated by numerical methods (described later in this report), all options in that group would be subsequently dropped via the "Unable to compute IV" filter.

Second, an unexplainable difference occurred upon the application of the Volume = 0 filter. In Table B1 of Constantinides, Jackwerth, and Savov (2013), no options have a Volume = 0 in their dataset. However, we observed over two millions options with a Volume = 0. Unfortunately, no more details were given in the manuscript describing this step. In order to not diverge from their data pool we chose to drop 0 options, this is reflected in Table 2.

Further data is included in Appendix B.

1.2 Level 2 Filters

Methodology The Level 2 filters comprised of the following:

- *Days to Maturity <7 or >180*: The objective of the Days to Maturity filter was to exclude options with less than a week to expiration (typically exhibiting erratic price behavior) or more than 180 days to expiration (typically exhibiting low liquidity).

- *IV <5% or >100%*: The objective of Level 2 IV filter was to exclude options with extreme IV values.

- *Moneyness <0.8 or >1.2*: The objective of the Moneyness filter was to exclude options with extreme moneyness values, due to these options having low liquidity and little value beyond the intrinsic value of the option.

- *Implied Interest Rate <0*: The objective of the Level 2 Implied Interest Rate filter was to exclude options with negative implied interest rates, which are likely due to quoting errors. To calculate the implied interest rate, we compute the put-call parity implied interest rate using each option pair's bid-ask midpoints as the price. The put-call parity implied interest rate is the interest rate (r) that makes the put-call parity equation hold (Equation 1).

- *Unable to compute IV*: The objective of this filter was to exclude options where the IV was not computable, due to missing or extreme parameters for numerical option price solvers. To note, we utilize the Black-Scholes-Merton option pricing formulae for European options[2].

Commentary on Results For the purposes of the IV filters, we tested multiple numerical methods for computing IV, including binary search, Newton-Raphson, and quasi-Newton methods using the `scipy.optimize` package in Python. We found that no single method could compute IV for all options, in all time periods, and that the percentage of options with incomputable IVs increased as time to maturity decreased. Importantly, the reasons for inability to compute IV were not explicitly

[2] The Black-Scholes formulae for the price of European call (C) and put (P) options are given by $C = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$ and $P = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1)$, respectively, where $d_1 = \frac{\ln(\frac{S_0}{K})+(r+\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$. Here, S_0 is the current stock price, K is the strike price, r is the risk-free interest rate, σ is the volatility of the stock's return, T is the time to expiration, Φ is the cumulative distribution function of the standard normal distribution.

addressed in the paper, and the impact of these options on the final option portfolios is unclear.

@Harrison Holt: Please send commentary / challenges on your results to @Viren. bullets are fine, I'll reword. just want to capture major points.

Further data is included in Appendix C.

1.2.1 1996-01 to 2012-01

figure tags: Figure 1 Figure 2 Figure 3 Figure 4 Figure 5

1.2.2 2012-02 to 2019-12

figure tags: Figure 6 Figure 7 Figure 8 Figure 9 Figure 10

1.3 Level 3 Filters

Methodology The Level 3 filters are comprised of an implied volatility filter (the "IV filter") and a put-call parity implied interest rate filter (the "put-call parity filter", or "PCP filter"). The Level 3 filters described in the paper were not as straightforward to replicate as the previous two levels. In particular, the intricate construction of these filters, and the lack specificity regarding critical filter parameters in Constantinides, Jackwerth, and Savov (2013), made the filtered option counts highly sensitive to our assumptions for these parameters.

Importantly, since the IV and PCP filters (as well as the Level 1 and Level 2 filters) are applied sequentially, differences in replication get compounded downstream.

- *IV Filter*: The objective of the IV filter was to reduce the prevalence of apparent butterfly arbitrage. A butterfly arbitrage occurs when there is a discrepancy in the IV structure across difference moneyness levels for options having the same expiration date. The construction of the IV filter involved fitting a quadratic polynomial to the observed log volatilities of puts and calls separately (a computationally intensive task). The original paper then measured the relative distance in percent between the fitted log IVs and the observed log IVs, grouped the options into moneyness bins, calculated the standard deviation of the entire sample of relative distances by moneyness bin, and then filtered out options where the relative distance was greater than a certain threshold.

- *PCP Filter*: The objective of the PCP filter was to ensure that put-call parity held for every put-call option pair with the same date, expiry date, and moneyness. This was done by utilizing the put-call parity equation (Equation 1) to calculate the implied interest rate based on each option pair's bid-ask midpoint prices.

$$C - P = S - Ke^{rT} \tag{1}$$

$$e^{rT} = \frac{(S - C + P)}{K} \tag{2}$$

$$r = \frac{1}{T} \cdot \log\left(\frac{S - C + P}{K}\right) \tag{3}$$

Commentary on Results Constantinides, Jackwerth, and Savov (2013) do not provide detail on the relative distance algorithms utilized and are also unclear on what standard deviation threshold was used for exclusion with the Level 3 filters. For our base case replication, we assumed a percentage relative distance, but sensitized the filtered option counts to Manhattan (absolute) distance, and Euclidean distance. For the standard deviation threshold base case, we excluded options >2 standard deviations away, but sensitized 2.0 to 5.0 standard deviations. Despite these efforts, we found that no combination of parameters could replicate the filtered option counts in the original paper, likely due to differences in the core assumptions of the relative distance algorithms and standard deviation thresholds, and potentially due to downstream effects from the Level 1 and Level 2 filters.

appendix: Appendix D

1.3.1 1996-01 to 2012-01

figure tags: Figure 11 Figure 12 Figure 13 Figure 14

1.3.2 2012-02 to 2019-12

figure tags: Figure 15 Figure 16 Figure 17 Figure 18

2 Replicating Table2

Table 3 describes how many options are found, go missing, or expire in the dataset. An option is found if it reappears the next trading day. An option is missing for if it does not reappear the next trading day. Multiple days missing, counts as multiple options missing. Lastly, if an option is lost and expires this is noted as expired.

We would like to note an interesting aspect of the 1996-01 to 2012-01 dataset. Over 80% of the options expire on a Saturday or a non-trading day. To handle this, we push the expiration day to the nearest Friday, presumably the nearest trading day. However, there are quite a few edge cases which would explain the discrepancy between our analysis and Constantinides, Jackwerth, and Savov (2013). Further investigation is required. A short summary of the day distribution is given in Table 1.

Table 1: Option Expiration days

	1996-01 to 2012-01	2012-02 to 2019-12
Total Options	461890	3164202
Trading Days	10%	86%
Saturdays	87%	12%
Other Days	3%	2%

Trading days are determined by the NYSE calendar provided by pandas market days.

3 Conclusion

Our findings illustrate that seemingly straight forward instructions for filtering data may lead to divergent results. To reduce these errors, we suggest that journals require that computationally intense manuscripts, such as Constantinides, Jackwerth, and Savov (2013), publish their code base along with their findings. Our data acquisition is briefly described in Appendix A, and our code base can be found our Github [3].

[3] https://github.com/harrypandas/finm-32900_final_project.git

4 References

Constantinides, George M., Jens Carsten Jackwerth, and Alexi Savov. 2013. "The Puzzle of Index Option Returns." *The Review of Asset Pricing Studies* 3 (2):229–257. URL https://doi.org/10.1093/rapstu/rat004.

Table 2: Table B1 Summary

		OptionMetrics: 1996-01 to 2012-01		OptionMetrics:2012-02 to 2019-12		Total	
		Deleted	Remaining	Deleted	Remaining	Deleted	Remaining
Starting	Calls		1,704,220		7,901,901		9,606,121
	Puts		1,706,360		7,901,427		9,607,787
	All		3,410,580		15,803,328		19,213,908
Level 1 filters	Identical	0		277,102		277,102	
	Identical except price	10		2,557,330		2,557,340	
	Bid = 0	272,078		1,069,116		1,341,194	
	Volume = 0	0		0		0	
	All		3,138,492		11,899,780		15,038,272
Level 2 filters	Days to expiration <7 or >180	1,297,729		3,080,910		4,378,639	
	IV <5% or >100%	16,432		63,639		80,071	
	K/S <0.8 or >1.2	550,227		1,987,486		2,537,713	
	Implied interest rate < 0	642,940		2,053,938		2,696,878	
	Unable to compute IV	37,733		385,913		423,646	
	All		593,431		4,327,894		4,921,325
Level 3 filters	IV filter	38,568		312,899		351,467	
	Put-call parity filter	92,973		850,793		943,766	
	All		461,890		3,164,202		3,626,092

Number of observations that are removed upon application of Appendix B filters.

Table 3: Table 2 Results

Observations	Calls				Puts			
	1996-01 to 2012-01		2012-02 to 2019-12		1996-01 to 2012-01		2012-02 to 2019-12	
All trading days								
Found	176,225	79%	1,185,069	76%	176,225	79%	1,185,069	76%
Missing	6,867	3%	21,026	1%	6,867	3%	21,026	1%
Expired	40,298	18%	352,870	23%	40,298	18%	352,870	23%
Last trading day of the month								
Found	19,126	82%	280,486	80%	19,126	82%	280,486	80%
Interpolated	4,104	18%	69,727	20%	4,104	18%	69,727	20%

Tracking the instances options are found, missing or expired.

A Data

Our option data is queried from OptionMetrics provided by Wharton Research Data Services (WRDS). We limit the query to SECID = 108105, S&P 500 Index - SPX. We use the three month Tbill as our interest rate, this is from the Federal Reserve Board's H15 report supplied by WRDS.

In comparison to their data, we have pulled 184 more options than them. It is unclear where the discrepancy lies. We assumed we were off by a day however this will truncate or elongate the dataset by over 300 points. We credit the discrepancy to OptionMetrics updating their data to be more accurate.

The following links contain the documentation and helpful links for the WRDS database:

- Option Metrics Overview
- Option Metric Keys
- Option Metrics Query
- Federal Reserve Report

B Level 1 Filter

Table 4: 1996-01 to 2012-01 Summary of Options with No Volume Nor Open Interest

	Before (N)	Before (%)	After B1 (N)	After B1 (%)
Volume = 0	2,324,250	68	226,205	49
Open Interest = 0	819,542	24	102,616	22
Overlap	786,912	23	93,842	20

Number of observations that remain in the 1996-01 to 2012-01 data with volume and open interest equal to zero, as well as the overlap.

Table 5: 2012-02 to 2012-01 Summary of Options with No Volume Nor Open Interest

	Before (N)	Before (%)	After B1 (N)	After B1 (%)
Volume = 0	11,482,922	73	1,921,400	61
Open Interest = 0	5,877,697	37	1,099,423	35
Overlap	5,667,076	36	1,022,397	32

Number of observations that remain in the data with volume and open interest equal to zero, as well as the overlap.

Table 4 and Table 5 illustrate that many of the options in the initial and post filtering dataset have no trading volume nor open interest. Importantly, but perhaps not unexpectedly, nearly half of the options in the 1996-01 to 2012-01 dataset have no trading volume, and around 20% of options in this dataset have no open interest. The impact of these options in the portfolios of Constantinides, Jackwerth, and Savov (2013), if they are not filtered out of the final analysis, are unclear to us and merits further investigation.

C Level 2 Filter

C.1 1996-01 to 2012-01

C.1.1 Effects of filtering Days to Maturity <7 or >180

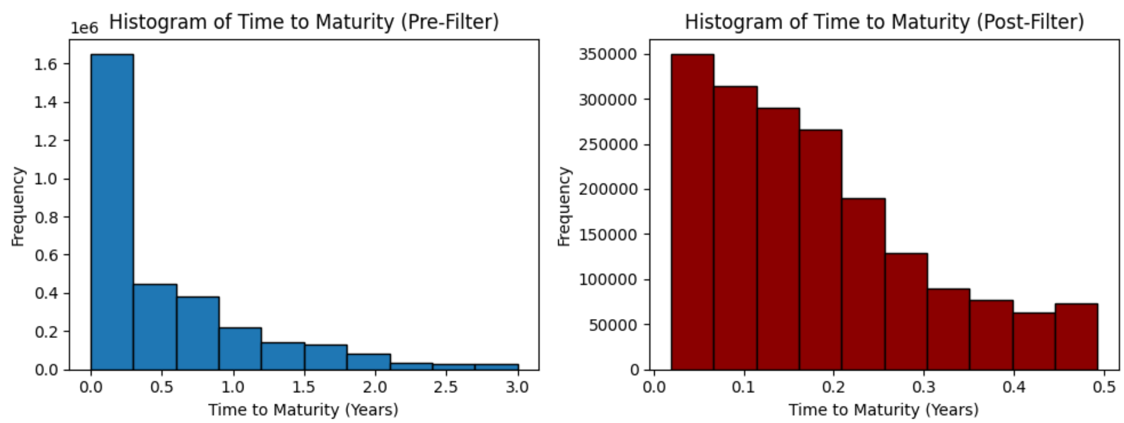

Figure 1: Distribution of time to maturity, measured in years from option initial date to expiration date. The graph on the left shows the distribution prior to applying the initial level 2 filter of excluding days to maturity less than 7 and greater than 180. Right shows distribution post filter.

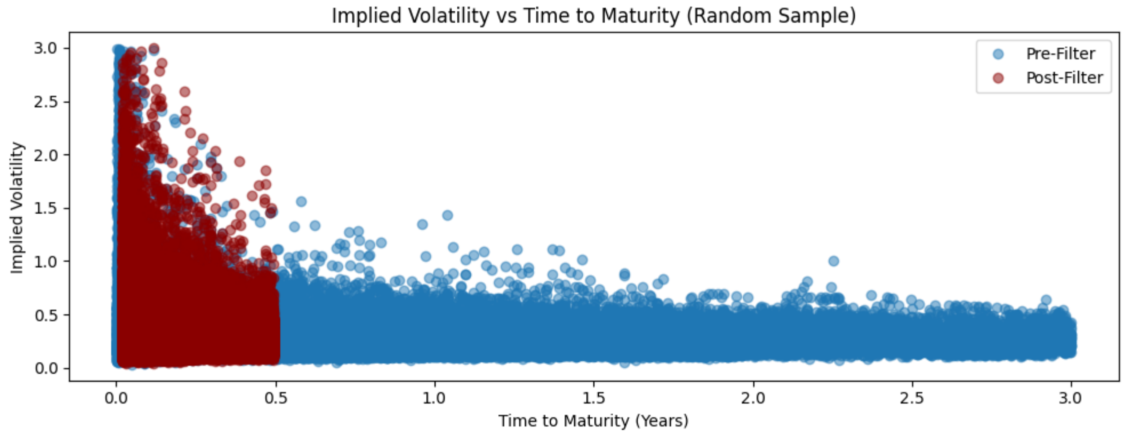

Figure 2: As noted in the paper, the short maturity options tend to move erratically nearing expiration. In Figure 2, post-filter (red) we see a slight reduction of short-term options with a high implied volatility.

C.1.2 Effects of filtering IV <5% or >100%

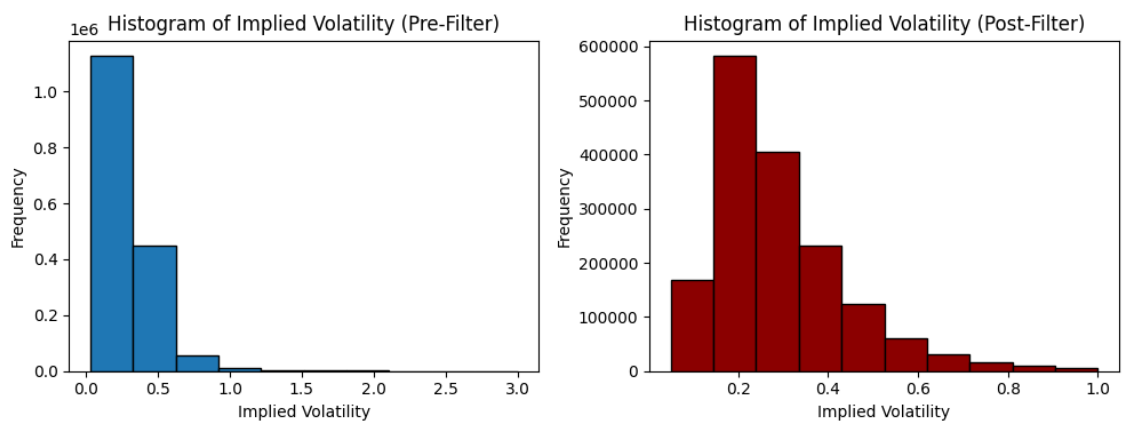

Figure 3: Removing option quotes with implied volatilities lower than 5% or higher than 100% eliminates extreme values and reduces the skewness of the implied volatility distribution.

C.1.3 Effects of filtering on Moneyness <0.8 or >1.2

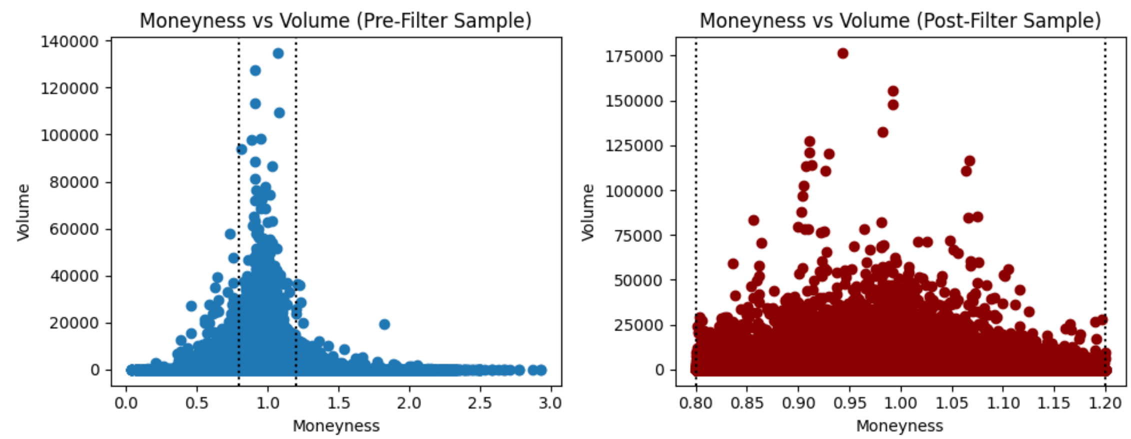

Figure 4: Removing option quotes with moneyness lower than 0.8 and higher than 1.2 eliminates extreme values. These extreme values potentially have quotation problems or low values.

C.1.4 Effects of filtering out options where we could not compute IV

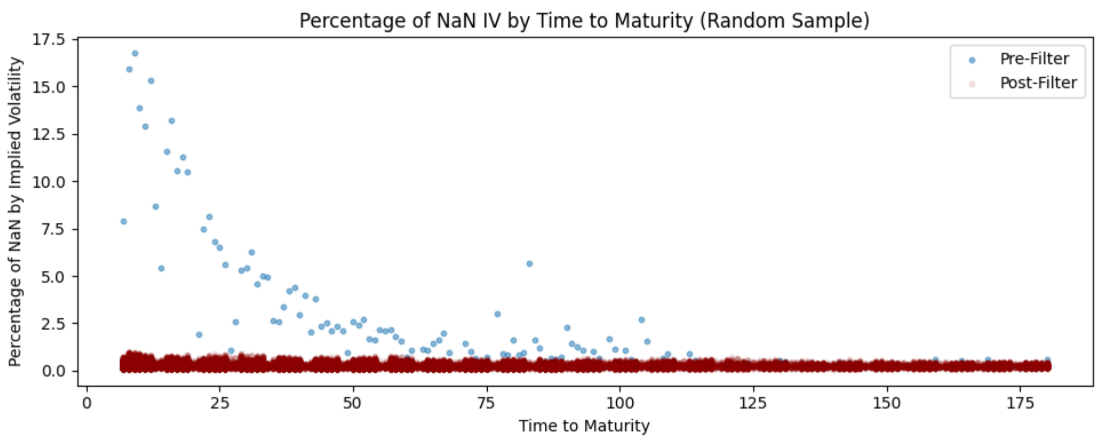

Figure 5: Through our analysis we found there are cases where we could not compute implied volatility (IV), as a result, the values were NaN. In figure above, there is a clear trend where the percentage of incomputable IVs increase as time to maturity decreases.

C.2 2012-02 to 2019-12

C.2.1 Effects of filtering Days to Maturity <7 or >180

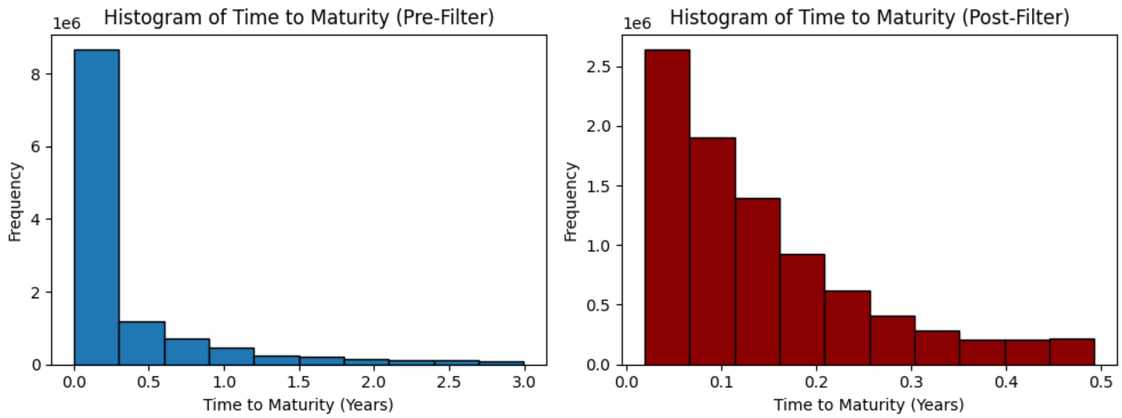

Figure 6: Distribution of time to maturity, measured in years from option initial date to expiration date. The graph on the left shows the distribution prior to applying the initial level 2 filter of excluding days to maturity less than 7 and greater than 180. Right shows distribution post filter.

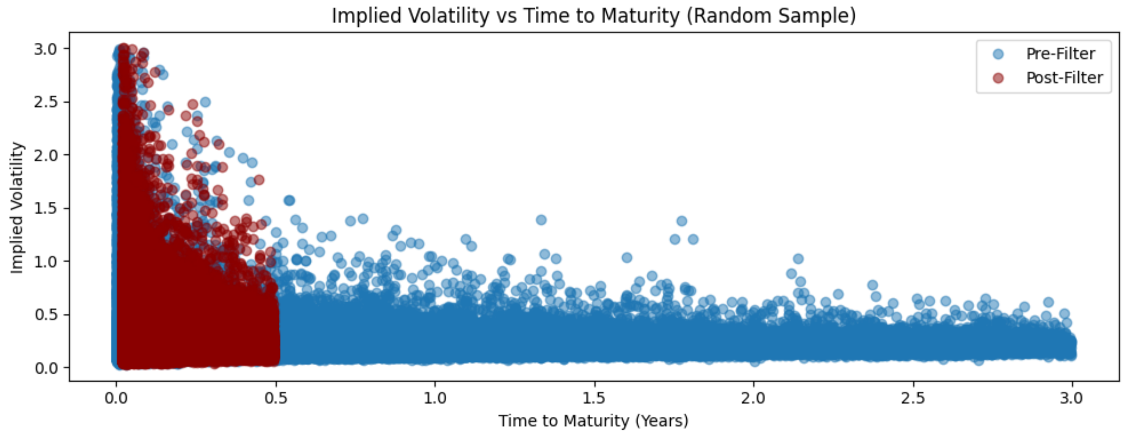

Figure 7: As noted in the paper, the short maturity options tend to move erratically nearing expiration. In Figure 2, post-filter (red) we see a slight reduction of short-term options with a high implied volatility.

C.2.2 Effects of filtering IV <5% or >100%

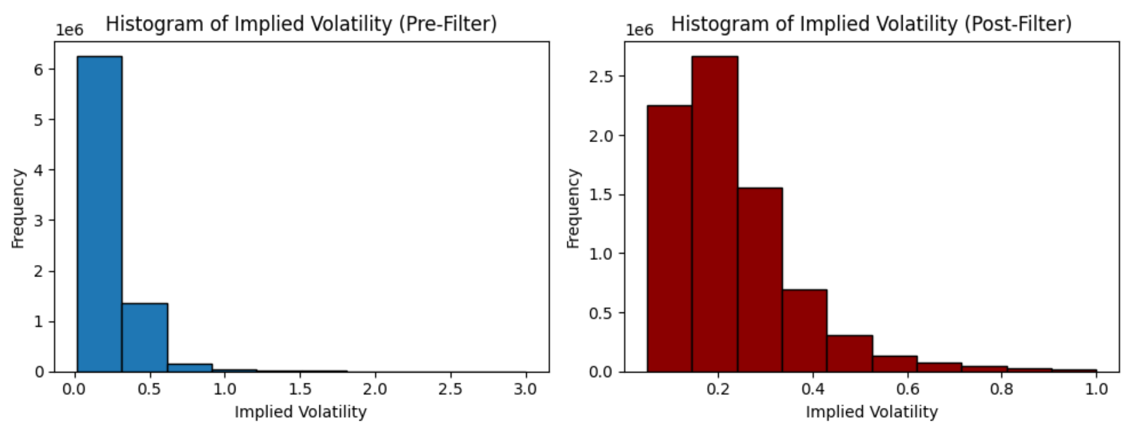

Figure 8: Removing option quotes with implied volatilities lower than 5% or higher than 100% eliminates extreme values and reduces the skewness of the implied volatility distribution.

C.2.3 Effects of filtering on Moneyness <0.8 or >1.2

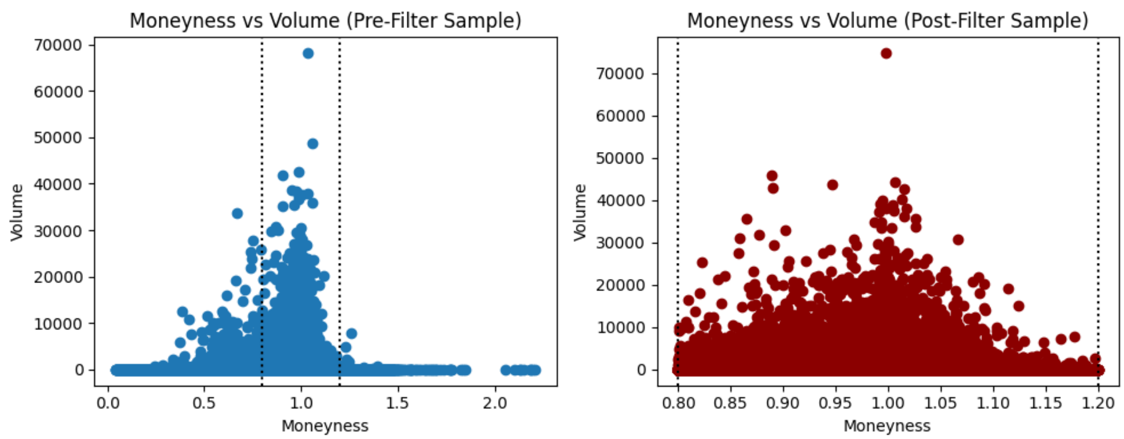

Figure 9: Removing option quotes with moneyness lower than 0.8 and higher than 1.2 eliminates extreme values. These extreme values potentially have quotation problems or low values.

C.2.4 Effects of filtering out options where we could not compute IV

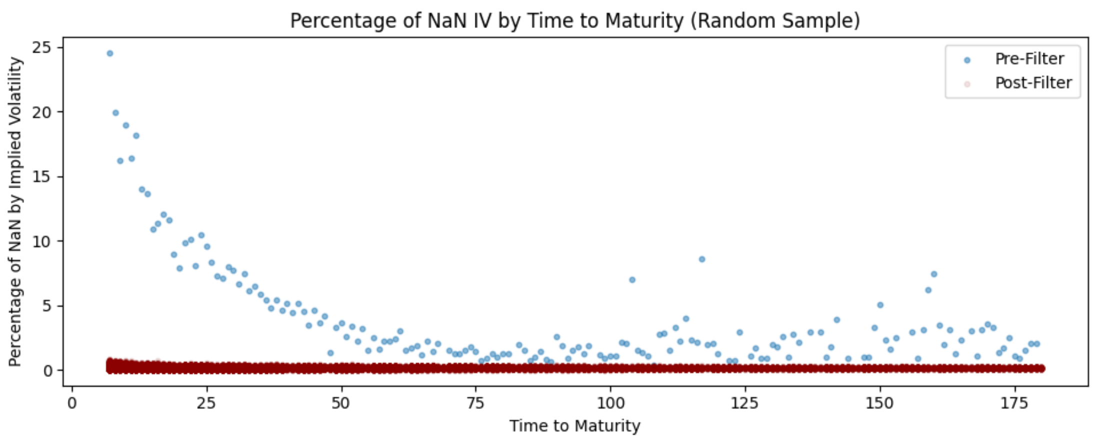

Figure 10: Through our analysis we found there are cases where we could not compute implied volatility (IV), as a result, the values were NaN. In figure above, there is a clear trend where the percentage of incomputable IVs increase as time to maturity decreases.

D Level 3 Filter

D.1 1996-01 to 2012-01

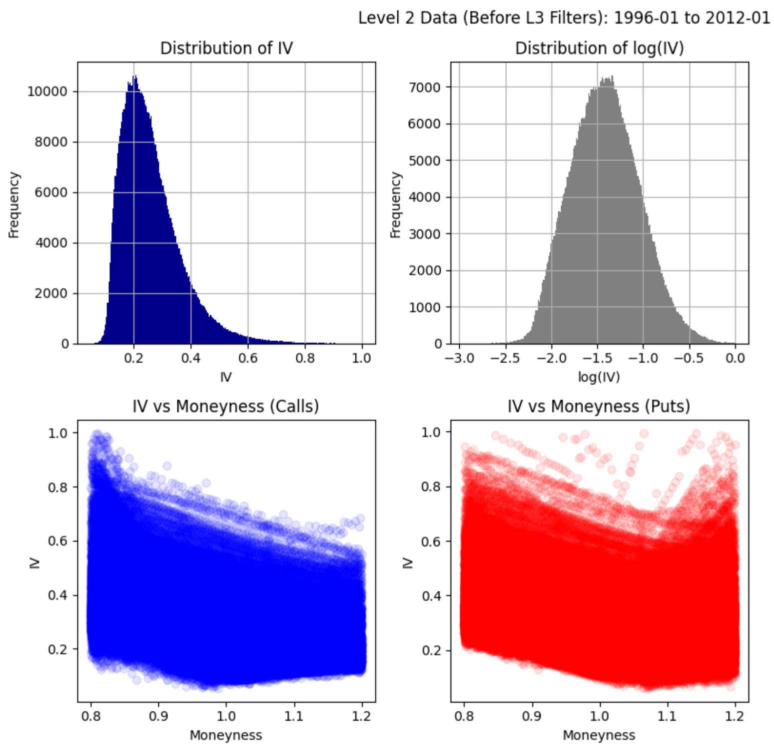

Figure 11: Your caption here

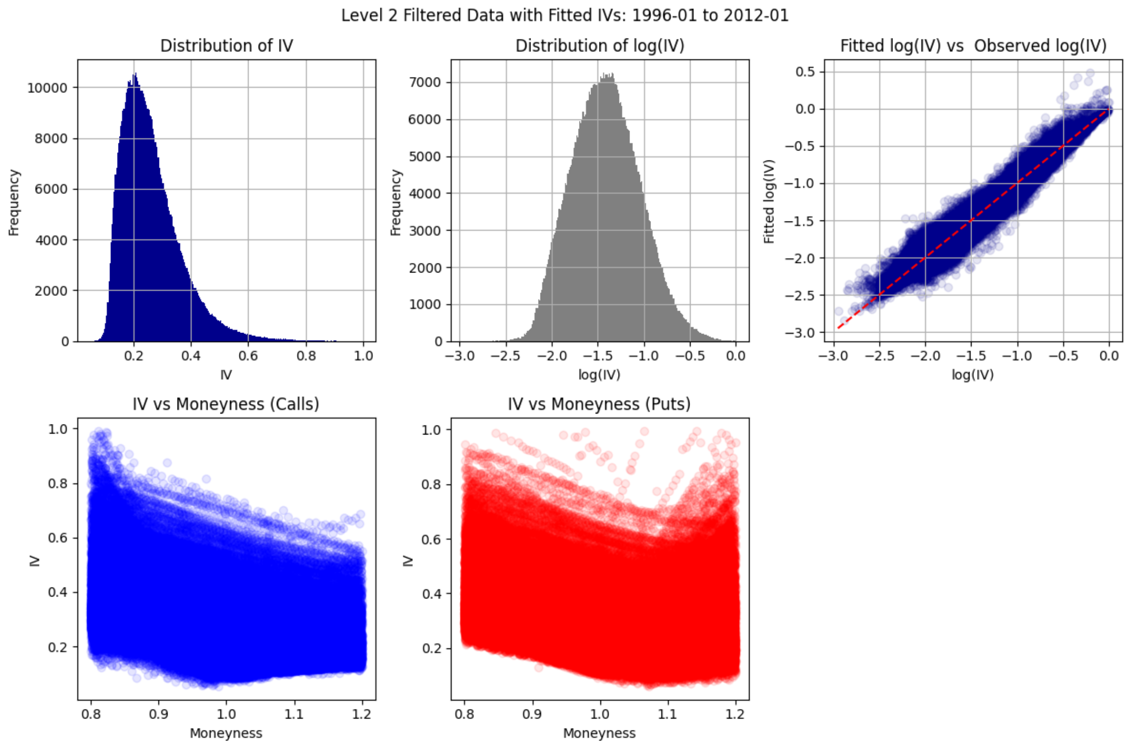

Figure 12: Your caption here

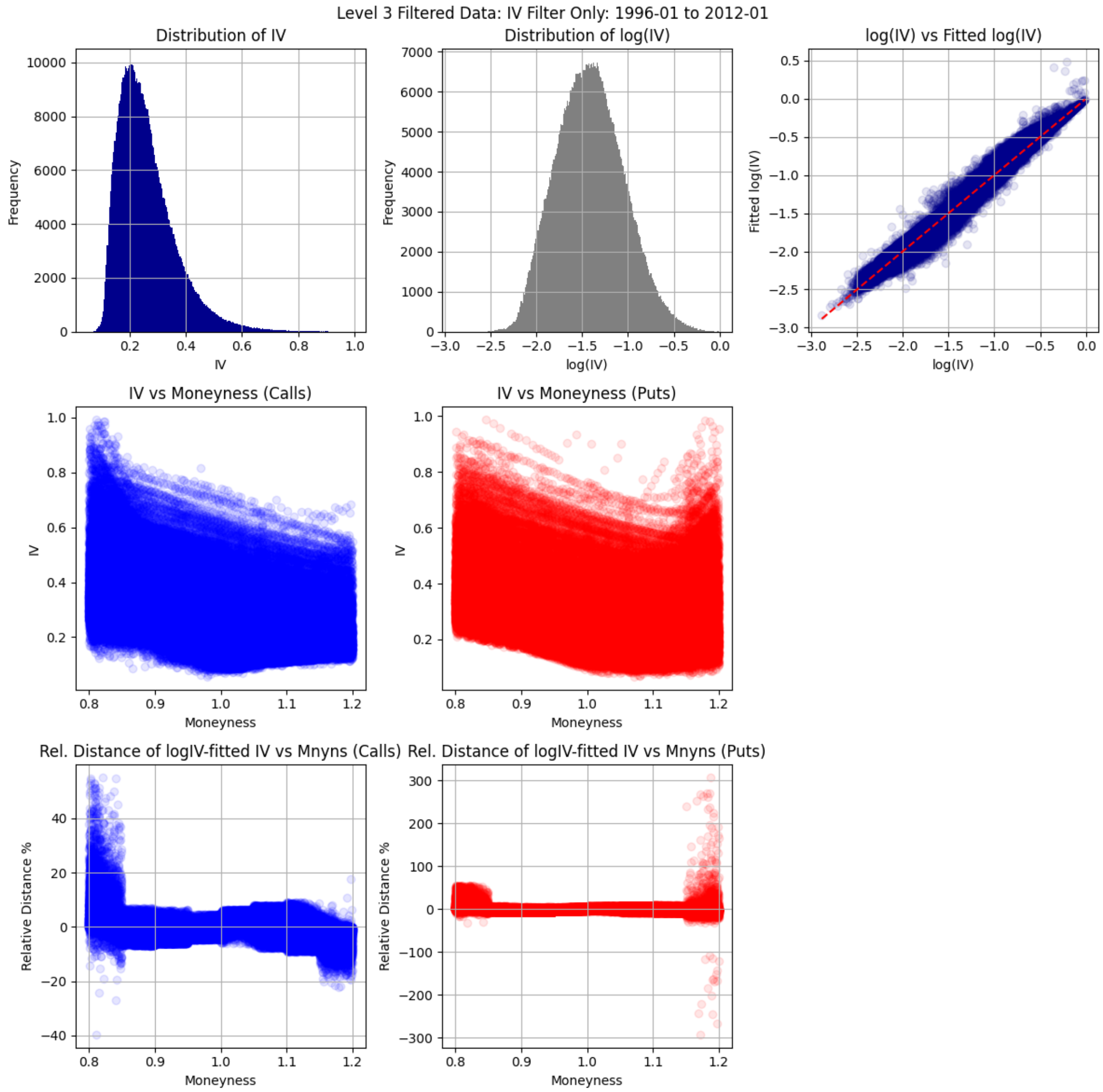

Figure 13: Your caption here

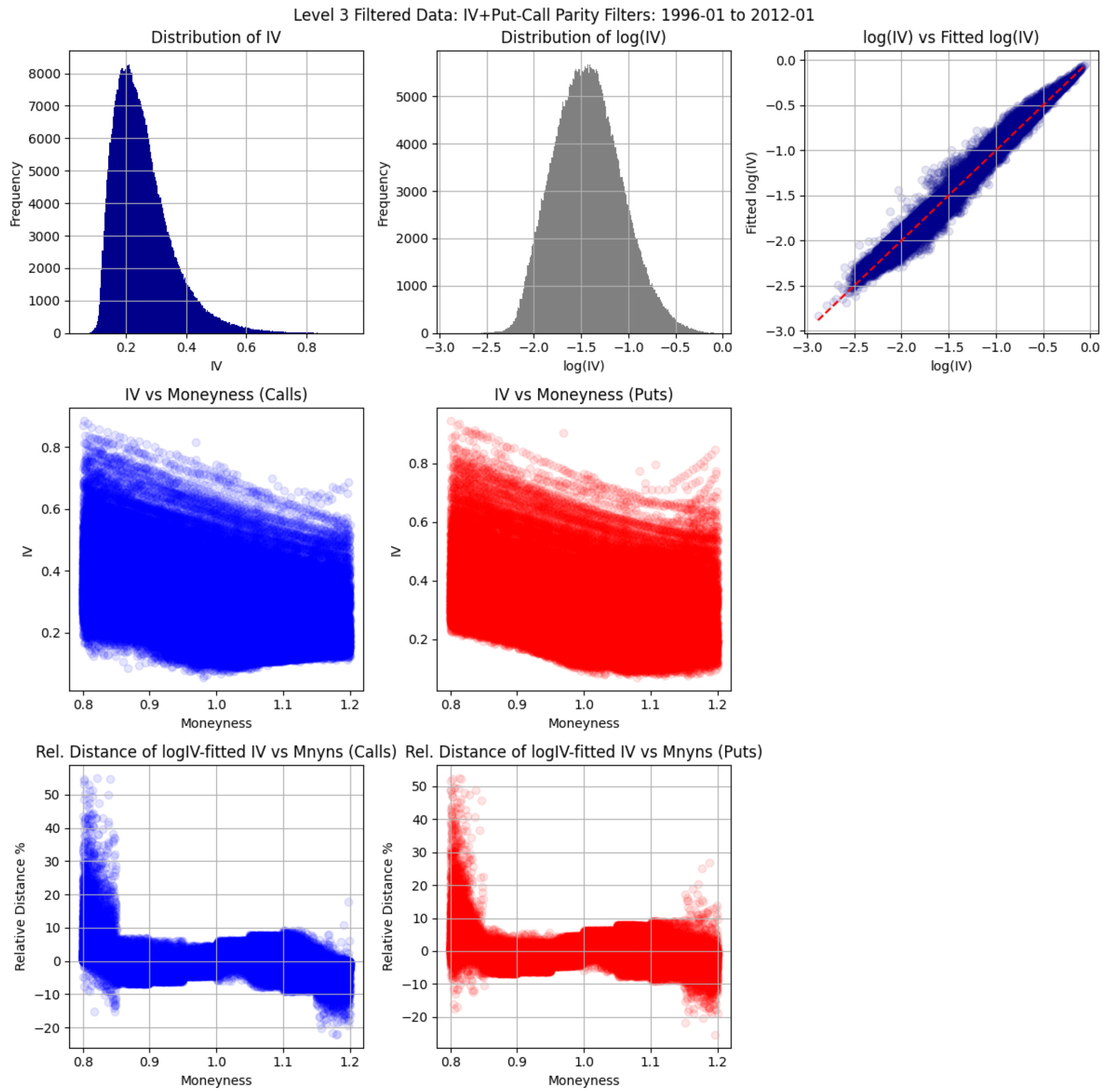

Figure 14: Your caption here

D.2 2012-02 to 2019-12

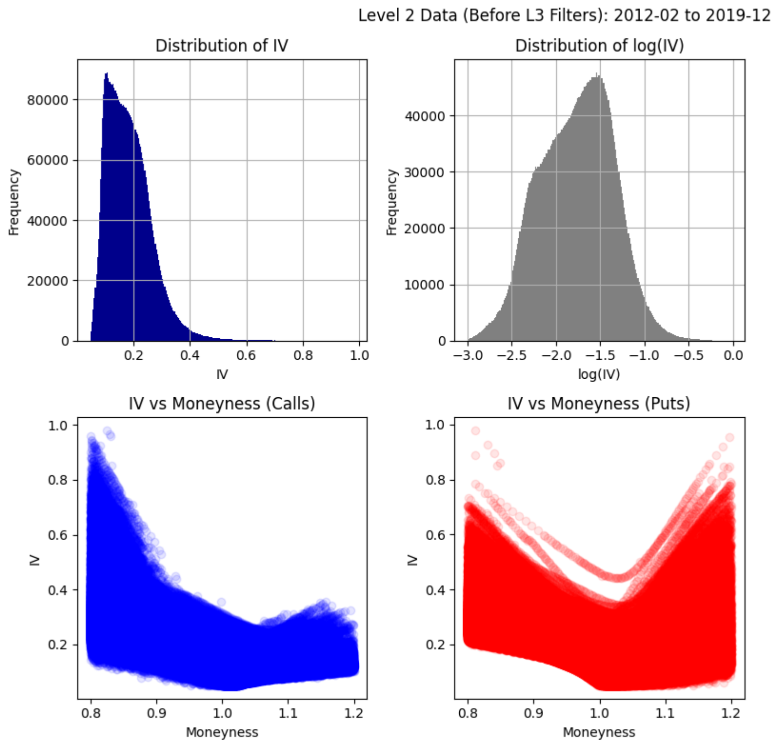

Figure 15: Your caption here

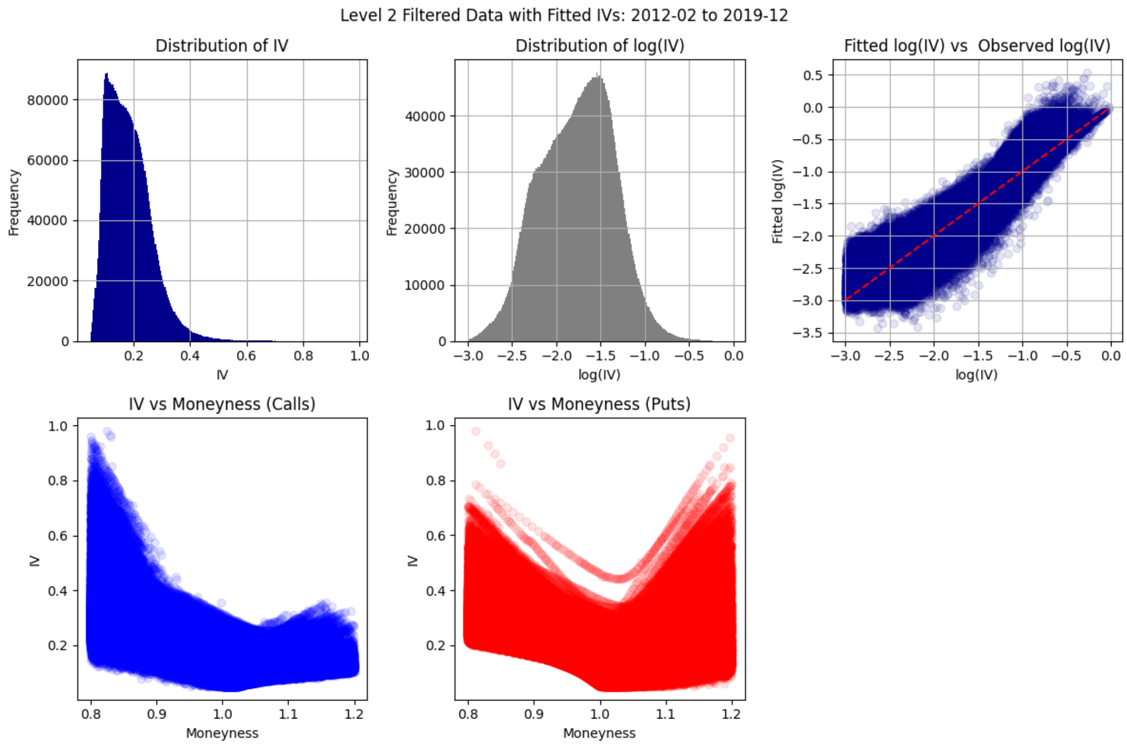

Figure 16: Your caption here

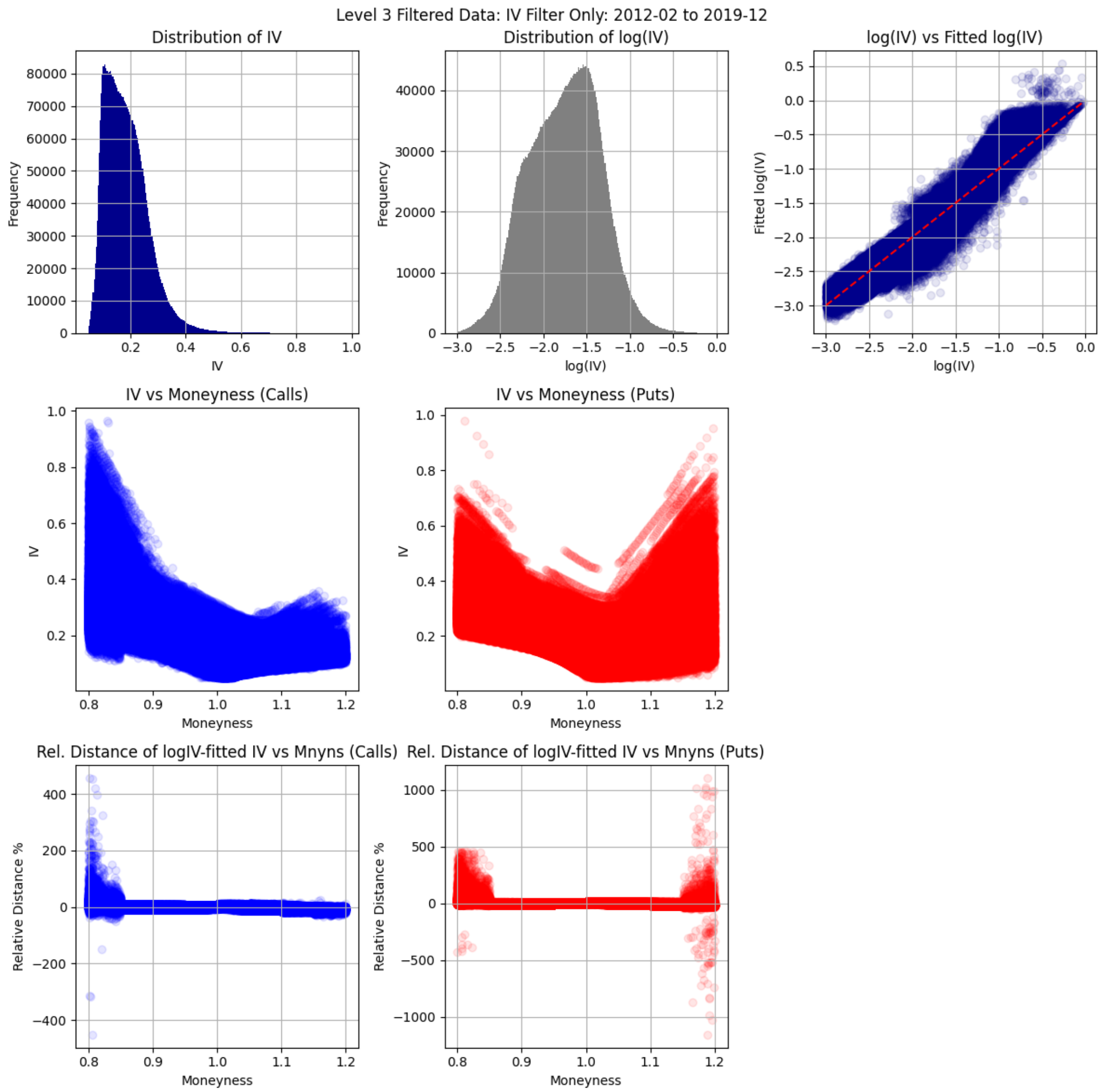

Figure 17: Your caption here

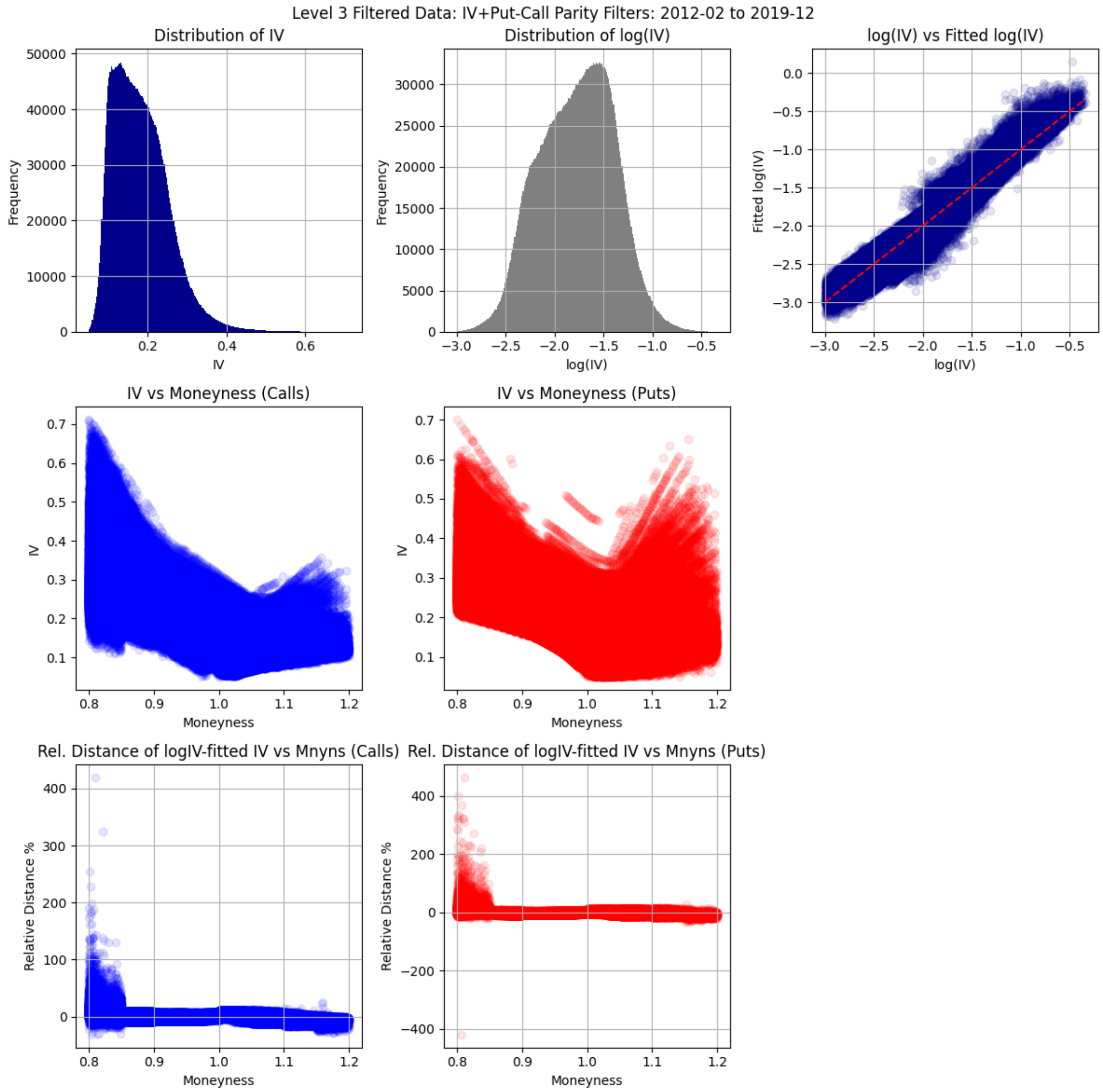

Figure 18: Your caption here