The Puzzle of Filtering Index Options

Viren Desai, Harrison Holt, Ian Hammock

March 6, 2024

Abstract

In this report we summarize our efforts to replicate the Appendix B data filtration process described in *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 paper. Due to the unavailability of SPX option data from 1985 to 1995, we focus our analysis on replicating the filtration results using data from 1996-01 to 2012-01. We then apply these filters to more recent data from 2012-02 to 2019-12. We show that the paper's results are challenging to replicate as described, and also that elegant results obtained from intricately constructed data filters are not always reproducible in other periods. The implications for option pricing models based on time-fragile data filters would be an interesting follow-up study. Our analysis and code can be readily found on Github[1].

---

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 final results are summarized in Table 1.

Table 1: 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. Note that we did not drop any options as a result of the Volume = 0 filter, as we observed over two million options with Volume = 0. Please see Table 2 for further details.

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.

Analysis of Results and Commentary   Two major 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. Table 2 and Table 3 illustrate that many of the options in the

3

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 final portfolios of Constantinides, Jackwerth, and Savov (2013), (if they are not excluded by other filters) is unclear to us and merits further investigation. 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 1.

Table 2: 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 3: 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 2012-02 to 2019-12 data with volume and open interest equal to zero, as well as the overlap.

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].

Analysis of Results and Commentary 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 uncomputable IVs increased as time to maturity decreased (see Figure 1 and Figure 2). Importantly, the reasons for inability to compute IV were not explicitly addressed in the paper, and despite the low remaining percentage of uncomputable IVs in the post-filter data, we do not know what impact the inclusion or exclusion of these options would have on the final option portfolios.

---

[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.

Figure 1: 1996-01 to 2012-01 Pre- and Post-Filter Uncomputable IVs by Time to Maturity
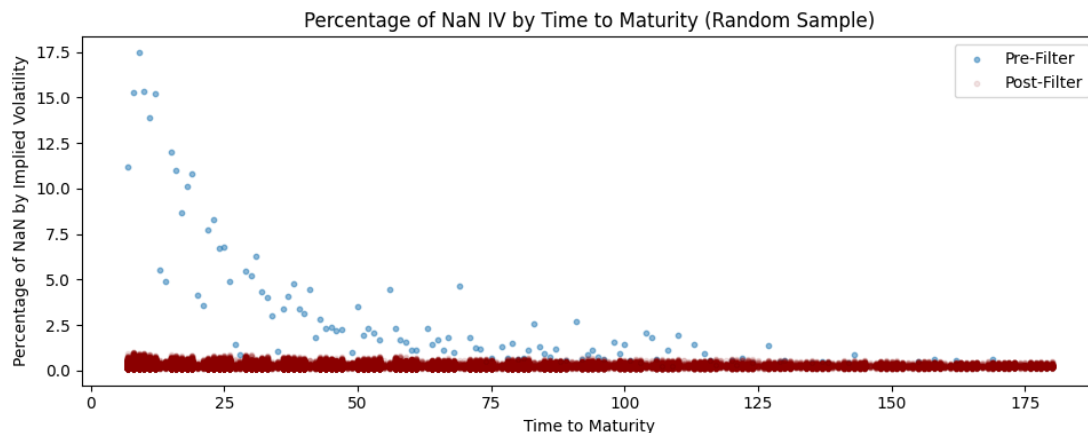
Through our analysis we found there are cases where we could not compute implied volatility (IV), as a result, the values were NaN. In figure above, there is a clear trend where the percentage of uncomputable IVs increase as time to maturity decreases.

Figure 2: 2012-02 to 2019-12 Pre- and Post-Filter Uncomputable IVs by Time to Maturity
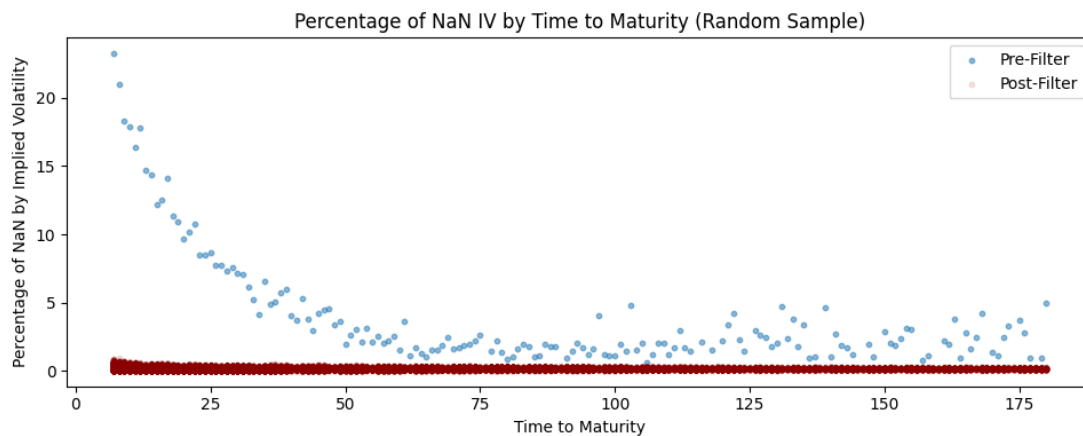
While the trend for uncomputable IVs increasing as time to maturity decreases is still present in the 2012-02 to 2019-12 dataset, the percentage of uncomputable IVs post-filter is lower than in the 1996-01 to 2012-01 dataset.

8

Further data on the results of the Level 2 filters can be accessed with the links below, and are included in Appendix C.

1.2.1 Level 2 Additional Data: 1996-01 to 2012-01

Figures: Figure 7, Figure 8, Figure 9, Figure 10

1.2.2 Level 2 Additional Data: 2012-02 to 2019-12

Figures: Figure 11, Figure 12, Figure 13, Figure 14

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 of 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 mon-

eyness. 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.

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

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

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

Analysis of Results and Commentary   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 other methods, including 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.

Figure 3 and Figure 5 will help reader understand the impact of the Level 3 filters applied to the Level 2 filtered data from 1996-01 to 2012-01.

11

Figure 3: 1996-01 to 2012-01 Level 3 Data (After Level 2 Filters), with Quadratic Polynomial Fitted IVs
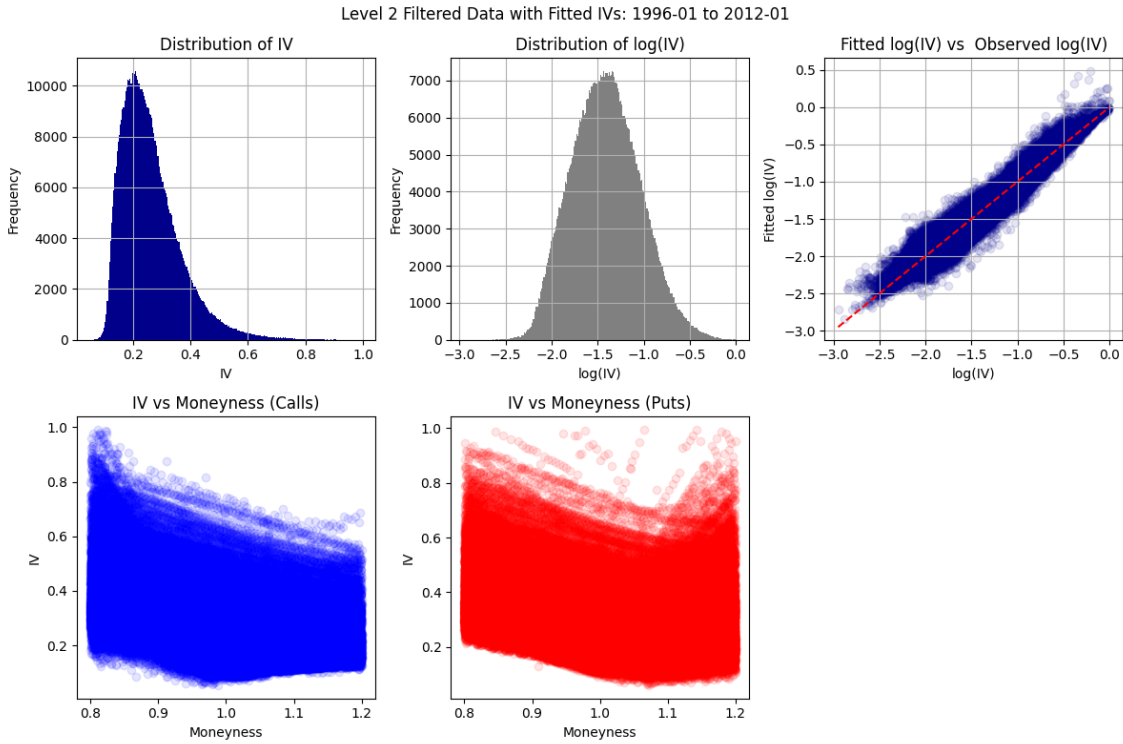

Level 2 Filtered Data with Fitted IVs: 1996-01 to 2012-01

We note that the IV charts for the remaining calls and puts after the Level 2 filters still exhibit some anomalous IV readings, particulary for options that are away-from-the-money. This is the apparent "butterfly arbitrage" that the Level 3 filters aim to remove.

Figure 4: 1996-01 to 2012-01 Level 3 Filtered Data: IV Filter Only
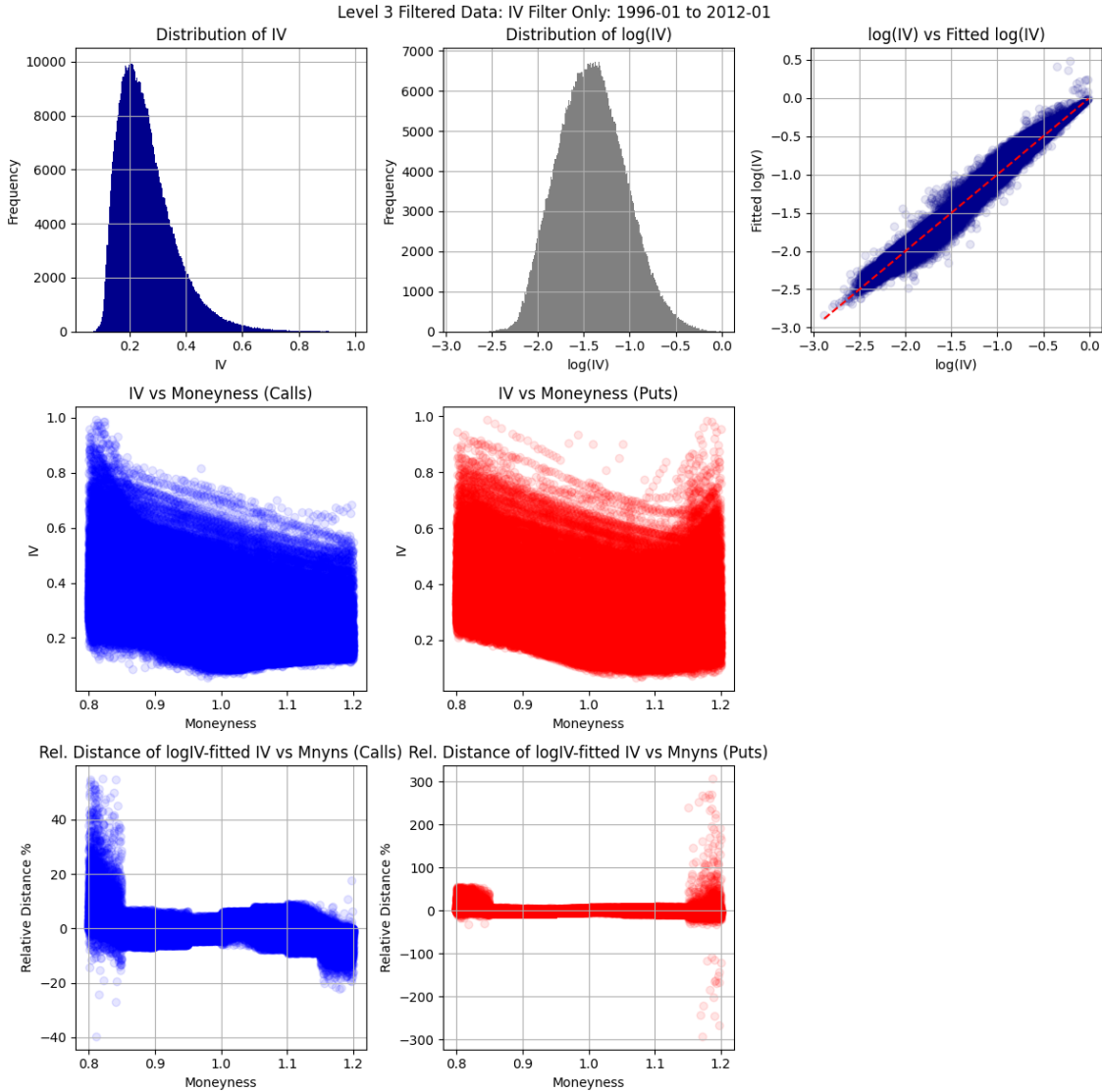
Using the Level 3 IV filters, we are able to remove a significant number of IV outliers. However, we observe that there still seem to be IV outliers for calls and puts. This necessitates the next Level 3 filter: the PCP filter.

Figure 5: 1996-01 to 2012-01 Level 3 Filtered Data: An Elegant Filtration
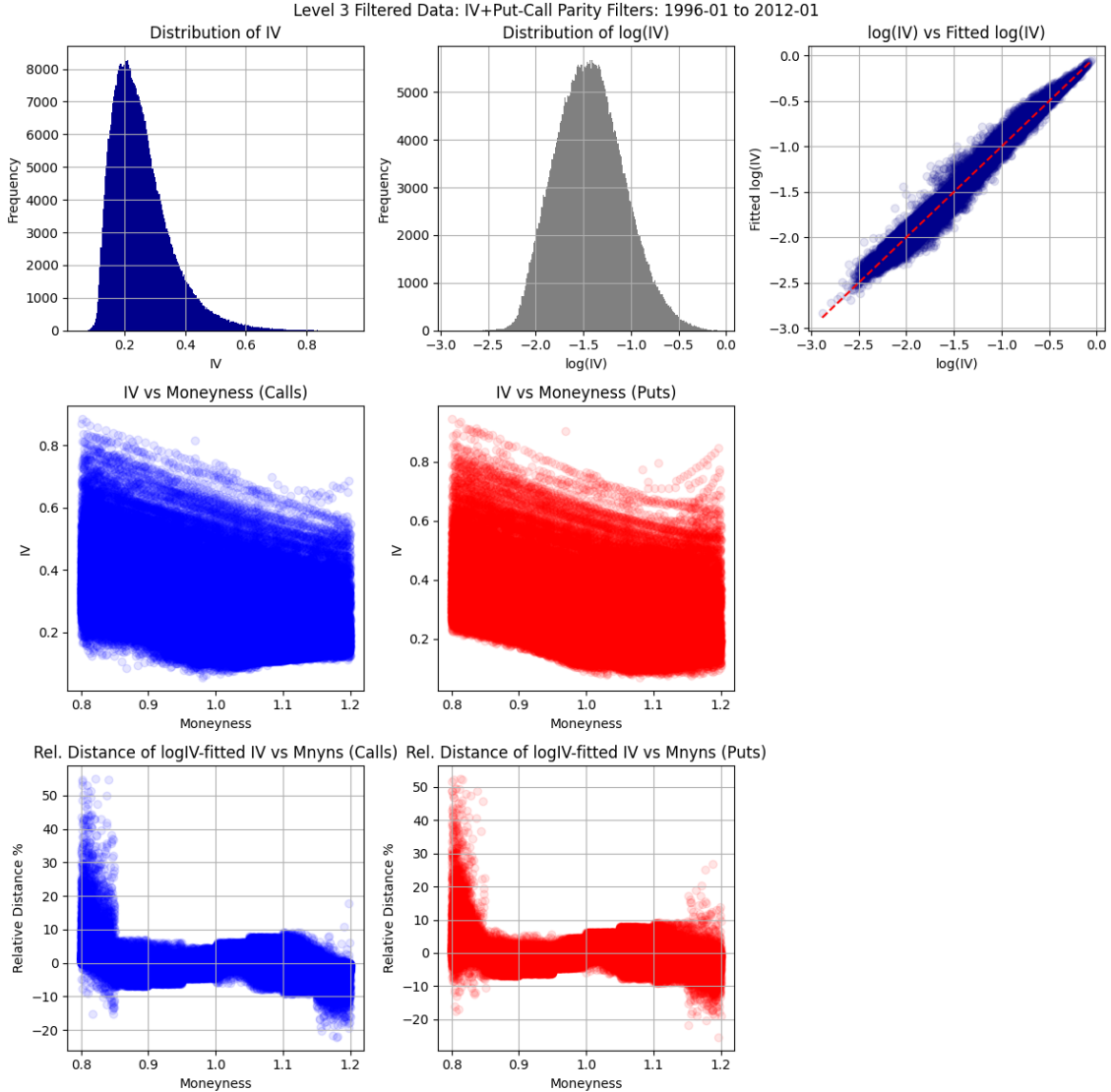
Using the Level 3 filters, we were able to remove a significant number of IV outliers and option pairs that do not satisfy put-call parity. Note that the final distribution of IV in the middle row charts is more akin to the theoretical "volatility smile", indicating that the filters are indeed quite elegant in their combined effect. Notice also how this filtration tightens the relationship between the fitted and observed log IVs in the top right charts.

Further data on the results of the Level 3 filters applied to the 2012-2019 data can be accessed with the links below, and are included in Appendix D

1.3.1 2012-02 to 2019-12

Figures: Figure 15, Figure 16, Figure 6

2 Replicating Table 2

Table 4 in the original paper describes how many options are "found", "go missing", or expire in the dataset, according to definitions developed by Constantinides, Jackwerth, and Savov (2013). An option is found if it disappears and then reappears in the data the next trading day. An option is missing if it disappears and does not reappear the next trading day. Multiple days missing count as multiple options missing. Lastly, if an option is lost and expires this is noted as expired. Our results are summarized in Table 4.

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 (10% expire on trading days, and 3% on other days). We suspect that this may be due to changes in data reporting norms, but this would need to be researched further. To handle this issue, we pushed the expiration day to the nearest Friday, presumably the nearest trading day. However, there were several edge cases that would need to be addressed individually. As a result, discrepancies exist between our analysis and Constantinides, Jackwerth, and Savov (2013), and further investigation is required. A short summary of the data distribution is given in Table 5 in Appendix B.

Table 4: 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.

3   Conclusion

Our findings illustrate that seemingly straight forward instructions for filtering data may lead to divergent results, with potentially significant downstream impacts, supporting the idea of a "reproducibility crisis" in academic finance. These findings can be broadly summarized as follows:

- *Compounding data filtration errors*: We see these divergences in the application of the Level 1, Level 2, and Level 3 filters to the 1996-01 to 2012-01 and 2012-02 to 2019-12 datasets, despite our best efforts to replicate the analysis in Constantinides, Jackwerth, and Savov (2013) as closely as possible. Due to the sequential application of these filters, errors get compounded downstream, and the final option portfolios may be significantly different from those in the original paper.

- *Errors due to lack of specificity in algorithms and parameters*: We believe that the differences in our results are due to the lack of specificity in the original paper regarding the exact methods used to apply the filters, and the lack of detail on the parameters used in the filters. There may be some potential impacts from changes in the underlying databases as well, but we have no information on this.

Importantly, one of the major stated objectives of the filters was to reduce the prevalence of apparent butterfly arbitrage. While the final filtration results after Level 3 show a relatively "well-behaved" implied volatility distribution (see Figure 5),

observe how the "butterfly wings" reappear when we apply the same filters, with the same parameters to the data from 2012-02 to 2019-12! (see Figure 6).

This is a clear indication that the filters are not robust to the passage of time, and as such, pricing models based on data filtered by the processes described in Constantinides, Jackwerth, and Savov (2013) should be subject to regular review and scrutiny.

To reduce these errors, we suggest that journals require authors of computationally intense manuscripts, such as Constantinides, Jackwerth, and Savov (2013), to publish their code base along with their findings. Our data acquisition process is briefly described in Appendix A, and our code base can be found on our Github[3] repository.

---

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

Figure 6: 2012-02 to 2019-12 Level 3 Filtered Data: Not So Elegant Filtration
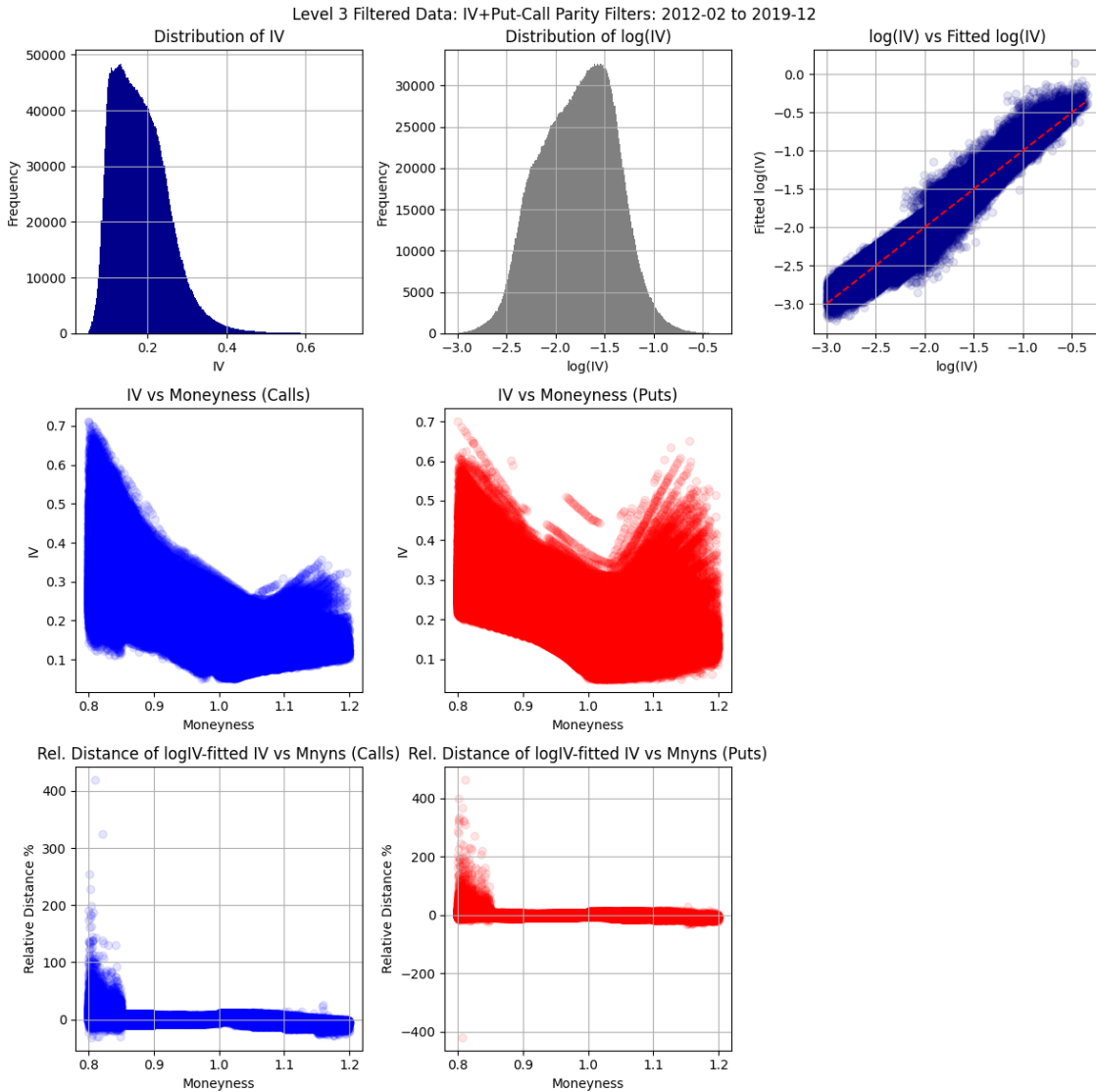
Applying the same filters to a different time period does not yield an elegant implied volatility distribution, and clearly does not eliminate opportunities for butterfly arbitrage! Notice how the identical filtration process does not yield lognormal IV, and the resultant volatility smile is irregular and somewhat V-shaped.

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.

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 Odd Expiration Dates

Table 5: 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.

C   Additional Data: Level 2 Filters

C.1   1996-01 to 2012-01

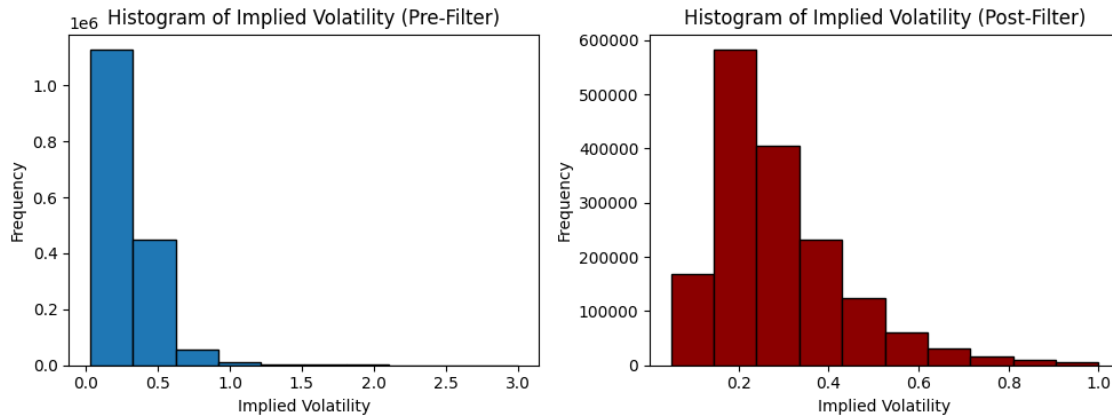**Figure 7: Effects of filtering Days to Maturity <7 or>180**
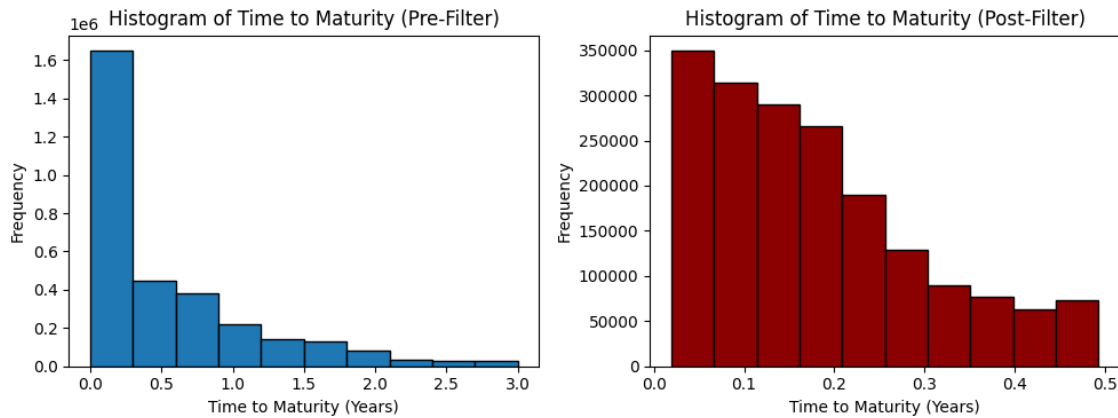
Distribution of time to maturity, measured in years from option initial date to expiration date. The graph on the left shows the distribution prior to applying the initial Level 2 filter of excluding days to maturity less than 7 and greater than 180. Right shows distribution post filter.

Figure 8: Comparison of Pre- and Post-Filter Data
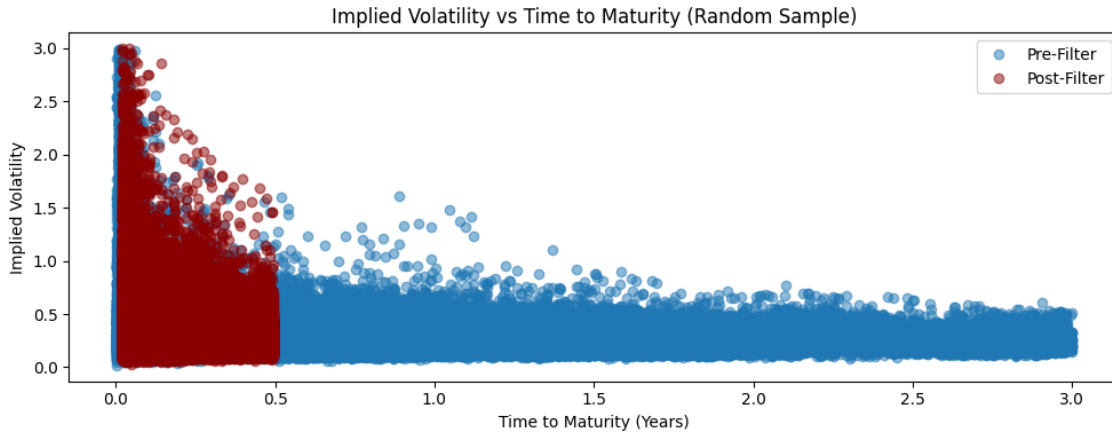
As noted in the paper, the short maturity options tend to move erratically nearing expiration. In Figure 8, post-filter (red) we see a slight reduction of short-term options with a high implied volatility.

Figure 9: Effects of filtering IV <5% or >100%

*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.*

Figure 10: Effects of filtering on Moneyness <0.8 or >1.2
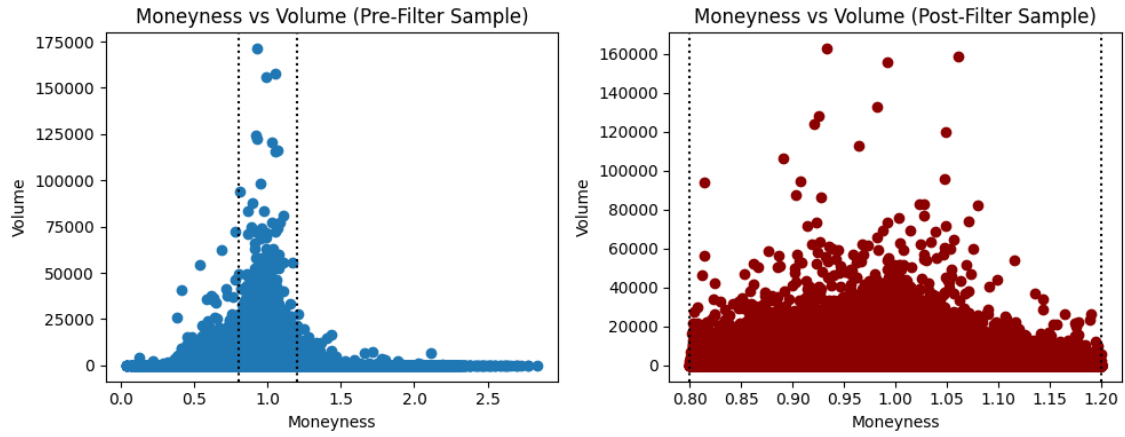
Removing option quotes with moneyness lower than 0.8 and higher than 1.2 eliminates extreme values. These extreme values potentially have quotation problems or low values.

C.2 2012-02 to 2019-12

Figure 11: Effects of filtering Days to Maturity <7 or >180
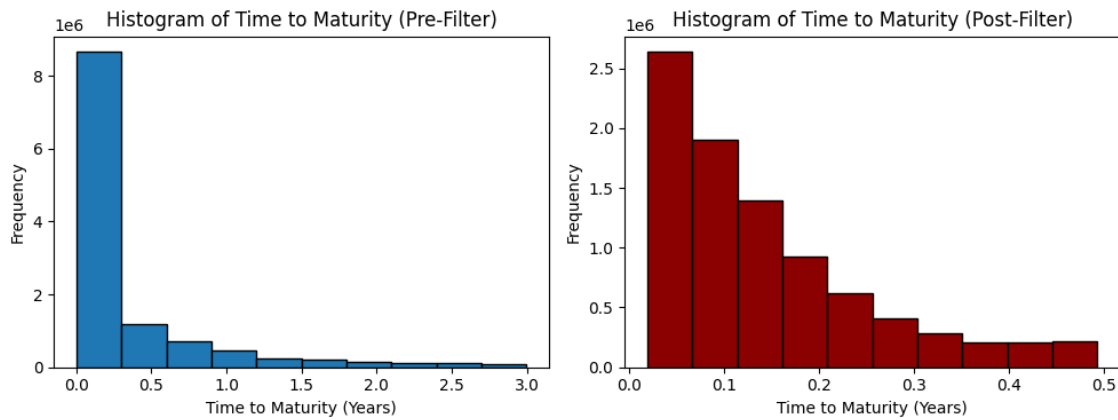
Distribution of time to maturity, measured in years from option initial date to expiration date. The graph on the left shows the distribution prior to applying the initial level 2 filter of excluding days to maturity less than 7 and greater than 180. Right shows distribution post filter.

Figure 12: Comparison of Pre- and Post-Filter Data
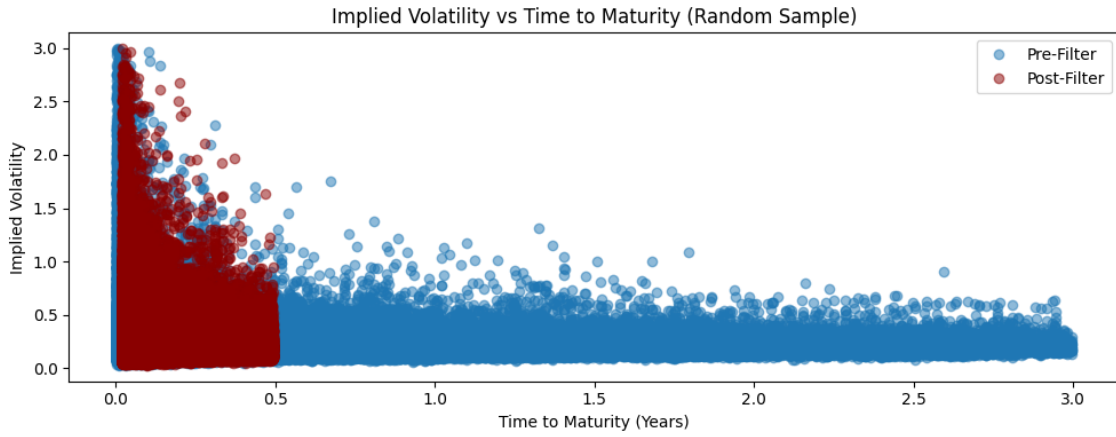
As noted in the paper, the short maturity options tend to move erratically nearing expiration. In Figure 12, post-filter (red) we see a slight reduction of short-term options with a high implied volatility.

Figure 13: Effects of filtering IV <5% or >100%
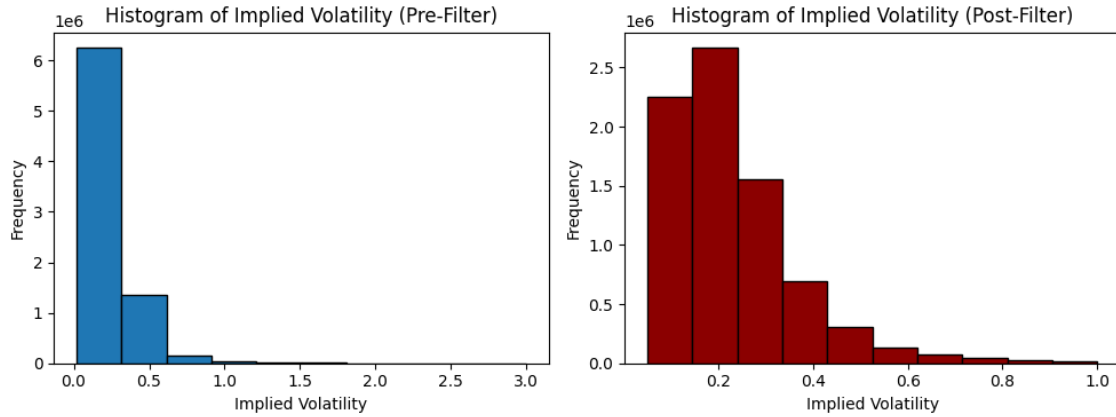
Removing option quotes with implied volatilities lower than 5% or higher than 100% eliminates extreme values and reduces the skewness of the implied volatility distribution.

Figure 14: Effects of filtering on Moneyness <0.8 or >1.2
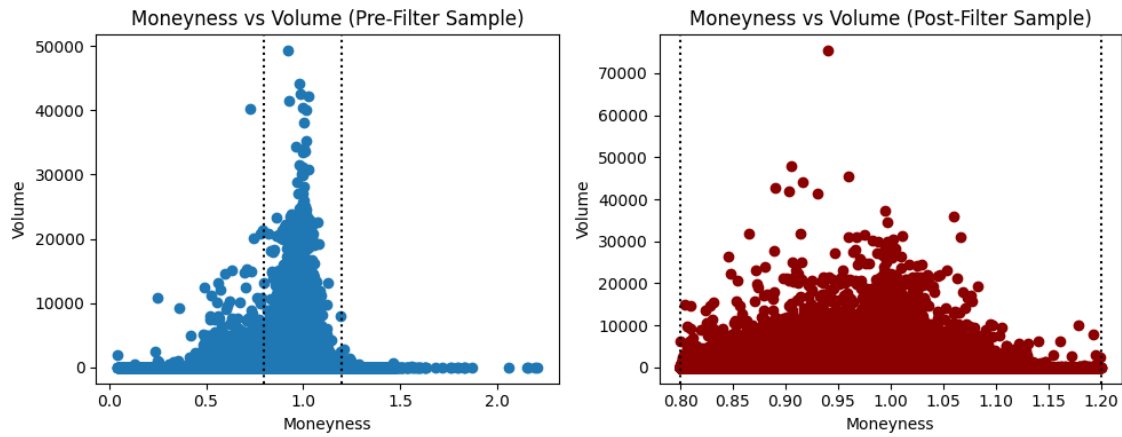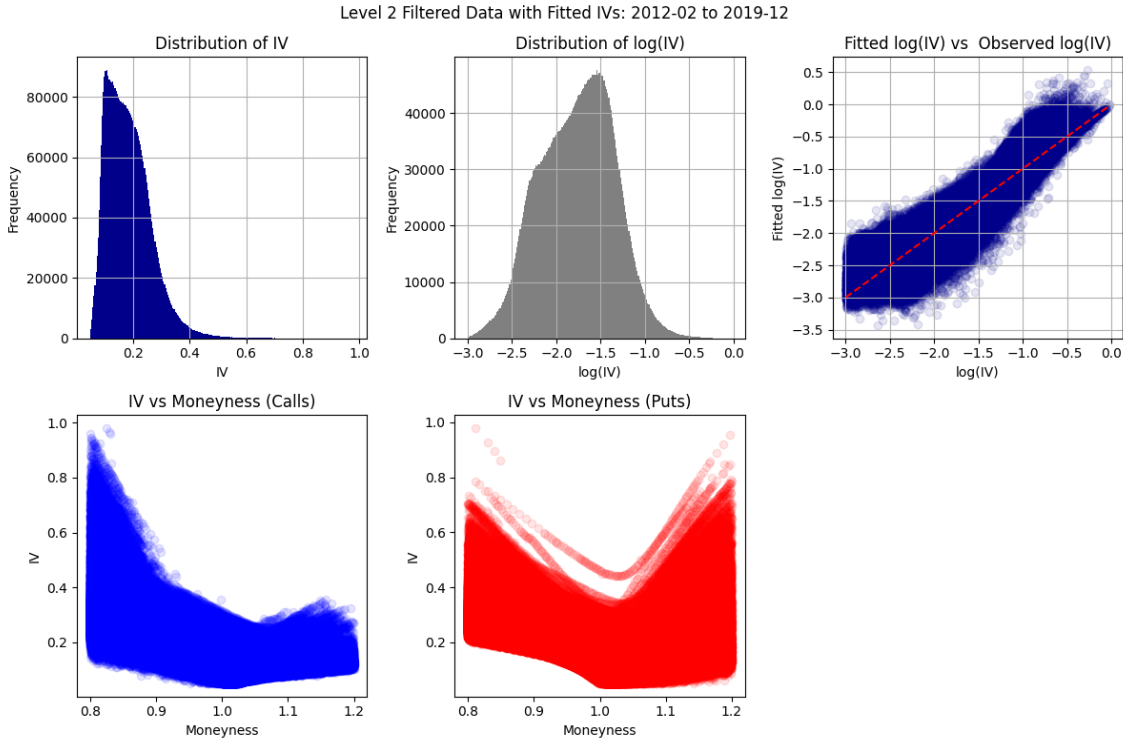

Removing option quotes with moneyness lower than 0.8 and higher than 1.2 eliminates extreme values. These extreme values potentially have quotation problems or low values.

D   Additional Data: Level 3 Filters

D.1   2012-02 to 2019-12

Figure 15: Level 3 Data (After Level 2 Filters): Messy IVs

We note that the IV charts for the remaining calls and puts after the Level 2 filters still exhibit some anomalous IV readings, particulary for options that are away-from-the-money. This is the apparent "butterfly arbitrage" that the Level 3 filters aim to remove.

Figure 16: Level 3 Data (After IV Filter Only): Not So Effective
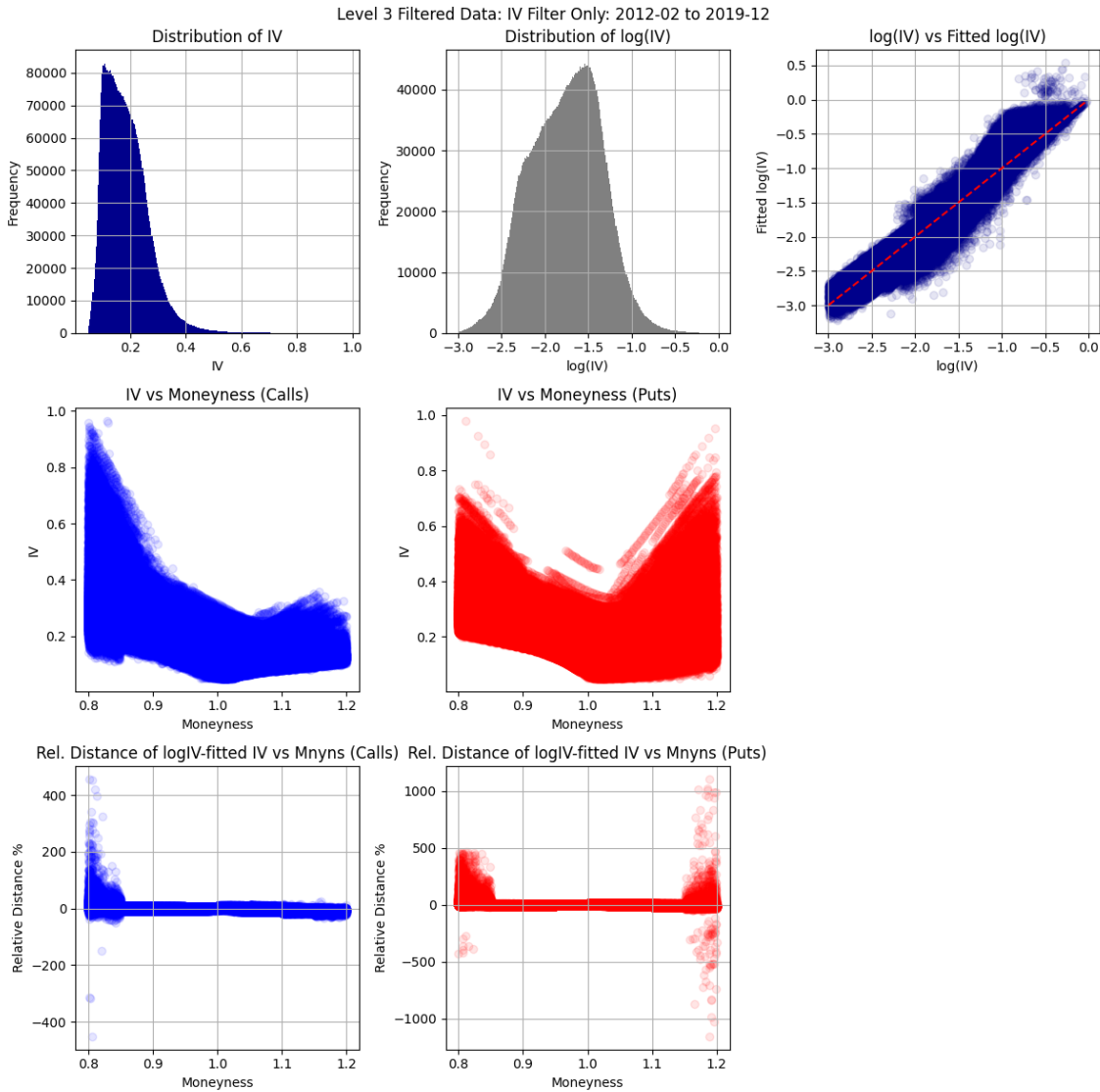
The Level 3 IV filter is not as effective at removing the outliers as we would like. As showing in Figure 6, even the PCP filter was unable to remove opportunities for butterfly arbitrage, highlighting the time-fragility of the filters.