Too Good to Be True: Look-ahead Bias in Empirical Options Research

Jefferson Duarte, Christopher S. Jones, Mehdi Khorram,
Haitao Mo, and Junbo L. Wang *

February 29, 2024

*We thank Yakov Amihud, Kevin Aretz, Hank Bessembinder, Amit Goyal, Kris Jacobs, Dmitriy Muravyev, Alessio Saretto, Avanidhar Subrahmanyam, Tarik Umar, Aurelio Vasquez, and seminar participants at the 2023 Conference on Derivatives and Volatility. Duarte is with the Jesse H. Jones School of Business at Rice University (jefferson.duarte@rice.edu). Jones is with the Marshall School of Business at the University of Southern California (christopher.jones@marshall.usc.edu). Khorram is with the Saunders College of Business at the Rochester Institute of Technology (mkhorram@saunders.rit.edu). Mo (haitaomo@ku.edu) is with the School of Business at the University of Kansas, and Wang (junbowang@lsu.edu) is with the Ourso College of Business at Louisiana State University.

Too Good to Be True: Look-ahead Bias in Empirical Options Research

Abstract

Numerous trading strategies examined in options research exhibit remarkably high mean returns and Sharpe ratios. We show some of these seemingly "good deals" are due to look-ahead biases. These biases stem from using information unavailable at the portfolio formation time to filter out observations suspected of being noisy or erroneous. Our results suggest that elevated Sharpe ratios may serve as potential indicators of such look-ahead biases. Furthermore, deviating from previous literature findings, we show that illiquidity is not strongly priced in stock options and that only a small set of stock characteristics are in fact associated with option expected returns.

Keywords: Options; Look-ahead bias

1 Introduction

Numerous trading strategies in the options literature exhibit high expected returns and Sharpe ratios. Indeed, the options literature often reports instances of substantial average returns, accompanied by Sharpe ratios well above two (e.g., Zhan, Han, Cao, and Tong (2021) and Christoffersen, Goyenko, Jacobs, and Karoui (2018)). These elevated Sharpe ratios imply that many option strategies persistently offer "good deals," which in turn suggests that the options market faces frictions that prevent investors from arbitraging away these seemingly advantageous opportunities.[1] However, before drawing this conclusion, it is imperative to assess the realism of these high mean returns and Sharpe ratios. In this paper, we investigate the extent to which these elevated mean returns and Sharpe ratios may be attributed to look-ahead bias.

We investigate how look-ahead biases influence two stylized facts in the empirical options literature. The first stylized fact is that numerous stock characteristics seem to be priced in the cross-section of individual equity option returns. In their study, Zhan, Han, Cao, and Tong (2021, hereafter ZHCT) explore the relation between the returns from delta-hedged call writing and ten stock characteristics commonly analyzed in the equity literature. At the end of each month, ZHCT classify all stocks within their sample into deciles based on these characteristics. They compare the returns over the following month of the portfolios of delta-neutral calls written on stocks falling within the top decile versus the bottom decile. Across the majority of the characteristics they analyze, they identify substantial mean returns for long/short strategies involving the top and bottom decile portfolios. In fact, the average long/short equal-weighted mean excess return across all the characteristics they investigate amounts to 2.4% per month, with an average t-statistic of about 19 and an average annualized

[1] To benchmark these elevated Sharpe ratios, note that Bernie Madoff's Ponzi scheme had Sharpe ratios between 2.5 and 4 (Carozza (2009)).

Sharpe ratio of 4.

The second stylized fact is that illiquidity seems to be strongly priced in individual equity options. Christoffersen, Goyenko, Jacobs, and Karoui (2018, hereafter CGJK) explore the link between the mean daily returns of delta-hedged options and option illiquidity. At the conclusion of each trading day, CGJK construct five portfolios of delta-hedged options, sorting them based on option illiquidity. They find significant mean returns for the strategy involving a long position in illiquid options and a short position in liquid options. Specifically, the mean long/short excess return for providing liquidity on at-the-money (hereafter ATM) call (put) options amounts to 3.4% (2.5%) *per day*, with a t-statistic of approximately 26 (14) and an annualized Sharpe ratio of 9 (5).

We employ a two-step methodology to investigate how look-ahead biases affect the stylized facts presented in these papers. In the initial step, we apply the same sample selection filters as those used by ZHCT and CGJK in order to replicate their findings. It is worth noting that within the options literature, it is customary to employ filters during sample selection to eliminate erroneous and noisy observations. However, differently from most of the literature (e.g. Driessen, Maenhout, and Vilkov (2009), Bollen and Whaley (2004), Goyal and Saretto (2009), Cao and Wei (2010), and Muravyev (2016)), some of the filters utilized in ZHCT and CGJK rely on information that is not available at the time of portfolio formation (t_0). Consequently, these filters result in infeasible trading strategies. For example, one of the sample selection filters entails removing all option prices that do not satisfy arbitrage bounds. This filter is infeasible for a trader who, at time t_0, does not possess the information to determine whether an option price will satisfy no-arbitrage bounds at the end of the holding period (t). In our second step, we replicate the option trading strategies in ZHCT and in CGJK using a feasible trading approach that relies solely on information available at the time of portfolio formation.

We find similar results to those in ZHCT and in CGJK in the first step of our methodology. Specifically, most of the mean returns and t-statistics of the long/short strategies based on stock characteristics match those in ZHCT very well. In our replication, the mean long/short monthly excess return across all characteristics they examine is 2.4% with a mean t-statistic of about 16. The mean annualized Sharpe ratio is 3.6. Moreover, as in CGJK, we find that the mean liquidity premium on options can be more than one percent per day. Specifically, the mean long/short excess return of providing liquidity on ATM call (put) options is 1.6% (0.9%) *per day* with a t-statistic of about 30 (18). The annualized Sharpe ratio for providing liquidity for ATM calls (puts) is 10.4 (6.5).

In the second step of our methodology, we uncover a notable shift in the results, as the elevated mean returns and t-statistics previously reported for the infeasible strategies dissipate when replaced with feasible strategies. When we replicate the trading strategies in ZHCT and CGJK using feasible trading strategies that rely exclusively on information available at the time of portfolio formation, the previously observed high mean returns, t-statistics, and Sharpe ratios do not persist. In the feasible trading strategies, we find that the mean long/short monthly excess return across all characteristics examined in ZHCT amounts to 0.4%, with an average t-statistic of approximately 2. The average annualized Sharpe ratio stands at around 0.5. Similarly, our analysis reveals that the mean long/short daily excess return for providing liquidity on ATM call and put options is 0.10% and 0.06%, respectively, with t-statistics of about 2.6 and 1.3. The annualized Sharpe ratio for providing liquidity for ATM calls and puts is 0.9 and 0.5, respectively.

The difference between the feasible and infeasible results indicates that the high Sharpe ratios and mean returns for the trading strategies we examine are due to look-ahead biases. The large biases we find result from sample selection filters based on information that is not available to traders at the time of portfolio formation. These filters are intended to remove erroneous and noisy observations, but if they are based on information that is not available at the time of portfolio formation they may result in large look-ahead biases. Naturally, applying these filters only at the time of portfolio formation would not result in such a bias.

Two characteristics of the procedure detailed in ZHCT introduce a positive look-ahead bias. Firstly, their sample selection criteria exclude call options that become either deep out-the-money (OTM) or deep in-the-money (ITM) at the end of the holding period. Since delta-neutral call writing, characterized by a negative gamma, yields negative returns when call options become either deep-ITM or deep-OTM, the exclusion of these options from their sample introduces a positive look-ahead bias. Secondly, the stock attributes they use for portfolio sorting exhibit a correlation with stock volatility. Options on high volatility stocks are more likely to end up either deep-OTM or deep-ITM by the conclusion of the holding period. Consequently, the positive bias resulting from the exclusion of deep ITM or OTM options from their sample is more pronounced in portfolios with higher volatility, thereby impacting the mean returns of their long/short strategies.

The procedure in CGJK introduces look-ahead bias because of two features. Firstly, CGJK's sample selection process includes only options that fall within a predefined moneyness range (e.g., ATM) in the *middle* of the holding period. As a result, their selection of call options includes a higher proportion of calls that were OTM prior to the holding period when the underlying stock – and therefore a call option – has a positive return during the holding period. Similarly, their selection of put options includes a greater number of puts that were OTM prior to the holding period when underlying stock returns are negative, implying positive put option returns. Secondly, their sorting variable — option illiquidity — is higher for OTM options than it is for ITM options. Consequently, sorting option portfolios based on illiquidity is tantamount to sorting them by their moneyness. When we combine these two features, the procedure in CGJK generates a look-ahead bias because sorting based on option illiquidity is effectively equivalent to sorting based on moneyness, which, in turn, is akin to sorting based on the returns of the options during the holding period.

It is important to note that some degree of infeasibility is natural in empirical research. For instance, removing options with prices that appear to be erroneous could induce some bias. Erroneous observations contribute to noise in the inferences and are commonplace in the datasets used in academic research. In the appendix, we show results with an infeasible sample that is cleaned for outliers. Our protocol for managing erroneous observations involves transparently disclosing their count within the sample and comparing the results with and without these data points. This is distinct from eliminating every observation with potential error. More crucially, the biases in ZHCT and CGJK arise from selecting a sample of options based on their moneyness after the portfolio formation period, not from the removal of clearly erroneous observations.

Our primary contribution is to the empirical option literature. We contribute to this literature in three different ways.

First, our paper adds to the expanding body of research on empirical option pricing, which explores the relationship between various stock characteristics and expected option returns. For example, Goyal and Saretto (2009), Aretz, Lin, and Poon (2022), Choy and Wei (2023), Bali, Beckmeyer, Mörke, and Weigert (2023), and Cao, Goyal, Zhan, and Zhang (2023) investigate how different stock attributes are linked to option returns. We extend this literature by showing that the number of stock characteristics associated with expected option returns is much lower than that found by ZHCT. In contrast to ZHCT, who identified a strong correlation between delta-hedged option returns and a wide array of previously unexplored stock characteristics (such as stock price, profit margin, firm profitability, cash holdings, cash flow variance, new share issuance, total external financing, distress risk, and dispersion of analyst forecasts), our findings suggest that most of these apparent relations are the result of look-ahead biases. Notably, within the feasible sample, we find that only stock price, cash holdings, and new share issuance are modestly related to the cross-section of delta-hedged option returns.

Second, our findings underscore the need for further examination of whether illiquidity is priced in stock options. The strength of the CGJK findings established a stylized fact in the literature that illiquidity is indeed priced into options. In fact, numerous papers in the field cite CGJK as evidence supporting the idea that options expected returns are correlated with illiquidity (e.g., Goncalves-Pinto, Grundy, Hameed, van der Heijden, and Zhu (2020); Cosma, Galluccio, Pederzoli, and Scaillet (2020); Goyal and Saretto (2022); Ramachandran and Tayal (2021)). However, our results suggest that this may not be the case, as we observe evidence of illiquidity being priced only marginally in call options when employing the bias correction procedure outlined by Duarte, Jones, and Wang (2022). Notably, we do not find evidence indicating that illiquidity is priced in put options in our feasible sample. Our weaker results on illiquidity are consistent with the fact that there is no definitive theoretical prediction suggesting that illiquidity should be positively priced in options, given their zero-net-supply nature. In case of options, the pricing of illiquidity can be either negative or positive, depending on whether sellers or buyers demand an illiquidity premium (Brenner, Eldor, and Hauser, 2001).

Third, a common practice in the empirical option literature is to address the effect of measurement errors in option prices with sample selection filters. These filters normally eliminate deep-OTM options and option prices that do not satisfy arbitrage bounds (e.g., Bollen and Whaley (2004), Driessen, Maenhout, and Vilkov (2009), Goyal and Saretto (2009), and Muravyev (2016)). We show that these filters may result in large look-ahead biases when they are based on information that is not available at the time of portfolio formation.

Our work is also related to the burgeoning literature addressing the many biases in the estimation of expected option returns. Eraker and Osterrieder (2018) show that the VIX is a biased estimate of implied volatility because the midpoint of the closing bid-ask prices is biased due to order flow pressure, while Goyenko and Zhang (2019) find biases in mean returns computed from closing prices that arise from differences between closing and intraday option prices. Duarte, Jones, and Wang (2022) extend the work by Blume and Stambaugh (1983) and Asparouhova, Bessembinder, and Kalcheva (2010, 2013) to options. They show that estimates of the volatility risk premium in individual equity options are sensitive to microstructure biases. Differently from these papers, we focus on the effect of look-ahead biases in the estimation of expected returns of option strategies.

We also extend the literature that analyzes look-ahead biases in the estimation of expected returns. This literature has largely focused on the effect of survivorship biases in the estimation of expected returns (e.g., Brown, Goetzmann, and Ross (1995), Carpenter and Lynch (1999), ter Horst, Nijman, and Verbeek (2001), and Baquero, ter Horst, and Verbeek (2005)). We add to this work by showing that procedures used to eliminate potentially noisy or erroneous data may also lead to look-ahead biases. Cochrane and Saa-Requejo (2000) theoretically extend the no-arbitrage principle to price options by assuming the investors engage in strategies with high Sharpe ratios – good deals. Our findings suggest that the same idea applies in empirical research. Specifically, we show that unreasonable Sharpe ratios may serve as an indicator of look-ahead biases and that strategies with high Sharpe ratios should be scrutinized.

2 Empirical Methodology

Our convention is that t_0 is the time that a trader forms a portfolio of options. We assume that the trader only has access to information that is publicly available at time t_0 to form her portfolio. Let R_t represent the return of the trader's portfolio between t_0 and the end of the holding period ($t > t_0$). The return R_t is the object of study of an econometrician analyzing a trading strategy. We also define t_E as the time of the information that the econometrician uses to create a portfolio that is supposed to be formed at time t_0.

We define a trading strategy as infeasible when $t_E > t_0$ and feasible otherwise. An econometrician can analyze an infeasible trading strategy by using information that is available at only time $t_E > t_0$ to place a given security in a pseudo portfolio at time t_0. In contrast, a trader does not have, at time t_0, the information necessary to select the options in the econometrician's pseudo portfolio. Therefore the trading strategy is infeasible.

Infeasibility could result from a number of different choices by the econometrician, but in this paper we focus on strategies that are infeasible as a result of sample selection. A feasible sample is one that is based on information that is observable at t_0. In contrast, an infeasible sample is based on information that is not observable at t_0. For instance, a sample is infeasible when it excludes all option observations with prices violating no-arbitrage conditions at time t. This sample is infeasible because a trader does not know at time t_0 whether any given option satisfies no-arbitrage conditions at time t.

A look-ahead bias in the estimation of expected returns occurs when the trading strategy is infeasible and the sample selection criteria covaries with the return $R_{i,t}$ of the securities in the portfolio. For example, assume that an econometrician wishes to estimate the expected return of options in the population ($E[R_{i,t}]$). To do so, the econometrician estimates the expected return of calls in the sample, which is $E[R_{i,t}|IS_i]$, where IS_i indicates that call i is within the econometrician's sample at time t_0. The look-ahead bias is

$$E[R_{i,t}|IS_i] - E[R_{i,t}] = (E[R_{i,t}] - E[R_{i,t}|FO_i]) \times Odds[FO_i], \tag{1}$$

where FO_i indicates that call i is filtered out of the econometrician's sample at time t_0, and $Odds[FO_i]$ is the odds that that option i is filtered out of the sample.[2] Equation 1 indicates

[2]This follows from $E[R_{i,t}] = E[R_{i,t}|IS_i] \times P[IS_i] + E[R_{i,t}|FO_i] \times P[FO_i]$, where $P[IS_i]$ and $P[FO_i]$ are

that when $E[R_{i,t}] = E[R_{i,t}|FO_i]$ the bias is zero. That is, if the sample selection process does not covary with $R_{i,t}$ there is no look-ahead bias even if the sample is infeasible.

We examine the look-ahead biases in two types of option strategies. First, we examine option strategies that are based on stock characteristics (ZHCT). Second, we examine the premia on liquidity provision in the option market (CGJK).

While the bias presented in Equation 1 impacts the estimation of expected returns for options, it may not necessarily affect the outcomes in ZHCT and CGJK. In these studies, the option strategies involve sorting options within the sample to construct portfolios and subsequently calculating the mean returns of long/short strategies. Consequently, the bias in these strategies is the difference in the biases of the top and bottom portfolios. If both these portfolios are equally susceptible to look-ahead biases, this difference is zero, and the mean return of the long/short strategy is unaffected by look-ahead bias. Conversely, if the top and bottom portfolios are impacted differently by look-ahead biases, then the mean return of the long/short strategy will be influenced by look-ahead bias. To clarify, by employing Equation 1, if the probabilities of excluding options ($Odds[FO_i]$) or the expected returns of the options that are excluded ($E[R_{i,t}|FO_i]$) from the bottom and top portfolios are dissimilar, the average difference in returns between these portfolios is subject to look-ahead bias. Therefore, look-ahead biases affect long/short strategies only when the sorting variable used to create the portfolios is correlated with the look-ahead bias in the mean return of each portfolio.

We employ a two-step methodology to investigate the presence of look-ahead biases in the mean returns of the long/short strategies described in ZHCT and CGJK. Initially, we reproduce the findings from these papers utilizing all available information. Subsequently, we emulate their long/short strategies by implementing a feasible trading strategy based

the probabilities that option i is in the sample and is filtered out from the sample respectively. $Odds[FO_i] = P[FO_i]/(1 - P[FO_i])$

solely on information accessible at the time of portfolio formation (t_0). In Section 3, we examine the results presented in ZHCT. In Section 4, we delve into the results presented in CGJK.

3 Option Return Predictability

ZHCT study the relation between monthly return to delta-hedged call writing and ten different stock characteristics. In their paper, the monthly excess return to writing a delta-hedged call is

$$\frac{\Delta_{t-1}S_t - C_t}{\Delta_{t-1}S_{t-1} - C_{t-1}} - 1 - r_{f,t-1}, \tag{2}$$

where Δ_{t-1}, S_{t-1}, and C_{t-1} are the call option delta, underlying price and call price respectively at the portfolio formation time ($t-1$). The risk-free rate is $r_{f,t-1}$. The value of the delta-hedged call writing portfolio at time $t-1$ is $\Delta_{t-1}S_{t-1} - C_{t-1}$. In ZHCT, t_0 is the last trading day of the month ($t_0 = t-1$). As ZHCT, we hold the position for one month without rebalancing the delta hedge.

We implement the same procedure as ZHCT to study the relation between the return to delta-hedged call writing and stock characteristics. At the end of each month and for each of the ten stock characteristics they study, ZHCT sort all stocks in their sample into deciles and compare the portfolios of delta-neutral calls written on the stocks belonging to the top decile to those written on the bottom decile.

The ten stock characteristics in ZHCT are as follows: CFV (cash flow variance from Haugen and Baker (1996)), CH (cash-to-assets as in Palazzo (2012)), DISP (analyst earnings forecast dispersion as in Diether, Malloy, and Scherbina (2002)), ISSUE1Y (one-year new issues as in Pontiff and Woodgate (2008)), ISSUE5Y (five-year new issues as in Daniel and Titman (2006)), PM (profit margin as in Soliman (2008)), ln(PRICE) (the log of stock price at the end of last month as in Blume and Husic (1973)), PROFIT (profitability as in

Fama and French (2006)), TEF (external financing for one fiscal year end as in Bradshaw, Richardson, and Sloan (2006)), and ZS (Z-score as in Altman (1968) and Dichev (1998)). We follow ZHCT closely for calculation of stock characteristics. We also consult with and follow the original papers that publish those characteristics as stock return predictors as well as Hou, Xue, and Zhang (2020), which provides construction details. Appendix A gives details on these characteristics. ZHCT also use three control variables: $ln(HV/IV)$ (volatility mispricing measure as in Goyal and Saretto (2009)), IVOL (idiosyncratic volatility as in Ang, Hodrick, Xing, and Zhang (2006) and Cao and Han (2013)), AMIHUD (the Amihud (2002) stock illiquidity measure). The construction of the control variables is also described in the appendix.

We use the same data source and time period as ZHCT to create our sample. The data on U.S. individual stock options are from OptionMetrics from January 1996 to April 2016. The data set includes the daily closing bid and ask quotes, trading volume, open interest, and delta of each option. We use the WRDS link table to merge this data set with CRSP, from which we obtain stock returns, prices, trading volume, market capitalization, and adjustments for stock splits. Our sample includes options on common stocks with CRSP share codes of 10 and 11 and a stock price at least five dollars at the time of portfolio formation ($t-1$). The Fama-French common risk factors and the risk-free rate are from Kenneth French's website. Annual accounting data are from Compustat, while analyst coverage and forecast data are from I/B/E/S.

The sample used in ZHCT is infeasible because it is based on information only available at the end of the holding period (time t). That is, $t_E = t$ which is greater than $t_0 = t-1$. Specifically, to build a sample that replicates ZHCT's results, we apply the following filters using call and put prices at time t and $t-1$: (1) Bid prices are positive. (2) The quote midpoint is at least \$1/8. That is, $C_{t-1} \geq \$1/8$, $P_{t-1} \geq \$1/8$, $C_t \geq \$1/8$ and $P_t \geq \$1/8$.

(3) Quote midpoints do not violate no-arbitrage conditions, which result in having missing implied volatilities in Optionmetrics.[3] (4) The underlying stock does not pay a dividend during the remaining life of the option. This reduces the potential effects of early exercise, as the options we examine are American.

In contrast, to build a *feasible* version of the ZHCT sample, we impose filters (1), (2), and (3) using only option prices at time $t-1$. Also, differently from ZHCT we exclude an option if the underlying stock paid a dividend during the holding period when the dividend was announced before portfolio formation.[4]

In addition to the filters above, we follow ZHCT and impose the following filters when selecting stock options for both the feasible and infeasible samples: (5) We only retain options with positive total trading volume in the month preceding the portfolio formation. (6) We remove all option observations with bid price larger or equal to the ask price at $t-1$. (7) We exclude options with moneyness lower than 0.8 or higher than 1.2 at time $t-1$, where moneyness is defined as stock price divided by strike.

After applying these filters to each optionable stock, we choose a pair of options (one call and one put) that are closest to being at-the-money (ATM) and have the shortest maturity among options with more than one month to expiration. The vast majority of the options selected each month have the same maturity. We drop options with maturity different from the majority. Finally, we only retain stocks with both call and put options available after filtering. As a result, if a call is not filtered out of the sample with filters (1) to (7) above, but the corresponding put is filtered out, then this last step results in the call being removed

[3]The no-arbitrage conditions imposed are $S \geq C \geq max(0, S - PV(K))$ and $K \geq P \geq max(0, K - S)$, where $PV(K)$ is the present value of the strike price.

[4]This procedure results in a sample with some options that have both the stock dividend announcement and the ex-dividend date during the holding period. Some of these call options may be potentially exercised before their ex-dividend date. We show robustness of our results to these potential early exercises in Appendix B. See Aretz and Gazi (2023) for a discussion of the early exercise premium in put options.

from the sample. To be specific, since the infeasible (feasible) sample applies filters at both time t and $t-1$ (only time $t-1$), this last step accordingly applies at both time t and $t-1$ (only time $t-1$).

Table 1, Panel A shows summary statistics of calls in the infeasible sample. The sample statistics in Panel A are in general close to those reported in ZHCT. For instance, the mean monthly return of writing delta-hedged calls is about 3% with a standard deviation of about 6%, while ZHCT report in their Table 1 a mean return of 3.5% with a standard deviation of 5.65%.

[Insert Table 1 Here]

The returns from the feasible sample, shown in Panel B, are very different. The mean monthly return in the feasible sample is close to zero, with a standard deviation of about 10%. The difference in the mean returns displayed in Panels A and B indicates that the look-ahead bias in the mean return of writing delta-hedged calls in the infeasible sample is about 3% per month. This is a very large bias in monthly returns. In addition, the standard deviation is more than 50% larger in the feasible sample.

The summary statistics in Table 1 reveal that the large bias in the infeasible sample is due to removal of a large number of call options that become either deep-OTM or deep-ITM at time t. Indeed, the infeasible sample has 151,756 observations while the feasible sample has almost twice the number of options, with 274,389 observations. That is, the filters at time t delete almost half of the option-month observations. Not only do these filters remove almost half of the sample, they also remove observations in which delta-hedged call writing has negative returns. In fact, the tenth (first) percentile of the returns in the feasible sample is about -8% (-35%), while it is much higher in the infeasible sample, at about -2% (-16%). Also, the tenth (first) percentile of the moneyness (closing price of the underlying divided by option strike price) of the options at time t is about 83% (59%) in the feasible sample, while it is 89% (73%) in the infeasible sample. That is, the infeasible filters remove call options that become deep-OTM from the sample. Moreover, the 90^{th} (99^{th}) percentile of the moneyness of the options at time t is about 116% (144%) in the feasible sample while it is 112% (131%) in the infeasible sample. Therefore, the infeasible filters also remove call options that become deep-ITM.

The infeasible sample filters out options that become either deep-OTM or deep-ITM at time t because of filters like the one requiring that $C_t \geq \$1/8$ and $P_t \geq \$1/8$. This filter in particular removes the returns of call options with price dropping below \$1/8 at time t, which occurs when the underlying stock price drops significantly ($S_t - S_{t-1} \ll 0$). Moreover, if the price of a put drops below \$1/8 at time t, the return of the matched call is dropped from the infeasible sample.[5] This happens when the underlying stock price increases significantly ($S_t - S_{t-1} \gg 0$).

This selection criterion eliminates many negative returns, resulting in a positive look-ahead bias. Approximating the return of a delta-hedged call writing with a Taylor expansion around $S_t = S_{t-1}$ helps formalize this point. The approximate return is

$$-\frac{\theta_{t-1}}{\Delta_{t-1}S_{t-1} - C_{t-1}} - \frac{\Gamma_{t-1}}{2(\Delta_{t-1}S_{t-1} - C_{t-1})}(S_t - S_{t-1})^2, \tag{3}$$

where θ_{t-1} and Γ_{t-1} are the derivative of the call price with respect to time (in months), and the second derivative of the call price with respect to the price of the underlying stock respectively. When $(S_t - S_{t-1})^2$ is small, Equation 3 results in a positive value since θ_{t-1} is negative. On the other hand, when $(S_t - S_{t-1})^2$ is large, the expression above results in a negative value since Γ_{t-1} of call options is positive. This happens when the price of the underlying moves dramatically because either $S_t - S_{t-1} \ll 0$ or $S_t - S_{t-1} \gg 0$. That is, when call options become deep-ITM or deep-OTM the return of a delta-hedging call writing

[5] Recall that the sample selection criteria require calls to be matched with puts.

is negative. Consequently, a sample selection criterion that removes options that become deep-ITM or deep-OTM results in a positive bias in the mean return to delta-hedged call writing.

While the bias in Table 1 is very large, as we point out in Section 2, this bias may have no effect on the ZHCT results in principle. Therefore, to examine the biases in the long/short strategies in ZHCT we replicate their results with both infeasible and feasible samples.

Table 2 shows the results of the replication of ZHCT. We follow ZHCT and adjust predictors (by multiplying by -1) so that the high portfolio has the highest mean return. The table shows the difference between the equally-weighted average returns of the high-minus-low portfolios.[6] The results of the columns labeled infeasible are very similar to those in ZHCT, with most of the mean returns and t-statistics a close match. As with ZHCT, most of the mean returns, t-statistics, and Sharpe ratios are quite large. Indeed, the mean returns are in general above 1.5% per month, while their respective t-statistics are above ten. With one exception, Sharpe ratios are above two, with several portfolios having values above five. For instance, a long/short position in the bottom/top decile portfolios sorted on the log of the price of the underlying (ln(PRICE)) at the end of the previous month generates a mean excess return of about 5% per month, with a t-statistic of about 26 and a Sharpe ratio of 5.88. The only sorting variable that does not generate the same magnitude of t-statistic and mean returns as ZHCT is the Z-score. Our Z-score results are similar to those in Goyal and Saretto (2022), though.

The large returns and t-statistics reported for the infeasible strategies disappear for the feasible strategies. In contrast to the results in the columns labeled infeasible, the columns labeled feasible in Table 2 show mean returns that are in general not statistically different

[6]The appendix shows results of weighting by market capitalization of the underlying stock and of weighting by the market value of option open interest at the beginning of the holding period. As ZHCT, we winsorize all the sorting variables each month at the 0.5% and 99.5% levels.

from zero. The economic significance of the mean returns is also strongly affected with the adoption of a feasible sample. For instance, the feasible strategy based on ln(PRICE) generates a mean excess return of about 0.6% per month, with a t-statistic of about 3 and a Sharpe ratio of 0.6. Overall, the results in Table 2 indicate that the large Sharpe ratios of the strategies described in ZHCT result from look-ahead biases. We examine these biases in more detail in Section 5.

[Insert Table 2 Here]

4 Illiquidity Premia

CGJK study the relation between option illiquidity and the mean daily returns on delta-hedged calls and puts. CGJK differ from standard practice and use intraday transactions prices to compute option returns, though they use closing prices to compute hedge ratios and stock returns. To formalize the CGJK methodology, define $O_{i,j,t}$ as the dollar volume-weighted average price of option i on stock j over all trades that happen between the open and close of trading day t. We represent the return of option i at time t, computed with these intraday prices, as $R_{O,i,j,t} = O_{i,j,t}/O_{i,j,t-1} - 1$. CGJK computes delta-hedged returns as $DHR_{i,j,t} = R_{O,i,j,t} - \beta_{i,j,t-1}R_{S,j,t}$ where $\beta_{i,j,t-1}$ is the beta of option i with respect to the underlying stock and $R_{S,j,t}$ is the return of the underlying stock j.[7] $\beta_{i,j,t-1}$ is computed from prices observed at the close of day $t-1$, while $R_{S,j,t}$ uses closing prices on both $t-1$ and t. The use of intraday prices to calculate option returns implies that portfolio formation starts with the first trade during period $t-1$. Also, note that because $\beta_{i,j,t-1}$ and $R_{S,j,t}$ are calculated with closing prices at time $t-1$, the positions are effectively unhedged during the portion of day $t-1$ that the trades are in place.

CGJK examine option returns at the stock level. That is, they calculate the equal average

[7] The β of an option is equal to its Δ divided by its price multiplied by the price of the underlying.

of $DHR_{i,j,t}$ across all options on stock j that are within a certain moneyness range at time $t-1$:

$$DHR_{j,t} = \frac{1}{N} \sum_{i=1}^{N} DHR_{i,j,t} \tag{4}$$

We implement the same procedure as CGJK to study the relation between option expected returns and illiquidity. Their main measure of illiquidity is the effective option spread, $ES^O_{j,t}$, which is computed daily at the stock level. The effective spread on day t for option i is twice the average of the absolute percentage difference between transactions prices and contemporaneous quote midpoints, where the average is weighted by the number of contracts transacted. As in the case of returns, CGJK take the equally-weighted average of effective spreads across all options within a certain moneyness range on the same stock j to calculate $ES^O_{j,t}$. At the end of each trading day $t-2$, CGJK create five portfolios of delta-hedged options sorted on $ES^O_{j,t-2}$ and compare the average returns of the top and bottom quintiles.

We merge three datasets in our empirical analysis: CRSP, OptionMetrics, and CBOE LiveVol. Our sample period is from January 2012 to December 2019. The data range for CGJK spans from 2004 to 2012. We selected a different sample period from CGJK for two main reasons. First, in line with Andersen et al. (2021), we observe numerous outliers in the LiveVol data that likely represent errors. While these outliers do not alter our primary conclusions, they introduce noise into our inferences, much more so in the early sample period from 2004 to 2012. Second, perhaps due to the volume of these questionable observations, CBOE LiveVol has ceased offering data from before 2012. In essence, CBOE LiveVol now exclusively provides data from 2012 onwards. As such, our findings would not be replicable if we started our sample in 2004. This poses a significant issue, as emphasized by Spiegel (2019), who notes that having accessible data "is the most fundamental requirement that any article claiming scientific validity must fulfill. Results must be verifiable." Nevertheless, we do have access to a sample of LiveVol data for the same time frame as CGJK with a smaller number of stocks than CGJK, and we detail results for this period in the Internet Appendix. Like CGJK, our sample only includes stocks in the S&P 500 Index. Option hedge ratios and closing prices are from OptionMetrics, and stock returns are from CRSP. LiveVol is used to compute effective option spreads and option returns based on intraday prices. We follow CGJK and consider put and call options with maturities between 30 and 180 days.

To build a sample that replicates the primary results of CGJK, we impose a number of filters. Using OptionMetrics data from the close of day $t-1$, we require that (1) the absolute value of delta is between 0.375 and 0.625, (2) the closing relative bid-ask spread is no larger than 50%, (3) the closing mid-point price is at least \$0.10, (4) the option's open interest is positive, and (5) the closing price, delta, and implied volatility are all non-missing.[8] Additional filters are applied to the LiveVol data on day $t-1$ and day t. Specifically, we only retain transactions for which (6) the transaction price satisfies no-arbitrage conditions and (7) transactions occur when the relative bid-ask spread is 50% or less. We also exclude (8) transactions with negative prices, (9) that occur when the bid-ask spread is not positive, or that have (10) time stamps that don't match the date on the file. Finally, (11) transactions data must exist, meaning that option volume must be positive, a condition that is also imposed on day $t-2$, following CGJK.

The sample in CGJK is therefore infeasible. The use of intraday prices in CGJK to calculate option returns implies that portfolio formation starts with the first trade on day $t-1$. That is, t_0 is the time of the first trade of day $t-1$. In comparison, the sample selection uses criteria based on prices from the close of $t-1$ and, in addition, requires that the option is traded on day t. That is, t_E is the closing of the trading day t. As a result,

[8] Much of CGJK is done using ATM options. Many studies (e.g., Driessen, Maenhout, and Vilkov (2009) and Bollen and Whaley (2004)) also define moneyness as the option delta from OptionMetrics. These filters are standard in the literature and do not result in look-ahead biases when imposed only prior to the holding period (e.g., Goyal and Saretto (2009), Cao and Wei (2010), and Muravyev (2016).

$t_E > t_0$ – the portfolio is selected after the start of the holding period.

In contrast, to build a feasible sample that follows CGJK in spirit, we impose filters (1) to (5) using only prices and deltas at the closing of day $t-2$. Moreover, we apply filters (6) and (7) only on day $t-1$ and (11) only on days $t-2$ and $t-1$. These filters are applied before the options are included in the trader's portfolio and hence are feasible. In the case in which an option is not traded on day t, we use the closing quote on day t to calculate its return.[9] These changes prevent the conditioning on information after the start of the holding period.

Our feasible sample still has some degree of infeasibility because, as in CGJK, we exclude data violating filters (8), (9), and (10) on day t. These filters eliminate only about 1% of the entire LiveVol dataset. Moreover, each of these violations unambiguously signals a data error rather than an actual market condition. Furthermore, violations of each of these conditions are often associated with large outliers. We emphasize that our main results are virtually unaffected by whether these filters are imposed or not, though they have a modest effect on the summary statistics.

Figure 1 clarifies the differences between the feasible and infeasible samples used to analyze the pricing of illiquidity in equity options. Figure 1, Panel A shows the time line of the sample selection in CGJK. Figure 1, Panel B shows the time line of the feasible sample selection. The difference between these two sample selection criteria is emphasized on the differences between these two panels. Specifically, in Panel A the sample selection happens with information available at the closing of day $t-1$, while the sample selection uses information available at the closing of day $t-2$ in Panel B. Moreover, in Panel A, it is required that an option is traded at time t to be included in the sample, while there is no such requirement in Panel B.

[9]This is analogous to the procedure that CRSP uses for computing stock returns. Indeed, when a stock is not traded in a given day, CRSP uses the closing mid-quote to compute the stock return.

[Insert Figure 1 Here]

Figure 2 shows that OTM options have higher effective spreads. Specifically, Figure 2 shows the average effective spreads of calls and puts as function of moneyness defined as the ratio of the underlying closing price (S) by the option strike price (K). In the case of calls (puts) effective spreads are decreasing (increasing) with S/K indicating that OTM options have larger effective spreads than ITM options. In both cases, effective spreads increase steeply from ATM ($S/K = 1$) to OTM options.

[Insert Figure 2 Here]

Table 3 displays the summary statistics of the delta-hedged option returns (DHR) and option effective spreads (ES^O) in our sample. Panel A shows summary statistics on mean delta-hedged returns and underlying stock returns. Panel B shows summary statistics on effective option spreads. Even though our sample period is different from that used in CGJK, our summary statistics on mean returns are not very different. As in CGJK, we find that the mean DHR is positive for both calls and puts. Moreover, the mean returns for the underlying stocks are also very similar to those in CGJK. The mean average effective spread in our sample is somewhat smaller than those in CGJK. This is consistent with the observation, described in CGJK, that effective spreads have been decreasing over time.

Table 3, Panel A shows that the mean returns of calls and puts have a strong positive look-ahead bias in the infeasible sample. Indeed, the difference between the mean DHR of calls in the infeasible and feasible sample is 0.46% per day (0.47-0.01) and is 0.66% per day (0.63+0.03) for puts. These are large values for daily returns.

The significant look-ahead biases arise from the fact that the infeasible sample only includes options at the time of portfolio formation (i.e., the time of the first trade on day $t-1$) that are ATM by the middle of the holding period (i.e., the close of day $t-1$). To understand this, consider that a call option which is ATM at the end of day $t-1$ may have been either OTM or ITM at the opening of $t-1$, where any change in moneyness is driven by movements in the underlying stock price. Call options transition from OTM to ATM when the underlying stock price rises, and from ITM to ATM when the underlying stock falls. Consequently, selecting call options that are ATM at the closing of $t-1$ results in the selection of more calls that were OTM (ITM) at the open of day $t-1$ when the underlying stock increases (decreases) between the open and close of day $t-1$.

More importantly, this selection criterion leads to a sample that is more (less) sensitive to the underlying stock return when the underlying stock rises (falls). This is because the returns of OTM options are more responsive to changes in the underlying stock price than those of ITM options.[10] Further, recall that a portion of DHR_t is the unhedged option return between the opening and the closing of $t-1$ and hence is substantially driven by the option sensitivity to the underlying stock return over the same period. Consequently, the mean return (DHR_t) of the infeasible sample becomes upward biased, as symmetric positive and negative stock returns have asymmetric effects on the options included in the sample.

Table 3, Panel C illustrates these effects for call options by showing the mean moneyness (S/K), β, and DHR separately for $R_{S,t-1} > 0$ and $R_{S,t-1} < 0$. Notably, when $R_{S,t-1} > 0$, moneyness is about two percentage points smaller, which results in a larger β and in a absolute value of the mean DHR 0.29% *higher* than the absolute value of the mean DHR when $R_{S,t-1} < 0$ (3.72% vs. 3.43%). That is, the selected infeasible sample of call options is more (less) sensitive to the underlying stock return when the underlying stock goes up (falls).

These look-ahead biases are absent in the feasible sample. As we emphasize in Session 2, look-ahead biases occur when the sample selection criterion exhibits correlation with the

[10]OTM options have higher absolute β_S values compared to ITM options.

returns during the holding period. These biases occur in the infeasible sample because call options that are more (less) sensitive to the returns of the underlying stock are included in the sample when the returns of the underlying stock are positive (negative) during the holding period. In contrast, the feasible sample selection criteria are uncorrelated with the returns during the holding period as they are based on information available prior to the beginning of the holding period (i.e., end of day $t-2$). Indeed, Table 3, Panel C demonstrates that when $R_{S,t-1} > 0$, moneyness is only 0.15 percentage points lower than when $R_{S,t-1} < 0$. As a result, the absolute value of the mean DHR in the feasible sample when $R_{S,t-1} > 0$ is 0.31% *lower* than the absolute value of the mean DHR when $R_{S,t-1} < 0$ (3.39% vs. 3.70%).

This bias is similar for puts, except that the transition from OTM or ITM to ATM requires a stock return of the opposite sign. Table 3, Panel D illustrates the results for puts. When $R_{S,t-1} < 0$, the absolute value of the mean put DHR is 1.03% greater than the absolute value of the mean put DHR when $R_{S,t-1} > 0$ (4.16% vs. 3.13%). As with calls, this generates an upward bias in the unconditional return on the infeasible portfolio. Returns on the feasible portfolio are unconditionally lower and less dependent on the sign of $R_{S,t-1}$.

Table 3, Panels C and D, also demonstrate that the selection of the infeasible sample also has an impact on the effective spreads of options. As a reminder, Figure 2 illustrates that OTM options exhibit larger effective spreads compared to ATM or ITM options. Consequently, the fact that the infeasible sample includes more OTM calls (puts) on days when $R_{S,t-1} > 0$ ($R_{S,t-1} < 0$) leads to higher mean spreads for calls (puts) when $R_{S,t-1} > 0$ ($R_{S,t-1} < 0$). Indeed, when $R_{S,t-1} > 0$, the mean effective spread of calls (Panel C) in the infeasible sample is about three quarters of a percentage point higher (4.85% vs. 4.12%) than the mean spread when $R_{S,t-1} < 0$. Similarly, when $R_{S,t-1} < 0$, the mean effective spread of puts (Panel D) in the infeasible sample is about half a percentage point (4.13% vs. 3.68%) higher than the mean spread when $R_{S,t-1} > 0$. The patterns are significantly less pronounced in the feasible sample.

[Insert Table 3 Here]

While the bias in Table 3 is very large, as we point out in Section 2, biases may have no effect on the CGJK results in principle. To examine the biases in the estimation of the illiquidity premium in options, we replicate their main result with both the infeasible and feasible samples.

Table 4 shows the results of the replication of CGJK with both samples. We report mean returns on portfolios that go long options with high effective spreads and short options with low effective spreads. Both equally-weighted (EW) and gross-return-weighted (GWR) means are reported, the latter to address the microstructure biases described in Duarte, Jones, and Wang (2022).[11]

[Insert Table 4 Here]

The results in the columns labeled infeasible in Table 4 are very similar to those in CGJK. Indeed, the equally-weighted mean returns, t-statistics, and Sharpe ratios for the infeasible strategy are large. The illiquid calls outperform the liquid calls by more than 1% per day, while differences for puts are also large, at around 0.9% per day. Corresponding t-statistics are all above eleven, and long/short portfolio Sharpe ratios are all above three. There is not much of difference between the EW and GRW results, at least for the infeasible sample.

The results with feasible strategy paint a different picture. Indeed, the large returns, t-statistics, and Sharpe ratios reported for the infeasible strategy disappear for the feasible

[11]Equal weighted portfolio returns can be written as $\sum_{j=1}^{M_t}\sum_{i=1}^{N_{j,t}}(1/M_t)(1/N_{j,t})DHR_{i,j,t}$, where M_t is the number of stocks at date t, $N_{j,t}$ is the number of options on stock j on that date, and $DHR_{i,j,t}$ is the delta-hedged return of option i on stock j. Gross return weighted portfolios are instead computed as $\sum_{j=1}^{M_t}\sum_{i=1}^{N_{j,t}} w_{i,j,t}DHR_{i,j,t}$, where $w_{i,j,t} \propto (1/N_{j,t}) \times (O_{j,i,t-1}/O_{j,i,t-2})$, O is the option price, and weights are normalized to sum to 1 for each t.

strategy. Moreover, the gross return-weighted put means are not statistically different from zero. Thus, the results in Table 4 indicate that liquidity is priced only in call options once we correct for microstructure biases described in Duarte, Jones, and Wang (2022). The large differences between the feasible and infeasible mean returns indicate look-ahead bias explains most of the pricing of illiquidity in options.

The economic significance of the estimated price of illiquidity is strongly affected with the adoption of a feasible sample. For instance, a feasible long/short strategy on the top/bottom quintile portfolio of delta-hedged calls sorted on illiquidity generates a mean excess return of about 0.19% per day, with a t-statistic of about 4.9 and a Sharpe ratio of 1.73. In contrast, the equivalent infeasible strategy generates a mean excess return of about 1.52% per day, with a t-statistic of about 31 and a Sharpe ratio of 11. Overall, the results in Table 4 indicate that the impressive Sharpe ratios earned from providing liquidity in options largely result from look-ahead biases. We discuss the precise source of these biases in Section 5.

5 Understanding the Look-Ahead Biases

To understand why the difference in the mean returns of infeasible and feasible strategies in Table 2 is due to look-ahead biases, recall that the results in Table 1 show that the infeasible sample in ZHCT filters out options that end up either deep-OTM or deep-ITM at time t. Removing options with extreme moneyness also drives the results in Table 2. To show this, we analyze how the infeasible filters affect changes in the moneyness of the options in the High and Low return portfolios of Table 2. The results are in Table 5.

Panel A of Table 5 shows the changes in the 10^{th} percentile of moneyness (S/K) between times $t-1$ and t. Because options start off ATM, changes in the 10^{th} percentile of moneyness are universally negative, but the changes are in most cases significantly larger for the feasible sample. This is because the infeasible sample removes a large fraction of the calls that become

OTM by time t.

More importantly, the effects of filtering OTM calls differs between the high and low deciles. For every characteristic, the "DiD" column shows that the impact of filtering is higher for the high decile portfolio than it is for the low decile portfolio. That is, in every case, more OTM options are removed from the top decile portfolio than from the bottom. This largely explains why it appears to be more profitable to write options for the high decile.

Similarly to Panel A, the results in Table 5, Panel B show that infeasible filters that remove ITM calls at time t also have a stronger effect in the High portfolios than in the Low portfolios. Specifically, Panel B displays the changes in the 90^{th} percentile of moneyness (S/K) between $t-1$ and t. The negative values in the DiD column in Panel B indicate that the infeasible filters result in a smaller increase in the 90^{th} percentile of moneyness for the High portfolio compared with the Low portfolio. Therefore, the infeasible filters remove more deep-ITM calls from the high decile portfolio, again raising the the returns to call writing for that decile.

[Insert Table 5 Here]

The differential effects presented in Table 5 should naturally be expected to result in look-ahead biases in returns. To understand the strength of this effect, we plot the bias of each strategy from Table 2 as a function of the DiD displayed in Table 5. The result, in Figure 3, Panel A, is an unmistakably positive and near-linear relation between the DiD in the changes in the 10^{th} percentile of moneyness and the look-ahead biases from Table 2. Similarly, Panel B of Figure 3 illustrates an extremely strong negative relation between the DiD in the change in the 90^{th} percentile of moneyness and the same look-ahead biases from Table 2. Indeed, the relations in Panels A and B represent correlations of approximately 98%

and -96%, respectively. This suggests that differences in the mean returns of the infeasible and feasible strategies can be almost entirely attributed to the filters that remove options with extreme moneyness from the feasible sample at time t.

[Insert Figure 3 Here]

Naturally, a lingering question pertains to why the infeasible filters exert differential impacts on the High and Low portfolios. To address this, Figure 3, Panel C plots the look-ahead bias as a function of the average difference in the implied volatilities of the call options in the High and Low portfolios. Panel C reveals a substantial correlation between the difference in implied volatility and the look-ahead bias, with a correlation coefficient of approximately 70%. Consequently, the look-ahead bias in ZHCT stems from the fact that the call options in their High portfolios are more likely on stocks that exhibit higher volatility, making them more prone to end up either deep-ITM or deep-OTM by the conclusion of the holding period and, as a result, being removed from the sample.

The disparity in the mean returns of the infeasible and strategies of CGJK and their feasible counterparts is also due to look-ahead biases. To understand this, we recall two key findings from the results in Table 3.

Firstly, the selection criteria results in the infeasible sample only including options at the time of portfolio formation (i.e., the time of the first trade on day $t-1$) that become ATM by the middle of the holding period (i.e., the close of day $t-1$). Consequently, the infeasible sample tends to select OTM call (put) options when the stock return on day $t-1$ is positive (negative) and ITM call (put) options when the stock return on day $t-1$ is negative (positive).

Secondly, OTM options tend to have higher spreads. Thus, the fact that the infeasible sample includes a greater number of OTM calls (puts) on days when $R_{S,t-1} > 0$ ($R_{S,t-1} < 0$) results in higher mean spreads (ES^O_{t-2}) for calls (puts) when $R_{S,t-1} > 0$ ($R_{S,t-1} < 0$).

These two characteristics of the infeasible sample also underlie the results in Table 4. To illustrate this, consider that stocks with positive (negative) returns during the holding period tend to have call options that were OTM (ITM) on day $t-1$ included in the sample. Since moneyness is linked to ES^O_{t-2}, stocks with positive (negative) returns during the holding period tend to have a larger representation of illiquid (liquid) call options within the sample. Likewise, stocks with positive (negative) returns during the holding period tend to feature a higher proportion of liquid (illiquid) put options in the sample. Consequently, sorting call portfolios on ES^O_{t-2} in the infeasible sample is akin to sorting them based on $R_{S,t-1}$. When we sort put portfolios in the infeasible sample by ES^O_{t-2}, it is akin to sorting them based on $-R_{S,t-1}$.

Table 6 illustrates these points by showing a variety of quintile averages for portfolios of delta-hedged calls and puts sorted by illiquidity (ES^O_{t-2}). Panel A (B) shows the results for the infeasible (feasible) sample.

[Insert Table 6 Here]

Panel A results show a strong relation between ES^O_{t-2} and delta-hedged returns (DHR). It also shows that calls with high ES^O_{t-2} are on average more OTM prior to the holding period (lower $(S/K)_{t-2}$) and are written on stocks with much higher stock returns on day $t-1$ ($R_{S,t-1}$). The result is that the mean DHR is higher as well.[12] In fact, Table 6 demonstrates that a trading strategy involving short selling stocks with liquid calls and buying stocks with illiquid calls between $t-2$ and $t-1$ yields an average return of 0.52% per day, with a t-statistic of 42.80. This strategy boasts a remarkable Sharpe ratio of approximately 15. However, it's important to note that this strategy is infeasible. The results for puts show

[12]Recall that DHR is driven by the unhedged option return on day $t-1$ since the delta hedge is not put in place until the close of day $t-1$. Hence the call return is mainly the result of a positive β times a positive $R_{S,t-1}$.

that puts with high ES^O_{t-2} are more OTM prior to the holding period (higher $(S/K)_{t-2}$) and are written on stocks with much lower mean stock returns $R_{S,t-1}$. Hence, the strong relation between ES^O_{t-2} and DHR in the infeasible sample can be attributed to the fact that $R_{S,t-1}$ is linked to both ES^O_{t-2} and DHR within this sample.

The findings in Panel B indicate that there exists only a weak relation between ES^O_{t-2} and DHR in the feasible sample. High ES^O_{t-2} calls are more likely to be OTM, characterized by lower $(S/K)_{t-2}$ values. However, in contrast to the results observed in the infeasible sample, these high ES^O_{t-2} calls are not written on stocks with significantly higher mean returns ($R_{S,t-1}$). Similar patterns emerge for puts within this sample. Therefore, the feasible sample selection process eliminates the spurious correlation between $R_{S,t-1}$ and ES^O_{t-2} observed in the infeasible sample. This, in turn, results in considerably weaker evidence suggesting that illiquidity is priced in options.

6 Conclusion

We demonstrate that look-ahead bias can explain some of the stylized facts in the options empirical literature. This bias arises from the application of sample filters that employ information not available at the time of portfolio formation. Notably, filters associated with option moneyness appear to play a particularly crucial role in generating look-ahead biases.

The first stylized fact is that numerous stock characteristics seem to be priced in the cross-section of individual equity option returns. Our findings indicate a narrower set of stock characteristics associated with option expected returns compared to ZHCT. In contrast to ZHCT, who identified a robust correlation between delta-hedged option returns and ten stock characteristics, our results suggest that many of these relations are due to look-ahead biases. Specifically, within the feasible sample, we find that only three stock characteristics exhibit a modest correlation with the cross-section of delta-hedged option returns.

The second stylized fact is that illiquidity seems to be strongly priced in individual equity options. We provide evidence suggesting that illiquidity is not nearly as strongly priced in the cross-section of options returns as implied by CGJK. Specifically, we observe evidence of illiquidity being priced only in call options when employing the bias correction procedure outlined by Duarte, Jones, and Wang (2022). Remarkably, we do not find evidence indicating that illiquidity is priced in put options within our feasible sample.

References

Altman, Edward I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23, 589–609.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Andersen, Torben, Ilya Archakov, Leon Grund, Nikolaus Hautsch, Yifan Li, Sergey Nasekin, Ingmar Nolte, Manh Cuong Pham, Stephen Taylor, and Viktor Todorov, 2021, A descriptive study of high-frequency trade and quote option data, *Journal of Financial Econometrics* 19, 128–177.

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Aretz, Kevin, and Adnan Gazi, 2023, The early exercise risk premium, *Management Science* Forthcoming.

Aretz, Kevin, Ming-Tsung Lin, and Ser-Huang Poon, 2022, Moneyness, underlying asset volatility, and the cross-section of option returns, *Review of Finance* 27, 289–323.

Asparouhova, Elena, Hendrick Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, *Journal of Financial Economics* 96, 215–237.

———, 2013, Noisy prices and inference regarding returns, *Journal of Finance* 68, 665–714.

Bali, Turan G, Heiner Beckmeyer, Mathis Mörke, and Florian Weigert, 2023, Option return predictability with machine learning and big data, *Review of Financial Studies* 36, 3548–3602.

Baquero, Guillermo, Jenke ter Horst, and Marno Verbeek, 2005, Survival, look-ahead bias, and persistence in hedge fund performance, *Journal of Financial and Quantitative Analysis* 40, 493–517.

Blume, Marshall, and Robert Stambaugh, 1983, Biases in computed returns, *Journal of Financial Economics* 12, 387–404.

Blume, Marshall E., and Frank Husic, 1973, Price, beta, and exchange listing, *Journal of Finance* 28, 283–299.

Bollen, Nicolas P. B., and Robert E. Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* 59, 711–753.

Bradshaw, Mark T., Scott A. Richardson, and Richard G. Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53–85.

Brenner, Menachem, Rafi Eldor, and Shmuel Hauser, 2001, The price of options illiquidity, *Journal of Finance* 56, 789–805.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance* 50, 853–873.

Cao, Jie, Amit Goyal, Xintong Zhan, and Weiming E. Zhang, 2023, Unlocking ESG premium from options, *Working Paper*.

Cao, Jie, and Bing Han, 2013, Cross section of option returns and idiosyncratic stock volatility, *Journal of Financial Economics* 108, 231–249.

Cao, Melanie, and Jason Wei, 2010, Option market liquidity: Commonality and other characteristics, *Journal of Financial Markets* 13, 20–48.

Carozza, Dick, 2009, Chasing Madoff An Interview with Harry Markopolos, *Fraud Magazine* https://www.fraud-magazine.com/article.aspx?id=313.

Carpenter, Jennifer N., and Anthony W. Lynch, 1999, Survivorship bias and attrition effects in measures of performance persistence, *Journal of Financial Economics* 54, 337–374.

Choy, Siu Kai, and Jason Wei, 2023, Investor attention and option returns, *Management Science* 69, 4845–4863.

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2018, Illiquidity premia in the equity options market, *Review of Financial Studies* 31, 811–851.

Cochrane, John H., and Jesus Saa-Requejo, 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79–119.

Cosma, Antonio, Stefano Galluccio, Paola Pederzoli, and Olivier Scaillet, 2020, Early exercise decision in American options with dividends, stochastic volatility, and jumps, *Journal of Financial and Quantitative Analysis* 55, 331–356.

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

Dichev, Ilia D., 1998, Is the risk of bankruptcy a systematic risk?, *Journal of Finance* 53, 1131–1147.

Diether, Karl, Christopher Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance* 57, 2113–2141.

Driessen, Joost, Pascal J. Maenhout, and Grigory Vilkov, 2009, The price of correlation risk: Evidence from equity options, *Journal of Finance* 64, 1377–1406.

Duarte, Jefferson, Christopher S. Jones, and Junbo Wang, 2022, Very noisy option prices and inference regarding the volatility risk premium, *Journal of Finance* Forthcoming.

Eraker, Bjørn, and Daniela Osterrieder, 2018, Market maker inventory, bid-ask spreads, and the computation of option implied risk measures, *Working paper.*

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stock bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2006, Profitability, investment and average returns, *Journal of Financial Economics* 82, 491–518.

Goncalves-Pinto, Luis, Bruce D. Grundy, Allaudeen Hameed, Thijs van der Heijden, and Yichao Zhu, 2020, Why do option prices predict stock returns? The role of price pressure in the stock market, *Management Science* 66, 3903–3926.

Goyal, Amit, and Alessio Saretto, 2009, Cross-section of option returns and volatility, *Journal of Financial Economics* 94, 310–326.

———, 2022, Are equity option returns abnormal? IPCA says no, *Working paper.*

Goyenko, Ruslan, and Chengyu Zhang, 2019, Option returns: Closing prices are not what you pay, *Working paper.*

Haugen, Robert, and Nardin Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Muravyev, Dmitriy, 2016, Order flow and expected option returns, *Journal of Finance* 71, 673–708.

Palazzo, Berardino, 2012, Cash holdings, risk, and expected returns, *Journal of Financial Economics* 104, 162–185.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *Journal of Finance* 63, 921–945.

Ramachandran, Lakshmi Shankar, and Jitendra Tayal, 2021, Mispricing, short-sale constraints, and the cross-section of option returns, *Journal of Financial Economics* 141, 297–321.

Soliman, Mark T., 2008, The use of DuPont analysis by market participants, *Accounting Review* 83, 823–853.

Spiegel, Matthew, 2019, Editorial: Replication? Do you even have access to the data?, *Critical Finance Review* 8, 11–13.

ter Horst, Jenke R., Theo E. Nijman, and Marno Verbeek, 2001, Eliminating look-ahead bias in evaluating persistence in mutual fund performance, *Journal of Empirical Finance* 8, 345–373.

Zhan, Xintong (Eunice), Bing Han, Jie Cao, and Qing Tong, 2021, Option Return Predictability, *Review of Financial Studies* 35, 1394–1442.

Figure 1: **Sample selection to analyze illiquidity premium in options.** Panels A shows the time line of the sample selection in Christoffersen, Goyenko, Jacobs, and Karoui (2018). Panel B shows the time line of the feasible sample selection.

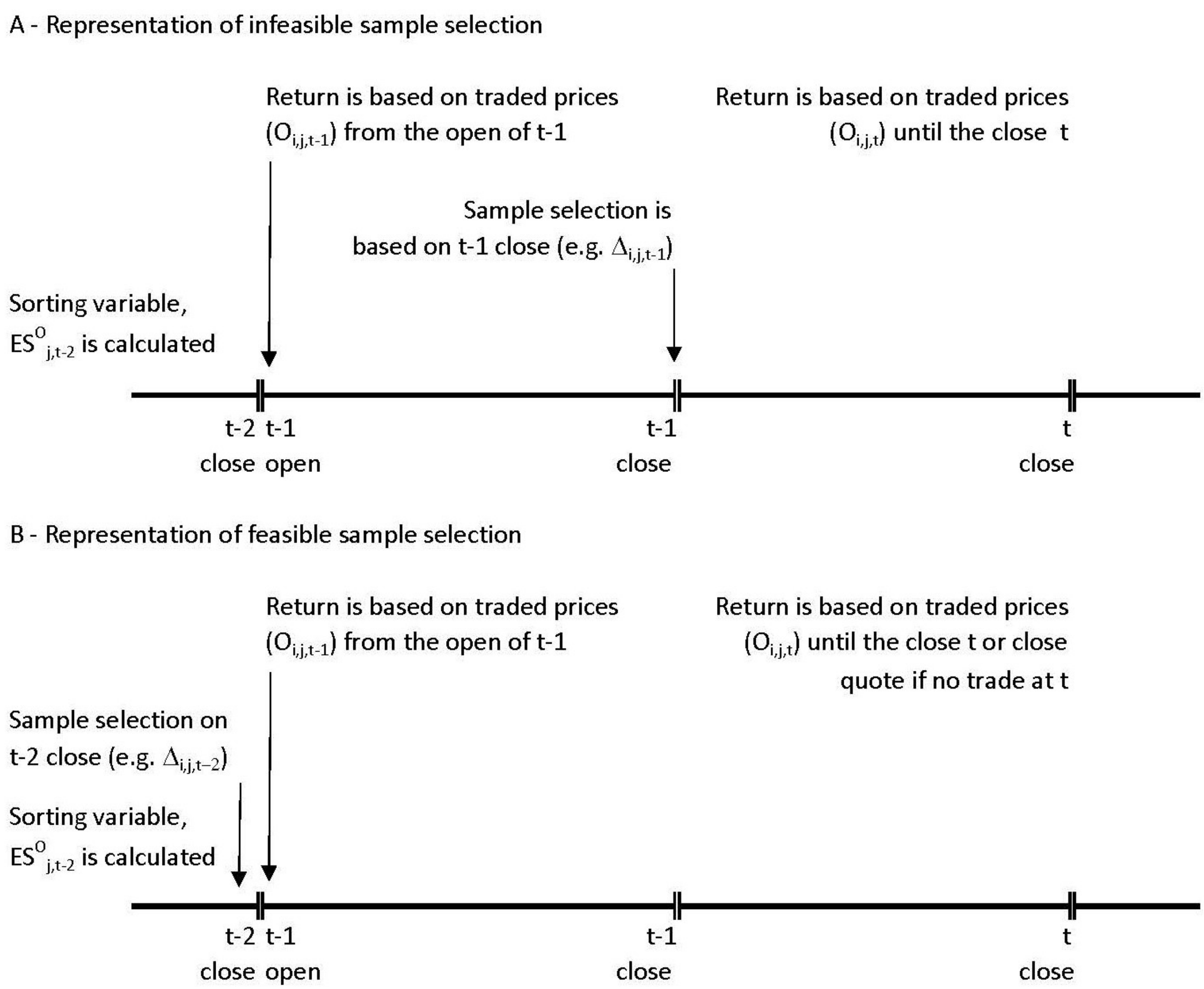

Figure 2: **Option Illiquidity as Function of Moneyness.** This figure shows the average of the effective spreads across all options within different moneyness ranges. Moneyness is defined as the closing underlying price (S) divided by the option strike price (K). Panel A (B) shows the results for calls (puts).

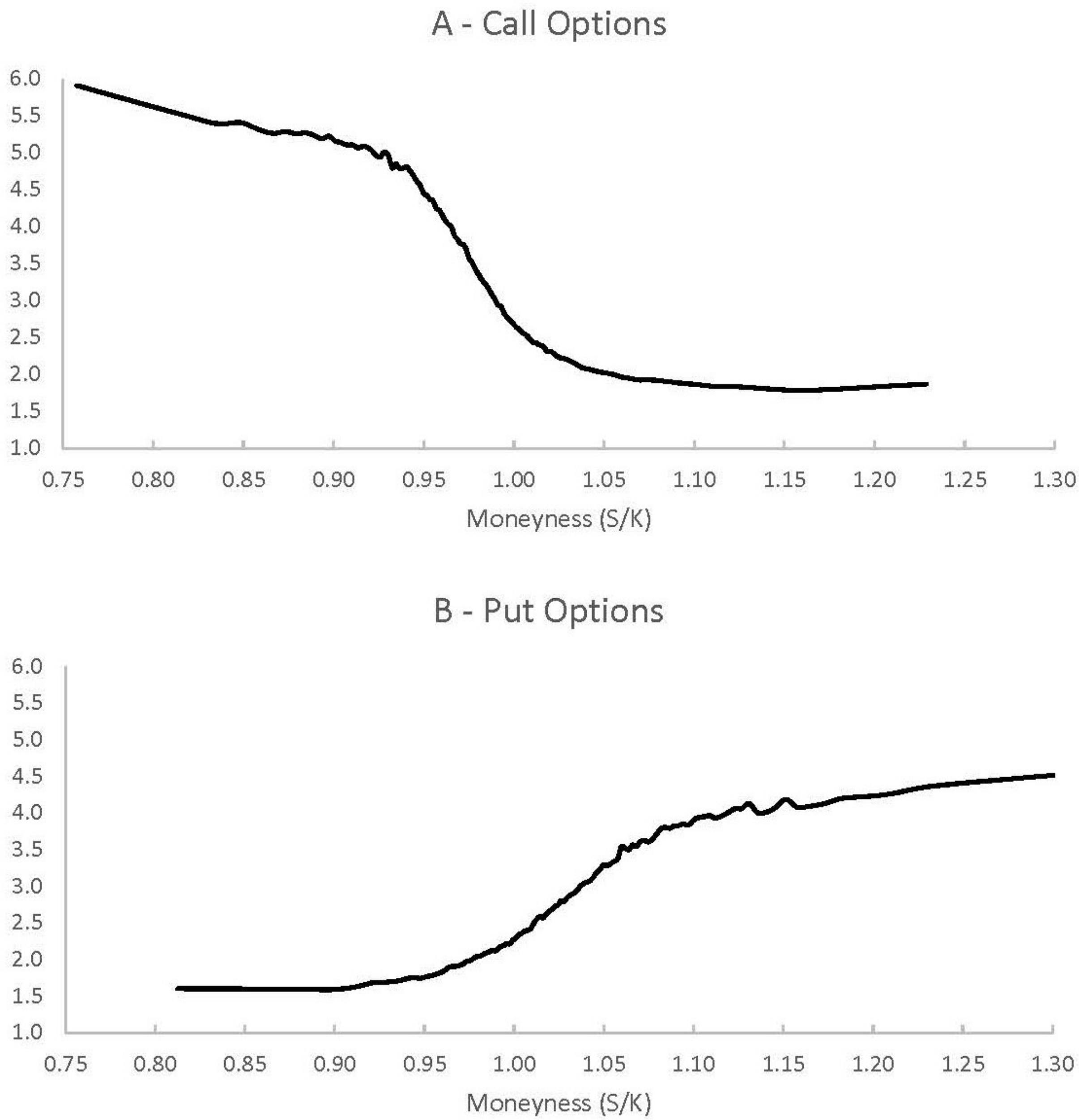

Figure 3: **Look-ahead bias in Zhan, Han, Cao, and Tong (2021) trading strategies.** This figure presents the look-ahead biases for the different strategies in Zhan, Han, Cao, and Tong (2021) as function of the DiD of the changes in the 10^{th} (Panel A) and 90^{th} (Panel B) percentiles of moneyness for the High and Low portfolios in Table 5. The look-ahead biases are the differences between the infeasible and feasible equal weighted mean returns in Table 2. Panel C presents the look-ahead biases as function of the difference in the mean implied volatilities of the call options in the High and Low portfolios.

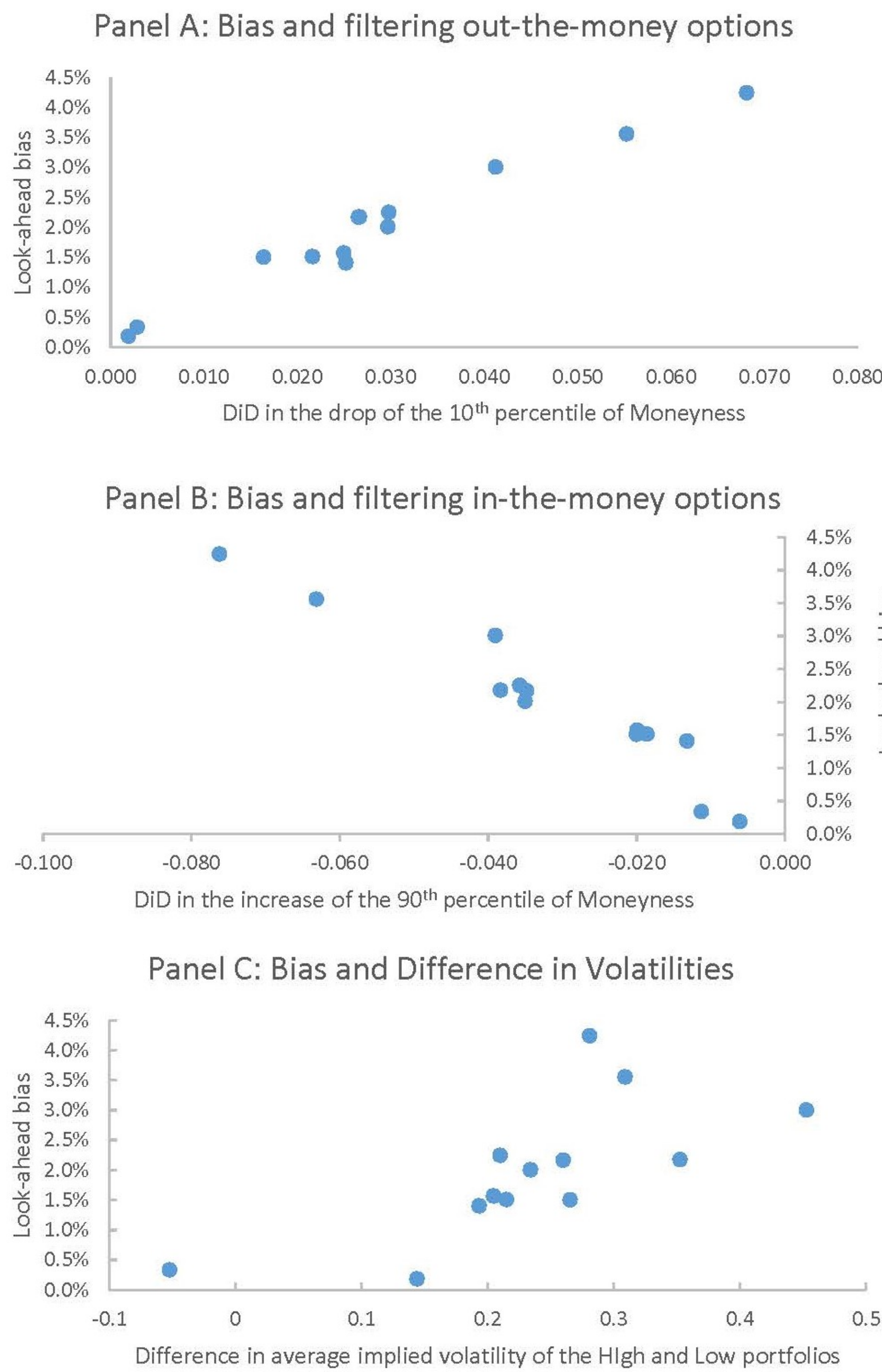

Table 1: **Summary statistics as in Zhan, Han, Cao, and Tong (2021).** This table displays the summary statistics of samples used to replicate the Zhan, Han, Cao, and Tong (2021) results. "Buy & hold until month-end" refers to the return of writing delta-hedged call options from the end of the previous month ($t-1$) to the end of the current month (t). We follow Zhan, Han, Cao, and Tong (2021) and define option moneyness as the closing stock price (S) divided by the option strike price (K). Panel A shows the pooled sample summary statistics for the infeasible sample. Panel B shows the pooled sample summary statistics for the feasible sample. The infeasible and feasible filters used to create these samples are described in Section 3. The data are from Optionmetrics database.

Panel A: Infeasible sample

	Mean	Std. Dev.	1^{st}	10^{th}	25^{th}	Median	75^{th}	90^{th}	99^{th}
Buy & hold until month-end (%)	3.19	6.29	-15.83	-2.13	0.95	3.14	5.78	9.17	18.87
Moneyness$_{t-1}$ (S/K) (%)	99.53	4.68	86.25	94.03	97.00	99.67	102.12	104.98	112.20
Moneyness$_t$ (S/K) (%)	100.43	10.67	73.00	89.00	94.73	100.09	105.75	112.00	130.77
Days to maturity	49.99	1.97	45.00	47.00	50.00	50.00	51.00	52.00	53.00
Vega	0.14	0.01	0.11	0.13	0.14	0.14	0.15	0.15	0.15
Quoted bid-ask spread$_{t-1}$ (%)	15.56	15.34	1.68	4.35	7.06	11.43	18.18	30.30	80.00
Quoted bid-ask spread$_t$(%)	29.19	31.48	1.05	4.88	8.70	17.14	38.46	66.67	142.86

Panel B: Feasible sample

	Mean	Std. Dev.	1^{st}	10^{th}	25^{th}	Median	75^{th}	90^{th}	99^{th}
Buy & hold until month-end (%)	0.25	9.56	-34.89	-8.45	-1.81	1.80	4.50	7.73	16.89
Moneyness$_{t-1}$ (S/K) (%)	99.18	05.10	84.50	93.09	96.47	99.44	102.02	105.00	112.80
Moneyness$_t$ (S/K) (%)	99.81	15.50	59.30	82.82	92.00	99.84	107.31	116.00	144.00
Days to maturity	49.91	2.02	45.00	47.00	50.00	50.00	51.00	52.00	53.00
Vega	0.14	0.01	0.09	0.13	0.14	0.14	0.15	0.15	0.15
Quoted bid-ask spread$_{t-1}$ (%)	18.93	20.52	1.72	4.71	7.69	12.77	22.22	40.00	114.29
Quoted bid-ask spread$_t$ (%)	52.43	65.78	1.34	4.96	9.09	20.47	66.67	200.00	200.00

Table 2: **Average return of long-short portfolios of delta-neutral written calls sorted by underlying stock characteristics.** This table shows the results of the replication of Zhan, Han, Cao, and Tong (2021) with infeasible and feasible samples. All portfolios are equal-weighted. This table displays the difference between the returns of the high and low decile portfolios formed on various characteristics. The signs of some characteristics are switched so that the high-low average return is always positive for the infeasible sample. Average returns are in percent. Annualized Sharpe ratios are within brackets and t-statistics are in parentheses.

	Infeasible	**Feasible**	**Bias**
CFV	2.12	-0.13	2.25
	(16.99)	(-0.94)	(16.29)
	[3.78]	[-0.21]	
CH	2.10	0.60	1.51
	(13.90)	(2.70)	(10.29)
	[3.09]	[0.60]	
DISP	1.91	-0.11	2.01
	(14.40)	(-0.69)	(14.23)
	[3.20]	[-0.15]	
ISSUE_1Y	1.62	0.20	1.41
	(13.68)	(1.59)	(12.99)
	[3.04]	[0.35]	
ISSUE_5Y	1.82	0.31	1.52
	(14.16)	(2.14)	(13.16)
	[3.15]	[0.48]	
TEF	1.59	0.01	1.58
	(11.07)	(0.08)	(12.01)
	[2.46]	[0.02]	
-PM	2.30	0.12	2.18
	(16.47)	(0.66)	(14.50)
	[3.66]	[0.15]	
-LN(PRICE)	4.85	0.61	4.25
	(26.48)	(2.74)	(21.70)
	[5.88]	[0.61]	
-PROFIT	2.27	0.10	2.17
	(19.18)	(0.54)	(14.81)
	[4.26]	[0.12]	
-ZS	0.44	0.10	0.34
	(2.63)	(0.58)	(2.02)
	[0.58]	[0.13]	
-VOL_deviation	3.00	2.81	0.19
	(13.59)	(11.53)	(1.49)
	[3.02]	[2.56]	
IVOL	3.78	0.77	3.01
	(25.18)	(3.75)	(16.72)
	[5.60]	[0.83]	
AMIHUD	3.78	0.22	3.56
	(24.40)	(1.21)	(23.35)
	[5.42]	[0.27]	

Table 3: **Summary statistics as in Christoffersen, Goyenko, Jacobs, and Karoui (2018).** This table presents summary statistics for the samples used to estimate the illiquidity premium in option returns. The end of the holding period is at the close of day t. Panel A presents summary statistics for delta-hedged returns (DHR_t) on calls and puts as well as on the excess returns of the underlying stocks ($R_{S,t}$). DHR_t is an average of the returns on all ATM options for a given stock on a single day. Each option's return is constructed from average transaction prices on days $t-1$ and t, but the delta hedge is not put into place until the close of day $t-1$. As in CGJK, we compute the descriptive statistics for each stock and then we take the averages of these statistics across stocks. We report the mean, standard deviation, skewness, kurtosis, first-order autocorrelation of delta-hedged returns ($\rho(1)$), and first-order autocorrelation of the absolute value of delta-hedged returns, ($|\rho(1)|$). We also report the average number of stocks in each cross-section and the average number of option series for each stock on each day. Panel B presents statistics on the effective option spreads on day $t-2$ (ES^O_{t-2}). For each stock and on each day, we compute ES^O_{t-2} from all the available options in the sample and then take the time series mean, standard deviation, minimum, maximum, and first-order autocorrelation ($\rho(1)$) for each stock. Panel B displays the averages of these statistics across stocks. Panels C and D present the means of moneyness $(S/K)_{t-2}$, β_{t-2}, DHR_t, and ES^O_{t-2} on days in which the returns on the underlying stock in the first day of the holding period ($R_{S,t-1}$) are either positive or negative. The moneyness of an option is defined as the closing price of the underlying stock (S) divided by the option's strike price (K). The option β is its delta divided by its price multiplied by the price of the underlying stock. Filters used in sample selection are described in Section 4. The sample only contains ATM options, defined for the infeasible (feasible) sample as options with the absolute value of δ_{t-1} (δ_{t-2}) between 0.375 and 0.625. Returns and effective spreads are in percent. The sample includes all S&P 500 constituents with any valid traded options data between January 2012 and December 2019.

Panel A: Returns

	Call		Put		Stock
	Infeasible	**Feasible**	**Infeasible**	**Feasible**	**Return**
Mean	0.47	0.01	0.63	-0.03	0.05
Std. Dev.	13.69	13.67	13.26	12.98	1.70
Skewness	1.12	1.14	1.27	1.23	0.18
Kurtosis	20.46	26.27	14.95	19.42	17.43
$\rho(1)$	-0.31	-0.36	-0.29	-0.36	0.00
$\lvert\rho(1)\rvert$	0.20	0.25	0.19	0.25	0.11
Average # of Stocks	353	429	282	388	475
Average # of Series	4.05	3.80	3.27	2.89	

Panel B: Effective Spread

	Call		Put	
	Infeasible	**Feasible**	**Infeasible**	**Feasible**
Mean	4.52	4.56	3.92	4.02
Std. Dev.	3.64	3.62	3.71	3.55
Minimum	0.09	0.13	0.03	0.07
Maximum	40.81	41.23	41.48	40.78
$\rho(1)$	0.23	0.25	0.20	0.22

Panel C: Averages for Calls when $R_{S,t-1} > 0$ and when $R_{S,t-1} < 0$

	Infeasible				Feasible			
	$(S/K)_{t-2}$ (1)	β_{t-2} (2)	DHR_t (3)	ES^O_{t-2} (4)	$(S/K)_{t-2}$ (5)	β_{t-2} (6)	DHR_t (7)	ES^O_{t-2} (8)
$R_{S,t-1} > 0$	98.03	13.89	3.72	4.85	99.04	12.51	3.39	4.60
$R_{S,t-1} < 0$	100.06	12.64	-3.43	4.12	99.21	12.53	-3.70	4.50

Panel D: Averages for Puts when $R_{S,t-1} > 0$ and when $R_{S,t-1} < 0$

	Infeasible				Feasible			
	$(S/K)_{t-2}$ (1)	β_{t-2} (2)	DHR_t (3)	ES^O_{t-2} (4)	$(S/K)_{t-2}$ (5)	β_{t-2} (6)	DHR_t (7)	ES^O_{t-2} (8)
$R_{S,t-1} > 0$	99.39	-11.46	-3.13	3.68	99.96	-10.99	-3.24	3.97
$R_{S,t-1} < 0$	101.53	-12.39	4.16	4.13	100.18	-11.09	3.70	4.08

Table 4: **Average return of long-short portfolios of hedged options sorted by option illiquidity.** This table shows the results of the replication of Christoffersen, Goyenko, Jacobs, and Karoui (2018) with infeasible and feasible samples. We first sort stocks into quintiles based on their option illiquidity at time $t-2$. Option illiquidity is obtained as volume-weighted effective spreads from intraday LiveVol data. The illiquidity of the options in a given stock is the average of the illiquidity of the individual options written on that stock. The table reports equal weighted (EW) and gross return weighted (GRW) means of the difference between the returns on the top and bottom quintile portfolios. The means weighted by lagged option gross returns (GRW) are reported to correct for microstructure biases described in Duarte, Jones, and Wang (2022). The table reports portfolio results for daily delta-hedged (DHR_t) call and put returns. DHR_t is the return on a portfolio of all ATM calls or puts for a given stock. Each option's return is computed based on average transaction prices on days $t-1$ and t. The portfolio is unhedged during regular trading hours on day $t-1$ and hedged starting at the close of day $t-1$. Average returns are in percent. Annualized Sharpe ratios are within brackets, and t-statistics are within parentheses. The sample includes the S&P 500 constituents with valid traded options data from January 2012 to December 2019.

	Calls			**Puts**		
	Infeasible	**Feasible**	**Bias**	**Infeasible**	**Feasible**	**Bias**
DHR_t (EW)	1.52	0.19	1.33	0.92	0.11	0.81
	(30.59)	(4.86)	(29.95)	(19.41)	(2.64)	(16.97)
	[10.87]	[1.73]		[6.9]	[0.94]	
DHR_t (GRW)	1.58	0.10	1.47	0.92	0.06	0.86
	(29.29)	(2.58)	(29.64)	(18.25)	(1.34)	(16.74)
	[10.41]	[0.92]		[6.49]	[0.48]	

Table 5: **Filtering extreme moneyness options.** This table presents the changes in the 10^{th} (Panel A) and 90^{th} (Panel B) percentiles of moneyness (S/K) within each portfolio between the time of portfolio formation ($t-1$) and the end of the holding period (t). The Difference column is the difference between Infeasble and Feasible. DiD is the high minus the low return portfolio differences, with t-statistics in parentheses.

Panel A: Changes in the 10^{th} percentile of moneyness

	Low Return Portfolio			High Return Portfolio			
	Infeasible	Feasible	Difference	Infeasible	Feasible	Difference	DiD
CFV	-4.09	-6.51	2.42	-4.50	-9.90	5.40	2.98 (8.90)
CH	-3.11	-6.75	3.64	-6.67	-11.95	5.28	1.64 (4.96)
DISP	-3.18	-6.03	2.85	-4.65	-10.47	5.82	2.97 (9.02)
ISSUE_1Y	-2.80	-5.83	3.03	-5.81	-11.36	5.54	2.52 (8.52)
ISSUE_5Y	-2.25	-5.26	3.02	-5.38	-10.55	5.18	2.16 (7.01)
TEF	-3.51	-6.49	2.98	-6.31	-11.78	5.48	2.50 (7.28)
-PM	-4.44	-7.59	3.15	-6.37	-12.18	5.81	2.66 (7.15)
-LN(PRICE)	-5.84	-7.42	1.58	-2.06	-10.45	8.39	6.81 (18.05)
-PROFIT	-4.30	-7.65	3.36	-6.17	-12.18	6.01	2.65 (7.49)
-ZS	-6.79	-11.60	4.81	-4.57	-9.66	5.09	0.28 (0.82)
-VOL_deviation	-4.78	-9.55	4.77	-4.81	-9.77	4.96	0.19 (0.58)
IVOL	-1.66	-3.55	1.89	-7.15	-13.16	6.01	4.12 (12.36)
AMIHUD	-4.94	-6.62	1.68	-2.63	-9.83	7.20	5.52 (17.79)

Panel B: Changes in the 90^{th} percentile of moneyness

	Low Return Portfolio			High Return Portfolio			
	Infeasible	Feasible	Difference	Infeasible	Feasible	Difference	DiD
CFV	6.26	8.15	-1.90	6.79	12.26	-5.47	-3.58 (-8.99)
CH	5.53	8.44	-2.91	9.08	13.99	-4.91	-2.00 (-4.74)
DISP	5.58	8.03	-2.45	6.71	12.67	-5.95	-3.50 (-8.42)
ISSUE_1Y	5.88	8.61	-2.72	7.74	11.79	-4.05	-1.32 (-4.21)
ISSUE_5Y	4.87	7.39	-2.52	6.98	11.36	-4.38	-1.86 (-6.52)
TEF	6.17	8.84	-2.67	8.05	12.72	-4.67	-2.00 (-4.77)
-PM	6.94	9.08	-2.14	7.92	13.90	-5.98	-3.84 (-8.02)
-LN(PRICE)	8.52	9.31	-0.80	3.61	12.03	-8.42	-7.62 (-14.36)
-PROFIT	6.92	9.92	-3.00	7.59	14.08	-6.49	-3.49 (-6.53)
-ZS	8.91	12.64	-3.73	6.47	11.33	-4.86	-1.13 (-2.89)
-VOL_deviation	6.80	10.68	-3.88	6.56	11.06	-4.49	-0.61 (-1.92)
IVOL	4.04	5.72	-1.68	9.00	14.58	-5.58	-3.91 (-8.58)
AMIHUD	7.34	8.40	-1.05	4.30	11.67	-7.37	-6.31 (-13.61)

Table 6: **Portfolios sorted on illiquidity.** This table displays the mean characteristics of portfolios sorted on illiquidity. Illiquidity is measured by the option effective spread (ES^O_{t-2}). Moneyness is defined as the closing price of the underlying stock divided by the option strike price ($(S/K)_{t-2}$). The return of the underlying stock between the closing of $t-2$ and the close of $t-1$ is $R_{S,t-1}$. DHR_t is the return on a portfolio of all ATM calls or puts for a given stock. Each option's return is computed based on average transaction prices on days $t-1$ and t. The portfolio is unhedged during regular trading hours on day $t-1$ and hedged starting at the close of day $t-1$. The table reports equal weighted (EW) and gross return weighted (GRW) means of DHR_t. The means weighted by lagged gross option returns (GRW) are reported to correct for microstructure biases described in Duarte, Jones, and Wang (2022). Panel A (B) shows results for the infeasible (feasible) sample. The infeasible and feasible filters used to create these samples are described in Section 4. Returns, moneyness and spreads are in percent. T-statistics are within parentheses.

Panel A: Infeasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.06	1.80	2.59	3.92	8.49	7.44	0.78	1.41	2.03	3.09	6.95	6.17
	(106.74)	(99.36)	(95.85)	(96.63)	(95.19)	(90.94)	(81.93)	(84.53)	(81.48)	(80.46)	(78.19)	(75.52)
$(S/K)_{t-2}$	99.46	99.26	99.11	98.93	98.57	-0.90	100.00	100.18	100.36	100.54	100.90	0.90
	(4931.30)	(4635.68)	(4172.48)	(3875.96)	(3617.87)	(-54.44)	(4791.86)	(4918.15)	(4998.99)	(4680.37)	(4602.80)	(52.96)
$R_{S,t-1}$	-0.17	0.00	0.09	0.16	0.35	0.52	0.22	0.08	-0.01	-0.08	-0.22	-0.44
	(-8.99)	(-0.23)	(4.64)	(8.44)	(16.88)	(42.80)	(10.69)	(3.99)	(-0.56)	(-3.82)	(-11.31)	(-38.42)
DHR_t (EW)	-0.49	-0.02	0.27	0.46	1.03	1.52	-0.20	0.02	0.19	0.32	0.73	0.92
	(-6.35)	(-0.19)	(3.20)	(5.60)	(12.16)	(30.59)	(-2.76)	(0.32)	(2.53)	(4.07)	(8.99)	(19.41)
DHR_t (GRW)	-0.48	0.00	0.24	0.46	1.10	1.58	-0.25	-0.04	0.12	0.25	0.67	0.92
	(-6.25)	(-0.00)	(3.16)	(5.04)	(12.47)	(29.29)	(-3.50)	(-0.48)	(1.67)	(3.25)	(8.06)	(18.25)

Panel B: Feasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.08	1.84	2.67	4.01	8.59	7.51	0.81	1.47	2.13	3.25	7.22	6.40
	(105.24)	(97.58)	(95.57)	(96.29)	(96.95)	(92.79)	(81.70)	(84.19)	(80.57)	(79.85)	(77.97)	(75.45)
$(S/K)_{t-2}$	99.32	99.25	99.17	99.07	98.86	-0.45	100.06	100.13	100.24	100.36	100.59	0.53
	(6626.82)	(6075.52)	(5278.44)	(4958.47)	(4427.69)	(-35.19)	(6658.39)	(6540.20)	(6977.84)	(7010.22)	(7080.13)	(42.58)
$R_{S,t-1}$	0.05	0.05	0.06	0.06	0.07	0.02	0.06	0.06	0.06	0.06	0.05	-0.01
	(2.52)	(2.35)	(2.86)	(3.03)	(3.58)	(2.29)	(3.16)	(3.19)	(3.18)	(3.18)	(2.56)	(-1.64)
DHR_t (EW)	-0.06	-0.10	-0.04	0.04	0.13	0.19	-0.06	-0.03	-0.10	-0.08	0.05	0.11
	(-0.84)	(-1.35)	(-0.52)	(0.49)	(1.69)	(4.86)	(-0.78)	(-0.35)	(-1.27)	(-1.07)	(0.68)	(2.64)
DHR_t (GRW)	-0.04	-0.07	-0.05	0.04	0.06	0.10	-0.10	-0.06	-0.14	-0.14	-0.04	0.06
	(-0.54)	(-0.87)	(-0.68)	(0.50)	(0.87)	(2.58)	(-1.39)	(-0.86)	(-1.87)	(-1.90)	(-0.56)	(1.34)

Internet Appendix to

Too Good to Be True: Look-ahead Bias in Empirical Options Research

A Details about stock characteristics

Here we provide construction details for the stock characteristics ZHCT use as option return predictors. The first ten variables are the main predictors of interest.

1. CFV: Cash flow variance, as in Haugen and Baker (1996), computed as the variance of the monthly ratio of annual cash flow to market value of equity over the last 60 months (requiring nonmissing observations in at least 36 months). Annual cash flow is Net Income (Compustat annual item NI) plus Depreciation and Amortization (item DP), all scaled by monthly updated market value of equity. We assume annual accounting items known publicly 4 months after the fiscal year end. To avoid stale information, we do not use annual accounting information from the fiscal year end that is older than 15 months.

2. CH: Following Palazzo (2012), we measure cash-to-assets, as cash holdings (Compustat quarterly item CHEQ) scaled by total assets (item ATQ). We assume quarterly accounting items known publicly 4 months after the fiscal quarter end. To avoid stale information, we do not use quarterly accounting information from the fiscal quarter end that is older than 6 months.

3. DISP: Analyst earnings forecast dispersion, as in Diether, Malloy, and Scherbina (2002). We measure dispersion in analyst earnings forecasts as the ratio of the standard deviation of earnings forecasts (IBES unadjusted file, item STDEV) to the absolute value of the consensus mean forecast (unadjusted file, item MEANEST). We use the earnings forecasts for the current fiscal year (fiscal period indicator = 1) and we require them to be

denominated in US dollars (currency code = USD). Stocks with a mean forecast of zero are assigned to the highest dispersion group. Firms with fewer than 2 forecasts are excluded.

4. ISSUE1Y: One-year new issues, as in Pontiff and Woodgate (2008), measured as the natural log of the ratio of the split-adjusted shares outstanding at one fiscal year end to the split-adjusted shares outstanding at the fiscal year end 12 months ago. The split-adjusted shares outstanding are shares outstanding (Compustat annual item CSHO) times the adjustment factor (item AJEX). We assume annual accounting items known publicly 4 months after the fiscal year end. To avoid stale information, we do not use annual accounting information from the fiscal year end that is older than 15 months.

5. ISSUE5Y: Five-year new issues, as in Daniel and Titman (2006), for each month t, measured as the log growth rate in the market equity not attributable to the stock return, $log(Me_t/Me_{t-60}) - r(t-60,t)$. $r(t-60,t)$ is the cumulative log stock return over the past 60 months including month t (i.e., the return spanning from the last trading day of month $t-60$ to the last trading day of month t), and Me_t is the market equity (from CRSP) on the last trading day of month t.

6. PM: Profit margin, as in Soliman (2008), calculated as Earnings Before Interest and Taxes (Compustat annual item EBIT) scaled by Revenue (item REVT). We assume annual accounting items known publicly 4 months after the fiscal year end. To avoid stale information, we do not use annual accounting information from the fiscal year end that is older than 15 months.

7. ln(PRICE): The log of stock price at the end of last month, as in Blume and Husic (1973).

8. PROFIT: Profitability, as in Fama and French (2006), calculated as earnings divided by book equity (the denominator is current, not lagged, book equity), in which earnings is defined as Income Before Extraordinary Items (Compustat annual item IB). Book equity is

stockholders' book equity, plus balance sheet deferred taxes and investment tax credit (item TXDITC) if available, minus the book value of preferred stock. Stockholders' equity is the value reported by Compustat (item SEQ), if it is available. If not, we measure stockholders' equity as the book value of common equity (item CEQ) plus the par value of preferred stock (item PSTK), or the book value of assets (item AT) minus total liabilities (item LT). Depending on availability, we use redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for the book value of preferred stock. We assume annual accounting items known publicly 4 months after the fiscal year end. To avoid stale information, we do not use annual accounting information from the fiscal year end that is older than 15 months.

9. TEF: Total external financing, as in Bradshaw, Richardson, and Sloan (2006), for one fiscal year end, scaled by the average of total assets (Compustat annual item AT) at the same fiscal year end and the prior fiscal year end. Total external financing is the sum of net equity financing and net debt financing. Net equity financing is the proceeds from the sale of common and preferred stocks (Compustat annual item SSTK) less cash payments for the repurchases of common and preferred stocks (item PRSTKC) less cash payments for dividends (item DV). Net debt financing is the cash proceeds from the issuance of long-term debt (item DLTIS) less cash payments for long-term debt reductions (item DLTR) plus the net changes in current debt (item DLCCH, zero if missing). We assume annual accounting items known publicly 4 months after the fiscal year end. To avoid stale information, we do not use annual accounting information from the fiscal year end that is older than 15 months.

10. ZS: Z-score. We follow Altman (1968) to construct the Z-score (Dichev (1998)) as $Z \equiv 1.2WCTA + 1.4RETA + 3.3EBITTA + 0.6METL + SALETA$, in which WCTA is working capital (Compustat annual item ACT minus item LCT) divided by total assets (item AT), RETA is retained earnings (item RE) divided by total assets, EBITTA is earnings before interest and taxes (item OIADP) divided by total assets, METL is the market equity

(from CRSP, at fiscal year end) divided by total liabilities (item LT), and SALETA is sales (item SALE) divided by total assets. For firms with more than 1 share class, we merge the market equity for all share classes before computing Z. We winsorize all non-dummy variables on the right-hand side of equation at the 1th and 99th percentiles of their distributions each year. At the end of June of each year t, we split stocks into deciles based on Z-score for the fiscal year ending in calendar year $t-1$. Monthly decile returns are calculated from July of year t to June of t+1, and the deciles are rebalanced in June of t+1.

The last three variables are controls.

11. VOL_deviation: Volatility deviation measure as in Goyal and Saretto (2009), calculated as the log difference between the rolling annual realized volatility and Black-Scholes implied volatility for at-the-money options at the end of the last month.

12. IVOL: Idiosyncratic volatility, as in Ang, Hodrick, Xing, and Zhang (2006) and Cao and Han (2013), computed as the standard deviation of the residuals of a regression of individual stock returns on the three Fama and French (1993) factors using daily data in the previous month.

13. AMIHUD: The Amihud (2002) stock illiquidity measure (calculated as the average of the daily ratio of the absolute stock return to dollar volume over the previous month).

B Robustness

Table A1 replicates the results of ZHCT using three different weighting schemes to calculate the average return of a delta-neutral call writing portfolio. In addition to equal weighting (EW), these include weighting by the market capitalization of the underlying stock (Stock-VW) and the market value of option open interest at the outset of the holding period (Option-VW). The table also presents results utilizing "true arbitrage" filters, which are based on bid or ask prices, as appropriate, instead of arbitrage filters based on mid-quotes.

[Insert Table A1 Here]

Table A2 displays replication results of ZHCT excluding options where the underlying stock pays dividends during the holding period.

[Insert Table A2 Here]

We originally obtained data from CBOE LiveVol to replicate CGJK in 2020, when CBOE was offering data starting from 2004. However, we observed some large discrepancies between the LiveVol data we procured and the data as reported by CGJK. Despite reaching out to both CGJK and CBOE LiveVol to address these inconsistencies, we were unable to resolve them. Although our 2004-2012 sample differs from that in CGJK, these variations do not influence our conclusions. We present our results for our 2004-2012 sample below.

Table A3 shows the summary statistics for the sample period running from January 2004 to December 2012. This table shows that the number of underlying stocks we have in our sample is smaller than the one in CGJK. Moreover, this table also shows that the mean and the standard deviation of the the delta-hedged call return are very large at 2.3% and 111.29% per day, respectively.

[Insert Table A3 Here]

Table A4 shows replication results of the CGJK portfolio sorts for the sample period running from January 2004 to December 2012. Mirroring the findings in Table 6 in the main paper, Table A4 confirms that look-ahead biases account for the price of illiquidity in options documented in CGJK.

[Insert Table A4 Here]

Contrary to CGJK and findings in Table 6, Panel B of Table A4 suggests a potential negative pricing of illiquidity in call options. Notably, the GRW mean return of the "5-1" portfolio in Panel B stands at -3.47% per day. Furthermore, the table shows that the high return on delta-hedged calls is heavily driven by the options in quintile 3, which has an average EW delta-hedged return of 16.19% per day.

The explanation for the large average returns, and possibly also for the finding that illiquidity might be negatively priced in call options between 2004-2012, appears to be driven by extreme outliers that remain in the feasible sample but that are eliminated by the filters used in the infeasible sample. We therefore introduce additional filters that exclude certain outliers from the LiveVol database. Specifically, our sample retains options that adhere to three criteria: (1) The mean underlying midpoint from LiveVol is less than 10% outside the daily high/low range from CRSP. (2) The mean option trade price from LiveVol is between 1/10 and ten times the IvyDB closing midpoint. (3) The mean LiveVol option trade price is within 50% of the mean LiveVol midpoint. These criteria are applied at t-2, t-1, and t, making them infeasible.

[Insert Table A5 Here]

Summary statistics using the new filters are shown in Table A5. They confirm that outliers drive the large mean return of delta-hedged call options in Table A3. Portfolio sorts using the new filters, shown in Table A6, no longer show a large negative price of illiquidity, and the anomalously large return on the third quintile portfolio is gone.

[Insert Table A6 Here]

The findings in Table 6 in the main text rely on a feasible sample without the outlier filters used in Table A6. For completeness, however, Table A7 replicates that analysis using those same outlier filters. The outcomes in Table A7 are fundamentally consistent with those in Table 6, signifying that outliers in LiveVol don't influence the 2012-2019 sample as significantly as in the 2004-2012 sample. Nevertheless, they do affect the results somewhat, as we find that there is no longer any significant difference between high and low liquidity options when portfolios use gross return weighting.

[Insert Table A7 Here]

C Impact of Different Filters

C.1 Option Return Predictability

To understand the sources of look-ahead bias in ZHCT, we separately analyze the effects of each filter in Table A8. For brevity, we only examine equally weighted portfolios.

ZHCT impose many filters on their sample, but we focus here on those based on data at the end of the holding period (t). As discussed above, these filters require that: (1) Option bid prices are positive. (2) The bid-ask midpoint is at least \$1/8. That is $C_{t-1} \geq \$1/8$, $P_{t-1} \geq \$1/8$, $C_t \geq \$1/8$ and $P_t \geq \$1/8$. (3) The midpoint does not violate no-arbitrage conditions. Options that violate such no-arbitrage conditions have missing implied volatility in the Optionmetrics database. (4) The underlying stock does not pay a dividend during the remaining life of the option. This mitigates early exercise effects, as individual stock options in the U.S. are of the American type.

[Insert Table A8 Here]

Table A8 begins by repeating the infeasible and feasible spread portfolios described in Table A8 The additional columns in the table implement portfolios that are feasible except for one filter, specified on the column name, which is imposed at the end of the holding period.

The differences between these columns and the feasible portfolios indicate the importance of each source of look-ahead bias.

The table shows, for example, that conditioning on firms that do not announce and pay dividends during the holding period has little effect on the results. Similarly, the no-arbitrage filter has an effect that in some instances is small.

In contrast, the filters that limit the minimum price of the option by requiring either a midpoint of at least 1/8 or a nonzero bid price turns out to be critical, as they are responsible for most of the difference between feasible and infeasible results. The results in Table A8 support our explaination of the source of the look-ahead biases in ZHCT as being the result of removing from the sample options that become either deep-OTM or deep-ITM at the end of the holding period. Deep-OTM options are more likely to have bid prices equal to zero or prices below 1/8. Since ZHCT require these conditions for each matching pair of put and call, these filters also affect deep-ITM call options if the corresponding deep-OTM put option does not satisfy them.

C.2 Illiquidity Premium

To understand the sources of look-ahead bias in CGJK, we separately analyze the effects of each filter in Table A9. For brevity, we again focus on ATM options (options with absolute value of delta between 0.375 and 0.625), and we only examine equally weighted portfolios.

CGJK impose many filters on their sample, but we focus here on those based on data at the end of the first day of the holding period, which would not be available to investors trading in the middle of the day. As discussed above, these filters include upper bounds on spreads, lower bounds on prices, lower and upper bounds on deltas, and positive open interest. They also include a requirement of positive volume, which is implicit, as CGJK rely on transaction prices for computing returns.

[Insert Table A9 Here]

Table A9 begins by repeating the infeasible and feasible spread portfolios constructed by buying options with high effective spreads and writing options with low effective spreads. The additional columns in the table implement portfolios that are feasible except for one filter, specified on the column name, which is imposed at the end of the first day of the holding period (day $t-1$). The differences between these columns and the feasible portfolios indicates the importance of each source of look-ahead bias.

The table shows, for example, that conditioning on closing bid-ask spread has little effect on the results, which is likely the result of ATM options rarely trading with spreads near 50%. Similarly, the end-of-day price filter also has little effect, as a $0.10 lower bound is rarely binding for ATM options. The requirement that option volume be positive on day t is also unimportant in this sample.

In contrast, the seemingly benign delta filter turns out to be critical, as it is responsible for most of the difference between feasible and infeasible results. Options pass this filter when their delta is between 0.375 and 0.625 in absolute value at the end of day $t-1$. The results in Table A9 support our explanation of the source of the look-ahead biases in CGJK as the result of selecting options that are ATM at time $t-1$.

In the interest of brevity, we do not report results for OTM options but note that they are more significantly affected by the filters on spreads and positive volume. The spread filter is naturally more important for OTM options, which often do trade with spreads above the maximum value of 50%. The requirement of positive trading volume also biases OTM returns upward. We find that calls tend to trade more when stock prices go up, while puts tend to trade more when stock prices decline. As a result, a sample of options, either calls or puts, that conditions on future volume being positive has an upwardly biased average return. These results emphasize that a look-ahead bias that is minimal in one context may be more problematic in another.

Table A1: **Weighted-Average return of long-short portfolios of delta-neutral call writing sorted by underlying stock characteristics.** This table shows the results of the replication of Zhan, Han, Cao, and Tong (2021) with infeasible and feasible samples. We use three different weighting schemes to compute the average return of a portfolio of delta-neutral call writing: weight by market capitalization of the underlying stock (Stock-VW), equal weight (EW), and weight by the market value of option open interest at the beginning of the holding period (Option-VW). This table displays the difference between the returns of the portfolio with High and Low returns. The consequence of this procedure is that, for some sorting variables, the displayed High minus Low mean returns are the difference in the mean returns of the portfolio in bottom decile and of the portfolio in the top decile of the stock characteristic. When this is the case, the first column shows a negative sign on the label of the sorting characteristic. The column "True Arb." shows the results of feasible strategies using a sample filter that removes from the sample options with prices that do not satisfy arbitrage conditions based on bid and ask option prices. In contrast, the columns with the feasible and infeasible results are based a sample filter that removes from the sample options with prices that do not satisfy arbitrage conditions based on the mid point of the bid and ask option prices. Average returns are in percent. Annualized Sharpe ratios are within brackets and t-statistics are within parentheses.

	EW			**Stock-VW**			**Option-VW**		
	Infeasible	**Feasible**	**True arb.**	**Infeasible**	**Feasible**	**True arb.**	**Infeasible**	**Feasible**	**True arb.**
CFV	2.12	-0.13	-0.11	1.52	-0.13	-0.10	2.53	0.34	0.35
	(16.99)	(-0.94)	(-0.76)	(9.90)	(-0.80)	(-0.60)	(10.11)	(1.14)	(1.17)
	[3.78]	[-0.21]	[-0.17]	[2.20]	[-0.18]	[-0.13]	[2.25]	[0.25]	[0.26]
CH	2.10	0.60	0.63	0.40	-0.05	-0.09	1.62	0.31	0.29
	(13.90)	(2.70)	(2.86)	(1.76)	(-0.20)	(-0.35)	(5.59)	(0.80)	(0.75)
	[3.09]	[0.60]	[0.63]	[0.39]	[-0.04]	[-0.08]	[1.24]	[0.18]	[0.17]
DISP	1.91	-0.11	-0.08	1.28	-0.21	-0.17	2.20	0.21	0.19
	(14.40)	(-0.69)	(-0.51)	(7.10)	(-1.17)	(-0.97)	(8.70)	(0.84)	(0.75)
	[3.20]	[-0.15]	[-0.11]	[1.5]	[-0.26]	[-0.22]	[1.93]	[0.19]	[0.17]
ISSUE_1Y	1.62	0.20	0.24	0.44	-0.25	-0.17	1.25	-0.11	-0.01
	(13.68)	(1.59)	(1.87)	(2.05)	(-1.24)	(-0.90)	(4.25)	(-0.46)	(-0.05)
	[3.04]	[0.35]	[0.42]	[0.46]	[-0.28]	[-0.20]	[0.95]	[-0.10]	[-0.01]
ISSUE_5Y	1.82	0.31	0.37	0.23	-0.31	-0.19	1.39	0.08	0.19
	(14.16)	(2.14)	(2.63)	(0.85)	(-1.52)	(-1.01)	(4.90)	(0.29)	(0.72)
	[3.15]	[0.48]	[0.59]	[0.19]	[-0.34]	[-0.22]	[1.09]	[0.07]	[0.16]
TEF	1.59	0.01	0.06	1.06	-0.11	-0.08	1.87	0.07	0.12
	(11.07)	(0.08)	(0.39)	(5.21)	(-0.51)	(-0.34	(6.74)	(0.25)	(0.42)
	[2.46]	[0.02]	[0.09]	[1.16]	[-0.11]	[-0.08]	[1.50]	[0.06]	[0.09]
-PM	2.30	0.12	0.10	1.73	-0.27	-0.25	2.81	-0.20	-0.21
	(16.47)	(0.66)	(0.55)	(6.10)	(-1.11)	(-1.09)	(10.84)	(-0.63)	(-0.66)
	[3.66]	[0.15]	[0.12]	[1.36]	[-0.25]	[-0.24]	[2.41]	[-0.14]	[-0.15]
-ln(PRICE)	4.85	0.61	0.64	4.43	0.23	0.24	5.62	0.84	0.84
	(26.48)	(2.74)	(3.01)	(19.09)	(1.11)	(1.24)	(18.88)	(2.70)	(2.82)
	[5.88]	[0.61]	[0.67]	[4.24]	[0.25]	[0.28]	[4.19]	[0.60]	[0.63]
-PROFIT	2.27	0.10	0.10	1.55	-0.16	-0.11	2.50	0.05	0.08
	(19.18)	(0.54)	(0.56	(6.20)	(-0.84)	(-0.58)	(9.03)	(0.16)	(0.28)
	[4.26]	[0.12]	[0.12]	[1.38]	[-0.19]	[-0.13]	[2.01]	[0.04]	[0.06]
-ZS	0.44	0.10	0.09	0.48	-0.08	-0.06	1.41	0.45	0.48
	(2.63)	(0.58)	(0.54)	(1.92)	(-0.34)	(-0.26)	(4.42)	(1.61)	(1.76)
	[0.58]	[0.13]	[0.12]	[0.43]	[-0.07]	[-0.06]	[0.98]	[0.36]	[0.39]
-VOL_deviation	3.00	2.81	2.79	1.78	1.89	1.75	2.88	2.23	2.09
	(13.59)	(11.53)	(11.27)	(6.94)	(8.02)	(7.96)	(9.12)	(7.28)	(7.17)
	[3.02]	[2.56]	[2.51]	[1.54]	[1.78]	[1.77]	[2.03]	[1.62]	[1.59]
IVOL	3.78	0.77	0.80	2.83	0.45	0.50	3.77	0.58	0.65
	(25.18)	(3.75)	(3.76)	(9.85)	(1.68)	(1.99)	(12.66)	(1.62)	(1.85)
	[5.60]	[0.83]	[0.84]	[2.19]	[0.37]	[0.44]	[2.81]	[0.36]	[0.41]
Amihud	3.78	0.22	0.23	3.41	0.0003	-0.05	4.52	1.01	1.00
	(24.40)	(1.21)	(1.21)	(22.35)	(0.002)	(-0.29)	(16.91)	(3.24)	(3.23)
	[5.42]	[0.27]	[0.27]	[4.97]	[0.0004]	[-0.07]	[3.76]	[0.72]	[0.72]

Table A2: **Average return of long-short portfolios of delta-neutral call writing sorted by underlying stock characteristics with dividend infeasible strategies.** This table shows the results of the replication of Zhan, Han, Cao, and Tong (2021) with infeasible and feasible samples. We use three different weighting schemes to compute the average return of a portfolio of delta-neutral call writing: weight by market capitalization of the underlying stock (Stock-VW), equal weight (EW), and weight by the market value of option open interest at the beginning of the holding period (Option-VW). This table displays the difference between the returns of the decile portfolios with High and Low returns. The consequence of this procedure is that, for some sorting variables, the displayed High minus Low mean returns are the difference in the mean returns of the portfolio in bottom decile and of the portfolio in the top decile of the stock characteristic. When this is the case, the first column shows a negative sign on the label of the sorting characteristic. The results are based on a sample without stocks that pay a dividend during the holding period. Average returns are in percent. Annualized Sharpe ratios are within brackets and t-statistics are within parentheses.

	EW	**Stock-VW**	**Option-VW**
CFV	-0.0013	-0.0013	0.0036
	(-0.87)	(-0.72)	(1.12)
	[-0.19]	[-0.16]	[0.25]
CH	0.0058	-0.0007	0.0035
	(2.69)	(-0.28)	(0.91)
	[0.60]	[-0.06]	[0.20]
DISP	-0.0010	-0.0022	0.0027
	(-0.63)	(-1.21)	(1.06)
	[-0.14]	[-0.27]	[0.23]
ISSUE_1Y	0.0019	-0.0029	-0.0010
	(1.47)	(-1.42)	(-0.39)
	[0.33]	[-0.32]	[-0.09]
ISSUE_5Y	0.0022	-0.0034	0.0002
	(1.54)	(-1.60)	(0.07)
	[0.34]	[-0.36]	[0.02]
TEF	-0.0002	-0.0022	0.0005
	(-0.13)	(-0.96)	(0.17)
	[-0.03]	[-0.21]	[0.04]
-PM	0.0009	-0.0037	-0.0026
	(0.49)	(-1.46)	(-0.79)
	[0.11]	[-0.32]	[-0.18]
-ln(PRICE)	0.0065	0.0031	0.0093
	(2.89)	(1.47)	(2.90)
	[0.64]	[0.33]	[0.65]
-PROFIT	0.0010	-0.0017	0.0001
	(0.55)	(-0.79)	(0.03)
	[0.12]	[-0.18]	[0.01]
-ZS	0.0014	-0.0009	0.0046
	(0.84)	(-0.38)	(1.57)
	[0.19]	[-0.08]	[0.35]
-VOL_deviation	0.0286	0.0183	0.0221
	(11.71)	(7.76)	(6.98)
	[2.60]	[1.73]	[1.55]
IVOL	0.0077	0.0040	0.0058
	(3.63)	(1.39)	(1.54)
	[0.81]	[0.31]	[0.34]
AMIHUD	0.0024	-0.0001	0.0097
	(1.34)	(-0.04)	(3.03)
	[0.30]	[-0.01]	[0.67]

Table A3: **Summary statistics as in Christoffersen, Goyenko, Jacobs, and Karoui (2018) (2004-2012).** This table presents summary statistics on the samples used to estimate the illiquidity premium in option returns. The end of the holding period is at the closing of t. Panel A presents summary statistics on delta-hedged returns (DHR_t) of calls and puts as well as on the excess returns of the underlying stocks $R_{S,t}$. DHR_t is the mean across all options on the same underlying stocks of the return of an unhedged option between the opening and the closing of $t-1$ with a hedge put in place from the closing of $t-1$ and t. As in CGJK, we compute the descriptive statistics for each stock and then we take the averages of these statistics across stocks. We report the mean, standard deviation (Std. Dev.), skewness, kurtosis, first-order autocorrelation of delta-hedged returns $\rho(1)$, and first-order autocorrelation of the absolute value of delta-hedged returns, $|\rho(1)|$. Panel B presents statistics on the option effective spreads at the closing of $t-2$ (ES^O_{t-2}). For each stock and on each day, we compute the ES^O_{t-2} with all the available options in the sample and then we compute across time the mean, the Std Dev, the minimum (min), the maximum (max), and the first-order autocorrelation, $\rho(1)$. Panel B displays the averages of these statistics across stocks. Panels C and D present the means of moneyness $(S/K)_{t-2}$, β_{t-2}, DHR_t, and ES^O_{t-2} on days in which the returns of the underlying stock in the middle of the holding period which is on the closing of $t-1$ ($R_{S,t-1}$) are positive or negative. The moneyness of an option is defined as the closing price of the underlying stock (S) divided by the option's strike price (K). The option β is its δ divided by its price. The sample that is built with the infeasible (feasible) filters are described in Section 4. The sample contains only ATM options which are defined in the infeasible (feasible) sample as options with the absolute value of δ_{t-1} (δ_{t-2}) between 0.375 and 0.625. Returns and effective spreads are in percent. The sample includes the S&P 500 constituents with valid traded options data from January 2004 to December 2012.

Panel A: Returns

	Call		**Put**		**Stock**
	Infeasible	**Feasible**	**Infeasible**	**Feasible**	**Return**
Mean	0.51	2.30	0.67	0.08	0.03
Std. Dev.	14.51	111.29	13.18	15.84	2.40
Skewness	0.93	1.93	0.83	1.20	0.43
Kurtosis	12.91	67.31	7.00	29.91	16.35
$\rho(1)$	-0.32	-0.38	-0.29	-0.37	-0.04
$\lvert\rho(1)\rvert$	0.23	0.26	0.19	0.25	0.21
Average # of Stocks	312	385	232	333	455
Average # of Series	2.83	2.92	2.38	2.49	

Panel B: Effective Spread

	Call		**Put**	
	Infeasible	**Feasible**	**Infeasible**	**Feasible**
Mean	6.00	5.99	4.69	4.82
Std. Dev.	3.83	3.80	3.27	3.28
min	0.19	0.20	0.11	0.13
max	38.29	41.14	30.99	31.44
$\rho(1)$	0.33	0.36	0.29	0.32

Panel C: Averages for Calls when $R_{S,t-1} > 0$ and when $R_{S,t-1} < 0$

	Infeasible				Feasible			
	$(S/K)_{t-2}$ (1)	β_{t-2} (2)	DHR_t (3)	ES^O_{t-2} (4)	$(S/K)_{t-2}$ (5)	β_{t-2} (6)	DHR_t (7)	ES^O_{t-2} (8)
$R_{S,t-1} > 0$	96.68	11.00	4.48	6.50	98.09	10.07	4.86	5.98
$R_{S,t-1} < 0$	99.52	10.09	-3.85	5.44	98.27	10.07	-0.29	5.99

Panel D: Averages for Puts when $R_{S,t-1} > 0$ and when $R_{S,t-1} < 0$

	Infeasible				Feasible			
	$(S/K)_{t-2}$ (1)	β_{t-2} (2)	DHR_t (3)	ES^O_{t-2} (4)	$(S/K)_{t-2}$ (5)	β_{t-2} (6)	DHR_t (7)	ES^O_{t-2} (8)
$R_{S,t-1} > 0$	98.05	-10.30	-3.64	4.27	99.04	-10.09	-3.84	4.79
$R_{S,t-1} < 0$	101.19	-10.74	4.40	5.03	99.23	-10.20	4.13	4.84

Table A4: **Portfolios sorted on illiquidity in the sample period from 2004 to 2012.** This table displays the mean characteristics of portfolios sorted on illiquidity. Illiquidity is measured with the option effective spread ES^O_{t-2}. Moneyness is defined as the closing price of the underlying stock divided by the option strike price $(S/K)_{t-2}$. The return of the underlying stock between the closing of $t-2$ and the close of $t-1$ is $R_{S,t-1}$. DHR_t is the mean across all options on the same underlying stocks of the return of an unhedged option between the opening and the closing of $t-1$ with a hedge put in place from the closing of $t-1$ and t. The table reports the equal weighted (EW) and the gross return weighted (GRW) mean DHR_t. The means weighted by lagged option gross returns (GRW) are reported to correct for microstructure biases described in Duarte, Jones, and Wang (2022). Panel A (B) shows results for the infeasible (feasible) sample. The infeasible and feasible filters used to create these samples are described in Section 4. Returns, moneyness and spreads are in percent. The sample period is from January 2004 to December 2012. T-statistics are within parentheses.

Panel A: Infeasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.73	2.62	3.50	4.96	10.03	8.30	1.28	2.03	2.66	3.65	7.23	5.94
	(47.66)	(53.28)	(59.57)	(72.56)	(104.55)	(122.64)	(43.51)	(49.80)	(53.90)	(62.07)	(88.44)	(102.06)
$(S/K)_{t-2}$	99.02	98.65	98.32	97.93	97.04	-1.98	98.79	99.20	99.58	99.95	100.58	1.79
	(1667.09)	(1552.29)	(1338.01)	(1128.32)	(789.59)	(-25.62)	(1367.04)	(1435.36)	(1610.52)	(1692.78)	(1839.94)	(45.43)
$R_{S,t-1}$	-0.32	-0.08	0.07	0.20	0.45	0.77	0.31	0.10	-0.03	-0.14	-0.36	-0.67
	(-10.69)	(-2.89)	(2.52)	(6.74)	(13.21)	(29.48)	(10.29)	(3.29)	(-0.94)	(-4.56)	(-11.59)	(-36.49)
DHR_t (EW)	-0.65	-0.18	0.17	0.51	1.28	1.93	-0.37	-0.11	0.15	0.39	0.83	1.20
	(-8.17)	(-2.19)	(2.03)	(5.76)	(12.45)	(31.54)	(-5.86)	(-1.55)	(2.02)	(4.92)	(9.18)	(19.76)
DHR_t (GRW)	-0.65	-0.20	0.15	0.51	1.30	1.95	-0.42	-0.17	0.08	0.29	0.73	1.15
	(-8.27)	(-2.41)	(1.83)	(5.75)	(12.68)	(30.38)	(-6.57)	(-2.48)	(1.08)	(3.77)	(8.13)	(18.62)

Panel B: Feasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.76	2.66	3.56	5.06	10.22	8.46	1.31	2.06	2.70	3.75	7.52	6.21
	(48.16)	(53.99)	(61.43)	(75.57)	(107.60)	(126.75)	(44.62)	(51.66)	(56.97)	(67.68)	(96.72)	(112.79)
$(S/K)_{t-2}$	98.70	98.53	98.37	98.07	97.40	-1.30	98.85	99.09	99.37	99.65	100.09	1.24
	(1620.35)	(1453.50)	(1311.12)	(1118.25)	(832.64)	(-20.24)	(1308.76)	(1372.17)	(1454.58)	(1462.68)	(1542.59)	(44.93)
$R_{S,t-1}$	0.03	0.03	0.03	0.04	0.05	0.03	0.04	0.03	0.04	0.05	0.04	0.00
	(0.91)	(0.99)	(1.04)	(1.21)	(1.68)	(2.09)	(1.32)	(1.03)	(1.46)	(1.52)	(1.21)	(-0.40)
DHR_t (EW)	3.74	0.71	16.19	0.75	0.41	-3.33	0.16	0.14	-0.08	0.03	0.02	-0.14
	(1.90)	(0.93)	(1.00)	(1.11)	(4.21)	(-1.69)	(0.68)	(0.54)	(-1.04)	(0.27)	(0.29)	(-0.62)
DHR_t (GRW)	3.76	0.80	17.17	0.71	0.29	-3.47	0.15	0.08	-0.14	-0.05	-0.12	-0.27
	(1.90)	(0.95)	(1.00)	(1.04)	(3.00)	(-1.76)	(0.56)	(0.32)	(-2.03)	(-0.49)	(-1.54)	(-1.09)

Table A5: **Summary statistics as in Christoffersen, Goyenko, Jacobs, and Karoui (2018) (2004-2012) with data matching between OptionMetrics and LiveVol.** This table presents summary statistics on the samples used to estimate the illiquidity premium in option returns. The end of the holding period is at the closing of t. Panel A presents summary statistics on delta-hedged returns (DHR_t) of calls and puts as well as on the excess returns of the underlying stocks $R_{S,t}$. DHR_t is the mean across all options on the same underlying stocks of the return of an unhedged option between the opening and the closing of $t-1$ with a hedge put in place from the closing of $t-1$ and t. As in CGJK, we compute the descriptive statistics for each stock and then we take the averages of these statistics across stocks. We report the mean, standard deviation (Std. Dev.), skewness, kurtosis, first-order autocorrelation of delta-hedged returns $\rho(1)$, and first-order autocorrelation of the absolute value of delta-hedged returns, $|\rho(1)|$. Panel B presents statistics on the option effective spreads at the closing of $t-2$ (ES^O_{t-2}). For each stock and on each day, we compute the ES^O_{t-2} with all the available options in the sample and then we compute across time the mean, the Std Dev, the minimum (min), the maximum (max), and the first-order autocorrelation, $\rho(1)$. Panel B displays the averages of these statistics across stocks. Panels C and D present the means of moneyness $(S/K)_{t-2}$, β_{t-2}, DHR_t, and ES^O_{t-2} on days in which the returns of the underlying stock in the middle of the holding period which is on the closing of $t-1$ ($R_{S,t-1}$) are positive or negative. The moneyness of an option is defined as the closing price of the underlying stock (S) divided by the option's strike price (K). The option β is its δ divided by its price. The sample that is built with the infeasible (feasible) filters are described in Section 4. The sample contains only ATM options which are defined in the infeasible (feasible) sample as options with the absolute value of δ_{t-1} (δ_{t-2}) between 0.375 and 0.625. Returns and effective spreads are in percent. The sample includes the S&P 500 constituents with valid traded options data from January 2004 to December 2012.

Panel A: Returns

	Call		**Put**		**Stock**
	Infeasible	**Feasible**	**Infeasible**	**Feasible**	**Return**
Mean	0.50	-0.04	0.67	-0.02	0.04
Std. Dev.	14.39	14.15	13.12	12.79	2.46
Skewness	0.73	0.41	0.73	0.57	0.41
Kurtosis	7.65	7.40	4.84	6.04	15.31
$\rho(1)$	-0.32	-0.34	-0.29	-0.30	-0.03
$\|\rho(1)\|$	0.23	0.25	0.20	0.21	0.21
Average # of Stocks	312	343	232	271	417
Average # of Series	2.83	2.92	2.38	2.49	

Panel B: Effective Spread

	Call		**Put**	
	Infeasible	**Feasible**	**Infeasible**	**Feasible**
Mean	6.00	5.88	4.68	4.67
Std. Dev.	3.83	3.61	3.27	3.12
min	0.19	0.24	0.11	0.16
max	38.29	35.82	31.00	27.66
$\rho(1)$	0.33	0.35	0.29	0.31

Panel C: Averages for Calls when $R_{S,t-1} > 0$ and when $R_{S,t-1} < 0$

	Infeasible				Feasible			
	$(S/K)_{t-2}$ (1)	β_{t-2} (2)	DHR_t (3)	ES^O_{t-2} (4)	$(S/K)_{t-2}$ (5)	β_{t-2} (6)	DHR_t (7)	ES^O_{t-2} (8)
$R_{S,t-1} > 0$	96.68	11.00	4.47	6.50	97.99	10.32	3.97	5.91
$R_{S,t-1} < 0$	99.52	10.09	-3.86	5.44	98.25	10.29	-4.13	5.85

Panel D: Averages for Puts when $R_{S,t-1} > 0$ and when $R_{S,t-1} < 0$

	Infeasible				Feasible			
	$(S/K)_{t-2}$ (1)	β_{t-2} (2)	DHR_t (3)	ES^O_{t-2} (4)	$(S/K)_{t-2}$ (5)	β_{t-2} (6)	DHR_t (7)	ES^O_{t-2} (8)
$R_{S,t-1} > 0$	98.04	-8.65	-3.65	4.26	99.22	-8.70	-3.84	4.64
$R_{S,t-1} < 0$	101.19	-9.29	4.40	5.03	99.50	-8.79	3.94	4.71

Table A6: **Portfolios sorted on illiquidity in the sample period from 2004 to 2012 and with additional filters.** This table displays the mean characteristics of portfolios sorted on illiquidity. Illiquidity is measured with the option effective spread ES^O_{t-2}. Moneyness is defined as the closing price of the underlying stock divided by the option strike price $(S/K)_{t-2}$. The return of the underlying stock between the closing of $t-2$ and the close of $t-1$ is $R_{S,t-1}$. DHR_t is the mean across all options on the same underlying stocks of the return of an unhedged option between the opening and the closing of $t-1$ with a hedge put in place from the closing of $t-1$ and t. The table reports the equal weighted (EW) and the gross return weighted (GRW) mean DHR_t. The means weighted by lagged option gross returns (GRW) are reported to correct for microstructure biases described in Duarte, Jones, and Wang (2022). Panel A (B) shows results for the infeasible (feasible) sample. The infeasible and feasible filters used to create these samples are described in Section 4. Returns, moneyness and spreads are in percent. The sample period is from January 2004 to December 2012. T-statistics are within parentheses.

Panel A: Infeasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.73	2.62	3.50	4.96	10.03	8.30	1.28	2.03	2.66	3.65	7.22	5.94
	(47.67)	(53.29)	(59.57)	(72.56)	(104.57)	(122.66)	(43.52)	(49.80)	(53.91)	(62.09)	(88.42)	(102.06)
$(S/K)_{t-2}$	99.02	98.65	98.32	97.93	97.04	-1.98	98.79	99.20	99.58	99.95	100.58	1.79
	(1666.28)	(1553.65)	(1338.50)	(1127.73)	(789.99)	(-25.65)	(1367.88)	(1434.26)	(1609.91)	(1692.19)	(1840.09)	(45.48)
$R_{S,t-1}$	-0.32	-0.08	0.07	0.20	0.45	0.77	0.31	0.10	-0.03	-0.14	-0.36	-0.67
	(-10.68)	(-2.88)	(2.52)	(6.73)	(13.21)	(29.50)	(10.29)	(3.29)	(-0.92)	(-4.57)	(-11.58)	(-36.47)
DHR_t (EW)	-0.66	-0.20	0.16	0.51	1.27	1.93	-0.39	-0.12	0.14	0.38	0.82	1.21
	(-8.36)	(-2.42)	(1.93)	(5.68)	(12.39)	(32.09)	(-6.02)	(-1.72)	(1.90)	(4.83)	(9.15)	(19.86)
DHR_t (GRW)	-0.65	-0.21	0.15	0.51	1.30	1.95	-0.42	-0.17	0.08	0.29	0.73	1.16
	(-8.32)	(-2.56)	(1.86)	(5.78)	(12.68)	(30.53)	(-6.66)	(-2.53)	(1.07)	(3.83)	(8.15)	(18.75)

Panel B: Feasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.73	2.61	3.45	4.86	9.69	7.96	1.29	2.02	2.63	3.59	7.01	5.71
	(47.37)	(52.98)	(59.10)	(72.38)	(104.32)	(124.69)	(43.30)	(49.67)	(53.87)	(62.00)	(90.29)	(106.66)
$(S/K)_{t-2}$	98.69	98.52	98.34	98.02	97.31	-1.38	98.99	99.22	99.50	99.78	100.19	1.19
	(1618.37)	(1454.75)	(1310.34)	(1119.74)	(811.35)	(-20.51)	(1367.11)	(1473.92)	(1550.78)	(1600.28)	(1637.03)	(40.28)
$R_{S,t-1}$	0.02	0.02	0.02	0.04	0.06	0.04	0.05	0.05	0.06	0.06	0.05	0.00
	(0.72)	(0.85)	(0.74)	(1.23)	(1.82)	(2.74)	(1.67)	(1.46)	(1.79)	(2.02)	(1.68)	(-0.00)
DHR_t (EW)	-0.19	-0.20	-0.18	-0.12	0.01	0.20	-0.14	-0.15	-0.16	-0.12	-0.13	0.01
	(-2.39)	(-2.47)	(-2.16)	(-1.46)	(0.07)	(4.61)	(-2.10)	(-2.35)	(-2.26)	(-1.73)	(-1.67)	(0.20)
DHR_t (GRW)	-0.15	-0.17	-0.16	-0.14	-0.06	0.09	-0.16	-0.18	-0.20	-0.18	-0.24	-0.08
	(-1.90)	(-2.15)	(-1.98)	(-1.67)	(-0.70)	(2.05)	(-2.58)	(-2.73)	(-2.97)	(-2.58)	(-3.24)	(-1.71)

Table A7: **Portfolios sorted on illiquidity in the sample from 2012 and 2019 and with additional filters.** This table displays the mean characteristics of portfolios sorted on illiquidity. Illiquidity is measured with the option effective spread ES^O_{t-2}. Moneyness is defined as the closing price of the underlying stock divided by the option strike price $(S/K)_{t-2}$. The return of the underlying stock between the closing of $t-2$ and the close of $t-1$ is $R_{S,t-1}$. DHR_t is the mean across all options on the same underlying stocks of the return of an unhedged option between the opening and the closing of $t-1$ with a hedge put in place from the closing of $t-1$ and t. The table reports the equal weighted (EW) and the gross return weighted (GRW) mean DHR_t. The means weighted by lagged option gross returns (GRW) are reported to correct for microstructure biases described in Duarte, Jones, and Wang (2022). Panel A (B) shows results for the infeasible (feasible) sample. The infeasible and feasible filters used to create these samples are described in Section 4. Returns, moneyness and spreads are in percent. T-statistics are within parentheses.

Panel A: Infeasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.06	1.80	2.59	3.91	8.49	7.43	0.78	1.41	2.03	3.09	6.94	6.16
	(107.27)	(99.56)	(95.99)	(96.79)	(95.04)	(90.71)	(82.32)	(84.82)	(81.70)	(80.58)	(78.41)	(75.71)
$(S/K)_{t-2}$	99.47	99.26	99.11	98.93	98.57	-0.90	100.00	100.17	100.36	100.53	100.89	0.90
	(4934.86)	(4630.58)	(4172.28)	(3877.47)	(3614.74)	(-54.65)	(4784.02)	(4920.67)	(5004.20)	(4676.20)	(4596.05)	(53.04)
$R_{S,t-1}$	-0.17	0.00	0.09	0.16	0.35	0.52	0.22	0.08	-0.01	-0.07	-0.22	-0.44
	(-8.97)	(-0.21)	(4.62)	(8.42)	(16.92)	(42.87)	(10.68)	(3.98)	(-0.60)	(-3.79)	(-11.30)	(-38.38)
DHR_t (EW)	-0.49	-0.02	0.22	0.44	1.01	1.50	-0.20	0.02	0.18	0.30	0.70	0.91
	(-6.45)	(-0.28)	(2.89)	(5.40)	(11.97)	(30.97)	(-2.87)	(0.28)	(2.47)	(3.86)	(8.70)	(19.58)
DHR_t (GRW)	-0.48	0.00	0.24	0.45	1.10	1.58	-0.25	-0.03	0.12	0.25	0.66	0.91
	(-6.32)	(-0.05)	(3.12)	(4.91)	(12.62)	(30.46)	(-3.56)	(-0.46)	(1.68)	(3.21)	(8.02)	(18.80)

Panel B: Feasible

	Calls						Puts					
	1	2	3	4	5	5-1	1	2	3	4	5	5-1
ES^O_{t-2}	1.04	1.75	2.51	3.79	8.13	7.09	0.78	1.38	1.98	2.99	6.66	5.88
	(111.08)	(103.96)	(100.56)	(100.61)	(97.31)	(92.52)	(86.36)	(88.09)	(84.53)	(83.53)	(78.25)	(74.97)
$(S/K)_{t-2}$	99.29	99.23	99.15	99.03	98.80	-0.49	100.14	100.21	100.32	100.44	100.67	0.53
	(6402.24)	(5982.16)	(5095.63)	(4648.72)	(4186.09)	(-35.43)	(6947.25)	(6623.57)	(6808.56)	(7008.17)	(6782.69)	(40.65)
$R_{S,t-1}$	0.05	0.05	0.04	0.05	0.07	0.01	0.06	0.08	0.08	0.09	0.07	0.01
	(2.64)	(2.36)	(2.24)	(2.75)	(3.51)	(1.81)	(2.87)	(3.80)	(4.07)	(4.46)	(3.73)	(1.63)
DHR_t (EW)	-0.12	-0.17	-0.17	-0.05	-0.02	0.09	-0.06	-0.08	-0.12	-0.16	-0.09	-0.02
	(-1.49)	(-2.12)	(-2.17)	(-0.64)	(-0.28)	(2.24)	(-0.92)	(-1.15)	(-1.62)	(-2.16)	(-1.20)	(-0.56)
DHR_t (GRW)	-0.08	-0.11	-0.12	-0.01	-0.01	0.07	-0.09	-0.09	-0.15	-0.19	-0.15	-0.06
	(-1.01)	(-1.39)	(-1.50)	(-0.14)	(-0.15)	(1.57)	(-1.33)	(-1.35)	(-2.11)	(-2.59)	(-2.07)	(-1.34)

Table A8: **Impact of different filters on option return predictability by underlying stock characteristics.** This table presents a comparison of the infeasible and feasible calculations from Table 2, where different filters are applied one at a time. All portfolios are equal-weighted. Average returns are in percent. Annualized Sharpe ratios are within brackets and t-statistics are within parentheses.

	Infeasible	**Feasible**	**Feasible except dividend**	**Feasible except mid arb.**	**Feasible except bid>0**	**Feasible except mid ≥ 1/8**
CFV	2.12	-0.13	-0.13	0.36	1.58	1.34
	(16.99)	(-0.94)	(-0.87)	(2.52)	(13.28)	(8.01)
	[3.78]	[-0.21]	[-0.19]	[0.56]	[2.95]	[1.78]
CH	2.10	0.60	0.58	1.11	1.91	1.44
	(13.90)	(2.70)	(2.69)	(5.24)	(10.59)	(6.26)
	[3.09]	[0.60]	[0.60]	[1.16]	[2.35]	[1.39]
DISP	1.91	-0.11	-0.10	0.36	1.39	1.29
	(14.40)	(-0.69)	(-0.63)	(2.60)	(10.80)	(7.50)
	[3.20]	[-0.15]	[-0.14]	[0.58]	[2.40]	[1.67]
ISSUE_1Y	1.62	0.20	0.19	0.63	1.42	1.05
	(13.68)	(1.59)	(1.47)	(4.79)	(12.23)	(7.32)
	[3.04]	[0.35]	[0.33]	[1.06]	[2.72]	[1.63]
ISSUE_5Y	1.82	0.31	0.22	0.71	1.52	1.32
	(14.16)	(2.14)	(1.54)	(4.84)	(11.58)	(8.33)
	[3.15]	[0.48]	[0.34]	[1.08]	[2.57]	[1.85]
TEF	1.59	0.01	-0.02	0.42	1.41	0.91
	(11.07)	(0.08)	(-0.13)	(2.54)	(9.75)	(5.13)
	[2.46]	[0.02]	[-0.03]	[0.57]	[2.17]	[1.14]
-PM	2.30	0.12	0.09	0.75	1.94	1.41
	(16.47)	(0.66)	(0.49)	(4.31)	(12.87)	(6.93)
	[3.66]	[0.15]	[0.11]	[0.96]	[2.86]	[1.54]
-LN(PRICE)	4.85	0.61	0.65	1.50	4.09	3.61
	(26.48)	(2.74)	(2.89)	(7.11)	(21.53)	(19.52)
	[5.88]	[0.61]	[0.64]	[1.58]	[4.78]	[4.34]
-PROFIT	2.27	0.10	0.10	0.64	1.99	1.37
	(19.18)	(0.54)	(0.55)	(3.78)	(14.57)	(6.83)
	[4.26]	[0.12]	[0.12]	[0.84]	[3.24]	[1.52]
-ZS	0.44	0.10	0.14	0.07	0.31	0.61
	(2.63)	(0.58)	(0.84)	(0.44)	(1.93)	(3.59)
	[0.58]	[0.13]	[0.19]	[0.10]	[0.43]	[0.80]
-VOL_deviation)	3.00	2.81	2.86	3.04	2.86	2.86
	(13.59)	(11.53)	(11.71)	(12.71)	(11.71)	(11.71)
	[3.02]	[2.56]	[2.60]	[2.82]	[2.60]	[2.60]
IVOL	3.78	0.77	0.77	1.53	3.20	2.46
	(25.18)	(3.75)	(3.63)	(7.65)	(20.08)	(11.28)
	[5.60]	[0.83]	[0.81]	[1.70]	[4.46]	[2.51]
Amihud	3.78	0.22	0.24	1.04	3.93	1.37
	(24.40)	(1.21)	(1.34)	(5.68)	(25.51)	(6.83)
	[5.42]	[0.27]	[0.30]	[1.26]	[5.67]	[1.52]

Table A9: **Impact of different filters on liquidity sorts for ATM options** This table presents average return of long-short portfolios of hedged and unhedged options sorted by option illiquidity. The table first repeats the infeasible and feasible calculations from Table 5 before showing the impact of imposing infeasible filters one at a time. All portfolios are equally weighted. T-statistics are in parentheses.

	Infeasible	Feasible	Feasible except spread	Feasible except price	Feasible except delta	Feasible except open int	Feasible except $\text{volume}_t > 0$
Delta-hedged calls	1.52	0.19	0.23	0.19	1.51	0.20	0.13
	(30.59)	(4.86)	(5.84)	(4.88)	(33.27)	(4.97)	(2.98)
Delta-hedged puts	0.92	0.11	0.11	0.11	0.95	0.10	0.03
	(19.41)	(2.64)	(2.70)	(2.66)	(17.12)	(2.33)	(0.63)
Unhedged calls	1.80	0.20	0.23	0.20	1.56	0.19	0.31
	(18.88)	(2.30)	(2.65)	(2.34)	(16.62)	(2.20)	(3.57)
Unhedged puts	0.85	0.06	0.07	0.06	0.94	0.04	-0.12
	(10.30)	(0.75)	(0.83)	(0.74)	(10.82)	(0.48)	(-1.34)