Investor Sophistication and Disclosure Clienteles;

Internet Appendix

Alon Kalay

Columbia Business School

Columbia University

2015

1. Measuring the presence of sophisticated investors via inefficient exercise activity in the options market

1.1 Option Exercise and Ex-Dividend Events – An Overview [1]

When a stock goes ex-dividend the owner of the stock retains the right to the dividend payment, and any future owner will no longer be eligible for the dividend payment. Since the dividend payment is no longer attributed to the ownership of the stock on the ex-dividend day (ex-day), the price of the stock should immediately adjust to reflect this fact, and decline by the value of the dividend amount (ceteris paribus). [2]

In contrast to stockholders, options holders are not entitled to receive the underlying dividend. Therefore, it is optimal to exercise an American call option on the last cum date under the following conditions: 1) the option is in the money 2) the underlying stock pays a dividend 3) the expected time value of the option on the last cum day (the day before the ex-date) is less than the expected drop in stock price (e.g., Roll [1977]). This is true, because in this scenario the price of the option cannot drop prior to the ex-date, to reflect the expected drop in stock price. Otherwise, the option price drops below its intrinsic value and presents an immediate arbitrage opportunity. This occurs precisely because the option is in the money, and the expected time value is less than the expected drop in stock price. As a result, there is an expected (predictable) drop in the option price between the last cum date and the ex-date, equal to the difference between the dividend amount to be paid and the expected time value. In cases where the expected time value is greater than the dividend amount, there is no expected (predictable) drop

[1] This section relies heavily on the discussions in Pool et al. [2008] and Hao et al. [2009].
[2] It is common to estimate the expected drop in stock price by the actual dividend amount per share (e.g., Hao et al. [2009] and Pool et al. [2008]).

in the price of the option, and early exercise is not optimal (see Hao et al. [2009] and Pool et al. [2008] for an extended discussion).

If we define the following:

- S_c = the cum stock price, including the dividend payment
- $E(S_e)$ = the expected stock price on the ex-dividend day
- D^* = the dividend per share (equal to the expected drop in stock price)
- $C_e(X)$ = the expected price of a call option with strike price X on the ex-date, which can be computed with an option pricing model and expected inputs (e.g., expected volatility)

Then:

- $E(S_e) = S_c - D^*$
- And the expected time value of the option on the ex-date can be defined as $\pi_e = C_e(X) - \max(0, E(S_e) - X)$

If $D^* > \pi_e$ then exercising the option on the last cum day is optimal since the expected time value of the option is less than the expected drop in the stock price. When an investor fails to exercise an option in this scenario, he will forfeit a profit = $D^* - \pi_e$ to the writer (seller) of the option, which is the expected drop in the option price between the last cum date and the ex-date.[3,4] The profit is essentially the loss incurred by the buyer of the option, which results in a direct wealth transfer to the seller of the option.

If all the option holders exercise their options optimally, then this price drop does not represent an exploitable profit opportunity to the short seller, because all the open interest in the call option disappears on the last cum date. In other words, the open interest drops to zero on the

[3] Options in which early exercise is optimal are characterized by the fact that they are deep in the money and have a relatively short remaining horizon. For example, the median (mean) option in the sample is ~ $8 ($12) in the money and has 16 (30) calendar days to maturity on the last cum date. Therefore, most of these options will need to be exercised soon in any case and the *incremental cost* associated with early exercise is likely to be insignificant. Furthermore, sophisticated institutional investors face limited transaction costs. Within the retail sector, more sophisticated customers who work with full-service brokers on a fee base do not pay commissions per transaction and would have no additional costs associated with this decision. Hence, even in the face of any transaction costs, *sophisticated investors are more likely to gain from early exercise due to the lower costs they face, and are more likely to exercise their options efficiently.*

[4] Since standard option contracts are based on 100 underlying shares, the profit per contract is 100 times the profit computed above.

last cum date. Therefore, it is optimal to exercise an option when the cost of early exercise (the forgone time value) is lower than the benefit of early exercise (receiving the dividend).

Past research documents a significant portion of American call options remain unexercised even when exercising them is optimal (Kalay and Subrahmanyam [1984], Hao et al. [2009], Pool et al. [2008]). Hao et al. [2009] conclude that approximately 40% of the call options that should have been exercised from 1996–2006 remain unexercised, and that this behavior persists throughout the period. Pool et al. [2008] further demonstrate that option investors have left approximately $491 million on the table over the same period. Given that profit opportunities exist around ex-dividend days, some arbitragers attempt to capture these profits. Hao et al. [2009] and Pool et al. [2008] describe an arbitrage strategy called the “dividend play”, which market makers in the options market engage in. One institutional detail that is important for the market maker’s strategy is that the OCC processes purchases first, then exercises, and finally sales. This allows the market makers to enter into offsetting positions, in order to capture a portion of the unassigned open positions (across all the options holders in the market). Essentially, the market makers buy and sell a large amount of options from each other, which effectively create no position. However, because the clearing house processes exercises before sales, the fact that the market maker sold all of his long positions does not stop him from joining the pool of exercises assigned by the clearing house. Since the market makers exercise 100% of their options, they are likely to end up with some unassigned positions; especially since the exercises are assigned proportionately to the size of the position. Hao et al. [2009] and Pool et al. [2008] provide more in depth discussions of the market makers activities and the related institutional features.

There are two important institutional details related to the ‘dividend play’ for the development of the option-based measure. First, even though the market makers engage in some arbitrage activity, they do not affect the total profit left on the table. They do not affect the amount of profit left on the table because the market makers enter into offsetting positions and exercise all their long (buy) positions immediately. *Hence the amount of profit left on the table is a function of pre-existing investors’ failure to exercise their options efficiently*. The measure developed in this paper is based on the investment activity of these pre-existing option holders. Second, because the market makers enter into offsetting positions, *they do not affect the total amount of open interest*. Their activity only affects the volume in the options market (Pool et al. [2008], footnote 18, page 20). Indeed both Pool et al. [2008] and Hao et al. [2009] document a significant increase in option volume on the last cum day, resulting from the market makers’ activity. This fact is important for the measure developed in this paper, which focuses on the proportion of *open interest* left over for options where early exercise is optimal.

1.2 Measuring the presence of sophisticated information processors

Sophisticated investors (information processors) spend more resources (e.g., time and attention) following their investments and are more capable at analyzing and gathering investment-related information. Therefore, they are more likely to determine if early exercise is optimal and exercise their options efficiently, in order to retain the profit. Conversely, less sophisticated investors (information processors) devote less time and attention to their investments and are less proficient in analyzing investment-related information. Thus they face higher information processing constraints and are less likely to make efficient exercise decisions, forgoing the profit. However, their behavior is not irrational; it results from information processing constraints (Bloomfield [2002], Hirshleifer and Teoh [2003]). Less sophisticated investors find tracking their investments in detail too costly, due to their time constraints or limited ability to analyze and gather information. Following this logic, I develop a new measure to quantify the relative presence of sophisticated investors in the firm's options based on inefficient exercise activity. Ceteris paribus, the lower the percent of unexercised contracts (the lower the percent of open interest that remains open at the close of the last cum date), the higher the proportion of sophisticated investors in the firm's options.

The potential occurrence of inefficient exercise activity can be measured for all firms that pay a cash dividend and have equity call options with positive open interest going into the last cum day. For this set of firms, some call options series will be characterized by the fact that they are in the money, and their expected time value is less than the expected drop in stock price. As a result, all of the options in these series should be exercised on the last cum day, and their open interest should decline to zero going into the ex-dividend day (by the close of the last cum day).[5]

For a particular firm-dividend event (underlying equity security), multiple option series will be traded with open interest going into the last cum day, some of which need to be exercised. To measure the proportion of sophisticated dollars invested in the firm's options at that point in time, I aggregate the percent of contracts that remain open on the close of the last cum day, in a particular series in which exercise is optimal, across all such series. For example, assume a firm has two option series in which exercise is optimal on the last cum day. Further

[5] One potential concern is that a sophisticated investor chooses not to exercise his option to maintain his portfolio allocation, for example, because he is capital constrained. Then the measure would not identify the sophistication of the investor base correctly. However, this constraint is less likely to affect the decision of a sophisticated investor for several reasons: (1) Options in which early exercise is optimal are characterized by having a relatively short remaining horizon, which would imply that an investor needs to rebalance his portfolio soon in any case. (2) Sophisticated investors have alternative instruments, such as equity swaps, which allow them to exercise their options without taking delivery of their shares, while still remaining exposed to the underlying movements in the stock. (3) If more sophisticated investors are wealthier and have more capital, these investors are less likely to be constrained by these frictions.

assume the first series has 100 open contracts going into the last cum day in which exercise is optimal, and 40 contracts remain open on the close of the last cum day, whereas the second series has 300 contracts going into the last cum day, and 150 contracts remain open at the close of the last cum day. Then, holders of 40% of the contracts in the first series and 50% of the contracts in the second series forgo some profit. These percentages are aggregated across the two series, and the final value of the measure for this firm-dividend event would equal $\frac{40+150}{100+300} = 47.5\%$, which is the average percentage of open interest across the two series, weighted by the prior level of open interest.

To aggregate the information present in all of the option series for a given firm-dividend event I make the following assumption: an investor either exercises all or none of her options for a particular firm-dividend event. This assumption is reasonable because a significant portion of the relevant processing costs are at the firm level and have a fixed nature. In untabulated robustness tests, I find similar results to those presented in the paper using alternative aggregation techniques. As one example, using only the most liquid option for each firm dividend event yields similar results.

Formally, the proposed measure equals:

$$(\%)\ \mathit{Open\ Interest}_j = \frac{\sum_i \text{open interest going into the ex day}_{ij}}{\sum_i \text{open interest going into the last cum day}_{ij}} \quad \text{for all relevant}$$

option series *i* in each firm-dividend event *j*.

To compute the measure for a given firm-dividend event, I estimate the expected time value for all of the call options with positive open interest on the open of the last cum day. I then identify all options in which the dividend amount to be paid is greater than the expected time value, and the potential profit is positive. Following Hao et al. [2009] and Pool et al. [2008] I use the actual dividend amount as an estimate of the expected drop in stock price. To accurately estimate the expected time value of the option on the ex-dividend day (π_e), an investor needs to use an option pricing model during the last cum day. An investor further needs to estimate the expected stock price on the ex-dividend day ($E(S_e)$) and the expected volatility on the ex-dividend day (σ_e), which serve as inputs for the model. The remaining required variables (time to maturity, strike price, and the interest rate) are deterministic and would not need to be estimated.

Following the methodology presented in Hao et al. [2009] and Pool et al. [2008], the expected time value of the option is estimated as follows:

1. $E(S_e)$, the expected stock price on the ex-dividend day equals the closing price on the last cum day minus the upcoming dividend amount.
2. $C_e(X)$, the expected price of the call option with strike price X on the ex-dividend day is computed using the Black-Scholes-Merton Model,[6] with the following inputs:
    a. $E(S_e)$: the expected stock price from above.
    b. σ_e: the expected volatility of the underlying security on the ex-dividend day equal to the annualized standard deviation of the logarithmic daily returns over the prior 60 days.[7]

---

[6] Adjusting the model for future dividend payments is not a primary concern for this sample, because options for which early exercise is optimal expiry relatively soon and are unlikely to be affected by future dividend payments (between the ex-day and expiry). In the final sample, approximately 90% of the options experience no future dividend payments. For the remaining 10%, adjusting for future dividends would decrease the expected time value of the option, thus increasing the profit from early exercise. For these few cases the profit from early exercise is understated. As a result, these options also require early exercise and are classified correctly.

[7] Pool et al. [2008] use historical volatility in their valuation, whereas Hao et al. [2009] use implied volatility where available and an estimated implied volatility parameter for the remaining cases. My approach follows Pool et al. because implied volatilities are likely to be less reliable when measured around the ex-dividend day, due to the dividend arbitrage activity documented in both papers.

c. The deterministic variables are based on values reported in the Option Metrics database: T – time to maturity, measured in years between the ex-dividend day and the expiry day; R – the zero coupon rate; and X – the exercise price.

3. Finally, the expected time value, $\pi_e = C_e - \max(E(S_e) - X, 0)$.

In practice, investors may use alternative pricing models to compute the time value of the option. To alleviate the concern that differences in pricing models are responsible for my identification of options in which early exercise is optimal, I require the open interest at the close of the last cum day (going into the ex-day) to be lower than the open interest at the open of the last cum day (close of the prior day) for an option series to be included in the measure. In other words, to classify an option as having potential profits from early exercise, I require at least some investors to unwind their positions so that my identification is more likely to be aligned with that of the market. Furthermore, to eliminate the effect of observations with few outstanding contracts, I also require a minimum open interest level of at least 50 contracts on the last cum day for a particular series to be included in the aggregate measure. Finally, I impose a minimum profit restriction of $0.05 for each option series that is included in the aggregate measure, which implies that an investor forgoes at least $5.00 (per contract) if he fails to optimally exercise an option contract in the series. Although this cutoff is admittedly arbitrary, I impose it to increase the likelihood that exercise is optimal in cases in which I mistakenly overestimate the profit from early exercise using my model. In untabulated robustness tests, I find that the results are generally unchanged when I relax these restrictions. The final sample used in the paper is presented in Table 1.

Variation in the option-based measure is expected to correlate with variation in investor sophistication because on average, sophisticated investors spend more resources following their investments and are more likely to identify cases where early exercise is optimal. However, variation in the profit from early exercise may also lead to variation in the option-based measure. For example, if sophisticated investors exercise their options early more often (on average) when the estimated profit is higher. In other words,

$$((\%)\ \textit{Open Interest}_j) = f\,(\textit{investor base, profit from early exercise}).$$

Therefore, I include two additional control variables in the regression analysis to capture the cross-sectional variation in open interest that results from the cross-sectional variation in profit from early exercise. First, I include the dividend amount as a firm level control variable for the cross-sectional variation in the gross profit from early exercise, Second, I include the bid-ask spread in the firm's stock as a firm level control variable for the potential variable cost associated with early exercise. In theory, only the incremental cost associated with early exercise should matter to a sophisticated investor and since she would pay half of the spread when exercising the option at expiry, the bid-ask spread in the stock does not necessarily result in any incremental costs. However, when the spread is relatively high on the last cum date, an investor may be more concerned with the level of the spread and willing to wait to exercise her option at expiry, when spreads tend to decline. In this case, higher spreads will result in higher incremental costs (expected/perceived costs) associated with early exercise, which offset the potential gain from early exercise.[8]

---

[8] Pool et al. [2008] discuss why exercise is preferable to selling an option, given the lower transaction costs present in the equity market. Hence the bid-ask spread in the equity market is the relevant potential cost.
In un-tabulated results, I find that for the entire sample, the mean (median) spread is not lower (drops by $0.01) between the last cum day and expiry. However, for observations in which spreads are in the top quartile, the mean (median) spread is lower by $0.04 ($0.03) between the last cum day and expiry (both significant). These findings support the idea that investors may perceive incremental transaction costs associated with early exercise when spreads are relatively high during the last cum day. As far as the magnitudes of the spreads are concerned, the 75th percentile for spreads in the sample equals approximately 20 cents / share. Therefore, spreads in the top quartile are sufficiently large to affect an investor's decision, because a standard option contract is on 100 shares and the average profit from early exercise per share is therefore approximately $0.24.

2. Descriptive Statistics

Table 2 presents descriptive statistics for the sophistication measure and the underlying options included in the measure. Panel A reveals that options in which early exercise is optimal are deep in the money and have a relatively short remaining horizon. The median option in the sample has 16 days to expiry and is $8.80 in the money. Hence, these options are likely to require exercise soon, irrespective of the early exercise decision. Therefore, any potential transaction costs (to exercise the option or renew the position) would have to be *incremental* to the transaction costs associated with closing out the position at expiry, and are less likely to affect the exercise decision. The mean (median) profit from early exercise is $24.50 ($17.20) per contract, representing approximately 2% of the price of an average contract in the sample. A large enough gain to motivate a sophisticated investor devoted to her investments. The average level of the measure reported in Panel B equals 35%. This level is similar to the levels reported by Hao et al. [2009] and Pool et al. [2008], who find that 40% of the call options that should have been exercised from 1996–2006, remain unexercised. In the empirical analysis employed in the paper, a log transformation of the measure equal to $\ln(1+\%(Open\ Interest_j))$ is employed. The transformed measure has a mean (median) value of 0.28 (0.25), where lower values represent a higher concentration of sophisticated investors.

To examine the time series properties of the measure I rank all the observations based on the sophistication measure in a given quarter-year. Then, I examine the Spearman correlation between the rankings of a given firm, measured over two (three) consecutive observations. If the measure captures an element of the firm’s ownership structure, then it should exhibit some level of persistence. However, it is difficult to establish an ex-ante expectation for the magnitude of the persistence. The results in Panel B show the measure is persistent through time. The correlation between the rankings across two (three) consecutive observations is approximately 35%, significant at the 1% level. In Panel C, observations are sorted into quintiles and the transition matrix between two consecutive observations is created. Firms in the top (bottom) quintile are more likely to remain in the top (bottom) two quintiles then they are to move to the remaining quintiles. Taken together, the results show the measure is persistent within the firm, consistent with the idea that the measure captures an element of the investor base.

REFERENCES

Bloomfield, R. J., 2002, The Incomplete Revelation Hypothesis and Financial Reporting, *Accounting Horizons* 16, 233–243

Hao, J., A. Kalay, and S. Mayhew, 2009, Ex-Dividend Arbitrage in Option Markets, *Review of Financial Studies,* 23, 271–303

Hirshleifer, D., and S. H. Teoh, 2003, Limited Attention, Information Disclosure and Financial Reporting, *Journal of Accounting and Economics* 36, 337–386

Kalay, A., and M. G. Subrahmanyam, 1984, The Ex-dividend Day Behavior of Option Prices, *Journal of Business 57,* 113–128

Pool, V. K., H. R. Stoll, and R. E. Whaley, 2008, Failure to Exercise Call Options; An Anomaly and a Trading Game, *Journal of Financial Markets* 11,1–35

Table 1. Sample Composition Process

Panel A describes the sample composition process used in the paper. The sample is determined by the information required to compute the option-based measure of investor sophistication (the sophistication measure). To create the sample, I begin by collecting information on all the cash dividend payments available on CRSP, made by firms listed in the U.S. (common share code 10, 11), over the period from 1996–2007. I then proceed to gather information on all the outstanding call options available in the Option Metrics database for each dividend payment. From this set of options, I retain all the options that are in the money, and have positive open interest at the open of the last cum date. Finally, I compute the profit from early exercise for the remaining options and retain all the options where there are at least 50 contracts open at the beginning of the last cum date, there is a decline in open interest during the last cum date, and the profit from early exercise is at least $0.05/share. Panel B presents the number of observations included in the final sample, broken down by year.

Panel A **Sample Composition Process**	Number of Options (Payments)	Number of Dividend Events	Number of Firms
Dividend payments (events)	66,279	66,279	2,528
Outstanding equity call options linked to dividend payments	757,699	28,356	1,269
Call options that are in the money, and have positive open interest going into the last cum day	307,329	27,035	1,255
Options where the profit from early exercise on the last cum day is positive and there is a reduction in open interest during the last cum date	43,351	10,443	1,033
Options where the profit is at least $0.05 per share, and the open interest going into the last cum day is at least 50 contracts and	22,712	7,860	756
Final number of option series included in the aggregate measure of investor sophistication	22,712	7860	756

Panel B
Distribution of Observations by Year

Year	Dividend Events	Firms
1996	239	118
1997	321	153
1998	241	134
1999	260	145
2000	294	163
2001	408	205
2002	604	271
2003	1019	392
2004	1150	455
2005	1105	463
2006	1108	458
2007	1111	471
Total	7860	756

Table 2. Descriptive Statistics for the Sophistication Measure and Underlying Options

This table reports descriptive statistics for the features of the options included in the final sample (see Table 1). Panel A reports the characteristics of the underlying options included in the measure. Panel B reports statistics for the sophistication measure. To compute the correlations, observations are ranked based on the sophistication measure each quarter. The spearman correlations for the rankings across consecutive observations within each firm are then reported in Panel B. Panel C presents a transition matrix for the sophistication measure across consecutive observations. This analysis is limited to all the observations where consecutive observations are within a calendar year. Results are reported for quintiles defined using the entire sample, and for quintiles that are re-defined each quarter-year. Panel C reports the percent of observations in each quintile at time T = t+1 for all the observations in quintile 1 and 5 at time T = t.

Panel A
Option Characteristics for Options Where Exercise is Optimal (n=22,712)

	Mean	Q1	Median	Q3	σ
Profit per option contract ($)	24.5	10.2	17.2	27.5	37.4
Open interest (last cum date)	1,445	116	285	929	4,946
Days to maturity	30.6	8.0	16.0	38.0	49.4
Depth in the money (S-X) in ($)	11.9	4.8	8.8	15.1	11.2

Panel B
Distribution of the option-based measure

	Mean	Q1	Median	Q3	σ	N
% Open Interest	0.35	0.09	0.28	0.57	0.29	7860

	Correlation	N
Correlation between measure ranks across consecutive observations at time [t, t-1]	0.36***	7104
Correlation between measure ranks across consecutive observations at time [t, t-2]	0.32***	6474

* significant at 10%; ** significant at 5%; *** significant at 1%

Panel C
Transition matrix across two consecutive observations [t, t+1]

Quintile at time T = t+1	1	2	3	4	5	N
Analysis for Quintile 1 (at time T=t)						
Quintiles based on the entire sample	30.1	23.4	17.1	14.6	14.9	1284
Quintiles re-defined every quarter	30.0	21.0	17.0	15.6	16.5	1280
Analysis for Quintile 5 (at time T=t)						
Quintiles based on the entire sample	14.1	16.5	17.8	21.7	30.0	1411
Quintiles re-defined every quarter	16.3	18.2	19.3	21.1	25.1	1436