

Data Dictionary

Variable	Variable Name	Variable Definition	Data Type
IPO Pricing – Used to derive Targets			
P(IPO)	Offer Price	This is IPO final offering price (in dollars) stated in the 424B4 filings.	Continuous/Float
P(H)	Price Range Higher Bound	This is the higher bound (in dollars) of the filing price range stated in the S-1 filings.	Continuous/Float
P(L)	Price Range Lower Bound	This is the lower bound (in dollars) of the filing price range stated in the S-1 filings.	Continuous/Float
P(1Day)	First Day Trading Price	This is the first day trading (closing) price of the stock (in dollars).	Continuous/Float
IPO Characteristics – Economic/Accounting/Financial Determinants of IPOs			
C1	Days	Calendar days between the initial S-1 filing and the filing of Form 424B4 from EDGAR.	Continuous/Float
C2	Top-tier Dummy	Dummy variable, set to 1 if leading underwriter of IPO has a rating of 8 or higher, otherwise 0	Binary/Boolean
C3	Earnings Per Share	Trailing Earnings Per Share (EPS) at the time of the IPO (in dollars)	Continuous/Float
C4	Prior Nasdaq 15-day Returns	The buy-and-hold returns of the CRSP Nasdaq value-weighted index on the 15-trading days prior to the IPO day (t), ending on day $t-1$	Continuous/Float
C5	Outstanding Shares	Defined as the number of outstanding common stock shares in the IPO	Continuous/Float
C6	Offering Shares	Defined as the number of offering common stock shares in the IPO	Continuous/Float
C7	Sales	Trailing annual firm sales (in millions of dollars) at the time of the IPO	Continuous/Float
Textual Characteristics of IPO Prospectus MD&A Sections			
T1	Number of Sentences	Number of sentences in the MD&A Sections	Continuous/Float
T2	Number of Words	Number of words in the MD&A Sections	Continuous/Float
T3	Number of Real Words	Number of textual words in the MD&A Sections	Continuous/Float
T4	Number of Long Sentences	Number of sentences with more than 20 words in the MD&A Sections	Continuous/Float
T5	Number of Long Words	Number of words with more than 10 letters in the MD&A Sections	Continuous/Float
Sentiment Characteristics of IPO Prospectus MD&A Sections			

S1	Number of Positive Words	Number of positive words in the MD&A Sections, according to the L&M dictionary	Continuous/Float
S2	Number of Negative Words	Number of negative words in the MD&A Sections, according to the L&M dictionary	Continuous/Float
S3	Number of Uncertain Words	Number of uncertain words in the MD&A Sections, according to the L&M dictionary	Continuous/Float
Targets Variables Need Calculation			
Y1	Pre-IPO Price Revision	Set to 1 if $P(\text{IPO}) < P(\text{mid})$, where $P(\text{mid}) = (P(\text{H}) + P(\text{L}))/2$, otherwise 0	Binary/Boolean
Y2	Post-IPO Initial Return	Set to 1 if $P(\text{IPO}) < P(1\text{Day})$, otherwise 0	Binary/Boolean
Control Variables Need Calculation			
C3'	Positive EPS Dummy	Dummy variable set to 1 if C3 is positive at the time of the IPO, otherwise 0	Binary/Boolean
C5'	Share Overhang	$C5' = C5/C6$	Continuous/Float
C6'	Up Revision	Set to $(P(\text{IPO}) - P(\text{mid}))/P(\text{mid}) \times 100\%$, if $P(\text{IPO}) > P(\text{mid})$, otherwise 0	Continuous/Float
IPO Identifiers			
I1	Ticker	Ticker of the IPO	Text
I2	Company Name	Name of the company offering the IPO	Text
I3	Standard Industry Classifier	Standardized Industrial Classifier of the offering company	Text

HINTS:

1. Your data should have 2 categorical variables (Y1 and Y2), and 7 IPO characteristics variables (C1 - C2, C3', C4, C5' - C6', C7); 2. You should avoid using actual counts (T1 - T5, S1 - S3), but ratios (i.e. percent of long sentences $T1' = T4/T1$, percent of positive words $S1' = S1/T2$); 3. You should avoid at any cost of dropping any variable, or any observations;
4. You should tread carefully with missing values ("").