Estimating IPO Underpricing

TeWhan Hahn
Auburn University at Montgomery

ABSTRACT

This study examines alternative regression methods that address common issues found in IPO underpricing data, specifically those that violate the assumptions of Ordinary Least Squares (OLS) regression. These issues can significantly impact the results, including the estimated values of coefficients and their statistical significance. The study demonstrates that using different estimation methods, such as OLS, median regression, and robust regression, can lead to substantially different findings in IPO underpricing research. Additionally, the way standard errors are calculated can also influence the significance of the results. Therefore, it is recommended that studies investigating IPO underpricing rigorously test the assumptions of their chosen estimation method before applying it. This will ensure they select the most appropriate technique for their data and achieve more reliable results.

Keywords: cluster-adjusted standard error, heteroscedasticity-adjusted standard error, IPO underpricing

INTRODUCTION

The phenomenon of IPO (Initial Public Offering) underpricing (i.e., IPO's high initial return which is the percentage change in stock price from the offering price to the closing price on the first day of trading) has been one of the most researched and published finance topics and has spawned interests among researchers and market participants. Ljungqvist (2007) provides an overview of different theories suggested to explain IPO underpricing.

Many theories and explanations have been suggested to explain the average 20% or so first day of trading return for US IPOs. While different studies used different estimation methods in IPO initial return regressions, many studies simply ignored various data problems in the IPO data and used Ordinary Least Squares (OLS) regression method to explain variations in IPO initial returns.

However, econometric literature suggests that researchers should check if the OLS assumptions are satisfied, especially normality of residuals and homoscedasticity (Neter, Kutner, Nachtsheim, & Wasserman (1996)). This study tested heteroskedasticity using Breusch–Pagan/Cook–Weisberg (Breusch and Pagan (1979)) and found that the test's null hypothesis of homoscedasticity was easily rejected, suggesting the heteroscedasticity in the studied IPO data (1995-2010). Also tested was normality of residuals using Shapiro–Wilk W test (Shapiro and Wilk (1965)) and the null hypothesis of normality was easily rejected, suggesting non-normality in the studied IPO data. In addition, the outlier test identifies about 3.4% of the data points are severe outliers in the studied IPO data. These preliminary tests ask for the alternative estimation methods to be used in IPO initial return studies.

This study does not suggest an alternative theory or explanation of IPO underpricing. Rather, motivated by the above simple test results, it presents the different estimation results of IPO underpricing regressions by applying different regression estimation methods for comparison purposes, which researchers can use as a guide when presenting their empirical results. This fills the gap in the IPO initial return literature.

ESTIMATION METHODS USED IN IPO UNDERPRICING STUDIES

IPO studies have employed predominantly linear regression to estimate the empirical model of IPO initial returns using US data. Researchers' approach to estimate empirical models could be determined in two distinct ways. First, when the researchers were interested in measuring accurate empirical models, at least three different regression techniques could be used to estimate the accurate relationship between IPO underpricing and its possible determinants. The early IPO initial return studies used the ordinary least squares (OLS) method. OLS assumes the dependent variable (predicted outcome) is continuous and normally distributed around the predicted values. It also assumes there's no censoring in the data, meaning all observations are complete and within the observed range. OLS estimates the mean relationship between the independent variables (predictors) and the dependent variable. Therefore, OLS is suitable for continuous, normally distributed data with no censoring. In OLS, the coefficients represent the average change in the dependent variable for a one-unit change in the independent variable, holding other variables constant. However, based on the identified data characteristics, there are two different estimation methods that could be used to measure the accurate relationship between IPO underpricing and its possible determinants.

IPO data shows many potential outlier observations, which asks for a remedy to the impact of outliers on normality and homoscedasticity of errors. For example, as table 1 shows, the range of initial return of IPOs is very wide. In an unreported plot, there are many data points that are far from most of other data points. Therefore, alternative estimation methods to OLS should be considered.

The first alternative is robust regression (Davies and Gather (2012)). When data might have outliers or violate assumptions of least squares regression (OLS), robust regression provides a more reliable alternative. It's less sensitive to these issues, leading to more accurate estimates of coefficients and standard errors. It employs maximum likelihood-type estimators that down-weight the influence of extreme observations. In robust regression, the estimated coefficients represent the estimated effects of the independent variables on the dependent variable, just like in OLS. The robust regression's standard errors are less sensitive to outliers, providing a more accurate estimate of the variability around the coefficients.

The second alternative to OLS is median regression (Koenker and Bassett (1978)). Median regression makes no assumptions about the distribution of the dependent variable. It simply identifies the middle value when the data is ordered from least to greatest. Median regression estimates the median relationship between the independent variables and the dependent variable, focusing on the central tendency rather than the mean. Median regression is useful for data with non-normal distributions, outliers, or when the focus is on the middle value rather than the mean. It's also robust to extreme values and can be less sensitive to outliers compared to OLS. In Median regression, the coefficient represents the change in the median of the dependent variable for a one-unit change in the independent variable, holding other variables constant.

The appropriate regression method depends on the specific research question, data characteristics, and assumptions a researcher is willing to make.

Second, when studying IPO initial returns, researchers might believe a standard linear regression (OLS) captures the basic relationships. However, their focus may be to determine if specific independent variables truly influence initial returns, as suggested by theories explaining high IPO returns. In this case, getting the most accurate standard errors becomes crucial. There are three main approaches for estimating standard errors in IPO initial return regressions:

Ordinary Least Squares (OLS): This is the standard method for linear regressions.

Robust Standard Errors: This technique addresses potential issues with OLS standard errors when the variance of the errors isn't constant across observations (heteroskedasticity).

Cluster-Adjusted Standard Errors: This method is used when observations are grouped (clustered) and errors within a group might be correlated. It provides more accurate standard errors in such situations.

The following section will discuss the key differences between these methods:

OLS estimates the coefficients (slopes and intercept) of a linear regression model. OLS assumes constant variance of errors (homoscedasticity) and normally distributed residuals. Early IPO underpricing studies used OLS. OLS is preferred when the assumptions of homoscedasticity and normal residuals are met, and no dependence within groups exists. Standard errors provide an estimate of the variability of the coefficient estimates and are used to construct confidence intervals and hypothesis tests.

The first alternative to OLS in estimating standard errors in IPO underpricing studies is robust standard errors. Robust standard errors adjust the standard errors of OLS coefficients to account for potential violations of the assumptions of OLS, such as heteroskedasticity (unequal variance of errors) or non-normality of residuals. In IPO underpricing literature, many studies such as Purnanandam and Swaminathan (2004) and Belghitar and Dixon (2012) used robust standard errors in estimating standard errors. Robust standard errors do not require the same assumptions as OLS. They are "robust" to violations of these assumptions. Robust standard errors are used when there is evidence of heteroskedasticity, non-normality of residuals, or both, regardless of dependence within groups. These are typically wider than OLS standard errors, reflecting the increased uncertainty due to violations of assumptions.

The second alternative to OLS in estimating standard errors in IPO underpricing studies was cluster adjusted standard errors (Cameron and Trivedi (2013)). Cluster adjusted standard errors adjust the standard errors of OLS coefficients to account for dependence within groups (clusters) of data. Ritter (1984) noted that IPOs are coming in waves and tend to cluster. Ljungqvist and Wilhelm (2003) used cluster adjusted standard errors in the IPO underpricing regression. In this case, the use of cluster adjusted standard errors is relevant because IPOs within a month or a year are likely to be more like each other than observations from different months or years. Cluster adjusted standard errors do not require the assumption of homoscedasticity or normality of residuals, but they assume independence within groups is not present. Cluster adjusted standard errors are used when there is potential dependence within groups, regardless of whether the assumptions of homoscedasticity or normality are met. These can be wider or narrower than OLS standard errors depending on the degree of dependence within groups.

OLS estimates the coefficients, while Robust and Cluster adjusted standard errors address potential issues with those estimates. Robust addresses violations of homoscedasticity or normality, while Cluster adjusted addresses dependence within groups. The choice depends on the specific data characteristics and potential issues present.

This study, using 2000-2015 IPO data, investigates whether different regression estimation methods and different standard error estimation methods yield materially different estimation results of IPO initial returns.

DATA AND VARIABLE CONSTRUCTION

Data is from SDC Global New Issues database for the period 1995-2010. The data contains the firm commitment offerings for the given data period. Extracted were the offer_price and number of shares offered by SDC. IPOs with offer prices below $5 were excluded because it is well known that penny stocks are significantly different from the others. The list of independent variables used in this study is explained in Butler, Keefe, and Kieschnick (2014) or in the studies cited in the study.

Variables used in this study are defined as follows. Initial return is defined as a percent change from IPO offer_price and the first trading day closing price. Offer_price is the offer_price of the IPO. prim_shs_pct is the percent of newly created and offered shares in the total shares offered. partadj is the percent change from mid filing price to the offer_price. lnlock is the natural log of days in lockup period of the IPO. Venture1 is the dummy that has value of one if the IPO is venture-backed and zero otherwise. Hot is the dummy that has value of one if the IPO goes public in the month when the average IPO initial return is higher than the initial return for the whole sample and zero otherwise. mktrf is the market return more than risk-free rate in the year before the IPO. lnage is the natural log of (1+firm age). rank is the underwriter' rank of the IPO complied by Loughran and Ritter. lnintensity is the natural log of number of

IPOs in the three-month period prior to the IPO. lnproceeds is the natural log of expected proceeds at offer_price and calculated as the natural log of (total shares outstanding times offer price). Tech is the dummy variables that has value of one if the IPO is classified as tech firm and zero otherwise.

EMPIRICAL RESULTS

Table 1 presents summary statistics of variables. Initial return has an average of 27.68% and standard deviation 61.83%. The range of initial return shows significant variation to be explained. offer_price is $13.29 on average with min $3.5 and max $97. prim_shs_pct has mean 91,15% with min 2,84% and max 100%. partajd has mean 3% with -98.4% and max 805.2%. lnlock has an average of 4.03 and min 0 to max 7.51. mktrf has 1.24% with min -16.2% and max 8%. lnage has a mean 1.94 with min 0 and max 5.11. rank has mean 7.19 with min 0 and max 9. lnintensity has mean 4,67 with min 0 and max 5.51. lnproceeds has a mean of 17.74 with min 14.91 and max 22.71. tech has a mean value of 0.33 with min 0 and max 1.

Table 2 presents results from OLS, Median, and Robust regressions. Since median regression estimates the median relationship between initial return and its independent variables, the effect of outliers on the estimated coefficients is mitigated compared to the case of OLS. If the effect of outliers is significant, we can expect that the significance of independent variables changes a lot. We see some evidence of the effect from offer_price, prim_shs_pct, lnage, rank, and lnproceeds. In the case of robust regression, we see some evidence of the effect from offer_price, prim_shs_pct, lnage, and rank. Note that coefficients of independent variables change significantly in robust and Median regressions. Especially the coefficients of offer_price and lnage change their signs in robust and median regressions.

Table 3 presents results of OLS, Robust, and Cluster adjusted standard errors. Note the coefficients are the same. Asterisks on the standard errors represent the statistical significance of coefficients. When a researcher is more interested in testing the significance of a key independent variable, the use of a more robust method to estimate standard errors may be the focus of the interest. The starting point is again OLS estimation. When robust standard error method was used, offer_price, lnage, and rank changed their significance. When cluster adjusted standard error method with IPO year as a cluster, offer_price, prim_shs_pct, venture1, lnage, rank, inintensity, change their statistical significance. When cluster adjusted standard error method with IPO year & month as a cluster, offer_price, lnage, and rank change their statistical significance.

CONCLUSION

The IPO data used in this study exhibited several issues that violate the assumptions of Ordinary Least Squares (OLS) regression. These issues included non-normal residuals, heteroscedasticity (unequal variance), and significant outliers in the IPO initial return data. To address these problems, the study employed alternative regression methods for analyzing IPO initial returns. The results showed significant differences in the estimated coefficients depending on the chosen method. Additionally, using different methods to calculate standard errors led to variations in the significance of independent variables.

Based on these findings, we recommend that future studies investigating IPO underpricing rigorously test the assumptions of OLS before applying the method. Researchers should also consider using appropriate remedial measures if these assumptions are violated.

REFERENCES

Belghitar, Y., & Dixon, R. (2012). Do venture capitalists reduce underpricing and underperformance of IPOs?. *Applied Financial Economics*, *22*(1), 33-44.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random effects. Econometrica, 47(5), 1287-1294.

Butler, A. W., Keefe, M. O. C., & Kieschnick, R. (2014). Robust determinants of IPO underpricing and their implications for IPO research. *Journal of Corporate Finance*, *27*, 367-383.

Cameron, A. C., & Trivedi, P. K. (2013). Microeconometrics using Stata (revised ed.). Stata Press.

Davies, L., & Gather, U. (2012). *Robust statistics* (pp. 711-749). Springer Berlin Heidelberg.

Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33-50.

Ljungqvist, A. (2007). IPO underpricing. *Handbook of empirical corporate finance*, 375-422.

Ljungqvist, A., & Wilhelm Jr, W. J. (2003). IPO pricing in the dot-com bubble. *The Journal of Finance*, *58*(2), 723-752.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models.

Purnanandam, A. K., & Swaminathan, B. (2004). Are IPOs really underpriced?. *The Review of Financial Studies*, *17*(3), 811-848.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete tables). Biometrika, 52(3/4), 591-611.

Ritter, J. R. (1984). The" hot issue" market of 1980. *Journal of Business*, 215-240.

APPENDIX

Table 1
Summary Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
initret	3,438	27.68%	61.83%	-98.3%	1304.2%
offer_price ($)	3,438	13.29	5.86	3.5	97
prim_shs_pct	3,351	91.15%	16.84%	2.84%	100%
partadj	3,342	3%	31%	-98.4%	805.2%
lnlock	3,438	4.03	2.25	0	7.51
venture1	3,438	0.43	0.50	0	1
hot	3,438	0.53	0.50	0	1
mktrf	3,294	1.24%	3.75%	-16.20%	8%
lnage	3,438	1.94	1.15	0	5.11
rank	3,350	7.19	2.44	0	9.00
lnintensity	3,438	4.67	0.70	0	5.51
lnproceeds	3,438	17.74	1.08	14.91	22.71
tech	3,438	0.33	0.47	0	1

Table 2
Comparison of results from OLS, Median, and Robust regressions

	OLS		Robust		Median	
Variable	Coeff	STD Err	Coeff	STD Err	Coeff	STD Err
offer_price	0.0174***	0.0032	-0.0012	0.0011	-0.0025	0.0016
prim_shs_pct	0.0026***	0.0006	0.0001	0.0002	0.0005*	0.0003
partadj	0.7445***	0.0475	0.5262***	0.0163	0.6531***	0.0233
lnlock	-0.0219***	0.0048	-0.0049***	0.0016	-0.0094***	0.0023
venture1	0.0801***	0.0215	0.0357***	0.0074	0.0505***	0.0105
hot	0.1368***	0.0218	0.0394***	0.0075	0.0513***	0.0107
mktrf	0.0037	0.0027	0.0017*	0.0009	0.0012	0.0013
lnage	-0.0200**	0.0093	0.0062*	0.0032	0.0039	0.0046
rank	0.0098*	0.0055	0.0051***	0.0019	0.0055**	0.0027
lnintensity	-0.0419**	0.0178	-0.0126**	0.0061	-0.0174**	0.0087
lnproceeds	-0.0926***	0.0172	-0.0138**	0.0059	-0.0158*	0.0084
tech	0.0935***	0.0226	0.0369***	0.0078	0.0461***	0.0111
constant	1.5405***	0.3259	0.3780***	0.1119	0.4307***	0.1597
N	3,124		3,124		3,124	
R-Square	23.28%		-		14.71%	

One, two and three asterisks indicate the significance at the 10, 5, and 1 percent levels, respectively.

Table 3
Comparison of results from OLS, Robust, and Cluster-adjusted Standard Errors

		OLS	Robust	Cluster(Y)	Cluster(M)
Variable	Coeff	STD Err	STD Err	STD Err	STD Err
offer_price	0.0174	0.0032***	0.0077**	0.0091*	0.0077**
prim_shs_pct	0.0026	0.0006***	0.0004***	0.0010**	0.0005***
partadj	0.7445	0.0475***	0.1133***	0.1496***	0.1195***
lnlock	-0.0219	0.0048***	0.0051***	0.0050***	0.0053***
venture1	0.0801	0.0215***	0.0229***	0.0439*	0.0243***
hot	0.1368	0.0218***	0.0210***	0.0252***	0.0253***
mktrf	0.0037	0.0027	0.0025	0.0039	0.0036
lnage	-0.0200	0.0093**	0.0109*	0.0141	0.0111*
rank	0.0098	0.0055*	0.0043**	0.0065	0.0045**
lnintensity	-0.0419	0.0178**	0.0176**	0.0295	0.0206**
lnproceeds	-0.0926	0.0172***	0.0285***	0.0273***	0.0285***
tech	0.0935	0.0226***	0.0248***	0.0229***	0.0266***
constant	1.5405	0.3259***	0.4626***	0.4374***	0.4560***
N		3,124	3,124	3,124	3,124
R-Square		23.28%	23.28%	23.28%	23.28%

One, two and three asterisks indicate the significance of coefficients at the 10, 5, and 1 percent levels, respectively.