Determinants of Amount Raised in Equity Crowdfunding Campaigns; an Application of Truncated Regression

Xuerui MA[a,*], Peter G. MOFFATT[a,†], Simon A. PETERS[b,‡]

[a] *School of Economics, University of East Anglia, Norwich, UK*

[b] *Department of Economics, University of Manchester, Manchester, UK*

June 6, 2025

Abstract

Secondary-market data on 1,189 crowdfunding campaigns, extracted from the leading UK equity crowdfunding platform Seedrs, is used to identify the determinants of amount invested. A key feature of the data set is that only successful campaigns are observed. For this reason, we depart from previous literature by applying the truncated regression estimator (Hausman & Wise 1977) instead of OLS regression. We cluster at the company level in order to allow for multiple campaigns per company. The importance of allowing for truncation is confirmed using the Hausman Test. Moreover, allowing for truncation alters the conclusions arising from previous literature concerning the effects of standard variables including target amount, amount of equity offered, and number of team members participating in the venture. Quadratic terms are introduced in order to identify "sweet spots". Textual dummies are included in order to identify words that enhance investor appeal. The Linktest misspecification testing procedure confirms the validity of our final model.

Keywords: Equity Crowdfunding; Truncated data; Truncated regression; Hausman test; Link test.
JEL: C24;O3

Word count: 3,972
Declaration: The authors have no financial or proprietary interests in any material discussed in this article.

* *Email*: p.xma@outlook.com
† Corresponding author; *Email*: p.moffatt@uea.ac.uk; ORCID: 0000-0002-3825-0537
‡ *Email*: simon.peters@manchester.ac.uk; ORCID: 0000-0002-7317-9182

1 Introduction

The popularity of investing by individuals has shown a marked increase in recent decades and the internet has played an important role in this increase. In the very recent past, online-based equity crowdfunding (ECF) has become one of the most popular vehicles for the raising of capital for start-up companies, and been instrumental in the "democratization" of entrepreneural finance. At the same time, ECF has emerged as a popular research area within the broader sphere of entrepreneurship. For excellent reviews of the literature, see Abuo Habla & Broby (2019) and Mochkabadi & Volkmann (2020).

From the perspective of the companies attempting to raise finance through crowdfunding, it is clearly very useful for them to be aware of the requirements for successful fundraising (Ahlers et al. 2015). For this reason, a high proportion of the equity crowdfunding literature is concerned with the factors affecting investors' behaviour and campaign success, and the fine-tuning of campaign strategies (Johan & Zhang 2021).

The present paper contributes to this literature by addressing the following econometric problem that has not, to our knowledge, been previously addressed. The data analysed are collected from Seedrs,[1] one of the UK's largest crowdfunding platforms. Each observation in the data set is a crowdfunding campaign, characterised by a "target amount", and a "pitch" providing a description of the venture to potential investors. The dependent variable in our model is the logarithm of the ratio of "amount raised" to target, and we will henceforth refer to this measure as the "level of success" of the campaign. The econometric problem which is the focus of the paper is that, because only successful campaigns appear on the platform, the data on successfulness is truncated from below.[2] It is well-known (Hausman & Wise 1977, Maddala 1983) that applying linear regression techniques to truncated data results in inconsistent estimation, and the truncated regression estimator is required. In addition to applying this estimator, we use the Hausman testing procedure (Hausman 1978) as a formal test for the presence of truncation bias in the linear regression estimates.

A further econometric problem that we face is that some companies are observed with multiple campaigns, and within-company dependence in the data is likely. To deal with this issue, we adjust the standard errors for clustering at the company-level.

[1] https://www.seedrs.com

[2] It is evident that many previous researchers have succeeded in obtaining data sets containing both successful and unsuccessful campaigns (see e.g. Ahlers et al. (2015)). However, it appears that the policy of certain crowdfunding platforms (e.g. Seedrs, Crowdcube) changed recently, to the effect that only successful campaigns now appear on the website, and no information on unsuccessful campaigns is available. This was confirmed in an email exchange with Seedrs management on 1 December 2020.

We make a number of other contributions. To each estimated model, we apply the link test (Pregibon 1980) as a test for general forms of misspecification. In our final specification, we include quadratic terms of some explanatory variables (e.g. target amount, team size), and this enables us to obtain an estimate of the optimal value of each variable, that is, the value of each variable that is required to maximise the predicted level of success of the campaign. We will refer to the optimal value as the "sweet spot". Finally, we investigate the importance of the textual content of the campaign pitches. In particular, we set out to identify words whose appearance in the pitch either increases or decreases the amount raised by the campaign. Textual analysis has been used previously, in order to identify the linguistic styles (Kaminski & Hopp 2020) or semantic framing (Su et al. 2024) that predict campaign success. To our knowledge, this is the first study that sets out to identify particular words that predict success.

The remainder of the paper is organised as follows. Section 2 provides a review of related literature. Section 3 describes the data set used in the analysis. Section 4 outlines the econometric approach taken. Section 5 presents results. Section 6 concludes.

2 Related Literature

The literature on crowdfunding has grown rapidly in the last decade. Mochkabadi & Volkmann (2020) recently provided a systematic review of the literature on equity crowdfunding. Even more recently, Deng et al. (2022) provided a more focused survey of articles identifying the determinants of crowdfunding success. Of the 94 articles they survey, 57 use a binary measure of funding success, while 22 use the "success ratio" (amount raised divided by target).

One study using a binary indicator of success is Ralcheva & Roosenboom (2020), who analyse a data set consisting of 2171 equity crowdfunding campaigns launched between 2012 and 2017 on Crowdcube and Seedrs. 44% of the campaigns were successful.[3] Using binary logistic regression, they find that target amount, equity ratio and age of company all have a negative effect on the probability of success. The negative effect of equity ratio has been found by many others including Ahlers et al. (2015), Vismara (2016) and Rossi et al. (2020). Ahlers et al. (2015) is another example of a study in which a binary indicator of success has been used, and another of their findings is that companies who provide more detailed information about risks achieve a higher probability of success.

One study using the success ratio as the dependent variable and applying OLS is Giga

[3]It appears that failed campaigns were observable on these platforms before 2017.

(2018). The focus of this study is the effect of the number of founders with contributed capital, and the key finding is that more funds are raised when this number is two or more, suggesting a signalling mechanism to investors, of founders' cross-validation of the project's prospects. Vismara (2016) considers two continuous measures of success: number of investors and funding amount. Key results from this study are that both measures depend negatively on the percentage of equity offered, and positively on social capital, the latter being measured using the number of Linkedin connections of the proponent. Other studies reporting the significance of the number of social network connections of entrepreneurs include Ahlers et al. (2015), Lukkarinen et al. (2016), and Nitani & Riding (2017).

It is clear from the above that a wide range of explanatory variables have been used in models of crowdfunding success, and the list goes on. Another group of variables come under the heading of human capital, for example team size (Li 2016, Giga 2017), education, industry, and entrepreneurial experience (Nitani & Riding 2017, Piva & Rossi-Lamastra 2018), age of entrepreneur (Seigner et al. 2024), trustworthiness-in-appearance of entrepreneur (Duan et al. 2020), and gender of team members (Cicchiello & Kazemikhasragh 2022, Kleinert & Mochkabadi 2021). Yet another set of variables is concerned with the characteristics of the pitch. For example, the length of business descriptions has been found to have a positive effect on fundraising success (Dority et al. 2021, Johan & Zhang 2021). Anglin et al. (2018) find that "narcissistic rhetoric" in campaign pitches is useful up to a point, but appears to lower successfulness when used to excess.

As mentioned in Section 1, a key theme of this paper is the truncation in crowdfunding success data, resulting from the removal of unsuccessful campaigns by the platform. To our knowledge, Ma (2023) is the only researcher to have allowed for truncation in crowdfunding data. It must be acknowledged that some authors have used data collection techniques that could circumvent this problem. For example, Vismara (2016) and Nevin et al. (2017), tracked data from the platforms in real time, between the launch of the campaign and the closing date, resulting in data sets containing unsuccessful as well as successful campaigns.

3 Data

The source of the data used in this study is the established Equity Crowdfunding platform, Seedrs, which has been in existence since 2011.[4] The data used in this study was

[4]The first equity crowdfunding platform to appear on the UK funding scene was CrowdCube, earlier in 2011.

extracted from the Seedrs website[5] in November 2023. The criterion for inclusion in the sample is being a funded company who as active in the "secondary market" at that time. There are 771 such companies, between which there are 1,189 campaigns, implying that the average number of campaigns per company is 1.54. In the scraping process, the following information was extracted from each individual campaign: target amount; raised amount; percentage of equity offered; and number of team members. Descriptive statistics of all of these variables are presented in Table 1.

Variable	#Obs	Mean	Std.Dev.	Min	Max
Target Amount(£)	1,189	576,592	1,111,423	688	14,700,000
Raised Amount(£)	1,189	798,432.7	1,293,332	813	15,000,000
Success Ratio	1,189	1.562	0.784	1.0001	8.291
log(Success Ratio)	1,189	0.368	0.361	0.0001	2.115
Equity Offered(%)	1,189	11.136	7.849	0.06	95.15
# Team members	1,189	6.334	3.914	1	23

Table 1: Summary Statistics for Quantitative Variables; "success ratio" is raised amount divided by target amount.

Other independent variables are also created from information scraped from the campaigns. Descriptive statistics for these variables are provided in Appendix A. These are a set of 16 sector dummies indicating the sector of each company, and a set of dummies indicating the presence of certain words in the campaign pitches.

It is noted in Table 1 that the variable "success Ratio" - defined as amount raised divided by target - has a minimum slightly above 1. This is because the secondary market, from whence the data were collected, only contains companies whose campaigns were successful. Hence the data on "Success Ratio" is truncated from below at 1. The dependent variable in our econometric models will be the log of the success ratio, which is, for the same reason, truncated from below at zero. A histogram of this variable is presented in Figure 1. The distribution of this variable is seen to be uni-modal, with a clear mode near to zero. This is fully consistent with the distribution being truncated from below at zero, in the sense that the distribution resembles the upper tail of a normal distribution. It is also noted from Figure 1 that, even after taking the log-transformation of the success ratio, the distribution has a long tail to the right. In the process of data cleaning, 20 observations with success ratio above 10.0 were removed from the data, on the grounds that these ratios were considered to be implausibly high. All of the figures reported in Table 1 were obtained following data cleaning.

[5] https://www.seedrs.com. The data-scraping algorithm was written in Python, and the Python code is available from the authors on request.

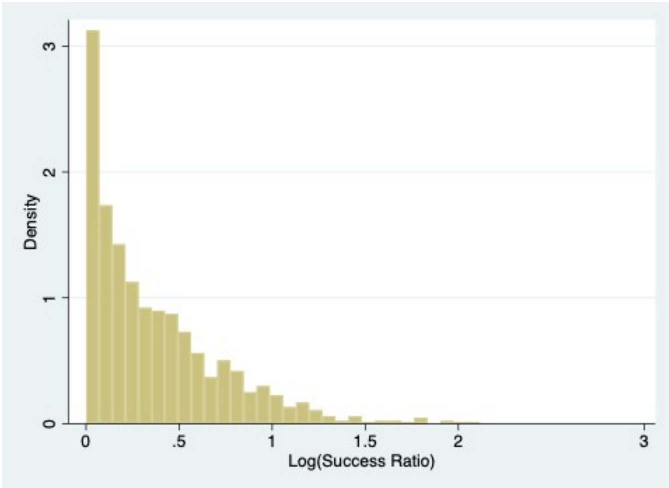

Figure 1: Histogram of $log\,(SuccessRatio)$

The relationships between log(success ratio) and each of the quantitative independent variables are shown in Figure 2. These relationships are presented as scatterplots with Lowess smoothers (Cleveland 1979) superimposed. The smoothers are useful in Figure 2 because some of them strongly suggest the presence of non-monotonic effects, particularly that of Target. For this reason, we include quadratic terms of this and other independent variables in our model, and then use the coefficients to deduce "sweet spots", that is, the values of the variables that maximise predicted level of success. In the case of Team Number, in addition to an inverted-u shape, we see a slight uptick when Team Number equals 1. To allow for this, we will include a dummy variable for single-entrepreneur campaigns.

One variable that appears to have a monotonic effect on the evidence of Figure 2 is the percentage of equity offered. What is striking about this plot is that the effect appears to be positive, and this is confirmed in the results reported in Section 5. This is surprising because it contradicts much previous research which finds that equity offered has a negative effect on crowdfunding success. In Section 5, we suggest possible reasons for this contrary finding.

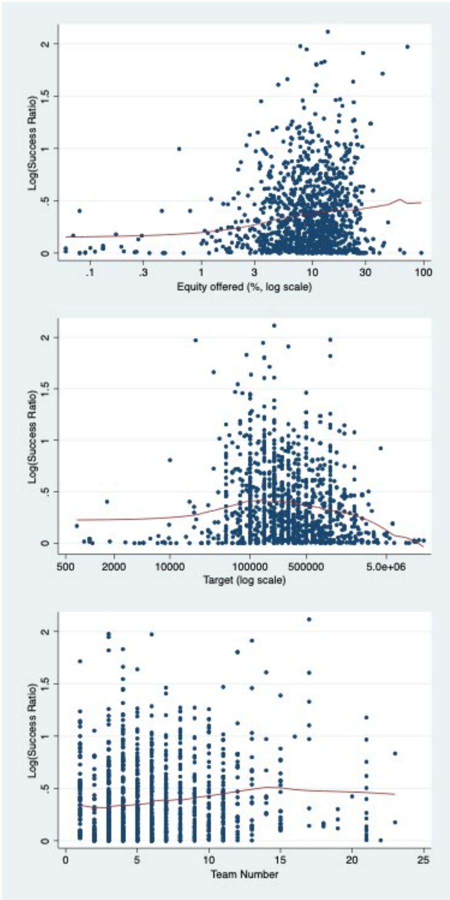

Figure 2: Relationship between $log\,(SuccessRatio)$ and Independent Variables

4 The Truncated Regression Model

In this section, we outline the truncated regression model (Hausman & Wise 1977), which is suitable for modelling truncated data of the type described in Section 3.

4.1 Model and Notation

Equity crowdfunding campaigns will be indexed i, with $i = 1, \ldots, n$. Let y_i^{target} be the target in campaign i, and let y_i be the total amount raised in campaign i. The level of success of a campaign may be measured using the "success ratio" $S_i \equiv y_i/y_i^{target}$. The dependent variable in our analysis will be $s_i \equiv \ln S_i$.[6]

We assume that this measure of success s_i depends linearly on a set of k independent variables contained in the $k \times 1$ vector x_i. The first element of x_i is 1.

$$\begin{aligned} s_i &= x_i'\beta + \epsilon_i \quad i = 1, ..., n \\ \epsilon_i &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{1}$$

β is a $k \times 1$ vector of parameters, the first of which is an intercept. ϵ_i is the equation error, assumed to be normally distributed with constant variance.

4.2 Estimation

Note that $S_i \geq 1$ (or $s_i \geq 0$) indicates a successful campaign, while $S_i < 1$ (or $s_i < 0$) indicates an unsuccessful one. The key feature of our data set is that only successful campaigns are observed. Hence, the dependent variable s_i is lower-truncated at zero.

An important point is that the level of success of the campaign and the process determining whether the observation is truncated are both given by Eq. (1). This is in contrast to a situation of "incidental truncation" (Wooldridge 1996), in which a separate auxiliary "selection equation" would be required to capture the truncation process.

If s_i were fully observed, we would proceed with estimation of model (1) using standard linear regression. However, since s_i is only observed if $s_i \geq 0$, the truncated regression model is required. Under this model, the likelihood contribution associated with campaign i is:

$$L_i\left(\beta, \sigma_\epsilon\right) = \frac{\frac{1}{\sigma_\epsilon}\phi\left(\frac{s_i - x_i'\beta}{\sigma_\epsilon}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma_\epsilon}\right)} \quad s_i \geq 0 \tag{2}$$

where $\phi(.)$ and $\Phi(.)$ are respectively the density and distribution functions of the standard normal. Maximum likelihood estimates of the parameters, β and σ_ϵ, are obtained by maximising the sample log-likelihood:

$$\begin{aligned} LogL\left(\beta, \sigma_\epsilon\right) &\equiv \sum_{i=1}^{n} \ln L_i\left(\beta, \sigma_\epsilon\right) \\ \left(\hat{\beta}, \hat{\sigma}_\epsilon\right) &= \underset{\beta, \sigma_\epsilon}{\operatorname{argmax}}\, LogL\left(\beta, \sigma_\epsilon\right) \end{aligned} \tag{3}$$

[6]The success ratio (S) has been used as a dependent variable by other authors, for example Giga (2018). We are not aware of any previous research that uses s, the log of the success ratio.

Equations (2) and (3) define the truncated regression model developed by Hausman & Wise (1977). Estimation of this model, with standard errors adjusted for clustering at the company level, is possible in econometric software packages.[7]

4.3 A Test for Truncation Bias

An important question is how serious are the consequences of ignoring the truncation in the data and proceeding with estimation on the assumption that the data is not truncated. This question can be addressed using the Hausman test (Hausman 1978). Aigner & Hausman (1980) have applied the Hausman test to test for truncation bias in the context of the model outlined in Section 4.2. In this section, we outline this testing procedure.

Let $\hat{\beta}_{ols}$ be the estimate of the vector β obtained by applying the OLS estimator to (1), and let $\hat{\beta}_{trunc}$ be that obtained by applying the truncated regression estimator defined in (2) and (3). If we also obtain estimates of the variance matrices of the two estimates, $\hat{V}_{ols}$ and $\hat{V}_{trunc}$ respectively, then the Hausman test statistic is given by:

$$H = \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right)' \left(\hat{V}_{trunc} - \hat{V}_{ols}\right)^{-1} \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right) \tag{4}$$

and $H \sim \chi^2(k)$ under the null hypothesis of no truncation bias, where k is the dimensionality of β. Hence, if the computed value of H is greater than the critical value $\chi^2_{k,0.05}$, we may conclude that the two estimates are significantly different and that the estimate obtained from standard OLS is inconsistent as a result of truncation bias.

When estimation is performed with standard errors adjusted for clustering at the company level, the formula for the Hausman Test is slightly different (Cameron et al. 2010). Let $\tilde{V}_{ols}$ and $\tilde{V}_{trunc}$ be the estimated variances of the OLS and truncated regression estimators obtained assuming clustering. Then the clustered version of the Hausman test is:

$$\tilde{H} = \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right)' \left(\tilde{V}_{trunc} - \tilde{V}_{ols}\right)^{-1} \left(\hat{\beta}_{trunc} - \hat{\beta}_{ols}\right) \tag{5}$$

Eq. 5 is the version of the Hausman test that will be used in Section 5.

4.4 Misspecification Test

As a misspecification test, the link test will be used. The link test (Pregibon 1980) is a version of the well-known RESET test (Ramsey 1969). The usefulness of the link test in

[7] For example, the **truncreg** command in STATA (StataCorp 2021) estimates the model defined in (2). The option **cluster(.)** provides standard errors adjusted for clustering.

micro-econometric models has been investigated by Peters (2000).[8]

The link test is performed in two stages. The first stage is to estimate the model under test with the set of independent variables contained in the vector x_i and then to generate the linear predictor of the dependent variable s_i:

$$\hat{s}_i = x_i'\hat{\beta} \tag{6}$$

The second stage is to estimate the model again, but using the two variables $\hat{s}_i$ and $\hat{s}_i^2$ as independent variables in place of the variables contained in x_i. This model should also include an intercept. If the first model was estimated using clustered standard errors, the second model should also be clustered at the same level. The link test statistic is the (asymptotic) t-test for testing the significance of $\hat{s}_i^2$ in this second model. If $\hat{s}_i^2$ shows significance, this indicates that the first model is misspecified in some way.

Peters (2000) has demonstrated that the link test can be used as an "omnibus test". Hence, when it rejects, it simply indicates that there is some sort of misspecification. This could be in the form of a missing independent variable, or an incorrect distributional assumption, or a failure to account for a data feature such as truncation.

5 Results

Five sets of results are presented in Table 2. The set of explanatory variables used has been selected via a general-to-specific model selection procedure. The four estimated models are: OLS with all explanatory variables; Truncated Regression with only linear terms in the quantitative variables; Truncated Regression without sector dummies; Truncated Regression without word dummies; Truncated regression with all explanatory variables. For all five models, standard errors are clustered at the company level.

The truncated regression model whose results appear in the final column nests the other three truncated regression models. Hence the restrictions embodied in the nested models can be tested using likelihood ratio (LR) tests. All of these restrictions are strongly rejected, confirming the considerable importance of the three groups of variables: the non-linear terms; the sector dummies; and the word dummies. The models can also be compared using Akaike's Information Criterion (AIC). This is a model selection criterion that adjusts for the number of parameters being estimated. The preferred model is the one with the lowest AIC. On this criterion, it is confirmed that the model appearing in

[8] In standard models, including truncated regression, the test can be applied easily using the **linktest** command in STATA (StataCorp 2021).

Table 2: Estimation Results. Dependent variable: log of success ratio. Truncated regression models estimated with lower truncation point zero. Asymptotic Standard errors clustered at the company level shown in parentheses. Hausman Test shown for each truncated model is for testing for truncation bias in a model that does not allow for truncation. The Hausman test statistic is $\tilde{H}$ defined in (5). AIC is a goodness-of-fit measure that adjusts for the number of parameters being estimated. The best model is the one with the lowest AIC.

VARIABLES	OLS	Truncated Regression	Truncated Regression	Truncated Regression	Truncated Regression
log (target)	0.293 ***	-0.463***	6.271***	6.637***	5.290***
	(0.0533)	(0.109)	(2.012)	(1.784)	(1.479)
log (target)^2	-0.0148***	-	-0.271***	-0.288***	-0.231***
	(0.00216)		(0.0829)	(0.0736)	(0.0607)
log (equity)	0.104***	0.807***	0.620***	0.717***	0.577***
	(0.0186)	(0.193)	(0.121)	(0.165)	(0.144)
I (TeamNumber = 1)	0.181***	-	1.458***	1.331***	1.0797***
	(0.0546)		(0.422)	(0.384)	(0.317)
TeamNumber	0.0302***	0.0958***	0.259***	0.236***	0.189***
	(0.0102)	(0.0239)	(0.0781)	(0.0750)	(0.0614)
TeamNumber^2	-0.000693	-	-0.00716**	-0.00666*	-0.00487*
	(0.000538)		(0.00341)	(0.00347)	(0.00284)
Automotive & Transport Sector	0.108*	0.672*	-	0.595*	0.623**
	(0.0626)	(0.403)		(0.352)	(0.310)
Data & Analytics Sector	-0.110**	-2.599	-	-2.137	-1.815
	(0.0463)	(1.797)		(1.479)	(1.266)
Energy Sector	0.122*	0.916**	-	0.649	0.706**
	(0.0664)	(0.425)		(0.397)	(0.340)
Finance & Payments Sector	0.166***	1.0126***	-	0.925***	0.882***
	(0.0418)	(0.297)		(0.264)	(0.234)
"health/healthy"	0.0943***	0.665***	0.399**	-	0.488***
	(0.0271)	(0.190)	(0.163)		(0.138)
"organic"	0.0486	0.290	0.350*	-	0.293*
	(0.0304)	(0.183)	(0.191)		(0.154)
"data"	0.0333	0.345**	0.261	-	0.202
	(0.0251)	(0.164)	(0.165)		(0.131)
"entertainment"	-0.125***	-1.242*	-1.488**	-	-1.0494*
	(0.0450)	(0.711)	(0.731)		(0.559)
"information"	-0.0574**	-0.724***	-0.514**	-	-0.402**
	(0.0266)	(0.245)	(0.234)		(0.185)
Constant	-1.4096 ***	1.485*	-40.119***	-42.024***	-33.479***
	(0.318)	(0.830)	(12.810)	(11.271)	(9.353)
$LogL$	-	90.84	100.83	107.01	124.99
$AIC (= 2k - 2LogL)$	-	-155.69	-177.67	-192.01	-217.98
Hausman Test Statistic	-	27.48**	24.62**	28.60***	39.19***
p-value	-	(0.011)	(0.010)	(0.003)	(0.001)
Link Test Statistic	3.23***	-0.92	-1.62	-1.14	-1.51
p-value	(0.001)	(0.356)	(0.105)	(0.255)	(0.130)
n	1189	1189	1189	1189	1189

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

the final column is the preferred model, even allowing for its higher number of parameters.

The next thing to note is that the truncated regression results (final column) differ quite markedly from the OLS results (first column), underlining the importance of finding the "correct" model. In particular, note that the OLS slope estimates tend to be much lower in magnitude than the corresponding truncated regression estimates. This bears out the general result that disregarding truncation leads to a bias towards zero in the slope estimates (Maddala 1983, Greene 2003). This difference is tested formally using the Hausman test, outlined in Section 4.3. For each Truncated Regression model, a Hausman test is reported. This test compares the set of estimates with the corresponding set of estimates from an OLS model with the same specification. In all cases we see evidence of truncation bias in the model which fails to account for truncation. In the case of the full model (final column), this evidence is particularly strong ($p = 0.001$).

The other test reported in Table 2 is the link test, outlined in Section 4.4. On the basis of this test, there appears to be strong evidence that the OLS model is misspecified (p=0.001), while there is no evidence of misspecification in the truncated regression model with the same set of independent variables (p=0.130). This difference suggests that the misspecification being detected for the OLS model is in the form of the failure to allow for truncation, rather than a misspecification of the regression function.

We next turn to the interpretation of the coefficients. For this purpose we focus on the final column of Table 2. The coefficients of the textual dummies reveal that the presence of the words "health/healthy" and "organic" in the campaign's pitch both have a significantly positive effect on the level of success. These results appear to underline the importance that investors attach to the development of healthy lifestyles and the environmental sustainability of the planet. It also appears that the presence of "entertainment" and "information" both appear to have a negative effect. Of course, the presence of certain words may take the role of a proxy for the type of product being produced by the company. However, we also note that the word dummies show significance even though sector dummies are also present.

Of the 17 sector dummies that we started with,[9] only four show significant effects and are included in the final specification: "Automotive & Transport", "Energy" and "Finance & Payments" sectors appear to have higher levels of success than the excluded sectors; the "Data & Analytics" sector appears to have a lower level of success.

Turning next to the quantitative explanatory variables, we first see that the log of the

[9] See Appendix A.

equity ratio appears to have a significantly positive effect on the level of success of a campaign. This appears to contradict previous literature. For example Vismara (2016) suggests that entrepreneurs' willingness to invest in their own project signals quality, and this signal is important because the entrepreneur has more information than the investor on the quality of the project. If investors pay attention to this signal, then the equity ratio (proportion of equity offered) is predicted to have a negative effect on the level of success. We suggest the following explanations for the positive effect that we have found. If the proportion of equity offered is very low, this might be perceived as a signal that the external investors are not regarded as important to the success of the project, and this might put investors off. Investors want to feel that they are making an important contribution. Another possibility is that a low equity ratio signals that the company's pre-money valuation is exaggerated, and this might make investors wary.[10]

For both log(target) and team number, motivated by the inverted-U shaped curves seen in Figure 2 above, we include quadratic terms.[11] In each case, the sign of coefficient on the quadratic term is negative, confirming the inverted u-shaped effect. Using the coefficients of the linear and quadratic terms, it is a simple matter to to deduce the optimal value (or "sweet spot") of each independent variable.[12] These sweet spot estimates are presented in Table 3. Noticeable differences are seen in the sweet spots when they are obtained from different models. For example, based on the OLS model (ignoring truncation) the sweet spot of target amount is £20,615, while for the truncated regression model, it is more than four times higher, at £93,655.

The sweet spot for the number of team members appears to be around 19. Notice from Table 2 that we have also included a dummy variable for campaigns with a single team member, and this appears to have a positive effect (again consistent with the plot shown in Figure 2 above). There is a theoretical explanation for the inverted-u shaped effect of team number. Hornuf & Schmitt (2017) have hypothesized as follows: "On the one hand, starting a business alone can be difficult and cumbersome because of lack of competences and capacity constraints. On the other hand, the larger the management team of the start-up becomes, the more likely are disputes among management team members to arise". Our results appear to support both of these hypothesis because the quadratic relationship implies that as team size increases, level of success first increases and then decreases. However, it is slightly more complicated than this because level of success is

[10] Pre-money valuation is essentially self-reported by the company. The Seedrs platform provides a "pre-money valuation calculator", and simply requests that the company is honest with their answers.

[11] Quadratic specifications in crowdfunding research have previously been usefully applied by Anglin et al. (2018).

[12] The "sweet spot" of a variable is computed by dividing the coefficient of the linear term by two times the coefficient of the quadratic term, and reversing the sign. The confidence interval for the sweet spot is obtained using the delta method (Oehlert 1992).

boosted when the campaign has just one team member.[13]

Variables	OLS	Truncated Regression
Target Amount(£)	20,615	93,655
	[1,270 , 39,960]	[39,513 , 147,797]
Number of Team Members	21.8	19.3
	[1.3 , 42.3]	[8.2 , 30.6]

Table 3: Point Estimates of "sweet spots" (interval estimates in square brackets; obtained using delta method, with cluster-robust variance matrix.)

6 Conclusion

In the past decade or so, crowdfunding has rapidly gained importance as a method for firms to raise funds. Over the same period, there has been a rapid increase in the volume of research carried out with the objective of identifying the determinants of crowdfunding success. There is therefore no doubt that this is an interesting and topical research area, that is growing in importance.

In this paper, we have made a number of new contributions to the literature on Equity Crowdfunding. Most importantly, we have taken account of the truncation in the data that arises because only successful campaigns are observed in the secondary market. The allowance for truncation has been seen to lead to very different results from a model that disregards truncation, and these differences have been confirmed formally using the Hausman test. Further evidence of the superiority of the truncated regression model over OLS has been obtained using the omnibus misspecification test, the link test. Second, since our data set contains multiple observations per company, we have allowed for company-level clustering in estimation. Third, we have applied the truncated regression estimator to a model specification that includes quadratic terms in some continuous explanatory variables. This has enabled us to deduce optimal values (or "sweet spots") for each of these variables, that is, values predicted to maximise the level of success of a campaign. Use of quadratic terms to pursue similar goals has been made by Anglin et al. (2018) who found an inverted-U effect of "narcissistic rhetoric" on crowdfunding performance. Information relating to "sweet spots" clearly has the potential to be very useful to entrepreneurs setting out on equity crowdfunding campaigns. Furthermore, we have found that the sweet-spots obtained from the truncated regression estimates can be very different to those from OLS, and this underlines the importance of allowing for truncation.

[13]It can easily be verified using the results in Table 2 that predicted level of success is higher when team size is at the sweet spot of 19, than when it is at 1.

Finally, we have investigated the importance of the presence of particular words in crowdfunding announcements. We have found that the presence of certain words do indeed have a significant effect on the level of success, some positive and some negative. This is a simple form of textual analysis. The interesting findings obtained here suggest that the use of more sophisticated methods of textual analysis in predicting crowdfunding success is a promising area for further research.

Based on the estimation results we have obtained, we are able to conclude that: a high equity ratio is beneficial; the optimal target is around £94,000; the optimal team size is around 19; and the presence in the pitch announcement of the words "healthy" and "organic" appear to improve the predicted level of success of the campaign.

Appendix A

Variable	#Obs	Proportion of sample (%)
Advertising & Marketing	1,189	1.5
Automotive & Transport	1,189	4.6
Clothing & Accessories	1,189	4.5
Content & Information	1,189	2.4
Data & Analytics	1,189	1.7
Energy	1,189	3.5
Entertainment	1,189	3.5
Finance & Payments	1,189	15.9
Food & Beverage	1,189	21.8
Games	1,189	0.8
Healthcare	1,189	5.1
Home & Personal	1,189	8.5
Programming & Security	1,189	0.3
Property	1,189	5.6
Recruitment & Procurement	1,189	1.4
SaaS/PaaS	1,189	11.4
Travel, Leisure & Sport	1,189	7.7

Table A.1: Summary Statistics for Sector Dummies. Final column shows the sample percentage of campaigns in each sector.

Variable	#Obs	Proportion of sample (%)
health/healthy	1,189	21.5
organic	1,189	14.6
quality	1,189	28.4
planet	1,189	6.1
plant	1,189	7.9
investment	1,189	67.2
data	1,189	29.6
analytics	1,189	5.6
sugar	1,189	5.7
statistics	1,189	2.3
analysis	1,189	4.6
property	1,189	9.9
home	1,189	23.7
personal	1,189	22.0
travel	1,189	10.3
leisure	1,189	2.1
sport	1,189	10.4
recruitment	1,189	4.4
procurement	1,189	1.7
healthcare	1,189	2.0
food	1,189	18.7
beverage	1,189	2.2
finance	1,189	15.4
entertainment	1,189	2.4
energy	1,189	7.5
information	1,189	25.2

Table A.2: Summary Statistics for Word Dummies. Final column shows percentage of campaigns for which the word appears in the pitch.

References

Abuo Habla, Z. & Broby, D. (2019), 'Equity crowdfunding: literature review', *University of Strathclyde Working Paper* pp. 1–19.

Ahlers, G. K., Cumming, D., Günther, C. & Schweizer, D. (2015), 'Signaling in equity crowdfunding', *Entrepreneurship theory and practice* **39**(4), 955–980.

Aigner, D. J. & Hausman, J. A. (1980), 'Correcting for truncation bias in the analysis of experiments in time-of-day pricing of electricity', *The Bell Journal of Economics* pp. 131–142.

Anglin, A. H., Wolfe, M. T., Short, J. C., McKenny, A. F. & Pidduck, R. J. (2018), 'Narcissistic rhetoric and crowdfunding performance: A social role theory perspective', *Journal of Business Venturing* **33**(6), 780–812.

Cameron, A. C., Trivedi, P. K. et al. (2010), *Microeconometrics using stata*, Vol. 2, Stata press College Station, TX.

Cicchiello, A. F. F. & Kazemikhasragh, A. (2022), 'Tackling gender bias in equity crowdfunding: an exploratory study of investment behaviour of latin american investors', *European Business Review* .

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American statistical association* **74**(368), 829–836.

Deng, L., Ye, Q., Xu, D., Sun, W. & Jiang, G. (2022), 'A literature review and integrated framework for the determinants of crowdfunding success', *Financial Innovation* **8**(1), 41.

Dority, B., Borchers, S. J. & Hayes, S. K. (2021), 'Equity crowdfunding: Us title ii offerings using sentiment analysis', *Studies in Economics and Finance* .

Duan, Y., Hsieh, T.-S., Wang, R. R. & Wang, Z. (2020), 'Entrepreneurs' facial trustworthiness, gender, and crowdfunding success', *Journal of Corporate Finance* **64**, 101693.

Giga, A. (2017), 'Firm financing through equity crowdfunding'. SSRN working paper. **URL:** *https://ssrn.com/abstract=2995592*

Giga, A. (2018), 'Founders' commitment and firm financing: Multiple'skins in the game'as a signal to investors', *Available at SSRN 3290378* .

Greene, W. H. (2003), *Econometric analysis*, Pearson Education India.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica: Journal of the econometric society* pp. 1251–1271.

Hausman, J. A. & Wise, D. A. (1977), 'Social experimentation, truncated distributions, and efficient estimation', *Econometrica: Journal of the Econometric Society* pp. 919–938.

Hornuf, L. & Schmitt, M. (2017), Equity crowdfunding in germany and the uk: Follow-up funding and firm survival, Technical report, CESifo working paper.

Johan, S. & Zhang, Y. (2021), 'Investors' industry preference in equity crowdfunding', *The Journal of Technology Transfer* pp. 1–29.

Kaminski, J. C. & Hopp, C. (2020), 'Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals', *Small Business Economics* **55**(3), 627–649.

Kleinert, S. & Mochkabadi, K. (2021), 'Gender stereotypes in equity crowdfunding: the effect of gender bias on the interpretation of quality signals', *The Journal of Technology Transfer* pp. 1–22.

Li, J. (2016), 'Equity crowdfunding in china: Current practice and important legal issues', *Asian Bus. Law.* **18**, 59.

Lukkarinen, A., Teich, J. E., Wallenius, H. & Wallenius, J. (2016), 'Success drivers of online equity crowdfunding campaigns', *Decision Support Systems* **87**, 26–38.

Ma, X. (2023), Econometrics of cyber-investment: Analysis of cryptocurrencies and equity-crowdfunding, PhD thesis, University of East Anglia.
URL: *https://ueaeprints.uea.ac.uk/id/eprint/93537/*

Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometrics*, number 3, Cambridge university press.

Mochkabadi, K. & Volkmann, C. K. (2020), 'Equity crowdfunding: a systematic review of the literature', *Small Business Economics* **54**(1), 75–118.

Nevin, S., Gleasure, R., O'Reilly, P., Feller, J., Li, S. & Christoforo, J. (2017), Large crowds or large investments?: How social identity influences the commitment of the crowd, *in* 'ECIS 2017 Proceedings', Association for Information Systems. AIS Electronic Library (AISeL), pp. 2802–2813.

Nitani, M. & Riding, A. (2017), On crowdfunding success: Firm and owner attributes and social networking, *in* '2017 Emerging Trends in Entrepreneurial Finance Conference'.

Oehlert, G. W. (1992), 'A note on the delta method', *The American Statistician* **46**(1), 27–29.

Peters, S. (2000), 'On the use of the reset test in microeconometric models', *Applied Economics Letters* **7**(6), 361–365.

Piva, E. & Rossi-Lamastra, C. (2018), 'Human capital signals and entrepreneurs' success in equity crowdfunding', *Small Business Economics* **51**(3), 667–686.

Pregibon, D. (1980), 'Goodness of link tests for generalized linear models', *Journal of the Royal Statistical Society Series C: Applied Statistics* **29**(1), 15–24.

Ralcheva, A. & Roosenboom, P. (2020), 'Forecasting success in equity crowdfunding', *Small Business Economics* **55**(1), 39–56.

Ramsey, J. B. (1969), 'Tests for specification errors in classical linear least-squares regression analysis', *Journal of the Royal Statistical Society: Series B (Methodological)* **31**(2), 350–371.

Rossi, A., Vanacker, T. R. & Vismara, S. (2020), 'Equity crowdfunding: New evidence from us and uk markets', *Available at SSRN* .

Seigner, B. D. C., McKenny, A. F. & Reetz, D. K. (2024), 'Old but gold? examining the effect of age bias in reward-based crowdfunding', *Journal of Business Venturing* **39**(3), 106381.

StataCorp, L. (2021), 'Stata statistical software: release 17 college station', *TX StataCorp LLC* .

Su, L., Sengupta, J., Li, Y. & Chen, F. (2024), '"want" versus "need": how linguistic framing influences responses to crowdfunding appeals', *Journal of Consumer Research* **50**(5), 923–944.

Vismara, S. (2016), 'Equity retention and social network theory in equity crowdfunding', *Small Business Economics* **46**(4), 579–590.

Wooldridge, J. M. (1996), 'Introductory econometrics: A modern approach, 3rd ed.'.