A Matrix Completion Approach to Policy Evaluation: Evaluating the Impact of the VCT Scheme on Investment in the U.K

Dennis Iweze*

April 26, 2021

Abstract

In this study, we estimate the causal effect of the Venture Capital Trust (VCT) scheme on investment (change in total-asset formation) for VCT-funded firms in the U.K. To this end, we hand-collect data on all firms to have ever received VCT funding since inception of the scheme. Thereafter, we develop a framework for estimating causal effects in settings where some firm-years are exposed to a binary treatment (VCT funding) and the goal is to estimate counterfactual outcomes for the VCT-funded (treated) firm-years combinations. This framework is based on a popular unsupervised machine learning task called matrix completion. In tandem with our hand-collected data, we apply this framework to estimate the causal effect of the venture capital trust (VCT) scheme on investment for VCT-funded firms. Our estimand is the Average Treatment Effect on the Treated (ATET). We find that the VCT scheme caused an increase in investment of 60.69% for VCT-funded firms; the ATET is 60.69%. We also document novel insights regarding the relationship between changes to the U.K governments' VCT policy and the investment patterns of VCTs. Finally, we show that our matrix completion estimator outperforms an unconfoundedness-based estimator.

Keywords: Venture Capital Trust, Causal Effects, Hand-Collected Data, Low-Rank Matrix Estimation

*School of Economics and Finance, Queen Mary University of London, e-mail: **d.iweze@qmul.ac.uk**. I am grateful to my supervisor, Jason Sturgess, for his supervision and guidance on this project.

1 Introduction

The Venture Capital Trust (VCT) scheme , introduced in 1995, is one of three tax-based venture capital schemes, the others being the Enterprise Investment Scheme (EIS) and the Seed Enterprise Investment Scheme (SEIS). The VCT scheme is a U.K government policy response to a perceived breakdown in the financial markets and their ability to provide risky capital to risky but promising U.K SMEs. The scheme is designed to encourage investors to invest (indirectly) in British, unquoted, smaller, and higher-risk firms - with a need for start-up, early stage or expansion capital - by investing through subscription to a VCTs shares. VCTs are U.K publicly-quoted and closed-ended funds and the U.K government encourages investment in these financial intermediaries by offering investors tax-rebates, which in turn serves as a subsidy to the cost of equity. In this study, we do two things. First, we develop a new[1] framework for estimating average causal effects in a setting where some firm-years are subject to a binary shock. To estimate the average causal effect of the shock on the shocked firm-years in this framework, we impute the "missing" potential control outcomes. Secondly, we provide in-depth details on VCTs and how changes to the governmental regulations guiding VCT activities has shaped the aggregate patterns seen in the data. We then use our developed framework to estimate the average causal effect of the VCT scheme on investment (growth in total-asset formation) for VCT-backed firms' in the U.K.

Generally, the importance of venture capital (VC) funding for SME's and by extension the economy has been extensively documented. From studies such as: Kaplan and Lerner (2010) who document that even though less than 0.25% of U.S firms receive VC-backing, an estimated one-half of IPOs are VC backed, Metrick and Yasuda (2011) who emphasise the relationship between VC funding, small firms, and innovation, Gornall and Strebulaev (2015) who show that U.S public firms with VC-backing represent one-fifth of the market capitalisation and 44% of the research and development expenditure of U.S. public firms and are an integral component of the

[1] To our knowledge, our study is the second article to employ a Matrix Completion approach to estimating causal effects in Economics. Our Matrix Completion framework and algorithm is different from but conceptually similar to that of Athey et al. (2018) - the first study to employ such an approach. One instance where we differ is that our algorithm also uses the observed values of treated/VCT-funded firms to impute their counterfactual outcomes, whereas theirs does not. The main reason for this is because the setting they focus on is one in which there is only a single treated unit-period, thus there is no pattern in the observed treatment data to exploit, whereas we have many treated unit-periods in our setting.

U.S. economy, and Gompers et al (2020) who document the VC-backed heritage of numerous innovative companies, their effects on the U.S and global economy, and with the aid of survey data - explore how these VCs make decision.

However, even though VCTs are analogous to VCs - the main difference between both primarily centres around the trust status and specific government regulations guiding the former - the specific importance of VCT funding for SMEs and in turn, the wider economy is practically unknown in academia [2]. The closest thing to a rigorous exploration of VCTs and their impact on SMEs is a 2008 HMRC commissioned report written by Cowling et al (2008). Here, with the aid of administrative data from Her Majesty's Revenue and Customs (HMRC), and a fixed-effects and random-effects model, they find a positive correlation between fixed-asset formation/employment and VCT-backing of SMEs. Nonetheless, their conclusion - based on their econometric analysis - is that on average, the VCT scheme[3] has had "little discernible impact on $\cdots$ or investment". Whilst it is useful to review existing literature, the contrast between our approach and that taken in this paper could not be clearer. As such, we do not use any of their results as a baseline, but perhaps our reader might find it useful to keep their conclusion in mind as we detail our approach and present our results.

Our first step in this study is the hand-collection of data on all VCT-backed firms in the U.K (both former and current). Here - with the aid of the hand-collected data - our aim is to quantify the average growth in total-asset formation of VCT-backed firms, as a result of VCT-backing. With this approach to data-collection, we analogously follow the pioneering survey works of Lintner (1956) on dividend policy, Graham and Harvey (2001) on CFO financial policies, and Gompers et al (2020) on VC decision-making.

Despite the numerous academic studies on VCs, which we can in turn relate to VCTs, our study - with the aid of hand-collected VCT data - attempts to provide some understanding of VCTs, VCT-backed firms, and their contribution to the economy. Our hand-collected data on VCT-backed firms is the only such data available, at least to our knowledge. Also, from of our hand-collection efforts, we are able to extract information that allow us to match the aggregate investment policies

[2] The bulk of knowledge on VCTs and their importance to macroeconomic considerations is limited to reports commissioned by various bodies such as: governmental agencies, VCTs themselves, and investment companies and their affiliates.

[3] Their study covers both the VCT and EIS scheme.

of VCTs at each point in time, to the contemporaneous changes in the VCT regulations. This particularly can serve as a template for regulators to enact effective changes to the VCT regulations. For instance, we contemporaneously observe how changes to the age criteria for first-time VCT-backed firms affected the median size of a first-time backed firm. This can inform regulators on what policies to implement if they want to affect the type of firm that receives VCT-funding.

We now summarise our data hand-collection effort. We begin in alphabetical order, we start with Albion Development VCT. We source and gather every semi-annual and annual report it had ever published (e.g. from 1995 - 2018). We painstakingly read through the reports and from each, we collect details on its investee firms: the first time they received VCT funding, categorical information such as the date of incorporation, and industry sector. We worked through the entire list of current and former VCTs'. For each and every VCT, we repeat the same painstaking process of reading through every financial report it ever produced, to gather a list of all firms it invested in, the date of investment, the industry the investee operates in - amongst other categorical data. Armed with a list of all VCT-backed firms - 1931 unique U.K firms, we extract their financial and categorical data from the FAME database. Specifically, the information on each VCT-backed firms annual total-assets feeds into the second phase of our project; the estimation of the effect of VCT-backing on the growth in total-asset formation of VCT-backed firms. This is the Average Treatment Effect on the Treated (ATET). There are two parts to estimating this estimand. The observed total-assets for VCT-backed firms, which we already have, and the missing potential control total-assets, which we obviously do not have, and is the focus of the second part of our study.

The literature on causal inference has several approaches to the problem of imputing the "missing" potential control outcomes. For instance, Imbens and Rubin (2015) take an unconfoundedness approach to the problem. This approach is akin to imputing the "missing" potential control outcomes for treated units with the observed outcomes for control units - which are units that share similar pre-treatment outcome values with the treated units. Another approach is the synthetic control approach employed in studies such as Doudchenko and Imbens (2016), Ben-Michael et al (2018). This approach is akin to imputing the saymissing potential control outcomes for treated units with weighted average outcomes for control units. Here, the weights are constructed such that the weighted lagged control outcomes are equal to the lagged outcomes for treated units.

Whilst seemingly similar, both approaches have very salient differences. They especially differ in the data-correlation patterns they exploit to impute the "missing" potential control outcomes. The unconfoundedness approach assumes that the outcomes for the treated and control units follow the same trend in the pre-treatment period. Also, the typical setting in the unconfoundedness approach is one in which the treated units are assumed to be treated all at the same time, in the last period. In contrast, the synthetic control approach assumes that the correlation between outcomes for both control and treated groups are steady over time. Also, the typical setting in the synthetic control approach is one in which there is only one or a few treated units, a significant more control units, and a substantial number of pre-treatment periods. We follow Athey et al (2018) in arguing that, given a particular setting, both approaches are interchangeable - after some regularisation.

In this study, to impute the "missing" potential control total-assets for VCT-backed firms (counterfactual total-assets), we take inspiration from the matrix completion literature. As shown in Athey et al (2018) - which is the only other study to adapt a matrix completion approach to a causal inference problem - our approach significantly differs from but incorporates some of the unconfoundedness and synthetic control features. In our matrix completion approach, given our observed matrix of outcomes for units - which could include both treated and untreated units - we assume that the data for treated units during treatment periods are missing. Our task is to impute the missing entries for treated units in our matrix. The imputed values represent the "missing" potential control outcomes. In other words, we impute the counterfactual total-assets for VCT-backed firms. This approach to imputing missing entries in a matrix assumes that the complete matrix has a low-rank, with said rank then implicitly realised bu regularisation methods - by adding a penaliser to the objective function. Seminal studies in the matrix completion literature include Cai, Candes, and Shen (2008), Candes and Recht (2009), Candes et al (2009), Candes and Plan (2009), Keshavan, Oh, and Montanari (2009).

We make four contributions in this paper. Firstly, we quantify the causal impact of the VCT scheme on total-asset formation (investment) for VCT-backed firms. Our result is the exact opposite of the results in Cowling et al (2008).Although not directly comparable - given that the data in their study spans a different period - the differences in our conclusions stems from the difference in approaches. In contrast to Cowling et al (2008), we find that the VCT scheme has had a very discernible, in-fact significant, effect on investment. It led to a 60.69% increase in investment for VCT-funded firms between 2003-2018. In our second contribution, we lay out our simple matrix completion algorithm (based on a Bregman algorithm), adapted to a setting where VCT-funded firms receive VCT funding at different times. In other words, treated units adopt the treatment at different times (staggered adoption). Our algorithm is efficient and converges at a rate of $1/K$, where K denotes the number of iterations.

Thirdly, our hand-collected data on all VCT-funded firms is the only available one of its kind, at least to our knowledge. We know this because other than searching and unable to find this data, we asked HMRC for the data, and they claimed to not keep such records. Seeing as they are only body with the reach to collect such information without necessarily going through our hand-collection approach, we are confident in our assertion.

Finally, with our hand-collected data, we provide a rich set of observations about VCT-funded firms. For instance, we show that the survival rate of these VCT-funded firms is higher than our understanding of the survival rate of small firms. We show also that there is a direct and very significant relationship between changes in government VCT regulations and changes in the investment patterns of VCTs.

The remainder of this study is organised as follows. Section 1.2 provides the framework for and presents our matrix completion estimator. In this section, we also show that our estimator has a closed form solution (a proximal map). In section 1.3, we derive our Bregman based algorithm for the numerical solution of our matrix completion problem. In section 1.4, we present detailed information on VCTs, the regulations guiding them, and examples of the state sponsored benefits that accrue to a VCT investor. In section 1.5, we present the data hand-collection process and our hand-collected data. We also present the additional financial data we used in tandem with our hand-collected data and present summary statistics on VCTs and VCT-funded firms. In section 1.6, we present our main results, how major VCT policy changes impacted the aggregate investment patterns of VCT-funded firms, and some additional results. In section 1.7, we summarise and conclude.

2 Matrix Factorisation: Singular Value Decomposition

In this section, we develop the underlying framework for our algorithm to estimate the average causal effect of venture capital trust (VCT) funding on investment in the U.K. The algorithm - which we lay-out in the next sub-section - is called a Matrix Completion algorithm.

The framework is based on a fundamental topic in unsupervised machine learning: the recovery of a low-rank matrix from high-dimensional data, which helps to uncover otherwise hidden information in data. Formally, the framework for low-rank matrix recovery or data dimensionality reduction is known as matrix factorisation. This framework is widely used in far ranging fields - from economics (Athey et al. (2018)), to computer vision (Candes and Plan (2010)), and control theory (Boyd et al. (1994)). It is used to solve many popular machine learning tasks such as matrix completion (Candes and Tao, (2009), Athey et al. (2018)), robust principal component analysis (Wright et al. (2009)), and matrix sensing (Zhong et al. (2015)). This framework is also employed in numerous real-world applications including causal panel data settings (Athey et al. (2018)), and building recommender systems (Koren et al. (2009)).

The idea behind matrix factorisation is that data is given in the form of a matrix Y, and we assume that the true dimensionality of the matrix (for example, the rank of the matrix) is much lower than the actual dimension of the matrix Y. This assumption can be formulated as:

$$\mathbf{Y} = WZ^{\top}, \tag{1}$$

for matrices $Y \in \mathbb{R}^{N\times T}$, $W \in \mathbb{R}^{N\times k}$ and $Z \in \mathbb{R}^{T\times k}$. If k is smaller than N and T, the rank of $\mathbf{Y}$ is k instead of N or T. Practically, this means we only store k(T + N) values of Y instead of NT values. The former being much smaller if k is chosen to be small.

To illustrate, assume we have a panel data of firm-years given by $Y \in \mathbb{R}^{70000\times 16}$, where every row is the vector representation of one firm, and assume that all firms can (approximately) be considered linear combinations of only 10 different firms, i.e. k = 10. This means we can store the data on all firm-years with only 10 × (16 + 70000) = 701,600 entries, as opposed to the NT=1,120,000 entries of the original dataset. This is approximately 62% of the original data entries.

There are numerous approaches to factorising matrices. In this paper, we focus on the singular value decomposition (SVD) approach; SVD generalises the concept of eigendecompositions of square matrices. It can be shown that every real matrix $Y \in \mathbb{R}^{N\times T}$ can be factorised into three matrices $U \in \mathbb{R}^{N\times N}$, $\Sigma \in \mathbb{R}^{N\times T}$ and $V \in \mathbb{R}^{T\times T}$ via

$$Y = U\Sigma V^\top, \tag{2}$$

where, both U and V are orthogonal matrices, i.e. $U^\top U = I_{N\times N}$, $UU^\top = I_{N\times N}$, $V^\top V = I_{T\times T}$ and $VV^\top = I_{T\times T}$, with their columns called left- and right- singular vectors of Y. In our case, where our matrix will always have $N > T$, the matrix Σ is a diagonal matrix of the form:

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

The entries $\{\sigma_j\}_{j=1}^{\min(N,T)}$ are called the singular values of Y, and are all non-negative i.e. $\sigma_j \geq 0 \; \forall \; j \in \{1, \cdots, \min(N,T)\}$.

Clearly, (1.2) is a special case of (1.1) i.e. $W = U\Sigma$ and $Z = V$. The SVD of our matrix Y allows us to easily compute the Frobenius norm of said matrix, given that the Frobenius norm is equivalent to the euclidean norm of the vector of singular values. Now, we can easily define our lower dimensional approximation approximation of matrix Y, with help from its SVD.

Supper we define a new matrix $U_k \in \mathbb{R}^{N\times T}$ as the first k columns of U. We thus have:

$$U_k U_k^\top Y = U_k U_k^\top \, U\Sigma V^\top = U_k \left(I_{k\times k} 0_{k\times(N-k)} \right) \Sigma V^\top = U\Sigma_k V^\top, \tag{3}$$

where $\Sigma_k \in \mathbb{R}^{N\times T}$ is defined as:

$$
\Sigma_k = \begin{pmatrix}
\sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \ddots & 0 & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_k & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}
$$

Therefore, $Y_k = U\Sigma_k V^\top = U_k U_k^\top Y$ is a rank k approximation of Y. In actuality, it is more than a rank-k-approximation, it is the *best rank-k-approximation* in the sense of the Frobenius norm. Note also that our depiction of Σ_k as square is simply for illustrative purposes.

Theorem 2.1. (Best rank-k-approximation). This theorem is based on the Eckart-Young-Minsky theorem.
For any matrix $\hat{Y} \in \mathbb{R}^{N\times T}$ with rank $(\hat{Y}) = k$, we have:

$$
\left\|Y - \hat{Y}\right\|_{FRO}^2 \geq \|Y - Y_k\|_{FRO}^2 = \left\|Y - U_k U_k^\top Y\right\|_{FRO}^2 = \sum_{j\geq k+1}^{\min(N,T)} \sigma_j^2.
$$

Therefore, Y_k is the best rank-k-approximation in the sense of the Frobenius norm.

Proof. See Eckart and Young (1936) for a proof of this theorem. □

2.1 Matrix Completion: Set Up

In this sub-section, we elaborate on the causal problem and lay out our matrix factorisation approach to a matrix completion task. This task falls under the unsupervised machine learning category, and is widely applied in ***recommender systems***.[4] As we develop the algorithm, our reader will begin to understand how causal problems with panel data settings can be formulated as matrix completion problems. Ultimately, the matrix completion algorithm allows us to impute

[4]Recommender systems are systems that aim to make recommendations to customers or users. This might entail ranking movies that a user has not seen, based on ratings of other movies seen by the same user, and ratings of the seen and unseen movies by other users. Completing missing ratings can be formulated as a matrix completion problem.

the missing potential control outcomes, which consequently allows us to estimate the average causal effect of VCT funding on VCT-funded firms.

We now elaborate on the set-up for our problem which draws inspiration from Athey et al (2018). Consider an $N \times T$ matrix $\mathbf{Y}$ which denotes our panel data of total-assets for N firms observed over T periods, with typical observation Y_{it} $\forall$ $i \in \{1, \cdots, N\}$ and $t \in \{1, \cdots, T\}$. Our adapted setup is motivated by a causal potential outcome setting (see Rubin (1974); Imbens and Rubin (2015); Athey et al (2018)), where at each time period, a firm either receives VCT funding or not. We characterise this as $W_{it} \in \{0,1\}$. In other words, W_{it} is an indicator for whether a firm received VCT-funding. We thus observe for each firm and time period, the pair $(Y_{it}W_{it})$, where the observed total-asset is $Y_{it} = Y_{it}(W_{it})$. We now turn to laying out our estimand: the average effect of the U.K venture capital trust funding scheme on investment - for firms who received said funding. This effect is formulated as:

$$\zeta = \frac{\Sigma_{i,t:W_{it}=1}[Y_{it}(1) - Y_{it}(0)]}{\Sigma_{i,t}W_{it}}. \tag{4}$$

To estimate this quantity, we need to impute the "missing" potential total-assets for all VCT-funded firms. Given the form of our estimand ζ, all the total-asset entries for $Y_{it}(1)$ are observed. We want to impute the counterfactual entries for VCT-funded: firms with $W_{it} = 1$. We will call this $Y_{it}(0)$ for firms with $W_{it} = 1$. For ease of notation and blending with the matrix completion literature, we will henceforth refer to our task as imputing the missing values of a partially observed matrix of total-assets Y. We do this safe in the knowledge that our reader understands that we are referring to the matrix of counterfactual outcomes. We are imputing the counterfactual total-assets of VCT-funded firms; the total-assets for VCT-funded firms had they not received VCT-funding. With this task completed, we can estimate our average causal effect of interest ζ.

To illustrate our imputation problem, we can think of our firm-year total-assets data as composed of two $N \times T$ matrices; one incomplete and the other complete:

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & Y_{12} & ? & \cdots & Y_{1T} \\ ? & ? & Y_{23} & \cdots & ? \\ Y_{31} & ? & Y_{33} & \cdots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N,1} & ? & Y_{N3} & \cdots & ? \end{pmatrix}, \qquad \mathbf{W} = \begin{pmatrix} 0 & 0 & 1 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \cdots & 1 \end{pmatrix},$$

with missing entries denoted by the question mark, thus

$$W_{it} = \begin{cases} 1 \\ 0 \end{cases}$$

is an indicator to indicate that the corresponding component of Y. i.e. Y_{it}, is missing.
Finally, we will henceforth interchangeably refer to VCT-funded firms as shocked, treated or investee firms.

2.2 Pattern of Missing Data: Staggered Adoption

In this study, we know that VCT-funded firms received said funding in a staggered fashion. In other words, there is a time-varying adoption of treatment (Athey et al, (2018), Shaikh and Toulis (2019)). Essentially, this means VCT-funded firms received VCT funding at different periods, and in some cases, over several years. Also, once a firm receives VCT funding i.e. becomes an investee, we assume it remains an investee forever. We illustrate with the below:

$$Y_{N\times T} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & \cdots & T \end{matrix} & \\ & \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \cdots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \cdots & \text{X} \\ \checkmark & \checkmark & \text{X} & \text{X} & \cdots & \text{X} \\ \checkmark & \checkmark & \text{X} & \text{X} & \cdots & \text{X} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \text{X} & \text{X} & \text{X} & \cdots & \text{X} \end{pmatrix} & \begin{matrix} \text{never received VCT funding} \\ \text{received VCT funding in period T} \\ \\ \text{received VCT funding in period 3} \\ \\ \text{received VCT funding in period 2} \end{matrix} \end{matrix}$$

Here, the check-mark (✓) represents observed values whilst the X represents missing values for total-assets. For instance, firm N (in the last row entry) received VCT funding in period 2. It may or may not have received further rounds of VCT funding in subsequent periods - up to period T. Regardless, the firm remains an investee firm from the moment it received VCT funding - hence the X in periods $3, 4, \cdots, T$.

2.3 The Matrix Completion Estimator

Given our $N \times T$ panel data/matrix of U.K firms' total-assets Y, which we model with the form:

$$\mathbf{Y} = \mathbf{L}\ , \tag{5}$$

our goal is to find a low-rank approximation to said matrix. The first a-priori assumption that we want to make is that the firms' in our matrix can be classified into types, and that the different types are less than N. Also, we assume that every observation in our matrix Y can be modelled as a linear combination of the observations from all firm types. Mathematically, this means that we assume that the matrix with all entries has a low-rank.
The task of finding a low-rank matrix approximation $\hat{L} \in \mathbb{R}^{N\times T}$ [5] to our total-assets matrix $Y \in \mathbb{R}^{N\times T}$ can be formulated as the convex optimisation problem:

$$\hat{L} = \underset{L\in\mathbb{R}^{N\times T}}{\arg\min} \left\{ \frac{1}{2}\|L - Y\|^2_{\text{Fro}} \ + \ \alpha\|L\|_* \quad \text{subject to } P_\Omega L = P_\Omega Y \right\}, \tag{6}$$

where $\|\cdot\|_*$ denotes the *nuclear-norm* [6], which is the one-norm or the sum of the vector of singular values of Y. i.e.

$$\|L\|_* = \sum_{j=1}^{\min(N,T)} \sigma_j,$$

[5]The low-rank matrix L has rank-r where $r << \min(N,T)$ so that it is low-rank

[6]The *nuclear-norm* is akin to a convex envelope of the rank without which the problem is non-convex. To elaborate, the nuclear norm is equal to the sum of the singular values of a matrix and is the best convex lower bound of the rank function on the set of matrices whose singular values are all bounded by 1. The intuition behind this heuristic is that whereas the rank function counts the number of non-vanishing singular values, the nuclear norm sums their amplitude, much like how the $\mathscr{L}_1$ norm is a useful surrogate for counting the number of non-zeros in a vector. Moreover, the *nuclear-norm* can be minimised subject to equality constraints via semi-definite programming. For the theoretical basis for when *nuclear-norm* produces the minimum rank solution. See Recht (2011).

where $\alpha > 0$ is a regularisation parameter and $\left\{\sigma_j\right\}_{j=1}^{\min(N,T)}$ denotes the singular values of L. Effectively, the *nuclear-norm* implicitly penalises the rank of the matrix $\hat{L}$ that we wish to recover. In order to ensure that the entries for which $\hat{L}$ is known matches the observed entries, we impose the constraint $P_\Omega L = P_\Omega Y$.

$P_\Omega : \mathbb{R}^{N\times T} \to \mathbb{R}^r$ denotes the projection onto the r observed entries of our total-assets matrix Y, provided by the set Ω. $P_\Omega Y$ are the known values of our total-assets matrix at these indices. To illustrate, we characterise our orthogonal projection operator P_Ω as

$$P_\Omega(Y)_{it} = \begin{cases} Y_{it}, & \text{if } (i,t) \in \Omega \\ 0, & \text{otherwise,} \end{cases}$$

and assume our incomplete total-assets matrix Y is given as

$$\mathbf{Y} = \begin{pmatrix} 1 & 4 & ? \\ ? & 2 & 7 \end{pmatrix}.$$

We know the indices $\Omega = \{(1,1),(1,2),(2,2),(2,3)\}$, and can therefore project them, i.e.

$$P_\Omega Y = (1 \;\; 4 \;\; 2 \;\; 7)^\top$$

Note that this operator is linear and its transpose operation $P_\Omega^\top : \mathbb{R}^r \to \mathbb{R}^{N\times T}$ is

$$P_\Omega^\top = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{pmatrix} z_1 & z_2 & 0 \\ 0 & z_3 & z_4 \end{pmatrix},$$

for $z := P_\Omega Y$.

In the sub-section after next, we will derive a computationally efficient algorithm for the numerical solution of our optimisation problem 6. However, it is pertinent to emphasise that 6. is a proximal mapping. In the following sub-section, we derive and show that this proximal map has a simple closed form solution.

2.4 Closed Form Solution: Proximal Map

We know that for $R(L) = \|L\|_*$ we are certain that $R(Q_1\, L\, Q_2) = R(L)$ for two orthogonal matrices $Q_1 \in \mathbb{R}^{N\times N}$ and $Q_2 \in \mathbb{R}^{T\times T}$. The intuition behind this heuristic is straightforward; the singular value decomposition (SVD) of $(Q_1\, L\, Q_2)$ is the same as the SVD of (L).
Now, let $Y = U\Sigma V^\top$ denote the SVD of Y, we can substitute $\hat{L}$ for $\tilde{L} = U^\top \hat{L} V$, with

$$\begin{aligned}\tilde{L} &= \underset{L\in\mathbb{R}^{N\times T}}{\arg\min} \left\{ \frac{1}{2}\left\|ULV^\top - U\Sigma V^\top\right\|^2_{\text{Fro}} \;+\; \alpha\left\|ULV^\top\right\|_* \right\}, \\ &= \underset{L\in\mathbb{R}^{N\times T}}{\arg\min} \left\{ \frac{1}{2}\|L-\Sigma\|^2_{\text{Fro}} \;+\; \alpha\|L\|_* \right\},\end{aligned} \tag{7}$$

where α is a regularisation parameter determined by cross-validation. Given Σ is a diagonal matrix with non-negative entries, the solution of (1.7) has to be a diagonal matrix with non-negative entries as well. Consequently, (1.7) simplifies to

$$\begin{aligned}\tilde{L} = &\underset{L\in\mathbb{R}^{\min(N,T)}_{\geq 0}}{\arg\min} \left\{ \frac{1}{2}\|l-\sigma\|^2 \;+\; \alpha \sum_{j=1}^{\min(N,T)} l_j \right\}, \\ = &\underset{L\in\mathbb{R}^{\min(N,T)}_{\geq 0}}{\arg\min} \left\{ \frac{1}{2}\sum_{j=1}^{\min(N,T)} (l_j-\sigma_j)^2 \;+\; \alpha \sum_{j=1}^{\min(N,T)} l_j \right\}.\end{aligned} \tag{8}$$

$l \in \mathbb{R}^{\min(N,T)}_{\geq 0}$ is the vector of diagonal entries of $\tilde{L}$, i.e. $\tilde{L} = \text{diag}(l)$, and also the vector of singular values of $\tilde{L}$. The vector $\sigma \in \mathbb{R}^{\min(N,T)}_{\geq 0}$ denotes the singular values of Σ, i.e. $\Sigma = \text{diag}(\sigma)$. Equation (1.8) has a closed-form solution - the soft-thresholding of the singular values σ! - given by:

$$\tilde{l}_j = \max(\sigma_j - \alpha,\; 0), \qquad \forall j \in \{1, \cdots, \min(N,T)\}.$$

We can thus compute the solution of (1.6) via

$$\hat{L} = U\tilde{L}V^\top, \qquad \text{for} \quad \tilde{L} = \text{diag}(\tilde{l}). \tag{9}$$

As with convention, we will express the solution to (1.9) in the proximal map notation as

$$\hat{L} = (I + \alpha\partial\|\,.\,\|_*)^{-1}(Y). \tag{10}$$

The intuitive implications of this proximal map are straightforward. Given our matrix of total-assets Y, all singular values below the threshold α will be set to zero, thus enforcing a lower rank of $\hat{L}$ compared to Y - if α is larger than at least the smallest singular value of Y. All other singular values are reduced by the factor α.

2.5 Numerical Solution: The Linearised Bregman Iteration

We now turn to deriving an efficient algorithm for the numerical solution of (1.6). Our algorithm is based on the following generalisation of the Bregman proximal algorithm (Algorithm 2 in Appendix A) to non-smooth functions, otherwise known as Bregman iteration:

$$w^{k+1} = \underset{w \in \mathbb{R}^N}{\arg\min} \left\{ E(w) \; + \; D_J^{p^k}(w, \; w^k) \right\}, \tag{11a}$$

$$p^{k+1} = p^k - \nabla E(w^{k+1}), \tag{11b}$$

for initial values w^0 and $p^0 \in \partial J(w^0)$, where ∂J denotes the sub-differential of J as defined in Definition (A.0.1. See Appendix A), and $D_J^p(w, \upsilon)$ is the generalised Bregman distance as defined in Definition (A.0.2. See Appendix A)

$$D_J^p(w, \upsilon) = J(w) - J(\upsilon) - \langle p, w - \upsilon \rangle,$$

for $p \in \partial J(\upsilon)$. From the original Bregman method, we can derive a linearised variant for the choice $J(w) = \frac{1}{\tau}\left(\frac{1}{2}\|w\|^2 + R(w)\right) - E(w)$. Bregman iteration (1.11) then reads

$$
\begin{aligned}
w^{k+1} &= \underset{w\in\mathbb{R}^N}{\arg\min}\left\{E(w) \;+\; \frac{1}{2\tau}\left\|w-w^k\right\|^2 + \frac{1}{\tau}D_R^{q^k}(w,w^k) - E(w) + E(w^k) + \langle\nabla E(w^k), w-w^k\rangle\right\},\\
&= \underset{w\in\mathbb{R}^N}{\arg\min}\left\{\frac{1}{2\tau}\left\|w-w^k\right\|^2 + \frac{1}{\tau}D_R^{q^k}(w,w^k) + \langle\nabla E(w^k), w-w^k\rangle\right\},\\
&= \underset{w\in\mathbb{R}^N}{\arg\min}\left\{\frac{1}{2\tau}\left\|w-w^k\right\|^2 + \frac{1}{\tau}D_R^{q^k}(w,w^k) + \frac{1}{\tau}\langle\tau\nabla E(w^k), w-w^k\rangle + \frac{1}{2\tau}\left\|\tau\nabla E(w^k)\right\|^2\right\},\\
&= \underset{w\in\mathbb{R}^N}{\arg\min}\left\{\frac{1}{2}\left\|w-\left(w^k-\tau\nabla E(w^k)\right)\right\|^2 + \; D_R^{q^k}(w,w^k)\right\},\\
&= (I+\partial R)^{-1}\left(w^k+q^k-\tau\nabla E(w^k)\right), \qquad (12a)\\
q^{k+1} &= q^k - \left(w^{k+1}-w^k-\tau\nabla E(w^k)\right), \qquad (12b)
\end{aligned}
$$

for sub-gradients $q^k \in \partial R(w^k)$ and the short-hand notation

$$(I+\partial R)^{-1}(z) := \underset{y\in\mathbb{R}^N}{\arg\min}\left\{\frac{1}{2}\|y-z\|^2 + R(y)\right\}.$$

Now, we focus on the special case $E(w) = \frac{1}{2}\|Aw-b\|^2$, for a matrix $A\in\mathbb{R}^{T\times N}$ and a vector $b\in\mathbb{R}^T$. For this special case, (1.12) reads

$$w^{k+1} = (I+\partial R)^{-1}\left(w^k+q^k-\tau\, A^\top(Aw^k-b)\right) \qquad (13a)$$

$$q^{k+1} = q^k - \left(w^{k+1}-w^k-\tau\, A^\top(Aw^k-b)\right). \qquad (13b)$$

If we assume that $(w^k+q^k)/\tau \in \mathbf{R}(A^\top)$, we can substitute $\tau\, A^\top b^k = w^k+q^k-\tau\, A^\top(Aw^k-b)$, which modifies (1.13) to

$$w^{k+1} = (I+\partial R)^{-1}\left(\tau\, A^\top b^k\right) \qquad (14a)$$

$$b^{k+1} = b^k - \left(Aw^{k+1}-b\right). \qquad (14b)$$

with initial value $b^0 = b$. Combining both equations of (1.14) into one yields:

$$b^{k+1} = b^k - \left(A(I+\partial R)^{-1}\left(\tau\, A^\top b^k\right) - b\right) \tag{15}$$

The motive behind re-characterising (1.13) to (1.15) is that (1.15) is simply gradient descent (See Algorithm 3 in Appendix A) applied to a very specific energy that we characterise with the following lemma:

Lemma 3. (Linearised Bregman iteration as gradient descent). The linearised Bregman iteration in the form of (1.15) is a gradient descent method with step-size one, i.e.

$$b^{k+1} = b^k - \nabla G_\tau(b^k),$$

applied to the energy

$$G_\tau(b^k) := \frac{\tau}{2}\left\|A^\top b^k\right\|^2 \;-\; \langle b^k, b\rangle \;-\; \frac{1}{\tau}\,\tilde{R}(\tau\, A^\top b^k).$$

Here, $\tilde{R}$ represents the Moreau-Yosida regularisation of the function R, i.e.

$$\begin{aligned}\tilde{R}(z) &:= \inf_{y\in\mathbb{R}^N}\left\{R(y) + \frac{1}{2}\|y-z\|^2\right\},\\ &= R\left((I+\partial R)^{-1}(z)\right) \;+\; \frac{1}{2}\left\|(I+\partial R)^{-1}(z) - z\right\|^2.\end{aligned}$$

Proof. The proof is reasonably succinct if we can compute the gradient of $\tilde{R}$, since the gradient of $\frac{\tau}{2}\left\|A^\top b^k\right\|^2 - \langle b^k,\ b\rangle$ [7] simply reads $\tau\, AA^\top b^k - b$. To compute the gradient $\nabla\tilde{R}$, we start by rewriting $\tilde{R}$ to

$$\begin{aligned}\tilde{R}(z) &= \inf_{y\in\mathbb{R}^N}\left\{R(y) + \frac{1}{2}\|y-z\|^2\right\},\\ &= \inf_{y\in\mathbb{R}^N}\left\{R(y) + \frac{1}{2}\|y\|^2 - \langle y,\ z\rangle + \frac{1}{2}\|z\|^2\right\},\\ &= \frac{1}{2}\|z\|^2 - \sup_{y\in\mathbb{R}^N}\left\{\langle y,\ z\rangle - R(y) - \frac{1}{2}\|y\|^2\right\},\\ &= \frac{1}{2}\|z\|^2 - \left(\frac{1}{2}\|\cdot\|^2 + R\right)^*(z),\end{aligned}$$

[7] $\langle\,\cdot\,\rangle$ denotes the inner product

where $F^*(z) := \sup_{y\in\mathbb{R}^N} \langle y,\ z\rangle - F(x)$ denotes the *convex conjugate* or *Fenchel conjugate* of a function F. We should point out that by definition, we observe that $\nabla F^*(z) = y^*$, where $y^* = argmax_{y\in\mathbb{R}^N}\{\langle y,\ z\rangle - F(y)\}$. When we have $F(y) = \frac{1}{2}\|y\|^2 + R(y)$, the problem reads as

$$
\begin{aligned}
y^* &= \arg\max_{y\in\mathbb{R}^N}\left\{\langle y,\ z\rangle - \frac{1}{2}\|y\|^2 - R(y)\right\}, \\
&= \arg\max_{y\in\mathbb{R}^N}\left\{-\frac{1}{2}\|y-z\|^2 - R(y)\right\}, \\
&= \arg\max_{y\in\mathbb{R}^N}\left\{\frac{1}{2}\|y-z\|^2 - R(y)\right\}, \\
&= (I+\partial R)^{-1}(z).
\end{aligned}
$$

With regards the gradient of $\tilde{R}$, we observe

$$
\nabla\tilde{R}(z) = z - (I+\partial R)^{-1}(z).
$$

As an immediate result of the chain rule, we have $\nabla\left(\left(\frac{1}{\tau}\tilde{R}\right)\circ\left(\tau A^\top\right)\right)(b^k) = A\nabla\tilde{R}(\tau A^\top b^k)$, [8] and thus conclude

$$
\begin{aligned}
\nabla G_\tau(b^k) &= \tau AA^\top b^k - b - \tau AA^\top b^k + A(I+\partial R)^{-1}\left(\tau A^\top b^k\right), \\
&= A(I+\partial R)^{-1}\left(\tau A^\top b^k\right) - b,
\end{aligned}
$$

which also serves as the proof. In the next section, we will apply Algorithm (1.15) to our optimisation problem (1.6). □

3.1 A Bregman algorithm for Matrix Completion

For notational convenience, let us rewrite $L - Y$ from Eq.1.6 as S. To solve (1.6), we have $w = (L, S)$, and E and J as

[8] $\circ$ is the Hadamard product/element-wise, entry-wise or Schur product. It is a binary operation that takes two matrices of the same dimensions and produces another matrix of the same dimension as the operands, where each element i, j is the product of elements i, j of the original two matrices. It should not be confused with the more common matrix product.

$$E(L,S) = \frac{1}{2}\| \, L+S-Y \, \|^2_{\text{Fro}} \qquad \text{and}$$
$$J(L,S) = \frac{1}{\tau}\left(\frac{1}{2}\| \, L \, \|^2_{\text{Fro}} \; + \; \gamma\alpha\| \, L \, \|_* + \frac{1}{2}\| \, S \, \|^2_{\text{Fro}} \; + \; \gamma\| \, S \, \|_*\right) - E(L,S),$$

for constants $\tau > 0$ and $\gamma > 0$.

Given our choices of E and J, we are in the exact framework of (1.13) for $A = (I \;\; I)$ and $b = Y$, thus we can numerically solve (1.6) by iterating the updates (1.14), which for our choice of E & J reads

$$L^{k+1} = (I \; + \; \gamma \, \alpha \, \partial\| \, . \, \|_*)^{-1} \left(\tau P_\Omega^\top z^k\right) \tag{16a}$$
$$S^{k+1} = \left(I \; + \; \gamma\partial\| \, . \, \|_{\text{Fro}}\right)^{-1} \left(\tau P_\Omega^\top z^k\right) \tag{16b}$$
$$z^{k+1} = z^k - \left(P_\Omega L^{k+1} - \; z\right), \tag{16c}$$

for $z^0 = z := P_\Omega Y, \;\; \tau \leq 1$ and $\alpha > 0$. Approach (1.16) is summarised in Algorithm 1 below.

Algorithm 1: Matrix Completion

Specify: parameters $\gamma > 0$, $\alpha > 0$, stopping index K

Initialise: Set of known matrix indices Ω, $z^0 = P_\Omega Y$ and $\tau \leq 1$

Iterate:

for $K = 0, \cdots, K-1$ **do**

$L^{k+1} = (I \; + \; \gamma \, \alpha \, \partial\| \, . \, \|_*)^{-1} \left(\tau P_\Omega^\top z^k\right)$;

$S^{k+1} = (I \; + \; \gamma \, \partial\| \, . \, \|_*)^{-1} \left(\tau P_\Omega^\top z^k\right)$;

$z^{k+1} = z^k - \left(P_\Omega L^{k+1} - \; z\right)$;

end

return L^K, S^K.

Algorithm 1 is relatively straightforward. It entails computing the singular value decomposition of $\tau P_\Omega^\top z^k$ in every iteration, and soft-thresholding the singular values as in (1.9) with threshold $\gamma \, \alpha$. Subsequently, we update the matrix z^k by subtracting the residual $P_\Omega L^{K+1} - \; z$ from it.

By applying an identical procedure to the proof of Theorem A.0.1 in appendix A, we can prove that Algorithm 1 converges at a rate of $1/K$ for $\tau \leq 1/\|A\|^2 = 1/2$, where K denotes the number of iterations.

4 All About VCT's

Before we get into the data collection, analysis and results, it is useful to provide a detailed insight into VCTs and what they are about.

The Venture Capital Trust (VCT) scheme, introduced in 1995, is one of three tax-based venture capital schemes, the others being the Enterprise Investment Scheme (EIS) and the Seed Enterprise Investment Scheme (SEIS). The VCT scheme is designed to encourage investors to invest (indirectly) in British, unquoted, smaller, and higher-risk firms - with a need for start-up, early stage or expansion capital - by investing through subscription to a VCTs shares. VCTs are U.K publicly-quoted and closed-ended funds. They typically fall under three broad sectors: generalist (which includes private equity and development capital), AIM, and specialist sectors e.g. renewable energy infrastructure, technology, or media. The typical VCT appoints a regulated investment manager who invests and manages the fund on a daily basis; very few VCTs are "self-run" by their directors. Conditional on the VCTs objectives and the VCT scheme regulations, the investment managers' goal is to invest in firms' that maximises returns to its shareholders. To that end, VCTs monitor, work with, and provide expert advise to their investee firms over the medium to long-term to help increase their value and therefore potentially maximise returns for its investors. We will provide more details on the VCT scheme and regulations - including how these regulations have evolved - when we present the main results in subsequent sections. But for now, the main highlights of the VCT scheme and its regulations are that VCTs: must be listed on a UK recognised Stock Exchange, are exempt from corporation tax on any capital gains from the disposal of an investment, can only invest in firms' carrying on a "qualifying"[9] trade with fewer than 250 full-time equivalent employees at the time shares are issued, and gross assets of no more than £15 million before investment and £16m immediately after investment. Potential investee firms can raise up to £5 million in any 12 month period from VCT investment, with this sum also inclusive of any investment via the other two venture capital schemes mentioned earlier; EIS and

[9]We will expand on qualifying trades when we analyse the average treatment effect on the treated (ATET)

SEIS.

A VCT will typically hold an investment for a period of three to seven years before looking to sell its stake in the investee firm. A very high percentage of the exit proceeds - subject to the VCTs investment policy and prevailing VCT scheme regulations - are re-invested into new investee firms'; VCT regulations require tax-free dividends be paid to investors where a gain is made. In very rare instances, some VCTs are set-up with a limited lifespan. These VCTs aim to exit from all of their investee firms, dissolve the VCT and return all capital to their investors after a defined period i.e. seven years. These limited-life VCTs typically focus on investee firms with guaranteed or contractual income, thus allowing for an easy exit within a defined period. We however note that with the introduction of new "risk-to-capital" guidelines for the VCT scheme in 2018, limited-life VCTs are now almost if not non-existent.

To encourage investment in VCTs, the U.K government offers significant tax advantages to VCT investors. An investor in VCT shares - purchased at launch, or during subsequent share class issues - receives up to 30% tax relief on their VCT share subscriptions of up to a maximum of £200,000, conditional on holding the investments for a minimum of five years. In addition to the tax-free dividends mentioned earlier, capital gains from VCT investments are also free of capital gains tax. If an investor purchases VCT shares on the secondary market i.e. after they are listed on the London Stock Exchange, there is no tax-relief on the purchase, but gains from such secondary market purchases are free of capital gains tax, in addition to any dividends from the investment being tax-free. Investors exit from VCTs by selling their shares on the London Stock Exchange, and or participating in any share buy-back scheme offered by VCTs.

Clearly, and in addition to tax-free savings with Individual Savings Accounts (ISA), and pension allowances, VCTs are an alternative for tax-efficient investing. We illustrate this point with a simple example. Assume a company with a share price of 200p pays a 10p dividend. The dividend yield is 5%. If an investor holds the shares of said company outside an ISA or pension, the net of tax yield is 3.38% for a higher-rate taxpayer and 3.1% for an additional-rate rate taxpayer, [10] assuming the £2,000 dividend allowance has been used. Analogously, if a VCT with an initial share price of 200p pays a 10p dividend, the yield is higher than 3.38% because the VCT investor gets up to 30% income tax-relief, hence the net purchase cost of the share is actually

[10] The tax rates are 32.5% and 38.1% respectively.

140p. A 10p dividend from VCT shares purchased at 140p results in a tax-free yield of 7.14%. To achieve an equivalent after-tax dividend yield of 7.14% on a taxable investment , a higher-rate tax-payer would need to earn a pre-tax yield of 10.6% whilst an additional rate tax-payer would need 11.5%.

4.1 VCT Tax Benefits: Illustrations

In this paper, we do not highlight nor analyse the risks inherent in subscribing to the equity issue of a VCT. [11] However - in recognition of these risks - the U.K government provides investors with a 30% income tax relief for subscriptions in new VCT fundraising. To illustrate with an illustration drawn from HMRC (2018) venture capital trust statistics, assume an investor invests £10,000 in a VCT fundraising round. This investor either receives a £3,000 cheque from the tax authority or a £3,000 reduction in her tax bill. We should emphasise that this is a tax rebate, hence restricted to the amount of income tax she paid. This means that (and given that the maximum annual VCT investment is £200,000) if she has only paid £2,000 in income tax, she would only receive a £2,000 instead of £3,000 tax rebate on her £10,000 investment. She must also hold her VCT shares for five years to permanently keep the tax rebate. Also, she does not get the rebate if she bought the shares on the secondary market. This example also illustrates the fact that the tax benefits from VCT investments are dependent on each individual investors' circumstance. We further illustrate with three more examples:

Example A

Francesca decides to invest £200,000 in a VCT offer for subscription. In the 2019/20 tax year she anticipates that she will pay £90,000 in income tax.

Example B

In the tax year 2019/20, Bukola decides to invest £10,000 in a VCT offer for subscription. She is a basic rate and non-Scottish tax-payer; she earns £30,000 annually hence will pay approximately £3,500 in income tax $([30,000 - 12,500(\text{Personal Allowance})] \times 20\%)$.

[11] VCTs are exposed to significantly higher risks than non-VCT equities. VCTs invest in smaller, fledgling firms, a lot of which will struggle or go into liquidation, resulting in losses for investors. Additionally, VCT shares are illiquid. Even though their shares are fully listed on the London Stock Exchange, they might be difficult to sell due

Investment	£200,000
Tax Rebate	(£60,000)
Effective Net Cost	£140,000
Tax Rebate as a percentage	30%

Investment	£10,000
Tax Rebate	(£3,000)
Effective Net Cost	£7,000
Tax Rebate as a percentage	30%

Example C

Adesua wants to invest £100,000 in a VCT offer for subscription. She is a higher rate and non-Scottish tax-payer; she earns £60,000 annually and has calculated that she will pay £11,500[12] in income tax in the tax year 2019/20.

Investment	£100,000
Tax Rebate	(£11,500)
Effective Net Cost	£88,500
Tax Rebate as percentage	11.5%

Adesua will not pay enough income tax to reclaim the full 30% tax rebate, hence will only receive the £11,500 in tax she paid as rebate.

to in some cases, there being a single "market maker" for the shares.

[12]Her tax liability is calculated as the sum of 0% on £12,500 personal allowance, basic rate of 20% on £37,500, and higher rate of 40% on £10,000

5 VCT Data

5.1 Measurement

In this section, we detail the hand-collection and measurement of our data on all U.K firms' (hereafter investees) that received funding via the VCT scheme - from VCT's.

Hand-collected data in various forms - surveys, scanning an entire electronic record of articles of incorporation and corporate legal filings, extracting data from a database - have become more commonplace in the financial economics literature. In designing our data measurement approach, we analysed the approaches used in studies such as Graham and Harvey (2001), Gompers et al (2016), Da Rin and Phalippou (2017), and Gompers et al (2019). Whilst our data measurement and collection approach is unique in the context of the data itself, it shares some similarity - at least in spirit - to the above mentioned studies.

Our first task is to collate data on all investees from the inception of the VCT scheme (1995) to present day. We started by sending emails to Her Majesty's Revenue and Customs (HMRC); UK's tax, payments and customs authority and governmental department responsible for the administration of the VCT scheme. Our request we thought was straightforward. We wanted a list of all investees that had ever received VCT funding since the inception of the scheme till present day. When the VCT section at HMRC were not forthcoming, we filed a freedom of information (FOI) request with the FOI section at HMRC. Our request was denied. Their legal justification was that it would take an inordinate amount of time and manpower to gather the information we sought. We then tried another approach.

We sent emails and placed phone calls to all VCTs' in the UK. Our request was simple. We asked each VCT to kindly provide a list of all the investees they had ever invested in to date, and most importantly, the date of the investment. This data would help us determine what firms' in the universe of U.K firms were "treated", and when they were "treated". The responses to our request were on average enthusiastic but reiterated what HMRC had told us. The typical VCT has invested in thousands of investees, and they claimed it would take an inordinate amount of time to gather the information we needed.

These response shaped our next plan of action. Firstly, we scoured the Companies House Service,

[13] the London Stock Exchange (LSE) website, and the Association of Investment Companies (AIC) website - to build a list of all current (62) VCTs' and former VCTs'.

Armed with this list, we began in alphabetical order - say we began with Albion Development VCT. We sourced and gathered every semi-annual and annual report it had ever published (e.g. from 1996 - 2018). We painstakingly read through the reports and from each, we collected details on its investees: the first time they received VCT funding, categorical information such as the date of incorporation, and industry sector. We focus on first-time investees because, in our matrix completion approach, once a firm is "treated", it remains in the "treated" group forever. This aspect is especially relevant because it means we do not need to track the subsequent funding rounds of each investee. The idea that once an investee is "treated", it remains in the treated group forever is really straightforward. Once you receive VCT funding, you can't "un-receive" it. The offer details of subsequent VCT or venture funding rounds are related to the offer details of preceding funding rounds. Also, there is a strong data-driven reason for our *once treated, you remain in the treated group forever* approach. The average investee will go through several funding rounds during its growth. This means that they undergo several "rounds of treatment", thus reemphasising our treatment approach. This approach has also been employed in the literature on causal potential outcomes, and has been dubbed "staggered adoption" - as highlighted in subsection (1.2.2). For more on "staggered adoption", see Athey and Stern (2002), Athey and Imbens (2018), and Athey et al (2018).

We worked through the entire list of current and former VCTs'. For each and every VCT, we repeat the same painstaking process of reading through every financial report it ever produced, to gather a list of all investees it invested in, the date of investment, the industry the investee operates in - amongst other categorical data. Where there is ambiguity regarding the date an investee was funded, we sent further emails and placed phone calls to the relevant VCT to find out when they funded so-and-so investee.

Armed with a list of all investees: 1931 unique U.K firms[14], we moved on the next stage of our

[13] A digital search service that provides free access to all public information stored on the UK register of companies

[14] The number of firms that received VCT funding for the first time is closer to 2000. However, due to data hand-collection difficulties - especially with regard to the exact date the investee received the VCT funding - we excluded some firms from this analysis.

data collection.

5.2 Data Source: FAME Data

Our panel/matrix data on total-assets and other categorical information for both VCT and non-VCT funded U.K firms are from the Financial Analysis Made Easy (FAME) database. FAME contains detailed financial, legal, and ownership information for public and privately incorporated firms in the U.K and Republic of Ireland.

The data-set is made up of 1931 investees (VCT-funded firms) and 80,000 randomly selected non-investee firms.

It is also worth mentioning that our sample is free of survivorship bias - as FAME reports historical information for up to 10 years regardless of whether a firm reports financial data or not.

5.3 Measurement and Implementation

Our data sample spans the time period 2003 - 2018 at an annual frequency. FAME data coverage starts from 2001, but we constricted our sample to start at 2003 because the 2001-2002 total-asset entries for a significant proportion of firms are missing; recall, our outcome variable is investment (the difference in natural log of total-assets).

For VCT-backed firms - one at a time - we search the FAME database for its financial and categorical information. This exercise was just as time consuming as the hand-collection of information on VCT-backed firms. Ensuring that the firm whose data is downloaded off of FAME is the exact same firm that received VCT backing is a time consuming and painstaking process. The data on non-VCT backed firms in the U.K is a FAME random sample which is representative of the universe of U.K firms. For non-VCT backed firms, we exclude firm-year observations with missing or zero values for total-assets between the periods 2003-2018. Of the 80,000 non-VCT backed firms in our random sample, the data cleaning exercise yields 73,520 non-VCT backed firms. Our final sample of 1,207,216 firm-years (investees and non-investees/non-VCT backed firms) contains information on each firms' annual total-assets, date of incorporation, Primary SIC Code, main activity, company status, and SME indicator.

Finally, we transform our total-asset values into natural logarithm (ln) before feeding it into our matrix completion algorithm.

5.4 Summary Statistics

Here, we present and analyse interesting summary statistics on investee firms and VCTs. In some instances, our summary statistics plots depicts the median aggregate; in some plots, we depict the mean aggregate. Our reasoning is very simply; the mean is sometimes drastically skewed by numerous outliers in the data. In such instances, we depict the median aggregate. The data for Fig.5 through Fig.10 are from the HMRC Venture Capital Trusts Statistics (2018). We begin with Fig.1. First thing to note is that the median pre-VCT-funding size of investee firms has varied over time. It ranges from circa £8.3m in 2008 to circa £1.7m in 2017. However, notice that post-2015, the median size has been at its lowest in all of the sample period (between 1.7m - 2.7m). In a latter section, we will analyse how changes to the policies guiding VCTs has impacted their aggregate investment decisions. Specifically, the 2015 rules prohibiting VCTs from investing in firms older than 7 years and the mandate that the potential investee must be an entrepreneurial firm with a genuine risk of loss of capital, and the objective to grow and develop. This policy change helps explain why the median pre-VCT-funding size of investees has shrunk since 2015.

We next present the number of firms that received funding per annum in Fig.2. We observe that VCTs invested in a record-breaking number of firms in both 2014 and 2018. What does this mean? Did VCTs fund-raise a record-breaking amount in both 2014 and 2018, and by implication, invest a record-breaking amount in both years - adopting a strategy of investing this record-breaking sum across a record-breaking large number of firms i.e. increase the extensive margin. Did VCTs fund-raise an average-record amount in 2014 and 2018, and by implication, invest an average-record amount, but spread this across a record-breaking large number of firms, hence the record number of new investees? We can answer this by jointly analysing our Fig.2. with Table.2. In tandem with Table.2. (column 2), it is clear that the extensive and intensive margin both increased. We observe that more money was raised in the periods 2013-2014 and 2014-2015 relative to the last 7-8 years. Also, 2018 was record-setting in terms of the amount of funds raised by VCTs - second behind the 2005 period. This leads us to conclude that not only did VCTs raise record-breaking amounts in both 2014 and 2018, they also invested in a record-breaking number of new firms. These patterns - investing in a larger number of firms - are related to the VCT policy changes in the respective periods. We will elaborate on these changes in subsequent sub-sections - especially with regards to annual and lifetime limits on investment in any one firm.

Fig.3. is a plot of the industry-sectoral distribution of investee firms, grouped by the year they received VCT investment. For example, the plot shows that the retail sector (red line) received the largest VCT investment in 2005, followed by the manufacturing sector (blue line). Overall, wholesale, followed by manufacturing, then services, were the largest recipients of VCT investment during the sample period; the median percentages are 8.4%, 5%, and 4.5% respectively.

Fig.4. categorises investee firms according to their current Companies House status. i.e. whether they are still Active or Dissolved/In Liquidation. For example, the first set of bars (blue then red) depicts the number of firms that received first-time VCT-funding in 2003, categorised according to their current status (Active or Dissolved/In Liquidation).

The first thing that stands out is that the majority of investees in every single cohort are still Active. In aggregate, of the 1,931 unique firms in our sample that received VCT funding for the first time between 2003-2018, 68% of them are still Active, with the remainder 32% classed as Dissolved/In Liquidation. To put these numbers in context, the Office for National Statistics (ONS) Business Demography data on the latest five year survival rate for British firms is 42.5%. It seems VCT-backed firms have out-performed the national average survival rate of new firms. However, we acknowledge that the average size of VCT-backed firms' - as measured by their total assets of 1.7 million - 2.7 million - is perhaps bigger than that of, for instance, the average startup in the Restaurants and Mobile Food Service Activities sector, and as such, using the national average survival rate to provide context might be misleading. We thus provide a more granular context by pointing out that the national average survival rate for the Computer programming, Consultancy and Related activities sector is 51.4%, a sector that is synonymous with large enterprises. This survival rate is still lower than that of VCT-backed firms' at 68%. Understanding why VCT-backed firms' have a relatively high survival rate is an important question, and will be the subject of future research.[15]

In Fig.5. the first thing to note is that since the 2008-2009 period, the annual amount of funds raised by VCTs has been consistently trending upwards. Between 2008-2018, there has been an

[15]The ONS Business Demography (2020) excel data file can be found at the following link: https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/datasets/businessdemographyreferencetable. The survival rate for Computer programming, Consultancy and Related activities is from Table 5.2a in the excel file.

almost 400% increase in the amount of funds raised - with this increase almost evenly spread across the period. In a subsequent section (Major VCT Policy Changes) where we analyse major VCT policy changes over time, we will go into detail on the specific policy changes and how they impacted VCT fund-raising activity and of course their onward funding of SMEs. For now, the highlights are: the increased income-tax relief from 20% to 40% in the 2004-2005 tax year explains the record setting amount of funds raised between 2004-2006; the 2017 patient capital review and reduction in lifetime pension allowances was the major determinant in the sustained upward trend in fund-raising since 2015-2016.

In Fig.6. the first thing to note is that the number of VCTs raising funds has almost always been less than the number of firms managing funds. It goes without saying that VCTs managing funds don't always feel the need to raise funds annually. The most interesting aspect of the fig. is the consistently decreasing number of VCTs managing funds since the 2010-2011 tax period. This period coincided with the tightening of VCT rules i.e. VCT policy changes that limited the types and size of firms a VCT could invest in. Consequentially, VCTs started to merge in response to these changes and of course, to achieve economies of scale. Additionally and as a further consequence of VCT policy changes and economies of scale, we note that the number of VCT's raising funds has been steadily declining since the 2013-2014 tax period, even though the amount of funds raised (Fig.1.5.) within the same period has been on the rise. The last thing to note is the sharp fall in the number of VCTs raising funds between 2005-2006 and 2006-2007. This was due to the decrease in the income tax-relief from 40% to 30% - for VCT investors.

From Fig.7. it is clear that the majority of VCT investors invest below £50,000. The distribution of the number of investors with regards to the amount they invest is very similar for the last two years of this study (2016-2018). The number of investors that invest between £5,000 - £10,000 represent 18.4% of the total number of VCT investors for both years and is the highest proportion. This is followed by the £25,000 - £50,000 range at 16.7% and 18% for both periods respectively. Also, only 4.8% and 5.3% of investors invested the maximum allowed for VCT investments; £200,000. However, with regards the actual amount invested in VCTs, we see from Fig.8. that investments between £150,000-£200,000 represent the highest proportion of the total amount invested in VCTs for both years.

In Fig.9. we group and then plot the number of VCT investors according to the size of their investment - for the 2016-2017 and 2017-2018 periods. The blue, red, yellow, and grey lines represent the different groupings. Firstly, the number of VCT investors for all groupings has been on the rise. Also, the £0 to £10,000, which is the lowest grouping (blue line) has the highest number of investors for both years. However and unsurprisingly, we see in Fig.10. that the actual amount invested is highest for the highest groupings, yellow and grey lines, which represent the £25,000-£100,000 and over £100,000 groupings respectively.

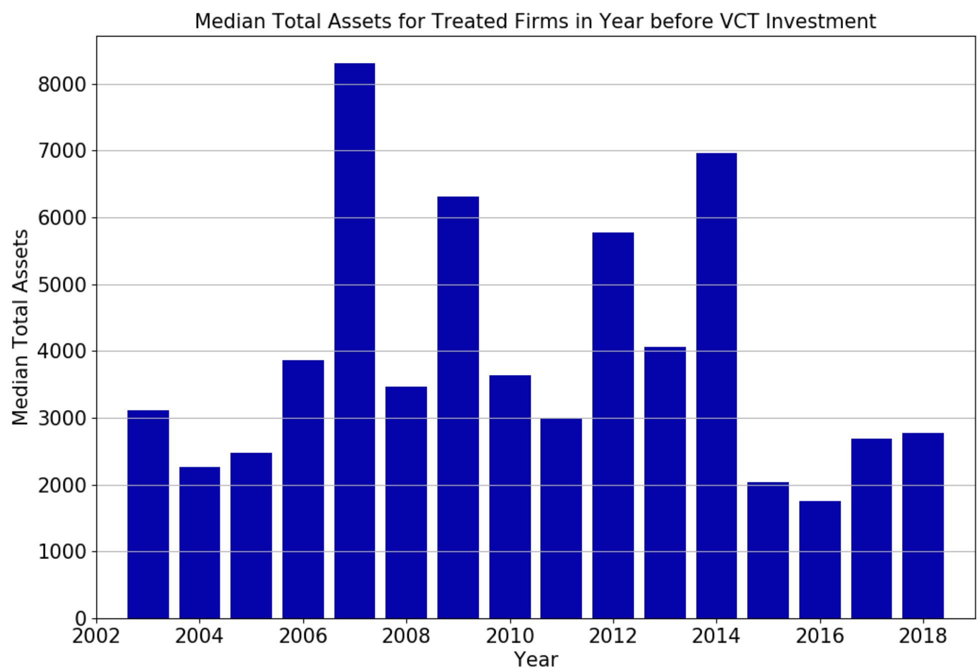

Figure 1: Median Total Assets for Investee Firms in Year before VCT Investment

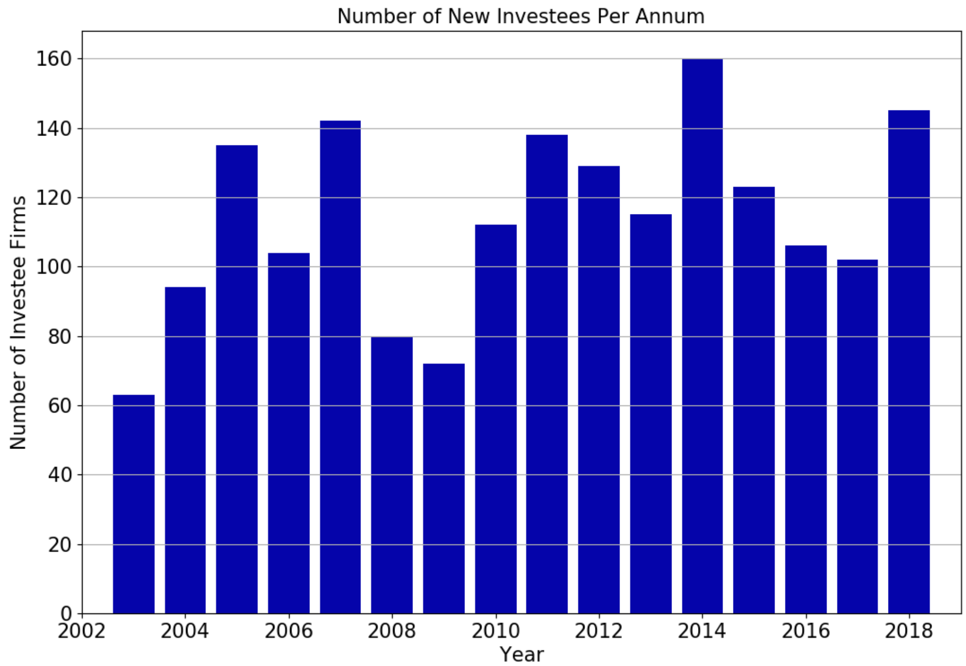

Figure 2: Number of New Investees Per Annum

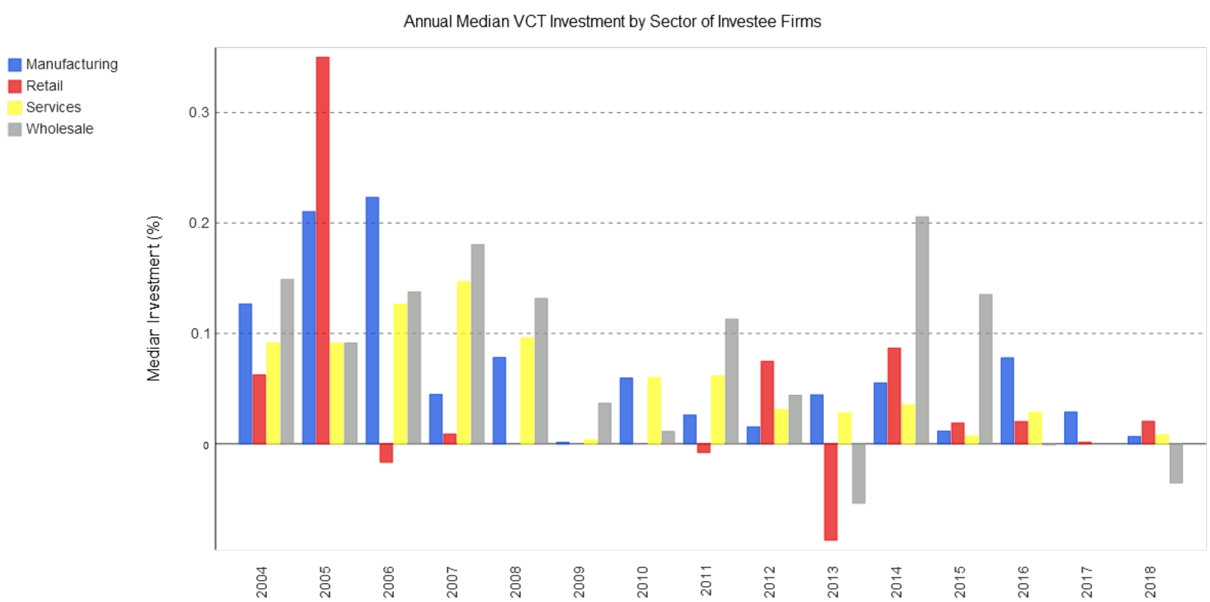

Figure 3: Annual Median VCT Investment by Sector of Investee Firms

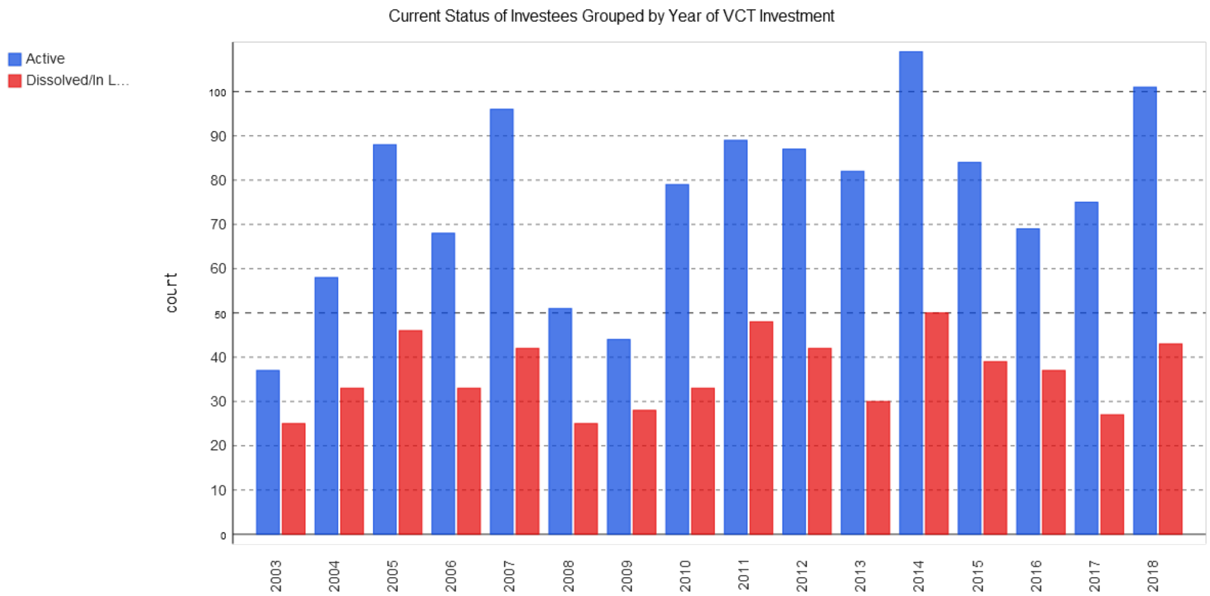

Figure 4: Current Status of Investees Grouped by Year of VCT Investment

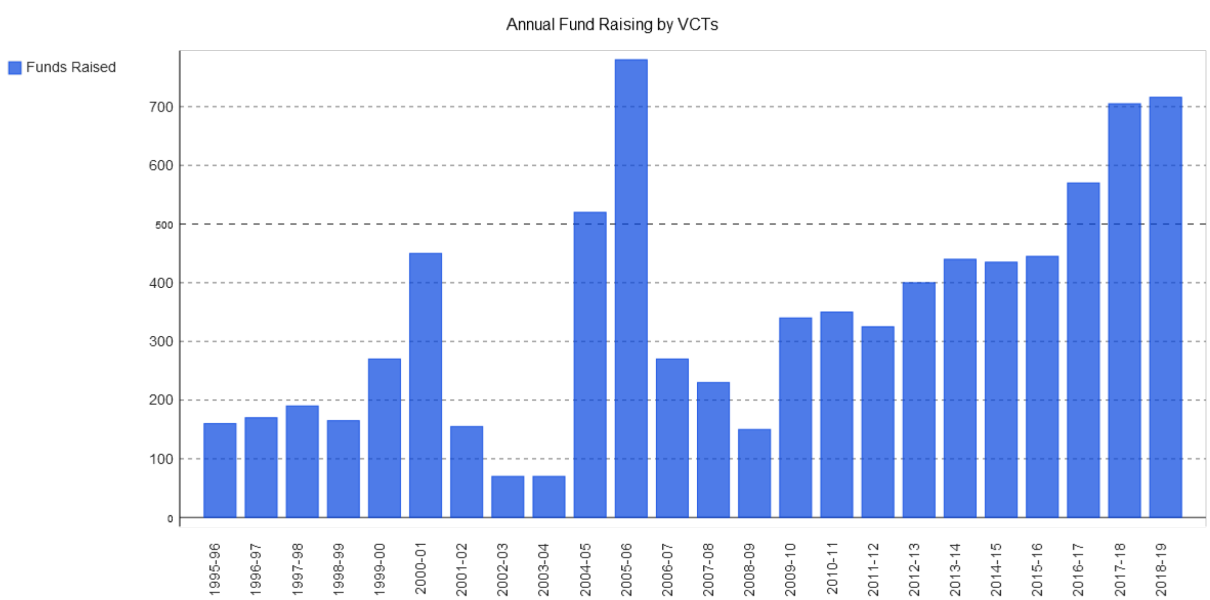

Figure 5: Annual Fund Raising by VCTs (£ Million)

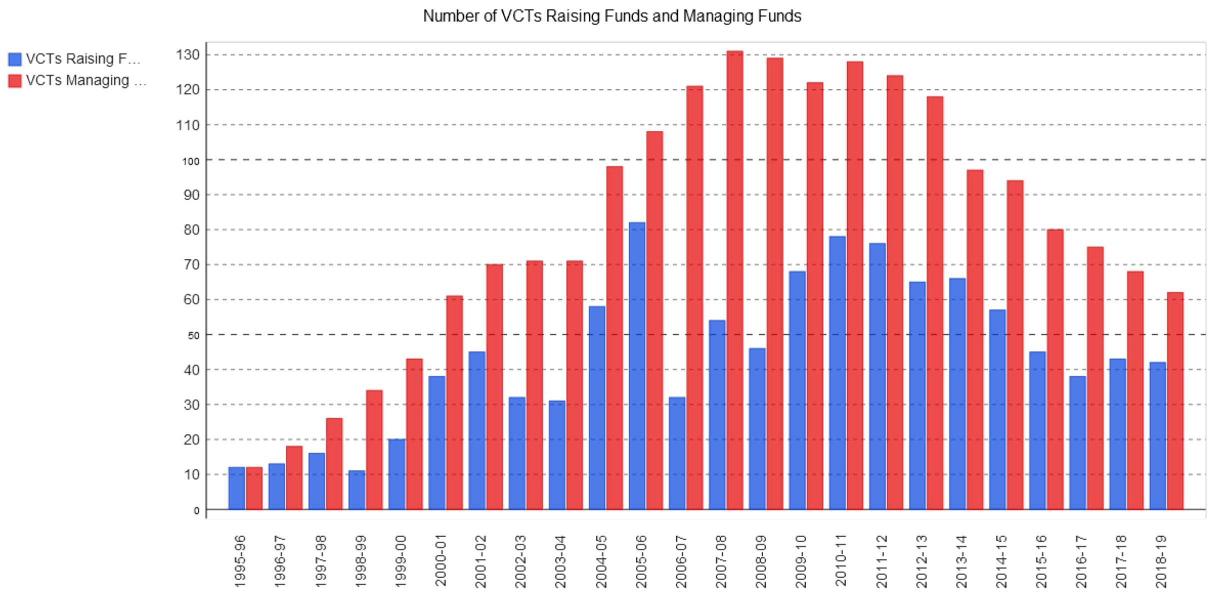

Figure 6: Number of VCTs Raising and Managing Funds

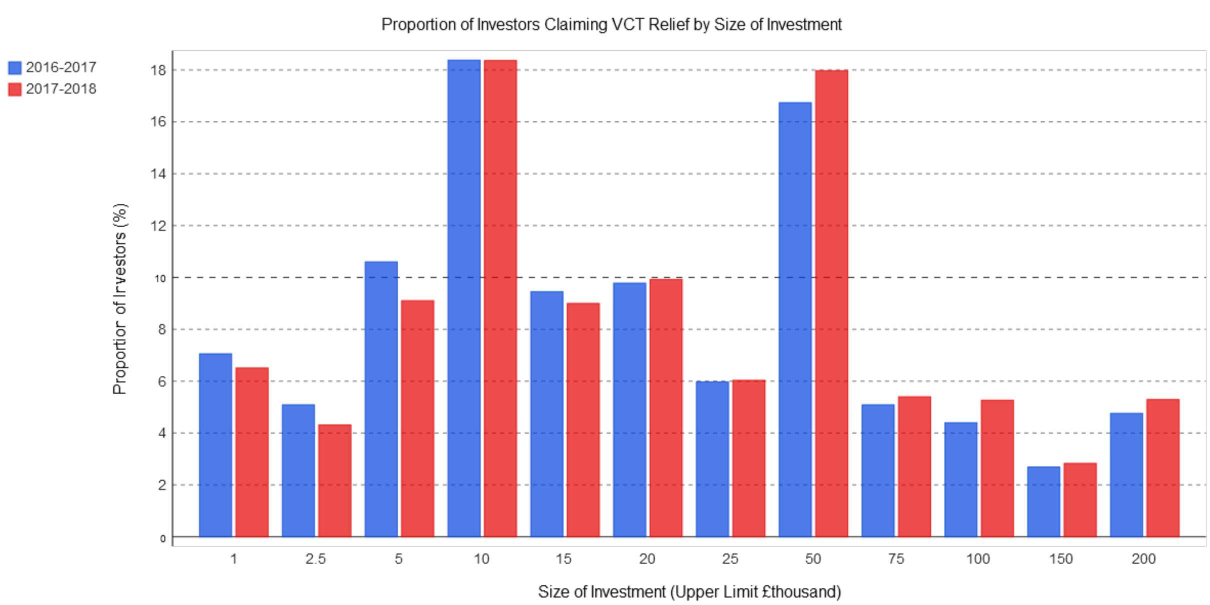

Figure 7: Distribution of Proportion of Investors Claiming VCT Relief by Size of Investment

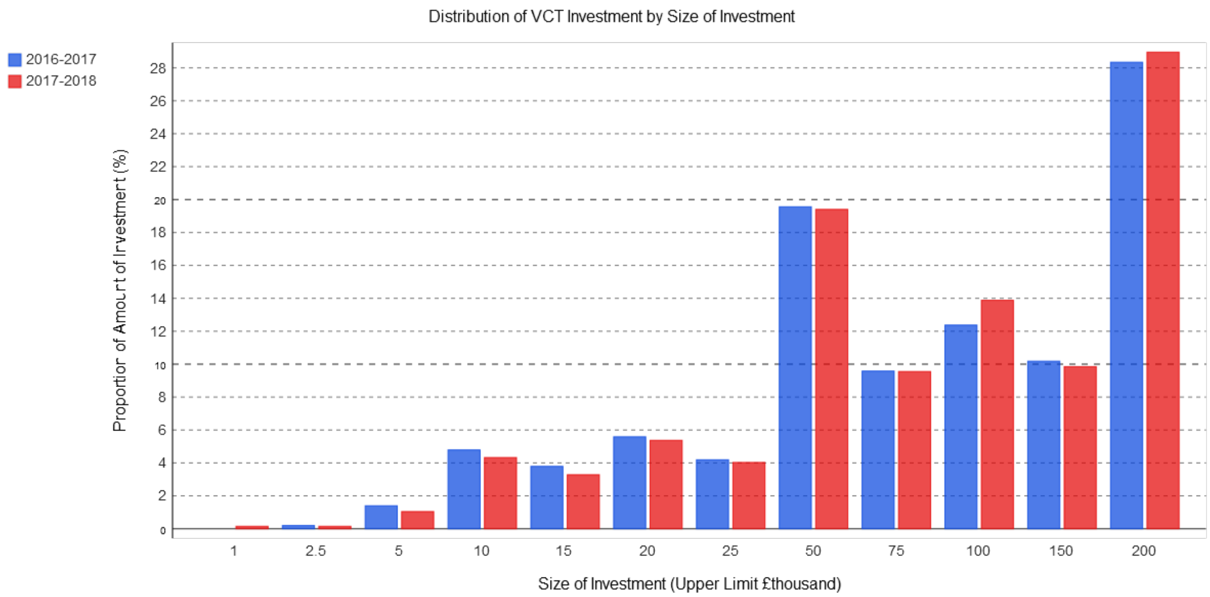

Figure 8: Distribution of Proportion of Investment for which VCT Relief was Claimed by Size of Investment

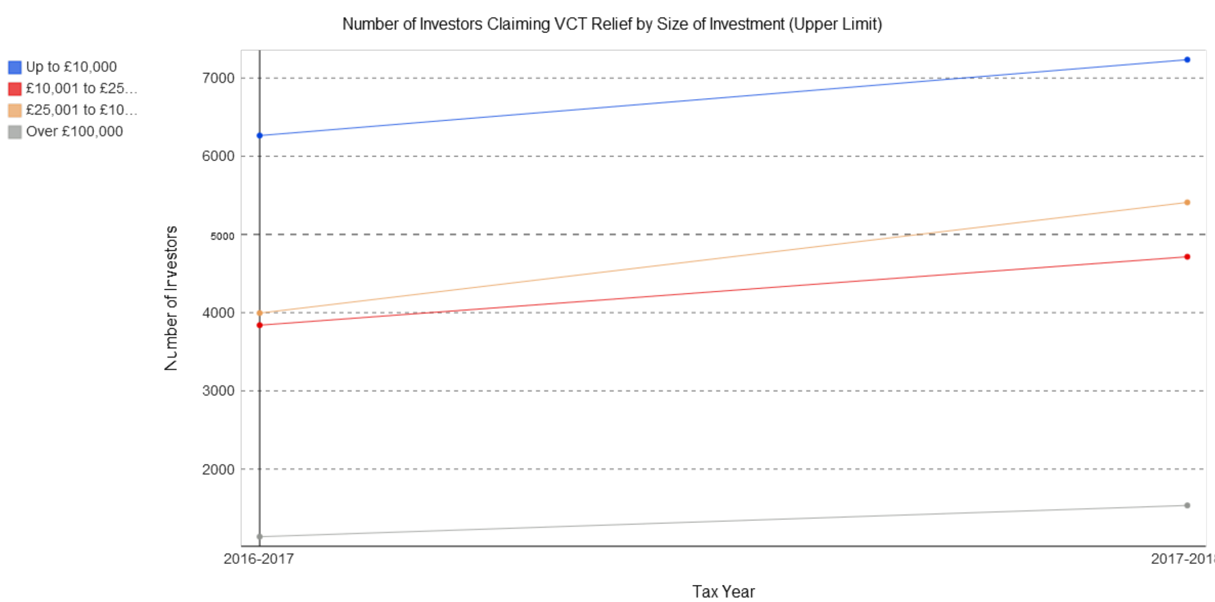

Figure 9: Number of Investors Claiming VCT Relief by Size of Investment

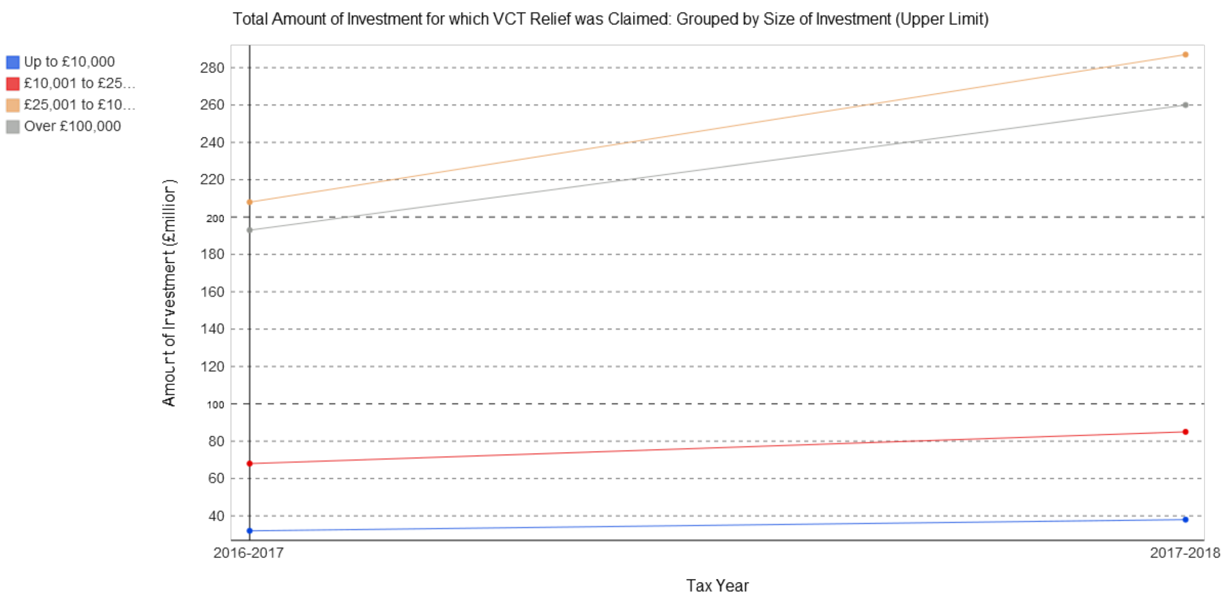

Figure 10: Total Amount of Investment for which VCT Relief was Claimed: Grouped by Size of Investment (Upper Limit)

6 Main Results

6.1 Counterfactual vs Observed Trends

In this section, we present the average treatment effect on the treated (ATET) we derived through the matrix completion method of imputing missing values in an incomplete matrix. This is the average causal effect of the VCT scheme on investment in the U.K economy - via its impact on the corporate asset formation of investees.

Given our approximated panel dataset/complete matrix (Y) of annual total-assets for all investee firm-years between 2003-2018 - wherein we reiterate that the entries for investees prior to their treatment are unchanged in the approximated Y matrix - we remind our reader that the ATET is calculated as:

$$\zeta = \frac{\Sigma_{i,t:W_{it}=1}[Y_{it}(1) - Y_{it}(0)]}{\Sigma_{i,t} W_{it}}, \tag{17}$$

where $Y_{it}(0)$ is the counterfactual investment for investee i in treatment period t. We calculate this as the difference in the natural logarithm (ln) of approximated total-asset value for firm i - between treatment period t and treatment period/period t - 1. i.e.

ln(total-assets)$_t$ - ln(total-assets)$_{t-1}$.

$Y_{it}(1)$ is the observed investment for investee i in period t, also calculated as

ln(total-assets)$_t$ - ln(total-assets)$_{t-1}$. W is the number of investees multiplied by the T number of years.

For cognitive ease purposes, all of our plots henceforth are colour coded. Red always depicts investees, blue always depicts non-investees. Where a plot depicts investee vs investee, red will depict counterfactual whilst blue will depict observed values. Fig.11. and Fig.12. depict two relationships. Fig.11. depicts the relationship between the counterfactual and observed investment for treated firms (investees). Fig.12. captures the relationship between the counterfactual investment for treated firms' (investees) vs the observed investment for the bottom 1931 non-investee firms.

The first plot has two Y-axes for illustrative purposes with both Y-axes in percentages. Given that we transformed the numbers to natural logarithm in our matrix completion algorithm - for reasons already specified - the percentages in and of themselves are not as important as the trends in the plot.

There is an ostensible divergence between the observed and counterfactual investment for investees - as seen in Fig.11. From the beginning in 2004 to 2007, we notice a precipitous drop in the counterfactual investment relative to the observed investment - red v blue o marker. From 2007 to 2009, we see a reversal in the trend of our counterfactual investment - in contrast to the observed values, which drop at an alarming rate in the same period. Within the period 2009-2013, both the counterfactual and observed values display similar trends - albeit a lagged trend for the observed values. Between 2016 - 2018, we see another notable contrasting trend between the counterfactual and observed investment. The counterfactual trend displays a steep decline whereas the observed values display solid growth.

The pertinent question then becomes: what is driving the predominantly divergent trend in our observed vs counterfactual values ? What is our Matrix Completion algorithm capturing in the counterfactual values. For that, we turn to Fig.12.

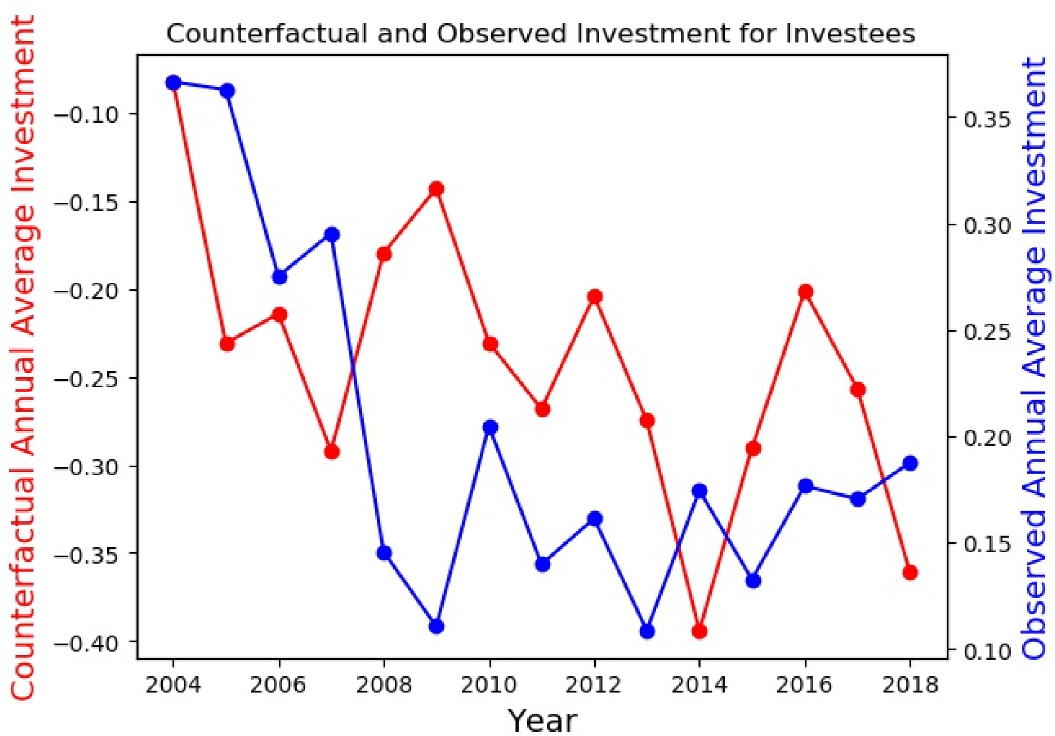

Figure 11: Counterfactual and Observed Investment for Investees

The blue o marker in Fig.12. is the observed annual average investment for the bottom 1931

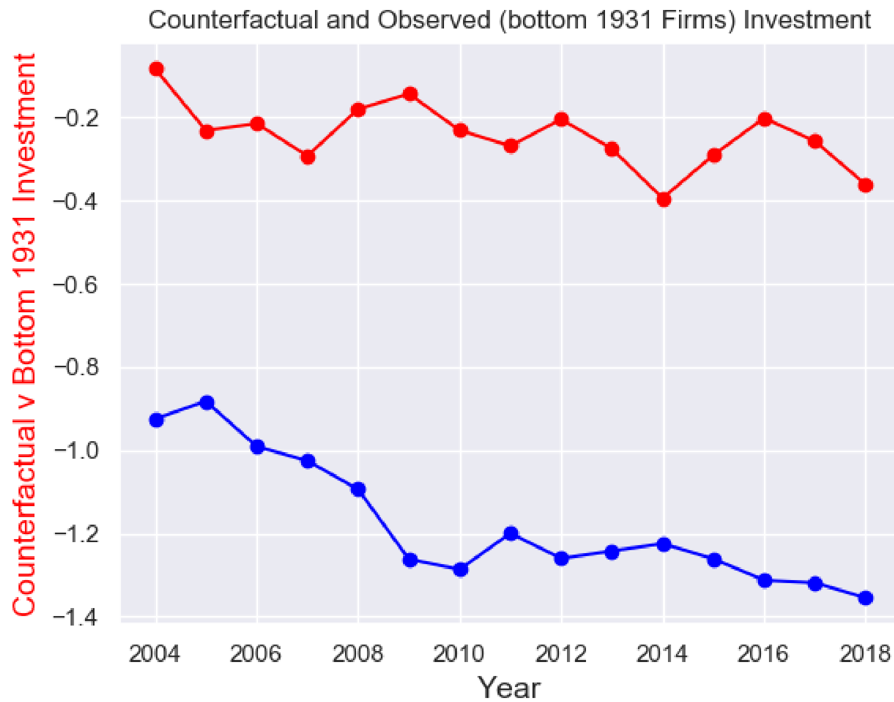

Figure 12: Counterfactual and Observed (bottom 1931 Firms) Investment

non-investee firms by investment rate between 2004-2018. We employ the following process to create the annual average investment for the bottom 1931 firms. For 2004, we select the bottom 1931 firms by investment rate then take the average. This represents the first data point on the blue line in Fig.12. We repeat the same process for each year to form the entire blue line. Note that the constructed panel data of bottom 1931 firms does not necessarily include the time-series of any one firm. We have simply constructed a panel data of the smallest 1931 firms by investment rate per annum. Firm A might be in the 2004 bottom 1931 firms, but not in the 2005 bottom 1931 firms. The red o marker is the same as in Fig.11; the counterfactual investment. Notice how the trends are similar between both. There is a general downward trend from the start of the sample till the end in both cases. The downward trend in the observed values are relatively more dogmatic in the sense that the decline in the annual investment rate is relatively steeper - relative to the red line. Now, why is all of this important. Well, our Matrix Completion algorithm through which the counterfactual values were approximated (as discussed earlier), was “fed” the data on all firms’ in the U.K - which includes both investee and non-investee firms. The algorithm teased out structural patterns in the data, patterns that on average, are shared by small firms, but less so for investees (which are firms' that differ in that they received VCT funding at one or numerous periods during the sample period). In other words, the counterfactual values are picking up the average trend in the observed data for the universe of U.K small firms, whereas, the observed values for a particular subset of small firms (investee firms') deviates from the average values for non-investee small firms.

6.2 Main Result

We now turn to presenting our main findings. Fig.13. is a plot of our main result - also tabulated in Table. 1. It depicts the annual average treatment effect on the treated (ATET). This effectively captures the average difference between the observed vs counterfactual values for investees as given in Eq.1.17 and Fig.11.

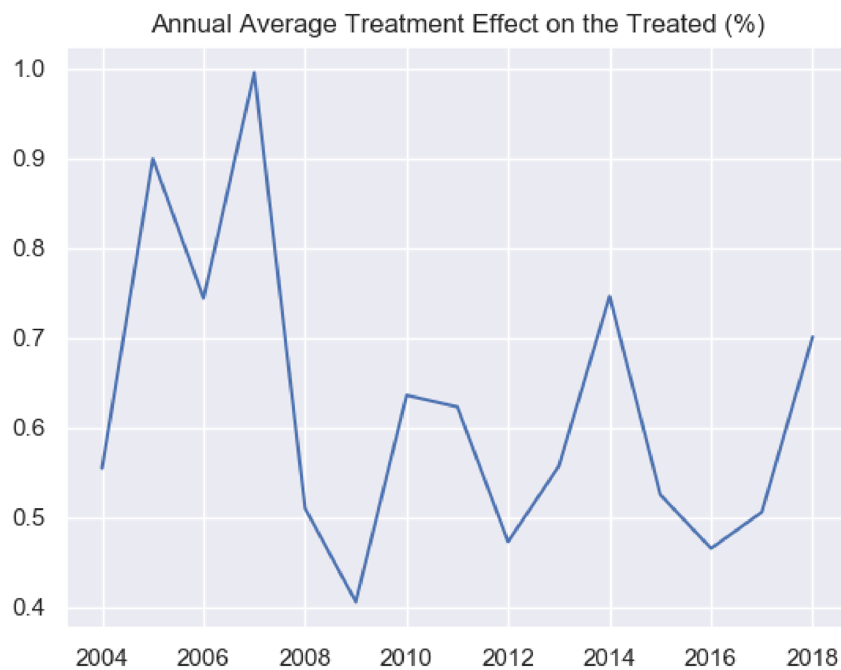

Figure 13: Annual Average Treatment Effect on the Treated

Table 1: **Annual Average Treatment Effect on the Treated**

Year	Average Treatment Effect on the Treated
2004	55.48%
2005	89.94%
2006	74.41%
2007	99.55%
2008	51.00%
2009	40.56%
2010	63.60%
2011	62.30%
2012	47.27%
2013	55.70%
2014	74.64%
2015	52.51%
2016	46.51%
2017	50.58%
2018	70.13%

Between 2004 - 2007, we see from the table that there was a substantial aggregate increase in corporate asset formation (investment) for investee - as a result of the VCT scheme - from 55.48% to 99.55%. 2007 heralds the beginning of a precipitous drop in VCT-induced investment. A drop that reaches its nadir in 2009 at 40.56%. Thereafter, we see a slightly sustained rise in investment as a result of the scheme, one that peaks in 2014 at 74.64%. From 2014, we have another sustained downward trend which lasts until 2017. Thereafter, the trend reverses - increasing from its 2017 value of 50.58% to 70.13% in 2018.

The average treatment effect on the treated (ATET) is the average of the values in Table.1.1. The ATET is 60.69%. This means, the VCT scheme led to a 60.69% increase in investment for VCT-backed firms' in the U.K. An important inquiry into this 60.69% increase is: At what cost has this increase come ? We will address this in a subsequent section (Additional Results).

The obvious question that arises out of observing Fig.1.13/Table.1.1. is: What is driving the fluctuating trends in Annual ATET ? There are two parts to the Annual ATET trends. The counterfactual and observed trends. Thus far, we have explicated on where the counterfactual trends come from. We now turn to Fig.1.14. to analyse the drivers of the observed trends for investees.

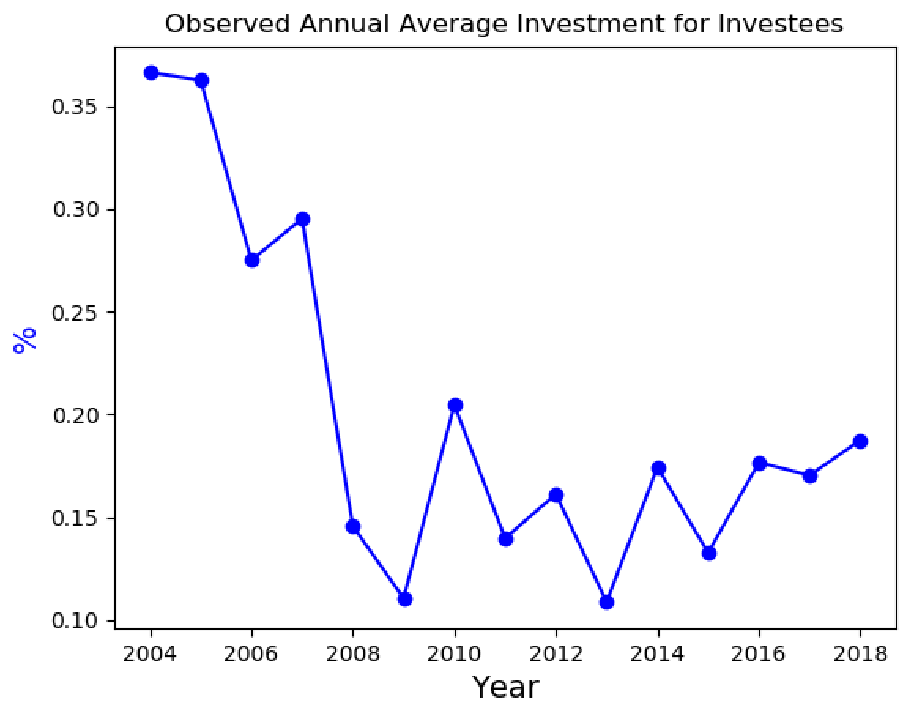

Figure 14: Observed Annual Average Investment for Investees

To analyse Fig.1.14., we will link the historic changes in the policies guiding the VCT scheme to the trends observed in the plot. Firstly and most importantly, we study and analyse the annual report of each VCT operating in the U.K - from the inception of the VCT scheme till present day. This is in addition to analysing policy changes as issued by HMRC. Finally, we also analyse how each VCT reacted to the HMRC-issued policy changes - with regards to their investment policy.

6.3 Major VCT Policy Changes

In this section, we will link changes in the observed annual average investment for investees depicted in Fig.1.14 to major VCT policy changes within our sample period (2004-2018). Text in Bold-Italics is our analysis of Fig.1.14 based on reading through each VCTs annual report in tandem with the major VCT policy changes.

- 2004-06: 6th April 2004 - Introduction of the 40% income tax relief rate for a two-year period starting on 6 April 2004 - prior to which income tax relief was given at 20%.
  Also, from 6th April 2004, the maximum amount individual investors could invest in VCTs to qualify for income tax relief increased from £100,000 to £200,000.
  However, the holding period - to keep your income tax relief - for VCT shares held by investors increased from three to five years.
  We attribute the highest points (2004-2005) in our Fig.1.14. of aggregate annual investment to the increased income tax relief. Our assertion is backed by the 244% average increase in the amount of funds raised in both 2004 and 2005 relative to the average raised in the two years prior (See Table 1.2). In the aggregate, VCTs attributed the high levels of funds raised and the subsequent high level of investment to the increased income tax relief.

- 2006-07: 6th April 2006 - The maximum gross assets of qualifying investees was reduced from £15 million to £7 million before investment and from £16 million to £8m immediately after investment. Also, the rate of income tax relief was reduced to 30% from 40%.

- 2007-08: 6th April 2007 - VCT qualifying investees must be firms with fewer than 50 full-time employees at the time shares are issued. 19th July 2007 - Investees can only raise a maximum of £2 million in any 12 month period under any or all of the tax-based venture capital schemes (Venture Capital Trusts, Enterprise Investment Scheme).
  Again, our analysis of the annual reports of each VCT managing funds within the 2006-2008 period reveals that the reduction in the rate of income tax relief - from 40% to 30% depressed their fund-raising activities within the period. Most importantly - as explicitly reported by VCT investment manager's in their annual reports - the reduction in the size of qualifying investees increased the risk profile of potential investees and further depressed their investment activities. All of this largely [16] explains the sustained downward trend in investment between 2006 - 2009 as seen in Fig.1.14. Our explanation

[16] We hedge by using the adverb "largely" because this time period also coincides with the height of the financial crisis and the attendant bear market. This however is an area we will not explore in this study.

also bears out in the numbers in Table.1.2. We see that the number of VCTs raising funds as a proportion of those managing funds drops from 68% in 2004-2006 to 34% in 2006-2008

- 2009-10: Capital raised by VCTs in a share issuance should be fully employed within two years of the issuance. However, if the issue takes place before commencement of the intended trade, then the capital raised should be fully employed within two years of commencement.
  Our analysis reveals that the 2009-10 major policy change did not drive the upward trend seen in the same period; Fig.1.14. During the period, VCT investment managers document their concerns about the impact of the economic downturn and tightened lending conditions on SMEs. They however saw this as an opportunity to further invest in their existing portfolios; tightened lending conditions meant VCTs were one of the few sources of financing for investee firms: through the provision of working capital to investees, by funding acquisitions carried out by investees, and funding the restructuring of investees existing stakeholders.
  Thus, we see from Table.1.2. that even though fundraising in the period was at a three-year high, the number of new investees that received VCT-funding was the lowest it had been since 2003 (See Fig.1.2.). This means, more money was being raised by VCTs relative to the last three years, but fewer new investees were receiving said funds. Therefore, the data in Fig.1.2. backs up the documented claim that VCTs viewed the tightened lending conditions for SMEs as an opportunity to solidify their existing positions under favourable terms, and hence, a large proportion of the three-year-record-breaking newly raised funds went to existing investee firms.

- 2010 - 2011: 6th April 2011 - VCTs must hold at least 70%, by VCT tax value , of its total investments (shares, securities and liquidity) in VCT qualifying holdings, within approximately three years of a fundraising. For VCTs whose accounting periods begin on or after 1 January 2020, this percentage increased to 80%. From that date, total investments also includes funds raised up to 31 December 2017.
  Also, a VCT can only invest a maximum of £1m per tax year in each of its investees, and no investment in a single investee or group of investees may constitute more than 15% (by

VCT tax value) of the VCTs total investments at the date of investment.

- 2011-12: For funds raised before April 2011: at least 30% of a VCTs qualifying investment by value must be held in "eligible shares" (do not carry any preferential rights). For qualifying investments made by VCTs after 5 April 2018, together with qualifying investments made by funds raised after 5 April 2011, they must in aggregate be comprised of at least 70% by VCT tax value in "eligible shares".
  At least 10% of each investment in qualifying investees is held in eligible shares (by cost at the time of investment).
  VCTs income must come wholly or primarily from shares and securities.
  VCTs must distribute sufficient revenue dividends from its revenue available for distribution so as not to retain more than 15% of its income from shares and securities in a year.
  VCTs must be listed on a UK recognised Stock Exchange.
  The requirement that a potential investees main trade be carried on wholly or mainly in the UK was cancelled, and replaced with a requirement that the investee have a permanent establishment in the UK.
  The restriction that prevented VCTs from investing more than £1m per annum in any single investee was also removed.

- 2012-13: 6th April 2012: The 2007 restriction on VCT qualifying investees having a maximum of 50 employees is increased to a maximum of 250 full time equivalent employees. Also, the 2006 reduction in gross assets of VCT qualifying investees was reversed. VCTs can once again invest in firms with a maximum gross assets of £15 million before investment and £16 million after investment.
  Additionally, the rule that an investee firm is restricted to an annual VCT investment limit of £2m - imposed in 2007 - is increased to £5 million, with a lifetime limit of £12 million (for knowledge intensive companies the annual limit is £10 million and the lifetime limit is £20 million).
  Regarding investments made by a VCT from capital it raised on or after 6 April 2012, if an investee firm uses the funds to acquire shares in another company, this will not be considered as using them for a qualifying purpose.
  The main theme of the policy changes between 2010-2013 was a reversal of the 2006-

2007 changes. These reversals were introduced to stimulate VCT fundraising and subsequent investment in UK SMEs.
However, all of the investment managers expressed concern in their annual reports about an uncertain and fragile UK economy. The main highlights of their concern was the sovereign debt crisis in the eurozone, upward inflationary pressures, and a sustained downward pressure on public sector spending.
These reasons help explain the downward trend we see in the period in Fig.1.14.

- 2014-15: From April 2014 VCTs can no longer return share capital to investors within three years of the end of the accounting period in which the VCT issued the shares.
  Additionally, legislation was introduced to prevent investors refreshing income tax relief on investments into VCTs by disposing of VCT shares and reinvesting the proceeds in new shares. The legislation allowed new investment into VCTs to still be eligible for income tax relief. However, investments that are:

  – conditional on a share buy-back or made within a six month period of a sale of shares in the same VCT, will not qualify for income tax relief. The measure does not affect subscriptions for shares where the monies being subscribed represent dividends which the investor has elected to reinvest. The legislation was also changed to allow individuals to subscribe for shares in a VCT via a nominee.
    These major policy changes are responsible for the downward trend depicted in the period in Fig.1.14.

- 2015-16: 8th July 2015 - Policy changes were introduced to bring the VCT scheme in line with the European Union's risk capital guidelines:

  1. VCTs may not: offer secured loans to investees, and any returns on loan capital above 10% must only represent a commercial return on the principal; invest in investees that do not meet the new "risk to capital" condition (which requires an investee, at the time of investment, to be an entrepreneurial company with the objective to grow and develop, and where there is a genuine risk of loss of capital).
  2. Restrictions on investments that VCTs can make, particularly with respect to the age of the business. Potential investees have been limited to firms that are less than 7

years old (ten years for knowledge intensive businesses).

Non-qualifying investments can no longer be made, except for certain exemptions in managing the Company's short-term liquidity. Exemptions are limited to investments in firms such as OEICs (Open Ended Investment Company), Investment Trusts or listed firms.

Investment managers report that this policy change (No's 2) will curtail their investment in Alternative Investment Market (AIM) shares; AIM shares form a significant proportion of VCT portfolio holdings. This line of reasoning is clearer when we consider that the London Stock Exchange require that firms' be at least 3 years old before they can registered on the AIM.

VCTs further interpret this particular policy change as likely to reduce the scope of investments they can make, potentially increasing the risk profile of their portfolio's. For instance, they claim that replacing the shares of AIM firms with that of smaller unquoted firms will increase the risk profile of their portfolios.

3. Ban on using funds raised by VCTs to finance management buy-out (MBO), Buy-In, Management Buy-Out (BIMBO), or company acquisitions. ***Investment managers report that this will eliminate the lower risk component of their portfolios***.

4. These policy changes were introduced with a ten-year sunset clause - providing a decade of stability with regards VCT policy changes.

In summary, we have two countervailing forces affecting VCTs. On the one hand, the narrower set of investment opportunities (No's 2, 3, 4) could potentially depress investment activity. To paraphrase the sentiments of numerous investment managers "These new inhibitions will curtail significant drivers of growth in the UK SME sector. They will curtail as opposed to encourage investment activity". On the other hand - and this sentiment was also explicitly expressed by VCT investment managers in their annual report - there is a high demand for VCTs to fund-raise as a result of a reduction in the pension lifetime allowance from £1,250,000 to £1,000,000, the tapering away of pension tax allowances for high earners earning £110,000 a year or more, which can gradually reduce said person's annual allowance from the standard £40,000 to as low as £10,000[17]***,***

[17] Prior to 2009, high earners could save up to £235,000 a year in a pension and receive nearly £100,000 in tax

and the launch of pension freedoms that allow for cash to be taken out of the pot for investment rather than buying an annuity. All of this has caused VCTs to become more attractive to investors seeking additional tax-advantaged investments.
The tax-advantage phenomena clearly dominated the narrower set of investment opportunities phenomena, and helps explain the upward trend we see in investment beginning in 2015 till the end of our sample in 2018.
What is clear from the major VCT policy changes between 2015-18 is the Government's desire to refocus investment towards young growth companies. We have argued that these changes have been successful in stimulating new investment, especially the ban on funding MBOs, BIMBOs and acquisitions. To reiterate the point, our reader might have noticed that prior to 2015, periods of rising growth in the rate of investment always preceded or followed periods of falling growth in investment. However, since 2015, the growth in investment has been trending upward.

- 2017-2018: Patient Capital Review:
  In the November 2017 budget, the U.K. Government reviewed the VCT scheme as part of its wider "Patient Capital Review"[18]. The outcome was a number of proposed changes to the VCT regulations in an effort to refocus investment on potentially higher risk sectors that require capital (Her Majesty's Treasury Policy Paper, 2017) - summarised below:

  1. Expand the VCT scheme to enable VCTs provide follow-on investment which will help to "scale-up" investees, thus easing the transition from a dependence on VCT funding to venture funding. For instance, increasing the current Knowledge Intensive Company allowance would help increase focus on science based firms.
  2. Increasing the annual and lifetime investment limits would allow for follow on investment from VCTs, thus slowing the transition away from tax-incentivised financing (Her Majesty's Treasury Policy Paper, 2017).

Anticipation of the above changes from the Patient Capital Review also influenced the

relief. As of 6th April 2016, that sum is limited to £10,000 in a pension and just £4,000 in tax relief,

[18]The review considered how to support innovative firms to access the finance that they need to scale up. Her Majesty's Treasury published a consultation seeking views on how to increase the supply of capital to growing, innovative firms.

increased growth rate in investment between 2017-2018.

- April 2020: Minimum of 80% of a VCTs funds must be invested in VCT qualifying investments - up from 70%.

Table 2: **Amount of Funds Raised and Number of VCTs. Amount:Millions. Number:Actual. Data from HMRC VCT Statistics (2018).**

Year	Funds Raised	VCTs Raising Funds in the Year	VCTs Managing Funds	Income-
	Amount	Number	Number	Tax Relief (%)
1995-1996	160	12	12	20%
1996-1997	170	13	18	20%
1997-1998	190	16	26	20%
1998-1999	165	11	34	20%
1999-2000	270	20	43	20%
2000-2001	450	38	61	20%
2001-2002	155	45	70	20%
2002-2003	70	32	71	20%
2003-2004	70	31	71	20%
2004-2005	520	58	98	40%
2005-2006	780	82	108	40%
2006-2007	270	32	121	30%
2007-2008	230	54	131	30%
2008-2009	150	46	129	30%
2009-2010	340	68	122	30%
2010-2011	350	78	128	30%
2011-2012	325	76	124	30%
2012-2013	400	65	118	30%
2013-2014	440	66	97	30%
2014-2015	435	57	94	30%
2015-2016	445	45	80	30%
2016-2017	570	38	75	30%
2017-2018	705	43	68	30%
2018-2019	716	42	62	30%
Total	8,375			

6.4 Additional Results

We now turn to presenting some simple plots and statistics to show how the investment pattern of investees compare to that of non-investee firms. In Fig.1.15. we plot the observed average investment for investee firms vs the observed average investment for all non-investee firms in the U.K. We observe an ostensible difference between the investment patterns of investee vs non-investee firms. Not only do investees - in the aggregate - invest at a much higher rate than non-investee firms, we also observe divergent aggregate patterns since 2009. For instance, from 2015 onward, the aggregate investment trend of investee firms (red line) has been steadily rising, whereas that of non-investee firms has steadily fallen. We however note the very similar declining investment trend for both investee and non-investee firms between the period 2004-2009.

Now our reader might wonder if aggregation i.e. plotting averages, masks other patterns that exist in the data for non-investee firms. Now this is a valid concern considering the firms in our sample range from the smallest firm with £1000 in total-assets, to the largest with £20 billion in total-assets. To allay this concern, we repeat Fig.1.15 with one crucial change. We instead plot investee firms vs the annual top 1931 non-investee firms. What do we mean by this? For 2004, we select the top 1931 firms by investment rate, then take the average. This represents the first data point on the blue line in Fig.1.16. We choose the top 1931 because the number of investee firms in our sample is 1931. We repeat the same process for each year to form the entire blue line. Note that the constructed panel data of top 1931 firms does not necessarily include the time-series of any one firm. We have simply constructed a panel data of the largest 1931 firms per annum. Firm A might be in the 2004 top 1931 firms, but not in the 2005 top 1931 firms. We construct this panel data to allay the concern that the total-asset size distribution of non-investee firms is too wide for aggregation to reflect the entire sample.[19]

We see a slightly different pattern in Fig.1.16 relative to Fig.1.15. Between 2004-2011, the top 1931 non-investee U.K firms had an ostensibly higher trend in their investment pattern relative to investee firms. However, from 2011, we see that the aggregate investment pattern of these non-investee firms begins to drop very sharply - a drop that carries on to the end of the sample in 2018. On the other hand, we see a continuously increasing aggregate investment rate for investee

[19]Our analysis still stands even if we extend the number of non-investee firms to the top 20,000 non-investee firms.

firms beginning in 2013 - albeit with a drop in 2015 - till the end of our sample. We have already tied this increased investment rate to the major VCT policy changes in the period, so we won't belabour the point.

For completeness, we also repeat the same exercise for the bottom instead of the top 1931 firms as depicted in Fig.1.17. This plot is also very interesting in the dynamic it depicts. There is an ostensibly similar trend in the aggregate investment pattern of investee firms and the bottom 1931 non-investee firms between 2003-2014. However, and similar to the top 1931 non-investee firms, we see that the bottom 1931 non-investee firms have been investing at a declining rate from 2014 till the end of our sample in 2018 - diverging from the investment pattern seen with investee firms in the same period.

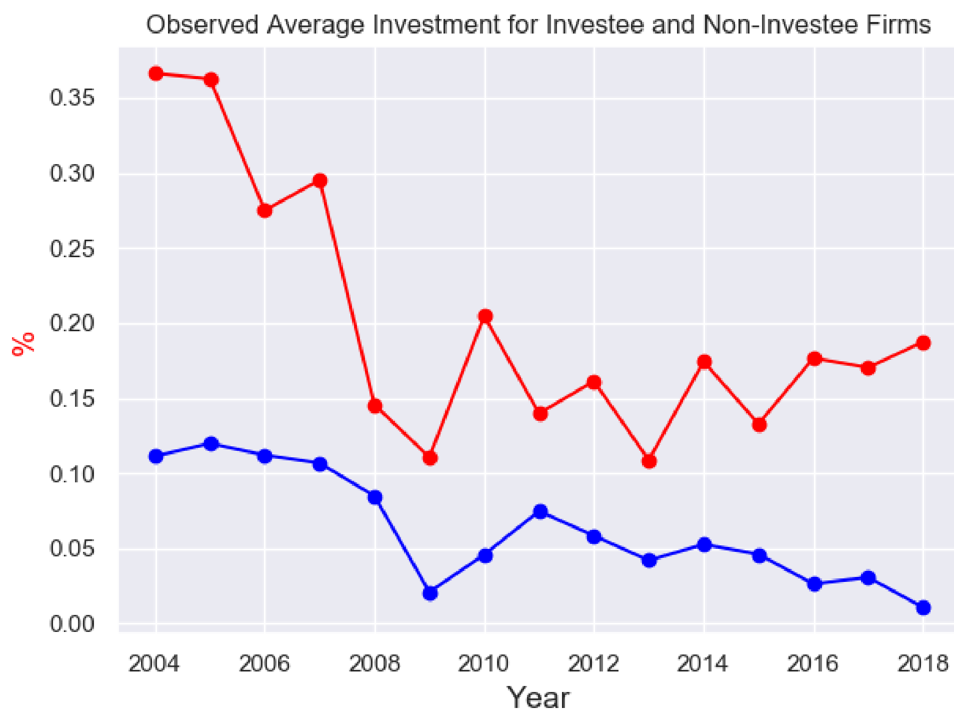

Figure 15: Observed Average Investment for Investee and non-Investee Firms

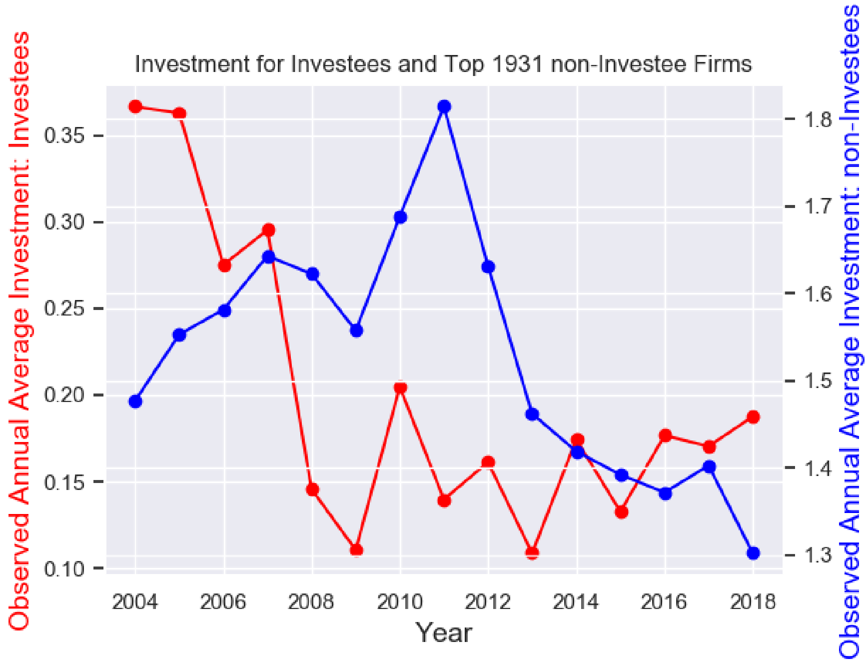

Figure 16: Observed Average Investment for Investee and Top 1931 non-Investee Firms

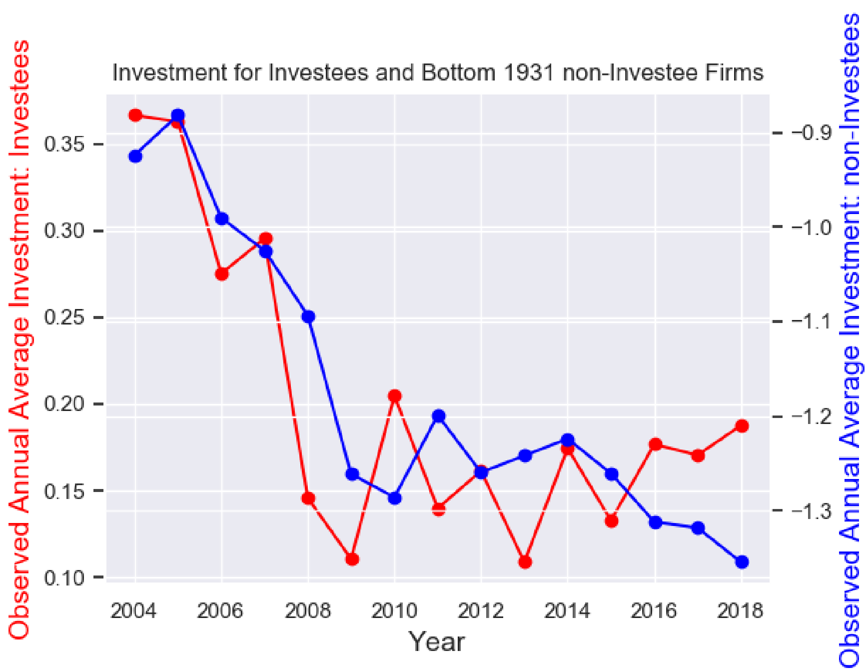

Figure 17: Observed Average Investment for Investee and Bottom 1931 non-Investee Firms

6.5 Cost to Taxpayers

From Table 1.2. we observe that circa £8.4 Billion pounds has been raised by over 200 VCTs [20] since inception of the VCT scheme in 1995. These funds have funded the activities of SMEs and increased corporate asset formation in the UK. Specifically, we have shown that these funds have had a measurable positive impact on investment in the U.K via its effect on VCT-backed firms - an increase in investment by VCT-backed firms of 63.63%. However, we also know that investors in the VCT scheme receive tax breaks such as: 30% upfront income tax relief, tax-free dividends, and exemption from capital gains tax. These tax breaks come at a non-trivial cost to the U.K tax-payer, as it involves the actual reduction of a VCT subscribers tax bill as detailed in subsection (All About VCTs). To illustrate, in fiscal year 2017-2018, the subsidy expenditure for the VCT scheme was £201 million. [21] This figure is extremely conservative as it does not take into account, investors making Income-Tax relief claims through Self-Assessment nor does it consider investors making claims through other systems e.g. PAYE. Also, it does not take into include other tax reliefs and exemption available through the VCT scheme such as capital gains and dividend-tax exemptions. For comparison sake, this conservative £201 million in subsidy expenditure is over half of the total managed expenditure, [22] over the same period, for the Department for International Trade (HM Treasury Public Expenditure Statistical Analysis, 2019), which was £394 million.

In the below table, we present data on the amount of investment on which relief was claimed on an annual basis between 2015-2018. Our reader should note that HMRC does emphasise that the investor-level information in the below table was prepared using Self Assessment (SA) returns. Thus, the information in the table will not cover investors making Income Tax relief claims through other channels (e.g. PAYE) or not making any claims. However, we know these

[20] To avoid duplication, we do not calculate the total for VCTs raising funds in column 3, Table 1.2. as VCTs can raise funds in multiple tax-years. Our total figure of 200+ comes from our hand-collected data on VCTs.

[21] £607 million - total from the last column in Table.1.3. which is the the amount of investment on which VCT investors claimed tax relief in the 2017-2018 fiscal year - multiplied by 30% tax relief.

[22] The total managed expenditure is the total amount the government spends. This is split up into: departmental budgets – the amount that government departments have been allocated to spend, also known as Departmental Expenditure Limits, and money spent in areas outside budgetary control – all spending that is not controlled by a government department and includes welfare, pensions and things such as debt interest payments, also known as Annually Managed Expenditure.

omissions are small - because we can compare them with the amount of funds raised by VCTs in the corresponding year (Table 1.2)

Table 3: **Venture Capital Trusts**

Income tax relief; distribution of investors and amount of investment on which relief was claimed from 2015-16 to 2017-18.

Data from HMRC VCT Statistics (2018).

(Upper Limit: £)	2015-2016		2016-2017		2017-2018	
	Investors	Amount of Investment (£ million)	Investors	Amount of Investment (£ million)	Investors	Amount of Investment (£ million)
1,000	1,240	1	1,075	0	1,230	1
2,500	630	1	775	1	815	1
5,000	1,365	6	1,615	7	1,720	7
10,000	2,545	22	2,800	24	3,470	29
15,000	1,195	16	1,440	19	1,700	22
20,000	1,285	24	1,490	28	1,875	36
25,000	775	18	910	21	1,140	27
50,000	2,155	83	2,550	98	3,395	130
75,000	640	40	775	48	1,020	64
100,000	635	60	670	62	995	93
150,000	330	41	410	51	535	66
200,000	620	122	725	142	1,000	194

6.6 Comparison of Results and Selection Bias

Our objective in this section is to highlight the performance of our matrix completion estimator at imputing the missing entries of our incomplete matrix of total-assets for VCT-backed firms (Y). We highlight its performance relative to a popular causal inference estimator used in the literature; the Difference-in-Differences (DID) estimator. This exercise has the added bonus of alleviating the clear selection issues inherent in a simple DID approach to estimating the ATET. Specifically, we predict that the ATET should be much lower, when estimated with the DID approach - relative to our matrix factorisation approach. Also, our aim is not necessarily to pinpoint the right or wrong algorithm. We simply want to uncover which algorithm works best in our causal potential outcome with staggered adoption setting. The DID estimator is constructed by regressing our observed outcome variable (total-assets) on firm and time fixed effect, plus a dummy for whether a firm received VCT-funding. Let us elaborate.

To construct our DID estimator in a comparable manner (comparable to our Matrix Completion estimator), we start with the observed data of total-assets for all firms Y. Our data has a $N \times T$ (where N is the the number of firms and T is the number of years in our sample) matrix form, which we partition into

- Y_* is a matrix with dimensions (N - treated firms) $\times$ (T - treated periods).
- y_{**} is a vector array with dimensions (N - treated firms).
- $y_{***}^{\top}$ is a vector array with dimensions (T - treated periods).

Recall, our matrix completion estimator characterises the solution to imputing the missing observations in $Y \in \mathbb{R}^{N \times T}$ as:

$$\hat{L} = \underset{L \in \mathbb{R}^{N \times T}}{\arg\min} \left\{ \frac{1}{2} \|L - Y\|_{\text{Fro}}^2 \ + \ \alpha \|L\|_* \quad \text{subject to } P_{\Omega} L = P_{\Omega} Y \right\},$$

as given in Eq.1.6. Therefore, the predicted value for the missing entries in our Y matrix is:

$$\hat{Y}^{MC} = \hat{L}.$$

The DID estimator is defined as:

$$\hat{\beta}^{DID} = \left(Y_*^{\top} Y_* \right)^{-1} \left(Y_*^{\top} y_{**} \right),$$

therefore the DID based prediction for imputing the missing entries in our $Y \in \mathbb{R}^{N \times T}$ matrix is:

$$\hat{Y}^{DID} \in \mathbb{R}^{N \times T} = y_{***}^{\top} \hat{\beta}^{DID} = y_{***}^{\top} \left(Y_*^{\top} Y_* \right)^{-1} \left(Y_*^{\top} y_{**} \right).$$

The DID estimator performs poorly with an ATET of 27.35% compared to the ATET of 60.7% from our MC estimator. The increased performance of our MC estimator is attributable to its use of additional observations i.e. pre-treatment total-assets of treated firms.

7 Conclusion

In this study, we sought to deepen our understanding of VCTs and estimate the causal effect of VCT-funding on total-asset formation for VCT-backed firms in the U.K. We hand-collected data on all former and the current 62 VCTs operating in the U.K. Specifically - to estimate the causal effect of VCT-funding on investment - we begun by hand-collecting data on all SMEs that ever received VCT-funding since inception of the VCT scheme in 1995.

We thereafter built and employed a Matrix Completion estimator - to estimate our causal effect - an estimator with intuitive computational properties. We found that between 2003-2018, the causal effect of the VCT scheme - the average treatment effect on the treated (ATET) - was 60.69%. We also find in an illustration, that our estimator outperforms a standard difference-in-difference estimator.

This study contributes to three broad spheres in Economics. Firstly, our results add to the literature on the importance of venture capital funding to SMEs, and is consistent with findings in Gonzalez-Uribe and Paravisini (2017), Gompers et al. (2020), and Iliev and Lowry (2020). Secondly, we contribute to the literature on causal potential outcomes. Our Matrix Completion estimator and algorithm enumerated in the paper - popular in the unsupervised Machine Learning literature and adapted for causal potential outcome settings - represents an additional tool in the estimation of causal effects toolkit. This estimator is also consistent with that proposed in Athey et al. (2018). Finally, our results are practically relevant for policy makers. The insights and results we provide can serve as a template to bolster the recommendations of the Patient Capital Review - in light of the current pandemic and its adverse impact on SMEs.

A Bregman Proximal Method

A.1 Set-Up

Let $\hat{w} = \underset{w \in \mathbb{R}^N}{\arg\min} \left\{ \frac{1}{2s} \|Xw - y\|^2 \right\}$ be the compact notation for a linear regression problem. We then have the corresponding optimality condition $\nabla E(\hat{w}) = 0$ which reads as

$$X^\top X \hat{w} = X^\top y \,, \tag{A.1}$$

which is also known as the normal equation associated with $X\hat{w} = y$

A.2 Bregman Proximal Method

If $X^\top X$ is invertible, we can solve (A.1) for $\hat{w}$ with any of the numerous algorithms that are used to numerically solve linear systems of equations. However, if N is very large, the oft-used algorithms that solve (A.1) to exacting numerical accuracy may require substantial computational time and large memory requirements.

We can however settle on approximate solutions of (A.1) by employing iterative algorithms such as gradient descent - which is an iterative procedure of the form

$$w^{k+1} = w^k - \tau \nabla E(w^k), \tag{A.2}$$

for some energy E, step-size parameter $\tau > 0$, and an initial value $w^0 \in \mathbb{R}^N$. For example, when $E(w) = \frac{1}{2s}\|Xw - y\|^2$, gradient descent reads as

$$\begin{aligned} w^{k+1} &= w^k - \frac{\tau}{s} X^\top \left(X w^k - y \right) , \\ &= \left(I - \frac{\tau}{s} X^\top X \right) w^k + \frac{\tau}{s} X^\top y. \end{aligned} \tag{A.3}$$

(A.3) is elegant in its simplicity. Iteratively solving (A.3) simply requires the computation of matrix multiplications and simple arithmetic operations. Additionally, with an algorithm like (A.2), we can deal with minimisation problems more generic than minimising the mean squared error (MSE).

Definition A.1 (Sub-differential)**.** Let $E : \mathbf{C} \subset \mathbb{R}^n \to \mathbb{R}$ be a convex and continuous function. Its sub-differential ∂E is characterised as the set:

$$\partial E(\upsilon) := \{g \in \mathbb{R}^n \mid E(w) - E(\upsilon) \geq \langle g, w - \upsilon \rangle, \ \forall \ w \in \mathbb{R}^n\}.$$

The elements $g \in \partial E(\upsilon)$ are known as sub-gradients.

Definition A.2 (Bregman distance)**.** Let $E : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, i.e. $\nabla E(w)$ exists $\forall\, w \in \mathbb{R}^n$ and is continuous. Then its corresponding Bregman distance $D_E : \mathbb{R}^n \times \mathbb{R}^n$ is defined as

$$D_E(u, \upsilon) := E(u) - E(\upsilon) - \langle \ \nabla E(\upsilon), \ u - \upsilon \ \rangle.$$

$\forall$ arguments $u, \upsilon \in \mathbb{R}^n$. We must emphasise that Bregman distances - defined in terms of a strictly convex function - are not necessarily distances in the sense of a metric. They are a statistical distance when the points are interpreted as probability distributions i.e. data-set of observed values. Bregman distances or divergences describe the distance of a function E at point u to its linearisation around υ, and are non-negative if and only if E is convex. A common Bregman distance is the squared Euclidean distance.

We now turn to deriving when and under what conditions (A.2) actually converges, what it converges to, and how quickly it converges. To maintain generality, we will consider a generalisation of gradient descent known as Bregman proximal method (BPM). This algorithm is based on Definition A.1 and an iterative procedure outlined in Algorithm 2 below.

Algorithm 2: Bregman proximal method

Specify: Energy function E: $\mathbb{R}^N \to \mathbb{R}$, Bregman function J: $\mathbb{R}^N \to \mathbb{R}$, index K

Initialise: $w^0 \in \mathbb{R}^N$

Iterate:

for $K = 0, \cdots, K-1$ **do**

$w^{k+1} = \arg\min_{w \in \mathbb{R}^N} \left\{ E(w) + D_J(w, w^k) \right\}$;

end

return w^K.

To understand how the BPM is supposed to help us minimise an energy E such as $E(w) = \frac{1}{2s}\|Xw - y\|^2$, we must emphasise that the choice of J is critical. For instance, if we choose

$J : \mathbb{R}^N \to \mathbb{R}$ with $J(w) := \frac{1}{2}\|w\|^2$, then solving the minimisation step in Algorithm 2 is just as difficult as minimising E itself. Thus, for our choice of J, we choose

$$J(w) := \frac{1}{2\tau}\|w\|^2 - E(w),$$

where $\tau > 0$ is a positive scalar. Computing the corresponding Bregman distance yields

$$\begin{aligned} D_J(w^{k+1},\, w^k) &= \frac{1}{2\tau}\left\|w^{k+1} - w^k\right\|^2 - \, D_E(w^{k+1},\, w^k), \\ &= \frac{1}{2\tau}\left\|w^{k+1} - w^k\right\|^2 - \, E(w^{k+1}) \, + \, E(w^k) \, + \, \langle \nabla E(w^k),\, w^{k+1} - w^k \rangle. \end{aligned}$$

Inserting our computed Bregman distance into the minimisation step in Algorithm 2 yields:

$$\begin{aligned} w^{k+1} &= \arg\min_{w \in \mathbb{R}^N} \left\{ E(w) \, + \, D_J(w,\, w^k) \right\}, \\ &= \arg\min_{w \in \mathbb{R}^N} \left\{ E(w^k) \, + \, \langle \nabla E(w^k),\, w - w^k \rangle \, + \, \frac{1}{2\tau}\left\|w - w^k\right\|^2 \right\}, \\ &= \arg\min_{w \in \mathbb{R}^N} \left\{ \langle \nabla E(w^k),\, w \rangle \, + \, \frac{1}{2\tau}\left\|w - w^k\right\|^2 \right\}. \end{aligned}$$

The objective function $L^k(w) := \langle \nabla E(w^k),\, w \rangle \, + \, \frac{1}{2\tau}\left\|w - w^k\right\|^2$ is convex and differentiable with gradient $\nabla L(w) = \nabla E(w^k) \, + \, \frac{1}{\tau}(w - w^k)$. Thus, the global minimiser can be obtained via $\nabla L(w^{k+1}) = 0$, which yields (A.2). Gradient descent is summarised in Algorithm 3 below.

Algorithm 3: Gradient Descent

Specify: Differentiable, convex function E: $\mathbb{R}^N \to \mathbb{R}$, step-size $\tau > 0$, index K

Initialise: $w^0 \in \mathbb{R}^N$

Iterate:

for $K = 0, \cdots, K-1$ **do**

  $w^{k+1} = w^k \, - \, \tau \nabla E(w^k)$;

end

return w^K.

The pertinent question now is this: Does Algorithm 2 (and by implication Algorithm 3) converge to a minimiser of the objective function E ? If it does, under what conditions does it converge ?

Theorem A.1. (Convergence of Algorithm 2). Let $E : \mathrm{C} \subset \mathbb{R}^N \to \mathbb{R}$ and $J : \mathrm{C} \subset \mathbb{R}^N \to \mathbb{R}$ be convex and continuously differentiable functions. Suppose $\hat{w}$ denotes a global minimiser of E. Thus, the iterates of Algorithm 2 satisfy

$$E(w^K) - E(\hat{w}) \leq \frac{D_J(\hat{w}, w^0) - D_J(\hat{w}, w^K)}{K}, \tag{A.4}$$

and therefore $\lim_{K\to\infty} E(w^K) = E(\hat{w})$.

Before we lay out the proof of Theorem A.1, we verify the following intermediate result.

Lemma B. We adopt the same assumptions as in Theorem A.1, and suppose w^* is defined as $w^* := \arg\min_{w\in\mathbb{R}^N}\{E(w) + D_J(w, \overline{w})\}$. Consequentially, the following identity holds:

$$E(w^*) \;+\; D_E(w, w^*) \;+\; D_J(w, w^*) \;+\; D_J(w^*, \overline{w}) \;=\; E(w) \;+\; D_J(w, \overline{w}).$$

Proof. Assume we can characterise w^* via the optimality condition:

$$0 = \nabla E(w^*) + \nabla J(w^*) - \nabla J(\bar{w}).$$

Taking an inner product with $w^* - w$ then yields:

$$\begin{aligned}
0 &= -\langle \nabla E(w^*), w - w^*\rangle - \langle \nabla J(w^*) - \nabla J(\bar{w}), w - w^*\rangle, \\
&= D_E(w, w^*) - E(w) + E(w^*) - \langle \nabla J(w^*), w - w^*\rangle + \langle \nabla J(\bar{w}), w - w^*\rangle, \\
&= D_E(w, w^*) - E(w) + E(w^*) + D_J(w, w^*) - J(w) + J(w^*) + \langle \nabla J(\bar{w}), w - w^*\rangle, \\
&= D_E(w, w^*) - E(w) + E(w^*) + D_J(w, w^*) - J(w) + J(w^*) + \langle \nabla J(\bar{w}), w - \bar{w} + \bar{w} - w^*\rangle, \\
&= D_E(w, w^*) - E(w) + E(w^*) + D_J(w, w^*) - D_J(w, \bar{w}) + D_J(w^*, \bar{w}),
\end{aligned}$$

which rounds off the proof. □

Proof. **Proof of Theorem A.1** By employing Lemma B for $w^* = w^{k+1}, \bar{w} = w^k$, and $w = \hat{w}$, we have:

$$\begin{aligned}
E(\hat{w}) + D_J(\hat{w}, w^k) &= E(w^{k+1}) + D_E(\hat{w}, w^{k+1}) + D_J(\hat{w}, w^{k+1}) + D_J(w^{k+1}, w^k), \\
&\geq E(w^{k+1}) + D_J(\hat{w}, w^{k+1})
\end{aligned}$$

given the convexity of E and J, which implies $D_E(\hat{w}, w^{k+1}) \geq 0$ and $D_J(w^{k+1}, w^k) \geq 0$.
Therefore, we have $E(w^{k+1}) - E(\hat{w}) \leq D_J(\hat{w}, w^k) - D_J(\hat{w}, w^{k+1})$.

Summing from $k = 0, \cdots, K-1$ then leads to

$$\sum_{k=0}^{K-1} E(w^{k+1}) - KE(\hat{w}) \leq D_J(\hat{w}, w^0) - D_J(\hat{w}, w^K). \tag{B.1}$$

We can also apply Lemma B for $w^* = w^{k+1}, \bar{w} = w^k$, and $w = \hat{w}$ to obtain:

$$\begin{aligned} E(w^k) + \underbrace{D_J(w^k, w^k)}_{=0} &= E(w^{k+1}) + D_E(w^k, w^{k+1}) + D_J(w^k, w^{k+1}) + D_J(w^{k+1}, w^k), \\ &\geq E(w^{k+1}), \end{aligned}$$

given the convexity of E and J, which implies $D_E(w^k, w^{k+1}) \geq 0$, $D_J(w^k, w^{k+1}) \geq 0$, and $D_J(w^{k+1}, w^k) \geq 0$.

We can thus conclude $E(w^{k+1}) \leq E(w^k) \ \forall \ k = 0, \cdots, K-1$, especially $KE(w^k) \ \leq \ \sum_{k=0}^{K-1} E(w^{k+1})$. If we plug this inequality into (B.1), it implies A.4. Given J is convex, we can also estimate:

$$E(w^K) \ \leq \ E(\hat{w}) + \frac{D_J(\hat{w}, w^0)}{K}$$

for a positive constant $D_J(\hat{w}, w^0)$ independent of K, thus concluding both $\lim_{K \to \infty} E(w^K) = E(\hat{w})$ and the proof.

As a little aside, it is clear that showing the convexity of E and J is sufficient to prove convergence of the objective E.

□

Remark 1. It is pertinent to emphasise that Theorem A.1 does more than guarantee the convergence of Algorithm 2. It also gives us a rate of convergence. This rate is $1/K$, which in convex optimisation is emphasised with the big O-notation, i.e.

$$E(w^k) - E(\hat{w}) = O(\frac{1}{K}),$$

which means the left-hand-side is proportional to $1/K$. To illustrate, assume $D_J(\hat{w}, w^0) = 10$, then we will require $K = 1000$ iterations to ensure $E(w^k) - E(\hat{w}) \leq 10^{-2}$ according to Theorem A.1.

References

[1] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. (12831), 2007.

[2] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59, 2012.

[3] Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132, 2003.

[4] United States Small Business Administration. Frequently asked questions: About small business. *U.S. Small Business Administration: Office of Advocacy*, 2018.

[5] Theodore Anderson. *An introduction to multivariate statistical analysis, volume 2*. Wiley New York, 1958.

[6] Joshua Angrist and Steve Pischke. *Mostly Harmless Econometrics: An Empiricists' Companion*. Princeton University Press, 2008.

[7] Manuel Arellano and Bo Honore. Panel data models: some recent developments. In J.J. Heckman and E.E. Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 53, pages 3229–3296. Elsevier, 1 edition, 2001.

[8] Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. Synthetic Difference in Differences. 2019.

[9] Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. 2018.

[10] Susan Athey and Guido W. Imbens. Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption. 2018.

[11] Susan Athey and Scott Stern. The impact of information technology on emergency health care outcomes. *The RAND Journal of Economics*, 33(3):399–432, 2002.

[12] Jushan Bai. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1):135–171, 2003.

[13] Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

[14] Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

[15] Jushan Bai and Serena Ng. Principal Components and Regularized Estimation of Factor Models. 2017.

[16] Peter Bates, Marc Cowling, Nick Jagger, and Gordon Murray. Study of the impact of the enterprise investment scheme (eis) and venture capital trusts (vcts) on company performance. *HM Revenue Customs Research Report*, 44, 2008.

[17] Eli Ben-Michael, Avi Feller, and Jesse Rothstein. The Augmented Synthetic Control Method. 2018.

[18] Jian-Feng Cai, Emmanuel J. Candes, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. 2008.

[19] E. J. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[20] Emmanuel Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98:925 – 936, 2010.

[21] Emmanuel Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Commun. ACM*, 55(6):111–119, June 2012.

[22] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *CoRR*, abs/0912.3599, 2009.

[23] Emmanuel J. Candès and Yaniv Plan. Matrix completion with noise. *CoRR*, abs/0903.3131, 2009.

[24] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *CoRR*, abs/0805.4471, 2009.

[25] G. Chamberlain. Feedback in Panel Data Medels. 1993.

[26] Gary Chamberlain. Panel data. In Z. Griliches† and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition, 1984.

[27] HM Revenue & Customs. Venture capital trusts statistics. *HM Revenue & Customs*, 2018.

[28] Ryan Decker, John Haltiwanger, Ron Jarmin, and Javier Miranda. The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24, 2014.

[29] Nikolay Doudchenko and Guido W Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. (22791), 2016.

[30] Office for National Statistics (ONS). Business demography. *Business Demography, UK*, 2020.

[31] David Gamarnik and Sidhant Misra. A note on alternating minimization algorithm for the matrix completion problem. *IEEE Signal Processing Letters*, 23(10):1340–1343, 2016.

[32] Laurent Gobillon and Thierry Magnac. Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics*, 98(3):535–551, 2016.

[33] Arthur S Goldberger. Structural Equation Methods in the Social Sciences. *Econometrica*, 40(6):979–1001, 1972.

[34] Paul A. Gompers, Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev. How do venture capitalists make decisions? *Journal of Financial Economics*, 135(1):169–190, 2020.

[35] Will Gornall and Ilya A. Strebulaev. The economic impact of venture capital: Evidence from public companies. *Stanford University, Graduate School of Business: Research Papers*, 2015.

[36] John R. Graham and Campbell Harvey. The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics*, 60(2-3):187–243, 2001.

[37] David Gross. Recovering low-rank matrices from few coefficients in any basis. *CoRR*, abs/0910.1879, 2009.

[38] John C Haltiwanger, Ron S Jarmin, and Javier Miranda. Who creates jobs? small vs. large vs. young. (16300), 2010.

[39] Nima Hamidi and Mohsen Bayati. On low-rank trace regression under general sampling distribution. *CoRR*, abs/1904.08576, 2019.

[40] Trevor Hastie, Rahul Mazumder, Jason Lee, and Reza Zadeh. Matrix completion and low-rank svd via fast alternating least squares. 2014.

[41] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2008.

[42] Cheng Hsiao, H. Steve Ching, and Shui Ki Wan. A panel data approach for program evaluation: Measuring the benefits of political and economic integration of hong kong with mainland china. *Journal of Applied Econometrics*, 27(5):705–740, 2012.

[43] PETER ILIEV and MICHELLE LOWRY. Venturing beyond the ipo: Financing of newly public firms by venture capitalists. *The Journal of Finance*, 75, 2020.

[44] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

[45] Prateek Jain and Inderjit S. Dhillon. Provable inductive matrix completion. *CoRR*, abs/1306.0626, 2013.

[46] R. H. Keshavan, S. Oh, and A. Montanari. Matrix completion from a few entries. pages 324–328, 2009.

[47] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries. *CoRR*, abs/0906.2027, 2009.

[48] Olga Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1):282–303, 2014.

[49] Vladimir Koltchinskii, Alexandre B. Tsybakov, and Karim Lounici. Nuclear norm penalization and optimal rates for noisy low rank matrix completion. 2016.

[50] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[51] John Lintner. Distribution of incomes of corporations among dividends, retained earnings, and taxes. *The American Economic Review*, 46(2):97–113, 1956.

[52] Pascal Massart. About the constants in talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.

[53] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11(80):2287–2322, 2010.

[54] Andrew Metrick and Ayako Yasuda. *Venture Capital and the Finance of Innovation*. John Wiley Sons, Hoboken, NJ, 2011.

[55] Hyungsik Roger Moon and Martin Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579, 2015.

[56] Hyungsik Roger Moon and Martin Weidner. Dynamic Linear Panel Regression Models With Interactive Fixed Effects. *Econometric Theory*, 33(1):158–195, 2017.

[57] Sahand Negahban and Martin J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. 2009.

[58] Sahand Negahban and Martin J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13(53):1665–1697, 2012.

[59] Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of *m*-estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557, 2012.

[60] M. Hashem Pesaran. Estimation and inference in large heterogeneous panels with a multi-factor error structure. *Econometrica*, 74(4):967–1012, 2006.

[61] Benjamin Recht. A simpler approach to matrix completion. *CoRR*, abs/0910.0651, 2009.

[62] Angelika Rohde and Alexandre B. Tsybakov. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

[63] Paul Rosenbaum and Donald Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.

[64] Donald B. Rubin. *Matched Sampling for Causal Effects*. Cambridge University Press, 2006.

[65] A. Shaikh and Panos Toulis. Randomization tests in observational studies with staggered adoption of treatment. *Becker Friedman Institute for Economics Working Paper Series*, 2019.

[66] Nathan Srebro, Noga Alon, and Tommi S. Jaakkola. Generalization error bounds for collaborative prediction with low-rank matrices. page 1321–1328, 2004.

[67] HM Treasury. Policy paper: Terms of reference for the patient capital review. *Department for Business, Energy Industrial Strategy*, 2017.

[68] HM Treasury. Public expenditure: Statistical analysis. *HM Treasury*, 2019.

[69] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2011.

[70] Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. The deconfounded recommender: A causal inference approach to recommendation. *CoRR*, abs/1808.06581, 2018.

[71] Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017.