# Risk and Return Characteristics of Entrepreneurial Companies

Arthur Korteweg

Morten Sorensen[†]

May, 2008

**Abstract:** Valuations of entrepreneurial companies are observed infrequently, but more often for well-performing companies. Consequently, estimators of risk and return must correct for heteroscedasticity and sample selection to obtain consistent estimates. We present a new dynamic sample selection model and estimate it with data for venture capitalists' investments in entrepreneurial companies. Correcting for biases leads to lower intercepts and markedly higher measures of systematic and idiosyncratic risks in both the one- and three-factor market models. Our results provide new insights into the performance of entrepreneurs and more generally suggest that estimation biases may inflate assessments of illiquidity premia.

---

[†] Stanford University GSB (korteweg@stanford.edu) and University of Chicago GSB and NBER (sorensen@uchicago.edu). We are particularly grateful to John Cochrane for comments and suggestions. We thank John Heaton, Seminar participants at University of Amsterdam, University of Chicago, University of Illinois–Urbana Champaign, Copenhagen University, University of Iowa and Tilburg University for helpful discussions, to Susan Woodward of Sand Hill Econometrics for generous access to data, and to the Center for Research in Security Prices (CRSP) and the Kauffman Foundation for financial support. All errors are our own.

A number of assets, such as privately held companies, real estate, corporate bonds, many structured products, and securities trading OTC, are only infrequently traded. Since their market values are only known when they trade, data with valuations and returns are necessarily infrequent. When these assets are kept "on the books" at the price of the previous trade, a "stale-price" problem arises (Scholes and Williams (1977) and Dimson (1979)). Furthermore, when the timing of the trades, and hence the timing of the observed valuations and returns, is endogenous, problems with heteroscedasticity and sample selection arise (see Woodward (2004) and Cochrane (2005)). Focusing on the problems arising from the endogenous timing of trades, we present a new dynamic sample selection model and use it to estimate the risk and return characteristics of entrepreneurial companies.

Our results suggest that entrepreneurial companies are more risky than previously found. We estimate a market beta ranging from 2.6 to 3.0. Estimates of the three-factor model (Fama and French (1995)) have loadings on the size factor (SMB) of 0.9 to 1.1 and loadings on the market-to-book factor (HML) of -1.7 to -2.0. Not surprisingly, entrepreneurial companies behave like very small, very high-growth companies. We find that the problems arising from the endogenous timing of the observed valuations are substantial. Valuations of entrepreneurial companies are only observed when these companies receive additional financing or have exits events (i.e. go public or are acquired). Both refinancing and exit events are more likely for companies doing well, and companies with better returns are overrepresented in the data. Controlling for this endogeneity, we see a significant decrease in the intercept of the market model and substantial increases in both systematic and idiosyncratic risk measures.

Our empirical model combines a dynamic asset pricing model with a selection model. The valuations of the entrepreneurial companies are assumed to develop according to a standard one- or three-factor market model, but most of these valuations are unobserved. A company's valuation is only observed when it receives new financing, goes public, or is acquired, and to account for the endogeneity of these events, we add a selection equation to the model. This equation defines a latent selection variable that determines the probability of observing the company's valuation as a function of the valuation itself, the time since the last refinancing, and general market conditions. Estimating this extended dynamic model is numerically challenging due to the large number of latent valuation and selection variables, and the estimation procedure must account for the joint distribution of all these variables. For example, evaluation of the likelihood function requires simultaneously integrating over all of the latent variables, which is numerically infeasible due to the curse of dimensionality. As an alternative, feasible estimation approach, we adopt a Bayesian methodology, relying on a Markov Chain Monte Carlo (MCMC) method known as Gibbs sampling. The simulation of the parameters' posterior distribution can be broken down into simulations from three simpler models: a Bayesian regression, a draw of truncated random variables, and a Kalman Filter. Each of these simpler models is well understood and numerically tractable, and the posterior distribution of the full model is obtained by simulating from these three smaller models in an iterative way, specified by the Gibbs sampling procedure.

### *A. Previous Literature*

Our analysis builds directly on Woodward (2004) and Cochrane (2005). Like them, we develop a selection model for valuations of entrepreneurial companies, but we extend their models in important ways. Woodward (2004) adopts the repeat-sales methodology that has been widely used to construct indices for real estate prices to construct a venture capital index, while correcting for selection bias using the Heckman (1979) sample selection model (see Hwang and Quigley (2003)). One potential limitation of this methodology is that it only adjusts the observed valuations for selection bias, and it does not correct the unobserved valuations. This is not a problem when focusing on the levels of valuations, but it raises empirical issues when estimating correlations between time-varying factors and returns. The standard selection methodology implicitly assumes that returns are earned evenly during the period between two observed valuations; however, the selection equation contains information about the timing of this return, and incorporating this information improves the statistical power of the method and may be necessary for consistency. Our model and the model by Cochrane (2005) capture this relationship.

Compared to the model by Cochrane (2005), our model is numerically more tractable, and we achieve this tractability by developing a new Bayesian estimation methodology. Tractability is important, since it allows us to estimate more flexible specifications. To illustrate, Cochrane (2005) assumes that the probability of observing a valuation (in a refinancing or exit event) is a function of the valuation alone. While we confirm that the valuation is an important determinant, this specification may be overly parsimonious. We further include the time since the previous financing round to better capture the frequency of refinancings, and we include controls for the aggregate venture capital activity in the market. This additional variation in the selection equation improves the statistical identification of the model, and it leads to estimates of the intercept and systematic risks that are higher than those in Cochrane (2005).

A related literature estimates aspects of the risk and return of private equity investments using the cash flows between the private equity funds and their limited partners, sometimes summarized as the funds' Internal Rate of Return (see Ljungqvist and Richardson (2003), Kaplan and Schoar (2005), and Driessen, Lin and Phalippou (2007)). Our approach complements these studies, but there are good reasons for starting from the valuations of the individual companies. First, we have more observations, since there are obviously more individual companies than private equity funds. Second, the return to a fund is earned across a portfolio of companies, typically over a ten- to thirteen-year period, making these procedures less powerful to separately consider changes over time, across industries, or for different kinds of entrepreneurial companies. Finally, estimation of risk and return from cash flows alone faces an identification problem,[1] making it necessary to assume that risk and returns are constant across funds

[1] To illustrate this problem, consider two VCs, both with an initial endowment of $1 and both investing for two periods, after which their entire capital is paid out. Their returns are given by the CAPM model, the market returns are 10% and -11.2% in the two periods, and both VCs earn zero $\alpha$. VC One has a $\beta$ of one and pays out 10% of his total capital after the first period. Hence, the two cash flows observed for this investor are (1 + 10%) x 10% = 1.1 and (1 + 10%) (1 - 10%) (1 - 11.2%) = 0.879. VC Two has a $\beta$ of negative one and pays out 12.2% of her total capital after the first period. Hence, the observed cash flows for VC Two are also (1 - 10%) x 12.2% = 1.1 and (1 - 10%) (1 - 12.2%) (1 + 11%) = 0.879. The identification problem is obvious. The two investors have opposite betas, yet return exactly the same cash

and/or time periods. Together with this assumption, the smaller number of private equity funds, and the concentration of private equity activity in relatively few years relative to the funds' ten-year life-spans further reduce the number of effective observations and may make estimates based on funds' cash flows more sensitive to the underlying assumptions.

## I. Econometric Model

Our model has two equations: a valuation equation and a selection equation. For each company, these equations specify the discrete-time evolution of the valuation and selection variables. To facilitate interpretation and comparison of the parameters with alternative models, we first derive the valuation equation from a continuous-time specification.

### *A. The Valuation Equation*

Let the value of company *i* be denoted $V_i$, and let it develop according to the continuous-time one-factor market model (the Fama-French three-factor model is a straightforward extension).

$$\frac{dV_i}{V_i} - r_f dt = \alpha_i dt + \beta_i \left( \frac{dM}{M} - r_f dt \right) + \sigma_w dw_i \tag{1}$$

flows to their investors. When these are the observed cash flows, it is not possible to determine whether the investor has a positive or a negative $\beta$.

Here $M$ is the value of the market portfolio, $r_f$ is the risk-free rate (continuously compounding), $\alpha_i$ is the drift of the valuation process in excess of $r_f$, and $w_i$ is a standard Brownian Motion. For simplicity, we suppress time subscripts.

To derive the discrete-time dynamics, use Itô's lemma to write

$$d\ln V_i + \left(\frac{1}{2}\sigma_i^2 - r_f\right)dt = \alpha_i dt + \beta_i\left(d\ln M + \left(\frac{1}{2}\sigma_m^2 - r_f\right)dt\right) + \sigma_w dw_i \quad (2)$$

where $\sigma_i^2$ and $\sigma_m^2$ are instantaneous volatilities of $V_i$ and $M$, and $\sigma_i^2 = \beta_i^2\sigma_m^2 + \sigma_w^2$. The discrete-time distribution of the return until time $t$ is

$$\ln\left(\frac{V_i(t)}{V_i(0)}\right) + \left(\frac{1}{2}\sigma_i^2 - r_f\right)t = \alpha_i t + \beta_i \ln\left(\frac{M(t)}{M(0)}\right) + \beta_i\left(\frac{1}{2}\sigma_m^2 - r\right)t + \sigma_w\int_0^t dw_i \quad (3)$$

Rearranging leads to

$$\ln\left(\frac{V_i(t)}{V_i(0)}\right) - r_f t = \left(\alpha_i - \frac{1}{2}\sigma_w^2 + \frac{1}{2}\beta_i\left(1-\beta_i\right)\sigma_m^2\right)t + \beta_i\left(\ln\left(\frac{M(t)}{M(0)}\right) - r_f t\right) + \varepsilon_i \quad (4)$$

where $\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$ for $\sigma_\varepsilon^2 = t\sigma_w^2$. Define $r_m(t) = M(t)/M(0)$, and let the one-period intercept be given by $\delta_i = \alpha_i - \frac{1}{2}\sigma_w^2 + \frac{1}{2}\beta_i\left(1-\beta_i\right)\sigma_m^2$ (this is *not* an abnormal return). The one-period valuation equation for the model is then

$$\ln V_i(t) = \ln V_i(t-1) + r_f + \delta_i + \beta_i\left(\ln r_m(t) - r_f\right) + \varepsilon_i(t) \quad (5)$$

*B. The Selection Equation*

A valuation is only observed when the entrepreneurial company has a refinancing or exit event, and the selection equation captures the endogeneity of these events. For company *i*, at time *t*, assume that $V_i(t)$ is observed when

$$w_i(t) \geq 0 \tag{6}$$

where the selection variable $w_i(t)$ is given by the selection equation

$$w_i(t) = X_i'(t)\gamma_0 + \ln\left(V_i(t)\right)\gamma_1 + \eta_i(t) \tag{7}$$

In this equation, $X_i(t)$ is a vector of characteristics (including a constant) that affects whether a refinancing or exit event happens. Below we include the time since the previous financing round and variables capturing general market conditions. Another important determinant is the company's valuation or return since the previous financing round, as captured by the second term in equation (7).[2] More successful companies, with higher valuations, should be more likely to be observed, and we expect $\gamma_1$ to be positive. We assume that $\eta_i(t)$ is distributed *i.i.d.* $N(0,1)$. The selection equation is equivalent to a binomial discrete choice model, so the scale of its parameters is unidentified and we normalize the scale by fixing the variance of the error term to equal one.

[2] The interpretation of $\ln(V_i(t))$ as a return since last financing round arises if we define *t* to be the time since the previous refinancing round and normalize $V_i(0) = 1$.)

*C. Estimation Procedure*

To summarize, the model consists of a valuation and a selection equation. These are

$$\ln\left(V_i(t)\right) = \ln\left(V_i(t-1)\right) + r_f + \delta + \beta\left(\ln\left(r_m(t)\right) - r_f\right) + \varepsilon_i(t) \tag{8}$$

$$w_i(t) = X_i'(t)\gamma_0 + \ln\left(V_i(t)\right)\gamma_1 + \eta_i(t) \tag{9}$$

The valuation $V_i(t)$ is observed when $w_i(t) \geq 0$ and is unobserved (latent) otherwise. The selection variable $w_i(t)$ is never observed. The distributions of the two residuals are *i.i.d.* $\varepsilon_i(t) \sim N\left(0, \sigma^2\right)$ and $\eta_i(t) \sim N\left(0,1\right)$. The parameters of interest are $\delta$, $\beta$, $\sigma^2$, and $\gamma = \left(\gamma_0, \gamma_1\right)$.

Estimation of this model is numerically challenging. To evaluate the likelihood function the latent variables ($2 \times N \times T$ in total) must be integrated out simultaneously, which is numerically intractable (see Judd (1998)). Instead, we turn to a Bayesian estimation procedure, using MCMC and Gibbs sampling. For this technique, the variables must be divided into blocks, and it is convenient to use three blocks: The first block contains the parameters, $\delta$, $\beta$, $\gamma$, and $\sigma^2$; the second block contains the selection variables, $w_i(t)$; the third block contains the valuation variables, $V_i(t)$. The Gibbs sampler simulates the joint posterior distribution of these variables by iteratively sampling the variables in each block conditional on the previous realization of the variables in the other blocks.[3] The numerical difficulty arises in the third block, which traces out the path of the unobserved valuations in the intermittent period between two observed valuations. This path must take into account realized market returns during this period as well as the fact that no valuation were observed, which reduces the conditional distribution of the unobserved valuations during this period.

*D. Specification of Gibbs Sampler*

Following Tanner and Wong (1987) the simulated posterior distribution is augmented with the latent selection and valuation variables. This augmentation technique significantly improves the numerical tractability of the model, and the original posterior distribution can be recovered by a simple projection of the augmented distribution. Our Gibbs sampler uses 500 iterations the estimated preceded by 500 iterations for burn-in. During the initial burn-in, the simulation appears to converge quickly.

In the first block, the parameters in the valuation and selection equations are simulated conditional on the valuation and selection variables and the data. Equations (8) and (9) show that after conditioning on $V_i(t)$, $V_i(t-1)$), and $w_i$, the parameters $\delta$, $\beta$, and $\gamma$ are defined by a linear regression. Consequently these parameters are sampled from the posterior distribution of a standard Bayesian linear regression.

---

[3] For more details about Gibbs sampling see Geman and Geman (1984) and Gelfand and Smith (1990). Related Bayesian methods are described more generally in Berger (1993), Geweke (2005), and Robert and Casella (2004).

Starting with the valuation equation, $\delta$ and $\beta$ are simulated from the linear regression where $Y_{V,i}(t) = \left(\ln\left(V_i(t)\right) - \ln\left(V_i(t-1)\right) - r_f\right)$ is regressed on a constant term and $\ln\left(r_m(t)\right) - r_f$. Accordingly, define $X_V(t) = \left[1 \quad \ln\left(r_m(t)\right) - r_f\right]$. Assuming an Inverse Gamma prior distribution for the variance, its conditional posterior distribution is also Inverse Gamma

$$\sigma^2 \sim IG(a,b) \tag{10}$$

with parameters

$$\begin{aligned} a &= a_0 + \frac{1}{2}T \\ b &= b_0 + \frac{e'e}{2} \end{aligned} \tag{11}$$

where $T$ is the number of valuations that are simulated, the vector $e$ contains the stacked error terms $\varepsilon_i(t)$, and the prior distribution of $\sigma^2$ is $IG(a_0, b_0)$.

Assuming a Normal prior, the posterior distribution of the coefficients conditional on the variance is

$$\left(\delta, \beta \mid \sigma^2\right) \sim N\left(\mu, \Sigma\right) \tag{12}$$

where

$$\begin{aligned} \Sigma &= \left(\Sigma_0^{-1} + X_V' X_V / \sigma^2\right)^{-1} \\ \mu &= \Sigma\left(\Sigma_0^{-1}\mu_0 + X_V' Y_V / \sigma^2\right) \end{aligned} \tag{13}$$

and the prior distribution of $(\delta, \beta)$ is $N(\mu_0, \Sigma_0)$.

Turning to the selection equation, the parameters are simulated using a similar Bayesian regression. From equation (9) we see that in this regression, $Y_S = w_i(t)$ is regressed on $X_S(t) = \left[ X_i'(t) \quad \ln\left(V_i(t)\right) \right]$. The selection equation is equivalent to a Probit model, and the scale of the parameters is normalized by fixing the variance of the error term to equal one. The conditional distribution of $\gamma$ is

$$\gamma \sim N(\theta, \Omega) \tag{14}$$

where

$$\begin{aligned} \Omega &= \left( \Omega_0^{-1} + X_S' X_S \right)^{-1} \\ \theta &= \Omega\left( \Omega_0^{-1}\theta_0 + X_S' Y_S \right) \end{aligned} \tag{15}$$

and the prior distribution of $\gamma$ is $N(\theta_0, \Omega_0)$.

In the second block, the selection variables are sampled conditional on the valuation variables and the parameters. The conditional distribution follows directly from equations (6) and (7)

$$w_i(t) \sim TN\left( X_i'(t)\gamma_0 + \ln\left(V_i(t)\right)\gamma_1, 1 \right) \tag{16}$$

where *TN* denotes a Normal distribution truncated at zero. The truncation is from above when the corresponding valuation is unobserved and from below when it is observed, as given by equation (6).

In the third, and final, block the latent valuations are sampled conditional on both the parameters and the selection variables. This is the most complex part of the procedure. It samples the entire path of latent valuations between any two observed valuations while conditioning on the information contained in the timing of the observed valuations and the market returns over this period. To accomplish this, we use the Forward Filtering Backwards Sampling (FFBS) procedure by Carter and Kohn (1994), which provides an efficient way to sample an entire path of latent valuations conditional on all the available information.

To understand this procedure, first note that conditional on the parameters and selection variables, the path of the latent valuations is defined by a Kalman Filter. In the filtering terminology, the latent valuation variables are state variables, and the transition rule is given by the valuation equation

$$\ln\left(V_i(t)\right) = \ln\left(V_i(t-1)\right) + \delta + r_f + \beta\left(r_m(t) - r_f\right) + \varepsilon_i(t) \tag{17}$$

where $\delta$ , $r_f$ and $\beta\left(r_m(t) - r_f\right)$ are "observed" controls acting on the state. This system has one or two observation equations depending on whether the valuation is observed or not. The first observation equation is given by the selection equation

$$w_i(t) = X_i'(t)\gamma_0 + \ln\left(V_i(t)\right)\gamma_1 + \eta_i(t) \tag{18}$$

When conditioning on the selection variables, these variables can be viewed as noisy "observations" of $\ln\left(V_i(t)\right)$. Moreover, when a valuation is observed, the observed valuation is given by the second observation equation

$$\ln\left(V_i(t)\right) = \ln\left(V_i^*(t)\right) \quad (19)$$

Here $V_i^*(t)$ is the observed valuation, and $V_i(t)$ is the latent valuation from the filter. In our specification we assume that the valuations are observed without error and these variables are identical, but it is straightforward to incorporate observation error as well.

*E. Prior Distributions*

We use diffuse priors for the parameters. For the valuation equation, the prior mean of $\delta$ is zero, and we use the GLS estimate of $\beta$ as its prior mean. The prior variances of $\delta$ and $\beta$ equal 1,000 times their estimated GLS variance (which is typically around 0.008). The prior distribution of $\sigma^2$ is an Inverse Gamma distribution with parameters 2.1 and 600. This distribution implies that $E(\sigma) = 0.12$ and $\sigma$ is between 1% and 50% with 99% probability.

In the selection equation, we use a Normal distribution with variance of 100 as the prior distribution of the parameters. The intercept has a prior mean of -1, and the mean of the log valuations is 1. The prior mean coefficient on time since last refinancing and its square are 1 and -0.1. All other coefficients have means of zero. The prior distributions are diffuse relative to the resulting posterior distributions, and the estimates are robust to changes in the means and variances of these distributions, indicating that the estimated coefficients reflect information in the data rather than in the priors.

## II. Data Description

We downloaded monthly market returns and Fama-French portfolios from Kenneth French's website. These factor returns are constructed from all NYSE, AMEX, and NASDAQ firms in CRSP. Monthly Treasury-Bill rates are from Ibbotson Associates and are also available on Ken French's website.

### *A. Venture Capital Data*

Data with VCs' investments in entrepreneurial companies are provided by Sand Hill Econometrics (SHE), a commercial data provider. The data contain the majority of US investments in the period from 1987 to 2005. SHE combines and extends two commercially available databases, Venture Xpert (formerly Venture Economics) and Venture One. These two databases are extensively used in the VC literature, and Gompers and Lerner (1999) and Kaplan, Sensoy and Strömberg (2002) investigate the completeness of the Venture Xpert data and find that they contain the majority of the investments. In addition, SHE has spent substantial time and effort to ensure the accuracy of the data. This includes removing investment rounds that did not actually occur, adding investment rounds that were not in the original data, and consolidating rounds, so that each round corresponds to a single actual investment by one or more VCs. Cochrane (2005) uses similar data, but our version is more recent and many of the previously encountered data problems have been resolved.

## *B. Calculating Valuations*

VCs distinguish between the pre- and post-money valuation of an investment, and the data contain both of these valuations for a large number of the investments. When a VC invests $I$ in a company with a total value of $PV_{POST}$ after the investment (the post-money valuation), the $PV_{PRE}$ (the pre-money valuation) is defined by $PV_{POST} = PV_{PRE} + I$.

To illustrate, imagine VC A invests $1m in a company with a pre-money valuation of $2m and a post-money valuation of $3m. This implies that VC A receives 1/3 of the shares of the company. In the next round, VC B invests $4m, at a pre-money valuation of $6m and a post-money valuation of $10m. The issuance of new shares to VC B dilutes VC A's ownership fraction to 1/5 ( = 1/3 x (1- 4/10) ), and the value of VC A's shares is 1/5 x $10m = $2m, for a return of 100%. This value can also be calculated using VC A's original ownership fraction and the pre-money valuation in the second round, as 1/3 x $6m = $2m, and hence we can calculate the returns from the round at time $t$ to the round at time $t'$ using the formula

$$r^{t,t'} = \frac{PV_{PRE}(t')}{PV_{POST}(t)} - 1 \tag{20}$$

We use this relationship to construct a new valuation variable that strips out the dilution of the original ownership. Starting from $V(0) = 1$, future values of this valuation are calculated using the equation

$$\frac{V(t')}{V(t)} = r^{t,t'} \tag{21}$$

where $r^{t,t'}$ is defined by equation (20), and the resulting valuations are used as the observed valuations in the estimation procedure.

Note that this calculation can only be performed when pre- and post-money valuations are observed for consecutive rounds. When an intermediate valuation is missing, it is impossible to adjust for the dilution, and the valuation variable is "restarted" after the break. This reduces the number of valuations used for the estimation, but it does not introduce any bias.

*C. Descriptive Statistics*

Table I presents descriptive statistics. We observe a total of 15,169 financing rounds of which 9,637 have valuation data. There are 3,237 unique start-ups, with the majority concentrated in the Biotech and IT industries. One-third of these firms ultimately go public, another 742 are acquired, and we have explicit information about 444 companies being liquidated. While we have no information about the fate of the remaining 940 companies, some of these may be alive and well, some may be "living zombies," but the majority is likely liquidated. The empirical model incorporates the uncertainty about these unobserved outcomes.

On average, an entrepreneurial firm receives 4.7 financing rounds (the median is 4 rounds), but this distribution is highly skewed, with some firms receiving as many as 9 rounds. On average, 12.4 months pass between consecutive rounds (the median is 10 months), but this distribution is also skewed. While 5% of follow-on investments occur after as little as 2 months, another 5% take more than 32 months. The observed returns between rounds are 129% on average, with a standard deviation as large as 340%.

**** TABLE I ABOUT HERE ****

## III. Effects of Heteroscedasticity and Sample Selection

Below we present estimates of the one-factor market model and discuss the econometric issues in the context of this model. Then we present estimates for more general specifications, noting that these suffer from the same econometric issues.

### *A. Heteroscedasticity, OLS, and GLS*

First, compare estimates arising from a standard OLS, a GLS, and our MCMC procedures. The OLS and GLS estimators are standard estimation procedures, which estimate the risk and return in a regression framework without accounting for the endogeneity of the observed returns. For these two estimators, we calculate the excess log return and regress it on the corresponding excess log return on the market. To facilitate the comparisons of the various estimators, these returns are calculated from the basic specification in equation (4),

$$\ln\left(V_i(t')/V_i(t)\right) - r_f^{t,t'} = \delta_{OLS}\left(t'-t\right) + \beta_{OLS}\left[\ln\left(r_m^{t,t'}\right) - r_f^{t,t'}\right] + \varepsilon_{OLS} \tag{22}$$

We estimate this regression pooled across firms.

Equation (22) is heteroscedastic, since valuations that are further apart have more volatile error terms. Formally, equation (4) implies

$$\varepsilon_{OLS} \sim N\left(0, (t'-t)\sigma_w^2\right) \tag{23}$$

The GLS estimator normalizes the specification in equation (22) by dividing by the square root of the time between observed valuations. Hence, we estimate the following GLS specification

$$\frac{\ln\left(V_i(t')/V_i(t)\right)-\ln(r_f^{t,t'})}{\sqrt{t'-t}}=\delta_{GLS}\left(\sqrt{t'-t}\right)+\beta_{GLS}\frac{\ln(r_m^{t,t'})-\ln(r_f^{t,t'})}{\sqrt{t'-t}}+\varepsilon_{GLS} \tag{24}$$

Table II presents estimates of the OLS, the GLS, and the MCMC models. These estimates are calculated without accounting for the endogeneity of the observed returns, and we see that the intercepts of monthly log returns from the OLS and the GLS models are $\delta_{OLS}=0.91\%$ and $\delta_{GLS}=2.39\%$. This difference is consistent with an insight from Cochrane (2005) that an important effect of selection is that the observed valuation process is not a Geometric Brownian motion. If the observed valuations did follow a Geometric Brownian motion, the OLS and GLS estimators would both be consistent, and heteroscedasticity would only affect the standard error of the OLS estimate. However, a valuation that is further away from the previous one has an error term with a greater variance, and the GLS procedure place less weight on these observations relative to the OLS procedure. The smaller OLS intercept shows that observations further apart underperform relative to those that are closer together, consistent with the valuations being observed when they cross a threshold condition, as specified by the selection equation.

Next, we observe that the GLS and MCMC estimates are similar. This is reassuring, since both methods capture the heteroscedasticity described above and are estimated using essentially the same information in the data. One difference is that the

MCMC estimator makes an explicit assumption about the distribution of the error terms whereas the GLS estimator is based on moment conditions. The similarity of the estimates suggests that the distributional assumptions are not unreasonable.

Finally, to turn these estimates into preliminary excess returns, we need to correct for the log-linearization and calculate $\alpha = \delta + 1/2\,\sigma^2 - 1/2\beta(1-\beta)\sigma_m^2$. The estimated standard error in the OLS regression is 88%, but due to the heteroscedasticity, it is not clear how to interpret this value (it does not represent a volatility over any given time period). The GLS estimation finds a monthly volatility of 32.66%, and correcting the intercept leads to an estimate of monthly abnormal return of 7.7%. Like Cochrane (2005), we find a substantial abnormal return, mainly driven by the large volatility of the process.

**** TABLE II ABOUT HERE ****

*B. Sample Selection*

To provide a sense of the magnitude of the selection problem, we compare MCMC estimates with and without the selection equation. Comparing Table II and III, we find that the intercept of the monthly log-market model declines from 2.13% to somewhere in the range from -2.11% to -3.45%, depending on the specification. The estimate of Beta increases from the initial estimate of 1.62 to somewhere between 2.64 and 3.02, which represents almost a doubling of the estimated systematic risk. The estimates of the volatility also increase from 32.43% to somewhere between 39.49% and 45.74%. The decrease in the intercept and the increases in the Beta and volatility are expected changes from including the selection equation in the model.

The estimated parameters in the selection equation are also reasonable. Across all specifications, the probability of observing a valuation depends positively on the return since the previous financing round ( $\ln(V_i(t')) - \ln(V_i(t))$ ). As the time since the previous financing round increases, a new financing round becomes more likely, but when the time becomes excessive, the square term dominates, and the chance of refinancing declines.

For the identification of our selection model, it is important to have independent variation in the selection equation. The identifying assumption is that this variation must be related to the probability of observing a valuation but independent of the error term in the valuation equation, $\varepsilon_i$ . Both the time since the previous refinancing and the number of acquisitions or public offerings in the overall market may provide this variation. These assumptions may be more reasonable for the time since previous refinancing than for the market wide variables. For the market wide variables, the implicit assumption is that the return to the market captures all the systematic variation in the entrepreneurial companies' return with no residual return left to be explained by the market wide variables.

## IV. Risk and Return Characteristics for Entrepreneurial Companies

### *A. Estimating a Three-Factor Model*

Table IV presents estimates of a Fama-French three-factor specification. Not surprisingly, we find a large positive loading on the market factor of 2.3 to 2.5, indicating that entrepreneurial companies are very exposed to the general market. For comparison

Davis, Fama, and French (2000) estimate factor loadings for publicly traded companies. For small growth companies (most similar to entrepreneurial companies), they find factor loadings on the market factor from 1.01 to 1.06 depending on the time period over which the loadings are calculated.[4] For the size factor, we estimate loadings from 0.9 to 1.1. This is comparable to the loadings from 1.22 to 1.39 reported by Davis, Fama, and French, suggesting that the largest entrepreneurial companies are not significantly smaller than the smallest publicly listed companies. For the book-to-market (HML) factor, we find negative loadings between -2.0 and -1.7, suggesting that our companies are extreme growth companies. Davis, Fama, and French find loadings between -0.14 and 0.23 for this factor for publicly traded small growth companies. Fama and French (1995), using a different definition of the small-growth portfolio, find loadings on the market portfolio of 1.06, of 1.04 on SMB, and of -0.31 on HML. Cochrane (2005) reports loadings on a sample of small Nasdaq firms. For value weighted portfolios, he reports loadings between 0.7 and 0.9 on the market portfolio, between 1.3 and 1.8 for SMB, and between 0.1 and 0.4 for HML.

*B. Risk and Returns at Different Stages*

Table V presents estimates with separate coefficient for investments in companies at different stages. The four stages are "seed," "early," "late," and "mezzanine." The table reveals interesting patterns. In both specifications, the intercept follows a U-shape.

---

[4] Note, these factor loading are not entirely comparable, since our loadings are defined from a continuous-time three-factor model and the comparison loadings are defined by a one-month discrete-time three-factor model. When alpha does not equal zero, the factor loadings change slightly with the time period over which the model is defined.

It is higher for seed and mezzanine investments, and smaller for early and late stage investments. The systematic risk increases with the stage of the investment. Seed investments have very little systematic risk exposure, suggesting that most of the risk at this stage may be idiosyncratic technological risks. After the seed stage, in the first specification, the risk exposure increases monotonically in the development of the company. This is consistent with the model by Berk, Green and Naik (2004), which suggest that entrepreneurial risks are idiosyncratic, but a real-option effect can gradually increase the exposure to market risks as the company matures. Again in the second specification the seed investments have no systematic risk exposure. However, as the company develops from the early, through the late, and to the mezzanine stage, the loading on the size factors declines from 1.27 to 0.56. This is consistent with the growth of the companies and that the size of companies receiving financing at the mezzanine stage is not substantially different from small publicly traded companies. The HML factor has small loadings at the seed and mezzanine investments, but very large (negative) loadings for early and late stage investments. Interpreting the HML factor exposure as a measure of growth options, this is consistent with the early and late stages being the most critical expansion stages for entrepreneurial companies. At the mezzanine stage, the factor loading becomes more moderate and more in line with the book-to-market loading for publicly traded small growth companies. Interestingly, the measure of idiosyncratic volatility remains largely constant across these four stages.

### *C. Industry Differences in Risk and Return*

Table VI presents estimates for investments separated into four broad industry classifications (using Sand Hill Econometrics' industry classifications), "Biotech," "IT," "Retail," and "Other." Again we find that controlling for sample selection leads to substantially lower estimates of the intercept and greater estimates of both systematic and idiosyncratic risks. We also find substantial differences across the four industry groups. Except for the GLS specification, Biotech and IT have consistently greater returns than Retail and Other. Retail investments show a particular high exposure to systematic factors, with an estimate of beta almost twice as large as for the other industries. In contrast, Biotech and the other industry group have less systematic exposure. While the industry classifications are perhaps overly broad, it is interesting to discover this degree of disparity across industries. It suggests that estimates of risk and return to private equity investments performed using fund level returns may be sensitive to the industry compositions of these funds.

## V. Conclusion

When estimating risk and return for assets with infrequently observed market valuations, empirical problems arise due to heteroscedasticity and sample selection. When the duration between observed valuations varies, the resulting heteroscedasticity makes it difficult to interpret estimated volatilities, and we show that a straightforward GLS correction can solve this problem. Aa sample selection problem arises when the timing of observed valuations is endogenous. We introduce and estimate a new sample selection model that overcomes both the sample selection and heteroscedasticity problems.

Our empirical model employs a Bayesian methodology, relying on insights from Gibbs sampling and Kalman Filtering, to generalize previous studies by Woodward (2004) and Cochrane (2005). It allows us to explicitly estimate the posterior distribution of the unobserved paths of valuations, and in this way allows us to estimate correlations with the market and other factors. It is possible to extend this model to facilitate applications to other infrequently-traded assets. For example, our valuation equation can directly incorporate hedonic characteristics of the underlying assets, which has been an important issue in the real-estate literature and the construction of repeat-sales indices.

We find that correcting for selection bias is important. It leads to substantial reductions in estimated intercepts and substantially increases estimated risk exposures, for both systematic and idiosyncratic risks. These findings are robust to different specifications of the pricing model and the selection equation. As a more general point, our findings may have implications for the estimation of illiquidity premia on infrequently traded assets. Illiquidity premia may be overstated if they are calculated from estimates of risk exposures that do not account for selection bias. Our methodology may help improve these estimates, which is important for understanding markets for illiquid assets more generally.

**Bibliography**


Berger, James, 1993. *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

Berk, Jonathan, Richard Green, and Vasant Naik, 2004, Valuation and Return Dynamics of New Ventures, *Review of Financial Studies* 17, 1-35.

Carter, C., and R. Kohn, 1994, On Gibbs Sampling for State Space Models, *Biometrika* 81, 541-553.

Cochrane, John, 2005, The Risk and Return of Venture Capital, *Journal of Financial Economics* 75, 3-52.

Dimson, E., 1979, Risk Measurement When Shares are Subject to Infrequent Trading, *Journal of Financial Economics* 7, 197-226.

Driessen, Joost, Tse-Chun Lin, and Ludovic Phalippou, 2007, Measuring the risk of private equity funds: A new approach, *working paper*.

Fama, Gene, and Kenneth French, 1995, Size and Book-to-Market Factors in Earnings and Returns, *Journal of Finance* 50, 131-155.

Gelfand, Alan, and Adrian Smith, 1990, Sampling Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association* 85, 398-409.

Geman, S., and D. Geman, 1984, Stochastic Relaxation, Gibbs Distributions, adn the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.

Geweke, John, 2005. *Contemporary Bayesian Econometrics and Statistics* (John Wiley & Sons, Hoboken, NJ).

Gompers, Paul, and Josh Lerner, 1999. *The Venture Capital Cycle* (MIT Press, Cambridge).

Heckman, James, 1979, Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-162.

Hwang, Min, and John M. Quigley, 2003, Selectivity, Quality Adjustment and Mean Reversion in the Measurement of House Values, *Journal of Real Estate Finance and Economics* 28, 161-178.

Judd, Kenneth, 1998. *Numerical Methods in Economics* (MIT Press, Cambridge).

Kaplan, Steven, and Antoinette Schoar, 2005, Private Equity Performance: Returns, Persistence, and Capital Flows, *Journal of Finance* 60, 1791-1823.

Kaplan, Steven, Berk Sensoy, and Per Strömberg, 2002, How Well do Venture Capital Databases Reflect Actual Investments? *working paper, University of Chicago*.

Ljungqvist, Alexander, and Matthew Richardson, 2003, The Cash Flow, Return and Risk Characteristics of Private Equity, *working paper*.

Robert, Christian, and George Casella, 2004. *Monte Carlo Statistical Methods* (Springer-Verlag, New York).

Scholes, M., and J. Williams, 1977, Estimating Betas from Nonsynchronous Data, *Journal of Financial Economics* 5, 309-328.

Tanner, Martin, and Wing Wong, 1987, The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association* 82, 528-549.

Woodward, Susan, 2004, Measuring Risk and Performance for Private Equity, *working paper*.

## Table I: Descriptive Statistics

Table I presents descriptive statistics for the VC sample. Separated into the four industry classifications, Panel A presents total number of rounds and number of rounds with valuations. Panel B contains the number of entrepreneurial firms sorted according to industry classifications and exit status. Panel C contains information about the average number of financing rounds per entrepreneurial firm. Total is the total number of rounds received, Tau contains the time since the previous round (in months), and return is the return earned since the previous round (in percent).

**Panel A: Total Rounds**

| | Total | Biotech | IT | Retail | Other |
|---|---|---|---|---|---|
| Number of rounds | 15,169 | 3,892 | 9,152 | 1,700 | 425 |
| With valuations | 9,637 | 2,485 | 5,759 | 1,126 | 267 |

**Panel B: Number of firms**

| | Total | Biotech | IT | Retail | Other |
|---|---|---|---|---|---|
| IPO | 1,111 | 343 | 623 | 121 | 24 |
| Acquisition | 742 | 137 | 507 | 74 | 24 |
| Liquidated | 444 | 70 | 288 | 77 | 9 |
| Unknown/still alive | 940 | 260 | 549 | 92 | 39 |
| Total | 3,237 | 810 | 1,967 | 364 | 96 |

**Panel C: Rounds per firm**

| | mean | median | St. Dev. | p5 | p95 |
|---|---|---|---|---|---|
| Total | 4.6861 | 4 | 2.2607 | 2 | 9 |
| w/ valid data | 2.9771 | 3 | 1.3099 | 2 | 6 |
| Tau | 12.3522 | 10 | 11.2771 | 2 | 32 |
| Return | 128.77 | 50 | 340.29 | -62.99 | 500.77 |

## Table II: OLS, GLS, and MCMC Estimates

The table presents OLS, GLS and MCMC estimates of the market model in monthly log returns. The GLS estimator scales each observation with the inverse of the square-root of the time since last financing round. MCMC estimates are mean and standard deviation of the parameters' simulated posterior distribution. The Intercept and Beta are the intercept and the slope from a regression of the log returns to the companies on the market log return. For the GLC and MCMC estimators, Sigma is the estimated monthly standard deviation of the error term from this regression. For the OLS estimator, Sigma is the estimated standard deviation of the error term (it does not have a time interpretation). The MCMC estimator use 500 iterations preceded by 500 discarded iterations for burn-in. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, or, for Bayesian estimates, whether zero is contained in the 1%, 5%, and 10% credible intervals, respectively.

| | OLS | | | GLS | | | MCMC | | |
|---|---|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | | Coef. | Std. Err. | | Mean | Std. Dev. | |
| Intercept | 0.0091 | (0.0008) | *** | 0.0239 | (0.0013) | *** | 0.0213 | (0.0012) | *** |
| Beta | 1.6104 | (0.0672) | *** | 1.3891 | (0.0869) | *** | 1.6152 | (0.0739) | *** |
| Sigma | 0.8777 | | | 0.3266 | | | 0.3243 | (0.0031) | *** |

## Table III: MCMC Estimates with Selection Correction

The table presents MCMC estimates of the one-factor market model in monthly log returns with selection correction. All reported estimates are mean and standard deviation of the simulated posterior distributions. In the valuation equation, Intercept and Beta are the intercept and the slope on the market log return. Sigma is the estimated monthly standard deviation of the error term. In the selection equation, Return is the log return earned since the previous financing event. Tau is the time since this event (in months). Rounds contains the total number of VC investment rounds in the market in the same month, and Dollars contains the total dollar volume of these investments. ACQ and IPO contain the number of VC backed acquisitions and IPOs observed in this month. The simulations use 500 iterations preceded by 500 discarded iterations for burn-in. ***, **, and * denote whether zero is contained in the 1%, 5%, and 10% credible intervals, respectively.

| | 1 | | | 2 | | | 3 | | | 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | | Mean | Std. Dev. | | Mean | Std. Dev. | | Mean | Std. Dev. | |
| Valuation Equation | | | | | | | | | | | | |
| Intercept | -0.0211 | (0.0011) | *** | -0.0345 | (0.0038) | *** | -0.0318 | (0.0046) | *** | -0.0295 | (0.0037) | *** |
| Beta | 2.8898 | (0.1087) | *** | 3.0157 | (0.1460) | | 2.7753 | (0.1635) | *** | 2.6377 | (0.1529) | *** |
| Sigma | 0.3949 | (0.0028) | *** | 0.4574 | (0.0193) | *** | 0.4552 | (0.0214) | *** | 0.4498 | (0.0173) | *** |
| Selection Equation | | | | | | | | | | | | |
| Return | 0.1810 | (0.0040) | *** | 0.2214 | (0.0085) | *** | 0.2165 | (0.0172) | *** | 0.2157 | (0.0137) | *** |
| Tau | | | | 0.2635 | (0.0117) | *** | 0.2664 | (0.0110) | *** | 0.2644 | (0.0115) | *** |
| Tau^2 | | | | -0.0290 | (0.0024) | *** | -0.0298 | (0.0028) | *** | -0.0295 | (0.0025) | *** |
| Rounds | | | | | | | | | | 0.4876 | (0.0662) | *** |
| Dollars | | | | | | | 11.9621 | (2.8409) | *** | -8.2378 | (4.1443) | * |
| ACQ | | | | | | | 2.7620 | (0.5679) | *** | 0.3078 | (0.5833) | |
| IPO | | | | | | | 5.0547 | (0.7148) | *** | 5.4601 | (0.6836) | *** |
| Constant | -1.4817 | (0.0052) | *** | -1.6304 | (0.0093) | *** | -1.8056 | (0.0164) | *** | -1.8604 | (0.0174) | *** |

## Table IV: MCMC Estimates of Fama-French 3-factor Model

The table presents the posterior distributions of the parameters of the Fama-French model in log-returns. Factor and risk-free returns are from Kenneth French's website. The table presents OLS, GLS, and MCMC estimates of the market model in monthly log-returns with selection correction. Alpha_tilde and RMRF are the intercept and the slope on the market log-return. SMB and HML are loadings on the size and book-to-market factors. Sigma is the monthly standard deviation of the error term. In the selection equation, Return is the return earned since the previous financing event, and Tau is the time since this event (in months). Rounds measures the total number of VC investment rounds in the market in the same month. Similarly, Dollars measures the total dollar volume of these investments, and ACQ and IPO contain the number of VC backed acquisitions and IPOs observed in this month. The GLS estimator scales each observation with the inverse of the square-root of the time since last financing round, as described in the text. Reported MCMC estimates are mean and standard deviation of the parameters' simulated posterior distribution. The simulations use 500 iterations preceded by 500 discarded iterations for burn-in. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, or, for Bayesian estimates, whether zero is contained in the 1%, 5%, and 10% credible intervals, respectively.

| | OLS | | | GLS | | | MCMC | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | 2 | | | 3 | | | 4 | | |
| | Coef. | Std. Err. | | Coef. | Std. Err. | | Mean | Std. Dev. | | Mean | Std. Dev. | |
| Valuation Equation | | | | | | | | | | | | |
| Intercept | 0.0082 | (0.0009) | *** | 0.0216 | (0.0014) | *** | -0.0316 | (0.0034) | *** | -0.0278 | (0.0036) | *** |
| RMRF | 1.4299 | (0.0694) | *** | 1.2327 | (0.0877) | *** | 2.5446 | (0.1562) | *** | 2.2662 | (0.1354) | *** |
| SMB | 0.4832 | (0.0803) | *** | 0.7230 | (0.1026) | *** | 1.0948 | (0.1343) | *** | 0.9155 | (0.1182) | *** |
| HML | -0.6322 | (0.0686) | *** | -0.8113 | (0.0856) | *** | -2.0071 | (0.1189) | *** | -1.7411 | (0.1580) | *** |
| Sigma | 0.8714 | | | 0.3240 | | | 0.4455 | (0.0126) | *** | 0.4401 | (0.0169) | *** |
| Selection Equation | | | | | | | | | | | | |
| Return | | | | | | | 0.2258 | (0.0109) | *** | 0.2244 | (0.0128) | *** |
| Tau | | | | | | | 0.2756 | (0.0124) | *** | 0.2803 | (0.0146) | *** |
| Tau^2 | | | | | | | -0.0296 | (0.0024) | *** | -0.0302 | (0.0026) | *** |
| Rounds | | | | | | | | | | 0.4614 | (0.0647) | *** |
| Dollars | | | | | | | | | | -10.8600 | (3.9612) | *** |
| ACQ | | | | | | | | | | 1.1112 | (0.6038) | ** |
| IPO | | | | | | | | | | 5.2952 | (0.6930) | *** |
| Constant | | | | | | | -1.6386 | (0.0090) | *** | -1.8691 | (0.0194) | *** |

**Table V: Risk-Return for Different Company Stages**

The table presents the posterior distributions of the parameters of a market model and Fama-French model. Factor and risk-free returns are from Kenneth French's website. The table presents MCMC estimates of the market and the FF3 models in monthly log-returns with selection correction. Alpha and Beta (RMRF) are the intercept and the slope on the market log-return. SMB and HML are loadings on the size and book-to-market factors. Sigma is the monthly standard deviation of the error term. Reported MCMC estimates are mean and standard deviation of the parameters' simulated posterior distribution. The simulations use 500 iterations preceded by 500 discarded iterations for burn-in. ***, **, and * denote whether zero is contained in the 1%, 5%, and 10% credible intervals, respectively.

| | | Mean | Std. Dev. | Mean | Std. Dev. |
|---|---|---|---|---|---|
| Intercept | seed | 0.0598 | (0.0099) *** | 0.0590 | (0.0112) *** |
| | early | -0.0095 | (0.0021) *** | -0.0045 | (0.0025) |
| | late | -0.0440 | (0.0026) *** | -0.0224 | (0.0029) *** |
| | mezz | -0.0091 | (0.0116) | 0.0024 | (0.0131) |
| RMRF | seed | -0.0011 | (0.7865) | 0.2990 | (0.6865) |
| | early | 2.8120 | (0.1416) *** | 2.1549 | (0.1515) *** |
| | late | 3.1726 | (0.1683) *** | 2.0133 | (0.1781) *** |
| | mezz | 4.4887 | (0.7729) *** | 2.7077 | (0.9561) ** |
| SMB | seed | | | -0.3445 | (0.6917) |
| | early | | | 1.2705 | (0.2608) *** |
| | late | | | 0.7819 | (0.2214) *** |
| | mezz | | | 0.5558 | (0.6187) |
| HML | seed | | | 0.4564 | (0.4809) |
| | early | | | -1.2211 | (0.1592) *** |
| | late | | | -2.1263 | (0.1803) *** |
| | mezz | | | -0.1868 | (0.6531) |
| Sigma | seed | 0.5148 | (0.0099) *** | 0.5237 | (0.0117) *** |
| | early | 0.5220 | (0.0054) *** | 0.5295 | (0.0088) *** |
| | late | 0.5353 | (0.0057) *** | 0.5429 | (0.0093) *** |
| | mezz | 0.4989 | (0.0150) *** | 0.5053 | (0.0165) *** |

**Table VI: Industry level Risk-Return Estimates**

The table presents the posterior distributions of the parameters of the Fama-French model in log-returns. Factor and risk-free returns are from Kenneth French's website. The table presents OLS, GLS, and MCMC estimates of the market model in monthly log-returns with selection correction. Intercept and Beta are the intercept and the slope on the market log-return. Sigma is the monthly standard deviation of the error term. In the selection equation, Return is the return earned since the previous financing event, and Tau is the time since this event (in months). Rounds measures the total number of VC investment rounds in the market in the same month. Similarly, Dollars measures the total dollar volume of these investments, and ACQ and IPO contain the number of VC backed acquisitions and IPOs observed in this month. The GLS estimator scales each observation with the inverse of the square-root of the time since last financing round, as described in the text. Reported MCMC estimates are mean and standard deviation of the parameters' simulated posterior distribution. The simulations use 500 iterations preceded by 500 discarded iterations for burn-in. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, or, for Bayesian estimates, whether zero is contained in the 1%, 5%, and 10% credible intervals, respectively.

| | | OLS | | | GLS | | | MCMC | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Coef. | Std. Err. | | Coef. | Std. Err. | | Mean | Std. Dev. | | Mean | Std. Dev. | |
| Valuation Equation | | | | | | | | | | | | | |
| Intercept | Biotech | 0.0114 | (0.0016) | *** | 0.0200 | (0.0025) | *** | -0.0285 | (0.0035) | *** | -0.0249 | (0.0036) | *** |
| | IT | 0.0091 | (0.0011) | *** | 0.0258 | (0.0017) | *** | -0.0318 | (0.0036) | *** | -0.0278 | (0.0039) | *** |
| | Retail | 0.0068 | (0.0030) | ** | 0.0326 | (0.0044) | *** | -0.0492 | (0.0058) | *** | -0.0476 | (0.0062) | *** |
| | Other | 0.0063 | (0.0053) | | 0.0105 | (0.0077) | | -0.0505 | (0.0081) | *** | -0.0462 | (0.0080) | *** |
| Beta | Biotech | 0.5927 | (0.1300) | *** | 0.3140 | (0.1736) | * | 2.0956 | (0.2288) | *** | 1.6607 | (0.1841) | *** |
| | IT | 1.8656 | (0.0843) | *** | 1.6240 | (0.1087) | *** | 2.9741 | (0.1647) | *** | 2.7108 | (0.1563) | *** |
| | Retail | 3.1028 | (0.2244) | *** | 2.7645 | (0.2757) | *** | 5.4072 | (0.3543) | *** | 5.1907 | (0.3538) | *** |
| | Other | 0.8432 | (0.4012) | ** | 0.6402 | (0.5138) | | 1.7414 | (0.6055) | *** | 1.2323 | (0.5800) | *** |
| Sigma | | 0.8680 | | | 0.3232 | | | 0.4508 | (0.0136) | | 0.4445 | (0.0172) | *** |
| Selection Equation | | | | | | | | | | | | | |
| Return | | | | | | | | 0.2204 | (0.0140) | *** | 0.2108 | (0.0136) | *** |
| Tau | | | | | | | | 0.2594 | (0.0123) | *** | 0.2610 | (0.0120) | *** |
| Tau^2 | | | | | | | | -0.0287 | (0.0021) | *** | -0.0299 | (0.0023) | *** |
| Rounds | | | | | | | | | | | 0.4714 | (0.0632) | *** |
| Dollars | | | | | | | | | | | -6.8734 | (3.8495) | * |
| ACQ | | | | | | | | | | | 0.2844 | (0.6360) | |
| IPO | | | | | | | | | | | 5.2208 | (0.7185) | *** |
| Constant | | | | | | | | -1.6293 | (0.0094) | *** | -1.8505 | (0.0175) | *** |