A New Model of Venture Capital Risk and Return

Michael Ewens*

December 2009

Abstract

I formulate a model and estimator of venture capital (VC) returns motivated by the entrepreneurial firm life-cycle and the extreme return outcomes of typical venture capital investments. The model incorporates tail events and the estimator corrects for sample selection bias and endogenous investment holding periods. I find that an asymmetric three-state mixture distribution is a better characterization of returns than the standard single-state model. Mixture states mimic typical VC outcomes: "winners," "break-even" and "failures." Imposing normality on venture capital investment returns understates downside risk and kurtosis. In contrast to earlier studies, the mixture model reveals a leptokurtic, negatively-skewed returns distribution. Two new implications follow from the results. First, volatility as an estimate of risk underestimates the frequency and magnitude of large, negative VC returns. Investors in venture capital may need to incorporate additional moments or semivariance into their allocation decisions. Second, a microcap index benchmark previously shown to mimic the means and CAPM alphas of VC returns lacks the downside risk or fat tails of the VC mixture distribution. Thus, VC investments offer some risk and return features unavailable in publicly traded equities.

1 Introduction

Accurately characterizing the risk and return of venture capital investments informs capital allocation and reveals how tail events are reflected in returns. Venture capital-backed firms play an important role in economic growth through the creation of new technologies and firms. For example, clean energy startups received some $7 billion in venture capital in 2008 and in 2009 clean-energy is on pace to be the most popular industry for venture capital. Major investors in this asset class – limited partners or LPs – include pension funds, investment banks and endowments. These institutions may invest 10 - 20% of their portfolio in venture capital and thus require an accurate measure of the risk and return for allocation decisions.[1] LPs also use returns estimates for comparison of venture capital to publicly-traded alternatives that lack the former's illiquidity and fees. Discovering if and how non-normality characterizes the returns distribution of venture capital investments should illustrate one way the asset class differs from standard public indices.

*Ph.D. candidate, Department of Economics, University of California San Diego. Contact: mewens@ucsd.edu. I thank VentureSource, and particularly Brendan Hughes, for access to the VentureOne database. I also thank Jun Lui, David Coats, Ben Gillen, Jacob LaRiviere, Grace Chui-Miller, Yixiao Sun, Allan Timmermann, Choon Wang, Hal White and seminar participants for helpful comments. All errors are my own.

[1]For example, in June 2009, the California public employee pension fund committed 14% of its $183 billion fund to private equity and venture capital.

Extreme outcomes characterize venture capital (VC) returns. Complete capital loss is common, while large positive returns can determine the survival of a VC fund. The frequency of large negative and positive returns typical of VC investments do not resemble those predicted by the log normal distribution. The current literature assumes normality for log returns, trims extremes and potentially ignores risk heterogeneity by pooling distinct outcomes such as bankruptcies and "home runs." Confident in the accuracy of observed extreme returns, I must address tail events in any risk and return estimator. Failure to account for the unique features of venture capital investments could result in misleading estimates of risk and return used by both investors in venture capital and venture capitalists.

I present a model of VC risk and return that incorporates typical extreme investment outcomes and I show that a mixture distribution best fits the data. Estimates implies a risk and return environment with negative skewness and excess kurtosis and thus suggest that the standard volatility risk measure offers poor inference about return outcomes. My work extends previous studies (Cochrane (2005), Korteweg and Sorensen (2009)) by accounting for non-normal features of the returns distribution. Two new implications emerge from this study. First, the significant and extreme downside risk revealed by the mixture is not replicated in the standard public equity microcap index. Venture capital returns present a unique investment opportunity distinct from the standard public benchmark. Second, the deviations from normality imply that standard risk measures do not adequately characterize venture capital returns as they underestimate the frequency and magnitude of tail events. Thus, investors in venture capital funds may need an alternative risk measure for asset allocation. The final shape of the returns distribution indicates that conventional, single-state models underestimate downside risk.

The paper's first contribution is a model of risk and return that incorporates tail events. A continuous time returns process approximates the evolution of entrepreneurial firm value and includes the Fama and French pricing factors. This structure requires the calculation of round-to-round returns, which measure the increase in valuation between either two private financing events or a financing and exit. Although not explicitly earned by venture capitalists, round-to-round returns approximate portfolio outcomes and increase the number of available return observations. Through the parameterization of the state means, a K-state normal mixture distribution allows return characteristics such as market risk to differ across state. Moreover, as a semi-parametric estimator, a mixture model incorporates multi-modal distributions, fat tails and skewness. Following Berk, Green and Naik (1999,2004), I let entrepreneurial firm size impact the mixing probabilities. The final state probabilities map to ex-ante likelihoods of each return outcome. I estimate the model with financing data covering 1987-2007 from the VentureOne database that includes over 50,000 financing events of over 15,000 entrepreneurial firms.

Despite the database's comprehensiveness and lack of systematic measurement error, sample selection bias and endogenous regressors hinder any attempt to estimate unbiased risk characteristics. Observed venture capital returns suffer from sample selection while investment holding period – a key regressor in this paper's returns model – is endogenous in the market model. Cochrane (2005) presents a novel, intuitive sample selection model whose incorporation into a returns process dramatically alters risk and return estimates. I use a rich set of entrepreneurial firm observables and a more flexible selection function to correct for selection bias. Using lagged capital stock as an instrument, I model cross-sectional sample selection in VC returns with a Heckman-like ordered probit proposed by Hwang et. al. (2005). The theory behind CAPM gives instrument validity and estimates indicate that the instrument is not weak. Next, the venture capitalist investment process generates shorter holding periods for large returns and thus a negative correlation between a regressor and the error term. This endogeneity results in too small a beta and too large an alpha. I incorporate endogenous holding periods in a duration model where historical industry exit times provide exogenous variation.

The paper's next major contribution combines each of these empirical problems into a tractable estimator from Wooldridge (2003). A three-stage model integrates sample selection, the endogenous holding period and the mixture returns process to produce unbiased risk and return estimates. Before estimating the full mixture model, I run a single-state model for comparison to previous results in the literature. The structure of the returns model requires a GLS correction that divides all variables in the discrete estimator by the square root of the holding period.[2] Single-state estimates give an arithmetic alpha of 40%, annualized log volatility of 99% and beta of 1.9. The corrections for endogenous holding periods and sample selection have the expected impacts on parameter estimates: beta increases and alpha falls.

Next, I test whether a K-state mixture (K > 1) better fits the observed returns. Both statistical tests and economic reasoning suggest that a three-state returns model is a better fit than the normal, single-state process widely used in the literature. The three-state mixture estimates present a venture capital returns distribution composed of tail events with distinct risk characteristics. Using the mean log return, mixture states map to "winners," "failures" and "break-even." The implied CAPM alpha and beta depend on the individual state. Right tail returns earn a 101% alpha with an insignificant market loading, compared to a statistically insignificant alpha for left tail outcomes. Ex-ante, 25% of returns draw from the "failure" state (mean log return of -200%), while 62% of returns stem from a distribution with a mean log return of 22%. The state probabilities predict that "winners" are four times as likely for entrepreneurial firms with a capital stock of less than $1 million than for firms that have raised more than $10 million. In contrast to Cochrane (2005), estimates show that systematic risk increases with entrepreneurial firm size. Surprisingly, the probability of the failure state is constant across an entrepreneurial firm's life. Although the population log volatility of 104% differs little from the single-state estimate, the final mixture estimates present a returns distribution with excess kurtosis and negative skew.

The non-normal features of the mixture distribution differ from current results in the literature and show that characterizing risk in venture capital requires more than volatility. As with any non-symmetric returns distribution, volatility of VC investment outcomes fails to characterize the frequency and magnitude of tail events. In particular, only 52% of returns are within one standard deviation of the mean return compared to 68% for a normal distribution. The negative skew also reveals a downside risk ignored by standard single-state estimators as shown by a semi-variance 50% larger than population variance. These non-normal features indicate that standard moment conditions used in asset allocation miss distinctive features of VC outcomes. LPs and even venture capitalists may need to incorporate different measures of risk such as semi-variance for optimal portfolio choice (e.g. Markowitz (1959)).

I compare the mixture distribution to that of a microcap index shown by Cochrane (2005) to mimic means and alphas of venture capital investments. I investigate whether the same conclusions about non-normality and downside risk are present in this index. I find that a single-state normal model rather than a three-state mixture characterizes all deciles of the microcap index. The microcap index lacks the skewness and kurtosis

[2] I show in the Appendix that one must divide by the predicted holding periods from the second stage duration model to ensure that the endogeneity does not hamper unbiased parameter estimates.

features found in VC investment returns. I conclude that VC assets add unique risk and return features to a portfolio of public stocks.

I check the robustness of the model's assumptions on estimates. All estimators require an assumed value for bankrupt firms earned by the VC investors, which I set to 25% of the original investment. Results are insensitive to a wide and sensible range of this cutoff values for out of business outcomes, while general conclusions about alphas, betas and distribution shape are only sensitive to very small cutoffs. As the out of business cutoff value falls, the returns distribution approaches a four-state mixture with the same basic population estimates of the three-state model. Finally, the results are robust to additional mixture states. No information is lost if the model is re-run with a four-state mixture. The tails shift outward slightly, but all major means, volatilities and risk parameters remain unchanged.

The paper proceeds as follows. Section II and III describe the literature and motivate the mixture model. Section IV details the mixture model and returns model, while Section V summarizes the data. I present an estimator in Section VI and the results in Section VII. Finally, Section VIII compares the estimated returns distribution to public markets and Section IX discusses robustness checks.

2 Literature

The distinction between venture capital funds and venture capital financings translates into two different types of returns. Cochrane (2005) is the most comprehensive investigation of the returns to venture capital investments. His innovative model illustrates the severe sample selection of investment-level VC data. With his estimates, he concludes that "there is nothing special about venture capital per se" as he demonstrates its returns closely mimic a unique NASDAQ micro-cap index over the same time period. Hwang, Quigley and Woodward (2005) and Peng (2001) build indexes of venture capital investments to gauge their risk and return against public equity returns. Hwang et. al. find that investments in venture capital do not significantly outperform the market index, though they show there exist optimal portfolio allocations with some share of private equity. Using a modified and defunct version of the VentureOne data, Peng finds a strong correlation between venture capital returns and the NASDAQ and thus a CAPM beta of over 4.

The literature on VC fund returns compares them to public assets and investigates its cross- sectional features. Kaplan and Schoar (2005) show that S&P 500 returns exceed equal-weighted VC fund returns, while size-weighted returns marginally exceed those of public markets. However, Phalippou and Zollo (2005) argue that the size weighting procedure creates an upward bias on return calculation as it over-weights most recently raised VC funds. Phalippou and Gottschlag (forthcoming) argue that risk-adjusted private equity fund returns underperform the S&P 500 by 6% a year. Jones and Rhodes-Kropf (2003) finds that VC fund returns earn a zero alpha on average, confirming their hypothesis that the investor market is perfectly competitive. Using a database weighted toward buyout funds and non-high-tech firms, Ljungqvist and Richardson (2003) conclude that fund returns exceed the S&P 500 for 1980 - 1996. Several studies investigate the cross-sectional distribution of VC fund returns.

Kaplan and Schoar (2005) also find significant performance heterogeneity and persistence for top funds. Sorensen (2007) shows that are sorting plays a leading role in explaining heterogeneity, while others (e.g. Ljungqvist, Hochberg and Lu (forthcoming)) show that an investor's social network predicts success. Finally, Gompers (1996) and Jones and Rhodes-Kropf (2003) suggest that a VC's early success at finding "winners" determines both survival and future returns.

3 Typical Entrepreneurial Outcomes

I present a new model of venture capital risk and return motivated by the typical development process of entrepreneurial firms. Consider a fictional biotechnology firm BioMix that just received its first venture investment. BioMix is developing a new gene therapy for sale to the public, which requires completion of three research and development stages as illustrated in Figure 1.

BioMix must first complete drug development. If successful, BioMix receives second VC investment and enters the next stage. If drug development fails, the firm shuts down and investors lose most of their capital. The second stage begins human trials of the drug and applying for FDA approval. Failure at this stage results in another loss for investors, but one where assets like patents can be sold to recoup some of the invested capital. If the FDA approves the drug for sale, Bio Mix begins production and marketing. The size of the drug's market determines the final return earned by investors. A market with little competition results in a huge return to investors, while a market with a new rival substitute drug generates below-average returns. This story illustrates three basic return outcomes: winners, failures and break-even. Many venture capitalists break their portfolio outcomes into such groups. For example, the venture capital firm Union Square Ventures presented prospective limited partners with expectations of one-third of their portfolio outcomes in each scenario.[3] The returns model incorporates this multi-state returns structure.

4 Model

4.1 Mixture distribution

The BioMix story and VC expectations suggests that VC return outcomes fall into three basic categories: winners, failures and break-even. A standard model for estimating an outcome equation when distinct processes generate observations is a mixture model. A cross-sectional mixture model supposes there are K unique states of the world, where each state follows a process defined by a pdf $f(\cdot)$. Estimation of a mixture seeks to identify the number of states, their probabilities and their characteristics. Typically, the states are distinguished by the first two moments of the distribution $f(\cdot)$. The outcome equation throughout this paper will be a returns process to investing in entrepreneurial firms. In the BioMix example, I hypothesize that the mean return to each of the possible success and failure outcomes differ.

Suppose the outcome of an investment in entrepreneurial firm i's j'th financing event is generated by one of K mixture states. Let S_{ij} indicate the mixture state of the return to this investment, R_{ij}. Then

$$P(S_{ij} = k) = p_k \qquad k = 1, 2, \ldots, K \tag{1}$$

with $\sum_{k=1}^{K} p_k = 1$. I assume throughout that returns to each state are normally distributed with mean μ_k and volatility σ_k. The returns process parameterizes the mean and thus produces the typical risk and return characteristics. The final pdf of all returns (R) is a linear combination of each state distribution:

[3] http://www.avc.com/a_vc/2008/08/venture-fund--1.html

$$f(R) = \sum_{k=1}^{K} p_k f_k(R) \tag{2}$$

where $f_k(R) = f(R|S_i = k)$ is a normal pdf with mean μ_k and volatility σ_k. As formulated, this mixture distribution resembles the measurement error model in Cochrane (2005).

Motivated by a significant amount of large annualized returns immediately prior to IPOs and data entry mistakes, Cochrane (2005) addresses return outliers with a uniform measurement error process on all returns that trims π% of the returns distribution (where π is estimated). Without the measurement error process, his log market model fits all observed returns with an enormous mean and variance influenced by a few outliers in both tails. The uniform trimming predominantly removes the positive and negative tail events. As I will discuss below, the database used in this paper suffers from little systematic measurement error, so a new approach is required to deal with extreme outcomes. A mixture accommodates extreme outcomes and potentially presents economically intuitive states that mimic those of the BioMix outcomes. A mixture model is also a flexible estimator that can incorporate non-normal returns, fat tails and skewness. Such features are an accepted fact of most VC investment returns, however, the literature has yet to incorporate them into an estimator.

4.1.1 State probabilities

Motivated by theory and past empirical results, I enrich the probability distribution over mixture states in (1). Berk, Green and Naik (1999) attempt to explain how and why the Fama and French pricing factors describe the cross-section of expected returns. They show that differences in a firm's proportion of "growth options" to assets in place can explain differences in returns between small and large firms. Small firms have a larger proportion of growth options which alters their expected returns. Berk, Green and Naik (2004) formulate a model of staged investments in R&D projects that predicts differences in risk premiums across the project's life. Finally, Cochrane (2005) estimates alpha's and betas across development stage and shows systematic risk falls as firms develop. These results suggest that the development stage of an entrepreneurial firm predicts outcome and should be incorporated into the probability a return is in a mixture state. The finance literature on state switching follows this same approach when it incorporates covariates into transition probabilities (see Hamilton and Owyang (2009) and Perez-Quiros and Timmermann (2001)).

In equation (2), the mixing probabilities p_k are fixed across returns. The results discussed above indicate that for given states of the mixture, entrepreneurial firm size impacts state membership. Formally, let there be a covariate $Z = (Z_1, \ldots Z_N)$ that proxies for firm size at each financing event. Then let state probabilities p_k follow a multinomial logit:

$$p_k = \frac{\exp(Z\rho_k)}{\sum_{k=1}^{K} \exp(\rho_k Z)} \tag{3}$$

If the coefficients $(\rho_1, \ldots \rho_K)$ are insignificant, the model reduces to fixed state probabilities. Finally, identification of the coefficients of the state probabilities requires (Jiang and Tanner (1999)) a covariate not included in the returns model. I discuss variable options in the data section below.

4.1.2 Caveats

I argue that the mixture model can identify unique groups within the returns distribution that map to typical outcomes in venture capital investments. McLachlan and Peel (2000) show that if this is a goal of estimation, then the number of components found for the mixture may not directly map to the number of groups in the underlying population. The disconnect stems from how a mixture of normals fits a skewed distribution. If some of the states of the mixture are in fact better approximated by a skewed distribution, then "more than one normal component may be needed to model a skewed group-conditional distribution." (pp 176) I return to this issue when testing for the number of states in the mixture.

4.2 Returns

I now discuss the returns process underlying each state of the mixture. I assume the development of entrepreneurial firm valuation follows a log-continuous process. The prevalence of outliers requires the use of log returns. I follow the model presented by Cochrane (2005):

$$d \ln V = (r^f + \gamma_k)dt + \beta_k(d \ln V^m - r^f dt) + \sigma_k dB^\nu \tag{4}$$

$$d \ln V^m = \mu_m dt + \sigma_m dB^{\tilde{m}} \tag{5}$$

where V and V^m are the values of the entrepreneurial firm and market and the subscript k represents the state of the K-state mixture. I assume that $E[dB^\nu dB^{\tilde{m}}] = 0$ and let $dB^\nu \sim N(0, dt)$ (i.e. Brownian motion). The dependence between the return variance and holding period dt will factor prominently in parameter estimation. Cochrane shows that (4) implies the following continuous time returns process in levels:

$$\frac{dV}{V} - r^f dt = [\gamma_k + \frac{1}{2}\beta_k(\beta_k - 1)\sigma_m^2 + \frac{1}{2}\sigma_k^2]dt + \beta_k(\frac{dV^m}{V^m} - r^f dt) + \sigma_k dB^\nu. \tag{6}$$

For the above, the mapping from the log market model parameters to the standard level-CAPM α and β is:

$$\alpha_k = \gamma_k + 1/2\beta_k(\beta_k - 1)\sigma_m^2 + 1/2\sigma_k^2 \tag{7}$$

and β_k is the same as in the levels market model.

4.2.1 Timing

Venture capital returns rarely occur over periodic intervals. Let entrepreneurial firm i have financing events at the following calendar times:

$$\tau^i = (\tau_{i0}, \tau_{i1}, \ldots, \tau_{iT^i}) \tag{8}$$

where $T^i + 1$ is the total number of financing events for firm i and τ_{ij} is the calendar date of each. For example, a τ^i might look like:

$$\tau^i = (4/4/98, 5/15/2000, 10/12/2000, 6/1/2004).$$

Define an exit for a VC financing as any subsequent capital infusion that involves transfer of equity. The resulting round-to-round returns includes IPOs, acquisitions, bankruptcies and private financings. The holding period for firm i's financing j is

$$\tau_{ij+1} - \tau_{ij}.$$

The holding period unit throughout the paper will be months. Note that if VC financings occurred at regular intervals, the holding period would be constant across j for each firm i. Write the T^i observed log returns as

$$r^i = (r_{\tau_{i1}}, r_{\tau_{i2}}, \ldots, r_{\tau_{iT^i}}) \tag{9}$$

where $j = 1, 2, \ldots, T^i$ and

$$r_{\tau_{ij}} = \ln \frac{V_{\tau_{ij+1}}}{V_{\tau_{ij}}}. \tag{10}$$

$V^i_{\tau_{ij}}$ is the value of the firm at time τ_{ij} and $r_{\tau_{ij}}$ is the log return to investing in the j financing for firm i and selling that stake $\tau_{ij+1} - \tau_{ij}$ months later at the $j + 1$th financing. I use non-annualized returns in all estimations (with holding period as an independent variable), while the returns incorporate ownership dilution across financing events and holding period as a dependent variable. I can now write the discrete version of (4), which takes the following form:

$$r^i_{\tau_{ij}} = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \epsilon_{k,\tau_{ij}} \tag{11}$$

where $r^m_{\tau_{ij}} = \ln \frac{V^m_{\tau_{ij+1}}}{V^m_{\tau_{ij}}}$ and $\epsilon_{k,\tau_{ij}} \sim N(0, \sigma^2(\tau_{ij+1} - \tau_{ij}))$. The distribution of $\epsilon_{\tau_{ij}}$ follows from the Brownian motion assumption and represents the prevalent heterogeneity in VC investment holding periods. I use (11) for all estimations of risk and return, which resemble the round-to-round returns in Cochrane (2005). The simplified returns model in (11) ignores overlapping value revelation across entrepreneurial firms and therefore possible correlations between them. Finally, inclusion of additional pricing factors like Fama and French (FF) are simply added onto (4) and the discrete model. The only change is the mapping of log parameters to level parameters, where α_k becomes:

$$\alpha_k = \gamma + .5 * \zeta' diag(\Gamma) - .5\zeta'\Gamma\zeta \tag{12}$$

with $\zeta = (\beta, SMB, HML)$ and Γ is the covariance matrix of market returns and FF factors "small minus big" and "high minus low." The population variance is a product of state means and volatilities:

$$\sigma^2 = \sum_{k=1}^{K} p_k(\sigma_k^2 + \mu_k^2) - \mu$$

where $\mu = \sum_{k=1}^{K} p_k \mu_k$ is the population return.

The returns model is one of value paths and the returns they generate, while the data is a cross-section of thousands of path realizations. Thus, I assume that each realization is an independent draw of the returns model. Simply, each observed return informs the estimator about the true path of entrepreneurial firm valuation.

4.3 Sample Selection

The ideal data set of venture capital financings would include periodic value revelations by all entrepreneurial firms. The database, however, suffers from two related problems:

1. Sample selection in returns: high value, high quality firms reveal their value more often.
2. Endogenous holding periods: the timing of the financing events – and thus value revelation – occurs more quickly for high value, high quality firms.

Hwang et al. (2005) summarizes the first problem: "[F]irms receiving financing at any point in time are those whose prospects are more promising than other firms'." Simply, investors invest in entrepreneurial firms whose expected return far exceeds the investor's cost of capital; a condition met more often for growing, high value firms. As returns are generated in the database with new exit events, it includes relatively more exits for financings of high-quality firms. Moreover, entrepreneurial firms that exit via initial public offering (IPO) typically require more capital and thus more financing events than firms that fail to go public. Thus, the database includes more returns associated with IPOs than failures and estimates of means and arithmetic α are too large.

Any corrections for sample selection must model VC exit events. Although an entrepreneurial firm can "exit" via bankruptcy, IPO or acquisition, the company's financings can exit in a slightly different way. The three possible outcomes for a VC investment are:

- new financing round $\implies$ firm can have more financing events
- IPO or acquisition $\implies$ firm has no more financing events
- bankruptcy $\implies$ firm has no more financing events

The fourth state is the non-exit event: the firm did not receive new capital by the end of the sample period and the last known financing event lacks an exit. IPOs, acquisitions and bankruptcies are absorbing states.

4.3.1 Model

I follow the sample selection correction approach of Hwang et. al. (2005) and model the exit events as an ordered probit. The model should predict outcomes that lead to observed values of the outcome equation. Ignoring missing data, new financing events generate observed returns when existing investors sell their equity stakes. One can think of financing events as times when the underlying value of the firm is revealed by new investors who buy out the existing investors.

Let each financing event have an unobserved latent variable defined by:

$$Y_{ij}^* = \theta Z_{ij} + \nu_{ij} \qquad \nu_{ij} \sim N(0,1) \tag{13}$$

Here Z_{ij} includes the independent variables from the returns model and an instrument for firm i's, j'th financing discussed below. Selection follows from $E[\epsilon_{\tau_{ij}}|\nu_{ij}] = \rho\nu_{ij}$ where $\epsilon_{\tau_{ij}}$ is the error term in the returns model. Financings that exited in bankruptcy have a low realization of Y^*_{ij} while those that exited via IPO, new financing or acquisition have a high realization. The remainder of financings have yet to exit at the end of the sample either because the firm's poor prospects generated little interest from investors or there was too little time to observe the exit. Define

$$Y_{ij} = \begin{cases} 0 & \text{if financing exits in bankruptcy } Y^*_{ij} < \alpha_1 \\ 1 & \text{if no financing occurs as of end of sample } \alpha_1 \leq Y^*_{ij} < \alpha_2 \\ 2 & \text{if financing exits with IPO, acquisition or new round } Y^*_{ij} \geq \alpha_2 \end{cases} \tag{14}$$

for unknown cutpoints $\alpha_1 < \alpha_2$.[4] Given the distributional assumptions for ν_{ij}, the conditional distribution of Y_{ij} given Z_{ij} is:

$$P(Y = 0|Z) = \Phi(\alpha_1 - Z\theta)$$

$$P(Y = 1|Z) = \Phi(\alpha_2 - Z\theta) - \Phi(\alpha_1 - Z\theta)$$

$$P(Y = 2|Z) = 1 - \Phi(\alpha_2 - Z\theta).$$

where $\Phi(\cdot)$ is the standard normal cdf. Finally, log returns are only observed when $Y_{ij} \in (0, 2)$, which will be the conditioning statement in the log market model. In particular,

$$E[r^i_{\tau_{ij}}|Y^*_{ij} > \alpha_2] = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \rho\sigma_k Mills_{ij2} \tag{15}$$

$$E[\ln r^i_{\tau_{ij}}|Y^*_{ij} < \alpha_1] = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \rho\sigma_k Mills_{ij1} \tag{16}$$

where

$$Mills_{ij2} = \frac{\phi(\alpha_2 - \theta Z_{ij})}{\Phi(-(\alpha_2 - \theta Z_{ij}))}$$

$$Mills_{ij1} = -\frac{\phi(\alpha_1 - \theta Z_{ij})}{\Phi(\alpha_1 - \theta Z_{ij})}$$

with ϕ and Φ are the standard normal pdf and cdf[5].

4.3.2 Selection and mixture

The combination of a mixture model and a sample selection model results in an important assumption:

$$E[\epsilon_{il}|S_i = j, Y_{ij} > \alpha_2] \neq E[\epsilon_{il}|S_i = j, Y_{ij} > \alpha_2]$$

for $j \neq k$. Simply, the sample selection effect depends on mixture outcome because I let the coefficients on the inverse Mills terms to depend on the mixture state. For example, conditional on a "bad" state outcome,

[4] I define the latent variable at the financing level, so it effectively restarts for each firm at each financing event.

[5] See Hwang et. al. (2005) for derivation.

the effect of selection may be relatively small. Incorporating this assumption requires a straightforward modification of the sample selection conditioning in (13):

$$E[\epsilon_{k,\tau_{ij}}|\nu_{ij}] = \rho_k \nu_{ij}$$

Thus, (15) and (16) require further conditioning on mixture state to become:

$$E[r^i_{\tau_{ij}}|S_i = k, Y^*_{ij} > \alpha_2] = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \rho_k \sigma_k Mills_2$$

$$E[r^i_{\tau_{ij}}|S_i = k, Y^*_{ij} < \alpha_1] = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \rho_k \sigma_k Mills_1$$

I test this assumption in the mixture estimates by imposing fixed Mills coefficients across states and testing whether the restriction improves model fit.

4.4 Endogenous holding periods

The second issue hindering unbiased parameter estimation stems from endogenous timing:

2. Endogenous holding periods: the timing of the financing events – and thus value revelation – occurs more quickly for high value, high quality firms.

Hall and Woodward (2008) summarize the problem as follows:

> Investments with longer holding periods tend to have lower per-period returns, so they have negative values of the idiosyncratic component [...]. Longer holding periods raise the cumulative [market] return. Thus the two variables are negatively correlated. (page 27)

Unlike public market investments, if you hold long in VC, the returns tend to fall. The resulting bias generates too low a CAPM beta and too high a CAPM alpha. The bias stems from a process where financing events occur after the underlying value of the firm reaches some unobserved threshold. Entrepreneurial firms whose value is growing will not only have relatively more financing exits than the average firm, but the time between the events will be relatively shorter. Cochrane (2005) models this idea explicitly with a selection function that resembles a duration model of venture capital exits conditional on valuation. The data on venture capital holding periods confirms the endogeneity problem.

Table 1 compares holding periods across different levels of returns. The latter illustrates that the mean holding period weakly falls as observed returns increase. Although median durations do not fall directly with returns, going from an investment that returns 200% to one that returns over ten times (150% annualized vs. 650% annualized) is significant. Doubling a return without doubling the holding period increases the annualized return dramatically.

5 Data

I use the VentureOne database maintained by the Dow Jones subsidiary VentureSource. It covers the venture capital market from 1987 to 2007 and tracks investments of $361 billion in 55,457 rounds and 16,849 companies. VentureOne collects their data on venture-capital backed companies with surveys of venture capitalists and entrepreneurial firms. The data lacks financings where corporations or institutions are the only investors in a particular company. For all financings that involve a venture capitalist, VentureOne claims near perfect coverage of 1992 – present. Smaller (in total assets) venture capitalists that invest alone in an entrepreneurial firm are less likely to be observed than the average size venture capitalist. However, if the small investor co-invests with larger venture capitalists already in the dataset, VentureOne starts to actively track the former's financings.

I corroborated 184 financings with two active venture capitalists to see if there are any missing rounds and found none.[6] Moreover, I only found a few instances ($< 1\%$) where the numbers reported by VentureOne differed from the true value reported by the VC. In a similar study, Kaplan, Sensoy and Stromberg (2002) use detailed deal data to cross-check 143 financings in the VentureOne database. The authors find a bias toward reporting California-based companies and estimated that approximately 15% of all financings were missing. Overall, they conclude that VentureOne represents a relatively more reliable source of venture capital data than the major alternatives.

CRSP and Global Financial Data provide the market return data. Using the investment and exit dates, I match the values of the S&P 500 and Wilshire 5000 index over a financing holding period. I settle on the Wilshire 5000 as the market factor because it contains a more diverse set of firms in terms of both size and industry, but results are similar with the S&P 500 as the pricing factor. Next, I describe the collection and features of each data set: the exit, holding period and returns data.

5.1 Exit data

Over their lifetime, venture capital-backed firms have multiple investments or staged financings. After a sequence of financings and several years, the firm has an IPO, acquisition or shuts down. Firms founded in the last few years are much less likely to have such exits simply because there has not been enough time.

Table 2 shows that the average firm has over two financing events. It takes an average of 2.8 financing rounds for firms to go public or be acquired. Bankruptcies occur more quickly, with an average of 2.2 financing rounds until exit. Firms go bankrupt an average of 5.6 years after founding, while newly-public firms are 7 years at the IPO.

Table 3 shows the distribution of exit types for financings, where exits are defined on a round-to-round basis rather than investment-to-liquidation. Over 85% of financings have an "exit" event. The statistics in Table 3 differ from the proportion of firms that are publicly-held or bankrupt in Table 5. Some 10% of financings exit via a bankruptcy or acquisition, while 16% have yet to exit as of the end of the sample. Standard financing exits account for 61% of all exits and include such events as a first round financing "exiting" in the second round.

5.2 Holding periods

The 20 years of data include over 37,000 non-exit, standard financing events with relevant censored and non-censored exit times. Table 4 presents basic statistics on the holding periods. On average, entrepreneurial firms receive new capital every year and a half. Figure 2 shows the estimated survivor function with a rapidly

[6] The two venture capital firms have more than $1 billion and less than $100 million in assets respectively.

increasing probability of exit for the first two years after the investment. Entrepreneurial firms without new rounds as of the end of the sample have censored holding periods. For all variables that require a start and end time, the last date in the sample – 12/31/2007 – is used as the end point.

5.3 Returns

A venture capitalist can liquidate her investment in several ways. IPOs and acquisitions represent explicit transfers of ownership to either the public or another corporation. The VentureOne database includes excellent information on the former. I use the market capitalization of the firm at offering as the sale price for any financing event exiting via IPO and if reported, the acquisition price either in shares of the buying firm or cash. Unfortunately, acquisition prices are only available for about 30% of all acquired firms. The other possible exit outcome is bankruptcy. The VentureOne database does not include explicit dates for these events, but does include dates when the firm's records were last updated. Updates for all firms occur quarterly, so I use these dates for bankruptcy exits. Over 200 entrepreneurial firms in the database have not had a new financing in over 5 years and lost contact with VentureSource. To identify any possible bankruptcies (or other exits) for this set of firms, I tracked each of them using a combination of web searches and phone calls. The final database represents a clear picture of the venture capital market for 1987 - 2007.

Round to round returns require either an exit with a known valuation or two subsequent financings with reported valuations. In particular, the post-money valuation ($Post$) of the financing of interest and the pre-money valuation (Pre) of the subsequent round are required to calculate a return. The log return for firm i's j'th financing is

$$\ln(R_{\tau_{ij}}) = \ln(\frac{Pre_{\tau_{ij+1}}}{Post_{\tau_{ij}}})$$

Out of business outcomes result in near to complete loss of capital. The latter cannot be used in a log returns model, so I follow Korteweg and Sorensen (2009) and the results of Cochrane (2005)'s selection model. I assume that the final exit value is 25% of the dollars invested if the investment fails. Robustness checks for this out of business cutoff demonstrate some sensitivity to results for cutoff values less than 10%, which logged are extremely negative. However, the basic results hold for a wide range of out of business cutoffs.

Table 6 summarizes round-to-round returns data for non-bankrupt firms. Due to missing data, only some 10,000 of 32,000 exits have known returns. Both arithmetic and log returns are large. Annualization has a dramatic impact on mean arithmetic returns, with an average 8548%. The value increases when I include those financings that exited in less than 1 month. IT firms have higher returns than health care firms, while "real" exits like IPOs or acquisitions result in higher mean returns than other round-to-round returns. Finally, the 89% return to private financing implies that the average new capital infusions represents a large increase firm valuation. Table 7 summarizes returns when the bankruptcy outcomes are included.

6 Estimation

Sample selection and endogeneity hinder unbiased estimation of risk and return parameters. Wooldridge (2002) presents a 3-stage estimator for just this problem. Let the population outcome equation be:

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + u_1$$

with endogenous y_2 defined as

$$y_2 = z\delta_2 + \nu_2. \tag{17}$$

where z includes z_1. In the context of the returns model, y_1 is

$$r^i_{\tau_{ij}} = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \epsilon_{k,\tau_{ij}}$$

and the endogenous variable y_2 is the investment holding period $\tau_{ij+1} - \tau_{ij}$.

Equation (17) is a simple linear projection of the endogenous variable on a set of instruments z. In venture capital, observation of $(r_{\tau_{ij}}, \tau_{ij+1} - \tau_{ij})$ (i.e. (y_1, y_2)) depends on the outcome of a third random variable y_3. Observable returns and holding periods occur when a firm has a financing event. Define this variable in the standard selection equation probit setting:

$$y_3 = 1\{z\delta_3 + \nu_3 > 0\}$$

For the venture capital setting with three possible outcomes, the y_3 model will be the ordered probit in equation (13) defined by the variable:

$$Y^*_{ij} = \theta Z_{ij} + \nu_{ij}$$

Wooldridge shows that identification of the model requires the following assumptions:

1. (z, y_3) is always observed. (y_1, y_2) is observed when $y_3 = 1$.
2. (u_1, ν_3) is independent of z
3. $\nu_3 \sim N(0, 1)$
4. $E[u_1|\nu_3] = \rho\nu_3$
5. 5.$E[\mathbf{z}'\nu_2] = \mathbf{0}$ with $z\delta_2 = z_1\delta_{21} + z_2\delta_{22}, \delta_{22} \neq 0$.

Estimation of this model follows two simple steps:

1. Estimate $\hat{\delta}_3$ from a probit of y_3 on z using all available observations. Calculate the inverse Mills term for each observed y_{1i}, $\hat{\lambda}_{i3}$.
2. Estimate $y_{i1} = \mathbf{z}_1\delta_1 + \alpha_1 y_{i2} + \rho\hat{\lambda}_{i3} + r_i$ with the selected sample using 2SLS with instruments $(z_i, \hat{\lambda}_{i3})$.

Two issues remain. First, identification of the population parameters requires two separate instruments: one each for the endogenous variable and selection model, a potentially high hurdle for venture capital data. Second, Wooldridge stresses that "all exogenous variables should appear in the selection equation, and all should be listed as instruments"[7] for the 2SLS linear projection. For the venture capital returns model, this requirement conflicts with unobserved explanatory variable market return for those firms that do not have new financing events.

[7] page 569

6.1 Applying the model to venture capital

The selection and endogeneity problems in venture capital data conform to the Wooldridge three-stage model. I make several modifications to the model for the venture capital data:

1. The endogenous variable y_2 is holding period $t_{ij} - t_{ij-1}$ in months. I thus estimate a duration model rather than a standard linear projection.

2. The selection equation for y_3 must account for multiple exit types: exit, bankruptcy and still private. Thus, the selection model is an ordered probit as in (13).

3. A fairly large set of regressors clearly impact both the exit type and the timing of the exit. These include firm age, location and basic investor characteristics. Excluding these regressors from the outcome equation – the returns model – requires some strong assumptions. I rely on the theory behind CAPM and only include the pricing factors in the returns model. Basic results are unchanged when I exclude all these variables from the estimator.

4. The other exogenous variable in the outcome/returns model – the market return $\ln \frac{V^m_{t_{ij}}}{V^m_{t_{ij-1}}}$ – is only observed when a financing occurs. Therefore, I cannot explicitly include it in the set of exogenous variables z. I use the market return as of the end of the sample as a replacement for this variable when the end point is unknown.

6.2 Instruments

The last identification condition of the three-stage model requires two instruments: one each for the duration and sample selection model.

6.2.1 Selection Model

Identification of the ordered probit stage of the estimator requires a financing-level variable that predicts exit events but does not belong in the returns model. Using the Census Bureau's database on business ownership, Fairlie and Robb (2008) show that the best indicator of entrepreneurial success is startup capital. For the sample selection instrument I propose the lag of the sum of the firm's capital stock. Let k_{ij} be the capital raised in firm i's j'th financing. Then the capital stock raised as of the j'th financing is:

$$K_{ij} = \sum_{l=1}^{j-1} k_{il}$$

This value will be zero for first financings, however, this lack of variability is not an issue because of the lack of sample selection for these events. All firms have a first financing event.

The likelihood of a new financing event – and thus a return observation – depends on the underlying health of the firm, itself closely related to the current capital stock. All else equal, I argue that the probability of a new financing (other than still private or bankruptcy) increases in the past capital stock, but the unobserved firm quality that correlates with IPO exits does not correlate with past capital stock. My focus on the exit event of the current financing rather than the firm as a whole helps to reinforce the exclusion restriction. Firm quality is presumably fixed, while the likelihood of a new financing event changes with time. A relatively large past capital stock indicates the firm can survive longer than average and also shows its investors' confidence that the expected return to investment was greater than their cost of capital. Moreover, I rely on the theory behind CAPM outcome equation which implies that only some pricing factors belong in the returns model. Table 8 shows the characteristics of this instrument across the firm development stage. Total capital raised increases as the firm develops, with significant variation within firm development stages.

6.2.2 Duration Model

The duration model requires a variable that correlates with durations but can be excluded from the returns model. The speed at which a similar firm in the same industry exited in the past should be indicative of how the current firm will exit, but not the return earned for the financing event. This variable predicts the state of the VC exit environment but does not correlate with firm quality. Aggregating at the industry may appear to introduce variation that could impact the return. However, given the unique features of entrepreneurial firms, their general industry label (e.g. Information Technology) rarely describes their business model.

Using the full sample of observed holding periods across four industries, I compute the average duration of all past financings in each industry. Let i represent an entrepreneurial firm and $\tau_i j$ be the calendar date of that firm's j'th financing. Define N_m as the set of financing events in industry m that occurred (i.e. exited) at least a month prior to τ_{ij}. Then the proposed instrument is calculated for industry m as follows:

$$D_{\tau_{ij},m} = \frac{1}{\#\{N_m\}} \sum_{(k,l)\in N_m,\tau_{kl}<\tau_{ij}} \tau_{kl} - \tau_{kl-1} \tag{18}$$

Note that the instrument varies for each financing event with a unique date and industry. Table 9 shows the basic characteristics of holding periods by each of the four major industries before and after 1998 (the beginning of the boom period). Holding periods have fallen since 1998 and have sufficient variation across industry. The instrument (18) averages these holding periods for each financing event date, while the lag ensures that it satisfies the exclusion restriction.

6.3 Modeling holding periods

Accurate modeling and estimation of VC holding periods may uncover features of the dynamics behind VC investments. Cumming (2006) demonstrates that total industry capital and entrepreneurial firm stage both explain cross-sectional differences in durations. Lunde, Timmermann and Blake (1999) study the closure rate of mutual funds and find that funds on the extreme ends of the age distribution are the least likely to close. Cochrane (2005) illustrates the importance of the time-to-exits or holding periods in accurately capturing sample selection in VC returns. His model's fit relied on its predictions of observed exit types over time, effectively estimating a competing risk model of venture capital exits.

The endogenous variable in the returns model is investment duration. Simple OLS will not suffice. Moreover, the estimation of the duration of financing events must capture right-censoring. Fortunately, there exist a wide variety of econometric models to estimate the impact of covariates on durations, which include fully parametric approaches and semi-parametric approaches like the Cox proportional hazard model. A popular hazard rate parametrization – the log-logistic – takes the following form:

$$\lambda(t; x(t)) = \frac{\exp(x(t)\kappa)\nu t^{\nu-1}}{1 + \exp(x(t)\kappa)t^{\nu}}$$

where $x(t)$ are the time-varying covariates, t is the holding period and $\lambda(\cdot)$ is the instantaneous probability of a failure at time t, conditional on survival up to t. There are several other parametric hazards available. I show below that the log-logistic fits the data the best. However, it is important to note that fit or accurate prediction is not an end for this stage of the estimation. Rather, identification simply requires that the instrumental variable produce predicted holding periods through this model whose variation is exogenous to the observed return.

The endogeneity correction requires an estimate of fitted holding periods. Parametric hazards give way to more sensible predicted values, so I focus on this set of models. Finally, I capture the richness of the exit types by using a competing risk model with continuous time durations. Observed returns stem from new financing events or exits such as IPOs and acquisitions. The model separates the two types of "risks" in order to better control for the unobserved heterogeneity. Combined with the sample selection model detailed above, the first two stages of the three-stage model are complete.

6.4 Mixture model estimation

The last estimation step of the 3-stage estimator is the returns equation. The Wooldridge (2002) formulation does not impose any functional form on the outcome equation, so a mixture model in the final stage is available. The sample selection and duration model steps produces Mills terms and predicted holding periods for each observed return. These outputs are incorporated in the mixture returns model (11), which reduces to single-state model when $k = 1$. Estimators of mixture models include the EM algorithm presented in Hamilton (1990) and Everitt and Hand (1981) or Bayesian methods presented in Fruhwirth-Schnatter (2006) and McLachlan and Peel (2000). I use the EM algorithm for all estimations.

6.5 GLS and holding periods

Recall the discrete model of venture capital returns:

$$r^i_{\tau_{ij}} = (\gamma_k + r^f)(\tau_{ij+1} - \tau_{ij}) + \beta_k(r^m_{\tau_{ij}} - r^f(\tau_{ij+1} - \tau_{ij})) + \epsilon_{k,\tau_{ij}}$$

where $\epsilon_{k,\tau_{ij}} \sim N(0, \sigma^2_k(\tau_{ij+1} - \tau_{ij}))$. The variance of returns is proportional to the endogenous holding period. If interest lies in efficient estimation and retrieval of the true conditional volatility σ_k, then one might think to follow the standard GLS procedure and divide all variables by the square root of the holding period. I show in the Appendix that such an adjustment leads to biased parameter estimates because the holding period is endogenous. Fortunately, the 2SLS estimator presents an alternative: the square root of the predicted holding period from the first stage regression. The following steps produce the modified GLS correction:

1. Estimate the duration model for the holding periods.
2. Compute the predicted holding periods for all observed returns: $(\tau_{ij+1} \hat{-} \tau_{ij})$.
3. Divide all variable in the second stage returns regression by $\sqrt{(\tau_{ij+1} \hat{-} \tau_{ij})}$.

4. Estimate the mixture returns model with the modified variables in step 3.

6.6 State probabilities

I allow the mixture state probabilities to depend on entrepreneurial firm size. Ideally, any covariate includes in the probability structure should not belong in the outcome (i.e. returns) model. Fortunately, the lagged capital stock from the selection model is both a proxy for entrepreneurial firm size and presumed exogenous in the log market model. So I let the covariates Z in equation (3) be the same lagged capital stock K_{ij} instrument from the ordered probit estimator. The results below are insensitive to simply using a dummy variable for stage, which is equal to one if the entrepreneurial firm is in the early stages of its development.

7 Results

7.1 Sample selection and duration results

Recall that the exit indicator takes the following form:

$$Y_{ij} = \begin{cases} 0 & \text{if bankruptcy} \\ 1 & \text{if no exit} \\ 2 & \text{if new financing event} \end{cases} \tag{19}$$

I include the log market return, holding period in months and dummies for firm development stage and firm industry in the ordered probit estimation (13). Table 10 presents the results. I do not include time dummies as they perfectly predict outcomes of 1987 - 1991 financings. The last three columns give implied marginal effects at the means of the independent variables. An increase in the market return lowers the probability of bankruptcy and no exit, while it increases the probability of a positive exit. A firm that moves from an early stage of development to a later stage sees a decrease in the probability of a positive exit because nearly all firms have at least two financing events when they are in the early development stages and bankruptcies typically occur after such time. Finally, the marginal effect signs for past capital raised show that an increase in total capital stock increases the probability of a good outcome and decreases the probability of a bad outcome. Thus, the t-stat of 3.8 shows that the instrument passes the relevancy test.

7.1.1 Hazard model choice

I run and compare five hazard function specifications. Although Cochrane (2005) does not explicitly label his selection function as a hazard function, his model for new financing events mimics an exponential hazard with a log-logistic functional form. Simply, time does not impact the probability of exit, while log firm value affects the rates of bankruptcy and new financing rounds.

Table 11 shows that the log-logistic formulation best fits the data using either the AIC or BIC statistic. In unreported results, the weibull, exponential and gompertz each have unintuitive and unrealistic signs and sizes for the coefficient estimates and hazard ratios. I settle on a log-logistic hazard with time-varying covariates and standard errors clustered at the firm level.

7.1.2 Duration model results

The duration model exploits the same set of variables included in the ordered probit with the additional historical duration instrument. I utilize all the available information about exits: censored and uncensored holding periods. A "failure" or "end of spell" occurs when a financing has a subsequent exit event. That event could be an IPO, a new round of financing or a bankruptcy. Figure 3 shows the estimated, non-monotonic hazard. The probability of a new financing increases for the first two years and falls thereafter. Table 12 presents the parameter estimates of a log-logistic hazard with the historical industry durations and Mills terms as instruments. An increase in the market return dramatically increases the probability of a financing event. The coefficient on the Mills term is difficult to interpret as it is only included for identification purposes. Finally, the instrument's relevancy is confirmed by its positive sign and significant coefficient estimate.

Implied hazard functions aid in interpretation of coefficient estimates. Figure 4 shows the difference in hazard rates between early stage firms and non-early stage firms. Younger, less developed firms have new financing events more quickly. I posited that an above average historical duration for the firm's industry would lead to a lower holding period. Figure 5 confirms that the estimated hazard is lower for those firms with an above average historical industry duration. The predicted holding periods from the duration model are now incorporated into the final stage of the 2SLS estimator.

7.2 Single-state estimates

Table 13 presents the full single state model parameter estimates with a one and three-factor model. I focus on the three-factor model throughout. The OLS column shows that ignoring the known form of the returns volatility and sample selection results in a huge volatility estimate and arithmetic alpha. The log intercept and alpha have no time dimension, so are difficult to interpret. The GLS column corrects all observed variables by dividing them by the square root of the (endogenous) holding period. With a more accurate estimate of volatility, it falls almost ten-fold. This decrease in turn lowers the implied arithmetic alpha to 60%. These estimates still suffer from both sample selection and endogenous holding periods.

Recall that sample selection produces too high an alpha and the endogenous holding periods amplify the problem by introducing downward bias in the systematic risk estimate. Comparing column 5 and column 7 of Table 13 shows that the inclusion of both the Mills term and fitted holding periods has an impact on parameter estimates. Systematic risk increases from 1.6 to 1.9, while the arithmetic alpha falls to 40%. With a significantly smaller volatility, one might think that a mixture model is no longer required. However, the returns model may still be misspecified if the true process follows a mixture.

7.2.1 Comparison to earlier results

Although estimated over 1987 - 2000, Cochrane (2005) presents a sensible benchmark to compare these single-state estimates. Recall that his final model included a measurement error process, a single factor and a parametric selection function. Table 14 shows that despite a similar arithmetic alpha and log returns volatility, the new model and data predict a significantly lower mean return than earlier results. Of course, the additional post-Nasdaq years contribute to this difference.

7.2.2 Predicted returns and pdf

Figure 6 plots the kernel estimator of observed log returns against the predicted pdf using the three-factor, single-state model. The differences between the GLS without selection corrections and the full model is volatility. As expected, inclusion of the Mills and endogeneity corrections puts less weight on the right tail and shifts the estimated returns distribution to the left. Overall, the selection and endogeneity corrections have the desired impacts on parameter estimates. I now estimate a mixture model for the last stage of the estimation.

7.3 Mixture step

Before estimating the full mixture model, I must choose the number of states and the underlying distribution for each state. The BioMix story and VC expectations suggest three mixture states for venture capital returns: "winners," "failures" and "break-even". Outliers computed from the residuals of the single-state estimator presented in Table 15 illustrate that extremes are in both the left and right tails.

7.3.1 Choosing the number of states

Testing for the number of states is a difficult problem. Selecting one or two states has been addressed by Cho and White (2007), however, choosing more than two states is an unresolved problem. McLachlan and Peel (2000) summarize the literature and available test statistics. They show that the penalized likelihood statistics AIC and BIC are often helpful, but tend to overestimate the number of states. Biernacki, Celeux and Govaert (2000) present a integrated classification likelihood (ICL) that suffers from less over-fitting problems than AIC and BIC. McLachlan and Peel (2000) show that ICL penalizes complex models and mixtures with significant overlap across states. Naik, Shi and Tsai (2007) present a test statistic that simultaneously selects the number of states and variables in a mixture regression. The returns model presented above gives a theoretical motivation for including all the pricing factors, so I focus on selecting the number states. Fortunately, the large sample of 12,000 returns mitigates the small sample problems of these test statistics. Table 16 presents the AIC, BIC and ICL statistics for two to five states.

The standard choice of using the lowest value AIC or BIC suggests that 4 or 5 states fit the data, while the ICL implies two states. However, I settle on three states for two reasons. First, the AIC and BIC tend to overfit the number of states. Also, recall that I assumed normal states within the mixture for simplicity and to retain mappings from log to levels. If some of the state densities are in fact non-normal (e.g. Student t)[8], then the AIC and BIC will overfit, while ICL is consistent (see Schannter p. 215). Next, the conclusions implied by the 4 and 5 states are very similar to those of three states and several of the state pdfs have significant overlap. In particular, the tail states are approximately the same in terms of risk and return with additional states, while the "middle" of the returns distribution is broken up into more and more states. A three-state world is also confirmed by an additional test.

I also test the null hypothesis of only two states (versus three) using the bootstrapped likelihood ratio test proposed by McLachlan (1987) and conclude that three states fir the data. This modified likelihood ratio statistic takes standard form $-2\ln\lambda$. I run 200 bootstraps and the estimate mixture model for both state types. The assumption of different variances within states means that the estimation must also sample

[8] See Giacomini, Gottschling, Haefke and White (2008) for a t-mixture alternative.

hundreds of possible initializations to find the largest local maximum. The result is a set of 200 test statistics which form the empirical distribution of the likelihood ratio. The value of the likelihood ratio test statistic$-2\ln\lambda$ for the original sample is .94. The statistic compares to the 93rd percentile of the bootstrapped test statistic distribution. Although this result does not pass the 5% rule, combined with the results of the AIC statistics and the economic motivation, I settle on three states. Even if the true number of states is four, little information is lost imposing three states.

7.4 Mixture model results

I estimate the three-state mixture on the three-factor returns model with the Mills terms and fitting holding periods from the first estimation stages. For comparison, Table 17 presents the parameter estimates for each mixture state. Column 1 presents the single-state model. Several specifications consistently produce a negative but insignificant beta estimate for one mixture states so I fix one state's beta at zero. This value is more sensible and actually increases the AIC and BIC statistics.

Across mixture states, the coefficient estimates on the pricing factors vary significantly. The market loading ranges from 3.6 to 1.7, while the HML factor is as low as -2.7. The pattern of mean log returns by state matches the Mills coefficient estimates. The state with the lowest predicted mean return also has the lowest Mills coefficient, suggesting that positive selection is relatively weaker conditional on that state. Next, the state-level predicted means closely match the "winners" and "failures" story posited above. One state has an alpha of 101%, a mean return of 185% and a prior state probability of 13%. In contrast, the "failure" state has a low mean of -202%, a mixing weight of 25% and an statistically insignificant alpha. The highest probability state produces a mean return of 22% and alpha of 30%, numbers very similar to those found for Cochrane (2005). Figure 7 displays the three state outcomes and the full pdf.

The mixture reveals distinct risk features of investment outcomes. The "winning" outcome has no systematic risk and its returns behave like growth stocks. The "failure" state has significant systematic risk and the expected returns look like small-cap stocks. An analysis of posterior probabilities reveals that many of the returns labeled "failures" were financed at the top of the Nasdaq bubble. These investments all failed when the market crashed. The introduction of the mixture also dramatically lowers the predicted mean returns, which falls to -11% in the three-state model from 4% in the single-state model. In contrast to a single-state estimator with normality, a mixture invites skewness and kurtosis. Compared to the single-state estimates I find a fatter left tail, skewness of -.2 and excess kurtosis of 4.4.

Computing arithmetic returns at the state-level is straightforward and follows from the log-normal assumption. The predicted means in Table 18 follow the same basic pattern as log returns, but we see that the log volatility dominates in the "winners" outcome. The predicted arithmetic return of 1300% stems from volatility and confirms the basic results in Cochrane (2005). The population arithmetic return is the linear combination of the state-means, however, there is no clear way to compute the total arithmetic volatility. Using the linear combination of state-level volatilities suggests a total volatility of 1100%. I believe these unreasonable numbers stem from the fact that I model log returns rather than arithmetic. It is doubtful that a mixture of normals for log-returns translates into a mixture of log-normals for arithmetic returns.

7.5 Comparing to single state

Figure 8 plots the full mixture pdf against the single-state estimate and kernel of observed returns. The most striking difference is in the left tail: the mixture model reveals a strong mode around -200% annualized log returns. The single-state model misses this feature of the distribution. This stark change likely stems from the conditioning selection correction and the size-dependent state probabilities. The selection problem presumes that there are too few bad returns relative to the good outcomes. If the instrument is working, the Mills term by state adjusts the means by state. Thus, conditional on the "failure" state, the Mills term effectively pushes down the mean and increases the relative likelihood of the state occurring. More importantly, the differences between the two models illustrates that volatility is not an informative measure of return outcomes.

The mixture and single-state model produce volatility estimates of 104% and 99%, however, the left and right tails on the former are fatter. The mixture has a skewness of -.2 (statistically significant) and an excess kurtosis of 4.4. Semi-variance is 50% larger than the population variance, indicating significant downside risk. By assumption, the single-state mixture with normality lacks any skewness or kurtosis. If the mixture represents the true returns distribution, the skewness and kurtosis demonstrate that simple volatility underestimates both the frequency of extreme outcomes and the magnitude of negative tail events. Simply, describing VC risk and return requires more than detailing the first two moments.

7.6 Stage and outcomes

Recall that I let the size of the entrepreneurial firm as proxied by the lagged capital stock impact the probability a return is in a given state. In unreported results, the coefficient on lagged capital stock is significant at the 5% level for each state. Thus, entrepreneurial firm size predicts state. The functional form of the state probabilities present ex-ante predictions of state membership given a firm's lagged capital stock. Figure 9 plots the predicted probability of each state implied by the state probability estimates. For firms with little or no capital stock — early-stage financings — the probability of a "winner" is about 30%. This likelihood is five times larger than for investments in firms that have raised over $10 million. In fact, the state probabilities say — ex-ante — that as a firm raises venture capital the returns distribution moves from a three-state mixture to a two-state. The stark difference in ex-ante probability of a winner based on firm size illustrates that mixture states map to (or correlate with) firm size. The state probabilities also allow for investigation of analysis of risk and return features as an entrepreneurial firms raises capital.

For a given entrepreneurial firm size, the state probabilities give the ex-ante likelihood of each mixture state. The characteristics of the states, however, are fixed. Figure 10 shows the implied changes in beta and alpha as firms raise capital. In contrast to Cochrane (2005), the mixture model predicts that the beta increases as firm develop. Entrepreneurial firms in the later stages of development are also nearer exits like acquisitions and IPOs. These outcomes are much more likely when the public markets are doing well. I find that this increase is not solely due to an decrease in return volatility from small to large firms, but rather from an increase in raw correlation between market and VC returns. Similarly, the arithmetic alpha falls as firms develop, suggesting much of the alpha in venture capital returns comes from early-stage investors. These investors take the most risk and also contribute more human capital to their investments than late-stage investors. Finally, Berk et. al. (2004) predict that risk premiums earned on investments in R&D projects (i.e. firms) should fall the project nears completion. Figure 11 shows a stark prediction from the mixing probabilities that confirms this hypothesis for VC returns. The ex-ante mean log return falls from 25% to -25% as one goes from minimal capital stock to over $50 million.

8 Comparing to public markets

Cochrane (2005) concludes that alphas and mean returns of VC investments resemble those of a micro-cap Nasdaq index. I now ask whether those same public indices are approximated by the mixture distribution that I found for VC returns. If not, then although basic means and average risk characteristics are similar, the skewness and multimodal features are unique to VC investment returns.

I construct the deciles of the Nasdaq index as detailed by Cochrane (2005). For comparison, I regress the log return to each public stock that matches the cutoff criteria on a constant, the excess log market return and the log FF factors. I strongly reject the multi-state model for all but the smallest decile. For this decile, the ICL statistics suggests a single state and the BIC maxes at 2. Given the overfit problems of BIC, this is a weak endorsement for a mixture model of small decile Nasdaq stocks. Moreover, Figure 12 shows the two states have significant overlap, while one has a volatility of 280%. These results imply that two states may be too many for this set of stocks. Figure 13 compares the single-state micro-cap distribution with that of VC returns and illustrates how downside risk and kurtosis are unique to latter. The two-state mixture for the smallest decile has an insignificant kurtosis and skewness. In contrast to venture capital returns, the microcap index volatility is informative about the relevant risks. Investors in venture capital find unique risk features unavailable in the standard benchmark.

The benchmark microcap index is not the only possible comparison asset. Harvey and Siddique (2000), Harvey and Siddique (1999) and Hwang and Satchell (1999) show that some public stock returns have negative skew and excess kurtosis. However, Kim and White (2004) show that estimators of both moments are sensitive to outliers. For example, the negative skew found in the daily S&P 500 index falls falls ten-fold when the 1987 market crash (one day) is excluded. The authors present robust estimators of skewness and kurtosis from the statistics literature. They estimate these alternative measures to the S&P 500 and find insignificant skew and diminished kurtosis even when outliers are included. These results suggest that the skewness and kurtosis of VC returns are likely not just unique when compared to the microcap index.

9 Robustness checks

9.1 Out of business returns

One important assumption for calculation of the out of business outcomes was the percent capital returned to original investors. I used 25% throughout the results discussed above as it was motivated by reality, how the rest of the literature dealt with these outcomes and the results of Cochrane (2005). Sensitivity analysis of this assumptions suggests that a wide range of values give the same basic results and implications, but at some point the log of a very small number generates an unrealistic negative left tail. Three basic changes occur as the cutoff value falls:

- β increases

- α falls
- Mean log returns fall dramatically

Table 19 shows the change in non-annualized log return for various levels of the cutoff. The mean falls dramatically: moving from a 15% to 10% cutoff, lowers the mean log return 42%. Moving the cutoff from 20% to 15% lowers the mean log return 50%. Surprisingly, these dramatic changes in mean returns has a small impact on single-state estimates. The beta estimate increases and the alpha estimate falls as the cutoff falls. However, the range of the latter is 1.8 - 2.1. Because the mixture model can fit unique tail returns processes, lowering the cutoff should change the left "failure" return process. I find that as the cutoff falls below 15%, the mean of the failure state falls, while the probability of the state decreases. The "winner" state also shifts: the mean increases and the state probability falls. As I lower the cutoff, the number of states moves towards four as the two tail states further separate. The ICL criterion suggests 4 states, while the AIC and BIC continue to imply 4 or 5. Overall, the mixture shape is sensitive to the smallest values of the cutoff criterion. The mean of the left tail state falls dramatically, while the risk characteristics like beta and the Fama and French loadings are constant. For example, the 4-state mixture with a 15% cutoff changes has a volatility of 113%, a mean log return of -40% and a alpha of 44%. Two states only differ in their intercept term. I conclude that cutoffs below 15% result in dramatically lower mean log returns and higher volatilities, but the current cutoff of 25% adequately represents the data and follows current practice.

9.2 State-dependent coefficients

Other than fixing one of the mixture state betas to equal zero, I allow all other coefficients on the pricing factors to differ across states. For example, I let $\beta_1 \neq \beta_2 \neq \beta_3$. Perhaps the Fama and French factor loadings are same across state, while the log market model intercept differs. Moreover, I argue that the "failure" and "winner" states have distinct systematic risks. I test whether the major coefficient estimates differ across state again using the BIC test statistics. I fix the states at three, so the information criterion does not suffer from the overfit problems. Table 20 shows that fixing coefficients for all market loadings leads to worse fit. The Mills term differs across states, confirming that selection effects depend on the return level. I conclude that the heterogeneity in the standard market and FF loadings across states is real.

10 Conclusion and further work

I present a new model of venture capital risk and return whose estimates demonstrate that normality is a poor approximation for investment outcomes. Mixture states that approximate the non-normality map to typical entrepreneurial firm outcomes. Negative skewness and excess kurtosis characterize a returns distribution where volatility underestimates left tail outcomes. I formulate a tractable, flexible estimator that corrects for sample selection and endogenous holding periods. The corrections require two instrument variables and have the expected impact on parameter estimates. The conclusions about non-normality and risk suggest that the research should move towards studying the optimal asset allocation decisions of the well-diversified, wealthy investors in venture capital. It is possible that the issues of weakly informative volatility may not inform their portfolio choices, however any preference for symmetric risk could alter allocation decisions.

The paper can be expanded in several ways. The market model setup may not fit VC returns. The basic derivation of all asset pricing models requires a "small" change in one's portfolio. As Cochrane explains, "a venture capitalist [...] must take all or nothing of a project" rather than a small, but diversified change in their portfolio.[9] The 3-stage model requires full observability of all exogenous variables, so I create market returns for financings without exits. Further work could test other assumptions for this variable. More pricing factors could enrich the model, but additional parameters may hinder estimation.

[9] Cochrane, John, Asset Pricing, Princeton University Press, 2005. page 37.

11 Tables and Figures

Table 1: Returns and durations

	Mean holding period	Median holding period	Observations
Bankruptcies	24.7 months	13.1 months	3636
0 < multiple < 2	15.5 months	13.8 months	7059
2<= multiple < 10	13.7 months	11 months	3214
multiple > 10	17.2 months	11.1 months	138

Includes all financing events with known dates. A holding period is defined as the time between two financing rounds or a financing round and an exit. Multiple is the gross increase in value of an entrepreneurial firm between two financing events. 5912 observations lack an exit date and are thus excluded from the table.

Table 2: Statistics on financings per entrepreneurial firm (with exits)

	Mean	Median
Full sample, 1987-2007		
Number of non-exit financings	2.5	2
Total financing events, including exits	3.1	3
Number of rounds for firms with IPO or acquisition	2.8	2
Number of rounds for bankrupt firms	2.2	2
Number of non-exit VC financings	2.9	3
Years to bankruptcy	5.6	5.6
Years to IPO	7	4.8
Years to acquisition	6.7	5.4

Counts the number of financing events per entrepreneurial firm with an exit by the end of the sample. Exits include IPOs, acquisitions
and bankruptcies. Non-VC rounds include corporate rounds and debt rounds.

Table 3: Exit type distribution for round-to-round returns, by financing

Exit type	**Percent of financings with exit type**
IPO	3.8%
Acquisition	9.7%
Bankruptcy	9.6%
Standard financing	61%
No exit	15.9%

Tabulation of exits for individual financing rounds. Includes all standard, non-debt venture capital rounds as of December 2007.

Table 4: Historical holding period across industry (in years)

Mean holding period	18.8 months
Median holding period	15.2 months
Number of financings	37,395
Number financings with exits	32,377

Holding period is the time between financing rounds. Averaged across all financing events with and without exits. Averages are across industry and financings. Industries are "Information Technology," "Healthcare," "Business and Consumer" and "Other." The sample includes all venture capital rounds as of December 31st, 2007.

Table 5: Exit type distribution for entrepreneurial firms

Exit type	**Percent of financings with exit type**
IPO	3.8%
ACQ	9.7%
Bankruptcy	9.6%
Standard financing	61%
No exit	15.9%

Tabulation of exits for individual financing rounds. Includes all standard, non-debt venture capital rounds as of December 2007.

Table 6: Round-to-round return characteristics, non-bankruptcies

	Arithmetic	Log	Obs.
Mean return (non-annualized)	108%	39%	10421
Median return (non-annualized)	50%	40%	10421
Mean annualized return	8548%	62%	10354
Median annualized return	141%	35%	10354
σ	252%	84%	10421
annualized σ	836267%	130%	
Mean return IPO/Acquisition	248%	93%	1171
Mean return private financing	89%	33%	9250
Mean return for IT firm	117%	91%	5753
Mean return for biotech firm	77%	35%	3157

Includes all venture capital rounds with holding periods greater than 1 month.
"Mean return IPO/Acquisition" is the return for a financing that exits via an IPO or acquisition.
"Mean return private financing" is the return with an "exit" via a standard financing round.

Table 7: Round-to-round return characteristics, all types

	Arithmetic	Log	Obs.
Mean return (non-annualized)	.72	-.067	12018
Median return (non-annualized)	.26	.23	12018
Mean annualized return	$2.1 * 10^{27}$	.326	12018
Median annualized return	.22	.109	12018
σ	232%	134%	12018
annualized σ	$2.4 * 10^{29}$	238%	12018
Mean return for IT firm	.81	-.056	6628
Mean return for bankruptcy	.55	-1.21	2820
Mean return for biotech firm	0	.07	3426

Includes all venture capital rounds with holding periods greater than 1 month, including out of business outcomes. "Mean return for bankruptcy" is the return for a financing that exits via bankruptcy."Mean return private financing" is the return with an "exit" via a standard financing round.

Table 8: Average historical capital raised by firm development and industry (millions)

Stage	Average	Standard deviation
Early	\$8.62	17.9
Mid-stage	\$20.8	24.1
Late	\$39	43.8
Information Technology	\$21.7	34
Healthcare	\$19.9	30
Consumer	\$19	30

Means are computed across all financing rounds that are of the particular stage and industry. "Early" stage includes the Series A financings and restart rounds that are 1st or Seed rounds. "Mid-stage" are all 2nd and 3rd rounds and those "Corporate" rounds that occur when a firm is 2 to 5 years old. "Late" is the remaining set of financing rounds not defined as early or mid-stage.

Thus, one firm will have multiple, distinct historical capital measures.

Table 9: Average historical historical holding period by industry (in years)

Industry	Average	Standard deviation
pre-1998		
IT	1.8	.12
Biotech	1.6	.12
Consumer	1.9	.14
Other	2	.1
post-1998		
IT	1.6	.09
Biotech	1.7	.08
Consumer	1.5	.1
Other	1.9	.1

Tabluates the time between financing rounds by industry and time period. A new financing round can be an exit or a standard financing.

Table 10: Ordered probit results

Independent variable	**Coefficient** (t-stat.)	Marginal effects: Bankruptcy (s.e.)	Marginal effects: no exit (s.e.)	Marginal effects: non-bankruptcy exit (s.e.)
past capital raised (millions)	.0128 (3.75)	-.0019 (.0005)	-.0018 (.0005)	.0037 (.001)
log market return	1.63 (31.98)	-.241 (.0071)	-.24 (.008)	.4813(.0148)
holding period	-.022 (-38.30)	.0032 (.0032)	.003 (.0001)	-.0064 (.0001)
stage (=1 if late stage)	-.1805 (-4.72)	-.0373 (.003)	-.03595 (.0029)	.0732(.006)
cutoff 1 (s.e.)	-1.57 (.026)			
cutoff 2 (s.e.)	-.939 (.025)			
Observations	37530			
Time dummies	N	N	N	N
Industry dummies	Y	Y	Y	Y
Interactions	N	N	N	N
pseudo-R^2	0.0527			

Dependent variable is 2 if financing had an non-bankruptcy exit as of the end of sample, 1 if no exit and 0 if bankruptcy exit. Data includes all standard venture capital rounds that impact dilution. Cutoffs are the α_1 and α_2 from the ordered probit specification.

Table 11: Selection of parametric hazard function

	Log-logistic	Weibull	Log normal	Exponential	Gompertz
AIC	**31475.388**	32772.11	32747.179	40201.231	37595.124
BIC	**31696.094**	32992.816	32967.885	40413.448	37815.83

Summary statistics from various duration models. Includes all censored and non-censoring holding periods for all standard venture capital rounds

Table 12: Log-logistic duration model estimates

Independent variable	**Coefficient** (t-statistic)	**Coefficient (t-statistic)**
log market return	2.31 (93)	2.46 (84)
mills	5.94 (103)	6.6 (89)
log historical industry duration	.5 (11.7)	.37 (8.5)
past raised capital (millions)	.003 (21.5)	.003 (21)
stage (=1 if late stage)	.32 (52)	.31 (45)
constant	-.36 (-10)	-.53 (-13)
Observations	37222	37222
Number of "failures"	32042	28564
Time dummies	N	N
Failure	New financing or bankruptcy	New financing
Industry dummies	Y	Y
Interactions	N	N

Standard errors are clustered at the firm level. Industry dummies are included in each regression but not reported. For the "New Financing" failure, the model treat bankruptcy exits as censored.

Table 13: Single-state estimates

	OLS	OLS-3	GLS	GLS-3	Full	Full-3
γ	-.0177(.0006)	-.01(.0008)	-.006(.0006)	.002(.0008)	-.007(.0002)	-.008(.0004)
β	2.4(.06)	1.6(.07)	2.6 (.06)	1.6(.08)	1.7(.04)	1.9(.06)
SMB		-.44(.09)		-.47(.1)		1.14(.07)
HML		-1.16(.08)		-1.5(.08)		-.12(.07)
α	824% (40)	814% (32)	60% (4.9)	70% (5.4)	40% (5)	40% (3.8)
Mills					1.3(.012)	1.4(.01)
σ	411%	406%	116%	114%	98%	99%

Standard errors in parentheses, clustered at the entrepreneurial firm. Estimates of equation (11) for $k = 1$ with a single-factor and three-factors (e.g. "OLS-3") using 12,007 returns to new financings and out of business outcomes. The market factor is the Wilshire 5000, while the SMB and HML are the standard FF factors "small minus big" and "high minus low". OLS does not include sample selection corrections and does not correct for volatility heterogeneity. GLS divides all variables by the square root of the observed holding periods. Full includes the Mills term from the ordered probit selection model, the fitted holding periods from 2SLS and divides all variables by the square root of the predicted holding periods. Alpha is computed using (), while the volatility estimate in columns 6 and 7 follow ().

Table 14: Single-state estimates and Cochrane (2005)

	Cochrane (2005)	Full
β	1.2	1.7 (.04)
α	35% (5.7)	40%(3.8)
$E[\ln R]$	15%	-12%
$\sigma(\ln R)$	91%	99%
$E[R]$	61%	38%
$\sigma(R)$	110%	103%

Notes: "Cochrane round-to-round" is from Table 4 of Cochrane (2005) and is estimated on a 1987 - 2000 round-to-round returns with the Nasdaq as the market factor (for proper comparison). Column 2 presents the full single-factor, single-state model estimates with the Wilshire 5000 as the market factor. $E[\ln R] = ga\hat{m}ma + \bar{\mu_{rf}} + \hat{\beta}(\bar{\mu}_m - \bar{\mu}_{rf})$ and $E[R]$ follows from the log-normal assumption.

Table 15: Return Outliers: Both large and small

log return	holding period (months)
1.2	.13
-5.4	5.9
4.6	3.02
-2.3	2.03
0	.1
2.8	2.5
1	.2
-4.2	1.2
2.8	2.5
2.7	6.5
-3.6	2.6
3.3	7.8
2.8	3.7
3.3	4
2.6	2.5

Summary of log returns for the observations (from the selected sample) with the largest residual in the 3-factor, single-state regression.

Table 16: Assessing the number of mixture states

States	AIC	BIC	ICL
2	-11838.54	-11735.03	-5529.421
3	-12256.77	-12094.11	-4279.314
4	-12678.06	-12456.25	546.809
5	-12845.59	-12564.64	3452.8016

Notes: Test statistics for various mixture formulations. Each mixture model estimated with 3 factors.

Table 17: Mixture Results: Three states

	Single-State	Mixture		
		Prob	*Estimate*	
α	39% (1.2)	14%	101%(11)	42%(3)
		61%	30%(6)	
		25%	7%(3)	
		Prob	*Estimate*	
β	1.86 (.06)	14%	0	1.9(.08)
		61%	1.7(.1)	
		25%	3.6(.2)	
		Prob	*Estimate*	
SMB	1.14 (.07)	14%	.14(.05)	1.2 (.1)
		61%	.9(.1)	
		25%	2.4(.3)	
		Prob	*Estimate*	
HML	-.12 (.06)	14%	-2.4(.2)	.05 (.1)
		61%	.28(.08)	
		25%	.9(.3)	
		Prob	*Estimate*	
E[ln R]	3%	14%	185%	-11%
		61%	22%	
		25%	-202%	
		Prob	*Estimate*	
σ(ln R)	99%	14%	128%	104%
		61%	90%	
		25%	96%	

Notes: Mixture estimates from EM algorithm of a 3-factor model with the Fama-French factors. Each sub-row is a state-coefficient/parameter. The final column represents the "average" or population estimate that is the weighted average of the state-level coefficient estimates. Standard errors for population parameters derived from the delta method.

Table 18: Implied arithmetic means

	"Winners"	"Break-even"	"Failures"	Population
$E[R]$	1340%	87%	-79%	219%
$\sigma(R)$	3000%	207%	20%	1300%

Note: Each state arithmetic return is calculated using $E[R] = \exp(E[\ln R] + .5\sigma^2) - 1$, while volatilities are $\sigma(R) = (E[R]+1)(\exp(\sigma^2) - 1)$

Table 19: Out of business cutoff and mean log returns

Cutoff	$E[\ln R]$	$\sigma(\ln R)$
.1	-17.2%	159%
.15	-12%	148%
.2	-8.4%	140%
.25	-5.7%	134%
.3	-3.3%	130%

Notes: Mean log returns are non-annualized percents. "Cutoff" is the assumption of the value of an investment at an out of business exit as a proportion of the original capital invested.

Table 20: BIC criterion for fixed coefficients across states

	Fixed	Unrestricted
β	-12053.29	-12099.38
SMB	-12082.65	-12099.38
HML	-11757.03	-12099.38
Mills	-12080.13	-12099.38

Notes: The table reports the BIC for each restriction. The second row fixes all betas to be identical across states. "Unrestricted" is the BIC from the 3-state mixture with beta = 0 for one state.

Figure 1: Fictional development of an entrepreneurial firm: BioMix

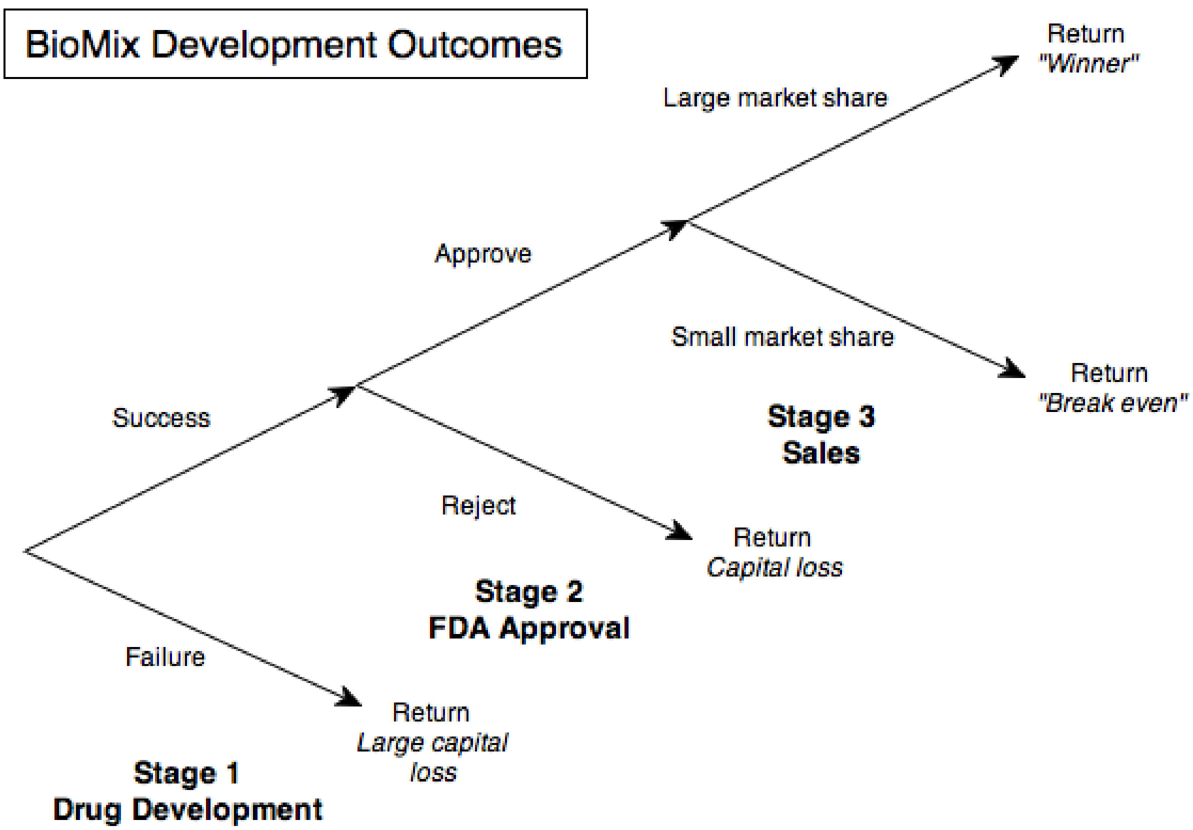


Figure 2: Survivor function estimate for investment holding periods

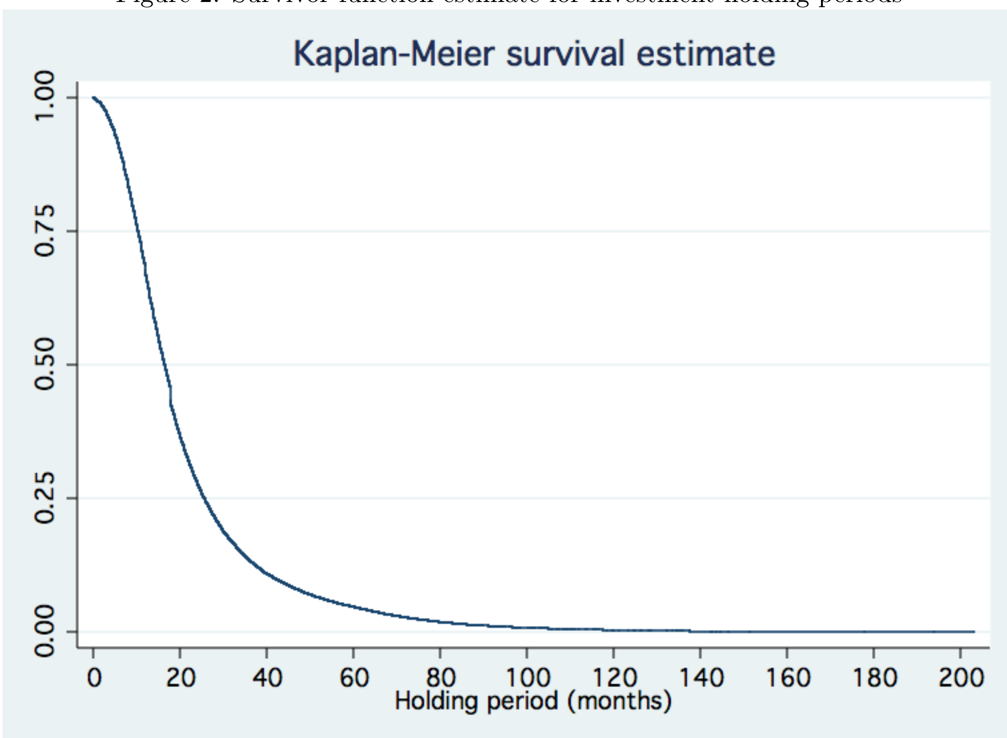


Figure 3: Estimated hazard

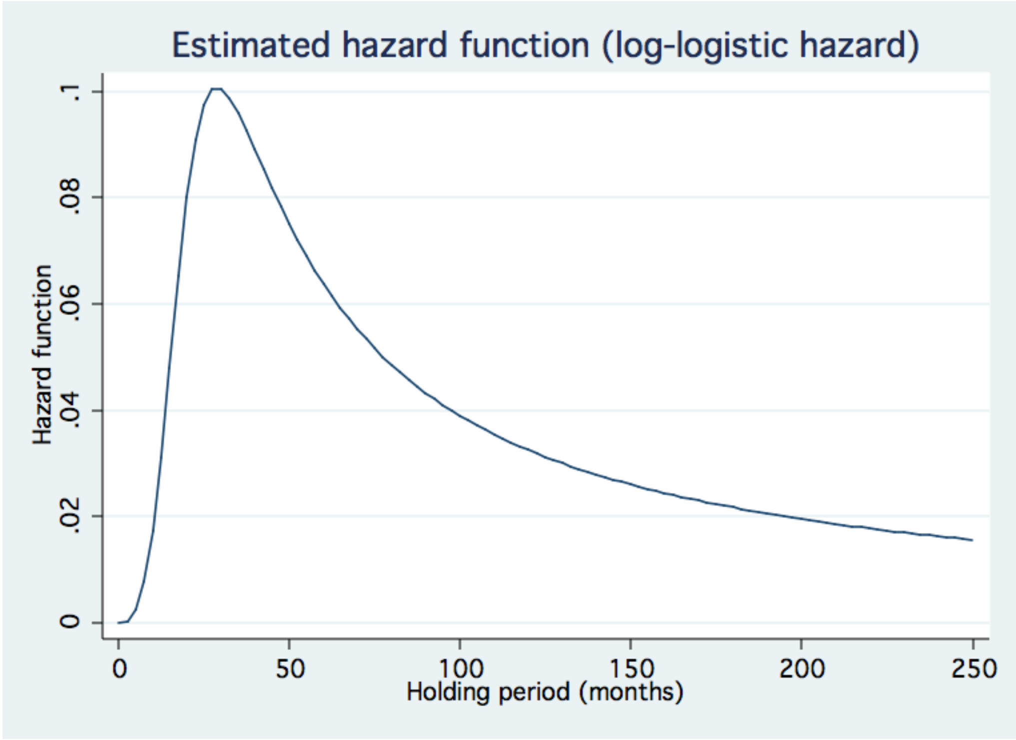


Figure 4: Estimated hazards by development stage

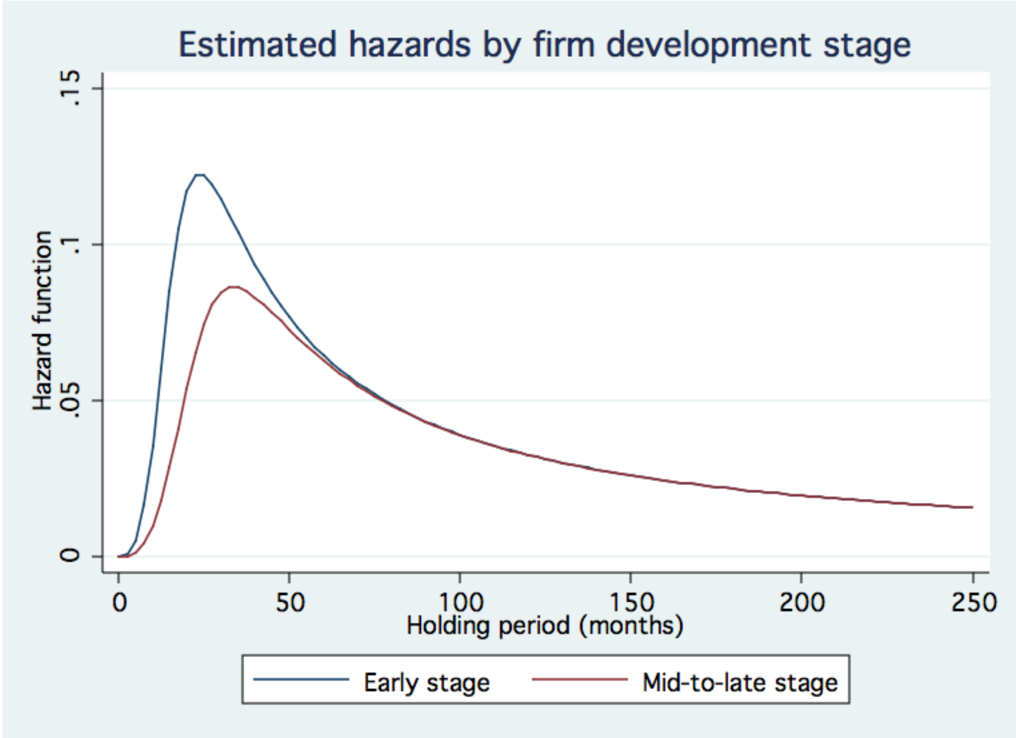


Figure 5: Estimated hazards for instrument (historical industry duration)

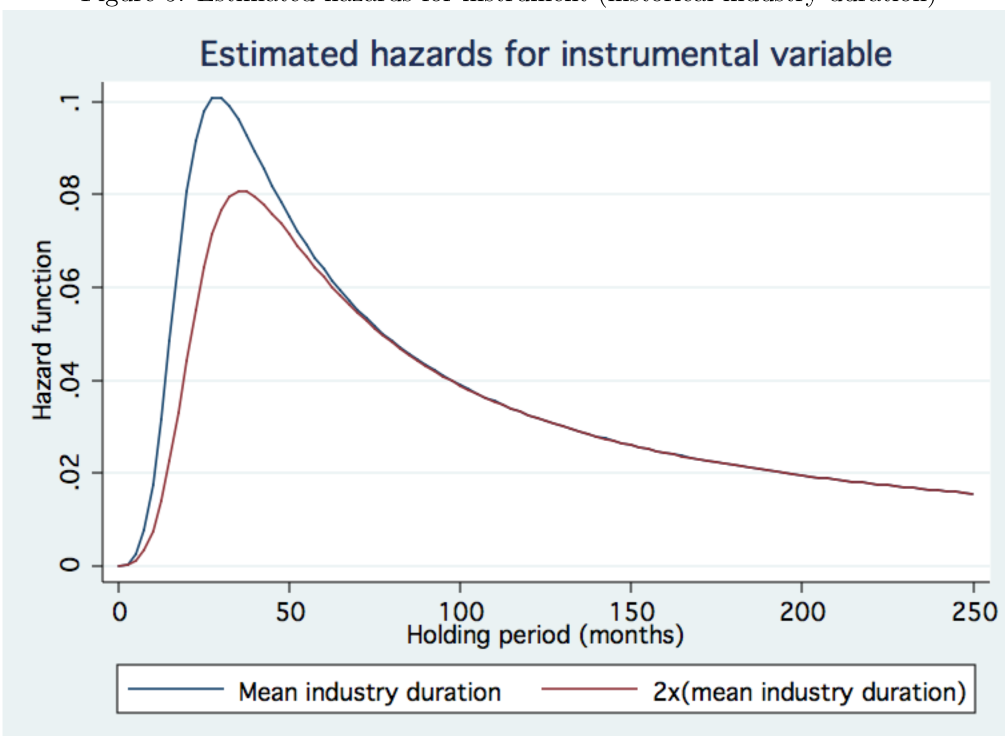


Figure 6: Kernel and full selection corrected pdf: Log returns

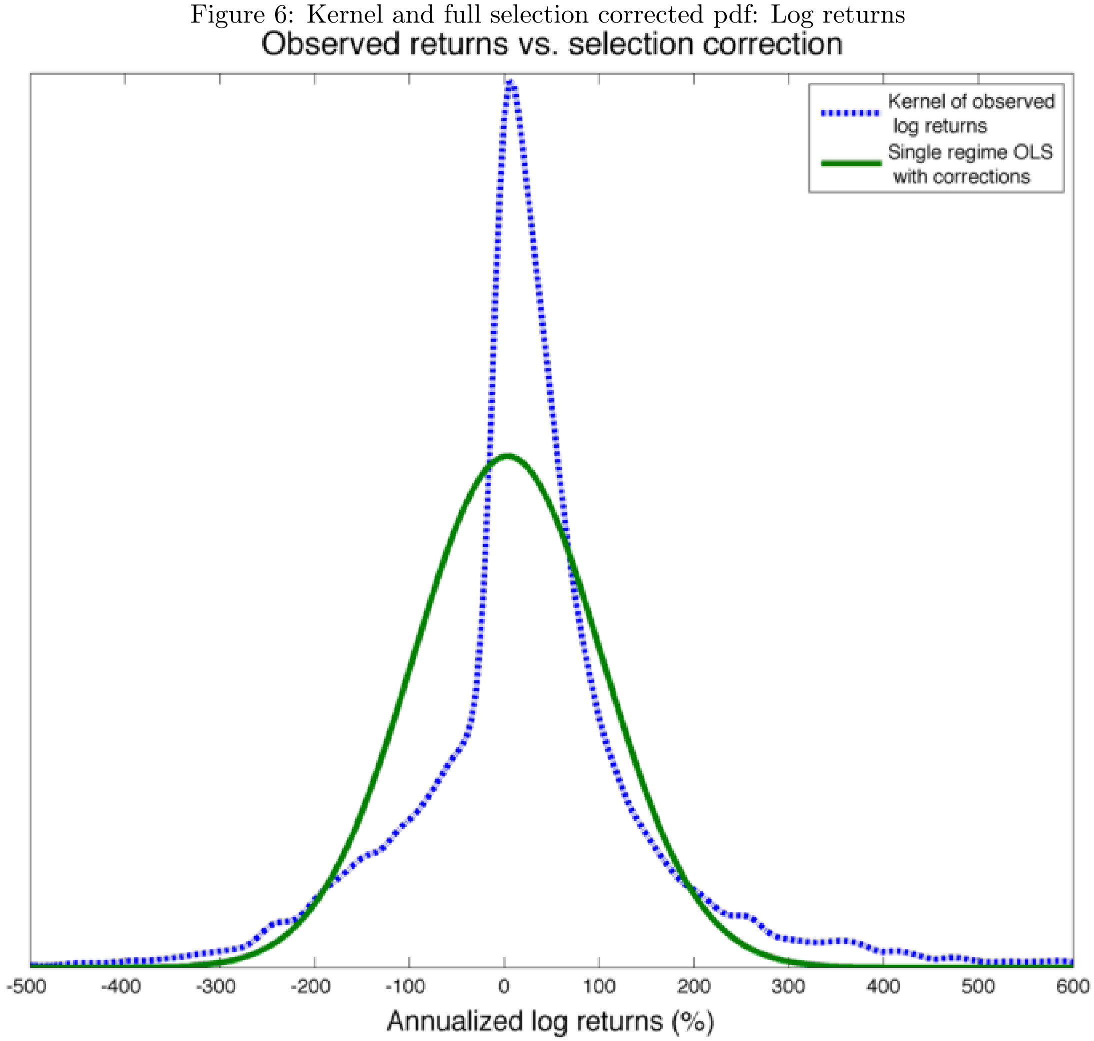


Figure 7: Full mixture and mixture states

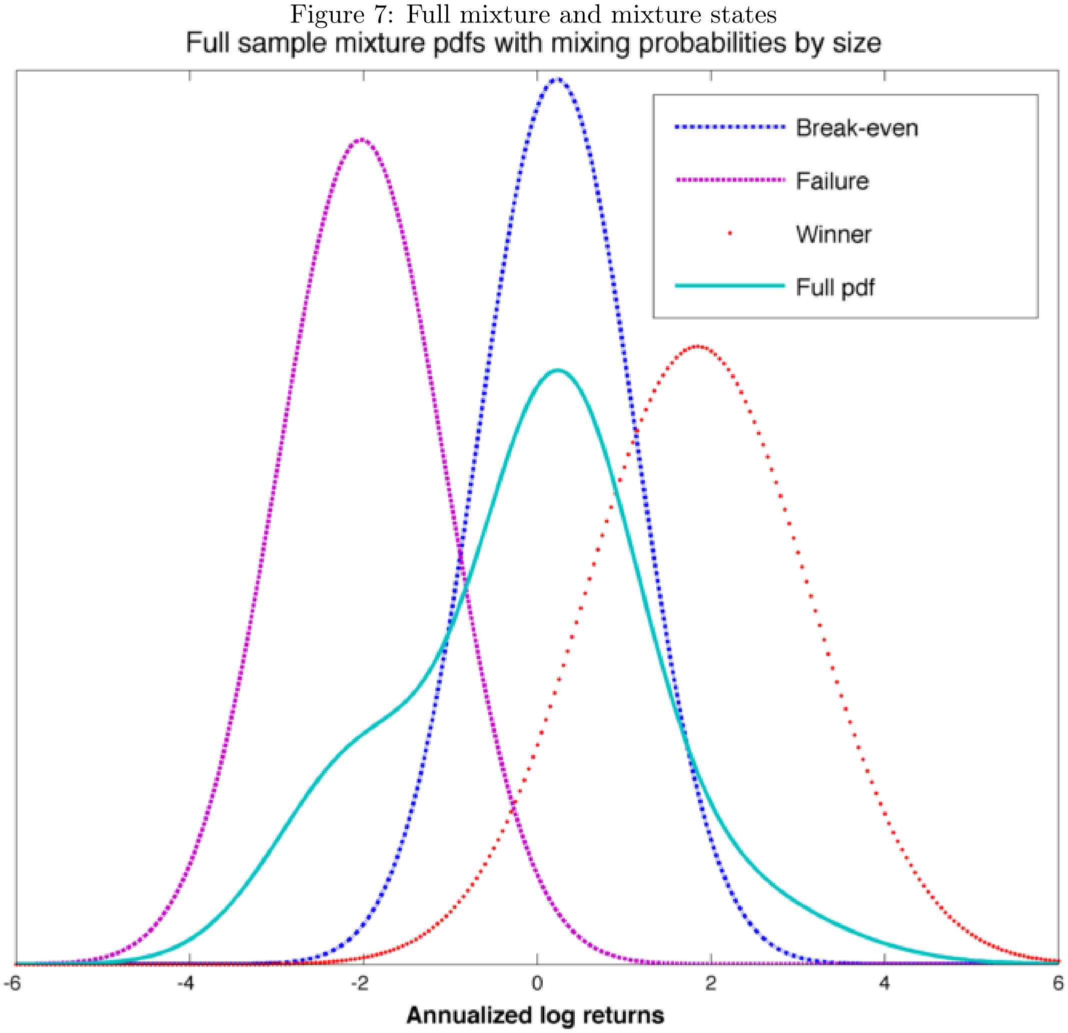


Figure 8: Progression from single-state to mixture

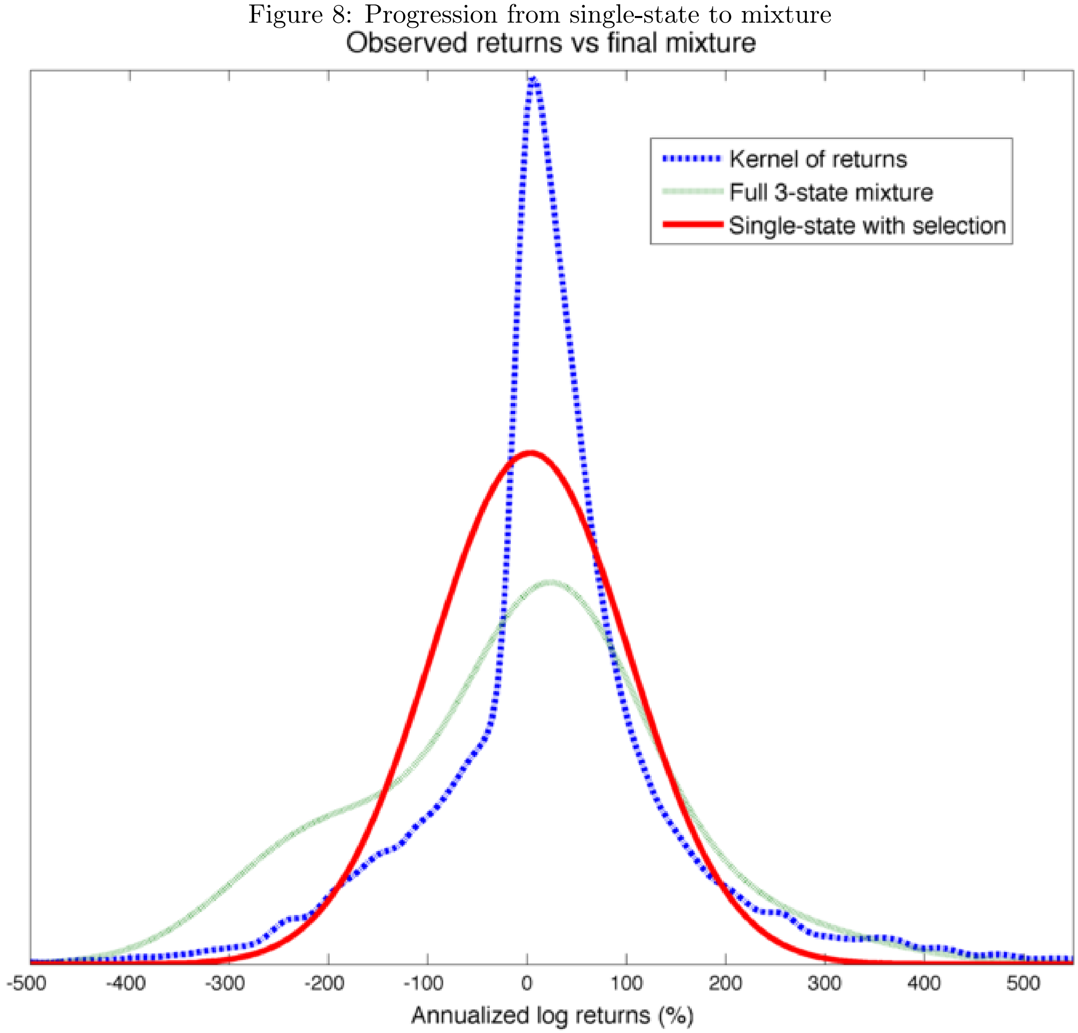


Figure 9: Ex-ante state predictions by size

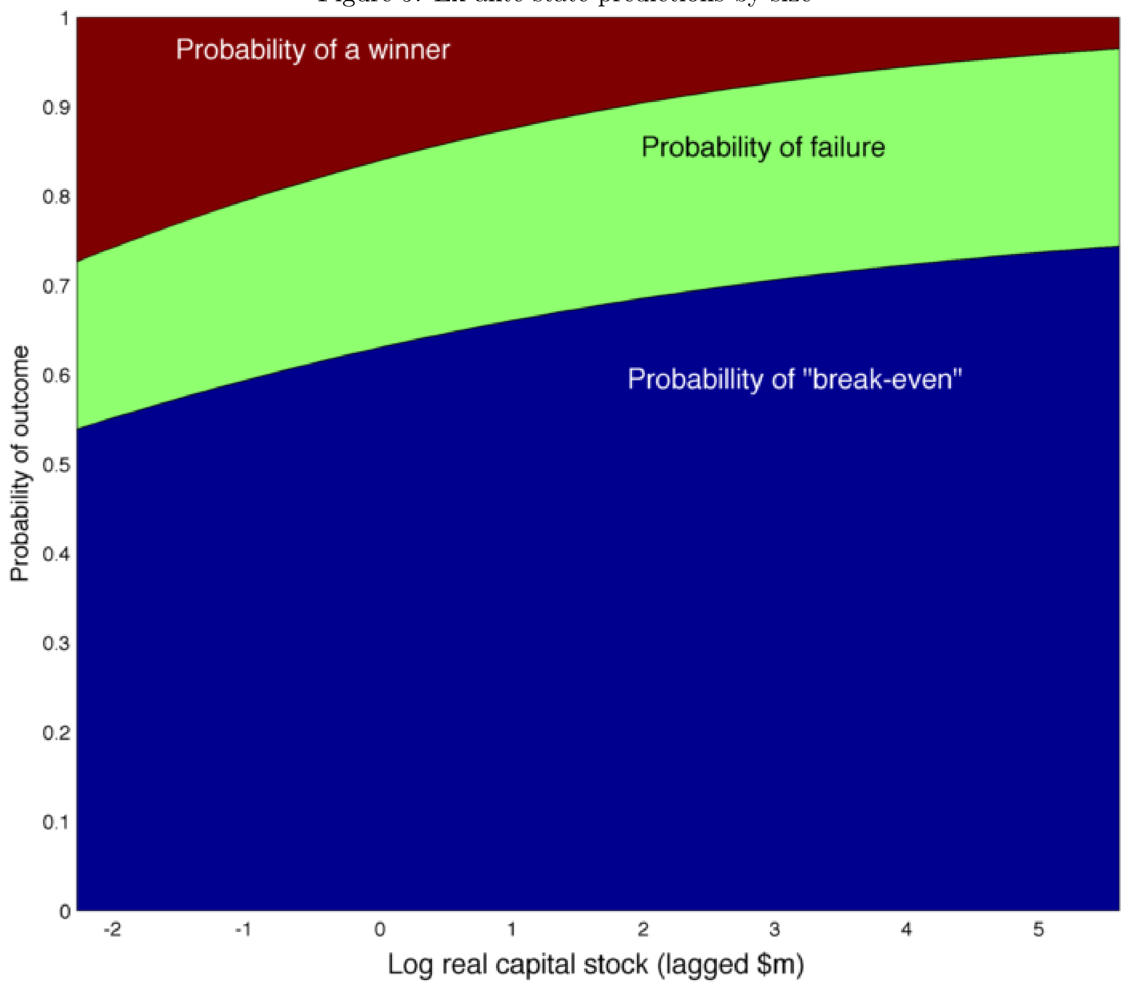


Figure 10: Alpha and beta by entrepreneurial firm size

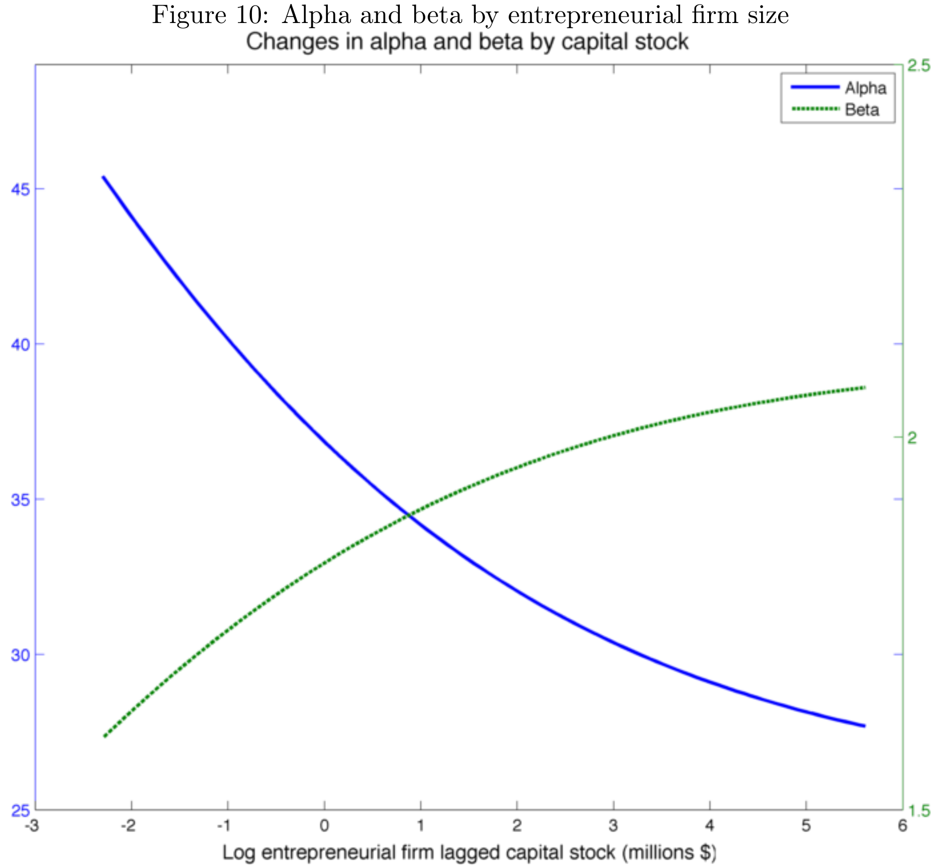
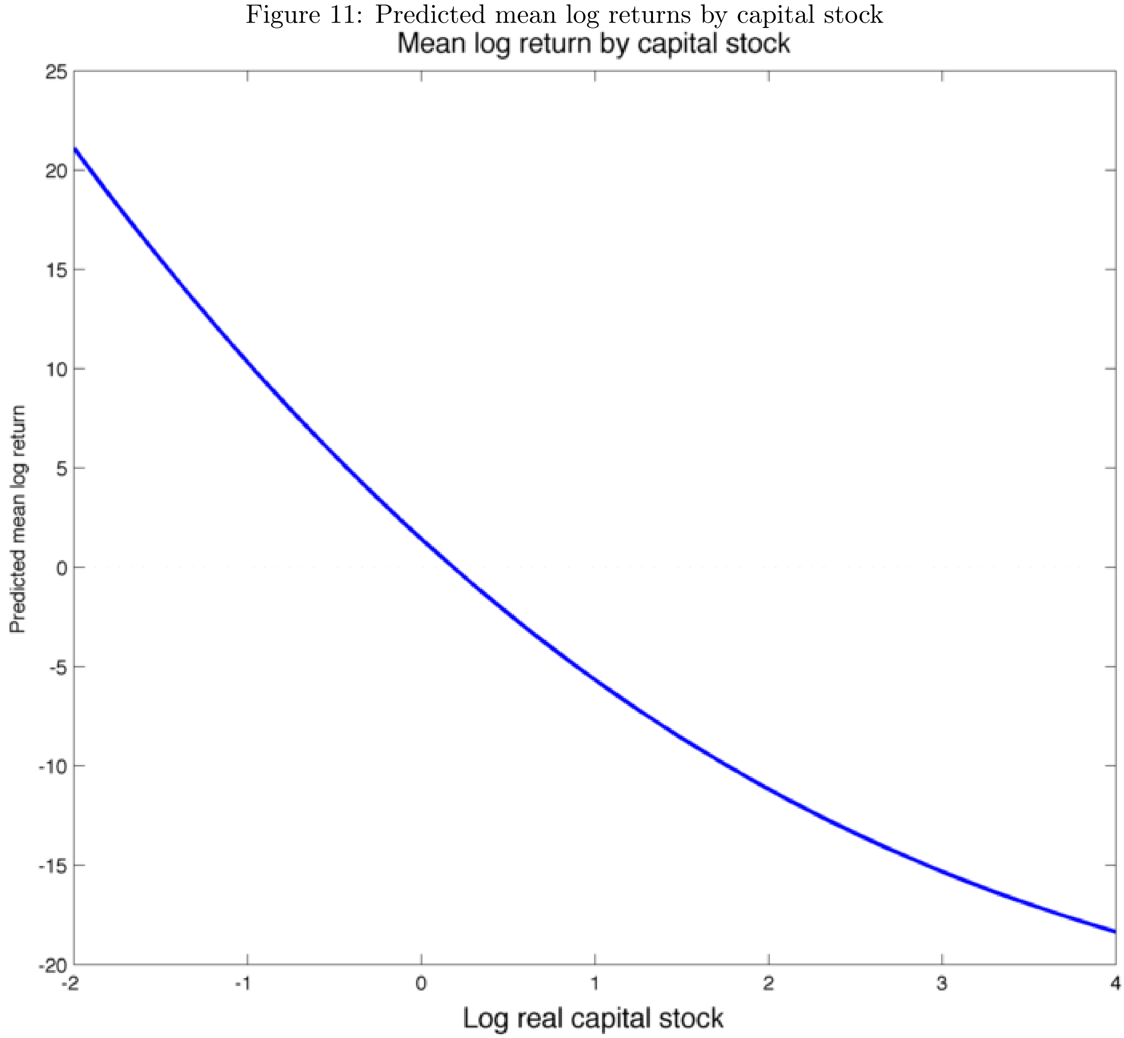


Figure 11: Predicted mean log returns by capital stock



Notes: predictions of mean log returns by the multinomial logit mixing probabilities. Capital stock is the lagged capital raised as of time t deflated by CPI.

Figure 12: Micro-cap Nasdaq mixture

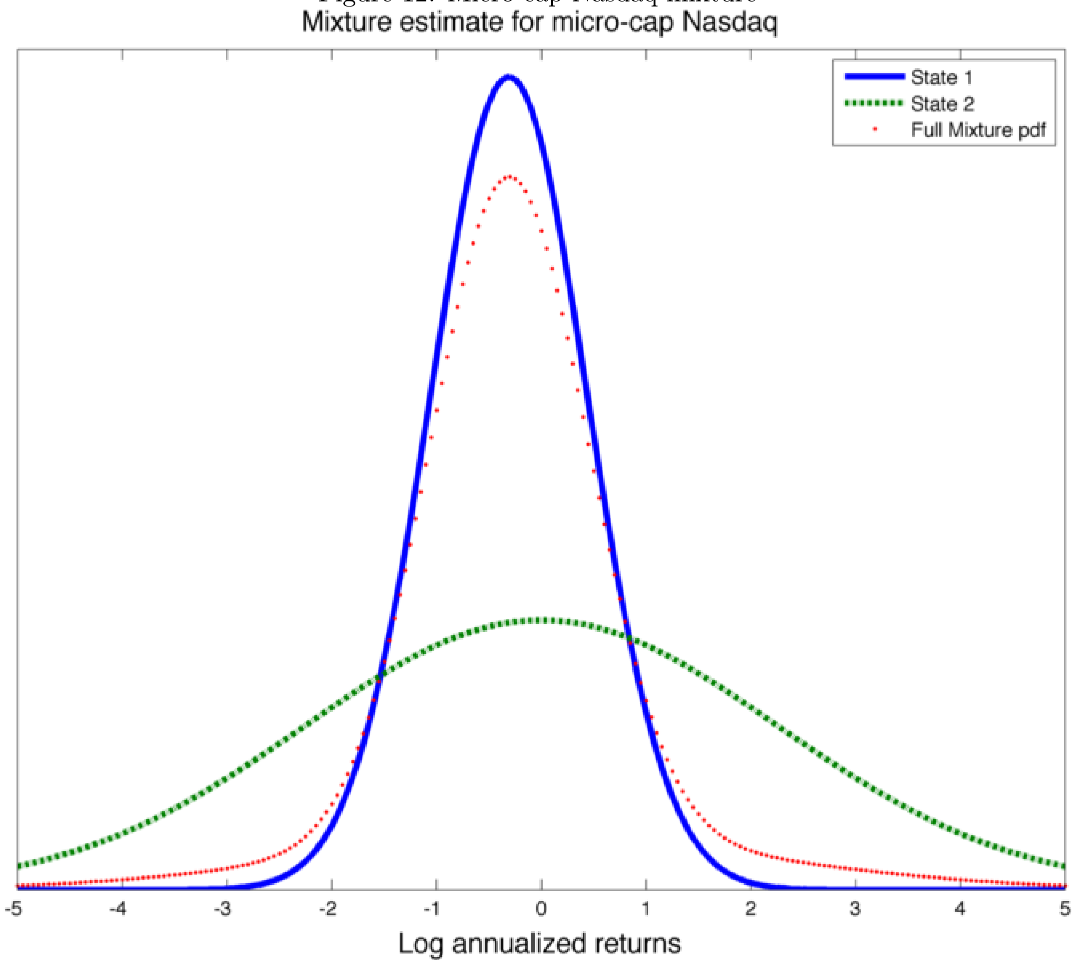


Figure 13: Micro-cap Nasdaq mixture vs VC returns mixture

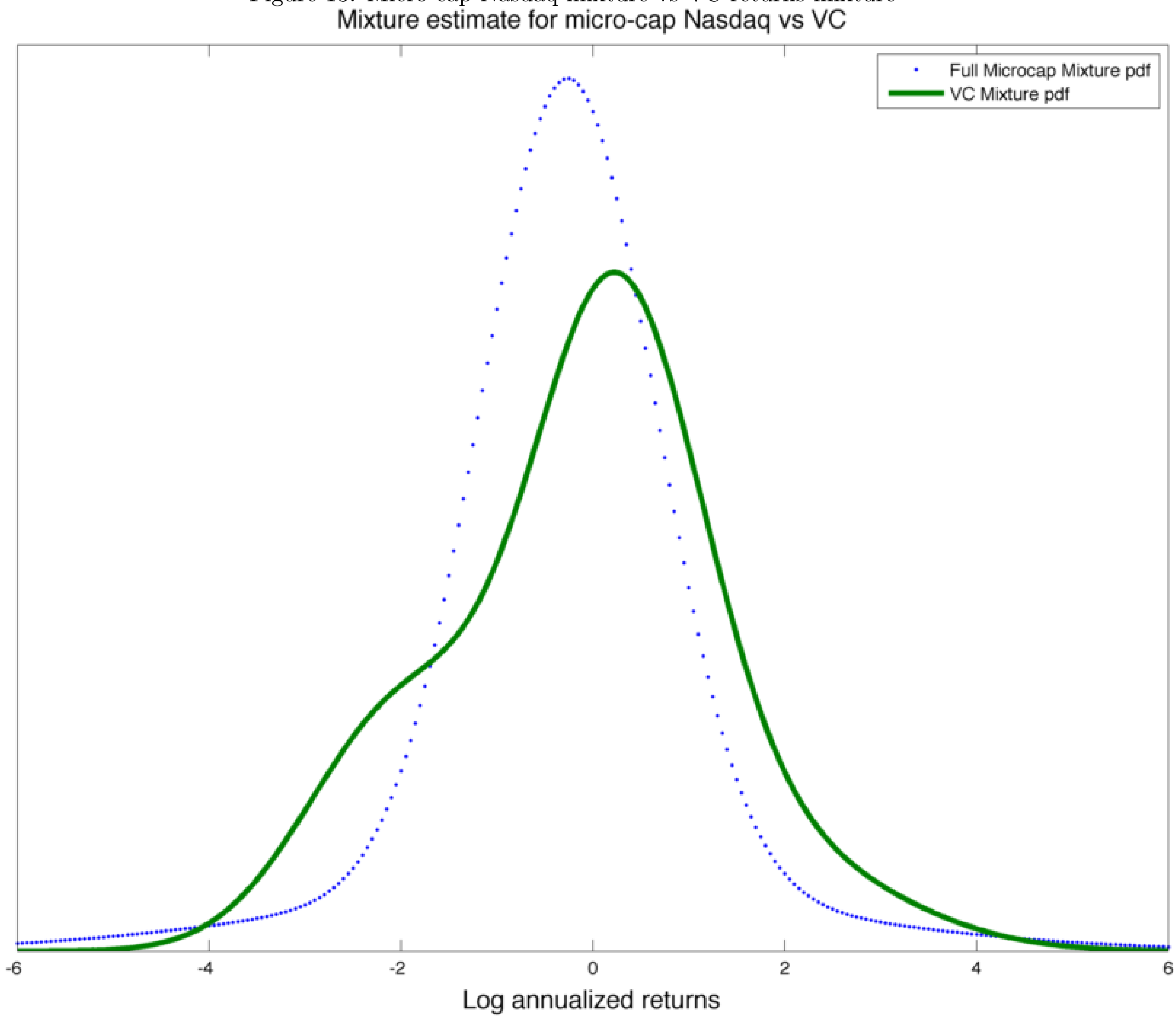


Figure 14: Bias of IV+GLS of β_1 with endogenous denominator

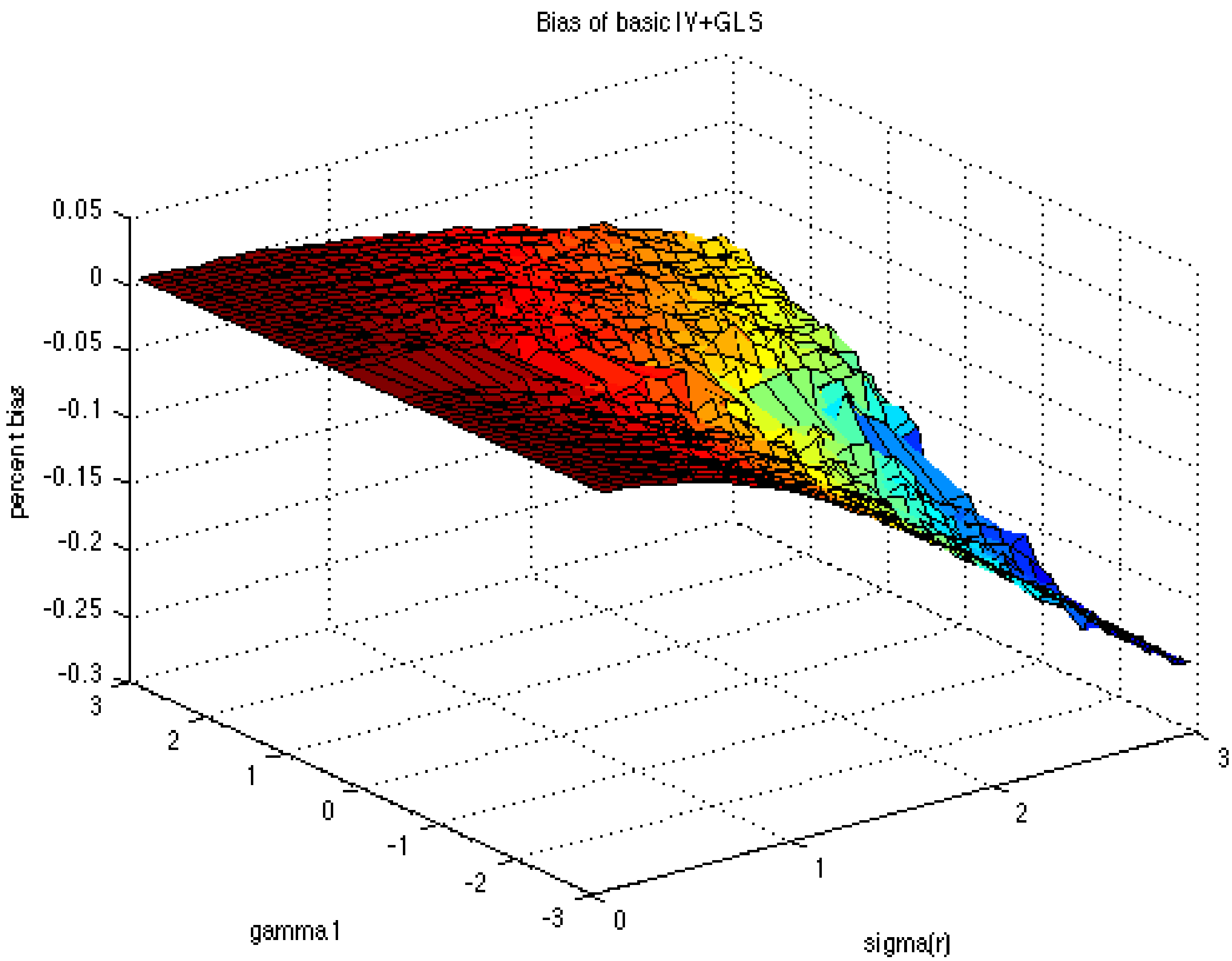
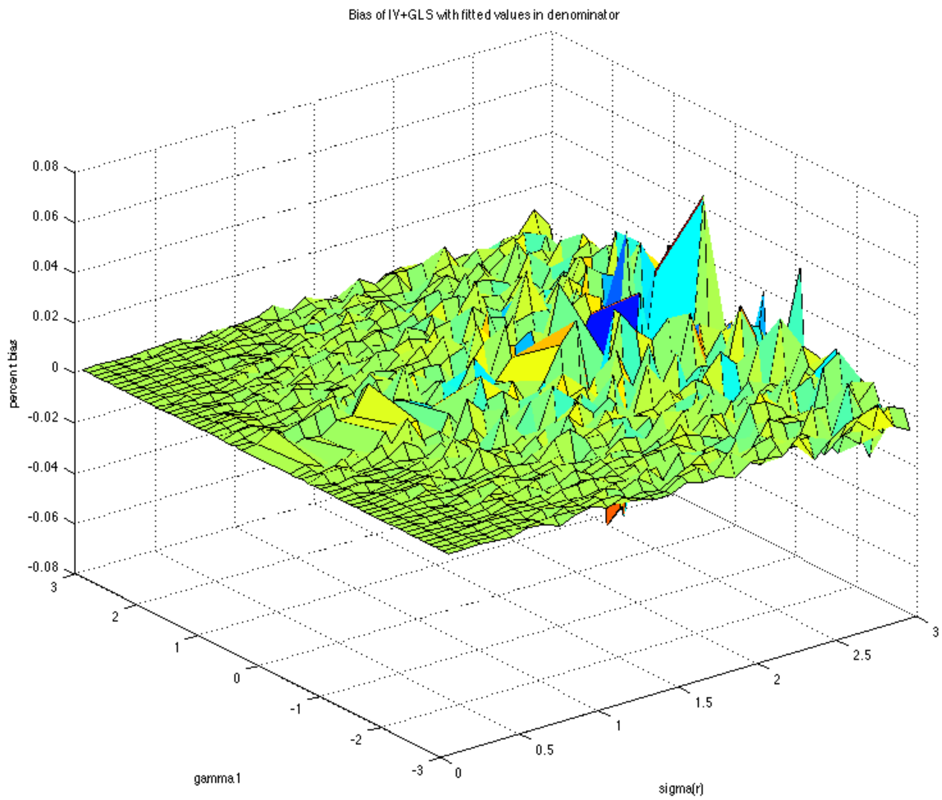


Figure 15: Bias of IV+GLS of β_1 with fitted denominator



References

Berk, Jonathan B., Richard C. Green, and Vasant Naik, "Optimal Investment, Growth Options, and Security Returns," *Journal of Finance*, October 1999, *54* (5), 1553–1607.

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

Biernacki, Christophe, Gilles Celeux, and Gerard Govaert, "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood," *Pattern Analysis and Machine Intelligence*, 2000, *22* (7), 719–725.

Cho, Jin Seo and Halbert White, "Testing for Regime Switching," *Econometrica*, November 2007, *75* (6), 1671–1720.

Cochrane, John, "The risk and return of venture capital," *Journal of Financial Economics*, 2005, *75* (1).

Cumming, Douglas J., "The Determinants of Venture Capital Portfolio Size: Empirica Evidence," *Journal of Business*, March 2006, *79* (3), 1083 – 1126.

Everitt, B.S and D.J. Hand, *Finite Mixture Distributions*, 1 ed., Chapman and Hall, 1981.

Fairlie, Rob and Alicia Robb, *Race and Entrepreneurial Success: Black-, Asian-, and White-Owned Businesses in the United States*, MIT Press, 2008.

Fruhwirth-Schnatter, Sylvia, *Finite Mixture and Markov Switching Models*, Springer, 2006.

Giacomini, R, A Gottschling, C Haefke, and H White, "Mixtures of t-distributions for finance and forecasting," *Journal of Econometrics*, May 2008, *144* (1), 175–192.

Gompers, Paul, "Grandstanding in the venture capital industry," *Journal of Financial Economics*, sep 1996, *42* (1).

Greene, William, *Econometric Analysis*, 5 ed., Prentice Hall, 2003.

Hall, Robert E. and Susan E. Woodward, "The Burden of the Nondiversifiable Risk of Entrepreneurship," NBER Working Papers 14219, National Bureau of Economic Research, Inc August 2008.

Hamilton, James, "Time Series Subject to Changes in Regime," *Journal of Econometrics*, 1990, *45*, 39–70.

____ , *Time Series Analysis*, Princeton University Press, 1994.

Hamilton, James D. and Michael T. Owyang, "The propagation of regional recessions," Working Papers 2009-013, Federal Reserve Bank of St. Louis 2009.

Harvey, Campbell R. and Akhtar Siddique, "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, December 1999, *34* (04), 465–487.

____ and ____ , "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 06 2000, *55* (3), 1263–1295.

Hwang, Min, John Quigley, and Susan Woodward, "An Index For Venture Capital, 1987-2003," *Contributions to Economic Analysis & Policy*, 2005, *4* (1), 1180–1180.

Hwang, Soosung and Stephen E Satchell, "Modelling Emerging Market Risk Premia Using Higher Moments," *International Journal of Finance & Economics*, October 1999, *4* (4), 271–96.

Jiang, Wenxin and Martin A. Tanner, "Hierarchical Mixtures-Of-Experts for Exponential Family Regression Models: Approximation and Maximum Likelihood Estimation," *The Annals of Statistics*, June 1999, *27* (3), 987–1011.

Jones, Charles and Matt Rhodes-Kropf, "The Price of Diversifiable Risk in Venture Capital and Private Equity," *Working Paper*, 2003.

Kaplan, Steven, Berk A. Sensoy, and Per Johan Stromberg, "How Well do Venture Capital Databases Reflect Actual Investments?," *Working Paper*, 2002.

Kaplan, Steven N. and Antoinette Schoar, "Private Equity Performance: Returns, Persistence, and Capital Flows," *Journal of Finance*, aug 2005, *60* (4).

Kim, Tae-Hwan and Halbert White, "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, March 2004, *1* (1), 56–73.

Korteweg, Arthur G. and Morten Sorensen, "Risk and Return Characteristics of Venture Capital-Backed Entrepreneurial Companies," Technical Report, AFA 2009 San Francisco Meetings Paper 2009.

Ljungqvist, Alexander and Matthew Richardson, "The Investment Behaviour of Private Equity Fund Managers," *RICAFE Working Paper*, oct 2003.

——, Yael Hochberg, and Yang Lu, "Networking as a Barrier to Entry and the Competitive Supply of Venture Capital," *Journal of Finance*, forthcoming.

Lunde, Asger, Allan Timmermann, and David Blake, "The hazards of mutual fund underperformance: A Cox regression analysis," *Journal of Empirical Finance*, April 1999, *6* (2), 121–152.

Markowitz, Harry, *Portfolio Selection, Efficiency Diversification of Investments*, Yale University Press, 1959.

McLachlan, Geoffrey and David Peel, *Finite Mixture Models*, Wiley-Interscience, 2000.

McLachlan, G.J., "On Bootstrapping the Likelihood Ration Test Statistic for the Number of Components in a Normal Mixture," *Applied Statistics*, 1987, *36* (3), 318–324.

Naik, Prasad A., Peide Shi, and Chih-Ling Tsai, "Extending the Akaike Information Criterion to Mixture Regression Models," *Journal of the American Statistical Association*, March 2007, *102*, 244–254.

Peng, Liang, "Building A Venture Capital Index," *Yale Working Paper*, aug 2001.

Perez-Quiros, Gabriel and Allan Timmermann, "Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities," *Journal of Econometrics*, July 2001, *103* (1-2), 259–306.

Phalippou, Ludovic and Maurizio Zollo, "Performance of Private Equity Funds: Another Puzzle?," *Working Paper*, June 2005.

—— and Oliver Gottschlag, "The Performance of Private Equity Funds," *Review of Financial Studies*, forthcoming.

Sorensen, Morten, "How Smart is Smart Money? A Two-Sided Matching Model of Venture Capital," *Journal of Finance*, 2007, *62*, 2725–2762.

Wooldridge, Jeffery M., *Econometric Analysis of Cross Section and Panel Data*, MIT, 2002.

A Estimating volatility

The size of the market model alpha hinges on the underlying volatility estimate, itself affected by corrections for sample selection and endogenous holding periods. For the standard Heckman selection correction, the coefficient on the Mills term is[10]

$$\beta_{Mills} = \rho\sigma$$

where σ is the true volatility of the returns and ρ is the sample selection correlation from assumption A4 in the 3-stage model. Observed volatility differs from σ because of sample selection. The conditional variance for each return j with the simple selection correction is

$$s^2 = \sigma^2(1 - \rho^2(\lambda(z\delta_3)(\lambda(z\delta_3) - z\delta_3))).$$

The σ falls out of the residual sum of squares from the first stage selection regression:

$$\sigma^2 = \frac{1}{N}s^2 + \hat{\hat{\delta}}\hat{\beta}^2_{Mills}$$

where

$$\hat{\hat{\delta}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\lambda}(z_i\hat{\delta}_3)(\hat{\lambda}(z_i\hat{\delta}_3) + z_i\hat{\delta}_3)).$$

As Greene (2003) shows, the $\delta_i \in (0, 1)$. If $\rho > 0$ (so that $\beta_{Mills} > 0$ and the data suffers from positive sample selection), then the conditional volatility estimate under-estimates the population volatility. The 2SLS step impacts the s^2 estimate by requiring the non-fitted holding periods in the error term estimation, slightly changing the estimate. I take this final estimate of s^2 and the new estimate of $\hat{\beta}_{Mills}$ to get the underlying volatility σ. The adjustments for selection and endogeneity result in a new mapping from the log market model parameters to the arithmetic alpha:

$$\alpha = \hat{\delta} + \frac{1}{2}\hat{\delta}(\hat{\delta} - 1) + \frac{1}{2}(\frac{1}{N}s^2_{2SLS} + \hat{\hat{\delta}}\hat{\beta}^2_{Mills}) \tag{20}$$

B GLS with endogeneous variables

I discuss below a generalized model of an endogenous variable that is also explicitly impacts the variance of the outcome equation.

B.1 Simple model

The following discussion borrows from Hamilton (1994). Suppose the outcome equation of interest is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

[10] I closely follow the discussion in Greene (2003).

and let x_1 be endogenous. There exists a variable x_3 that is exogenous in the outcome equation and correlated with x_1. Write

$$x_{1i} = \gamma_0 + \gamma_1 x_{3i} + \gamma_2 x_{2i} + u_i$$

Define $z' = (1, x_2, x_3)$. Then the estimate of $\gamma = [\gamma_0, \gamma_1, \gamma_2]$ is

$$\hat{\gamma} = \left[\sum_{i=1}^{N} z_i z_i'\right]^{-1} \left[\sum_{i=1}^{N} z_i x_{1i}\right]$$

From this estimate define

$$\hat{x} = [\hat{\gamma}' z, x_2]$$

The standard 2SLS estimator for $\beta = [\beta_1 \beta_2]$ is

$$\beta_{2SLS} = \left[\sum_{i=1}^{N} \hat{x}_i x_i'\right]^{-1} \left[\sum_{i=1}^{N} \hat{x}_i y_i\right]$$

B.2 Why GLS is biased

Suppose that we utilize the information that $\epsilon \sim N(0, \sigma^2 x_1)$ and divide all the regressors in the second stage by $\sqrt{x_{i1}}$. The derivation below shows that this leads to a biased estimate of β.

$$\tilde{\hat{x}}_i' = [\hat{\gamma} z_i / \sqrt{x_{1i}}, x_2 / \sqrt{x_{1i}}]$$

$$\tilde{x}_i = x_i / \sqrt{x_{1i}}$$

So the 2SLS-GLS estimator is

$$\beta_{2SLS-GLS} = \left[\sum_{i=1}^{N} \tilde{\hat{x}}_i \tilde{x}_i'\right]^{-1} \left[\sum_{i=1}^{N} \tilde{\hat{x}}_i \tilde{y}_i\right] = \left[\frac{1}{N}\sum_{i=1}^{N} \tilde{\hat{x}}_i \tilde{x}_i'\right]^{-1} \left[\frac{1}{N}\sum_{i=1}^{N} \tilde{\hat{x}}_i (\tilde{x}_i \beta + \tilde{\epsilon}_i)\right] = \beta + \left[\frac{1}{N}\sum \tilde{\hat{x}}_i \tilde{x}_i'\right]^{-1} \frac{1}{N}\sum_{i=1}^{N} \tilde{\hat{x}}_i \tilde{\epsilon}_i$$

This estimator is biased because

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \begin{bmatrix} \hat{\gamma} z_i / x_{1i} \\ x_{2i} / x_{1i} \end{bmatrix} \epsilon_i = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \begin{bmatrix} \hat{\gamma} z_i \\ x_{2i} \end{bmatrix} \frac{\epsilon_i}{x_{1i}} \neq 0$$

as $E[x_1 \epsilon] \neq 0$.

GLS with fitted values

I now repeat the exercise above, but replace the term $\sqrt{x_{1i}}$ with the fitted value from the first stage regression, $\sqrt{\hat{x}_{1i}}$. Note that

$$\hat{x}_{1i} = \hat{\gamma} z_i$$

We have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \hat{\gamma} z_i/\sqrt{\hat{x}_{1i}} \\ x_{2i}/\sqrt{\hat{x}_{1i}} \end{bmatrix} \tilde{\epsilon}_i = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \hat{\gamma} z_i \\ x_{2i} \end{bmatrix} \frac{\epsilon_i}{\hat{x}_{1i}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \hat{\gamma} z_i \\ x_{2i} \end{bmatrix} \frac{\epsilon_i}{\hat{\gamma} z_i} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} 1 \\ x_{2i}/\hat{\gamma} z_i \end{bmatrix} \epsilon_i = 0$$

where the last step follows from $E[\epsilon] = E[x_2\epsilon] = 0$. So the GLS correction should use the fitted values of the endogenous variable from the first stage regression. Next, I run some simple simulations showing the bias of each estimator.

B.3 Simulation results

The model presented above is simulated using N random draws S times. The simulation works as follows:

1. Generate N draws of $r \sim N(0, \sigma_r^2)$.
2. Generate N draws of $u = \sigma_u * r$
3. Generate N draws of $z = \mu_z + \sigma_z Z \qquad Z \sim N(0,1)$ and $x_2 = \mu_2 + \sigma_2 X_2, X_2 \sim N(0,1)$.
4. Compute $x_1 = \gamma_0 + \gamma_1 x_3 + \gamma_2 x_2 + u$ where γ is fixed.
5. Generate N draws of $\epsilon = \sigma_\epsilon x_1 E \qquad E \sim N(0,1)$
6. Compute $y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \epsilon$
7. Run OLS of y on a constant, x_1 and x_2. Save $\hat{\beta}$.
8. Run 2SLS. Compute $\hat{x}_1 = \hat{\gamma}_0 + \hat{\gamma}_1 x_3 + \hat{\gamma}_3 x_2$ from OLS of x_1 on a constant, x_3 and x_2.
9. Run OLS of y on a constant, $\hat{x}_1$ and x_2. Save $\hat{\beta}_{IV}$.
10. Compute $\tilde{y} = y/\sqrt{x_1}$, $\tilde{c} = 1/\sqrt{x}$, $\tilde{x}_1 = \hat{x}_1/\sqrt{x_1}$ and $\tilde{x}_2 = x_2/\sqrt{x_1}$.
11. Run OLS of $\tilde{y}$ on $\tilde{c}$, $\tilde{x}_1$ and $\tilde{x}_2$. Save $\hat{\beta}_{IVGLS}$.
12. Compute $\bar{y} = y/\sqrt{\hat{x}_1}$, $\bar{c} = 1/\sqrt{\hat{x}_1}$, $\bar{x}_1 = \hat{x}_1/\sqrt{\hat{x}_1}$ and $\bar{x}_2 = x_2/\sqrt{\hat{x}_1}$.
13. Run OLS of $\bar{y}$ on $\bar{c}$, $\bar{x}_1$ and $\bar{x_2}$. Save $\hat{\beta}_{IVGLS2}$.

Repeat these steps S times. We then have S estimates of $(\hat{\beta}, \hat{\beta}_{IV}, \hat{\beta}_{GLS}, \hat{\beta}_{GLSIV})$. We are ultimately interested in any bias of these estimators. First, we know that the standard OLS estimate of β will be biased. I am interested in the the bias of β_1 and β_2 for a range of γ_1 and σ_r. These two constants measure the relevancy of the instrument and the covariance of the two error terms. The fixed parameters are

$$\beta_1 = 1.5 \quad \beta_2 = 4 \quad \mu_z = 1 \quad \mu_{x_2} = 2.5 \quad \sigma_{x2} = 2.4 \quad \sigma_z = 1.1 \quad \sigma_\epsilon = .6\sigma \quad \gamma_2 = 2.2 \quad \sigma_u = .5$$

Finally, I set $N = S = 1000$. Figure 14 shows that the bias of the IV estimator increases (in absolute terms) as the variance of r increases, while for very large γ_1, the bias is minimal. Figure 15 shows the bias of the adjusted IV+GLS estimator that uses the fitted values of the endogenous variable in the denominator. The most significant absolute bias occurs when the instrument is weak: $\gamma_1 \approx 0$. Overall, using the fitted values for the GLS correction eliminates most of the pervasive bias.