14 Appendix

This appendix gives formulas and details of some of the empirical procedures. All programs and data are maintained on my website,

http://gsbwww.uchicago.edu/fac/john.cochrane/research/Papers/.

The main program is mlest3.m. This program runs the maximum likelihood estimation, sets up all the different cases and calls appropriate functions as needed. The other standalone programs are datachar2.m, which does much of the data characterization not requiring estimates, graphit.m, which makes graphs characterizing the estimates, such as graphs of predicted return distributions, and crspreg.m, which runs regressions and other analysis of the CRSP small Nasdaq data.

14.1 Data selection

The basic data loading and transformation for the maximum likelihood estimate are found in the top of doit3.m (called by mlest3.m), and the data loading and transformation for the data characterization is found in the top of datachar2.m. The venture one data are in returns.txt. The underlying venture one data without returns, but with more details including company names are in data.xls.

Starting with 16852 observations in the base case of the IPO/Acquisition sample (numbers vary for subsamples), I eliminated 99 observations with more than 100 or less than 0% inferred shareholder value, and 107 vc investments in the last period, the second quarter of 2000, since the model can't say anything until at least one period has passed. In 25 observations, the exit date comes before the VC round date, so I treat the exit date as missing.

For the maximum likelihood estimation, I treat 37 IPO, acquisition or new rounds with zero returns as out of business (0 blows up a lognormal), and I delete 4 observations with anomalously high returns (over 30000%) that were hand checked and found to be errors due to missing intermediate rounds. I similarly deleted 4 observations with a log annualized return greater than 15 ($100 \times (e^{15} - 1) = 3.269 \times 10^{8}\%$) on the strong suspicion of measurement error in the dates. All of these observations are included in the data characterization, however. I am left with 16,638 data points.

Datachar2.m makes the plots of the fates of venture capital investments. To estimate the fractions at, say, a 4 year age, I started with all rounds with a start date earlier than 4 years before the end of the sample – rounds that had a chance to achieve the 4 year age before the end of the sample. The fraction out of business is then the fraction of all these rounds that went out of business at an age less than or equal to 4 years. For example, a round that started in Jan 1991 and went out of business in June 1993 would be counted.

There is a selection bias with this measure: only firms that go out of business, are acquired or go public can have bad exit dates; firms that are still private cannot be removed from the sample for bad exit dates, and so are overrepresented. To account for this bias, I calculated the out of business fraction at, say, 4 years, as

$$\frac{\text{out}_4 \times \text{out ratio}}{\text{total}_4 \times \text{total ratio}}$$

where

$\text{out}_4 =$ number of rounds that started more than 4 years before the end of the sample, and went out of business in less than or equal to 4 years

$\text{total}_4 =$ number of rounds that started more than 4 years before the end of the sample.

out ratio = number of out-of-business rounds after selections/ number of out-of-business rounds before selections

total ratio = total number of rounds after selection / total number of rounds before selection .

I followed the same procedure to reweight the IPO or acquired category. I calculated the "still private" category as one less the last two categories. The weights do not make a difference noticeable to the eye in the Figures.

14.2 Logs to levels in the CAPM

This section derives the formulas for arithmetic α and β implied by the lognormal market model in Table 3-4 and 8-9. From the estimated market model in logs,

$$\ln\left(\frac{V_{t+1}^i}{V_t^i}\right) - \ln R_t^f = \gamma + \delta(\ln R_{t+1}^m - \ln R_t^f) + \varepsilon_{t+1}^i. \tag{8}$$

We want to find the implied CAPM in levels, i.e.

$$\frac{V_{t+1}^i}{V_t^i} - R_t^f = \alpha + \beta(R_{t+1}^m - R_t^f) + v_{t+1}^i.$$

Results. In the continuous time limit, $\beta = \delta$ and $\sigma(\varepsilon) = \sigma(v)$, but

$$\alpha_c = \gamma + \frac{1}{2}\delta\left(\delta - 1\right)\sigma_m^2 + \frac{1}{2}\sigma^2. \tag{9}$$

The major effect is a familiar $1/2\sigma^2$ term. In discrete time, we obtain instead

$$\begin{aligned}
\beta &= e^{\gamma+(\delta-1)\left(E(\ln R^m)-\ln R^f\right)+\frac{1}{2}\sigma^2+\frac{1}{2}\left(\delta^2-1\right)\sigma_m^2}\frac{\left(e^{\delta\sigma_m^2}-1\right)}{(e^{\sigma_m^2}-1)} \\
\alpha &= e^{\ln(R^f)}\left\{\left(e^{\gamma+\delta\left(E(\ln R^m)-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2}-1\right)-\beta\left(e^{\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma_m^2}-1\right)\right\}
\end{aligned}$$

I present these computations in the table; the continuous time α_c β_c are are quite similar.

Algebra for continuous-time limit. We start with the continuous time version of the log market model,

$$\begin{aligned}
d\ln V &= \left(r^f + \gamma\right)dt + \delta\left(d\ln P^m - r^f dt\right) + \sigma dz \\
d\ln P^m &= \mu_m dt + \sigma_m dz^m \\
E(dzdz^m) &= 0
\end{aligned}$$

Substituting,

$$
\begin{aligned}
d\ln V &= \left(r^f(1-\delta)+\gamma\right)dt+\delta\left(\mu_m dt+\sigma_m dz^m\right)+\sigma dz \\
&= \left(r^f+\gamma+\delta\left(\mu_m-r^f\right)\right)dt+\delta\sigma_m dz^m+\sigma dz
\end{aligned}
$$

Now, we can transform to levels. Using Ito's lemma,

$$
\begin{aligned}
\frac{dV}{V} &= \frac{d\left(e^{\ln V}\right)}{V}=d\ln V+\frac{1}{2}d\ln V^2 \\
&= \left[r^f+\gamma+\delta\left(\mu_m-r^f\right)\right]dt+\delta\sigma_m dz^m+\sigma dz+\frac{1}{2}\left(\delta^2\sigma_m^2+\sigma^2\right)dt
\end{aligned}
$$

$$
\frac{dP^m}{P^m}=\left(\mu_m+\frac{1}{2}\sigma_m^2\right)dt+\sigma_m dz^m
$$

Using the latter expression to substitute for dz^m,

$$
\frac{dV}{V}=\left[r^f+\gamma+\delta\left(\mu_m-r^f\right)+\frac{1}{2}\left(\delta^2\sigma_m^2+\sigma^2\right)\right]dt+\delta\frac{dP^m}{P^m}-\delta\left(\mu_m+\frac{1}{2}\sigma_m^2\right)dt+\sigma dz
$$

or, finally,

$$
\frac{dV}{V}-r^f dt=\left[\gamma+\frac{1}{2}\delta\left(\delta-1\right)\sigma_m^2+\frac{1}{2}\sigma^2\right]dt+\delta\left(\frac{dP^m}{P^m}-r^f dt\right)+\sigma dz
$$

We see that $\beta=\delta$, and the errors are the same, but we derive formula (9) relating the log intercept to the intercept in levels.

Algebra for the discrete-time lognormal calculation. From the model (8), we want to find the implied regression in levels,

$$R_{t+1} - R_t^f = \alpha + \beta(R_{t+1}^m - R_t^f) + \varepsilon_{t+1}^i$$

where

$$R_{t+1} \equiv \frac{V_{t+1}}{V_t}.$$

(It does not matter that the conditional expectation of V_{t+1}/V_t is a nonlinear function of R^m. The CAPM specifies the projection or linear regression.) We start with beta,

$$\beta = \frac{cov\left[R_{t+1}, R^m\right]}{var(R^m)}.$$

The denominator is

$$\begin{aligned}
var(R^m) &= E\left(R^{m2}\right) - \left[E(R^m)\right]^2 \\
&= E\left[e^{2\ln R^m}\right] - \left[E\left(e^{\ln R^m}\right)\right]^2 \\
&= e^{2\mu_m + 2\sigma_m^2} - \left(e^{\mu_m + \frac{1}{2}\sigma_m^2}\right)^2 \\
&= e^{2\mu_m + 2\sigma_m^2} - e^{2\mu_m + \sigma_m^2} \\
&= e^{2\mu_m + \sigma_m^2}\left(e^{\sigma_m^2} - 1\right).
\end{aligned}$$

The numerator is

$$cov\left(R,\ R^m\right) = E\left[RR^m\right] - E(R)E(R^m) = E\left[E(R|R^m)R^m\right] - E\left[E(R|R^m)\right]E(R^m)$$

Now,

$$E(R|R^m) = E\left\{e^{\gamma+\ln R^f+\delta\left[\ln R^m-\ln R^f\right]+\varepsilon^i}|R^m\right\} = e^{\gamma+\ln R_t^f+\delta\left[\ln R^m-\ln R^f\right]+\frac{1}{2}\sigma^2}.$$

Thus,

$$\begin{aligned}
cov\,(R,\ R^m) &= E\left[e^{\gamma+\ln R^f+\delta\left(\ln R^m-\ln R^f\right)+\frac{1}{2}\sigma^2+\ln R^m}\right] - E\left[e^{\gamma+\ln R^f+\delta\left(\ln R^m-\ln R^f\right)+\frac{1}{2}\sigma^2}\right]E\left[e^{\ln R^m}\right] \\
&= E\left[e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\ln R^m+\frac{1}{2}\sigma^2}\right] - E\left[e^{\gamma+(1-\delta)\ln R^f+\delta\ln R^m+\frac{1}{2}\sigma^2}\right]E\left[e^{\ln R^m}\right] \\
&= e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\mu_m+\frac{1}{2}(1+\delta)^2\sigma_m^2+\frac{1}{2}\sigma^2} - e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\mu_m+\frac{1}{2}\left(\delta^2+1\right)\sigma_m^2+\frac{1}{2}\sigma^2} \\
&= e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\mu_m+\frac{1}{2}\sigma^2}\left(e^{\frac{1}{2}(1+\delta)^2\sigma_m^2} - e^{\frac{1}{2}\left(\delta^2+1\right)\sigma_m^2}\right) \\
&= e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\mu_m+\frac{1}{2}\sigma^2}\left(e^{\frac{1}{2}(1+2\delta+\delta^2)\sigma_m^2} - e^{\frac{1}{2}\left(1+\delta^2\right)\sigma_m^2}\right) \\
&= e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\mu_m+\frac{1}{2}\sigma^2+\frac{1}{2}(1+\delta^2)\sigma_m^2}\left(e^{\delta\sigma_m^2}-1\right).
\end{aligned}$$

Putting it all together,

$$\begin{aligned}
\beta &= \frac{e^{\gamma+(1-\delta)\ln R^f+(1+\delta)\mu_m+\frac{1}{2}\sigma^2+\frac{1}{2}(1+\delta^2)\sigma_m^2}\left(e^{\delta\sigma_m^2}-1\right)}{e^{2\mu_m+\sigma_m^2}\left(e^{\sigma_m^2}-1\right)} \\
&= e^{\gamma+(\delta-1)\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma^2+\frac{1}{2}\left(\delta^2-1\right)\sigma_m^2}\frac{\left(e^{\delta\sigma_m^2}-1\right)}{\left(e^{\sigma_m^2}-1\right)}.
\end{aligned}$$

Continuing for α,

$$
\begin{aligned}
\alpha &= E(R) - R^f - \beta\left[E(R^m) - R^f\right] \\
&= E\left[E\left(R|R^m\right)\right] - R^f - \beta\left[E(R^m) - R^f\right] \\
&= e^{\gamma+\ln R^f+\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2} - e^{\ln R^f} - \beta\left[e^{\mu_m+\frac{1}{2}\sigma_m^2} - e^{\ln R^f}\right] \\
&= e^{\ln R^f}\left(e^{\gamma+\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2} - 1\right) - \beta\left(e^{\mu_m+\frac{1}{2}\sigma_m^2} - e^{\ln R_t^f}\right) \\
&= e^{\ln R^f}\left\{\left(e^{\gamma+\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2} - 1\right) - \beta\left(e^{\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma_m^2} - 1\right)\right\}
\end{aligned}
$$

Imposing $\alpha = 0$ by choice of γ

In the $\alpha = 0$ tests of Table 8 and 9, I choose γ to impose $\alpha = 0$ on the estimate. To impose $\alpha = 0$ we have to solve for γ with $\alpha = 0$,

$$
\begin{aligned}
e^{\gamma+\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2} - 1 &= \beta\left(e^{\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma_m^2} - 1\right) \\
e^{\gamma+\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2} - 1 &= e^{\gamma+(\delta-1)\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma^2+\frac{1}{2}\left(\delta^2-1\right)\sigma_m^2}\frac{\left(e^{\delta\sigma_m^2}-1\right)}{\left(e^{\sigma_m^2}-1\right)}\left(e^{\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma_m^2} - 1\right) \\
e^{-\gamma} &= e^{\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2} - \\
&\quad - \left(e^{(\delta-1)\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma^2+\frac{1}{2}\left(\delta^2-1\right)\sigma_m^2}\right)\frac{\left(e^{\delta\sigma_m^2}-1\right)}{\left(e^{\sigma_m^2}-1\right)}\left(e^{\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma_m^2} - 1\right)
\end{aligned}
$$

$$
e^{-\gamma} = e^{\delta\left(\mu_m-\ln R^f\right)+\frac{1}{2}\delta^2\sigma_m^2+\frac{1}{2}\sigma^2}\left[1 - \frac{\left(e^{-\left(\mu_m-\ln R^f\right)-\frac{1}{2}\sigma_m^2}\right)\left(e^{\delta\sigma_m^2}-1\right)\left(e^{\left(\mu_m-\ln R^f\right)+\frac{1}{2}\sigma_m^2}-1\right)}{\left(e^{\sigma_m^2}-1\right)}\right]
$$

$$
\begin{aligned}
e^{-\gamma} &= e^{\delta\left(\mu_m - \ln R^f\right) + \frac{1}{2}\delta^2\sigma_m^2 + \frac{1}{2}\sigma^2}\left[1 + \frac{\left(e^{\delta\sigma_m^2} - 1\right)\left(e^{-\left(\mu_m - \ln R^f\right) - \frac{1}{2}\sigma_m^2} - 1\right)}{\left(e^{\sigma_m^2} - 1\right)}\right] \\
\gamma &= -\ln\left\{e^{\delta\left(\mu_m - \ln R^f\right) + \frac{1}{2}\delta^2\sigma_m^2 + \frac{1}{2}\sigma^2}\left[1 + \frac{\left(e^{\delta\sigma_m^2} - 1\right)\left(e^{-\left(\mu_m - \ln R^f\right) - \frac{1}{2}\sigma_m^2} - 1\right)}{\left(e^{\sigma_m^2} - 1\right)}\right]\right\}
\end{aligned}
$$

$$
\gamma = -\left(\delta\left(\mu_m - \ln R^f\right) + \frac{1}{2}\delta^2\sigma_m^2 + \frac{1}{2}\sigma^2\right) - \ln\left\{\left[1 + \frac{\left(e^{\delta\sigma_m^2} - 1\right)\left(e^{-\left(\mu_m - \ln R^f\right) - \frac{1}{2}\sigma_m^2} - 1\right)}{\left(e^{\sigma_m^2} - 1\right)}\right]\right\}.
$$