Online Appendix: Details of Estimation Procedure

For each company, the econometric model is given by

$$v(t) = v(t-1) + r + \delta + \beta\left(r_m(t) - r\right) + \varepsilon(t), \tag{1}$$

$$w(t) = Z'(t)\gamma_0 + v(t)\gamma_v + \eta(t), \tag{2}$$

with *i.i.d.* distributions $\varepsilon(t) \sim N(0, \sigma_v^2)$ and $\eta(t) \sim N(0,1)$. For inference, we are interested in the posterior distribution of the parameters $\theta = (\delta, \beta, \sigma^2, \gamma)$. For numerical tractability, however, it is convenient to expand this distribution with the latent valuations and selection variables. The resulting augmented distribution $F[v(1), v(2), ..., v(T), w(1), w(2), ..., w(T), \theta \mid data, prior]$ is simulated using a Gibbs sampler (Gelfand and Smith (1990)) with three blocks, containing the valuations, the selection variables, and the parameters, respectively, as described below. By the Clifford-Hammersley theorem , sampling from the three blocks in turn generates a Markov chain whose stationary distribution if the augmented posterior *F*. For simplicity, we suppress the dependence of the risk-free rate *r* on *t*.

A. Draw Latent Valuation Variables Using FFBS

The latent valuation variables for the interim period between two observed valuations are sampled conditional on the parameters, the selection variables, and the realized market (and other factor) returns. In other words, we need to draw from the distribution of the valuations *v*(1)...*v*(*T*) conditional on *w*(1)...*w*(*T*), the parameters, and the observed data. We use the Forward Filtering Backwards Sampling (FFBS) procedure (Carter and Kohn (1994) and Fruhwirth-Schnatter (1994)), which provides an efficient way to sample a path of state variables defined by a linear state space model. Since the error terms are assumed *i.i.d.* across firms, we can sample $v(t)$ separately for each firm.

Interpreting the econometric model as a linear state space model, $v(t)$ is the state variable, and the outcome equation (1) is the transition rule. Conditional on the parameters, $r+\delta+\beta(r_m(t)-r)$ is an "observed" control acting on the state, and conditional on $w(t)$, the selection equation (2) is a noisy observation equation for the state.

This setup allows us to calculate the filtered distribution of $v(1)...v(T)$, using the Kalman filter. The Kalman filter produces the distribution of $v(t)$ conditional on $w(1)...w(t)$, for any time *t*. However, $v(t)$ needs to be sampled conditional on the entire time series $w(1)...w(T)$. This is achieved by a backward smoother, which effectively runs a Kalman filter backwards, starting at time *T*. The conditional distribution of the state vector of latent valuations is given by the identity (Lemma 2.1 in Carter and Kohn (1994))

$$p(v(1)\ldots v(T)\,|\,w^T) = p\{v(T)\,|\,w^T\}\prod_{t=1}^{T-1} p\{v(t)\,|\,w^t, v(t+1)\}, \qquad (3)$$

where $w^t=(w(1),\ldots,w(t))$ contains the selection variables up to time *t*. We will now describe the forward filter and backward sampling steps in detail.

Define $m(t \mid j) = E\{v(t) \mid w^j\}$ and $s(t \mid j) = \mathrm{var}\{v(t) \mid w^j\}$ as the mean and variance of $v(t)$ conditional on the selection variables up to time *j*. Note that all conditional distributions are Normal and hence fully characterized by their means and variances (see Kalman (1960) and Anderson and Moore (1979)).

For the forward filtering step, for $t = 1, \ldots, T$, we calculate $m(t \mid t)$ and $s(t \mid t)$ by iterating on the forward filter, through a forecasting and an updating part. The forecasting part involves the two equations

$$m(t+1 \mid t) = m(t \mid t) + \left(r + \delta + \beta(r_m - r)\right), \tag{4}$$

and

$$s(t+1 \mid t) = s(t \mid t) + \sigma^2 . \tag{5}$$

For the updating part, as long as $v(t)$ remains unobserved, we update

$$m(t \mid t) = m(t \mid t-1) + K \cdot [w(t) - Z'(t)\gamma_0 - m(t \mid t-1)\gamma_v], \tag{6}$$

where the Kalman gain *K* is given by

$$K = \frac{\gamma_v s(t \mid t-1)}{1 + \gamma_v^2 s(t \mid t-1)} . \tag{7}$$

When *K* is large, more weight is placed on the information from the selection equation. This happens when either γ_v or $s(t \mid t-1)$ is large, i.e. when either the selection equation is more informative about the valuations or when the valuations are more uncertain. Further,

$$s(t\,|\,t) = s(t\,|\,t-1)\cdot(1-\gamma_v K)\,. \tag{8}$$

To estimate the model without correcting for selection, we force $\gamma_v = 0$. Then $m(t\,|\,t) = m(t\,|\,t-1)$ and $s(t\,|\,t) = s(t\,|\,t-1)$, and no information is used in periods where $v(t)$ is unobserved. In periods where $v(t)$ is observed, $m(t\,|\,t) = v_{OBS}(t)$ and $s(t\,|\,t) = 0$.

For the backward sampling part, $v(T)$ is first simulated from the Normal distribution with mean $m(T\,|\,T)$ and variance $s(T\,|\,T)$, as given by the Kalman filter. For $t = T-1,\ldots,1$, we simulate $v(t)$ from the conditional distribution $p\{v(t)\,|\,w^t, v(t+1)\}$. This distribution can be derived from a filtering problem where the draw of $v(t+1)$ provides an additional observation of $v(t)$. The distribution is

$$p\{v(t)\,|\,w^t, v(t+1)\} \sim N(r,q)\,. \tag{9}$$

where

$$r = m(t\,|\,t) + G\cdot[v(t+1) - m(t+1\,|\,t)]\,. \tag{10}$$

$$q = s(t\,|\,t)\cdot(1-G)\,. \tag{11}$$

with

$$G = \frac{s(t\,|\,t)}{s(t\,|\,t)+\sigma^2}\,. \tag{12}$$

From equation (12), G can be interpreted as a Kalman gain similar to K in equation (7). As such, the backwards sampler weighs the information from the filtered distribution $v(t) \mid w^t$ and the information in $v(t+1) \mid w^T$ to obtain a draw of $v(t) \mid w^T$, with the weight depending on the relative variance of the filtered estimate $s(t \mid t)$ and the variance of $v(t+1)$. If the filtered estimate $m(t \mid t)$ is very precise relative to the variance of the valuation change from one period to the next, then G is close to zero, and most weight is put on the distribution of $v(t)$ from the Kalman filter. The more imprecise the Kalman filter distribution relative to how much the valuation can possibly change (as captured by sigma), the more weight is put on the "observed" $v(t+1)$.

B. Draw Selection Variables from Truncated Normal Distributions

The selection variables are sampled conditional on the valuations, parameters, and whether the valuation is observed or not. Simulating this block is similar to simulating the (augmented) posterior distribution of a Probit model (Albert and Chib (1993)).When the valuation is observed, the posterior distribution of the selection variable is

$$w(t) \mid Z, v, \gamma \sim LTN\left(Z'(t)\gamma_0 + v(t)\gamma_v, 1\right). \tag{13}$$

When it is unobserved, it is

$$w(t) \mid Z, v, \gamma \sim UTN\left(Z'(t)\gamma_0 + v(t)\gamma_v, 1\right). \tag{14}$$

Here, $LTN(\mu,\sigma^2)$ denotes a Normal distribution with mean μ and variance σ^2 truncated at zero from below, and *UTN* is the same distribution truncated at zero from above.

C. Draw Parameters Using a Bayesian Linear Regression

Conditional on $v(t)$ and $w(t)$ the distributions of δ, β, σ^2, and γ are given by two Bayesian linear regressions. Since $\varepsilon(t) \perp \eta(t)$ by assumption, we estimate the two equations separately.

In the valuation equation, δ, β, and σ^2 are defined by the regression of the excess returns $Y_v(t) = v(t) - v(t-1) - r$ (stacked for all companies) on a constant term and $r_m(t) - r$. Let $N(t)$ be the number of companies for which $v(t)$ exists, so $Y_v(t)$ is a $N(t)\times 1$ vector. Correspondingly, let $X_v(t)$ be a $N(t)\times 2$ matrix with a constant term and $r_m(t) - r$. Let Y_v and X_v contain $Y_v(t)$ and $X_v(t)$ stacked over all periods. The standard conjugate Normal-Inverse Gamma prior with prior parameters α_0, β_0, μ_0, and Σ_0 is

$$\sigma^2 \sim IG(a_0, b_0) \tag{15}$$

$$\delta, \beta \mid \sigma^2 \sim N(\mu_0, \sigma^2 \Sigma_0^{-1}) . \tag{16}$$

The posterior distributions for the parameters in the valuation equation are then (e.g. Rossi, Allenby, and McCulloch (2005)):

$$\sigma^2 \mid Y_v, X_v \sim IG(a, b) \tag{17}$$

$$\delta, \beta \mid \sigma^2, Y_v, X_v \sim N(\mu, \sigma^2 \Sigma^{-1}), \tag{18}$$

with parameters

$$a = a_0 + \sum_t N(t), \tag{19}$$

$$b = b_0 + e'e + (\mu - \mu_0)' \Sigma_0 (\mu - \mu_0), \tag{20}$$

$$\Sigma = \Sigma_0 + X_v' X_v, \tag{21}$$

$$\mu = \Sigma^{-1}(\Sigma_0 \mu_0 + X_v' Y_v), \tag{22}$$

The vector *e* contains the stacked error terms $e = Y_v - X_v \mu$.

The selection equation is simpler. The parameters are given by a linear regression of $Y_s(t) = w(t)$ on $X_s(t) = [Z'(t) \quad v(t)]$, again stacked over companies. To identify the scale of the parameters, we normalize the variance of the error term to one. The prior distribution of $\gamma = [\gamma_0 \ \ \gamma_v]$ is

$$\gamma \sim N(\theta_0, \Omega_0^{-1}), \tag{23}$$

and the posterior distribution becomes

$$\gamma \mid Y_s, X_s \sim N(\theta, \Omega^{-1}), \tag{24}$$

with

$$\Omega = \Omega_0 + X_s' X_s, \tag{25}$$

$$\theta = \Omega^{-1}(\Omega_0\theta_0 + X_s'Y_s)\,. \tag{26}$$