INFORMATION ACQUISITION AND UNDER-DIVERSIFICATION TECHNICAL APPENDIX

Stijn Van Nieuwerburgh and Laura Veldkamp[*]

November 13, 2008

A Information and Portfolio Choice in General Equilibrium with Correlated Assets

This section extends the model with mean-variance preferences and an entropy learning technology by endogenizing the prices of the risky assets. In such an equilibrium model, an investor must consider the information acquisition and investment strategies of other investors. Information is a strategic substitute in this setting: Investors want to learn about assets that others are not learning about. In equilibrium, this means that ex-ante identical investors will choose to observe different signals and will hold different assets. The nature of the individual's problem does not change. After accounting for the actions that other investors will take and how these will affect asset prices, an investor chooses one risk factor and concentrates all his capacity on learning about that one factor.

A.1 Modifications to the Setup

There is now a continuum of atomless investors, indexed by $j \in [0, 1]$. Preferences are mean-variance (equation 18 of the main text) and payoffs are identical to the model described in section 1 of the main text. The one change to the timing of events is that prices of the risky assets are revealed at time 2. We continue to hold the risk-free rate fixed. An equilibrium model requires two additional assumptions. First, the per capita supply of the risky asset is $\bar{x} + x$, a constant plus a random $(n \times 1)$ vector with known mean and variance, and zero covariance across assets: $x \sim N(0, \sigma_x^2 I)$. This risky asset supply creates noise in the price level that prevents investors perfectly inferring the private information of others. Without it, there would be no private information, and no incentive to learn. We interpret this noise as liquidity shocks, life-cycle needs of traders, or errors that investors make when inverting prices.[1] Second, we assume that each investor processes his own information; he observes his own signal whose noise is independent from the noise in other investors' signals.

   A third assumption replaces the the previous requirement that signals about each asset must be independent with the requirement that signals about each independent risk factor must be independent. (We relax this assumption in a partial equilibrium model in the next section.) To

[*]Van Nieuwerburgh: New York University, Stern School of Business, 44 West Fourth St., 9-190, New York, NY 10012; email:svnieuwe@stern.nyu.edu, Phone:212-998-0673, Fax: 212-995-4256. Veldkamp: New York University, Stern School of Business, 44 West Fourth St., 7-180, New York, NY 10012; email:lveldkam@stern.nyu.edu.

[1]See Biais, Bossaerts and Spatt (2004) for an interpretation in terms of risky non-tradeable endowments.

state this assumption formally, we need new notation. When payoffs co-vary, obtaining a signal about one asset's payoff is informative about other payoffs. To describe what a signal is about, it is useful to decompose asset payoff risk into orthogonal risk factors and the risk of each factor. This decomposition breaks the prior variance-covariance matrix Σ up into a diagonal eigenvalue matrix Λ, and an eigenvector matrix Γ: $\Sigma = \Gamma\Lambda\Gamma'$. The Λ_i's are the prior variances of each risk factor i. The ith column of Γ (denoted Γ_i) contains the loadings of each asset on the ith risk factor. Formally, the new assumption is that they choose the eigenvalues $\hat{\Sigma}$ of their posterior variance matrix, but not its eigenvectors Γ. This is equivalent to assuming that investors observe independent signals ($\Gamma'\eta$) about risk factor payoffs ($\Gamma'f$), or that they learn about the underlying Arrow securities independently. Nothing prevents investors from learning about many risk factors. The only thing this rules out is signals with correlated information about independent risks.

Asset prices p are determined by market clearing: the sum of investors' demands for each asset equals its supply.

$$\int_0^1 \widehat{\Sigma}_j^{-1}(\hat{\mu}_j - pr)dj = \bar{x} + x \tag{1}$$

A.2 Individual's Asset Allocation in Equilibrium

As before, we work backwards, starting with the optimal portfolio decision. In period 2, investors have three pieces of information that they aggregate to form their expectation of the assets' payoffs: their prior beliefs (common across investors), their signals (draws from distributions chosen in period 1), and the equilibrium asset price.

Proposition 1. *Asset prices are a linear function of the asset payoff and the unexpected component of asset supply.*

$$p = \frac{1}{r}(A + Bf + Cx) \tag{2}$$

This price is also a function of the posterior mean and variance of the 'average' investor:

$$p = \frac{1}{r}\left(\hat{\mu}_a - \rho\widehat{\Sigma}_a(\bar{x} + x)\right) \tag{3}$$

where the average posterior mean is $\hat{\mu}_a = \int_0^1 \hat{\mu}_j dj$ and the 'average' posterior variance is a harmonic mean of all investors' variances $\widehat{\Sigma}_a = \left(\int_0^1 \widehat{\Sigma}_j^{-1} dj\right)^{-1}$.

Proof. From Admati (1985), we know that equilibrium price takes the form $rp = A + Bf + Cx$ where

$$A = \left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1}\left(\Sigma^{-1}\mu - \rho\bar{x}\right), \tag{4}$$

$$B = \left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1}\left(\Psi + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi\right), \tag{5}$$

$$C = -\left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1}\left(\rho I + \frac{1}{\rho\sigma_x^2}\Psi'\right). \tag{6}$$

Here, Ψ is the average of investors' signal precision matrices $\Psi = \int_0^1 \Sigma_{\eta j}^{-1} dj$, and $\Sigma_{\eta j}$ is the variance-

2

covariance matrix of the signals that investor j observes.[2]

Using (10), note that $\left(\Sigma^{-1} + \frac{1}{\rho^2 \sigma_x^2} \Psi' \Psi + \Psi\right)^{-1} = \hat{\Sigma}_a$, the posterior variance for an investor with the average of all investors' posterior precisions:

$$\hat{\Sigma}_a \equiv \left(\int_0^1 \hat{\Sigma}_j^{-1} dj\right)^{-1} \tag{7}$$

Note also that $\Sigma_p \equiv \sigma_x^2 B^{-1} CC' B^{-1'} = (\frac{1}{\rho^2 \sigma_x^2} \Psi' \Psi)^{-1}$.

Then, the price equation can be rewritten as

$$rp = \hat{\Sigma}_a (\Sigma^{-1} \mu + \Psi f + \Sigma_p^{-1}(f - \rho \Psi^{-1} x) - \rho(\bar{x} - x)$$

Simple algebra reveals that $(f - \rho \Psi^{-1} x) = B^{-1}(rp - A)$, the unbiased signal that investors observe from the price level. The first three terms are equal to the posterior mean of the 'average' investor's beliefs (see equation (9)):

$$\hat{\mu}_a \equiv E_a[f_i] \equiv \int_0^1 \hat{\mu}_j dj \tag{8}$$

Thus, $rp = \hat{\mu}_a - \rho \hat{\Sigma}_a (\bar{x} + x)$. The price level is increasing in the posterior belief of the average investor about the mean payoff, and decreasing in risk aversion, the amount of risk the average investor bears, and the supply of the asset. □ □

If prices take this form, then inverting (2), the mean and variance of the asset payoff, conditional on prices are $E[f|p] = B^{-1}(rp - A)$ and $V[f|p] = \sigma_x^2 B^{-1} CC' B^{-1'} \equiv \Sigma_p$. Denote its eigenvalue matrix by Λ_p. Then, posterior beliefs about the asset payoff f, conditional on prior beliefs $\mu \sim N(f, \Sigma)$, signals $\eta \sim N(f, \Sigma_\eta)$, and prices, can be expressed using standard Bayesian updating formulas. The posterior mean belief about f is

$$\hat{\mu} \equiv E[f|\mu, \eta, p] = \left(\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1}\right)^{-1} \left(\Sigma^{-1} \mu + \Sigma_\eta^{-1} \eta + \Sigma_p^{-1} B^{-1}(rp - A)\right) \tag{9}$$

with variance that is a harmonic mean of the three signal variances.

$$\hat{\Sigma} \equiv V[f|\mu, \eta, p] = \left(\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1}\right)^{-1}. \tag{10}$$

These are the conditional mean and variance that investors use to form their portfolios in period 2. As in the partial equilibrium problem, optimal portfolios are $q^\star = \hat{\Sigma}^{-1}(\hat{\mu} - pr)/\rho$. The only change is in how the asset price affects the posterior mean and variance.

---

[2]The Lebesgue integral may not be well defined when $\{\eta_j\}$ are processes of independent random variables for a continuum of investors j, because realizations may not be measurable with respect to the joint space of parameters and samples. Also, the sample function giving each investor's individual shock may not be Lebesgue measurable, and thus the fraction of investors associated with each shock may not be well defined. Making independence compatible with joint measurability requires defining an enriched probability space, where the one-way Fubini property holds. Then the exact law of large numbers is restored. See Hammond and Sun (2003), and Duffie and Sun (2004) for recent solutions.

A.3 Information Capacity Allocation in Equilibrium

In period 1, the investor chooses a covariance matrix for his posterior beliefs $\widehat{\Sigma}$, just as in the partial equilibrium problem. The difference is that at time 1, payoffs and prices are unknown. The time-2 expectation of excess return is distributed $(\hat{\mu} - pr) \sim N\left((I - B)\mu - A, V_{ER}\right)$ where $V_{ER} \equiv \Sigma - \widehat{\Sigma} + B\Sigma B' + CC'\sigma_x^2 - 2\Sigma B'$. Following the same steps as in the proof of proposition 6, we compute period-1 expected utility:

$$U = \frac{1}{2}Tr(\widehat{\Sigma}^{-1}V_{ER}) + \frac{1}{2}((I - B)\mu - A)'\widehat{\Sigma}^{-1}((I - B)\mu - A). \tag{11}$$

Just as in partial equilibrium, the choice of posterior covariance matrix $\widehat{\Sigma}$ is subject to three restrictions. The first constraint is that the total information the investor sees cannot exceed his capacity. Entropy constrains the distance between the posterior belief variance $\widehat{\Sigma}$ and the prior belief variance Σ. In doing so, it requires that some capacity must be devoted to price discovery; the remaining capacity can be optimally allocated to signals. The second constraint prevents investors from forgetting information that is either contained in priors or in prices. Since $\Sigma^{-1} + \Sigma_p^{-1} = V[f|\mu, p]^{-1}$ is what the conditional precision of asset payoffs would be if the investor observed priors and prices, but no private signals, this constraint requires that $\widehat{\Sigma}^{-1} - (\Sigma^{-1} + \Sigma_p^{-1})$ is positive semi-definite. The third restriction, discussed previously, is that investors can choose only the eigenvalues $\hat{\Sigma}$ of their posterior variance matrix.

To state the information choice problem neatly, we first define two pieces of notation. The prior squared Sharpe ratio of risk factor i is

$$\theta_i^2 \equiv \frac{(E[\Gamma_i'(f - pr)])^2}{Var[\Gamma_i'f]} = \frac{(((I - B)\mu - A)'\Gamma_i)^2}{\Lambda_{ii}}. \tag{12}$$

The expected pricing error on risk factor i is

$$X_i = (1 - \Lambda_{Bi})^2 + \Lambda_i^{-1}\Lambda_{Ci}^2\sigma_x^2, \tag{13}$$

where Λ_{Bi} and Λ_{Ci} are the ith eigenvalues of B and C in proposition 1. The idea is that the noise in prices depends on how much asset prices are affected by true payoffs (fundamentals) and by asset supply shocks. The Λ_{Bi} and Λ_{Ci} are the weights of fundamentals and supply shocks in the equilibrium pricing equation.

Finally, we can rewrite the objective in terms of an eigenvalue choice problem. Note that $V_{ER} = \Gamma\Lambda(1 - \Lambda_B)^2\Gamma' - \sigma_x^2\Gamma\Lambda_C^2\Gamma' - \widehat{\Sigma}$. Using this and substituting (12), (13) and the identity $\widehat{\Sigma} = \Gamma\hat{\Sigma}\Gamma'$ into the objective (11) and rearranging, we can write investors' period-1 information choice problem as

$$max_{\{\hat{\Sigma}_1,\cdots,\hat{\Sigma}_N\}}\frac{1}{2}\{-N + \sum_{i=1}^{N}(X_i + \theta_i^2)\frac{\Sigma_i}{\hat{\Sigma}_i}\}. \tag{14}$$

$$s.t. \quad \prod_{i=1}^{N}\frac{\Sigma_i}{\hat{\Sigma}_i} = K \quad \text{and} \quad \hat{\Sigma}_i^{-1} \geq \Sigma_i^{-1} + \Lambda_{pi}^{-1} \, \forall i$$

If an investor becomes informed, his valuation of a risk factor, based on his private information, will deviate from the realized price. The expected squared deviation is X_i, which we call the *pricing*

error. The first term of (13) shows that pricing errors increase when prices are less reflective of true payoffs. Λ_{Bi} captures the covariance of the ith risk factor's price with its true payoff. When Λ_{Bi} is small, a low covariance makes X_i large. For example, if a well-informed investor sees a low price and knows that the true payoffs are likely to be high, he can exploit this by buying the asset. The uninformed investor, who knows little about the true payoff, cannot exploit this difference, and is taken advantage of by the informed investor. The second term of (13) shows that pricing errors increase when prices are more reflective of asset supply shocks. Λ_{Ci} is the weight of the ith risk factor's supply shock on the factor's price. When Λ_{Ci} is high, supply shocks create noise in prices that is exploitable by a well informed investor. For example, if such an investor sees a low price and knows it is due to a high supply shock, he can exploit this by buying the asset. The uninformed investor, on the other hand, attributes this low price to fundamentals. In sum, risk factors with a higher X_i are more valuable to learn about, because informed investors can make more profit.

Proposition 2. *In general equilibrium with a continuum of investors, each investor's optimal information portfolio uses all capacity to learn about one linear combination of asset payoffs. The linear combination weights are given by the eigenvector Γ_i associated with the highest value of the general equilibrium learning index: $\theta_i^2 + X_i$.*

Proof. The posterior variance of the average investor has the same eigenvectors as investors' prior beliefs do. This is true because sums, products and inverses of matrices with identical eigenvectors preserve those eigenvectors. This tells us that Ψ can be rewritten as $\Psi = \int_0^1 \Gamma^{-1'}\Lambda_{\eta j}^{-1}\Gamma^{-1}dj$. Since eigenvector matrices have the property that $\Gamma^{-1} = \Gamma'$, and defining $\Lambda_{\eta a}^{-1} = \int_0^1 \Lambda_{\eta j}^{-1}dj$, this is equivalent to $\Psi = \Gamma\Lambda_{\eta a}\Gamma'$. Because Σ_p, $\hat{\Sigma}_a$, B, and C are result from a combination of sums, products and inverses of Σ and Ψ, all have eigenvectors Γ.

Since $\hat{\Sigma}^{-1}$ and V_{ER} have the same eigenvectors as Σ, the proof of the proposition follows immediately from the proof of proposition 4, where $E[f - pr]$ is now based on prior beliefs ($E[f - pr] = (I - B)\mu - A$), instead of on ($\mu - pr$). $\square$ $\square$

Just as in partial equilibrium, each investor continues to specialize in learning about one risk factor. Again, the reason is that the objective function is convex in $\hat{\Sigma}_i$. The learning index $\theta_i^2 + X_i$ determines which risk factor the investor learns about. The most valuable risk factor to learn about has (i) a high expected return $\Gamma_i'E[f - pr]$, (ii) a large expected portfolio share $\Gamma_i'E[q]$, and (iii) a large exploitable pricing error. Expected returns and exploitable pricing errors are affected by the fraction of investors who learn about risk factor i. This new general equilibrium effect makes learning a strategic choice.

A.4 Aggregate Information Choice and Asset Returns

Information that investors learn is partly revealed through the asset price level. This learning externality makes learning a strategic substitute: investors prefer to learn information that others are not learning. Without substitutability, investors would specialize in learning about the same risk factor and would want to hold the same risk factor. All investors cannot hold more of one risk because assets are in fixed supply. Each investor would hold an equal expected share of the diversified market portfolio. Thus, substitutability is critical for specialization to sustain under-diversification, in general equilibrium.

Asset Returns   In order to characterize aggregate information choices, we must first understand the source of the externality: asset returns. Taking the expectation of the price in proposition 1 and rearranging delivers an expression for the expected return. It tells us that as long as assets are in positive net supply ($\bar{x} > 0$), the increase in information about an asset (fall in $\widehat{\Sigma}_a$) will cause its expected return to fall: $E[f - pr] = \rho\widehat{\Sigma}_a\bar{x}$.

Since learning choices are over risk factors, it is useful to write the expected return on a risk factor. The equilibrium risk premium of factor i is then the sum of the asset returns, each weighted by their loading on the risk factor Γ_i:

$$\Gamma_i'E[f - rp] \;=\; \rho\hat{\Lambda}_{ai}(\Gamma_i'\bar{x}). \tag{15}$$

The equilibrium factor risk premium depends on (i) the risk aversion of the economy ρ, (ii) the supply of the risk factor $\Gamma_i'\bar{x}$, and most importantly (iii) the weight $\hat{\Sigma}_{ai}$, the ith eigenvalue of aggregate variance matrix $\widehat{\Sigma}_a$. This weight measures how much the economy (the average investor) learns about risk factor i. A risk factor that the economy does not learn about has weight $\hat{\Lambda}_{ai} = \Lambda_i$. A risk factor that the economy learns about has a weight $\hat{\Lambda}_{ai} < \Lambda_i$. In other words, as more investors learn about risk factor i, $\hat{\Lambda}_{ai}$ decreases and its risk premium falls.

Strategic Substitutability in Information   When more investors learn about a risk factor, assets that load heavily on that factor experience a decline in expected return and exploitable pricing errors. Both effects make assets less valuable to learn about (lower $\theta_i^2 + X_i$). This generates strategic substitutability in learning.[3] Let Ψ be the average of investors' signal precision matrices $\Psi = \int_0^1 \Sigma_{\eta j}^{-1} dj$, where $\Sigma_{\eta j}$ is the variance-covariance matrix of the signals that investor j observes. Let $\Lambda_{\Psi i}$ be the eigenvalue of Ψ corresponding to the ith risk factor. The following proposition says that when the average investor becomes less uncertain about a risk factor i ($\Lambda_{\Psi i}$ falls), it becomes less valuable to learn about.

Proposition 3. *There is strategic substitutability in learning: $(X_i + \theta_i^2)$ is a strictly decreasing, monotonic function of average signal precision $\Lambda_{\Psi i}$.*

Proof. X_i and θ_i^2 are both decreasing in $\Lambda_{\psi i}$. Thus, their sum is decreasing. We start by deriving the expression for X_i. The first part of the objective is $Tr\left(\widehat{\Sigma}^{-1}V_{ER}\right)$, which we rewrite as $Tr\left(\widehat{\Sigma}^{-1}\Sigma\Sigma^{-1}(V_{ER} + \widehat{\Sigma} - \widehat{\Sigma})\right)$. This is $Tr\left(\widehat{\Sigma}^{-1}\Sigma\Sigma^{-1}(V_{ER} + \widehat{\Sigma}) - I\right)$ or $Tr\left(\widehat{\Sigma}^{-1}\Sigma\Sigma^{-1}(V_{ER} + \widehat{\Sigma})\right) - N$. The trace is the product of the eigenvalues. Let $\frac{\Sigma_i}{\widehat{\Sigma}_i}$, be the ratio of the precision of the posterior to the precision of the prior, i.e. it is the ith eigenvalue of $\widehat{\Sigma}^{-1}\Sigma$: $\frac{\Sigma_i}{\widehat{\Sigma}_i} \equiv \hat{\Sigma}_i^{-1}\Lambda_i$. Let X_i be the ith eigenvalue of $\Sigma^{-1}(V_{ER} + \widehat{\Sigma})$. Then the ith eigenvalue of the matrix inside the trace is $\frac{\Sigma_i}{\widehat{\Sigma}_i}X_i$, and $Tr\left(\widehat{\Sigma}^{-1}\Sigma\Sigma^{-1}(V_{ER} + \widehat{\Sigma})\right) = \sum_{i=1}^{N} X_i\frac{\Sigma_i}{\widehat{\Sigma}_i}$. This is because Σ, $\widehat{\Sigma}$, B and C all share the same eigenvectors Γ.

X_i, the ith eigenvalue of $\Sigma^{-1}(V_{ER} + \widehat{\Sigma})$, is $X_i = \Lambda_i^{-1}\left[\Lambda_i\left(1 + \Lambda_{Bi}^2 - 2\Lambda_{Bi}\right) + \Lambda_{Ci}^2\sigma_x^2\right]$, where Λ_{Bi} and Λ_{Ci} are the ith eigenvalue of B and C respectively. Using the definition of B and $\hat{\Sigma}_a$, we can rewrite B as $I - \hat{\Sigma}_a\Sigma^{-1}$, which has eigenvalues $\Lambda_{Bi} = 1 - \hat{\Lambda}_{ai}\Lambda_i^{-1}$, where Λ_{ai} is the ith

---

[3]In Barlevy and Veronesi (2000), a different assumption on the distribution of noise in prices generates complementarity in information acquisition. This effect comes from exploitable pricing errors that rise as more investors become informed ($\partial X_i/\partial \Lambda_{\Psi i} > 0$). In their model, the effect on expected returns is still a force for strategic substitutability.

eigenvalue of $\hat{\Sigma}_a$. Also, using the definitions of B and C, we have $C = -\rho B \Psi^{-1}$, and hence $CC'\sigma_x^2 = B\left(\frac{1}{\rho^2\sigma_x^2}\Psi'\Psi\right)^{-1}B'$, which in turn equals $B\Sigma_p B'$. The ith eigenvalues of the matrix $\Sigma^{-1}CC'\sigma_x^2$, $\Lambda_i^{-1}\Lambda_{Ci}^2\sigma_x^2$, are thus equal to $\Lambda_{ii}^{-1}\Lambda_{Bi}^2\Lambda_{pi} = \Lambda_i^{-1}(1 - \hat{\Lambda}_{ai}\Lambda_i^{-1})^2\Lambda_{pi}$. Now we can rewrite X_i as:

$$X_i = \left(\frac{\hat{\Lambda}_{ai}}{\Lambda_i}\right)^2 + \left(\frac{\Lambda_{pi}}{\Lambda_i}\right)\left(1 - \frac{\hat{\Lambda}_{ai}}{\Lambda_i}\right)^2.$$

An important property of X_i is that it is decreasing in the average signal precision of risk factor i, $\Lambda_{\Psi i}$, the ith eigenvector of Ψ. To simplify notation, define $a \equiv \frac{1}{\rho^2\sigma_x^2}$. To show strict substitutability is to show $\frac{\partial X_i}{\partial \Lambda_{\Psi i}} < 0$. We first recall that $\Lambda_{pi}^{-1} = a\Lambda_{\Psi i}^2$ and $\hat{\Lambda}_{ai}^{-1} = \Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i}$. We can rewrite X_i in our new notation as:

$$
\begin{aligned}
X_i &= \Lambda_{ii}^{-2}(\Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})^{-2} + \Lambda_{ii}^{-1}a^{-1}\Lambda_{\Psi i}^{-2}(a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})^2(\Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})^{-2}, \\
&= \Lambda_{ii}^{-1}(\Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})^{-2}[\Lambda_{ii}^{-1} + a^{-1} + 2\Lambda_{\Psi i} + a\Lambda_{\Psi i}^2].
\end{aligned}
$$

Taking a partial derivative with respect to $\Lambda_{\Psi i}$, we get:

$$
\begin{aligned}
\frac{\partial X_i}{\partial \Lambda_{\Psi i}} &= -2\Lambda_{ii}^{-1}(\Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})^{-3} \\
&\quad \cdot[(\Lambda_{ii}^{-1} + a^{-1} + 2\Lambda_{\Psi i} + a\Lambda_{\Psi i}^2)(2a\Lambda_{\Psi i} + 1) - (\Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})(a\Lambda_{\Psi i} + 1)], \\
&= -2\Lambda_{ii}^{-1}(\Lambda_{ii}^{-1} + a\Lambda_{\Psi i}^2 + \Lambda_{\Psi i})^{-3}[a^2\Lambda_{\Psi i}^3 + 3a\Lambda_{\Psi i}^2 + (3 + a\Lambda_{ii}^{-1})\Lambda_{\Psi i} + a^{-1}].
\end{aligned}
$$

The partial derivative is strictly negative because $\Lambda_{ii}^{-1} > 0$, $a > 0$, $\Lambda_{\Psi i} > 0$, and hence the term in parentheses and the term in brackets are strictly positive. Using L'Hôpital's rule, it is easy to show that $lim_{\Lambda_{\Psi i} \to 0} X_i = 1 + a^{-1}\Lambda_{ii}$, which equals $1 + \rho^2\sigma_x^2\Lambda_i$. Because of the new source of risk induced by noisy asset supply (σ_x^2), X_i is strictly greater than 1 when nobody learns about risk factor i ($\Lambda_{\Psi i} = 0$). Note that this is consistent with $X_i = 1$ in partial equilibrium, where prices we taken as given ($\sigma_x^2 = 0$).

We conclude the proof by showing that $\theta_i^2 = \frac{(((I-B)\mu - A)'\Gamma_i)^2}{\Lambda_i}$ is decreasing in $\Lambda_{\psi i}$. The denominator Λ_{ii} is exogenous. Using the formulas for A and B (4),(5), the expected return is $((I-B)\mu - A) = \rho\hat{\Sigma}^a\bar{x}$. Thus, $((I-B)\mu - A)'\Gamma_i = \rho\hat{\Sigma}_i^a(\Gamma_i'\bar{x})$. Since, $\hat{\Sigma}_i^a = (\Lambda_i^{-1} + \frac{1}{\rho^s\sigma_x^2}\Lambda_{\psi i}^2 + \Lambda_{\psi i})^{-1}$, the expected return and its square are decreasing in $\Lambda_{\psi i}$. $\square$ $\qquad\qquad$ $\square$

Aggregate Information Allocation In equilibrium, ex-ante identical investors are indifferent between learning about any one of the risk factors that the economy learns about. Each investors randomizes over which factor to learn about. So, which risk factor a particular individual learns about is indeterminate. But the equilibrium mixed strategy, or the fraction of agents learning about any one risk factor, is unique.

The equilibrium information allocation depends on investors' capacity. As capacity rises, the number of risk factors that the economy learns about increases, although each individual investor only learns about one risk. From proposition 2, we know that investors always allocate their capacity to the asset with the highest value of $(\Gamma_i'E[f - pr])^2(\Lambda_i)^{-1} + X_i$. Begin by ordering risk factors by their learning index values when $K = 0$, such that $(\Gamma_i'E[f - pr])^2(\Lambda_i)^{-1} + X_i \geq (\Gamma_{i+1}'E[f - pr])^2(\Lambda_{i+1})^{-1} + X_{i+1}$. For small levels of K, capacity is allocated only to the most

valuable risk factor, say risk factor 1, and to additional risk factors, only if their initial learning index value is equal to that of factor 1: $(\Gamma_1' E[f - pr])^2 (\tilde{\Lambda}_1)^{-1} + X_1 = (\Gamma_i' E[f - pr])^2 (\tilde{\Lambda}_i)^{-1} + X_i$.

As capacity rises, the learning index of the assets being learned about falls. As $K \to \infty$, precision of beliefs about asset 1 becomes infinite: $\Psi_{11} \to \infty$. Equation (15), shows that that $(\Gamma_1' E[f - pr])^2 \to 0$. Furthermore $X_i \to 0$ because there are no more exploitable pricing errors when beliefs are infinitely precise. Since index values are non-negative, there is some K_j for each asset j s.t. $\forall K > K_j$, investors learn about risk factor j.

To understand this result, consider three economies where investors have different capacities. In the low-capacity economy, all investors learn only about the risk factor with the highest learning index. In the medium-capacity economy, investors randomize between 2 risk factors. The proportions of investors that learn about 1 and about 2 is such that all investors are indifferent. Finally, in the high-capacity economy, investors are indifferent between learning about many risk factors. This type of result is referred to as 'water-filling' in the information theory literature.

A.5   Equilibrium Asset Prices

When the average investor learns about an asset, its risk and return fall. A model that does not account for information choice should have systematic pricing errors. This is a testable relationship between learning indices and prediction errors of the capital asset pricing model (CAPM).

In both a heterogeneous information model and a standard CAPM, an asset's expected excess return is proportional to its market beta and to the market risk premium. But the betas in the two models differ. The CAPM beta is the coefficient of a least-squares estimate of $R_{it} - r = \alpha + \beta_{CAPM}^i (R_{mt} - r)$; it measures the *unconditional* relationship between asset i returns and market returns. The learning model generates a beta that is *conditional* on information the average investor knows.

Proposition 4 states that asset prices and returns in our model are identical to an economy where a representative investor believes that payoffs are distributed $N(E_a[f], \widehat{\Sigma}_a)$; $E_a[f]$ is the average expectation and $\widehat{\Sigma}_a$ is the harmonic average covariance of our heterogeneously informed investors. Moments conditional on these average beliefs are denoted with an a subscript.

Proposition 4. *If the market payoff is defined as $f_m = \sum_{k=1}^N (\bar{x} + x_k) f_k$, the market return is $R_m = f_m \left(\sum_{k=1}^N (\bar{x} + x_k) p_k \right)^{-1}$, and the return on an asset i is $R_i = f_i / p_i$, then the equilibrium price of asset i can be expressed as $p_i = \frac{1}{r} (E_a[f_i] - \rho\, Cov_a[f_i, f_m])$. The equilibrium return is $E_a[R_i] - r = \beta_a^i (E_a[R_m] - r)$, where $\beta_a^i \equiv Cov_a[R_i, R_m]/Var_a[R_m]$.*

Proof. We can rewrite equation (3) for each asset $i \in \{1, 2, \cdots, N\}$ separately:

$$p_i = \frac{1}{r} \left(\hat{\mu}_a^i - \rho \sum_{k=1}^N Cov_a[f_i, f_k](\bar{x} + x_k) \right) = \frac{1}{r} \left(\hat{\mu}_a^i - \rho\, Cov_a[f_i, \sum_{k=1}^N (\bar{x} + x_k) f_k] \right)$$

where $Cov_a[f_i, f_k]$ denotes the (i, k) element of $\hat{\Sigma}_a$. Using the definition of f_m stated in the proposition, we obtain the first equation mentioned in the proposition: $p_i = \frac{1}{r} (E_a[f_i] - \rho\, Cov_a[f_i, f_m])$. To rewrite this equilibrium price function in terms of returns divide both sides by the price. Denote the return on security i by $r^i \equiv \frac{f_i}{p_i}$. Simple manipulation leads to:

$$E_a[r_i] - r = \rho\, Cov_a[r_i, f_m]. \tag{16}$$

8

This is true for each asset i, and hence also for asset m:

$$E_a[r_m] - r = \rho\, p_m Cov_a[r_m, r_m]. \tag{17}$$

Solving (17) for the risk aversion coefficient ρ, and substituting it into (16), we get the second equation in the proposition. $\qquad\square$

Assets the average investor learns more about should have lower betas, and therefore lower returns, than what the CAPM predicts. To see why, note that $\beta_{CAPM} = Corr[R_i, R_m]Std(R_i)/Std(R_m)$ and $\beta_a = Corr[R_i, R_m]Std_a(R_i)/Std_a(R_m)$. The correlation terms are the same because learning does not change the correlation structure, it reduces the standard deviation. If the average investor learned as much about asset i as about all other assets in the market, then the ratio of standard deviations would be the same, with our without learning. But, if the average investor learned more about asset i, then $Std_a(R_i)/Std_a(R_m)$ would fall. For these assets, $\beta_a^i < \beta_{CAPM}^i$. Information the average investor knows should lower $Std_a(R_i)$, lower the asset's β_a^i, and lower its predicted return.

To test this hypothesis, compute the expected return on each risk factor i according to a standard CAPM, subtract the mean observed return to form the prediction error, and test for a positive relationship between prediction errors and learning indices. This exercise is similar to work by Biais et al. (2004) who also derive a modified CAPM for a market with asymmetric information. They show how this estimation strategy can be used to infer the amount of private information present in the market.

B Allowing correlated signals

We now return to the partial equilibrium model and consider models where signals are not required to be independent. Rather than reconsider every combination of preferences and technologies, we work through two examples of the correlated assets and correlated signals model. We show that the indifference result for CARA and entropy and the specialization result for the mean-variance investor with an entropy learning technology survive.

Indifference with CARA preferences and entropy persists The proof simply does not rely on signal independence. The expression for expected utility relies only on the normality of the signal distribution. Since signals only affect $\widehat{\Sigma}$ and this enters in as a function of $|\widehat{\Sigma}^{-1}|/|\Sigma| = K$, any set of signals that satisfy the capacity constraint deliver equal expected utility, whether those signals are correlated or not.

Optimal specialization with mean-variance preferences and entropy persists For tractability, the main text assumes that investors learn about independent risks independently. In the case of correlated assets (a non-diagonal Σ), this means that signals are about the payoffs of the principal components $f'\Gamma$, where Γ comes from the decomposition $\Sigma = \Gamma'\Lambda\Gamma$. In other words, investors can learn about N orthogonal risk factors. The first such risk factor (principal component) could be interpreted as "the market." So, investors are allowed to learn about the market. The same is true for all other risk factors, which could represent industry-specific risk factors or company-specific risk factors. When asset payoffs are independent, the matrix Γ is the identity matrix, so that investors are learning about asset payoffs, not linear combinations of asset payoffs.

Let the posterior covariance matrix be $\widehat{\Sigma} = \widehat{\Gamma}'\hat{\Sigma}\widehat{\Gamma}$, where $\widehat{\Gamma}$ is an *arbitrary but fixed* posterior eigenvector matrix of $\widehat{\Sigma}$. In particular, $\widehat{\Gamma} \neq \Gamma$. The mean-variance objective function then becomes:

$$\frac{1}{2}Tr\left(\widehat{\Gamma}\hat{\Sigma}^{-1}\widehat{\Gamma}'\Sigma - I\right) + \frac{1}{2}(\mu - pr)'\widehat{\Gamma}\hat{\Sigma}^{-1}\widehat{\Gamma}'(\mu - pr).$$

One can write the ii element of the matrix $\widehat{\Sigma}^{-1}\Sigma$ as: $\left(\widehat{\Sigma}^{-1}\Sigma\right)_{ii} = \sum_{j=1}^{N}\sum_{l=1}^{N}\widehat{\Gamma}_{il}\hat{\Sigma}_l^{-1}\widehat{\Gamma}_{lj}\Sigma_{ji}$. Since the trace of a matrix is the sum of its diagonal elements, the objective function can be written as a sum of the $\{\hat{\Sigma}_l^{-1}\}$ each weighted by a scalar τ_l: $\frac{1}{2}\sum_{l=1}^{N}\tau_l\hat{\Sigma}_l^{-1} - \frac{N}{2}$, where τ_l is the learning index of the lth risk factor:

$$\tau_l = \sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\Gamma}_{il}\left(\widehat{\Gamma}_{lj}\Sigma_{ji} + \widehat{\Gamma}_{jl}(\mu_i - p_i r)(\mu_j - p_j r)\right)$$

Under the assumption $\widehat{\Gamma} = \Gamma$, we recover the learning index written in the main text: $\tau_l = \Lambda_l + ((\mu - pr)'\Gamma_l)^2$. But, whatever the exact specification of the learning index is, the investor always ranks the risk factors according to their learning index τ_l, and chooses to specialize in the risk factor with the highest τ_l. The reason is that the objective function is still linear in the $\{\hat{\Sigma}_l^{-1}\}$. Since the entropy constraint is still a constraint on the product of the posterior eigenvalues, it is a bound on $\prod_i \hat{\Sigma}_i^{-1}$. Thus, the problem is still maximizing a sum subject to a product constraint, which delivers a corner solution.

C Results for CRRA < 1

Expected utility is given bin section 2.2 of the main text. Working with the log of this objective simplifies the calculations. As before, we collect the constants in a new term (a). Since all the matrices in the problem are diagonal, the objective can be expressed as sums of the matrix diagonal

$$EU = a + \frac{-1}{2}\sum_{i=1}^{N}\log\left(1 + (\gamma - 1)\Sigma_{ii}\hat{\Sigma}_{ii}^{-1}\right) + \frac{\gamma - 1}{2}\sum_{i=1}^{N}\frac{(\mu_i - r)^2}{\hat{\Sigma}_{ii} + (\gamma - 1)\Sigma_{ii}}. \tag{18}$$

Solution with entropy costs We redefine the choice variable as in appendix A.6: The investor chooses $(K_1, \ldots, K_N) \geq 0$ where the choice variable measures the increase in precision, $\hat{\Sigma}_{ii}^{-1} = e^{K_i}\Sigma_{ii}^{-1}$, subject to the constraint that $\sum_i K_i \leq \ln(K)$. The objective and the first order condition are identical to the case with $\gamma > 1$, except with opposite sign. The second derivative is

$$\frac{\partial^2 EU}{\partial K_i^2} = \frac{1 - \gamma}{2}e^{-K_i}\left(\frac{(1 + (\mu_i - r)\Sigma_{ii}^{-1})e^{-K_i} + \gamma - 1 + (\gamma - 1)(\mu_i - r)^2\Sigma_{ii}^{-1}}{(e^{-K_i} + \gamma - 1)^3}\right). \tag{19}$$

$$= \frac{1 - \gamma}{2}e^{-K_i}\left(\frac{(1 + (\mu_i - r)^2\Sigma_{ii}^{-1})(e^{-K_i} + \gamma - 1)}{(e^{-K_i} + \gamma - 1)^3}\right).$$

Canceling out the $(e^{-K_i} + \gamma - 1)$ term leaves and expression that is positive for all assets. The positive results means there are increasing returns to learning about one asset. Thus, as long as the asset under investigation has an expected return that exceeds the risk-free rate, investors want to deepen their knowledge of that asset.

10

Solution with linear precision costs

$$\frac{\partial EU}{\partial \hat{\Sigma}_{ii}^{-1}} = \frac{1-\gamma}{2} \left[\frac{\Sigma_{ii}(1 + (\gamma-1)\Sigma_{ii}\hat{\Sigma}_{ii}^{-1}) + (\mu_i - r)^2}{(1 + (\gamma-1)\Sigma_{ii}\hat{\Sigma}_{ii}^{-1})^2} \right] \tag{20}$$

The second derivative of this expression is

$$\frac{\partial^2 EU}{\partial \hat{\Sigma}_{ii}^{-2}} = \frac{-(1-\gamma)^2\Sigma_{ii}}{2} \left[\frac{\Sigma_{ii}((1-\gamma)\Sigma_{ii}\hat{\Sigma}_{ii}^{-1} - 1) - 2(\mu_i - r)^2}{(1 + (\gamma-1)\Sigma_{ii}\hat{\Sigma}_{ii}^{-1})^3} \right] \tag{21}$$

The sign of this expression depends on the coefficient of relative risk aversion, the squared Sharpe ratio $\Sigma_{ii}^{-1}(\mu_i - r)^2$, and the amount of capacity the investor has because it determines $\Sigma_{ii}\hat{\Sigma}_{ii}^{-1}$. For very low capacity and very high capacity, this second derivative is positive. That means there are increasing returns to learning about one asset. Thus, investors deepen their knowledge. For an intermediate level of capacity, they broaden their knowledge.

D   Portfolio and Information Choice with Log Preferences

A model with log-normally distributed returns and log preferences has properties like those of the CRRA utility function with $\gamma > 1$, but with a simpler indirect utility function that looks just like the one from mean-variance utility. Consider the following utility function

$$U = E_1[E_2[\ln(W)]],$$

where the process for wealth W is given by

$$W_{t+1} = W_t \exp\{r + q'(\mu - r) - \frac{1}{2}q'\Sigma q + q'\Sigma^{1/2}z_t\} \quad \text{where} \quad z_t = Z_{t+1} - Z_t \sim N(0, I_n). \tag{22}$$

The period-2 portfolio choice problem is

$$\max_q E_2 \left[\ln(W_0 \exp(r + q'(\hat{\mu} - r) - \frac{1}{2}q'\hat{\Sigma}q + q'\hat{\Sigma}^{1/2}z_t) \right]$$

$$= \max_q \ln(W_0) + r + q'(\hat{\mu} - r) - \frac{1}{2}q'\hat{\Sigma}q$$

The first order condition is

$$q = \hat{\Sigma}^{-1}(\hat{\mu} - r).$$

Substituting the optimal portfolio into the objective yields a period-1 objective

$$\max_{\hat{\Sigma}} E_1 \left[\ln(W_0) + r + \frac{1}{2}(\hat{\mu} - r)'\hat{\Sigma}^{-1}(\hat{\mu} - r) + (\hat{\mu} - r)'\hat{\Sigma}^{-1/2}z_t \right]$$

This is an expectation of a chi-square. The random variable in the chi-square is $\hat{\mu} \sim N(\mu, \Sigma - \hat{\Sigma})$, plus a mean-zero random variable z_t, which drops out.

$$\max_{\hat{\Sigma}} \ln(W_0) + r + \frac{1}{2}Tr\left(\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma})\right) + \frac{1}{2}(\mu - r)'\hat{\Sigma}^{-1}(\mu - r)$$

Using the assumption that $\Sigma, \hat{\Sigma}$ are diagonal,

$$\max_{\hat{\Sigma}} \ln(W_0) + r + \frac{1}{2} \sum_{i=1}^{N} \left(\frac{\Sigma_{ii} + (\mu_i - r)^2}{\hat{\Sigma}_{ii}} - 1 \right)$$

This objective takes the form of a weighted sum of precisions, exactly like the mean-variance utility function. With an entropy learning technology, maximizing the sum subject to a product constraint will produce a corner solution where the investor specializes in learning about one asset. With a linear learning technology, the solution will be to specialize. The only exception to the specialization results arises when there are equal weights in the objective (the assets' squared Sharpe ratios are equal). In that case, the investor is indifferent about how to allocate capacity.

References

Admati, Anat, "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets," *Econometrica*, 1985, *53(3)*, 629–57.

Barlevy, Gadi and Pietro Veronesi, "Information Acquisition in Financial Markets," *Review of Economic Studies*, 2000, *67*, 79–90.

Biais, Bruno, Peter Bossaerts, and Chester Spatt, "Equilibrium Asset Pricing Under Heterogenous Information," 2004. Working Paper.