Information Acquisition and Portfolio Under-Diversification

Stijn Van Nieuwerburgh and Laura Veldkamp*

December 2004

ABSTRACT

Individual investors hold a fraction of their equity portfolio in a diversified mutual fund and another fraction in a small number of highly-correlated assets. Such behavior is optimal for investors who face constraints on how much they can learn. Optimal under-diversification arises because of increasing returns to scale in learning: As an investor holds more of an asset, the value of learning about it increases, but as he learns more about the asset, it becomes less risky, and more desirable to hold. The interaction of the learning problem and the portfolio problem causes investors to hold some fraction of their assets in a well-diversified fund, about which they learn nothing, and to hold the other fraction in a small set of highly-correlated assets that they learn about. In equilibrium, ex-ante identical investors specialize in learning about different risk factors. Assets whose returns correlate strongly with risk factors that many investors learn about have low expected returns.

*Stijn Van Nieuwerburgh is in the finance department at New York University, Stern School of Business, 44 West Fourth St., 9th floor, New York, NY 10012 (email:svnieuwe@stern.nyu.edu). Laura Veldkamp is in the economics department at New York University, Stern School of Business, 44 West Fourth St., 7th floor, New York, NY 10012 (email:lveldkam@stern.nyu.edu). We thank Dave Backus, Ned Elton, Bob Hall, Pascal Maenhout, Urs Peyer, Tom Sargent, Pierre-Olivier Weill, and seminar participants at NYU, Cornell, USC, UCLA, Stanford and UCSB for helpful comments. JEL classification: G11, D83, G12, G14. Keywords: Asymmetric information, learning, information theory, asset pricing, portfolio theory.

Recent empirical research has shown that many individuals hold under-diversified portfolios of common stock, in addition to a well-diversified mutual fund. The median retail investor at a large on-line brokerage company holds only 2.6 stocks (Barber and Odean, 2000). These portfolios of directly-held equity not only contain too few stocks, but the stocks they contain are positively correlated (Goetzman and Kumar, 2003). But directly-held equities are only 40% of the median household's portfolio; the remaining 60% is in stock and bond mutual funds (Polkovnichenko 2003). What could explain this under-diversified component of households' portfolios, when they know about the existence of, and are currently owning, better diversified, low-cost mutual funds?

We propose a model where rational investors choose how to allocate limited capacity to learn asset-relevant information, before forming their asset portfolios. This capacity limitation represents a time constraint on learning, a budget constraint for research, or a size constraint on the number of bits that a modem can download. When deciding how to allocate their information capacity, investors can choose to learn a small amount about a large number of assets or to specialize and learn precise information about a few assets. Once they begin to learn about a particular asset, investors will want to hold more of that asset in their portfolio because, being risk-averse, they prefer to hold assets that they are informed about. As asset holdings rise, returns to information increase; one signal applied to one share generates less profit than the same signal applied to many shares. Specialization arises because the more an investor holds of an asset, the more valuable it is to learn about that asset; but the more an investor learns about the asset, the more valuable that asset is to hold.

The interaction of the information portfolio problem and the asset portfolio problem creates a trade-off between diversification and specialization through learning. The result is that investors hold some fraction of their assets in a well-diversified fund, about which they learn nothing, and hold the other fraction in a small set of highly-correlated assets that they specialize in learning about. While the idea of gains to specialization has been discussed before (see Ivkovic, Sialm and Weisbenner, 2004), we introduce a new set of information choice tools to formalize this problem. Analyzing information choice sheds light on how diversification and specialization trade-off, what investors choose to specialize in, how investors' learning choices interact, and how aggregate learning affects asset prices.

An investor in our model chooses how much to learn about principal components of asset payoffs, subject to a constraint on his information capacity. As in Shannon (1948) and Sims (2003), capacity governs the bits of data that can be used to reduce the conditional variance of a random variable. After deciding how to allocate his capacity, the investor observes signals about next-period asset payoffs drawn from distributions that he has chosen. Conditional on these signals, he solves a standard CARA-normal portfolio problem. For the investor with zero information capacity, it is optimal to hold a diversified portfolio; our theory collapses to the standard model. As the investor's information capacity increases, holding a perfectly diversified portfolio is still feasible, but no longer optimal. We examine the predictions of this model in both partial equilibrium and general equilibrium settings. This setting is distinct from Peng (2004) because Peng's representative investor must hold all the assets for the market to clear; there is no portfolio choice. In contrast, the focus of our paper is on the interaction of asset portfolio and information choices.

Section II analyzes a partial equilibrium model where the investor takes asset prices as given. When asset returns are uncorrelated (section A), the investor chooses to learn about one asset. He holds a diversified portfolio and adds to that a "learning portfolio" consisting of a single asset. Specialization in learning arises because of increasing returns to learning: One piece of information is most profitably used when it is applied to many shares. Therefore, the investor allocates her capacity to assets she expects to hold in large quantity. By reducing risk, information makes the investor more likely to take an even larger position in the asset. The larger expected position feeds back, magnifying the incentive to learn. In the case of correlated assets (section B), the investor learns about a single risk factor instead of a single asset. She holds assets in her "learning portfolio" that are highly correlated with the risk factor, and hence with each other. We explore the case where not all risk is learnable (section C), and find that the returns to specialization are bounded. Given sufficient information capacity, the investor will learn about more than one risk factor.

Section III investigates a general equilibrium model where a continuum of investors interact, as in Admati (1985). Investors' actions now affect prices. Prices act as an additional source of information: they are a noisy signal of what other investors know (section B). We find that agents still have an incentive to specialize in one risk factor (section C). However, they also have an incentive to specialize in a different risk factor from the ones other agents are learning about. The new insight is that ex-ante identical investors may choose to learn about different risk factors and hold different concentrated asset portfolios (section D). We investigate the implications for the cross-section of asset prices (section E). We find that the risk premium on an asset is low when its correlation with the risk factors that the economy learns about is high. Asset returns are also described by a CAPM; the CAPM that would hold if each investor had the average of all investors' signal precisions. Finally, section IV discusses the implications of the theory for institutional portfolio management.

Why is it relevant to think of investors as information-constrained when information has never been so abundant? It is true that the internet, discount brokers, and real time price quotes give individual investors unparalleled access to financial information. By one estimate, on-line investors have access to 3 billion bits of information for free and 280 billion bits for sale.[1] But, it is precisely because information is overwhelming that capacity constraints on the ability to process that information have become more relevant. Psychologists have long known about human limitations on information absorption (see e.g. Miller, 1956, or Just and Carpenter, 1992). Individuals, as well as institutions, must make decisions about which stocks to follow, which reports to read and what research to do. These are the kinds of choices that our model captures.

Fixed costs of learning (Merton 1987) and fixed costs to trading have been used to explain this under-diversification puzzle. However, fixed costs cannot explain the split nature of portfolios. If diversification is prohibitively expensive, why would households hold so much of their portfolio in diversified funds? Our model keeps diversification feasible by allowing investors to trade in assets without learning about them and to allocate capacity continuously. Moreover, evidence suggests that the degree of diversification only slightly improved over the last decade, in spite of a large drop in (fixed and proportional) transaction and information costs. While the ease of access and the speed of dissemination of financial information have dramatically improved over the last decade, the processing capacity of the investor has not.

If investors concentrate their portfolios because they have informational advantages, then concentrated portfolios should outperform diversified ones (corollary 3). In contrast, if transaction costs or behavioral biases are responsible, then concentrated portfolios should offer no advantage. Ivkovic, Sialm, Weisbenner (2004) find that concentrated investors outperform diversified ones by as much as 3% per year. This excess return is even higher for investments in local stocks, where natural informational asymmetries are most likely to be present. Likewise, mutual funds with a higher concentration of assets by industry outperform diversified funds (Kacperczyk, Sialm and Zheng, 2004). In addition, if investors are earning these excess returns because of their superior information processing skill, then investors who have earned high returns in the past should earn high returns in the future. This prediction is confirmed by Coval, Hirshleifer and Shumway (2002). Finally, if asymmetric information exists in the market, then investors who learn from prices should outperform investors who buy and hold a market index. Using CRSP data (1927-2000), Biais, Bossaerts and Spatt (2004) show that price-contingent strategies generate annual returns (Sharpe ratios) that are 3% (16.5%) higher than the indexing strategy. These results highlight the quantitative importance of asymmetric information.

Many theories in economics and finance have predictions that depend crucially on what information agents have. But, this information is usually treated like an endowment. By asking

[1] Barber and Odean (2001) cite this estimate from Inna Okounkova at Scrudder Kemper. Downloading daily open, high, low, close and volume data for 10,000 stocks over a period of 5 years amounts to 63 million bits of information.

what information rational agents would want to acquire, predictions contingent on information sets can be turned into more general predictions. This paper provides a tractable framework and set of tools for analyzing optimal information choices and incorporating those choices into commonly-used models of portfolio composition and asset pricing.

I. Setup

This is a static model which we break up into 3-periods. In period 1, the investor chooses the distribution from which to draw signals about the payoff of the assets. The choice of signal distributions is constrained by information capacity, which bounds the total informativeness of the signals an investor can observe, and by principal components analysis, which limits the linear combinations of signals the investor can choose and keeps the problem tractable. In period 2, the investor observes signals from the chosen distribution and then chooses what assets to purchase. In period 3, he receives the asset payoffs and realizes his utility. Signal choices and portfolio choices in this setting are circular: What an agent wants to learn depends on what he thinks he will invest in and what he wants to invest in depends on what he has learned. To ensure that beliefs and actions are consistent, we use backwards induction. We first solve the period 2 portfolio problem for arbitrary beliefs. Then, we substitute the solution to that problem in to the period 1 information optimization problem.

The vector of unknown asset payoffs $f \sim \mathcal{N}(\mu, \Sigma)$ is what the investor will devote capacity to learning about. The investor's objective is to maximize his expected exponential utility of period-3 profits.

$$U = -E_1\{exp(-\rho q'(f - pr))\}$$

where ρ is risk aversion, r is the risk-free return and q and p are Nx1 vectors of the number of shares the investor chooses to hold and the asset prices. Following Admati (1985), the excess return on asset i is $f_i - rp_i$.

Period-2 investment problem Let $\hat{\mu}$ and $\widehat{\Sigma}$ be the mean and variance of payoffs, conditional on all information known to the investor in period 2. The choice of optimal portfolio $q^{\star}$, given that $f \sim \mathcal{N}(\hat{\mu}, \widehat{\Sigma})$, is standard:

$$q^{\star} = \frac{1}{\rho}\widehat{\Sigma}^{-1}(\hat{\mu} - pr). \tag{1}$$

The model does not rule out short sales: $q^{\star} < 0$ when $\hat{\mu} - pr < 0$.

In order to substitute this solution into the period-1 learning problem, we need to know the expected utility that results from having beliefs $\hat{\mu}$, $\widehat{\Sigma}$ and investing optimally. Then, by the law of iterated expectations, we can state the period-1 problem as choosing a signal distribution to maximize the period-1 expectation of this period-2 expected utility.

$$-E_1\left[E_2\left[\exp\left(-\rho q^{\star\prime}(f-pr)\right)\right]\right] = -E_1\left[exp\left[-\frac{1}{2}(\hat{\mu}-pr)'\widehat{\Sigma}^{-1}(\hat{\mu}-pr)\right]\right] \tag{2}$$

The right side is a moment generating function of a quadratic function of the normal random variable $(\hat{\mu}-pr)$. Hence, the right side of equation 2 can be expressed as

$$-|I+\widehat{\Sigma}^{-1}Var[\hat{\mu}-pr]|^{-1/2}\exp\left\{-\frac{1}{2}E[\hat{\mu}-pr]'\left[I+(Var[\hat{\mu}-pr]^{-1}\widehat{\Sigma}+I)^{-1}\right]\widehat{\Sigma}^{-1}E[\hat{\mu}-pr]\right\} \tag{3}$$

(see Mathai and Provost, 1992) where $|X|$ denotes the determinant of matrix X.

Period-1 learning problem At time 1, the investor chooses how to allocate his information capacity. When evaluating information, it is important to recognize that a signal about one asset's payoff is also informative about all the assets that are correlated with that asset. Thus the relevant choice is not what assets to learn about, but what orthogonal risk factors.

Our investor chooses how much to learn about each principal component of asset payoffs. He does this by choosing a distribution from which he will draw a signal about asset payoffs. Since principal components are the eigenvectors of a variance(-covariance) matrix, signals about principal components have variance (Σ_η) with the same principal components (eigenvectors) as Σ. At time 2, the investor will combine his signal $\eta \sim N(f,\Sigma_\eta)$ and his prior belief $\mu \sim N(f,\Sigma)$, using Bayes' law. His posterior belief about the asset payoff f has a mean

$$\hat{\mu} \equiv E[f|\mu,\eta] = \left(\Sigma^{-1}+\Sigma_\eta^{-1}\right)^{-1}\left(\Sigma^{-1}\mu+\Sigma_\eta^{-1}\eta\right) \tag{4}$$

and a variance that is a harmonic mean of the prior and signal variances:

$$\hat{\Sigma} \equiv V[f|\mu,\eta] = \left(\Sigma^{-1}+\Sigma_\eta^{-1}\right)^{-1}. \tag{5}$$

These are the conditional mean and variance that agents use to form their portfolios in period 2. Transforming equation (5) reveals that if Σ and Σ_η share the same eigenvectors, then $\hat{\Sigma}$ must share the same eigenvectors too. Let Ω denote the set of posterior beliefs $\hat{\Sigma}$ with the same eigenvectors as Σ.

Writing down a choice problem over signal distributions is complicated by the fact that not learning about an asset is equivalent to getting a signal with infinite variance. Since every signal variance has a unique posterior belief variance associated with it, we can avoid the problem of infinite-valued choice variables by optimizing over posterior belief variance $\widehat{\Sigma}$ directly.

Taking prices as given, the period-1 investor forms expectations about period-2 expected return: $(\hat{\mu} - pr) \sim N(\mu - pr, \Sigma - \widehat{\Sigma})$. Substituting this mean and variance into equation (3) and taking logs yields the following period 1 optimization problem:

$$\max_{\widehat{\Sigma}\,\epsilon\,\Omega} \; \log(|\Sigma|) - \log(|\widehat{\Sigma}|) + (\mu - pr)'(2\widehat{\Sigma}^{-1} - \Sigma^{-1})(\mu - pr). \tag{6}$$

There are 2 constraints governing how the investor can choose his signals. The first constraint governs the total capacity the investor is allowed to use to transmit information. Our measure of information *capacity* is the standard measure in information theory: the *reduction in entropy*. The entropy of a random variable x with probability density function $p(x)$ is $-E[\log(p(x))]$. If this log is in base 2, then entropy is the number of bits, the length of the binary string, required to describe the variable. For an n-dimensional multivariate normal, with variance-covariance matrix V, entropy is $\frac{1}{2}\log\left((2\pi e)^n|V|\right)$. Like variance, entropy is a measure of uncertainty about a variable. It is a stock; capacity is its flow.

Following Sims (2003), we bound the entropy of posterior beliefs, relative to prior beliefs. The more information a signal contains, the more the posterior variance falls below the prior variance, and the more information capacity is required to transmit the signal. Capacity K is the maximum amount by which entropy can be reduced; for normal variables, it is one-half the difference between the logs of the determinants of the prior and posterior variances.[2]

$$\frac{1}{2}\left[\log(|\Sigma|) - \log(|\widehat{\Sigma}|)\right] \leq K \tag{7}$$

Another way to interpret the capacity constraint is as a bound on the Kullback-Leibler distance between prior and posterior beliefs. In statistics, this distance is used as a measure of how difficult it is to distinguish one distribution from another. The capacity constraint bounds the reduction in uncertainty of payoffs due to the knowledge of the signal η.[3]

The second constraint is that the variance-covariance matrix of the signals must be positive

[2]The determinant of a variance-covariance matrix is also called the generalized variance of the multi-variate process.

[3]To see the role of the signal, the capacity constraint can be restated as a bound on the precision Σ_η^{-1} of signals η: $1/2\log\left(|\Sigma_\eta^{-1}\Sigma + I|\right) \leq K$.

semi-definite. This implies the following restriction on $\widehat{\Sigma}$:

$$\Sigma - \widehat{\Sigma} \quad \text{positive semi-definite} \tag{8}$$

Without this constraint, the investor could choose to increase entropy in one variable so that he could decrease entropy further in other variables without violating the capacity constraint.

The sequence of events is summarized in figure I.

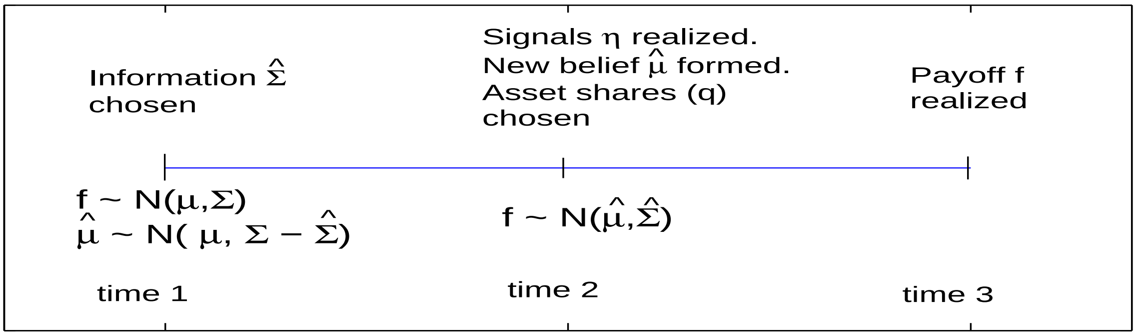

Figure 1. Sequence of events in partial equilibrium model

We have assumed in this formulation of the problem that the investor will choose to draw his signals from a normal distribution. Normal distributions have the property that they maximize the entropy of a variable, over all distributions with the same variance. So, if the objective is to minimize the variance of posterior beliefs, subject to the constraint that the entropy of those beliefs cannot be reduced more than a certain amount, then choosing normally distributed signals is optimal.[4]

A solution to the investor's problem is a choice of $\widehat{\Sigma}$ that maximizes (6) subject to (7) and (8), and portfolio positions that satisfy (1).

II. Partial Equilibrium Results

A. Independent assets

To gain intuition, it is helpful to first consider a simple case with N assets whose payoff variance-covariance matrix Σ is diagonal. Choosing signals with the same principal components as asset payoffs implies that signals are independent as well. The next section will generalize the problem to allow for correlated assets and signals.

[4]For more on the optimality of normal signals, see Cover and Thomas (1991), chapter 10.

There is free disposal of information in this model; an investor could always waste capacity. Therefore, we can assume that the information capacity constraint always binds and substitute out the first term of the objective function. Eliminating constant terms, $2K$ and $(\mu - pr)'\Sigma^{-1}(\mu - pr)$ and dividing by 1/2, we can rewrite the problem as

$$\max_{\{\widehat{\Sigma}_{ii}\}} \sum_{i=1}^{N} (\mu_i - p_i r)^2 \widehat{\Sigma}_{ii}^{-1} \tag{9}$$

$$s.t. \quad \prod_{i=1}^{N} \widehat{\Sigma}_{ii} = e^{-2K} \prod_{i=1}^{N} \Sigma_{ii}$$

$$\widehat{\Sigma}_{ii} \leq \Sigma_{ii} \quad \forall\, i$$

The first constraint results from (7) and the fact that the determinant of a diagonal matrix is the product of the diagonals. The second constraint uses (8) and the fact that a diagonal matrix is positive semi-definite if and only if all its elements are non-negative. Due to the linearity of the objective in $\widehat{\Sigma}^{-1}$, the maximum is a corner solution.

Proposition 1. *The optimal information portfolio with N independent assets uses all capacity to learn about one asset, the asset with the highest squared Sharpe ratio $(\mu_i - p_i r)^2 \Sigma_{ii}^{-1}$.*

Proof is in appendix A. Consider the problem of sequentially assigning units of capacity that can reduce the variance of an asset's payoff from Σ_{ii} to $\widehat{\Sigma}_{ii} = (1 - \epsilon)\Sigma_{ii}$. The greatest utility gain is obtained by assigning the first unit of capacity to the asset with the highest value of $(\mu_i - p_i r)^2 \Sigma_{ii}^{-1}$. The value of assigning the next unit of capacity to asset i is then even greater: $(\mu_i - p_i r)^2 \widehat{\Sigma}_{ii}^{-1} > (\mu_i - p_i r)^2 \Sigma_{ii}^{-1}$. The value of assigning each subsequent unit of capacity to i rises higher and higher, while the value of assigning capacity to all other assets remains the same. Therefore, the optimal choice of posterior variance is $\widehat{\Sigma}_{ii} = e^{-2K}\Sigma_{ii}$, and $\widehat{\Sigma}_{jj} = \Sigma_{jj}$ for all $j \neq i$.

The value of learning about an asset is indexed by its squared Sharpe ratio $(\mu_i - p_i r)^2 \Sigma_{ii}^{-1}$. Another way to express the same quantity is as the product of two components: $(\mu_i - p_i r)$ and $(\mu_i - p_i r)/\Sigma_{ii}$, which is $\rho E[q_i]$ for an investor who has zero capacity. An investor wants to learn about an asset that has (i) high expected excess returns $(\mu_i - p_i r)$, and (ii) features prominently in his portfolio. The fact that an investor wants to invest all capacity in one asset comes from the anticipation of his future portfolio position $E[q]$. The more shares of an asset he expects to hold, the more valuable information about those shares is, and the higher the index value he assigns to learning about the asset. But, as he learns more about the asset, the amount he expects to hold $E[q_i] = (\mu_i - p_i r)/\widehat{\Sigma}_{ii}$ rises. As he learns, devoting capacity to the same asset becomes more and more valuable. This is the increasing return to learning.

How does this learning strategy affect the investor's portfolio? For the assets that the investor does not learn about, the number of shares does not change. For the asset he does learn about, the expected number of shares increases by $E[q^{learn}] = \frac{1}{\rho\Sigma_{ii}}(\mu - pr)(e^{2K} - 1)$. Call the portfolio of shares that the investor would hold if he had zero-capacity and could not learn, q^{div}. This is the benchmark portfolio predicted by the standard CARA-normal model. Since it contains no signals, it is not random: $E[q^{div}] = q^{div}$. The portfolio of an investor with positive capacity is the sum of q^{div} and the component due to learning, q^{learn}.

Proposition 2. *As long as there is at least one asset for which* $(\mu - pr) \neq 0$*, then when capacity rises, the expected fraction of the optimal portfolio consisting of fully-diversified assets* $(|q^{div}|/(|q^{div}| + |E[q^{learn}]|)$ *falls.*

Proof: As capacity (K) increases from zero, the zero-capacity portfolio q^{div} is, by definition, unchanged. As long as there is an asset s.t. $(\mu_i - p_i r) \neq 0$, then proposition 1 tells us that an investor will learn about an asset $i^\star$ s.t. $(\mu_{i^\star} - p_{i^\star} r) \neq 0$. The only quantity that changes in K is the expected amount of asset $i^\star$ held due to learning: $|E[q^{learn}_{i^\star}]| = \frac{1}{\rho\Sigma_{i^\star i^\star}}|\mu_{i^\star} - p_{i^\star} r|(e^{2K} - 1)$. Since $\mu_{i^\star} - p_{i^\star} r \neq 0$, $|E[q^{learn}_{i^\star}]|$ is strictly increasing in K. $\square$

Only expected portfolio holdings can be predicted. Since actual signal realizations and therefore posterior beliefs $\hat{\mu}$ are random variables, the true portfolio chosen in period 2 could be either larger or smaller in absolute value, than it would have been without the signal. But, for any given belief about payoffs $\hat{\mu}_i$, having more capacity to reduce the variance of that belief $\widehat{\Sigma}_{ii}$, makes the investor take a larger position in the asset $|q_i|$.

This result can be easily restated in terms of the more familiar value-weighted fraction of shares in the learning and diversified funds. As long as the expected excess return and price for the learning asset i are positive, then the expected value-weighted fraction of shares held in the diversified portfolio falls. This is the sense in which learning and diversification trade off.

Corollary 3. *An investor who optimally chooses a less diversified portfolio earns a higher expected return than an investor who chooses a more diversified portfolio.*

Proof in appendix B. Proposition 2 tells us that investors who have high information capacities K choose highly under-diversified portfolios. Such investors makes more informed investment choices and obtain a higher expected profit. The reason is that these investors achieve a higher correlation between asset payoffs and portfolio shares. This prediction is corroborated by the findings of Ivokovic, Sialm, and Weisbenner (2004) and Kacperczyk, Sialm, and Zheng (2004), that under-diversified portfolios significantly outperform diversified ones.

Data Example with Independent Assets

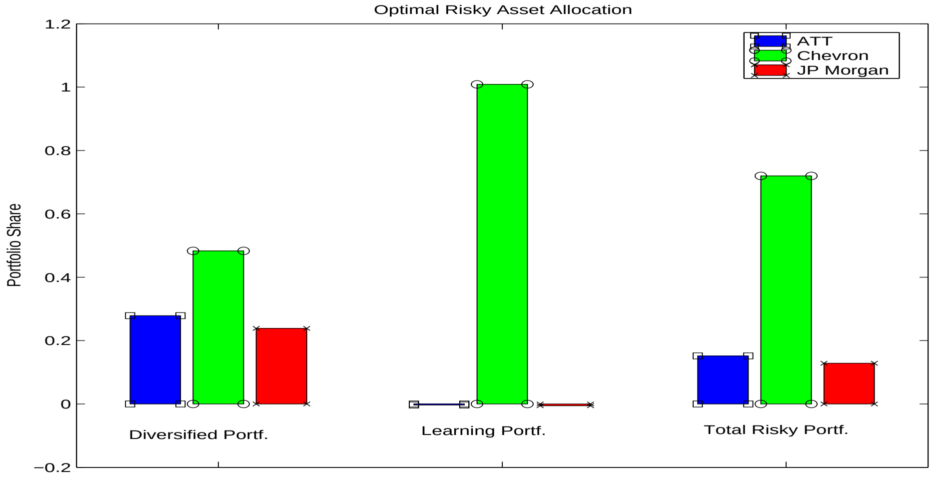

Figure 2. Under-Diversification and the Increasing Returns to Learning: Uncorrelated Assets.

We illustrate the portfolio composition with a numerical example. Figure 2 illustrates the case of uncorrelated assets for three assets from the S&P 500 index.[5] The monthly excess returns on AT&T, Chevron, and JP Morgan were nearly orthogonal in the sample period. Chevron had the highest Sharpe ratio (.58 annualized). When faced with the mean excess returns and the covariance matrix of returns of three assets, an investor with zero information capacity would hold an optimally diversified portfolio, consisting of 28% AT&T, 48% Chevron, and 24% JP Morgan ('diversified portfolio'). When given some information capacity, the investor specializes in learning about Chevron. ($K = .5$ here, which allows the investor to reduce the standard deviation of one asset by 39%.) The 'learning fund' is fully invested in Chevron. As a result, the total portfolio is under-diversified: 15% AT&T, 72% Chevron, and 13% JP Morgan.

B. Correlated assets

When assets are correlated, signals about individual asset payoffs are no longer principal components. Instead, principal components are linear combinations of asset payoffs with weights on each asset given by an eigenvector of Σ. Rather than choose how to reduce the risk of independent assets, investors choose how to reduce the variance of these independent risk factors. The factors could represent risks such as business cycle risk, pharmaceutical industry risk, or idiosyncratic risk. The variance of each risk factor is given by its eigenvalue (Λ_{ii}). After transforming assets into

[5]Monthly return data runs from November 1986 and December 2003 (206 observations). Excess returns are constructed by subtracting the return on a 1-month T-bill.

independent risk factors, the results for independent assets can be restated for the correlated assets case.

When an investor learns about principal components, his posterior belief variance $\widehat{\Sigma}$ has the same eigenvectors (Γ) as Σ. Therefore, the investor's choice is over the diagonal eigenvalue matrix $\hat{\Lambda}$, where $\widehat{\Sigma} = \Gamma\hat{\Lambda}\Gamma'$. Imposing that the capacity constraint (7) holds with equality, substituting it into the objective (6) and applying the definition of $\widehat{\Sigma}$, the optimum can be found by maximizing:

$$\max_{\hat{\Lambda}}(\mu - pr)'\Gamma^{-1'}\hat{\Lambda}^{-1}\Gamma^{-1}(\mu - pr). \tag{10}$$

$$s.t. \quad \prod_{i=1}^{N}\hat{\Lambda}_{ii} = e^{-2K}|\Sigma| \tag{11}$$

$$\Sigma - \hat{\Sigma} \quad \text{positive semi-definite} \tag{12}$$

where the first constraint follows from a determinant being the product of eigenvalues.

Proposition 4. *The optimal information portfolio with N correlated assets uses all capacity to learn about one linear combination of asset payoffs. The linear combination coefficients are given by the eigenvector Γ_i, with the highest factor squared Sharpe ratio: $(\Gamma_i'(\mu - pr))^2\,\Lambda_{ii}^{-1}$.*

Proof is in appendix C. There are two components of this result. The first component tells us how the investor initially ranks learning about each risk factor Γ_i. The second tells us that he wants to specialize completely in whatever risk factor he wants to learn about. What direction an investor decides to learn in is determined by the magnitude of the expected return on the risk factor $\Gamma_i'(\mu - pr)$ and by ρ times the expected holding of that risk factor: $\rho\Gamma_i'E[q]$. The fact that the investor wants to devote all capacity to learning about one risk factor comes from increasing returns. As the investor learns more about Γ_i, the investor expects to hold more of that risk factor: $\Gamma_i'E[q]$ grows. As he expects to hold more of the risk factor, the value of learning more about it rises.

What does this result mean for portfolio allocation? The investor will hold shares of each asset given by $\frac{1}{\rho}(\Gamma\hat{\Lambda}\Gamma')^{-1}(\hat{\mu}-pr)$. Again, this portfolio can be decomposed into the diversified benchmark portfolio that an investor with no capacity would hold $q^{div} = \frac{1}{\rho}(\Gamma\Lambda\Gamma')^{-1}(\hat{\mu} - pr)$, and the number of extra shares of assets that will be held due to learning,

$$q^{learn} = \frac{e^{2K} - 1}{\rho\Lambda_{ii}}\Gamma^{-1'}(i,:)(\hat{\mu} - pr)$$

where i is the factor the agent optimally learns about. This learning portfolio puts more weight on assets in proportion to how correlated they are with the risk factor that the investor is learning about. Since the 'learning' assets are highly correlated with a common risk factor, they are also highly correlated with each other. As K grows, the expected weight on this highly-correlated component of the portfolio rises exponentially. As learning increases, diversification falls.

Data Example with Correlated Assets

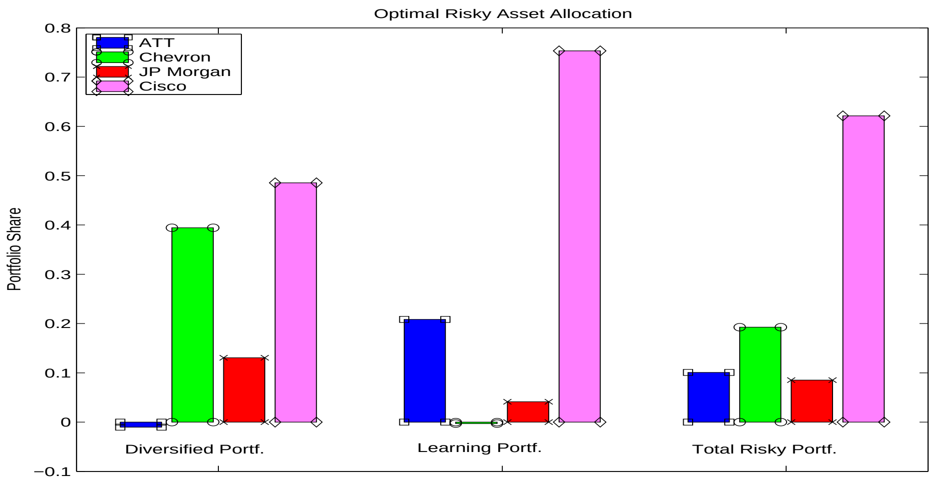

Figure 3. Under-Diversification and the Increasing Returns to Learning: Correlated Assets.

Figure 3 illustrates the case of correlated assets. It adds to the three uncorrelated assets described above a fourth asset, Cisco. Cisco has a low correlation with Chevron (-.008) and with JP Morgan (.068), but a high correlation with AT&T (.296). Cisco has a much higher Sharpe ratio than the other three firms. When offered these four assets, an investor with zero information capacity would hold an optimally diversified portfolio, consisting of -1% AT&T, 39% Chevron, 13% JP Morgan, and 49% Cisco ('diversification fund'). When given some information capacity (K is still .5 here), the investor learns about Cisco, the most valuable asset to learn about, but also about AT&T. The reason is that both Cisco and AT&T load positively on the most valuable risk factor (correlations .96 and .27 respectively). The 'learning fund' is invested for 75% in Cisco and 21% in AT&T. As a result of the specialization in learning, the total portfolio is under-diversified: 10% AT&T, 19% Chevron, 9% JP Morgan, and 62% Cisco. The new optimal portfolio has a variance (conditioning on past public information) that is 25% higher than the diversified portfolio variance; it is under-diversified.

C. Un-learnable risk

In the previous results, investors never diversify their information because learning substitutes for diversification. As learning increases and risk falls, the value of diversification falls as well. With un-learnable risk, there is some risk that learning cannot eliminate, but diversification can. This risk revives the benefits to diversification and makes high-capacity investors learn about multiple risk factors.

Un-learnable risk increases information portfolio diversification because it makes the returns to learning bounded. When all risk is learnable and capacity approaches infinity, the payoff variance of some portfolio approaches infinity, an arbitrage arises, and profit becomes infinite. Un-learnable risk imposes a finite, maximum benefit to learning. To reduce an asset's learnable payoff variance to near zero costs an unbounded amount of information capacity and yields only a finite benefit. Therefore, learning an arbitrarily large amount about a single asset is never optimal.

To examine the effects of introducing un-learnable risk, consider the following model. The investor's preferences, the sequence of events, and the optimal period-2 portfolio remains unchanged. The period-1 choice of signal distributions is constrained by the fact that of the total variance in the prior beliefs Σ, $\alpha\Sigma$ is un-learnable, and only $(1-\alpha)\Sigma$ can be learned $(0 < \alpha < 1)$.[6] The new period-1 problem is to maximize (6) subject to a constraints on the reduction in entropy of the *learnable* component of asset payoffs. This constraint is formulated so that eliminating all learnable risk (reducing $\widehat{\Sigma}$ to $\alpha\Sigma$) requires infinite capacity. When $\widehat{\Sigma} = \Sigma$, the investor is not learning anything, and no capacity is required.

$$\frac{1}{2}\log(|\Sigma - \alpha\Sigma|) - \log(|\widehat{\Sigma} - \alpha\Sigma|) \leq K \tag{13}$$

$$\Sigma - \widehat{\Sigma} \quad \text{positive} \quad \text{semi-definite.} \tag{14}$$

As in the case with learnable risk, we solve the problem by considering separately the eigenvalues $\hat{\Lambda}$ and eigenvectors Γ of the posterior variance matrix $\widehat{\Sigma}$. Following the steps outlined in the proof of proposition 4, we obtain a first-order condition with respect to $\hat{\Lambda}_i$. It describes an interior solution to the maximization problem.

$$\xi \frac{1}{\hat{\Lambda}_i - \alpha\Lambda_i} - \phi_i = \frac{1}{\hat{\Lambda}_i} + 2\frac{(\Gamma_i(\mu - pr))^2}{\hat{\Lambda}_i^2} \tag{15}$$

where ξ is the Lagrange multiplier on (13), and ϕ_i is the Lagrange multiplier on (14). Taking a

[6]For every result, except proposition 8, α can be a matrix where every element is $0 < \alpha_{ij} < 1$.

second derivative confirms that a solution to (15) is a maximum.[7]

Proposition 5. *When there is un-learnable risk, the number of risk factors that the investor learns about is an increasing step function of K*

Corollary 6. *When there is un-learnable risk and asset payoffs are independent, the number of assets held in the 'learning fund,' q^{learn}, is an increasing step function of capacity K*

Proofs are in appendix D.

The reason for learning about additional assets can be seen by examining the benefit and cost of learning for an asset where $\hat{\Lambda}_i < \Lambda_i$. The marginal benefit is $\frac{1}{\hat{\Lambda}_i} + 2\frac{(\Gamma_i(\mu - pr))^2}{\hat{\Lambda}_i^2}$. As the investor learns more and $\hat{\Lambda}_i$ falls, this benefit increases. Increasing returns to scale in learning are still present. However, in the limit as $\hat{\Lambda}_i$ approaches $\alpha\Lambda_i$, these benefits are finite. The marginal cost to learning is $\xi\frac{1}{\hat{\Lambda}_i - \alpha\Lambda_i}$. As the investor gets closer to learning all the learnable risk, this cost approaches infinity. Therefore, there is some finite cutoff level of $\hat{\Lambda}_i$ such that when the investor reaches this level of learning for asset i, he begins to allocate some capacity to another risk factor. In the case of independent assets, allocating capacity to another risk factor means learning about another asset. This means that another asset is included in the investor's learning fund.

Proposition 7. *When there is un-learnable risk and there is some asset i with non-zero expected excess return $(\mu_i - p_i r) \neq 0$, then, as capacity rises, the fraction of the expected optimal portfolio consisting of fully-diversified assets $(|q^{div}|/(|q^{div}| + |E[q^{learn}]|)$ falls.*

Proof is in appendix E.

Just as in the case where all risk is learnable, when the investor learns more about an asset, he expects to hold a larger position in that asset. Since the zero-capacity portfolio q^{div} does not change as capacity increases and more shares are held in the learning portfolio, the fraction of the expected portfolio that is diversified falls.

Proposition 8. *When there is un-learnable risk and capacity is infinite, the expected learning portfolio is fully diversified: $\lim_{k\to\infty} E[q^{learn}] = (1/\alpha - 1)q^{div}$.*

Proof: An agent with an infinite capacity would eliminate all learnable risk, setting $\hat{\Lambda} = \alpha\Lambda$, which implies $\widehat{\Sigma} = \alpha\Sigma$. In this limit, the learning fund is $E[q^{learn}] = \frac{1}{\rho}(1/\alpha - 1)\Sigma^{-1}(\mu - pr)$, a scaled-up copy of the diversified mutual fund. $\square$

Putting the results together tells us that as capacity increases, diversification falls, and then rises again. An agent with zero capacity holds only the diversified fund. An agent with infinite capacity

[7] There may be multiple solutions to (15). However, the smallest Λ_i that solves (15) is always a maximum.

holds a perfectly diversified learning fund. In between the two perfectly diversified extremes, the investor with positive, finite capacity to learn is optimally under-diversified.

III. Equilibrium Information and Investment Choices

In general equilibrium, 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. When all risk is learnable, the nature of the solution to the individuals problem does not change. After accounting for the actions that other agents 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. We begin by describing modifications to the setup.

A. Equilibrium Model

There is now a continuum of investors, indexed by $j \in [0, 1]$. Preferences, payoffs, and timing are identical to the model described in section I. The risk-free rate is still fixed. There are two additional assumptions required to model agents' strategic interactions. 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: $x \sim N(0, \sigma_x^2 I)$. The reason for having a risky asset supply is to create some noise in the price level that prevents investors from being able to perfectly infer the private information of others. Without this noise, there would be no private information, and no incentive to learn. We interpret this extra source of randomness in prices as due to liquidity or life-cycle needs of traders.[8] It could also represent errors that agents make when trying to invert prices.

Second, when investors draw their noisy asset payoff signals from the distributions that they have chosen, we assume that these draws are *independent*. This assumption corresponds to a decentralized view of information transmission. The truth is being sent to all investors. But each observes that truth after it has been transmitted through his own limited-capacity channel, which adds independent noise to the signal. The independent noise can also be thought of as an error that each investor adds when he interprets his information. We believe that this is the relevant physical constraint that humans are facing when trying to process financial information (Sims 2003). An alternative view of information transmission is that it is a centralized process. A news agency gets

[8] See Biais, Bossaerts and Spatt 2003 for an interpretation in terms of risky non-tradeable endowments.

a noisy signal of the truth and transmits that signal through noiseless channels all of us. We revisit the idea of centralized information processing in the conclusion.

Asset prices p are determined by market clearing. Price is set such that the sum of investors' demands for each asset equals its supply. In vector notation:

$$\int_0^1 \widehat{\Sigma}_j^{-1}(\hat{\mu}_j - pr)dj = \bar{x} + x \tag{16}$$

B. 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 must aggregate to form their expectation of the assets' payoffs: their prior beliefs (common across agents), their signals (draws from distributions chosen in period 1), and the equilibrium asset price.

Proposition 9. *Asset prices are a linear function of the asset payoff and the unexpected component of asset supply.*

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

This price can be expressed as a function of the posterior mean and variance of the 'average' investor:

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

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 is in appendix F, along with the formulas for A, B and C.

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

$$\hat{\mu} \equiv E[f|\mu, \eta, p] = \left(\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1}\right)^{-1}\left(\Sigma^{-1}\mu + \Sigma_\eta^{-1}\eta + \Sigma_p^{-1}B^{-1}(rp - A)\right) \tag{17}$$

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

$$\hat{\Sigma} \equiv V[f|\mu, \eta, p] = \left(\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1}\right)^{-1}. \tag{18}$$

These are the conditional mean and variance that agents use to form their portfolios in period 2.

The portfolio is

$$q^\star = \frac{1}{\rho}\widehat{\Sigma}^{-1}(\hat{\mu} - pr).$$

Given a posterior belief about the asset's payoff and variance of that belief, we can compute the period 2 expected utility of the agent. It is the same expected utility as in the partial equilibrium problem (equation 2).

C. Individual's Information Capacity Allocation in Equilibrium

In period 1, the investor chooses a covariance matrix for his posterior beliefs $\widehat{\Sigma}$. Using the law of iterated expectations, we can replace the objective function with its time-2 expectation. The solution proceeds as in the partial equilibrium problem except that asset prices are now random variables with known mean and variance.

As before, we will use a moment generating function to solve for expected utility in period 1. The time-2 expected excess return $(\hat{\mu} - pr)$ is a normal variable at time 1, with mean $(I-B)\mu - A$ and variance $V_{ER} \equiv \Sigma - \widehat{\Sigma} + B\Sigma B' + CC'\sigma_x^2 - 2B\Sigma$. Using the moment generating function for the quadratic form of $(\hat{\mu} - pr)$ (equation 3), we can solve for period-1 expected utility. The period-1 optimization problem of an investor is

$$\max_{\widehat{\Sigma}} \log\left(|V_{ER}|\right) - \log\left(|\widehat{\Sigma}|\right) + ((I-B)\mu - A)'(2\widehat{\Sigma}^{-1} - V_{ER}^{-1})((I-B)\mu - A) \tag{19}$$

Just as in partial equilibrium, choice of the covariance matrix of the posterior belief $\widehat{\Sigma}$ is subject to two constraints. These constraints are different from the ones in section I: They are still constraints on the distance between the posterior belief variance $\widehat{\Sigma}$ and a reference variance, but now the reference variance is $\widetilde{\Sigma}$, instead of the prior belief variance Σ. $\widetilde{\Sigma} = V[f|\mu, p] = (\Sigma^{-1} + \Sigma_p^{-1})^{-1}$ is what the conditional variance of asset payoffs would be if the agent observed no private signals, but only learned through the price level. As in the partial equilibrium model, the reference variance is the conditional variance of asset payoffs that an investor with zero capacity faces.

The first constraint is that the information the investor sees cannot reduce entropy by more than his capacity K, (the analog to equation 7):

$$\frac{1}{2}\left[\log(|\widetilde{\Sigma}|) - \log(|\widehat{\Sigma}|)\right] \leq K \tag{20}$$

The second constraint is the equivalent of (8). It prevents the investor from acquiring negative information that would make him more uncertain.

$$\widetilde{\Sigma} - \widehat{\Sigma} \quad \text{positive semi-definite} \tag{21}$$

The sequence of events is summarized in figure 4.

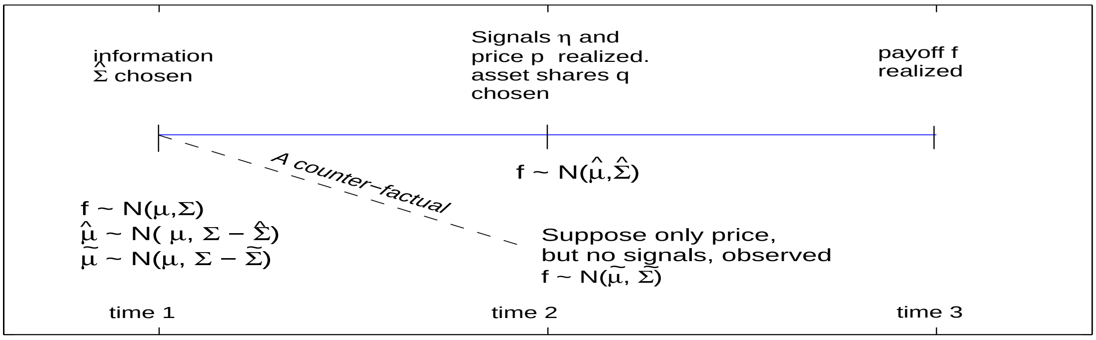

Figure 4. Sequence of events in general equilibrium model

As in partial equilibrium, learning about principal components of asset payoffs implies that prior and posterior variances have the same eigenvectors. Since $\widetilde{\Sigma}$ is what the zero-capacity investor would know about asset payoffs, learning about its principal components implies that $\widehat{\Sigma}$ has the same eigenvectors as $\widetilde{\Sigma}$. The proof of proposition 10 shows that $\widetilde{\Sigma}$ has the same eigenvectors Γ as Σ. Thus, define the eigenvalue decomposition of the 'no-signal' variance matrix as $\widetilde{\Sigma} = \Gamma\widetilde{\Lambda}\Gamma'$.

Eliminating constant terms in (19), yields an objective function of the form $\max_{\widehat{\Sigma}}((I-B)\mu - A)'\widehat{\Sigma}^{-1}((I-B)\mu - A)$. This objective function is of the same form as the objective function in the partial equilibrium model (equation 10). The constraints are also of the same form.

Proposition 10. *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 $\widetilde{\Gamma}_i$ associated with the highest value of $(\Gamma_i' E[f - pr])^2 (\widetilde{\Lambda}_{ii})^{-1}$.*

Proof is in appendix G. As before, the most valuable risk factor to learn about is the risk factor with (i) a high expected return $\Gamma_i' E[f - pr]$ and (ii) a large expected portfolio share $\widetilde{\Gamma}_i' E[q]$. In the special case of independent assets, this result tells an investor to devote all his capacity to learning about the asset with the highest expected squared Sharpe ratio. The Sharpe ratio's standard deviation of returns is conditional on the investors prior beliefs and what he will learn from observing asset prices.

The formulation of the entropy constraint in this section, using $\widetilde{\Sigma}$ as a benchmark variance, implicitly assumes that asset prices are freely observable with infinite precision. While it is certainly less time-consuming to find an asset's price than to find, read, and process a research report, it seems natural to assume that investors need to devote some processing capacity to price discovery. However, if we imposed the same form of learning constraints on price discovery, investors would never know exactly what the price of the assets they are buying is. This is difficult to analyze because it would introduce a new source of risk. Furthermore, investors who are holding many assets in a diversified mutual fund do not treat their ignorance about the underlying asset prices as risk. They can know the price of the mutual fund without learning about all its component prices.

D. Aggregate Information Portfolios and Asset Prices

The previous section characterized the optimal information and asset allocation for an individual investor. This section describes how these choices aggregate across investors.

Aggregate Information Allocation In equilibrium, ex-ante identical investors may learn about different risk factors and hold heterogenous portfolios, but they would get the same expected utility from learning about any of the risk factors that the economy learns about. Investors will choose to specialize in different risk factors because of strategic substitutability. When other investors learn more about a set of assets, the expected prices of those assets rise.

Proposition 11. *The number of risk factors that the economy learns about is weakly increasing in economies' aggregate capacity K.*

Proof in appendix H. How much investors learn about an asset is summarized by the aggregate precision of beliefs $\widehat{\Sigma}_a^{-1}$. Manipulating the price in proposition 9 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:

$$E[f - pr] = \rho \widehat{\Sigma}_a \bar{x}. \tag{22}$$

A reduction in expected return makes assets less valuable to learn about. Proposition 10 tells us that the value of learning about a risk factor is given by the expected return on the factor, times the variance of the factor, conditional on prices: $(\Gamma_i' E[f - pr])^2 (\widetilde{\Lambda}_{ii})^{-1} = \rho \Gamma' \widehat{\Sigma}_a \bar{x} (\widetilde{\Lambda}_{ii})^{-1}$. When more agents learn about a factor, the expected return on the assets that load heavily on that factor falls. This makes that factor less desirable to learn about.

The equilibrium information allocations follow a cutoff rule. Consider a thought experiment where all investors have the same capacity and we let them sequentially choose how to allocate

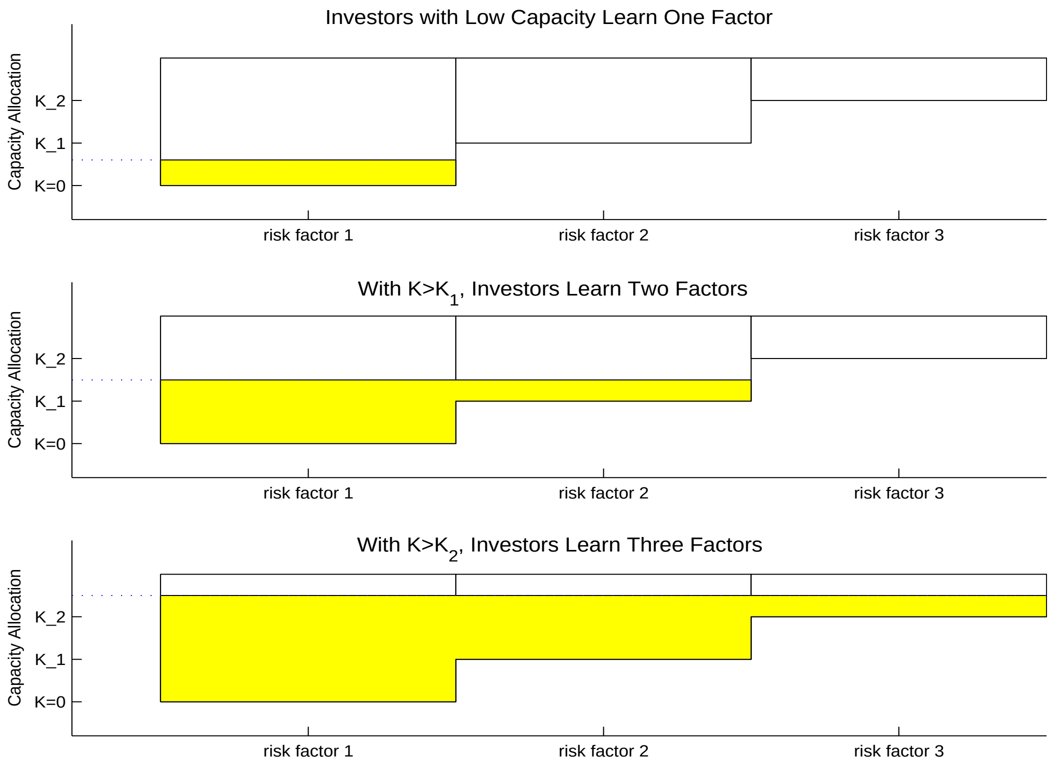

Figure 5. Aggregate allocation of information capacity for low, medium, and high-capacity investors.

it. The first investor learns about the risk factor that is most valuable when no other learning takes place. This is the same risk factor i that our partial equilibrium investor would learn about. Subsequent investors will continue to allocate their capacity to factor i until its value of $(\Gamma_i' E[f - pr])^2 (\widetilde{\Lambda}_{ii})^{-1} = (\Gamma_l' E[f - pr])^2 (\widetilde{\Lambda}_{ll})^{-1}$ for some other risk factor l. This cutoff is when capacity $K = K_1$ in figure 5. Then, some investors will find it beneficial to learn about risk factor l. The proportions of investors that learn about i and about l is such that all investors remain indifferent. Subsequent investors will continue to allocate capacity to these two risk factors, until all investors become indifferent between learning about i, l and some third risk factor (where $K = K_2$ in figure 5). This process continues until all capacity is allocated. This type of result is referred to as 'water-filling' in the information theory literature.

Asset Holdings in Equilibrium The cross-section of asset holdings is fully pinned down by the cross-section of information allocation. The mapping is just as described in proposition 2. Each investor holds a diversified portfolio, plus a learning portfolio that contains assets in proportion to the one risk factor he learns about.

Atomless Investors and Limits to Arbitrage We assumed that there is a continuum of atomless investors, who by definition, cannot impact asset prices. This turns out to matter for equilibrium learning strategies because it makes the returns to learning unbounded. An as investor learns more about an asset, he can take larger and larger positions in that asset to fully exploit what he has learned, without worrying about his information being revealed through the price level. In contrast, an investor that is large in the market will move the asset price level when he trades. If he tries to exploit very precise information by taking large asset positions, his impact on the market price will partially reveal what he knows. This diminishes the value of his information and re-introduces decreasing returns to learning about a single risk factor. In figure 5, the investor is filling a bin on his own. For example, his capacity may exceed cutoff K_1.

Similar to the case where some risk is not learnable (section II.C), giving investors some mass in the market will make them want to specialize for low levels of capacity, but broaden their learning to multiple factors as capacity increases. In order to analyze a setting where large capacity investors interact, we need to model investors who consider the effect of their own learning on the price level. This question is beyond the scope of the current paper. In the conclusion, we return to the idea of modelling large portfolio managers.

E. Cross-Section of Asset Returns

In this section we study the model's predictions for the cross-section of asset returns. We re-discover some old asset-pricing models.

An APT Representation of Asset Prices Our theory revives an old arbitrage-free pricing theory practice of using the principal components of the asset payoff matrix as priced risk factors (Ross 1976). We can rewrite the risk premium on an asset i as the sum of its loading on each principal component k times the equilibrium risk premium of that principal component:

$$E[f_i - rp_i] = \sum_{k=1}^{n} \Gamma_{ik} \left(\Gamma'_k E[f - rp] \right)$$

The equilibrium risk premium of factor k can be rewritten, using equation (22) and the result that $\widehat{\Gamma} = \Gamma$, as:

$$\Gamma'_k E[f - rp] \;=\; \rho \hat{\Lambda}_{ak} \Gamma'_k \bar{x}. \tag{23}$$

The equilibrium risk premium depends on (i) the risk aversion of the economy ρ, (ii) the supply of the risk factor $\widetilde{\Gamma}'_k \bar{x}$, and most importantly (iii) on the weight $\hat{\Lambda}_{ak}$, the eigenvalues of aggregate variance matrix $\widehat{\Sigma}_a$. This weight measures how much the economy learns about risk factor k. A risk factor that the economy does not learn about has weight $\hat{\Lambda}_{ak} = \tilde{\Lambda}_k$. A risk factor that the economy learns about has a weight $\hat{\Lambda}_{ak} < \tilde{\Lambda}_k$. In other words, as more agents learn about risk factor k, $\hat{\Lambda}_{ak}$ decreases.

Our theory has sharp predictions for which risk factors are learned about in equilibrium. Their risk premia are lower. An asset that loads heavily on those risk factors has a low risk premium.

A CAPM Representation of Asset Prices The equilibrium asset prices and returns are equivalent to the prices and returns that would arise in a representative agent economy. That representative agent is endowed with the belief that payoffs f are normally distributed with mean $E_a[f]$ and covariance $\hat{\Sigma}_a$: the heterogeneously informed investors' arithmetic average mean and harmonic average covariance.[9] In our model with heterogenous information and partial information aggregation, a version of the Capital Asset Pricing Model holds.

Proposition 12. *If the market payoff is defined as* $f_m = \sum_{k=1}^{N}(\bar{x} + x_k)f_k$, *the market return is* $r_m = \frac{f_m}{\sum_{k=1}^{N}(\bar{x}+x_k)p_k}$, *and the return on* i *is* $r_i = \frac{f_i}{p_i}$, *then the equilibrium price of asset* i *can be expressed as*

$$p_i = \frac{1}{r}\left(E_a[f_i] - \rho Cov_a[f_i, f_m]\right). \tag{24}$$

The equilibrium return equals

$$E_a[r_i] - r = \frac{Cov_a[r_i, r_m]}{Var_a[r_m]}(E_a[r_m] - r) \equiv \beta_a^i(E_a[r_m] - r). \tag{25}$$

The proposition states that the equilibrium expected return on a security is proportional to its beta and to the market price of risk expressed in beta units. Whereas Admati (1985) characterizes the expected equilibrium price, averaged over realizations of random variables, this pricing function holds in each state of nature. The standard CAPM is based on the assumption that all agents agree on the distribution of payoffs and returns. With heterogenous-information, each investor has a different assessment of risk-return tradeoffs, depending on his signal realization and on the signal's distribution. Yet a CAPM re-emerges.

The information of the representative agent is in no single agent's information set. As such, equations (24) and (25) do not hold for any one agent in the model. In addition, the representative

[9] Consistent with Bayesian updating, the arithmetic mean $E_a[f_i] \equiv \int_0^1 \hat{\mu}_j dj$ is used to update the conditional expectations and the harmonic mean $\hat{\Sigma}_a$ is used to update conditional variances.

agent's beliefs and the random asset supply cannot be observed by an econometrician, a problem highlighted by Roll (1977). Our theory offers a solution to the problem of unobservable information sets: using observable prior information, it can predict what investors will learn.

IV. Institutional Portfolio Management

While the paper's original motivation was the composition of individual investor portfolios, the model also dictates optimal allocations of research and financial resources for institutions. Through the lens of our theory, we see a specialized fund, such as a hedge fund or 'alpha-fund,' as an optimally under-diversified component of an institution's portfolio. Their investment strategy is to hold assets along one risk dimension in order to exploit the increasing returns to learning. Such specialized funds can take riskier positions, as their information capacity, and therefore their ability to manage that risk, grows.

Optimal portfolio management is a long-standing issue in the mutual fund literature. The seminal paper by Treynor and Black (1973, henceforth TB) is motivated by the idea that security analysts can analyze only a limited number of stocks. It departs from the efficient markets hypothesis by assuming that securities can deviate from their equilibrium price. Individual portfolio managers can exploit mispricing to make abnormal returns. The security analyst estimates the alpha of a security k as $\alpha_k = r_k - r - \beta_i'(r^{div} - r) - \varepsilon_k$, where r^{div} represents diversified portfolio returns and ε_k is idiosyncratic risk, with variance $\sigma^2(\varepsilon_k)$. The optimal portfolio tilts away from the diversified one, towards securities with a high 'information ratio': $\alpha_k/\sigma^2(\varepsilon_k)$.[10]

TB and our paper are similar in that both recognize the fundamental trade-off between diversification and specialization. However, the theories differ along several dimensions. First, ours is an equilibrium pricing model. There is no mispricing. A TB regression in our model will produce αs that capture *public* information already impounded in prices. If a portfolio manager followed the TB strategy, and purchased stocks with a positive (public) information ratio, his stocks would have prices that were depressed by privately informed investors' bad news. However, there is another notion of α that does reflect a profit opportunity. Investors demand different risk premia for the same asset because they have an individual-specific α, arising from private information.

Second, while TB endow investors with a set of securities that they can analyze, we examine the choice of what to learn. As in our model, TB investors who learn about an asset's α want to

[10] With a theory of the optimal q^{learn}, we avoid a non-uniqueness problem of TB's portfolio decomposition. To understand the non-uniqueness, suppose that the optimal diversified portfolio contains shares of asset 1 and 2 in the ratio of 1 to 2. The market (asset supply) is 2 shares of each asset. The asset supply can be decomposed into one share of the diversified portfolio, plus one share of asset 1 in the learning portfolio, or alternatively into 2 shares of the diversified portfolio and two shares sold short of asset 2. We are grateful to Ned Elton for pointing out this issue.

take a large position in that asset. But the feedback mechanism, where taking that large position makes an investor want to learn more about the asset, is unique to our setting.

V. Conclusion

Most theories of portfolio allocation and asset pricing take investors' information sets as given, and usually common. Investigating what information agents would choose has the potential to yield valuable insights into many portfolio and asset pricing puzzles. This paper has shown that when investors can choose what information they want to learn, given a fixed information capacity, they optimally invest all capacity into learning about one risk factor. When some risk is not learnable, they learn about a small number of risk factors. Since risk-averse investors prefer to take large positions in assets that they are well-informed about, higher-capacity investors hold larger quantities of the assets that they specialize in, causing their portfolios to be less diversified. In equilibrium, investors still prefer to specialize. However, they want to specialize in different assets from other investors. Ex-ante identical agents may optimally hold different portfolios.

The model has new cross-sectional asset pricing predictions. Investors want to learn about principal components of asset payoffs, formalizing an old idea in arbitrage-free pricing theory. Only a few principal components are learned about in equilibrium. The principal components that the economy learns about have lower risk premia. Individual assets that load heavily on those principal components command lower risk premia.

A minor alteration of the model would allow us to think about a setting where information capacity was not fixed, but was costly. For example, time spent increasing capacity might trade off with time spent working. A natural question to pose in this setting is "Why can't an investor delegate his portfolio management to someone who processes information for many investors?" If a manager can tell an investor precisely what portfolio weights to assign to various assets, then he can sell his information processing services and a market for information processing can arise. This is an environment where information processing is centralized. However, competitive forces would fight against complete centralization of information processing. Section D tells us that investors prefer to observe information that other investors are not purchasing. This strategic substitutability will compete with the efficiency gains to centralization. We conjecture an equilibrium with portfolio management that is neither fully centralized, nor fully decentralized. In such a world, competing portfolio managers' learning incentives would be well-described by this model. The question of what the number of portfolio managers will be, how they will be compensated, as well as how the compensation would overcome the inherent informational asymmetry between manager and investor, are topics we plan to address in future work.

References

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

[2] Barber, Brad; Odean, Terrance. "The Internet and the Investor." *Journal of Economic Perspectives*, 2001, vol.15(1), pp.41-54.

[3] Biais, Bruno; Bossaerts, Peter; Spatt, Chester. "Equilibrium Asset Pricing Under Heterogenous Information." *Working Paper*, 2004.

[4] Coval, Joshua; Hirshleifer, David and Shumway, Tyler. "Can Individual Investors Beat the Market?" Working paper, 2002.

[5] Cover, Thomas; and Thomas, Joy. *Elements of Information Theory.* John Wiley and Sons, New York, NY, 1991.

[6] Duffie, Darrell; Sun, Yeneng. "The Exact Law of Large Numbers for Independent Random Matching." *Working Paper*, August 2004.

[7] Goetzmann, William; Kumar, Alok. "Diversification Decisions of Individual Investors and Asset Prices." Working paper, 2003.

[8] Hammond, Peter; Sun, Yeneng. "Monte Carlo Simulation of Macroeconomic Risk with a Continuum of Agents: The Symmetric Case." *Economic Theory*, 2003, vol.21, pp.743-766.

[9] Ivkovic, Zoran; Sialm, Clemens and Weisbenner, Scott. "Portfolio Concentration and the Performance of Individual Investors." Working paper, 2004.

[10] Just, Marcel Adam; Carpenter, Patricia. "A Capacity Theory of Comprehension: Individual Differences in Working Memory." *Psychological Review*, 1992, vol.99(1), pp.122-149.

[11] Kacperczyk, Marcin; Sialm, Clemens, and Zheng, Lu. "On the Industry Concentration of Actively Managed Equity Mutual Funds." *Journal of Finance* forthcoming, 2004.

[12] Mathai, Arak and Provost, Serge. *Quadratic Forms of Random Variables: Theory and Applications.* Marcel Dekker, New York, 1992.

[13] Merton, Robert. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance*, 1987, vol.42(3), pp.483-510.

[14] Miller, G.A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capcacity for Processing Information." 1956, vol.63, pp.81-97.

[15] Peng, Lin. "Learning with Information Capacity Constraints." *Journal of Financial and Quantitative Analysis*, July, 2004.

[16] Polkovnichenko, Valery. "Household Portfolio Diversification: A Case for Rank Dependent Preferences." Federal Reserve Bank of Minneapolis Working Paper, 2003.

[17] Roll, R. "A Critique of the Asset Pricing Theory's Test. Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics*, 1977, vol.4, pp.129-176.

[18] Ross, Stephen. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory*, 1976, vol.13, pp.341-360.

[19] Shannon, C.E. "A Mathematical Theory of Communication." *Bell System Technology Journal*, 1948, vol.27, pp.379-423 and 623-656.

[20] Sims, Christopher. "Implications of Rational Inattention." *Journal of Monetary Economics*, v.50(3), 2003, pp.665-690.

[21] Treynor, Jack; Black, Fischer (1973). "How To Use Security Analysis to Improve Portfolio Selection". *The Journal of Business*, vol.46(1), pp. 66-86

A. Proof of Proposition 1

Consider a deviation from this solution that would allocate some capacity to another asset j, s.t. $\widehat{\Sigma}_{jj} = (1-\epsilon)\Sigma_{jj}$. Keeping total capacity constant implies that $\widehat{\Sigma}_{ii}$ must be increased by a factor of $1/(1-\epsilon)$. This deviation produces a net utility change

$$(\mu_j - p_j r)^2 \Sigma_{jj}^{-1} \left((1-\epsilon)-1\right) + (\mu_i - p_i r)^2 \Sigma_{ii}^{-1} \left(1-(1-\epsilon)\right)$$

Since i is the asset for which $(\mu_i - p_i r)^2 \Sigma_{ii}^{-1} > (\mu_j - p_j r)^2 \Sigma_{jj}^{-1}$, for all $j \neq i$, the net utility change from the deviation is negative. $\square$

B. Proof of Corollary 3

Proposition 2 shows that an investor optimally chooses a portfolio with a low level of diversification, meaning a low $(|q^{div}|/(|q^{div}| + |q^{learn}|)$, if and only if he has a higher information capacity. What remains to be shown is that a higher information capacity entails a higher expected profit: $E[q'(f - rp)]$.

The portfolio weights q can be decomposed into q^{div}, the zero-capacity portfolio and $q_i^{learn} = \frac{1}{\rho\Sigma_{ii}}(\hat{\mu}_i - p_i r)(e^{2K} - 1)$. The profit from the diversified portfolio $E[q^{div'}(f - rp)]$ does not vary in the information capacity K. The profit from the learning portfolio is $E\left[\frac{1}{\rho\Sigma_{ii}}(\hat{\mu}_i - p_i r)(e^{2K}-1)(f_i - rp_i)\right]$. This is increasing in K if $E[(\hat{\mu}_i - p_i r)(f_i - rp_i)] > 0$. Since the difference between f_i and $\hat{\mu}_i$ is a mean-zero, orthogonal expectation error,

$$E[(\hat{\mu}_i - p_i r)(f_i - rp_i)] = E[(\hat{\mu}_i - p_i r)^2] + 0 > 0.$$

$\square$

C. Proof of Proposition 4

Proof: Recalling that a determinant is the product of eigenvalues, and that therefore, pre- and post- multiplying a determinant by Γ leaves it unchanged, the constraint in equation 12 can be rewritten as: $|\Lambda - \hat{\Lambda}|$ is positive semi-definite. This is true if and only if $\Lambda_i - \hat{\Lambda}_i \geq 0 \ \forall i$. Replacing the vector $(\mu - pr)$ in proposition 1 with the vector $(\mu - pr)\Gamma$, and replacing Σ_{ii} with $\hat{\Lambda}_{ii}$, the result follows from the proof of proposition 1.$\square$

D. Proof of Proposition 5

Proof: An investor learns about a risk factor whenever the marginal benefit of allocating the first increment of capacity to that risk factor $\frac{1}{\Lambda_i} + 2\frac{(\Gamma_i'(\mu - pr))^2}{\Lambda_i^2}$ exceeds its marginal cost: $\xi\frac{1}{(1-\alpha)\Lambda_i} - \phi_i$. K enters this inequality only through the Lagrange multiplier ξ. When an investor learns about asset i, constraint (10) is no longer binding and $\phi = 0$. For each risk factor i, there is a cutoff value $\xi_i^\star = (1-\alpha)\left(1 + 2\frac{(\Gamma_i'(\mu - pr))^2}{\Lambda_i}\right)$ where marginal benefit and cost are equal. For all $\xi < \xi_i^\star$, the investor will learn about risk factor i. We know from the proof of proposition 7 that $\partial\xi/\partial K \leq 0$. Therefore, the number of factors i for which $\xi < \xi_i^\star$ must be an increasing step function in K. □

Proof of Corollary 6

Proof: From the proof of proposition 2, we know that a non-zero quantity of an asset is held in the learning fund whenever the investor learns about the asset and the expected excess return is not equal to zero. Getting a signal from a continuous distribution that implies a zero excess return is a zero probability event. Since each risk factor (eigenvector) puts weight on only one asset, the number of assets i for which $\Lambda < \Lambda_i^\star$, which is the number of assets the agent learns about, must be an increasing step function in K. □

E. Proof of Proposition 7

Proof: The diversified portfolio q^{div} is what the investor would hold with zero capacity. It does not change as capacity rises. So, it suffices to show that the absolute value of $q^{learn} = \frac{1}{a}\sum_{i=1}^{N}(\widehat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})(\mu_i - p_i r)$ rises in K. K does not enter directly in the first-order condition, but enters through its affect on the lagrange multiplier λ. Solving for $(\widehat{\Sigma}_{ii} - \alpha\Sigma_{ii})$ and substituting it into the capacity constraint (7) yields an expression for the multiplier $\lambda = e^{-2K/N}(1-\alpha)\prod_{i=1}^{N}[\Sigma_{ii}/\widehat{\Sigma}_{ii}^2(\phi\widehat{\Sigma}_{ii}^2 + \widehat{\Sigma}_{ii} + 2(\mu_i - p_i r)^2]^{1/N}$, which is decreasing in K. The first-order condition (15) is decreasing in K and the second-order condition tells us that it is decreasing in $\widehat{\Sigma}_{ii}$. Therefore, by the implicit function theorem, $\partial\widehat{\Sigma}_{ii}/\partial K < 0$. Since $(\mu_i - p_i r)$ depends on prior beliefs, which do not change with information capacity, and $\widehat{\Sigma}_{ii}^{-1}$ rises, $\partial|q^{learn}|/\partial K > 0$. □

F. Proof of Proposition 9

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

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

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

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

Ψ is the average of agents' 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 agent j observes.[11] Using (18), note that $\left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1} = \hat{\Sigma}_a$, the posterior variance

[11] The Lebesgue integral may not be well defined when $\{\eta_j\}$ are processes of independent random variables for a continuum of agents j, because realizations may not be measurable with respect to the joint space of parameters and

for an investor with the average of all investors' posterior precisions:

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

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

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

Simple algebra reveals that $(f - \rho\Psi^{-1}x) = B^{-1}(rp - A)$, the unbiased signal that agents observe from the price level. From equation 2, we note that the first three terms are equal to the posterior mean of the 'average' agent's beliefs, where average means the agent who has variance $\widehat{\Sigma}_a$. Thus,

$$rp = \hat{\mu}_a - \rho\widehat{\Sigma}_a(\bar{x} + x). \tag{26}$$

The price level is increasing in the posterior belief of the average agent about the mean payoff, and decreasing in risk aversion, the amount of risk the average agent bears, and the supply of the asset.□

G. Proof of Proposition 10

Proof: The result follows 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 being $(\mu - pr)$, and Σ is replaced by $\widetilde{\Sigma}$.

Then, it only remains to be shown that Γ is the eigenvector matrix is *stil*. By definition, $\widetilde{\Sigma} = (\Sigma^{-1} + \Sigma_p^{-1})^{-1}$. We know from appendix F that $\Sigma_p^{-1} = \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi$. Decomposing $\Sigma_{\eta j}$ into its eigenvectors and eigenvalues, 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 $\Psi = \Gamma\Lambda_{\eta a}\Gamma'$. Then, we can rewrite $\widetilde{\Sigma}$ as

$$\widetilde{\Sigma} = (\Gamma^{-1'}\Lambda^{-1}\Gamma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Gamma\Lambda_{\eta a}\Gamma'\Gamma\Lambda_{\eta a}\Gamma')^{-1}.$$

Using the facts that $\Gamma'\Gamma = I$ and $\Gamma^{-1} = \Gamma'$, and collecting terms, we get

$$\widetilde{\Sigma} = \Gamma(\Lambda^{-1} + \frac{1}{\rho^2\sigma_x^2}\Lambda_{\eta a}\Lambda_{\eta a})^{-1}\Gamma'.$$

which has eigenvectors Γ and a diagonal eigenvector matrix $(\Lambda^{-1} + \frac{1}{\rho^2\sigma_x^2}\Lambda_{\eta a}\Lambda_{\eta a})^{-1}$. □

H. Proof of Proposition 11

From proposition 10, we know that investors always allocate their capacity to the asset with the highest value of $(\Gamma_i' E[f - pr])^2(\tilde{\Lambda}_i)^{-1}$. Begin by ordering risk factors by their learning index values when $K = 0$, s.t. $(\Gamma_i' E[f -$

samples. Also, the sample function giving each agent's individual shock may not be Lebesgue measurable, and thus the fraction of agents 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.

$pr])^2(\tilde{\Lambda}_i)^{-1} \geq (\Gamma'_{i+1}E[f-pr])^2(\tilde{\Lambda}_{i+1})^{-1}$. For small levels of K, capacity is allocated only to risk factor 1 and to additional risk factors, only if their initial learning index value is equal to that of factor 1.

Investors will learn about any risk factor i only when that factor is as valuable to learn about as factor 1: $(\Gamma'_1E[f-pr])^2(\tilde{\Lambda}_1)^{-1} = (\Gamma'_iE[f-pr])^2(\tilde{\Lambda}_i)^{-1}$. Is there some level of capacity K_j such that these two index levels are equal? For any non-zero index, there must be. As $K \to \infty$, precision of beliefs about asset 1 becomes infinite: $\psi_{11} \to \infty$. Equation 22, shows that, $rp_1 \to \mu$, which implies that $(\Gamma'_1E[f-pr])^2 \to 0$. 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. $\Box$

I. Proof of Proposition 12

We can rewrite equation (26) for each asset $i \in \{1, 2, \cdots, N\}$ separately:

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

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:

$$p_i = \frac{1}{r}\left(E_a[f_i] - \rho\, Cov_a[f_i, f_m]\right). \tag{27}$$

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:

$$E_a[r_i] - r = \rho\, Cov_a[r_i, f_m]. \tag{28}$$

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

$$E_a[r_m] - r = \rho\, p_m Cov_a[r_m, r_m]. \tag{29}$$

Solving (29) for the risk aversion coefficient ρ, and substituting it into (28), we get the second equation in the proposition.$\Box$