Information Acquisition and Portfolio Under-Diversification

Stijn Van Nieuwerburgh and Laura Veldkamp*

May 17, 2005

Abstract

We solve the problem of an investor who chooses which assets' payoffs to acquire information about before making an investment decision. Investors specialize because information has increasing returns: As an investor learns more about an asset, it becomes less risky and more desirable to hold; as he holds more of the asset, the value of information about it increases. 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. In equilibrium, ex-ante identical investors acquire different information. Information is a strategic substitute because assets that many investors learn about have low expected returns. The theory can explain the empirical evidence that individual investors hold part of their equity portfolio in diversified mutual funds and the rest in a small number of highly-correlated assets. While such portfolios may appear under-diversified, they are optimal for investors who face constraints on how much information they can acquire.

*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, Massimo Massa, Pascal Maenhout, Urs Peyer, Matthew Pritsker, Tom Sargent, Chris Sims, Eric Van Wincoop, Pierre-Olivier Weill, and seminar participants at NYU, Cornell, USC, UCLA, Stanford, UCSB, INSEAD, Ohio State, and the Fed Board of Governors for helpful comments. **JEL** classification: G11, D83, G12, G14. **Keywords**: Asymmetric information, learning, information theory, asset pricing, portfolio theory.

Standard portfolio theory tells an investor how to allocate a fixed amount of wealth across assets. Before they form portfolios, investors face an equally important decision: which assets to research. This model tells an investor how to allocate a fixed capacity to observe information about future asset payoffs. Considering joint information and investment strategies explains why investor portfolios are more concentrated than standard portfolio theory would predict.

When deciding how to allocate their capacity, investors can choose to observe noisy signals about a large number of assets or to specialize and observe more precise information about a few assets. Once they choose to learn about a particular asset, but before they observe their chosen signal, risk-averse investors expect to hold more of that asset in their portfolio, because 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 choice 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 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. In equilibrium, ex-ante identical investors specialize in different risks. Assets that many investors learn about command a lower risk premium. This makes it optimal for different investors to research different assets.

The force behind specialization is a general one: increasing returns to information. Radner and Stiglitz (1984) show that the value of information increases as more of it is obtained, while Wilson (1975) developed the related idea that information value is increasing in a firm's scale of operation. These basic economic insights re-emerge in a setting where the choice is not the quantity of information, but its allocation across assets. Examining information choice in the context of a portfolio problem tells us what investors should specialize in and how specialization and diversification trade-off. Embedding this choice in a general equilibrium model tells us how investors' learning choices interact and how aggregate learning affects asset prices.

Starting with identical prior beliefs, investors can obtain additional signals about what the realizations of future asset payoffs will be. Information is not required to hold an asset (as in Merton, 1987); rather it is a tool to reduce the conditional variance of the asset's payoff. Because we focus on the allocation of information, rather than the quantity chosen, we endow investors with a fixed budget of signal precision to allocate across assets. This budget, which we call capacity, is quantified as an increase in the generalized precision of posterior beliefs about asset payoffs, relative to prior beliefs. After allocating capacity, the investor observes signals drawn from a distribution whose precision he has chosen. Conditional on these signals, he solves a standard CARA-normal portfolio problem. We examine the predictions of this model in both partial equilibrium and general equilibrium settings.

Section 2 analyzes a partial equilibrium model where the investor takes asset prices as given. When asset returns are uncorrelated, the investor chooses to learn about one asset. Because a piece of information is most profitable when it is applied to many shares, the investor allocates her capacity to the asset with the highest squared Sharpe ratio, the asset she is likely to hold most of. She invests in a diversified portfolio and adds to that a "learning portfolio" consisting of a single asset. When asset payoffs are correlated, the investor learns about a single risk factor instead of a single asset. Her "learning portfolio" contains assets in proportion to their covariance with the risk factor. In both cases, it is optimal for the investor with zero information capacity 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.

Specialization arises because of the interaction of the information and portfolio choice problems, not because of the form of the information constraint. Even with a learning technology that exhibits decreasing marginal returns, specialization, though moderated, still persists. With a small amount of capacity, investors fully specialize in learning about one risk. Given sufficient information capacity, the investor will learn about more than one risk factor (section 2.3). The increase and then decrease in the marginal value of information is similar to Keppo, Moscarini and Smith's (2005) more general learning results.

Section 3 investigates a general equilibrium model where a continuum of investors interact (as in Admati, 1985). Endogenous prices act as an additional source of information: they are a noisy signal of what other investors know. While agents still have an incentive to specialize in one risk factor, they also have an incentive to specialize in a different risk factor from the ones other agents are learning about: Learning is a strategic substitute.

When ex-ante identical investors choose to learn about different risk factors, they end up holding different concentrated asset portfolios. Asset returns in our model are described by the CAPM that would hold if each investor had the average of all investors' signal precisions. By characterizing the aggregate allocation of capacity, we can determine what this heterogeneous-information CAPM predicts for the cross-section of asset prices. When an asset's value is correlated with large, high-return risk factors, its price should be higher than the standard CAPM predicts. Finally, section 4 connects the theory with traditional theories of institutional portfolio management.

Recent empirical research confirms the predictions of our theory. 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, 2001). 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). Using Swedish data on investors' complete wealth portfolio, Massa and Simonov (2003) document similar facts. They rule out the explanation that this concentration optimally hedges labor income risk.

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 (see also Coval and Moskowitz, 1999; Massa and Simonov, 2003; Ivkovic and Weisbenner, 2005). Likewise, mutual funds with a higher concentration of assets by industry outperform diversified funds (Kacperczyk, Sialm and Zheng, 2004). If some investors have higher capacity than others, they should consistently earn higher returns. Indeed, the top 10% most successful investors do consistently earn higher excess returns (Coval, Hirshleifer and Shumway, 2002), as do institutional investors with degrees from more selective universities (Chevalier and Ellison, 1999). 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 economic importance of asymmetric information and help to rationalize the multi-billion dollar financial management industry.

Why is it relevant to consider information constraints when information has never been so abundant and so much investment is professionally managed? 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 (e.g. Miller, 1956; Just and Carpenter, 1992). While individuals can avoid processing information by paying a mutual fund manager, even fund managers must decide which stocks to follow, which reports to read and what research to do. The model could be reinterpreted as solving the fund manager's problem. We return to this idea in the conclusion.

Many theories in economics and finance have predictions that depend crucially on what information agents have. This information is usually treated as an endowment. By asking 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.

1 Setup

This is a static model which we break up into 3-periods. In period 1, the investor chooses the variance of signals about asset payoffs. That choice is constrained by information capacity, which bounds the total precision of the signals, 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 and then chooses what assets to purchase.

[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.

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. After learning, the investor will have posterior beliefs about asset payoffs: $f \sim N(\hat{\mu}, \widehat{\Sigma})$. Let r be 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), we call $f_i - rp_i$ the excess return on asset i. Investors have mean-variance utility with absolute risk aversion parameter ρ:

$$U = E\left[q'(f - pr) - \frac{\rho}{2} q'\widehat{\Sigma}q \;\middle|\; \mu\right]. \tag{1}$$

Mean-variance utility arises in settings where agents have negative exponential utility and face normally distributed payoffs.[2] It has the advantage of allowing a tractable solution to an equilibrium model. It treats learned information and prior information as equivalent. This investor chooses information that maximizes his expected utility at the time when he must make his portfolio decision. When choosing what to learn, our investor asks himself, "When I invest, what information would I most like to know?"

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 optimal portfolio $q^\star$ is

$$q^\star = \frac{1}{\rho}\widehat{\Sigma}^{-1}(\hat{\mu} - pr). \tag{2}$$

The model does not rule out short sales: $q^\star < 0$ when $\hat{\mu} - pr < 0$. Any remaining initial wealth is invested in the risk-free asset.

Substituting the optimal portfolio choice into (1) delivers the utility that results from having any beliefs $\hat{\mu}$, $\widehat{\Sigma}$ and investing optimally. The period-1 problem is choosing a signal

[2]It is equivalent to $U = E\left[-\log\left(E\left[\exp\left(-\rho q'(f - pr)\right) | \hat{\mu}, \mu\right]\right) | \mu\right]$. This formulation of utility is related to Epstein and Zin's (1989) preference for early resolution of uncertainty and Hansen and Sargent's (2004) models of risk-sensitive control.

distribution to maximize this expected payoff:

$$U = E\left[\frac{1}{2}(\hat{\mu} - pr)'\widehat{\Sigma}^{-1}(\hat{\mu} - pr)|\mu\right] \tag{3}$$

Period-1 learning problem At time 1, the investor chooses how to allocate his information capacity by choosing a normal distribution from which he will draw an $N \times 1$ signal about asset payoffs.[3] 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. Since every signal variance has a unique posterior belief variance associated with it, we can economize on notation and optimize over posterior belief variance $\widehat{\Sigma}$ directly. Prior (unconditional) variances and covariances are not random; they are given. Posterior (conditional) variances are also not random; they are choice variables that summarize the investor's optimal information decision.

In period 1, posterior means are random: $(\hat{\mu} - pr) \sim N(\mu - pr, V_{ER})$, where $V_{ER} \equiv Var[\hat{\mu} - pr|\mu]$. Inside the expectation of equation (3) is a non-central χ^2-distributed random variable. The solution is to maximize

$$max_{\widehat{\Sigma}} \;\; \frac{1}{2}Tr(\widehat{\Sigma}^{-1}V_{ER}) + \frac{1}{2}(\mu - pr)'\widehat{\Sigma}^{-1}(\mu - pr). \tag{6}$$

There are 2 constraints governing how the investor can choose his signals. The first constraint the *capacity constraint*. The work on information acquisition with one risky asset quantified information as the ratio of variances of prior and posterior beliefs (Verrecchia, 1982). The more information a signal contains, the more the posterior variance of the asset falls below the prior variance, and the more information capacity is required to observe the

[3] Choosing normal signals is optimal. When an objective is quadratic, normal distributions maximize the entropy over all distributions with a given variance (see Cover and Thomas (1991), Chapter 10).

signal. We generalize the metric to an multi-signal setting by calling capacity the ratio of the generalized prior variance to the generalized posterior variance, where generalized variance refers to the determinant of the variance-covariance matrix.

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

The amount of capacity K bounds the reduction in uncertainty of payoffs due to the knowledge of the signal η.[4]

This capacity constraint is one possible description of a learning technology. We think it is a relevant constraint because it is a commonly-used distance measure in econometrics (a log likelihood ratio) and in statistics (a Kullback-Liebler distance[5]); it is equivalent to a bound on entropy reduction, which has a long history in information theory as a quantity measure for information (Shannon 1948[6]); it can be re-interpreted as a constraint on the length of the binary code needed to describe signals; it is a measure of information complexity (Cover and Thomas 1991); and it has been previously used in economics (Sims 2003) and finance (Peng 2004) to model limited mental processing ability.[7]

That having been said, this particular formulation of the learning technology is not crucial for the results. Our capacity constraint is simply a way to describe a feasible set of learning possibilities that is rich enough to analyze the trade-off between diversification and specialization in learning. One alternative is to endow an investor with a fixed number of signals with equal precision, and let him choose how many signals to apply to each asset. Section 2.1 shows that this technology also generates specialization. Section 2.3 considers a second alternative learning technology, one with decreasing returns. The incentive to specialize persists, but is moderated. Finally, while studying the extensive margin of information

[4] 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: $1/2 \log\left(|\Sigma_\eta^{-1}\Sigma + I|\right) \leq K$.

[5] In statistics, this distance is used as a measure of how difficult it is to distinguish one distribution from another.

[6] In information theory, capacity is the standard measure of information: the reduction in entropy. 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. 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.

[7] 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.

acquisition is interesting, adding a cost for capacity won't change the nature of the capacity allocation decision. For every cost, there is an amount of capacity K that produces an identical result. In sum, because the increasing returns to specialization show up in the objective, through the endogenous portfolio choice, a broad class of learning technologies preserve specialization.

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

$$\Sigma_\eta \quad \text{positive semi-definite} \tag{8}$$

Without this constraint, the investor could increase uncertainty about one variable in order to obtain a more precise signal about another, without violating the capacity constraint. Ruling out increasing uncertainty implies that investors cannot forget nor see signals with negative information content.

Learning about correlated risks When asset payoffs co-vary, learning about one asset's payoff is informative about others. To keep track of what is being learned about, we study synthetic assets that are linear combinations of underlying assets, and that do not co-vary with each other. These synthetic assets are principal components, or risk factors. The coefficients of these linear combinations are given by an eigen-decomposition. This decomposition splits the prior variance-covariance matrix Σ into a diagonal eigenvalue matrix Λ, and an eigenvector matrix Γ: $\Sigma = \Gamma\Lambda\Gamma'$. The Λ_i's are the variances of each risk factor i. The i^{th} column of Γ (denoted by Γ_i) gives the loadings of each asset on the i^{th} risk factor.

Investors obtain signals about the payoffs of risk factors ($f'\Gamma_i$). Studying principal component risks is a well-established idea in the portfolio literature (Ross, 1976). Nothing prevents an investor from learning about many risks. The only thing this rules out is signals with correlated information about risks that are independent. For example, if one risk factor represented oil price risk and another represented rain-related risk, an investor cannot observe a linear combination of future oil prices and rainfall. Since rain and oil prices are independent, events that cause joint movements in rain and oil do not occur. We assume that the investor cannot choose to learn about zero-probability events. He can learn about rainfall and oil price risk only by acquiring a signal about each. While the key results, specialization and strategic substitutability, hold for any given set of orthogonal risk factors, this particular decomposition keeps the problem tractable. Investors will have posterior beliefs with the same eigenvectors as their prior beliefs ($\widehat{\Sigma} = \Gamma\hat{\Lambda}\Gamma'$), but with lower weights $\hat{\Lambda}_i$ on some risks they chose to learn about. The decrease in risk factor variance $\Lambda_i - \hat{\Lambda}_i$ captures how much an investor learned about that risk.

The sequence of events is summarized in figure 1.

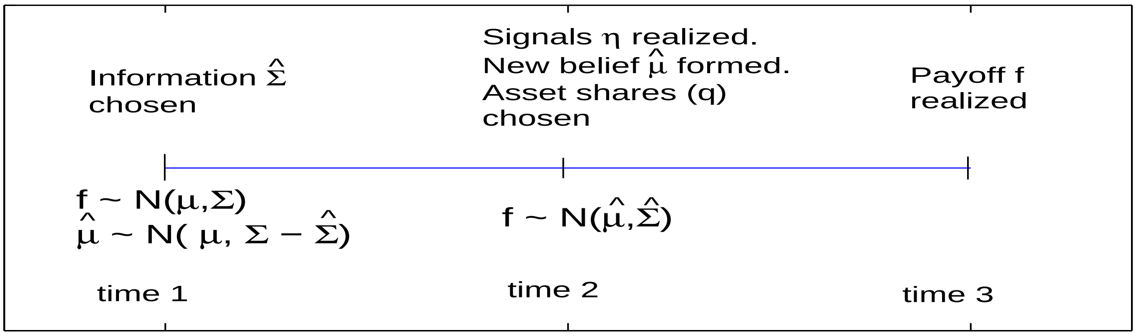

Figure 1: Sequence of events in partial equilibrium model

A solution to the investor's problem is a choice of the eigenvalues of $\widehat{\Sigma}$ that maximizes (6) subject to (7) and (8), and portfolio positions that satisfy (2).

2 Partial Equilibrium Results

2.1 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 correlated assets.

When investors takes prices as given, $V_{ER} = var[\hat{\mu}|\mu] = \Sigma - \widehat{\Sigma}$. Define the ratio of posterior to prior precisions of an asset i: $y_i \equiv \frac{\widehat{\Sigma}_{ii}^{-1}}{\Sigma_{ii}^{-1}}$. We can rewrite the problem in equation (6) as

$$max_{\{y_1,\cdots,y_N\}} \frac{1}{2}\{-N + \sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \theta_i^2 y_i\}. \tag{9}$$

$$s.t. \quad \prod_{i=1}^{N} y_i = \exp(2K)$$

$$y_i \geq 1,\ \forall i$$

where θ_i^2 is the prior squared Sharpe ratio of asset i: $\theta_i^2 \equiv \frac{(\mu_i - p_i r)^2}{\Sigma_{ii}}$. The first constraint results from (7) and the fact that the determinant of a diagonal matrix is the product of the diagonal entries. The second constraint uses (8), (5) and the fact that a diagonal matrix is positive semi-definite if and only if all its elements are non-negative.

The key feature of the learning problem (9) is that it is linear in precision y_i. It is the linearity of the objective that delivers a corner solution. The corner solution is to increase precision on the risk factor with the highest weight (θ_i^2), as much as possible. This solution would arise from a wide range of learning constraints. One example would be a constraint on the *sum* of the posterior or signal precisions. This is equivalent to endowing an investor with a fixed number of signals of equal precision, and letting him choose how many signals to apply to each asset. Increasing returns to learning is not a result that is specific to the entropy constraint.

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 $\theta_i^2 = (\mu_i - p_i r)^2 \Sigma_{ii}^{-1}$.*

Proof is in appendix A.1. 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 $E[q_i] = (\mu_i - p_i r)/(\rho\widehat{\Sigma}_{ii})$ 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 $E[q^{learn}] = \frac{1}{\rho\Sigma_{ii}}(\mu_i - p_i r)(e^{2K} - 1)$. 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}, (plus his position in the risk free asset).

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 A.2. 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

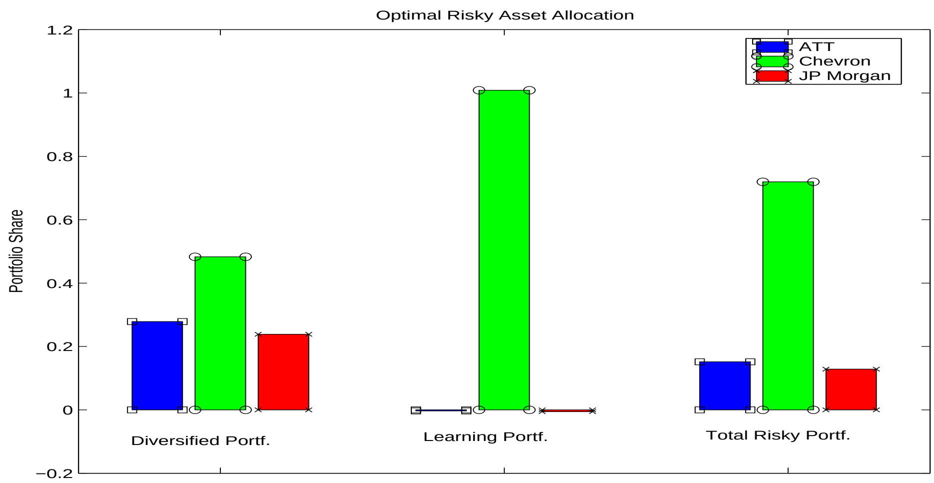

Figure 2: Under-Diversification and the Increasing Returns to Learning: Uncorrelated Assets.

We illustrate the portfolio composition with a simple numerical example. Figure 2 illustrates the case of three uncorrelated S&P 500 assets.[8] 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,

[8]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.

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.

2.2 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 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 $\widehat{\Sigma} = \Gamma\hat{\Lambda}\Gamma'$. Equivalently, the investor chooses the precision ratios of the risk factors i, y_i, which we redefine as $y_i \equiv \frac{\hat{\Lambda}_{ii}^{-1}}{\Lambda_{ii}^{-1}}$. The investor solves problem (9), where the prior squared Sharpe ratios θ_i^2 now refer to risk factors: $\theta_i^2 \equiv \frac{((\mu - pr)'\Gamma_i)^2}{\Lambda_{ii}}$. The capacity constraint still takes the form $\prod_{i=1}^{N} y_i = \exp(2K)$ because the determinant of $\Sigma\widehat{\Sigma}^{-1}$ is the product of its eigenvalues $\Lambda_{ii}\hat{\Lambda}_{ii}^{-1}$ and because Σ and $\widehat{\Sigma}$ share eigenvectors Γ. The no-negative learning constraint, which requires $\Sigma - \widehat{\Sigma}$ to be positive semi-definite, or equivalently that all its eigenvalues are non-negative becomes $y_i \geq 1$.

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 $\theta_i^2 = ((\mu - pr)'\Gamma_i)^2 \Lambda_{ii}^{-1}$.*

The proof follows immediately from proposition 1 and the new definition of y_i and θ_i^2. 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 specializes completely in whatever risk factor he ranks first. 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,

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

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 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).

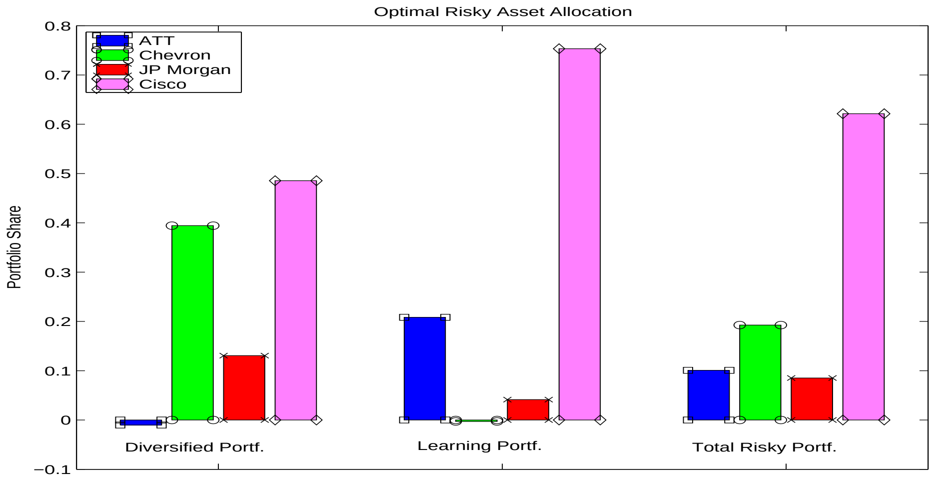

Figure 3: Under-Diversification and the Increasing Returns to Learning: Correlated Assets.

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.

2.3 Un-Learnable Risk and Decreasing Returns to Learning

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 some benefits to diversification and makes high-capacity investors learn about multiple risk factors. Adding un-learnable risk is also a way of introducing decreasing returns to learning. A capacity constraint that embodies decreasing returns to specialization still does not restore full diversification. This reinforces the point that specialization is driven by the increasing returns property of information, not the form of the capacity constraint.

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 zero, 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 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 $(0 < \alpha < 1)$.[9] 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.

$$\log(|\Sigma - \alpha\Sigma|) - \log(|\widehat{\Sigma} - \alpha\Sigma|) \leq 2K \tag{10}$$

Rewrite this constraint in terms of the precision ratios $y_i \equiv \frac{\hat{\Lambda}_{ii}^{-1}}{\Lambda_{ii}^{-1}}$:

$$-\sum_{i=1}^{N} \log(y_i^{-1} - \alpha) + \log(1-\alpha) \leq 2K. \tag{11}$$

The no-negative learning constraint is as before: $y_i \geq 1 \; \forall i$.

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 y_i. It describes an interior solution to the maximization problem.

$$(1+\theta_i^2) = \xi \frac{1}{y_i - \alpha y_i^2} - \phi_i, \tag{12}$$

where ξ is the Lagrange multiplier on (11), and ϕ_i is the Lagrange multiplier on $y_i \geq 1$. The left hand side is the marginal benefit of learning about risk factor i, the right hand side is the marginal cost. Taking a second derivative confirms that a solution to (12) is a maximum in the region $y_i > \frac{1}{2\alpha}$.

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

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 A.3.

The reason for learning about additional assets can be seen by examining the marginal benefit and the marginal cost of learning for an asset where $y_i > 1$ ($\phi_i = 0$). The marginal benefit is constant at $1 + \theta_i^2$. The marginal cost is convex in y_i; it first declines until $y_i = \frac{1}{2\alpha}$, and then increases. As the investor learns more and y_i increases, the marginal cost decreases. Increasing returns to scale in learning are still present. However, as y_i surpasses $\frac{1}{2\alpha}$, the marginal cost starts to increase. In the limit, as y_i approaches $\frac{1}{\alpha}$, and the investor gets closer to learning all the learnable risk, the marginal cost approaches infinity. Therefore, there is some finite cutoff level of y_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 A.4.

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:* $\lim_{k\to\infty} E[q^{learn}] = \left(\frac{1}{\alpha} - 1\right) q^{div}$.

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 $E[q^{learn}] = \frac{1}{\rho}\left(\frac{1}{\alpha} - 1\right)\Sigma^{-1}(\mu - pr)$, a scaled-up copy of the diversified mutual fund. □

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 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.

3 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.

3.1 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 1. 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.[10] 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

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

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

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

$$\int_0^1 \widehat{\Sigma}_j^{-1}(\hat{\mu}_j - pr)dj = \bar{x} + x \tag{13}$$

3.2 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 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 A.5, 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

$$\hat{\mu} \equiv E[f|\mu, \eta, p] = \left(\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1}\right)^{-1} \left(\Sigma^{-1}\mu + \Sigma_\eta^{-1}\eta + \Sigma_p^{-1}B^{-1}(rp - A)\right) \tag{14}$$

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

$$\hat{\Sigma} \equiv V[f|\mu, \eta, p] = \left(\Sigma^{-1} + \Sigma_\eta^{-1} + \Sigma_p^{-1}\right)^{-1}. \tag{15}$$

These are the conditional mean and variance that agents use to form their portfolios in period 2. 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. Optimal portfolios and expected utility are the same as in the partial equilibrium problem (equations 2 and 3). Only the conditioning information changes.

3.3 Individual's 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 the time-2 expected excess return $(\hat{\mu} - pr)$ conditional on μ is now a normally distributed variable at time 1 with mean $(I - B)\mu - A$ and variance $V_{ER} \equiv \Sigma - \widehat{\Sigma} + B\Sigma B' + CC'\sigma_x^2 - 2\Sigma B'$:

$$max_{\widehat{\Sigma}} \left\{ \frac{1}{2} Tr(\widehat{\Sigma}^{-1} V_{ER}) + \frac{1}{2}((I - B)\mu - A)'\widehat{\Sigma}^{-1}((I - B)\mu - A) \right\}. \tag{16}$$

Just as in the partial equilibrium problem, the choice of the covariance matrix of the posterior belief $\widehat{\Sigma}$ is subject to two constraints. The constraints are formally the same as in section 1, but require re-interpretation. The first constraint is that the total information the investor sees cannot reduce entropy by more than his capacity K. Being a constraint on the distance between the posterior belief variance $\widehat{\Sigma}$ and the prior belief variance Σ, it assumes that investors use capacity to extract payoff relevant information both from private signals η and from prices. Some capacity must be devoted to price discovery; the remaining capacity can be optimally allocated to signals.[11] The second constraint is the equivalent of (8). This

[11] In the partial equilibrium problem the capacity constraint on signals was $\log\left(|I + \Sigma\Sigma_\eta^{-1}|\right) \leq 2K$; in the

no-negative learning constraint prevents investors from forgetting information that is either contained in priors or in prices.

$$\widetilde{\Sigma} - \widehat{\Sigma} \quad \text{positive semi-definite} \tag{17}$$

where $\widetilde{\Sigma} = V[f|\mu, p] = (\Sigma^{-1} + \Sigma_p^{-1})^{-1}$ is what the conditional variance of asset payoffs would be if the agent observed no private signals, but only learned through the price level.

The sequence of events is as in the partial equilibrium problem, except that at time 2, prices p are revealed, in addition to private signals η.

As in partial equilibrium, learning about principal components of asset payoffs implies that prior and posterior variances have the same eigenvectors. This allows us to recast the problem in terms of eigenvalues. Recall the definitions of the precision ratios of the risk factors relative to the prior variance matrix: $y_i \equiv \frac{\hat{\Lambda}_{ii}^{-1}}{\Lambda_{ii}^{-1}}$ and the prior squared Sharpe ratio of risk factor i:

$$\theta_i^2 \equiv \frac{E[\Gamma_i'(f - pr)]}{Var[\Gamma_i' f]} = \frac{(((I - B)\mu - A)'\Gamma_i)^2}{\Lambda_{ii}}. \tag{18}$$

The problem can be written as:

$$max_{\{y_1, \cdots, y_N\}} \frac{1}{2}\{-N + \sum_{i=1}^{N} X_i y_i + \sum_{i=1}^{N} \theta_i^2 y_i\}. \tag{19}$$

$$s.t. \quad \prod_{i=1}^{N} y_i = \exp(2K)$$

$$y_i \geq \frac{\tilde{\Lambda}_{ii}^{-1}}{\Lambda_{ii}^{-1}}, \ \forall i$$

where X_i measures the magnitude of the exploitable pricing errors in risk factor i. If an investor becomes informed, his valuation of an asset, based on his private information, will deviate from the realized price. We call this deviation an 'exploitable pricing error'. X_i measures the period-1 expected squared pricing error.

Appendix A.7 shows that exploitable pricing errors depend on how much asset prices are

general equilibrium setting it becomes $\log\left(|I + \Sigma\Sigma_\eta^{-1} + \Sigma\Sigma_p^{-1}|\right) \leq 2K$.

affected by true payoffs (fundamentals) and by asset supply shocks:

$$X_i = (1 - \Lambda_{B,i})^2 + \Lambda_{C,i}^2 \sigma_x^2 \tag{20}$$

The first term shows that pricing errors increase when prices are less reflective of true payoffs. Λ_{Bi} is the weight of the i^{th} risk factor's true payoff on the factor's price. It is the i^{th} eigenvalue of the matrix B in proposition 9. When Λ_{Bi} is small, $(1 - \Lambda_{Bi})^2$ is large, prices don't co-vary much with investors' posterior beliefs $\hat{\mu}$. A low covariance makes exploitable pricing errors 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, on the other hand, knows little about the true payoff and cannot exploit this difference.

The second term shows that pricing errors increase when prices are more reflective of supply shocks. Λ_{Ci} is the weight of the i^{th} risk factor's supply shock on the factor's price. It is the i^{th} eigenvalue of the matrix C in proposition 9. When Λ_{Ci} is high, $(\Lambda_{Ci})^2$ is high, X_i is high. Supply shocks create noise in prices that is exploitable by a well informed investor. For example, if such 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.

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 Γ_i associated with the highest value of the squared Sharpe ratio plus exploitable pricing error: $\theta_i^2 + X_i$.*

Proof in appendix A.6.

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. The size of exploitable pricing errors is determined by the fraction of investors who learn about risk factor i. This is a new effect that shows up in general equilibrium only.

Learning is a strategic substitute. The more precise the posterior beliefs of the average investor about risk factor i, the less valuable it is to learn about. Let Ψ be 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. Let $\Lambda_{\Psi i}$ be the eigenvalue of Ψ corresponding to the i^{th} risk factor.

Proposition 11. *There is strategic substitutability in learning:* $(X_i + \theta_i^2)$ *is a strictly decreasing, monotonic function of* $\Lambda_{\Psi i}$.

Proof is in appendix A.7. When other investors learn more about a risk factor, the size of its exploitable pricing errors and its expected return fall.

The substitutability result is crucial to preventing diversification in general equilibrium. Because of substitutability, investors want to learn about risk factors that other investors are not learning about. If learning were a strategic complement, investors would want to specialize in learning about the same risk factor. This would make information symmetric: All investors would face the same payoff variance. As a result, they would all want to hold a lot of the risk factor they learn about. Since markets must clear, and investors each have an equal benefit of holding the asset they specialized in, they end up holding an equal share of the supply in expectation. Full diversification would arise.

3.4 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 will get the same expected utility from learning about any of the risk factors that the economy learns about. The reason they choose to specialize in different risk factors is because learning is a strategic substitute.

Proposition 12. *The number of risk factors that the economy learns about is weakly increasing in economies' aggregate capacity* K.

Proof in appendix A.8. 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:

$$E[f - pr] = \rho \widehat{\Sigma}_a \bar{x}. \tag{21}$$

A reduction in expected return makes assets less valuable to learn about (lower θ_i^2 in proposition 10). 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.

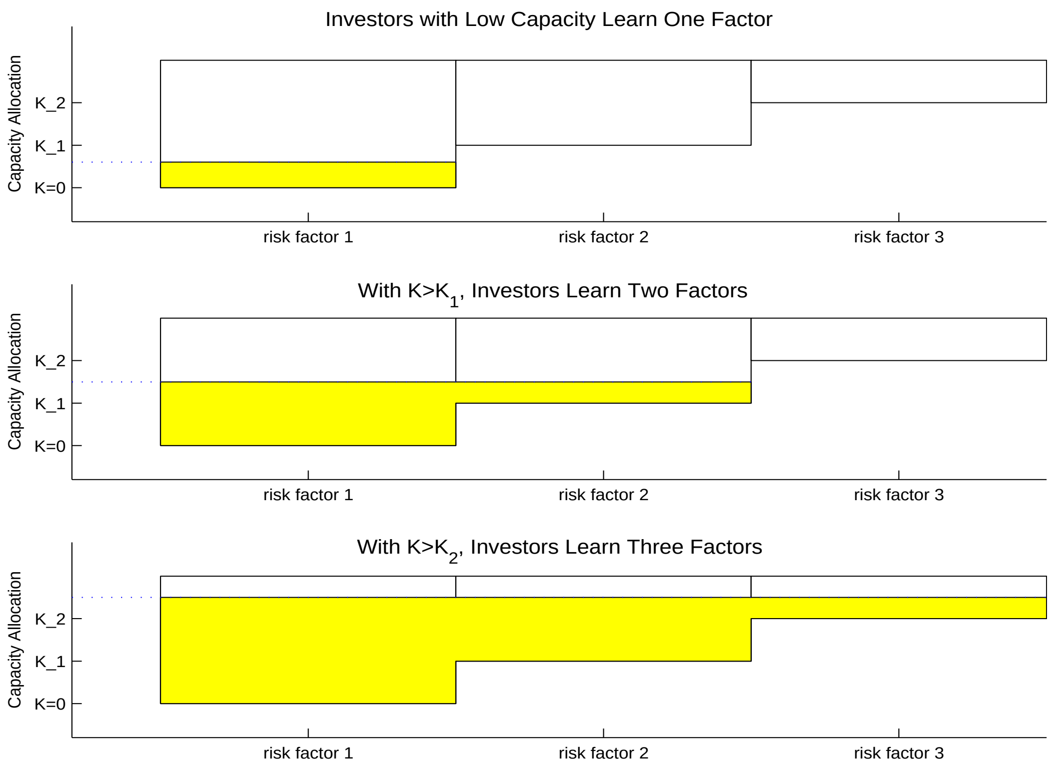

Figure 4: Aggregate allocation of information capacity for low, medium, and high-capacity investors.

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 it. The first investor learns about the risk factor that is most valuable when no other learning takes place. This is the risk factor with the highest learning index, the sum of its squared Sharpe ratio (θ_i^2) and $\rho^2\sigma_x^2\Lambda_{ii}$. This is the same risk factor as the one the partial equilibrium investor would learn about for $\sigma_x^2 = 0$. Subsequent investors will continue to allocate their capacity to factor i until the value of learning about risk factor i has dropped sufficiently that it equals the value of learning about the next most valuable risk factor l. This cutoff is when capacity $K = K_1$ in figure 4. 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. The reason that the first risk factor becomes gradually less valuable to learn about is that, as more investors become informed about it, $\hat{\Lambda}_{ai}^{-1}$ increases, thereby reducing both X_i (proposition 11) and θ_i^2, the latter through a decrease in expected return (equation 21). 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 4). 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 as described in proposition 2. Each investor holds a diversified portfolio, plus a learning portfolio. The diversified portfolio needs to be adjusted for learning from prices. The learning portfolio 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 4, 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 2.3), 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 modeling large portfolio managers.

3.5 Cross-Section of Asset Returns

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 (21) and the result that $\widehat{\Gamma} = \Gamma$, as:

$$\Gamma'_k E[f - rp] \;=\; \rho \hat{\Lambda}_{ak} \Gamma'_k \bar{x}. \tag{22}$$

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 (see equations 26 and 25 in appendix A.5). In our model with heterogenous information and partially revealing prices, a version of the Capital Asset Pricing Model holds.

Proposition 13. *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*

$$p_i = \frac{1}{r}\left(E_a[f_i] - \rho\, Cov_a[f_i, f_m]\right). \tag{23}$$

The equilibrium return is

$$E_a[r_i] - r = \frac{Cov_a[r_i, r_m]}{Var_a[r_m]}(E_a[r_m] - r) \equiv \beta_a^i(E_a[r_m] - r). \tag{24}$$

The proposition, similar to Lintner (1969), 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 of a representative agent. Without a theory of information acquisition, this pricing relationship is not testable. The information of the representative agent used in equations (23) and (24) cannot be observed by an econometrician or deduced from prices. Our contribution is to predict the information set of the representative agent.

Incorporating our results into the CAPM (equation 24) can explain why a public-information based CAPM under-prices large assets. In their seminal paper, Fama and French (1992) show that large firms offer lower average returns than small firms *for a given beta*. The standard CAPM fails to explain the cross-section of size portfolio returns because the beta for large (small) firms is 'too high' ('too low') to account for the return difference. This beta is based on public (prior) information. When investors can learn, the true risk of an asset depends on its 'learning beta' β_a^i, which is based on public information and private information investors have chosen to learn. Combining equations (18), (21), and proposition 10, the value of learning about risk factor i is given by $\rho^2\left(\frac{\hat{\Lambda}_{ii}^a}{\Lambda_{ii}}\right)(\Gamma_i'\bar{x})^2\hat{\Lambda}_{ii}^a + X_i$. Learning value is increasing in the size of the risk factor $(\Gamma_i'\bar{x})$. If large assets load heavily on these large risk factors, the representative investor will be well-informed about large assets. This lowers the conditional covariance of large assets with the market, not because it reduces the correlation, but because it reduces the conditional variance of the asset's return. Our findings suggest that any assets that load heavily on the largest principal components should have returns that are lower than the standard CAPM predicts.

4 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.

Optimal portfolio management is a long-standing issue in the mutual fund literature. In the seminal paper by Treynor and Black (1973, henceforth TB), security analysts can analyze only a limited number of stocks. It departs from the efficient markets hypothesis by assuming that individual portfolio managers can exploit mis-pricing 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': $\alpha_k/\sigma^2(\varepsilon_k)$.[12]

TB and our paper 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 irrational mis-pricing. A TB regression in our model will produce α's 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. Our theory suggests another notion of α: Investors demand different risk premia for the same asset because they have an individual-specific α, arising from private information.

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

[12] By defining q^{div} as the zero-capacity portfolio, 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.

5 Conclusion

Most theories of portfolio allocation and asset pricing take investors' information as given. Investigating optimal information choice has the potential to yield valuable insights into many portfolio and asset pricing puzzles. When investors can choose what information to acquire, given a fixed information capacity, they optimally devote all capacity to 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 larger positions in assets they are better-informed about, high-capacity investors hold larger 'learning portfolios', causing their total portfolio to be less diversified. In equilibrium, investors 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. Assets that many investors learn about command lower risk premia, than standard asset pricing models predict. These assets are ones that co-vary with the largest principal components.

While this model has focussed on a static information allocation problem, it could be extended along many dimensions. The quantity of information could be endogenized with a capacity production function. A model of dynamic information choice, incorporating recent advances in the dynamics of information value (Bernhardt and Taub, 2005), could be used to explain the persistence and turnover of investor portfolio holdings and time variation in expected asset returns. Finally, analyzing a market for information capacity could answer questions about the organization of the portfolio management industry.

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 were to sell information, information resale would undermine their profits. To avoid this problem, they should manage investors' portfolios directly. If information capacity is costly, then managers maximize profit by each specializing in a different risk factor. Whether an investor's portfolio will also be concentrated hinges on how portfolio managers set prices. Suppose that the fee is a fixed fraction of assets under management. This linear pricing scheme would undermine incentives to specialize. In this paper, individuals held concentrated portfolios because of increasing returns to information: investors could apply any signal to many shares of an asset, at no extra cost. But if they pay per share of asset, they will invest in many funds and diversify. However, such a linear pricing scheme is not a competitive equilibrium. Offering quantity discounts induces more investment in a fund. The additional investment reduces the fund manager's per-share cost and allows him to compete linear-price suppliers out of the market. Quantity discounts make investing small amounts in many funds costly. Competitive pricing of portfolio management services forces investors to internalize increasing returns to specialization; optimal under-diversification reappears.

A theory of information choice in financial markets is vital to understanding or justifying the active portfolio management industry. While a formal analysis of the industry and its effect on investor portfolios is left for future work, understanding an individual's information choice problem is a necessary first step.

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 and Odean, Terrance. "The Internet and the Investor." *Journal of Economic Perspectives*, 2001, v.15(1), pp.41-54.

[3] Bernhardt, Dan and Taub, Bart. "Strategic Information Flows In Stock Markets." *Working Paper*, 2005.

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

[5] Chevalier, Judith and Ellison, Glenn. "Are Some Mutual Fund Managers Better than Others? Cross-Sectional Patterns in Behavior and Performance." *Journal of Finance*, 1999, v.54 (3), pp.875-899.

[6] Coval, Joshua; Hirshleifer, David and Shumway, Tyler. "Can Individual Investors Beat the Market?" *Working Paper*, 2003.

[7] Coval, Joshua and Moskowitz, Tobias. "Home Bias at Home: Local Equity Preference in Domestic Portfolios." *Journal of Finance*, 1999, v.54, pp.1-54.

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

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

[10] Epstein, Larry and Stanley Zin. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica*, 1989, v. 57, pp. 937-69.

[11] Fama, Eugene and French Kenneth. "The Cross-Section of Expected Stock Returns." *The Journal of Finance*, 1992, v.47(2), pp.427-465.

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

[13] Hansen, Lars and Sargent, Thomas. *Recursive Models of Dynamic Linear Economies.* Monograph, June 2004.

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

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

[16] Ivkovic, Zoran and Weisbenner, Scott. "Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments." *Journal of Finance*, 2005, v.60, pp.267-306.

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

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

[19] Keppo, Jussi; Moscarini, Giuseppe, and Smith, Lones. "The Demand For Information: More Heat Than Light." *Working Paper*, 2005.

[20] Lintner, John. "The Aggregation of Investors' Diverse Judgements and Preferences in Purely Competitive Markets." *Journal of Financial and Quantitative Analysis*, 1969, v. 4(4), pp.347-400.

[21] Massa, Massimo and Simonov, Andrei. "Hedging, Familiarity and Portfolio Choice." *The Review of Financial Studies*, forthcoming, 2003.

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

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

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

[25] Polkovnichenko, Valery. "Household Portfolio Diversification: A Case for Rank Dependent Preferences." *The Review of Financial Studies*, August 2004.

[26] Radner, Roy and Stiglitz, Joseph. "A Nonconcavity in the Value of Information." In: *Bayesian Models in Economic Theory.* eds. M. Boyer and R.E. Kihlstrom, Elsevier Science Publishers B.V., 1984.

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

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

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

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

[31] Verrecchia, Robert. "Information Acquisition in a Noisy Rational Expectations Economy." *Econometrica*, 1982, v.50(6), pp.1415-1430.

[32] Wilson, Robert. "Informational Economies of Scale." *Bell Journal of Economics*, 1975, v.6, pp.184-195.

A Appendix

A.1 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. $\square$

A.2 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,

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

$\square$

A.3 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 $1+\theta_i^2$ exceeds its marginal cost: $\xi\frac{1}{y_i(1-\alpha y_i)}-\phi_i$. K enters this inequality only through the Lagrange multiplier ξ, the shadow cost of capacity. When an investor learns about asset i, the no-negative learning constraint is no longer binding and $\phi_i=0$. For each risk factor i, there is a cutoff value $\xi_i^\star = y_i(1-\alpha y_i)\left(1+\theta_i^2\right)$ where marginal benefit and cost are equal. For all $\xi<\xi_i^\star$, the marginal benefit is greater than the marginal cost and 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. $\square$

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 asset payoffs are independent, each risk factor corresponds to one and only one asset. Proposition 5 shows that when capacity increases, the number of risk factors learned about rises. Thus the number of assets learned about rises, and the number of different assets held in the learning fund rises. $\square$

A.4 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. When K rises, the absolute value of $E[q^{learn}]=\frac{1}{\rho}(\widehat{\Sigma}^{-1}-\Sigma^{-1})(\mu-pr)$ is affected only through $\widehat{\Sigma}$. How K affects $\widehat{\Sigma}$ can be seen in the first-order condition; it enters through the Lagrange multiplier ξ. Solving for $(y_i-\alpha y_i^2)$ from equation (12) and substituting it into the capacity constraint (11) yields an expression for the multiplier $N\log(\xi) = \sum_{i=1}^N \left(2\log(y_i)+\log(1+\theta_i^2+\phi_i)\right)+\log(1-\alpha)-2K$, which is decreasing in K. Applying the implicit function theorem to the first-order condition (12), we get $\frac{\partial y_i}{\partial K}\geq 0$, for every risk factor i, with strict inequality for those risk factors that are learned about. Since $\{y_i\}$ are the eigenvalues of $\widehat{\Sigma}^{-1}\Sigma$, and Σ is a constant, $\frac{\partial y_i}{\partial K}\geq 0$ implies that each element of the eigenvalue matrix $\hat{\Lambda}^{-1}$ of $\widehat{\Sigma}^{-1}=\Gamma\hat{\Lambda}^{-1}\Gamma'$, weakly rises with K. As a result, $\frac{\partial |E[q^{learn}]|}{\partial K}\geq 0$, with strict inequality for the risk factors that the investor learns about. $\square$

A.5 Proof of Proposition 9

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

$$\begin{aligned}
A &= \left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1}\left(\Sigma^{-1}\mu - \rho\bar{x}\right),\\
B &= \left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1}\left(\Psi + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi\right),\\
C &= -\left(\Sigma^{-1} + \frac{1}{\rho^2\sigma_x^2}\Psi'\Psi + \Psi\right)^{-1}\left(\rho I + \frac{1}{\rho\sigma_x^2}\Psi'\right).
\end{aligned}$$

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.[13]

Using (15), 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:

$$\hat{\Sigma}_a \equiv \left(\int_0^1 \hat{\Sigma}_j^{-1} dj\right)^{-1} \tag{25}$$

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 4, we note that the first three terms are equal to the posterior mean of the 'average' agent's beliefs:

$$\hat{\mu}_a \equiv E_a[f_i] \equiv \int_0^1 \hat{\mu}_j dj \tag{26}$$

Thus,

$$rp = \hat{\mu}_a - \rho\widehat{\Sigma}_a(\bar{x} + x). \tag{27}$$

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.□

[13]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 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.

A.6 Proof of Proposition 10

As long as $\widehat{\Sigma}^{-1}$ and V_{ER} have the same eigenvectors as Σ, then 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 $(\mu-pr)$. Sums, products and inverses of matrices with identical eigenvectors preserve those eigenvectors. This tells us that Ψ can be rewritten as $\Psi=\int_0^1 \Gamma^{-1\prime}\Lambda_{\eta j}^{-1}\Gamma^{-1}dj$. 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 (see appendix A.5), all have eigenvectors Γ. $\square$

A.7 Proof of Proposition 11

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 y_i, be the ratio of the precision of the posterior to the precision of the prior, i.e. it is the i^{th} eigenvalue of $\widehat{\Sigma}^{-1}\Sigma$: $y_i\equiv\hat{\Lambda}_{ii}^{-1}\Lambda_{ii}$. Let X_i be the i^{th} eigenvalue of $\Sigma^{-1}(V_{ER}+\widehat{\Sigma})$. Then the i^{th} eigenvalue of the matrix inside the trace is y_iX_i, and $Tr\left(\widehat{\Sigma}^{-1}\Sigma\Sigma^{-1}(V_{ER}+\widehat{\Sigma})\right)=\sum_{i=1}^N X_iy_i$. This is because Σ, $\widehat{\Sigma}$, B and C all share the same eigenvectors Γ.

The expression for X_i, the i^{th} eigenvalue of $\Sigma^{-1}(V_{ER}+\widehat{\Sigma})$, is:

$$X_i=\Lambda_{ii}^{-1}\left[\Lambda_{ii}\left(1+\Lambda_{Bi}^2-2\Lambda_{Bi}\right)+\Lambda_{Ci}^2\sigma_x^2\right],$$

where Λ_{Bi} and Λ_{Ci} are the i^{th} 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 $\Lambda_{Bi}=1-\hat{\Lambda}_{ai}\Lambda_{ii}^{-1}$, where Λ_{ai} is the i^{th} 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 $B\Sigma_pB'$. The i^{th} eigenvalues of the matrix $\Sigma^{-1}CC'\sigma_x^2$, $\Lambda_{ii}^{-1}\Lambda_{Ci}^2\sigma_x^2$, are thus equal to $\Lambda_{ii}^{-1}\Lambda_{Bi}^2\Lambda_{pi}=\Lambda_{ii}^{-1}(1-\hat{\Lambda}_{ai}\Lambda_{ii}^{-1})^2\Lambda_{pi}$. Now we can rewrite X_i as:

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

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

$$
\begin{aligned}
X_i &= g^2(g+ax^2+x)^{-2} + ga^{-1}x^{-2}(ax^2+x)^2(g+ax^2+x)^{-2}, \\
&= g(g+ax^2+x)^{-2}[g+a^{-1}+2x+ax^2].
\end{aligned}
$$

Taking a partial derivative with respect to x, we get:

$$
\begin{aligned}
\frac{\partial X_i}{\partial x} &= -2g(g+ax^2+x)^{-3}[(g+a^{-1}+2x+ax^2)(2ax+1)-(g+ax^2+x)(ax+1)], \\
&= -2g(g+ax^2+x)^{-3}[a^2x^3+3ax^2+(3+ag)x+a^{-1}].
\end{aligned}
$$

The partial derivative is strictly negative because $g > 0$, $a > 0$, $x > 0$, 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 $lim_{x\to 0} X_i = 1+a^{-1}g^{-1}$, which equals $1+\rho^2\sigma_x^2\Lambda_{ii}$. 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 ($x = \Lambda_{\Psi i} = 0$). Note that this is consistent with $X_i = 1$ in partial equilibrium, where prices we taken as given ($\sigma_x^2 = 0$).

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

□

A.8 Proof of Proposition 12

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. $(\Gamma_i'E[f-pr])^2(\tilde{\Lambda}_i)^{-1} \geq (\Gamma_{i+1}'E[f-pr])^2(\tilde{\Lambda}_{i+1})^{-1}$. 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: $(\Gamma_1'E[f-pr])^2(\tilde{\Lambda}_1)^{-1} = (\Gamma_i'E[f-pr])^2(\tilde{\Lambda}_i)^{-1}$. 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 21, shows that, $rp_1 \to \mu$, which implies that $(\Gamma_1'E[f-pr])^2 \to 0$. 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. □

A.9 Proof of Proposition 13

We can rewrite equation (27) 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:

$$p_i = \frac{1}{r}\left(E_a[f_i] - \rho\, Cov_a[f_i, f_m]\right). \tag{28}$$

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:

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

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

$$E_a[r_m] - r = \rho\, p_m Cov_a[r_m, r_m]. \tag{30}$$

Solving (30) for the risk aversion coefficient ρ, and substituting it into (29), we get the second equation in the proposition.□