Relative Valuation and Information Production

Armando Gomes, Alan Moreira, and David Sovich*

April 26, 2018

Abstract

We study the problem of an investor that allocates analysts to assets to learn about future asset values. We show that when analysts are better at relative rather than absolute asset valuations the optimal matching of analysts to assets displays a balancedness property in which pairs of distinct assets are covered by a similar number of analysts. A balanced allocation allows the investor to efficiently aggregate information using the relative value between assets, eliminating the effect of the analyst-specific component. We show that the optimal matching of analysts to assets and the optimal portfolio decision depends on the structure of the analyst coverage network - the bipartite graph where the vertices are the firms and the edges are all the pairs of distinct firms that are covered by at least one common analyst. For example, capital is only reallocated between firms that are connected in the network, and the intensity of the reallocations depends on both the value of relative asset recommendations and the strength of the connection between the assets.

---

*Gomes and Sovich are from the Olin Business School, Washington University in St. Louis, and Moreira is from the Simon Graduate School of Business at the University of Rochester. The authors can be reached at gomes@wustl.edu, dsovich@wustl.edu, and alan.moreira@simon.rochester.edu. We thank seminar participants at Olin Business of Business, University of Rochester, University of California at Irvine, and Finance Theory Group for comments.

1. Introduction

How should investors aggregate information produced by a wide variety of approaches? Theory typically model the information acquisition problem as a cost function that is increasing in the precision of the acquired signal. In this paper, we go inside the information production black box and study how an investor decentralizes information production and aggregates information of a variety of sources. The key ingredient of our analysis is the idea that we are better at producing relative value signals. For example, a financial analyst that covers two different firms can provide a more informative signal about the relative value across firms, i.e. how much the first firm is overpriced relative to the second firm, than a signal about their absolute value, i.e how much each firm should worth. While it is well known that relative valuation is ubiquitous in finance, both among practitioners and academics, the implications are not completely understood.

In this paper, we study the importance of relative valuation from the perspective of an investor that must decide how to produce and aggregate information. The investor problem starts with the problem of how to allocate analysts to firms. Relative valuation implies that the production of information depends not only of the total number of hours analysts spend researching each firm, but on the entire information production network, the bipartite graph where the vertices are the firms and the edges are all the pairs of distinct firms that are covered by at least one common analyst. The entire information production network matters because it changes the information content of a signal produced by a particular information source.

To understand the importance of the information production network an example is useful. Consider three firms (A,B,C) and three analysts (1,2,3) with capacity to analyze two firms each and a total time budget of one day. When signals produced are absolute, lets say a better forecast of the firm discounted cash-flows, the following two arrangements are equivalent from the perspective of investors: one can concentrate one day of each analyst in each firm, or have each analyst learn about any two firms with a total time allocation of one day per firm. As a matter of fact, in the case of absolute valuation, there is no notion of the information production unit, the analyst is this example. All that matters in this case is the total allocation of analyst time per firm. When the information is relative the first arrangement would produce no information at all, and the information produced in the second arrangement would depend how exactly each analyst splits her time between firms.

While our theoretical analysis will focus on the portfolio problem of a stock market investor, the paper insights apply whenever signals generated by multiple information production units need to be aggregated by a principal. For example, Metrick and Yasuda (2010) show that General Partners in Private Equity and Venture Capital funds often are responsible for no more that four firms. In the credit market, Fracassi, Petry, and Tate (2016) show that credit analysts inside credit rating firms cover eleven firms on average. In the context of the stock market, Gomes, Gopalan, Leary, and Marcet (2016) document that financial analyst cover on average seven different firms. The fact that analysts concentrate on analyzing only a small subset of firms, together with the fact that analysts do not seem to produce informative absolute signals (for example, see Fracassi et al. (2016)) implies that the information production network is essential to the optimal aggregation of information.

Our models is as follows. An investor deploys agents whom can only produce information about a subset of assets; each agent produces a noisy signal of asset values containing an agent specific-component common across all assets followed by the agent. The variance of this agent specific component controls the degree of relative valuation. A very high variance of this agent specific component captures situations when agents have models that are so distinct that render comparison of signals produced my different agents uninformative. The investor chooses which assets each agent will cover and how much time each agent allocates to the assets being covered. The model has three periods: in period one, the investor chooses how to allocate agents to assets; in period two, the investor observes the signals produced by the agents, and then the investor makes his investment decision; finally, in period three, the investor receives the payoffs from his investments.

We begin the analysis by obtaining the optimal investment decision given the signals produced by the agents. We obtain closed-form expressions for the posterior mean and variance of asset returns as function of the signals and the information production network. Specifically, we show that the posterior covariance between two different stock returns is a function of the analyst connection strength between these firms. The posterior covariance is in turn the key determinant of how signals are interpreted. We show that a positive signal about firm i produced by analyst a implies positive (negative) information about firm j if the connection between firm i and j (i.e., the posterior covariance) is stronger (weaker) than the average connection between firm i and all the other firms covered directly by analyst a. Because the connection strength between any two firms depend on paths connecting both firms, the mapping from signal to information requires knowledge of the entire information production network.

These insights have sharp implications for how the investor responds to new information. First, we show that wealth is only reallocated within a connected component of the network. That is, the investor only reallocate his portfolio across firms that have a chain of analysts linking them. Second, we show that the extent to which investors make reallocations of capital across connected stocks depends on the the strength of the connection between the stocks. For example, a positive signal about a single firm i causes a reallocation of capital throughout firm i's entire network component. The firms which are most closely connected (but not linked) to firm i receive positive reallocations of capital when firm i receives a positive signal. Since no capital can leave the component, the capital is taken from firms whom which i is less connected. We show further that in the knife edge case that all the information is potentially learnable, the increase in the posterior covariance between connected (but not linked firms) exactly offsets the changes in expected returns due the information produced. In this case capital is only reallocated across firm covered by the same analyst.

In order to study the optimal allocation of information resources, we first characterize in closed form the investor ex-ante utility as function of the information production network. We show that investor's utility function can be expressed as a monotonically increasing and concave function of the information production network. The monotonicity and concavity properties allow us to show that there is a unique global optimum for information production network.

We find it useful to illustrate the implications of relative valuation for the optimal arrangement of information production resources by contrasting it with the absolute valuation case. We start by showing that relative valuation is a strong force for diversification in information production even when preferences are conductive to specialization. Relative valuation introduces complementaries in information production because the value of the signal produced about a single firm increases the more other firms' signals it can be compared with. Another distinctive implication of relative valuation is that the investor allocates analysts to pair of firms that are ex-ante expected to form profitable long-short portfolios. Specifically, they start by covering pairs that the return difference have the highest (uncentered) second moments.

Our analysis emphasizes relative valuation,the situation where comparisons of signals across analysts is completely uninformative. But our framework is flexible to consider less extreme cases. For example, our framework coincides with the popular Van Nieuwerburgh and Veldkamp (2010) model of information production and portfolio allocation in the extreme case where signals only have absolute information. We show how our framework can be used to study the choice of relative versus absolute information production. Specifically, we show how an increase in the analyst capacity to analyze multiple firms, for example due to a reduction in the fixed costs associated with information production, leads the investor to optimally shift production towards relative valuation. Intuitively, when the analyst can look at more firms, relative valuation becomes more powerful as it allows the investor to reallocate capital across more firms. This again emphasizes the important role that the information production unit plays in a relative valuation world.

Having studied the problem of how to allocate one single information production unit, we study the problem of how to design the entire information production network. We show that relative valuation introduces a strong force for balancedness in the network. We start by showing that when the firms are symmetric, the optimal network has exactly the same number of analysts covering any pairs of stocks, i.e. the network is said to be balanced. We then extend our analysis to show this property holds even in less symmetric environments. Specifically, we show that in a multi-industry setting that the information production network is block balanced, with intra-industry pairs having more analyst coverage than across-industry pairs.

The result that the optimal assignment should exhibit a balancedness property, can be illustrated by means of the following simple example. A balanced allocation is one where every pair of distinct assets is covered by exactly the same set of analysts. For example, suppose that we have 6 stocks and 10 analysts that can each cover 3 stocks. Then the unique balanced allocation is $\{123, 124, 135, 146, 156, 236, 245, 256, 345, 346\}$ where each triple *abc* denotes the stocks, labelled 1 to 6, covered by each of the 10 analysts. The structure is said to be balanced because each pair of stocks is covered by exactly 2 analysts (for example, stocks 4 and 5 are covered by analysts 7 and 9). Note that each asset is covered by 5 analysts, but there are many other allocations of 10 analysts to 6 stocks where each asset is covered by exactly 5 analysts, such as for example the unbalanced allocation $\{123, 123, 123, 123, 123, 456, 456, 456, 456, 456\}$. We show that under some parametrizations, the unique balanced allocation above yields 25% higher investor utility than the unbalanced structure. Economically, balanced allocations of analysts to assets help increase investor's utility because it improve the investor's ability to efficiently explore relative valuation.

Our paper contributes to several strands of literature. First, our paper is related to the literatures on information acquisition and investment (Veldkamp (2011)) and endogenous analyst network formation (Hong and Chang (2016)). Second, our paper highlights an intricate link between the literature on Bayesian portfolio choice (Black and Litterman (1992); Zhou (2009); Goffman and Manela (2012)) and the literature on the use of graph theory and networks in finance (Anton and Polk (2013), DeGroot (1974), Golub and Jackson (2010), Kelly et al. (2013)). We show how the analyst coverage network impacts information aggregation and portfolio choice in a Bayesian setting. In the optimal Bayesian investment strategy, reallocation across industries depends on the structure of the Laplacian matrix of the analyst coverage network. In addition, the strength of the connections within the network determines how to adjust the weights in the optimal portfolio in response to changes in analyst recommendations. One of the key contributions of our paper is to show that the structure of the coverage network provides the information necessary for this weighting on the information.

Third, our paper develops a portfolio approach that mitigates the known bias in analyst recommendations. A large literature documents that analyst recommendations may be biased because of career concerns (Hong and Kubik (2004)), investment banking relationships (Michaely and Womack (1999); Kadan, Madureira, Wang, and Zach (2009)), and preferences for stocks with certain quantitative characteristics (Jegadeesh, Kim, Krische, and Lee

(2004)). Our paper provides a formal method for efficiently extracting information about excess future returns even when analyst recommendations display systematic biases.

Fourth, our model helps explain some of the extant empirical findings related to analyst stock recommendations. Boni and Womack (2006) show that analysts create value only by ranking stocks within industries. Jegadeesh et al. (2004) find that the level of the consensus analyst recommendation contains no marginal predictive power about returns. In other words, the extant literature finds that the value of analyst recommendations comes from their ability to rank stocks relatively rather than absolutely. Our model admits this empirical finding. When analyst recommendations are biased and investors have uninformative priors, only relative valuations matter. Moreover, when all industries belong to disconnected components of the analyst coverage network, the optimal portfolio only reallocates wealth relatively among stocks within industries. Reallocation across industries only occurs when industries are "bridged" by a common analyst. This supports the ideas in Kadan, Madureira, Wang, and Zach (2012) and Boni and Womack (2006) that firm recommendations only contain information about industry level prospects when analysts use a market benchmark.

2. The Model

The model has three periods. In period $t = 1$, an investor chooses how to allocate m information production units (analysts) to n risky assets in order to produce information about period $t = 3$ asset values. In period $t = 2$, the investor receives the information produced by the analysts, and then chooses how to invest her wealth across n risky assets and one risk-free asset. In period $t = 3$, the investor receives the returns from her investment.

2.1 Assets

The n tradable assets in the economy are labeled $i = 1, .., n$. We denote by R_i the return from investing in asset i from period $t = 2$ to period $t = 3$, and by R the vector of asset returns. Throughout the analysis, we assume that asset prices are given and normalized to 1 in period $t = 2$. Furthermore, asset returns follow a normal distribution, $R \sim N(\bar{R}, \Sigma)$, with prior expected return $\bar{R}$ and variance Σ.

We decompose individual asset returns, R_i, into a learnable return component, r_i, and an unlearnable return component, η_i:

$$R_i = \bar{R}_i + r_i + \eta_i, \tag{1}$$

where $\mathrm{E}(R_i) = \bar{R}_i$ is the unconditional expected return. The term r_i captures the variation in fundamentals that can be learned by analysts through research, while the term η_i captures risk fundamentals that cannot be learned.[1] Both r_i and η_i are normally distributed with zero mean and independent distributions:

$$r \sim N(0, \Sigma_l) \text{ and } \eta \sim N(0, \Sigma_u) \text{ and } R \sim N(\bar{R}, \Sigma) \text{ with } \Sigma = \Sigma_l + \Sigma_u. \tag{2}$$

A case of particular interest is when only a fraction $\alpha \in [0, 1]$ of the unconditional return variation is learnable. In this case, the learnable component of returns has variance $\Sigma_l = \alpha\Sigma$ and the unlearnable component has variance $\Sigma_u = (1 - \alpha)\,\Sigma$. Our model collapses to a fully learnable case, where $\Sigma_l = \Sigma$ and $\Sigma_u = 0$, whenever $\alpha = 1$. This is the case that is most commonly studied in the literature (e.g., Van Nieuwerburgh and Veldkamp (2010)).

2.2 Information Structure

In period $t = 1$, the investor can deploy m analysts, labeled $a = 1, ..., m$, to produce information (signals) about period $t = 3$ asset returns. Specifically, each analyst a can produce a signal y_{ia} about the learnable component of asset returns, r_i:

$$y_{ia} = r_i + u_a + \varepsilon_{ia}, \tag{3}$$

where u_a is an analyst-specific error term, and ε_{ia} is an asset-analyst error term which is independent of u_a.

The term u_a can be conceptualized as a measurement error (or bias) that is common across all signals produced by an analyst.[2] We assume that u_a follows a normal distribution

[1] The introduction of an unlearnable component implies an upper bound to the Sharpe ratio the investor can obtain by producing information.

[2] For example, an analyst may be an optimist and consistently report upward biased signals.

with mean zero and precision ϕ_a, where the precision controls the degree of relative valuation.[3] Specifically, as the variance of u_a grows, analyst signals become less informative about the level of asset returns, while the information content of signal differences, i.e. relative comparisons, remains unchanged. We refer to the case when the analyst-specific error is very precise (i.e., when $\phi_a \to \infty$) as absolute valuation, and we refer to the case when the error is very imprecise (i.e., when $\phi_a = 0$) as relative valuation.

The asset-analyst error term, ε_{ia}, follows a normal distribution with mean zero and precision $\tau_a\theta_{ia}$, where the term $\tau_a > 0$ denotes the total precision of analyst a and $\theta_{ia} \geq 0$ denotes the fraction of time spent by analyst a researching asset i (s.t. $\sum_{i=1}^{n}\theta_{ia} = 1$). While τ_a is given exogenously, the investor chooses how to allocate the research efforts of the analysts (given by the set $\{\theta_{ia}\}_i^a$) in period $t = 1$. Therefore, $\tau_a\theta_{ia} \geq 0$ is the amount of analyst a's total precision that the investor allocates to asset i. We assume that the distribution of ε_{ia} is independent across i, and that each analyst a can produce information for at most $q_a < n$ assets.[4].

To summarize, the signal structure is given by:

$$u_a \sim N\left(0, \phi_a^{-1}\right) \text{ and } \varepsilon_{ia} \sim N\left(0, (\tau_a\theta_{ia})^{-1}\right) \text{ with } \sum_{i=1}^{n}\theta_{ia} = 1. \quad (4)$$

The signal structure can be represented in vector notation as $y_a = r + u_a 1 + \varepsilon_a$, where $\varepsilon_a \sim N(0, (\tau_a diag\,(\theta_a))^{-1})$ and ε_a and u_a are independent across analysts.[5]

2.3 Portfolio Choice and Utility

The investor has wealth normalized to W=1. In period $t = 2$, the investor observes the analyst signals $y = [y_1', ..., y_m']'$ and chooses portfolio weights $\omega = [\omega_1, ..., \omega_n]'$ for the n risky assets, with the remainder $(1 - \sum_{i=1}^{n}\omega_i)$ allocated to the risk-free asset. The investor's total

[3] As is standard in the literature, we define precision as the inverse of variance. That is, for a random variable x, precision(x) = variance$(x)^{-1}$.

[4] Such a constraint arises naturally in a setting with firm specific fixed costs in information production.

[5] The important aspect of this information structure is what it implies for the information content of the signal produced by the analyst. Furthermore, the signal content is invariant to any affine transformation of the signal y_a. Therefore, relative valuation does not imply that the actual volatility of the signals reported to investor by the analyst have infinite variance. In fact, because the model is invariant to any affine transformation, the variance of the signal is not determined. We thank David Hirshleifer for this point.

return, R_p, is equal to $R_p = \omega(y)'(R - r_f) + r_f$, where $\omega(y)$ is the portfolio choice conditional on the signals and r_f is the risk-free interest rate.

We consider cases in which the investor has either mean-variance utility or CARA utility. The investor's ex-ante utility is given by the expectation taken over the joint distribution of returns and signals:[6]

$$U = E\left[E\left[R_p|y\right] - \frac{\gamma}{2}Var\left[R_p|y\right]\right] \text{ Mean-variance preference} \tag{5}$$

$$U = -\frac{1}{\gamma}\log\left[E\left[\exp\left(-\gamma R_p\right)\right]\right] \text{ CARA preference} \tag{6}$$

3. The Portfolio Decision

3.1 Optimal Portfolio Weights and the Precision Matrix

We solve the model backwards starting from the period $t = 2$ asset allocation problem. The investor takes as given the signals produced by the analysts, and the optimal portfolio weights are obtained from the standard first-order condition:[7]

$$\omega^*(y) = \frac{1}{\gamma}\left(var\left(R|y\right)\right)^{-1}\left(E\left(R|y\right) - r_f\right). \tag{7}$$

The posterior mean, $E(R|y)$, and variance, $var(R|y)$, of returns are obtained from Bayesian updating. Their closed-form expressions are given by the following Proposition.

Proposition 1. *Suppose that each analyst $a = 1, ..., m$ produces a vector of signals $y_a = r + u_a 1 + \varepsilon_a$ about asset returns $R = \bar{R} + r + \eta$. Let the distributions of the random variables be given by: $r \sim N(0, \Sigma_l)$, $\eta \sim N(0, \Sigma_u)$, $u_a \sim N(0, \phi_a^{-1})$, and $\varepsilon_a \sim N(0, (\tau_a diag(\theta_a))^{-1})$, where each distribution is mutually pairwise independent and $\bar{R} = E(R)$. Then each analyst produces a quantity $\Theta_a = \tau_a\left(diag(\theta_a) - \frac{\tau_a}{\tau_a + \phi_a}\theta_a\theta_a'\right)$ of information about asset returns, and*

[6] See appendix B.1 for this derivation.

[7] The portfolio wights are the same for both preferences because we normalize prices and investor wealth to one. See appendix B.1 for details.

the precision matrix of signals is given by:

$$\Theta = \sum_{a=1}^{m} \Theta_a. \tag{8}$$

Moreover, the posterior mean and variance of asset returns is given by:

$$\begin{aligned} var(R|y) &= \hat{\Sigma}(\Theta) = \Sigma_u + \left(\Theta + \Sigma_l^{-1}\right)^{-1} \\ E(R|y) &= \left(\Theta + \Sigma_l^{-1}\right)^{-1} \left(\sum_{a=1}^{m} \Theta_a y_a\right) + \bar{R}. \end{aligned} \tag{9}$$

All proofs are in the Appendix. The Proposition highlights the importance of the precision matrix, Θ, in the investor's posterior beliefs and asset allocation decision. Critically, as Θ becomes "larger" or "more informative", the posterior variance declines and the posterior mean assigns more weight to the analyst signals (relative to the prior mean of $E(r) = 0$). This, in turn, refines the investor's asset allocation decision. Later in the paper, we show that more informative Θ's strictly increase the investor's ex-ante utility, and hence Θ plays a key role in the period $t = 1$ information production decision as well.

The precision matrix, Θ, is intricately linked to the degree of relative valuation in the economy. To see this, note that Equation 8 can be re-written as follows:

$$\Theta = \sum_{a=1}^{m} \left[\frac{\phi_a}{\tau_a + \phi_a} \underbrace{\left[\tau_a diag(\theta_a)\right]}_{\text{absolute valuation}} + \frac{\tau_a}{\tau_a + \phi_a} \underbrace{\left[\tau_a \left(diag(\theta_a) - \theta_a \theta_a'\right)\right]}_{\text{relative valuation}} \right]. \tag{10}$$

The first term in the summation, $\tau_a diag(\tau_a \theta_a)$, is a diagonal matrix with elements equal to the precision added by analyst a about each asset i. This matrix captures the amount of information produced about the level of asset returns. Due to the analyst-specific error term u_a, only a fraction $\frac{\phi_a}{\tau_a + \phi_a}$ of this information is learned by the investor. In the extreme case of absolute valuation (i.e, $\phi_a \to \infty$), the precision matrix is given by:

$$\Theta_A = \sum_{a=1}^{m} \tau_a diag(\theta_a) : \text{ (absolute valuation case when all } \phi_a = \infty). \tag{11}$$

The second term in the summation, $\tau_a \left(diag(\theta_a) - \theta_a\theta'_a\right)$, is a square matrix with column and row sums adding up to zero. For a fixed analyst, this matrix captures the amount of information produced about the relative value of the assets. Moreover, if an analyst only covers one asset, then the amount of information she produces about the relative value of the assets is zero.[8] In the extreme case of relative valuation (i.e, $\phi_a = 0$), the precision matrix contains is given by:

$$\Theta_R = \sum_{a=1}^{m} \tau_a \left(diag(\theta_a) - \theta_a\theta'_a\right) : (\text{relative valuation case when all } \phi_a = 0). \tag{12}$$

Whenever the analysts are symmetric (i.e., $\tau_a = \tau$ and $\phi_a = \phi$ for all a), the precision matrix can be expressed as a convex combination of the absolute and relative valuation cases:

$$\Theta^* = \left(\frac{\phi}{\tau + \phi}\right)\Theta_A + \left(\frac{\tau}{\tau + \phi}\right)\Theta_R. \tag{13}$$

3.2 Understanding the Portfolio Decision - Relative Valuation

We now analyze in more detail the investor's portfolio choice for the case of relative valuation: $\phi_a = 0$ for all a.[9] We begin by establishing a useful connection between relative valuation and the network induced by the allocation of analysts to assets. We then show that this network determines how the investor learns from observable signals, in addition to how she reallocates her capital across the assets. Knowing these facts are crucial for understanding the investor's optimal allocation of analysts to assets in period $t = 1$.

[8]Consider an allocation $\{\theta_{ia}\}_i^a$ in which each analyst a produces information about just one asset i:

$$\text{for all a: } \theta_{ia} \in \{0, 1\} \text{ subject to } \sum_i \theta_{ia} = 1.$$

In the case of absolute valuation, the investor learns a total of $\sum_a \tau_a\theta_{ia}$ information about each asset. In the case of relative revaluation, the investor does not learn any information about any of the assets: $diag(\theta_a) - \theta_a\theta'_a = 0 \; \forall a$. Economically, this allocation is useless to the investor because she does not have a "benchmark" to compare each analyst's signal to. She cannot separate the information about returns from the analyst-specific errors embedded within the signals.

[9]We analyze the less interesting case of absolute valuation in Section **??**.

3.2.1 The Information Production Network

The information production network is defined as the graph, G, where the vertices, $V(G)$, are the n assets and the edges, $E(G)$, are the pairs of distinct assets that are covered by at least one common analyst.[10] Two assets i and j are adjacement in the network, $i \sim j$, if there is a common analyst covering both assets - i.e., i and j are joined by an edge. More generally, two assets i and j are connected in the network if there exists a set of firms $i = l_1, ..., l_N = j$ such that $l_{k-1} \sim l_k$ for $k = 1, ..., N$. Connected assets that are adjacent are considered directly connected, while connected assets that are non-adjacent are considered indirectly connected. Only connected assets can be compared in a relative valuation setting.

Connections between assets facilitate the flow of information throughout the network. To measure the strength of direct connections between pairs of assets, we define the $n \times n$ weighted adjacency matrix, $\mathsf{A}(G)$, as follows:

$$\mathsf{A}(G)_{ij} = \begin{cases} 0 & \text{for } i = j \\ \sum_a \tau_a \theta_{ia} \theta_{ja} & \text{for } i \neq j \end{cases}$$

The value of $\mathsf{A}(G)_{ij}$ is increasing in (1) the number of common analysts covering both assets i and j and (2) the amount of analyst precision allocated to covering both assets. The matrix has null entries along its diagonal, and assets which are non-adjacent also satisfy $\mathsf{A}(G)_{ij} = 0$.

The weighted adjacency matrix can be used to measure the strength of indirect connections in the network as well.[11] To see this, let $\left[\mathsf{A}^k\right]_{ij}$ denote the ij^{th} element of $\mathsf{A}(G)$ raised to the k^{th} power. A well-known result from graph theory (e.g., Newman (2010)) states that $\left[\mathsf{A}^k\right]_{ij}$ is equal to the weighted number of paths of length k connecting vertices/assets i and j. Therefore, the sum $\sum_{k=1}^{\infty} \left[\mathbf{A}^k\right]_{ij}$ measures the strength of indirect connections between assets i and j in the network. Similarly, $\sum_{k=0}^{\infty} \left[\mathbf{A}^k\right]_{ij}$ measures the strength of all (direct and indirect) connections between i and j in the network.

[10] Formally, let $\mathcal{A}_{ij} = \{a \in \mathcal{A} : i \in \mathcal{N}_a \text{ and } j \in \mathcal{N}_a \}$ denote the set of all analysts covering both firms i and j, where $\mathcal{N}_a = \{i : \theta_{ia} > 0\}$ is the set of all assets covered by analyst a. The graph G consists of a set of vertices $V(G) = \mathcal{N} \equiv \{1, ..., n\}$ and a set of edges $E(G) = \{\{i, j\} \in \mathcal{N} \times \mathcal{N} : i \neq j \text{ and } \mathcal{A}_{ij} \neq \varnothing\}$.

[11] Two non-adjacent assets i and j can be indirectly connected if, for example, there is an asset l such that $i \sim l$ and $j \sim l$. More generally, two non-adjacent assets are indirectly connected if there is a sequence of assets $l_1, ..., l_k$, such that $i \sim l_1$, $l_1 \sim l_2$, ... , $l_{k-1} \sim l_k$, $l_k \sim j$.

We define the degree of an asset i as $d(i) = \sum_j \mathsf{A}(G)_{ij}$ and we define the $n \times n$ degree matrix as $\mathsf{D}(G) = diag(d(1), ..., d(n))$. Finally, we define the $n \times n$ (weighted) Laplacian matrix, $\mathsf{L}(G)$, as the difference between the degree matrix and the weighted adjacency matrix:

$$\mathsf{L}(G) = \mathsf{D}(G) - \mathsf{A}(G).$$

It is immediate to verify that the Laplacian matrix is equal to the precision matrix in the case of relative valuation:[12]

$$\Theta_R = \mathsf{L}(G) = \sum_a \tau_a \left(diag(\theta_a) - \theta_a \theta_a' \right).$$

These connections to graph theory will allow us to better our understanding of how the investor incorporates the signals into her posterior beliefs and portfolio decision. Further details are provided below.

3.2.2 Implications for Learning - Relative Valuation

The following Proposition allows us to interpret the posterior

Asset Correlations: The following proposition is helpful in interpreting the posterior correlations among assets in terms of the strength of connections between firms in the analyst coverage network $\mathsf{A}(G)$.

We show below that relative valuation creates positive correlation among assets. Moreover, we also show that posterior correlations among assets i and j are larger when the direct link between the two assets (as given by the weights $\mathsf{A}_{ij} = \sum_a \tau_a \theta_{ia} \theta_{ja}$) are strong plus there are many indirect links connecting the two assets.

Proposition 2. *Let $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$ and posterior variance matrix $var(R|y) = \widehat{\Sigma} = \left(\Sigma^{-1} + \Theta\right)^{-1}$, where $\Theta = \sum_a \tau_a \left(diag(\theta_a) - \theta_a \theta_a' \right)$ and let $A = \Theta - diag(\Theta)$ be the adjacency matrix (also given by $\mathsf{A}_{ij} = \sum_a \tau_a \theta_{ia} \theta_{ja}$, for $i \neq j$). Then:*
(i) The posterior correlations among assets are non-negative $\widehat{\Sigma}_{ij} \geq 0$, for all i and j, and moreover, $\widehat{\Sigma}_{ii} \geq \widehat{\Sigma}_{ij}$;

[12] Note that $\mathsf{D}(G) = diag(\Theta_R)$ because $d(i) = \sum_j \mathsf{A}(G)_{ij} = \sum_a \tau_a \theta_{ia} \sum_j \theta_{ja} = \sum_a \tau_a \theta_{ia} - \sum_a \tau_a \theta_{ia}^2$, and that $\mathsf{A}(G) = diag(\Theta_R) - \Theta_R$.

(ii) The posterior variance can be expressed as $\widehat{\Sigma} = N^{-1/2}\left(\sum_{k=0}^{\infty}\mathbf{A}^k\right)N^{-1/2}$, *where* $\mathbf{A} = N^{-1/2}\mathsf{A}N^{-1/2}$ *is the normalized adjacency matrix, and* $N = \Sigma^{-1} + diag(\Theta)$ *is a normalizing diagonal matrix.*

Thus the stronger (weaker) is the connection between two firms i *and* j*, the larger (smaller) is the posterior correlation among* i *and* j*.*

The Proposition above essentially uses that $(I - \mathbf{A})^{-1} = \sum_{k=0}^{\infty}\mathbf{A}^k$ which holds when all eigenvalues of $\mathbf{A}$ have norm less than one which we show holds in the proof (i.e., $\rho(\mathbf{A}) < 1$). Moreover, we know from graph theory (see for example, Newman (2010)), that the number of paths of length k connecting two vertices i and j is given by $\left[\mathbf{A}^k\right]_{ij}$ (i.e., the ij^{th} element of the k^{th} power of the adjacency matrix $\mathbf{A}$) and that the sum $\sum_{k=0}^{\infty}\left[\mathbf{A}^k\right]_{ij}$ is the total number of paths connecting i and j.

Thus Proposition 2 formally shows that we can interpret the posterior correlations between assets i and j as larger (or smaller) depending on how strong (or weak) are the connection between i and j, as measured by the number of the paths connecting i and j in the graph with adjacency matrix $\mathbf{A}$.

Posterior Expected Returns: The posterior mean of returns (9) can be expressed also as the sum of each analyst signal averaged by the precision of each signal multiplied by the posterior variance:

$$E(R|y) = \widehat{\Sigma}\left(\sum_{a=1}^{m}\Theta_a y_a\right) + \bar{R},$$

We now show that how investor update the mean returns after observing investors' signals depend both on direct and indirect connections. The posterior expected return of asset i is greater when either asset i receives a favorable analyst recommendation y_{ia}, or when an asset j, which asset i is strongly connected to, receives a favorable analyst recommendation y_{ja}. In other words, each signal provided by the analysts affects the expected returns all the assets for which they cover. We state this result below.

Proposition 3. *The sensitivity of returns to asset recommendations satisfy*

$$\frac{\partial E\left(R|y\right)_j}{\partial y_{ia}} = \tau_a \theta_{ia} \left(\widehat{\Sigma}_{ji} - \sum_k \widehat{\Sigma}_{jk} \theta_{ka} \right).$$

The following comparative statics results hold:

(i) The asset i return increase when any analyst following asset i revise it upwards. That is $\frac{\partial E(R|y)_i}{\partial y_{ia}} \geq 0$.

(ii) When the strength of the connection between asset j and asset i is stronger (weaker) than the average strength of the connection among asset j and the other assets covered by analyst a, then asset j return increases when an analyst a revise asset i upwards. Formally, $\frac{\partial E(R|y)_j}{\partial y_{ia}} \gtrless 0$ *if and only if* $\widehat{\Sigma}_{ji} \gtrless \sum_k \widehat{\Sigma}_{jk} \theta_{ka}$.

3.2.3 Implications for Portfolio Choice

We now show how the information network provides insight about how capital is reallocated in response to asset recommendations. Remind that capital is reallocated based on the information produced as follows:

$$\frac{\partial \omega^*\left(y\right)}{\partial y_{ia}} = \frac{1}{\gamma} \left(\Sigma_u + \widehat{\Sigma} \right)^{-1} \frac{\partial E\left(R|y\right)}{\partial y_{ia}}. \tag{14}$$

In the information network graph G, two firms i and j are defined as *connected* if and only if there is a *path* connecting them. That is, there is a distinct set of firms $i = i_0, i_1, i_2, ...i_m = j$ such that i_{k-1} and i_k share a common analyst for $k = 1, ..., m$. Note that connection we formally define before is an equivalence relation. A graph, G, is *connected* if any two firms can be joined by a path, and is otherwise disconnected. A maximal connected *subgraph* of the graph is defined as a (connected) *component*, where a subgraph is any graph S formed from a subset of the vertices and edges of G. We will also refer to the *components* of the graph as the maximal equivalence classes of the connection relation.

We show below that no capital flows in or out of the separate connected components of the information production network. This result directly relates to the concept of relative valuation. Since asset analyst recommendations have no level anchor, investors can only make inferences about the relative value of assets that are evaluated by the same analyst or through a chain of analysts. For example, suppose the assets in one component receive on average more positive signals than the assets in second component Then investors are unable to infer whether the assets in first component have higher expected returns, or if the signals produced in second component have a different level, i.e. information produced by completely disconnected information sources cannot be combined. Only when there is a connection between the assets can inferences be made about expected returns and wealth reallocated.

Proposition 4. *(Wealth Reallocations) Consider the pure relative valuation setting with $\Sigma_u = (1-\alpha)\Sigma$ and $\Sigma_l = \alpha\Sigma$, for some $\alpha \in [0,1]$, which includes the full-learnable case. Then:*

(i) The assets can be partitioned into connected (and disjoint) components of the graph G defined above;

(ii) There is no reallocation of wealth among disconnected components. Formally, let $\omega(y)$ be the optimal portfolio choice upon learning signal y and let $\omega^{no-learn}$ be the portfolio choice under no learning. The wealth allocated to each connected component, say component $\mathcal{G}$, satisfy $\sum_{i\in\mathcal{G}} \omega(y) = \sum_{i\in\mathcal{G}} \omega_i^{no-learn}$, for all possible signals y;

(iii) In particular, there is no reallocation of wealth between risky and riskless assets, i.e. $\sum_{i\in\mathcal{N}} \omega(y) = \sum_{i\in\mathcal{N}} \omega_i^{no-learn}$ for all signals y.

In the knife-edge case of full learning, the wealth reallocations take a particularly simple form because $\left(\Sigma_u + \widehat{\Sigma}\right)^{-1} \widehat{\Sigma} = I$ when $\Sigma_u = 0$. In response to a signal received by analyst a for asset i, the optimal investment on assets $j \neq i$ is

$$\frac{\partial \Delta\omega_j}{\partial y_{ia}} = -\frac{1}{\gamma}\tau_a\theta_{ia}\theta_{ja} \leq 0 \text{ and } \frac{\partial \Delta\omega_i}{\partial y_{ia}} = \frac{1}{\gamma}\tau_a\left(\theta_{ia} - \theta_{ia}^2\right) \geq 0. \tag{15}$$

From equation (15) there is a reallocation of wealth across assets in response to signals in the presence of relative valuation. In particular, in response to a signal received by analyst a for asset $i \neq j$, the optimal loading on asset j marginally decreases. Note that the decrease is larger the more analyst a follows both assets i and j, and if the analyst does not follow both assets there is not reallocation between the assets.

Therefore the intensity of capital reallocation depends on the strength of the analyst following connection between assets. Whenever, the information between the pair of assets i and j, $\tau_a \theta_{ia} \theta_{ja}$ is high, good news about asset i leads the investor to pull more capital from asset j.

To conclude this section, no capital flows across connected components. In addition, the extent of intra-component capital reallocations depends on *both* the strength of connections and the value of asset recommendations. A favorable signal about a firm propagates throughout the entire component, and causes capital to reallocate to the assets in the component to which the firm shocked is most closely connected.

4. Investors Ex-ante Preferences over Alternative Information Production Networks

In the this section we focus on characterizing the investor ex-ante utility associated with allocations of information producing resources. Specifically in the Proposition 5 below we obtain the ex-ante expected utility and show that it depends essentially only on the information matrix Θ for both the CARA and mean-variance investor.

Proposition 5. *Let the prior excess return vector be $\mu = \bar{R} - r_f \mathbf{1}$ and variance of returns $R \sim N\left(\bar{R}, \Sigma\right)$ with learnable and unlearnable variances Σ_l and Σ_u. Whenever the information matrix about the learnable component is Θ, so that the posterior precision is $\hat{\Sigma}^{-1}(\Theta) = \left(\Sigma_u + \left(\Sigma_l^{-1} + \Theta\right)^{-1}\right)^{-1}$, then the ex-ante investor utility is:*

(i) In the CARA preference case, the ex-ante utility is

$$U(\Theta) = \frac{1}{2\gamma}\left(\log\det\left(\hat{\Sigma}^{-1}(\Theta)\Sigma\right) + \mu'\Sigma^{-1}\mu\right). \tag{16}$$

(ii) In the mean-variance preference case, the ex-ante utility is

$$U(\Theta) = \frac{1}{2\gamma}\left(Tr\left(\hat{\Sigma}^{-1}(\Theta)\Sigma\right) + \mu'\hat{\Sigma}^{-1}(\Theta)\mu - n\right). \tag{17}$$

Note that in full learning case, i.e. $\Sigma_u = 0$, the posterior precision is simply $\hat{\Sigma}^{-1}(\Theta) = \Sigma^{-1} + \Theta$, and thus linear in the information matrix Θ. Moreover, in the mean-variance preference case with full-learning, the ex-ante utility is also linear in the information matrix,

$$U(\Theta) = \frac{1}{2\gamma}\left(Tr(\Theta\Sigma) + \mu'\Theta\mu + \mu'\Sigma^{-1}\mu\right). \tag{18}$$

Thus, the investor optimal solution in this special case is determined without any regards for interdependencies among analysts' choices.

In all other cases, there are important interdependencies among analysts' information production choices. We show below the posterior precision matrix $\hat{\Sigma}^{-1}(\Theta)$ and the investor ex-ante utility $U(\Theta)$ are strictly monotonic and strictly concave mappings of the information matrix Θ. We introduce below definitions to formalize these concepts:

- *(Informativeness)* Matrix Θ is more informative than Θ^*, i.e. $\Theta \succeq \Theta^*$, if the matrix $\Theta - \Theta^*$ is positive semidefinite.

- *(Concavity)* The utility function and posterior precision mapping are *concave* if $U(\Theta_\lambda) \geq \lambda U(\Theta) + (1-\lambda)U(\Theta^*)$ and $\hat{\Sigma}^{-1}(\Theta_\lambda) \succeq \lambda\hat{\Sigma}^{-1}(\Theta) + (1-\lambda)\hat{\Sigma}^{-1}(\Theta^*)$, for any $\lambda \in [0,1]$ and $\Theta_\lambda = \lambda\Theta + (1-\lambda)\Theta^*$.

- *(Monotonicity)* The utility function and posterior precision mapping are *monotonic* if $U(\Theta) \geq U(\Theta^*)$ and $\hat{\Sigma}^{-1}(\Theta) \succeq \hat{\Sigma}^{-1}(\Theta^*)$, for any $\Theta \succeq \Theta^*$.

We provide more details in the Appendix B, including the natural extensions to strict concavity and monotonicity.

Lemma 1. *Consider the posterior precision mapping* $\hat{\Sigma}^{-1}(\Theta) = \left(\Sigma_u + \left(\Sigma_l^{-1} + \Theta\right)^{-1}\right)^{-1}$ *and the ex-ante utility* $U(\Theta)$*. Then:*
(i) The posterior precision mapping is strictly concave and strictly monotonic in the partial learning case (with Σ_u *and* Σ_l *invertible); and it is linear and strictly monotonic in the full learning case (with* $\Sigma_u = 0$*).*
(ii) The investor ex-ante utility function $U(\Theta)$ *of a CARA or mean-variance investor is strictly monotonic and strictly concave in the information matrix* Θ *[except that in the mean variance and full learning case* $\Sigma_u = 0$*, it is linear in the information matrix* Θ*].*

The strict concavity for the CARA utility arises both from the strict concavity of the log determinant mapping $X \rightarrowtail \log\det(X)$ and the strict concavity of the posterior precision with partial learning, while the stricty concavity for the mean-variance utility arises only from the latter.

We will show later on how these general strict monotonicity and concavity properties imply the strict optimally of balanced designs among analysts (see Section 6).

Rather than use the investor utility directly, it will save on notation going forward to focus on the utility gain relative to no-information learning,

$$\mathcal{U}(\Theta) = U(\Theta) - \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu, \tag{19}$$

subtracting away the constant utility term $\frac{1}{2\gamma}\mu'\Sigma^{-1}\mu$ due to no-learning.

The evaluation of investor utility in applications is made easier by the following result which provides the investor utility gain as a function of the eigenvalues of the weighted information matrix $\Theta\Sigma$.

Lemma 2. *Consider the partial learning setting, $\Sigma_u = (1-\alpha)\Sigma$ and $\Sigma_l = \alpha\Sigma$, for some $\alpha \in [0,1]$. Define the strictly increasing and concave function $f(x)$ by*

$$f(x) = \frac{1+\alpha x}{1+\alpha(1-\alpha)x}. \tag{20}$$

The ex-ante utility gain, in the CARA preference case, is

$$\mathcal{U}(\Theta) = \frac{1}{2\gamma}\sum_{i=1}^{n}\log f(\lambda_i). \tag{21}$$

where $\lambda_i \geq 0$ are the eigenvalues of the weighted information matrix $\Theta\Sigma$.

The proof is in the Appendix, where we also provide the similar expression for the mean-variance case in terms of the eigenvalues of the weighted information matrix.

5. Relative versus Absolute Valuation

In this section we contrast the properties of information production under relative and absolute valuation. To keep the analysis as simple as possible and better provide the intuition for the results we consider the mean-variance preference and full-learning in this section.

5.1 Broad Coverage versus Specialization Decision

Using the properties that the relative valuation component of the information matrix is the Laplacian of the graph we can establish a series of identities that gives us a more intuitive understanding of the how relative valuation works.

Proposition 6. *Suppose an analyst produce information about q assets $i = 1, ..., q$ with attention θ and precision ϕ and τ, so that the precision matrix is $\Theta = \tau \left(diag(\theta) - \frac{\tau}{\tau+\phi} \theta\theta' \right)$. Then the incremental utility gain, $\mathcal{U} = \frac{1}{2\gamma} \left(Tr \left(\Sigma\Theta \right) + \mu'\Theta\mu \right)$, can be expressed as*

$$\mathcal{U} = \frac{1}{2\gamma} \left[\underbrace{\tau \sum_{i,j=1:i<j}^{q} \theta_i\theta_j \left[\left(var(r_i - r_j) + (\mu_i - \mu_j)^2 \right) \right]}_{\mathcal{U}_R} \frac{\tau}{\phi+\tau} + \underbrace{\tau \sum_{i=1}^{2} \theta_i \left(var(r_i) + \mu_i^2 \right)}_{\mathcal{U}_A} \frac{\phi}{\phi+\tau} \right]. \tag{22}$$

The utility gain is the weighted average of the gain coming from absolute valuation and relative valuation related to the respective components of the precision matrix. Intuitively, under relative valuation the utility gain is proportional to the sum of variances of the long-short portfolio comparing all pair of assets weighted by the precision $\theta_i\theta_j$ allocated to the pair. These results follow directly from the fact that the precision matrix Θ_R is the Laplacian matrix of the information production network.

We now show that relative valuation increases the gains of diversification in the production of information. Consider a symmetric setting such as an industry with variance $\Sigma = \sigma^2 (I + \rho J)$ and equal excess returns $\mu = \bar{R}_i - r_f$ so that $var(r_i - r_j) = 2\sigma^2$ and $(\mu_i - \mu_j)^2 = 0$ for all pair of assets. This implies that the summation appearing in the expression for the utility gain is

$$\sum_{i,j=1:i<j}^{q} \theta_i \theta_j \left[\left(var(r_i - r_j) + (\mu_i - \mu_j)^2 \right) \right] = \sigma^2 \sum_{i,j=1:i<j}^{q} 2\theta_i \theta_j = \sigma^2 \left[\left(\sum_{i=1}^{q} \theta_i \right)^2 - \sum_{i=1}^{q} \theta_i^2 \right].$$

Since $\sum_{i=1}^{q} \theta_i = 1$ then the optimal time allocation problem is the unique solution of:

$$\begin{array}{cc} \max_\theta & 1 - \sum_{i=1}^{q} \theta_i^2 \\ \text{s.t.} & \sum_{i=1}^{q} \theta_i = 1 \text{ and } \theta_i \geq 0 \end{array}.$$

Therefore spreading out equally the time allocation to each asset, $\theta_i = \frac{1}{q}$, is a unique global optimal allocation of time by each analyst whenever relative valuation is important (i.e., $\phi < \infty$). The maximum value is $\sigma^2 \left(1 - \sum_{i=1}^{q} \frac{1}{q^2}\right) = \sigma^2 \left(\frac{q-1}{q}\right)$. Note that the time allocation above is the unique optimal regardless of the choice of precision ϕ and τ. The proposition below makes this point formally.

Proposition 7. *Consider the optimal information production design of a mean-variance investor and the assets have covariance $\Sigma = \sigma^2 (I + \rho J)$ and equal excess returns $\mu = \bar{R}_i - r_f$, and analyst can cover up to q assets with precision τ and ϕ at a convex cost $c(\tau, \phi)$. Then the analyst optimally spread their allocation of time equally across all q assets, i.e., $\theta_{ia} = \frac{1}{q}$, and the optimal relative and absolute valuation precision ϕ^* and τ^* are given by the unique global solution of the utility maximization problem*

$$\mathcal{U} = \frac{\tau}{2\gamma} \left(\left(\sigma^2 (1 + \rho) + \mu^2 \right) \frac{\phi}{\tau + \phi} + \sigma^2 \left(\frac{q-1}{q} \right) \frac{\tau}{\tau + \phi} \right) - c(\tau, \phi)$$

subject to $\tau, \phi \geq 0$. Moreover, whenever $\phi^* < \infty$ this information production choice is the unique global optimum.

With absolute valuation the gain in utility is composed of the sum of the utility coming from $\mu' \Theta_A \mu$, which is equal to $\tau \mu^2$, and the gain in utility coming from $Tr(\Sigma \Theta_A) = \tau \sigma^2 (1 + \rho)$ which is the total volatility times the total precision. With strict relative valuation the gain in utility is $Tr(\Sigma \Theta_R) = \tau \sigma^2 \left(\frac{q-1}{q}\right)$ equal to the idiosyncratic asset volatility times the total precision times the fraction $\left(\frac{q-1}{q}\right)$ lost due the fact that only relative signals have information. For intermediate cases the utility gain is the average of the absolute and relative valuation cases with weights respectively $\frac{\phi}{\tau+\phi}$ and $\frac{\tau}{\tau+\phi}$.

The results above illustrate that in general it will be optimal for an analyst to sacrifice some specialization and precision in exchange for a broader and (more imprecise) asset coverage. The utility gain attributed to the relative valuation is proportional to $\tau \times \left(\frac{q-1}{q}\right)$. Thus if analysts could cover more assets without any overall precision loss there would be an utility gain. However, the marginal gain is decreasing with the asset coverage. In particular, increasing coverage from 2 to 3 assets increases utility from $\tau \times \left(\frac{1}{2}\right)$ to $\tau \times \left(\frac{2}{3}\right)$, which is an utility gain of 33%, but the utility gain from increasing coverage from 9 to 10 assets would only be 1%.

More generally it is likely that the overall precision is decreasing in the total number of assets covered as information production is likely to have substantial fixed costs associated with leaning about a particular firm. Specifically, let us assume that the total analyst precision is decreasing in the number of assets covered, $\tau'(q) < 0$, then the decision of how many firms to cover would be determined by $\max_q \tau(q) \times \left(\frac{q-1}{q}\right)$. The first order condition yields that the optimal quantity is a declining function of the precision elasticity, $q^* = 1 - (\partial log(\tau(q))/\partial log(q))^{-1}$. In particular, if the total quality of the information decays very slowly, the analyst will optimally choose to cover a large number of firms.

5.2 Asymmetric Assets: Long-Short Portfolios

We now address the optimal design under asymmetric assets in the mean-variance utility case. We focus on the case that the analyst can only choose two assets, where the investor expected utility is as follows

$$\mathcal{U} = \frac{1}{2\gamma}\left[\underbrace{\tau\theta_i\theta_j\left[\left(var(r_i - r_j) + (\mu_i - \mu_j)^2\right)\right]}_{\mathcal{U}_R}\frac{\tau}{\phi+\tau} + \underbrace{\tau\left(\sum_{i=1}^{N^{\text{`}}}\theta_i\left(var(r_i) + \mu_i^2\right)\right)}_{\mathcal{U}_A}\frac{\phi}{\phi+\tau}\right],$$

In the case where there is only relative valuation the analyst choose the assets i and j

to cover, and always optimally choose to dedicate equal attention to them, i.e. $\theta_i = \theta_j = \frac{1}{2}$, and choose the pair of assets in order to maximize $\max_{i,j} var(r_i - r_j) + (\mu_i - \mu_j)^2$, i.e. the pair of assets with the long-short portfolio with the largest (uncentered) second moment. In, contrast in the case of absolute valuation, the analyst should dedicate all attention to the asset that maximizes $\max_i var(r_i) + \mu_i^2$. In the Appendix B, we generalize this result for a setting where the analyst covers more than two assets ($q > 2$).

6. Optimality of Balanced Designs

Having studied in Section 5 the problem of how to allocate one single information production unit, now we study the problem of how to design the entire information production network. We show that relative valuation introduces a strong force for balancedness in the network. We start by showing that when the firms are symmetric, the optimal network has exactly the same number of analysts covering any pairs of stocks, i.e. the network is said to be balanced. We then extend our analysis to show this property holds even in less symmetric environments. Specifically, we show that in a multi-industry setting that the information production network is block balanced, with intra-industry pairs having more analyst coverage than across-industry pairs.

6.1 Balanced Designs

We first start introducing the formal definition of balanced allocation.

Definition: *(Balanced Network) Consider a triple (n, m, q) where n denotes the number of assets, m the number of agents, and $q < n$ the maximum number of assets that an agent can cover. Let N_a be the subsets of assets that agents $a = 1, ..., m$ are covering. We say that the coverage is balanced, or a balanced design, if all subsets N_a have exactly q assets, and every pair of assets is covered by exactly λ agents.*

Consider the following example to illustrate this point.

Example 1: Consider a symmetric economy with $n = 6$ assets and $m = 10$ analysts which can cover $q = 3$ assets, and let the matching of analysts to assets be $\mathcal{A} = \{123, 124, 135, 146, 156, 236, 245, 256, 345, 346\}$, where we denote the subset of assets followed by each of the $m = 10$ analysts by triples abc. The structure is a balanced design with $\lambda = 2$ analysts covering each pair of assets. Note that each asset is followed by exactly $c = 5$ analysts. There can be multiple structures with the same allocation of analyst coverage per asset but with variation in the amount of analyst coverage per pair. For example consider the allocation allocation of resources $\mathcal{B} = \{123, 123, 123, 123, 123, 456, 456, 456, 456, 456\}$. The total utility of investors with the first structure, the balanced alloaction, is more than 25% higher than the second structure (see details in Appendix B).

A central results of our paper is that the maximization of investor's utility is uniquely achieved with balanced allocations of information production resources.

Proposition 8. *(Optimality of symmetric balanced designs) Consider the symmetric problem above where the investor have CARA preference (with full or partial learning) or mean-variance preference (with partial learning) with n assets and m agents that can cover $q \leq n$ assets with prior return $R - r_f\mathbf{1} \sim N(\mu\mathbf{1}, \Sigma)$ where $\Sigma = \sigma^2 I + \sigma_f^2 J$ such that there exists a balanced design.Then the most efficient allocation among all possible feasible allocations is a balanced design in which all analysts choose the same precision τ and ϕ. Whenever at the optimum $\phi < \infty$, then the balanced design is the unique maximum efficient allocation of resources.*

We establish in our next result the closed-form solution for the investor utility under CARA preferences when using a balanced allocation of information production resources.

Proposition 9. *(Investors' utility under a balanced design) Suppose the investor has CARA preference, there are n assets with prior excess return with variance $\Sigma = \sigma^2(I + \rho J)$, and full or partial learning with $\Sigma_u = (1-\alpha)\Sigma$ and $\Sigma_l = \alpha\Sigma$, for some $\alpha \in [0, 1]$. Suppose that all m analysts choose precision τ and ϕ and each can produce information about q assets and let the agents be organized according to a balanced design with $\lambda = \frac{mq(q-1)}{n(n-1)}$ analysts per pair of assets, and $c = \frac{mq}{n}$ analysts per asset. Then:*
(i) The precision of the signal obtained by investors is the $n \times n$ matrix Θ equal to

$$\Theta = \frac{\tau m}{n}\left[\frac{\phi}{\tau+\phi}I + \frac{\tau}{\tau+\phi}\frac{n(q-1)}{(n-1)q}\left(I - \frac{1}{n}J\right)\right], \tag{23}$$

where I and J are, respectively, the $n \times n$ identity matrix and matrix of ones in all entries.

(ii) The expected investor utility gain with a balanced design is equal to:

$$\mathcal{U}(\tau, \phi) = \frac{1}{2\gamma}\left((n-1)\log\left(f\left(\lambda_1\right)\right) + \log f\left(\lambda_2\right)\right) \tag{24}$$

where $f(\cdot)$ is give by equation (20), and λ_1 and λ_2 are the eigenvalues of the weighted information matrix $\Theta\Sigma$ given by:

$$\begin{aligned} \lambda_1(\tau, \phi) &= \sigma^2 \frac{m}{n} \frac{\tau}{\tau + \phi}\left(\phi + \tau \frac{n(q-1)}{(n-1)q}\right): \text{ (with multiplicity } n-1\text{)}, \\ \lambda_2(\tau, \phi) &= \sigma^2 \frac{m}{n} \frac{\tau\phi}{\tau + \phi}(1 + n\rho): \text{ (with multiplicity } 1\text{)}. \end{aligned} \tag{25}$$

(iii) There is a unique τ^* and ϕ^* that maximize the utility net of costs $\mathcal{U}(\tau, \phi) - mc(\tau, \phi)$ given by (24) subject to $\tau, \phi \geq 0$. This solution is such that:
(a) an increase in q leads to more weight being placed on relative valuation $\partial\left(\frac{\tau^*}{\phi^*}\right)/\partial q \geq 0$;
(b) an increase in asset correlation ρ leads to less weight being placed on relative valuation $\partial\left(\frac{\tau^*}{\phi^*}\right)/\partial\rho \leq 0$.

The optimal precision τ^* and ϕ^* maximize the utility $\mathcal{U}(\tau, \phi)$ given by (24) subject to $\tau, \phi \geq 0$. Any other allocation with an information matrix different than the one given by (23) yields strictly less utility to the investor.

Note that the term $\frac{(q-1)}{q(n-1)}$ multiplying the relative valuation part of the information matrix is increasing in q, the number of assets analyst can follow. Thus if analyst can spread out the same total precision τ across multiple assets the information matrix becomes more informative. However, the rate of utility gain is decreasing in the number of assets covered and for a large number of assets q, such as $q = 10$ assets, over 90% of the possible gains are already achieved.

6.2 Information Production within and across Industries: Block Balanced Designs

While analysts tend to be industry specialists, there is also a significant across-industry coverage by analysts. For example, Boni and Womack (2006, Table 3) document that sellside analysts cover approximately 10 firms and, on average, 76% of the companies covered by an analyst are in the same industry. We show now that relative valuation provides a new rational for the existence of across-industry analyst coverage.

We assume that analysts have to spend a fixed amount of time to understand an industry in order to produce an informative signal about assets belonging to that industry. This fixed production cost generate incentives to become industry specialists, and we show that, in the case of absolute valuation, analysts would be fully specialized in an industry. However, in the context of relative valuation, there are incentives for some analysts to follow assets across industries, otherwise different industries would be in separate disconnected components, and investors would not be able to make significant reallocations or comparisons across industries.

To formalize the concepts, in a simple setting suppose there are two industries each with n_j firms in each industry, and returns are $R_{ij} = \varepsilon_{ij} + f_j + f$, where f_j is a common industry factor and f is a common aggregate factor, and ε_{ij} are idyiosincratic risks with the same variance. Assume that analysts can only produce signals for assets after spending time to learn about each industry the assets covered belong; and there is a decay in total analyst precision as she covers more industries since there is less time remaining to learn about individual assets. Formally, if the q assets covered by the analyst are in n_I different industries, the total precision is $\tau\left(1 - n_I \kappa\right)$ and the information matrix contributed by each analyst in the absolute and relative valuation cases are, respectively

$$\Theta_a = \tau\left(1 - n_I \kappa\right) diag(\theta_a) \text{ (absolute valuation)}$$
$$\Theta_a = \tau\left(1 - n_I \kappa\right)\left(diag(\theta_a) - \theta_a \theta_a'\right) \text{ (relative valuation)}$$

Proposition 10. *Consider a setting where: (i) the investors either have a CARA preference (with full or partial learning) or mean variance preference (with partial learning); (ii) there aretwo industries $j = 1, 2$ each with n_j firms; (iii) returns are $R_{ij} = \varepsilon_{ij} + f_j + f$, where f_j is a common industry factor and f is a common aggregate factor; and (iv) there are m analysts producing information about q assets with total precision $\tau\left(1 - \kappa\right)$ or $\tau\left(1 - 2\kappa\right)$ depending on whether assets covered belong to one or two industries. Then:*

(i) In the absolute valuation setting, all analysts are industry specialists if $\kappa > 0$.

(ii) In the relative valuation setting, there will be always be some cross-industry analysts if there are enough analysts (formally, for each $\kappa > 0$, there exists $\underline{m}$ such that there are cross-industry analysts for all $m \geq \underline{m}$).

In the absolute valuation setting, there is no role for cross-industry analysts. And to the extent that covering assets in multiple industries reduce the overall precision of the assets covered we would not expect in the model above cross-industry coverage to exist.

The proof essentially consist of analyzing the eigenvalues of $\Theta\Sigma$, which by Lemma 2 determines the utility gain. Any allocation without cross-industry analysts result in information network with at least two disconnected components, and the weighted information matrix $\Theta\Sigma$ has at least two zero eignenvalues. Cross-industry analysts, by connecting the industries, add another strictly positive eigenvalue. The marginal benefits of each industry analyst converges to zero as the number of analysts increase, thus the addition of cross-industry analysts strictly dominates a structure with only industry analysts, with a large enough number of analysts.

Example: block balance This example illustrates that the coverage choice can be important part of the analyst contribution to investors. Consider the following information production structures with 4 analyst covering 8 assets $\mathcal{A} = \{1234, 1234, 5678, 5678\}$. A fifth analyst that can cover $q = 4$ assets can choose to cover assets 1234 (or 5678) such as the other 4 analysts, or assets 1235, or cover assets 1256. We show below that the later choice is the optimal choice and the one that maximizes investors' welfare. This choice improves the connectivity among assets the most and leads to a more balanced coverage with more interconnection among assets.

Using the parametrization $\left(v = 2, \alpha = \frac{1}{2}, \phi = 10, \tau = 10\right)$ the following are the associated investors' utility with each structure:

$$\begin{aligned} & \Delta\mathcal{U}(\mathcal{A} + 1234) = 0.226 \\ \mathcal{U}(\mathcal{A}) = 4.285 \rightarrow & \Delta\mathcal{U}(\mathcal{A} + 1235) = 0.465 \\ & \Delta\mathcal{U}(\mathcal{A} + 1256) = 0.525 \end{aligned}$$

An analyst by making choices that are more balanced, can improve the connectivity of assets and the performance of relative valuation. The more balanced choice can more than double her potential contribution to investors as illustrated above.

7. Conclusion

In this paper, we study the implication of relative valuation for how information is processed and how information producing resources should be organized. We study this problem from the perspective of an investor that must decide how to produce and aggregate information. The investor problem starts with the problem of how to allocate analysts to firms. Relative valuation implies that the production of information depends not only of the total number of hours analysts spend researching each firm, but on the entire information production network, the bipartite graph where the vertices are the firms and the edges are all the pairs of distinct firms that are covered by at least one common analyst. The entire information production network matters because it changes the information content of a signal produced by a particular information source.

We then use our framework to derive implications for portfolio choice and the optimal organization of the information production network. We find that relative valuation is a strong force for diversification in information production.

References

Black, F., and R. Litterman, 1992, Global portfolio optimization, *Financial Analysts Journal* 28–43.

Boni, L., and K. Womack, 2006, Analysts, industries, and price momentum, *Journal of Financial and Quantitative Analysis* 41, 85–109.

Fracassi, Cesare, Stefan Petry, and Geoffrey Tate, 2016, Does rating analyst subjectivity affect corporate debt pricing?, *Journal of Financial Economics* 120, 514–538.

Goffman, M., and A. Manela, 2012, An empirical evaluation of the black-litterman approach to portfolio choice, *Working Paper* .

Gomes, Armando R, Radhakrishnan Gopalan, Mark T Leary, and Francisco Marcet, 2016, Analyst coverage network and corporate financial policies .

Hong, H., and B. Chang, 2016, Stock market coverage, *Working Paper* .

Hong, H., and J. Kubik, 2004, Analyzing the analysts: Career concerns and biased earnings forecasts, *The Journal of Finance* 58, 1083–1124.

Jegadeesh, N., S. Kim, S.D. Krische, and C.M.C. Lee, 2004, Analyzing the analysts: When do recom mendations add value?, *Journal of Finance* 59, 1083–1124.

Kadan, O., L. Madureira, R. Wang, and T. Zach, 2009, Conflicts of interest and stock recommendations: The effects of the global settlment and related regulations, *Review of Financial Studies* 22, 4190–4217.

Kadan, O., L. Madureira, R. Wang, and T. Zach, 2012, Analysts' industry expertise, *Journal of Accounting and Economics* 54, 95–120.

Metrick, Andrew, and Ayako Yasuda, 2010, The economics of private equity funds, *The Review of Financial Studies* 23, 2303–2341.

Michaely, R., and K. Womack, 1999, Conflict of interest and the credibility of underwriter analyst recommendations, *Review of Financial Studies* 12, 653–686.

Newman, M.E.J., 2010, *Networks: An Introduction* (Oxford University Press).

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information acquisition and under-diversification, *The Review of Economic Studies* 77, 779–805.

Veldkamp, L., 2011, Information choice in macroeconomics and finance .

Zhou, G., 2009, Beyond black-litterman: Letting the data speak, *The Journal of Portfolio Management* 36, 36–45.

A. Appendix

Appendix A: Proofs

Proof of Proposition 1: We can write the signals in vector form as $y = Xr + \varsigma$ where $\varsigma = Bu + \varepsilon$, by stacking up all analyst signals. Let the $nm \times n$ dimensional indicator matrices $X = 1_m \otimes I_n$ (the Kronecker product of the vector of m ones and the n-dimensional identity matrix) and the $nm \times m$ dimensional indicator matrix $B = I_m \otimes 1_n$. The signal distribution, conditional on the return r, is:

$$y \sim N\left(Xr, \Sigma_\varsigma\right)$$

where

$$\Sigma_\varsigma = B\Sigma_u B' + \Sigma_\varepsilon \text{ and } Q = \Sigma_\varsigma^{-1}$$

Moreover,

$$r \sim N\left(0, \Sigma_l\right)$$

Lindley and Smith (1972) prove that the distribution of r, given y, is

$$r|y \sim N\left(Zw, Z\right)$$

with

$$\begin{aligned} Z^{-1} &= X'\Sigma_\varsigma^{-1}X + \Sigma_L^{-1} \\ w &= X'\Sigma_\varsigma^{-1}y \end{aligned}$$

Rewriting we get

$$\begin{aligned} var(r|y) &= \left(X'QX + (\Sigma_l)^{-1}\right)^{-1} \\ E(r|y) &= \left(X'QX + (\Sigma_l)^{-1}\right)^{-1} X'Qy. \end{aligned}$$

which implies (9).

By a similar argument we obtain the posterior distribution for u. Let $y = Bu + \delta$ where $\delta = Xr + \varepsilon$ with variance $\Sigma_\delta = X'\Sigma_l X + \Sigma_\varepsilon$. We have that $y \sim N\left(Bu, \Sigma_\delta\right)$ and that $u \sim N\left(0, \Sigma_u\right)$. Applying Lindley and Smith (1972) theorem yields $u|y \sim N\left(Td, T\right)$ where $T^{-1} = B'\Sigma_\delta^{-1}B + \Sigma_u^{-1}$ and $d = B'\Sigma_\delta^{-1}y$. Thus

$$\begin{aligned} var(u|y) &= \left(B'\Sigma_\delta^{-1}B + \Sigma_u^{-1}\right)^{-1} \\ E(u|y) &= \left(B'\Sigma_\delta^{-1}B + \Sigma_u^{-1}\right)^{-1} B'\Sigma_\delta^{-1}y. \end{aligned}$$

with $\Sigma_\delta^{-1} = \left(X'\Sigma_l X + \Sigma_\varepsilon\right)^{-1}$.

Note that if we ignore the prior distributions for r and u (i.e., consider uninformative priors with $\Sigma_u^{-1} \to 0$ and $\Sigma_l^{-1} \to 0$), we get the classical GLS estimates for r and u :

$$\begin{aligned} var(r_{GLS}) &= (X'Q_1X)^- \\ r_{GLS} &= (X'Q_1X)^- X'Q_1y, \end{aligned}$$

where

$$Q_1 = \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}B\left(B'\Sigma_\varepsilon^{-1}B\right)^{-1}B'\Sigma_\varepsilon^{-1}.$$

Note that by the Woodbury formula

$$Q_1 = \lim \Sigma_\varsigma^{-1} = \lim\left(B\Sigma_uB' + \Sigma_\varepsilon\right)^{-1} = \lim \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}B\left(\Sigma_u^{-1} + B'\Sigma_\varepsilon^{-1}B\right)^{-1}B'\Sigma_\varepsilon^{-1}.$$

Also, the estimate for the analyst-specific component is

$$\begin{aligned} var(u_{GLS}) &= (B'Q_2B)^- \\ u_{GLS} &= \left(B'Q_2B + \Sigma_u^{-1}\right)^{-1}B'Q_2y, \end{aligned}$$

where

$$Q_2 = \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}X\left(X'\Sigma_\varepsilon^{-1}X\right)^{-1}X'\Sigma_\varepsilon^{-1}.$$

Again by the Woodbury formula

$$Q_2 = \lim \Sigma_\delta^{-1} = \lim\left(X'\Sigma_lX + \Sigma_\varepsilon\right)^{-1} = \lim \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}X\left(\Sigma_l^{-1} + X'\Sigma_\varepsilon^{-1}X\right)^{-1}X'\Sigma_\varepsilon^{-1}.$$

(ii) The matrix Q has a block diagonal structure $Q = diag(Q_1, ..., Q_m)$, composed of the m blocks given explicitly by the $n \times n$ matrices

$$Q_a = \tau_a\left(diag\left(\theta_a\right) - \frac{\tau_a}{\phi_a + \tau_a}\theta_a\theta_a'\right) \tag{26}$$

Indeed, let $\zeta_{ia} = u_a + \varepsilon_{ia}$, where $cov(\zeta_{ia}, \zeta_{ja}) = \phi_a^{-1}$ for all $i \neq j$. In matrix notation $var\left(\zeta\right) = \Sigma_\varepsilon + B\Sigma_uB'$ where we remind that $\Sigma_u = diag(\phi_1^{-1}, ..., \phi_n^{-1})$ and $\Sigma_\varepsilon = diag\left(\Sigma_1, ..., \Sigma_m\right)$ with blocks $\Sigma_a = \tau_a^{-1}diag\left(\theta_{1a}^{-1}, ..., \theta_{na}^{-1}\right)$. Therefore, the matrix $\Sigma_\varepsilon + B\Sigma_uB'$ is an $nm \times nm$ matrix with block diagonal structure with blocks of dimension n given by

$$\tau_a^{-1}diag\left(\theta_{1a}^{-1}, ..., \theta_{na}^{-1}\right) + \phi_a^{-1}1_n1_n'.$$

Thus the matrix $Q = \left(\Sigma_\varepsilon + B\Sigma_uB'\right)^{-1}$ also has a block diagonal structure. By the Sherman–Morrison formula each block Q_a is

$$Q_a = \tau_a diag\left(\theta_a\right) - \frac{\phi_a^{-1}\tau_a^2}{1 + \phi_a^{-1}\tau_a1_n'diag\left(\theta_a\right)1_n}diag\left(\theta_a\right)1_n1_n'diag\left(\theta_a\right),$$

where $diag\left(\theta_a\right) = diag\left(\theta_{1a}, ..., \theta_{na}\right)$. Because $1_n'diag\left(\theta_a\right)1_n = 1$ and $diag\left(\theta_a\right)1_n = \theta_a$ this yields equation (26).

Finally, to obtain the precision matrix $\Theta = X'QX$ note that $X = 1_m \otimes I_n$ then $\Theta = X'QX$ becomes

$$\Theta = \left[\begin{array}{ccc} I_n & \cdots & I_n \end{array} \right] \left[\begin{array}{ccc} Q_1 & 0 & 0 \\ 0 & Q_a & 0 \\ 0 & 0 & Q_m \end{array} \right] \left[\begin{array}{c} I_n \\ \vdots \\ I_n \end{array} \right] = \sum_{a=1}^{m} Q^{(a)}.$$

Therefore,

$$\Theta = \sum_{a=1}^{m} \tau_a \left(diag\left(\theta_a\right) - \frac{\tau_a}{\phi_a + \tau_a} \theta_a \theta_a' \right).$$

Observe that

$$diag\left(\theta_a\right) - \frac{\tau_a}{\phi_a + \tau_a} \theta_a \theta_a' = \frac{\phi_a}{\tau_a + \phi_a} diag\left(\theta_a\right) + \frac{\tau_a}{\tau_a + \phi_a} \left(diag(\theta_a) - \theta_a \theta_a' \right)$$

which shows that the information matrix can be expressed as the weighted sum:

$$\Theta = \sum_{a=1}^{m} \tau_a \left[\frac{\phi_a}{\tau_a + \phi_a} diag(\theta_a) + \frac{\tau_a}{\tau_a + \phi_a} \left(diag(\theta_a) - \theta_a \theta_a' \right) \right].$$

Q.E.D.

Proof of Proposition 2: (i) Let $N = \Sigma_l^{-1} + diag(\Theta)$, thus $\Sigma_l^{-1} + \Theta = N - \mathsf{A}$ and $\widehat{\Sigma} = (N - \mathsf{A})^{-1}$.

Observe that $N = diag(\Theta) \geq 0$ and $\mathsf{A} \geq 0$. The matrix $N - \mathsf{A}$ is a nonsingular M-matrix. Indeed, its off-diagonal elements are non-positive and moreover, the matrix $N - \mathsf{A}$ is strictly dominant diagonal,

$$(N - \mathsf{A})_{ii} = \left[\Sigma_l^{-1}\right]_{ii} + \sum_a \tau_a \left(\theta_{ia} - \theta_{ia}^2\right) >$$
$$> \sum_a \tau_a \theta_{ia} \left(1 - \theta_{ia}\right) = \sum_a \sum_{j \neq i} \tau_a \theta_{ia} \theta_{ja} = \sum_{j \neq i} |\left[N - \mathsf{A}\right]_{ij}|.$$

Any nonsingular M-matrix is invertible and has a non-negative inverse, which proves that the inverse of $(N - \mathsf{A})$ exists and is nonnegative. In order to show that the inverse satisfy $\widehat{\Sigma}_{ii} \geq \widehat{\Sigma}_{ij}$ for all i and j, we use the following lemma (Berman and Plemmons, 1994, pp. 254): Let W be a nonsingular M-matrix whose row sum are all nonnegative. Then $V = (W)^{-1}$ satisfy $V_{ii} \geq V_{ij}$ for all i and j.

(ii) We can normalize $\widehat{\Sigma}$ multiplying by the diagonal matrix $N^{-1/2}$,

$$\widehat{\Sigma} = (N - \mathsf{A})^{-1} = N^{-1/2} \left(I - N^{-1/2} \mathsf{A} N^{-1/2}\right)^{-1} N^{-1/2} = N^{-1/2} \left(I - \mathbf{A}\right)^{-1} N^{-1/2}.$$

But we established above that the spectrum radius of $\mathbf{A}$ is less than one, $\rho(\mathbf{A}) < 1$, which implies that

$$(I - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \mathbf{A}^k,$$

which completes the proof.

Q.E.D.

Proof of Proposition 3: The conditional return is given by $E(R|y) = \widehat{\Sigma}\left(\sum \Theta_a y_a\right) + \bar{R}$, where $\widehat{\Sigma} = \left((\alpha\Sigma)^{-1} + \Theta\right)^{-1}$, is a linear function of the asset recommendations. The derivative with respect to y_{ia} is equal to

$$\frac{\partial E(R|y)}{\partial y_{ia}} = \widehat{\Sigma} v$$

where $v = [\Theta_a]_{\cdot i} \in \mathbb{R}^n$ is the i-th column of the matrix $\Theta_a = \tau_a \left(diag(\theta_a) - \theta_a \theta_a'\right)$. Thus $v_j = -\tau_a \theta_{ia} \theta_{ja}$ for $j \neq i$ and $v_i = \tau_a \theta_{ia}(1 - \theta_{ia})$.

Therefore,

$$\begin{aligned} \frac{\partial E(R|y)_j}{\partial y_{ia}} &= \left(\widehat{\Sigma} v\right)_j = \sum_k \widehat{\Sigma}_{jk} v_k \\ &= \widehat{\Sigma}_{ji} v_i + \sum_{k \neq i} \widehat{\Sigma}_{jk} v_k \\ &= \tau_a \theta_{ia} \left(\widehat{\Sigma}_{ji}(1 - \theta_{ia}) - \sum_{k \neq i} \widehat{\Sigma}_{jk} \theta_{ka}\right) \\ &= \tau_a \theta_{ia} \left(\widehat{\Sigma}_{ji} - \sum_k \widehat{\Sigma}_{jk} \theta_{ka}\right) \end{aligned}$$

Moreover, $\frac{\partial E(R|y)_j}{\partial y_{ia}} \gtrless 0$ if and only if $\widehat{\Sigma}_{ji} \gtrless \sum_k \widehat{\Sigma}_{jk} \theta_{ka}$. Also $\frac{\partial E(R|y)_i}{\partial y_{ia}} = \tau_{ia} \left(\sum_k \left(\widehat{\Sigma}_{ii} - \widehat{\Sigma}_{ik}\right) \theta_{ka}\right) \geq 0$ because $\widehat{\Sigma}_{ii} \geq \widehat{\Sigma}_{ik} \geq 0$, which completes the proof.

Q.E.D.

Proof of Proposition 4: (i) In the information network graph G, two firms i and j are defined as *connected* if and only if there is a *path* connecting them. Connection is an equivalence relation. A maximal connected *subgraph* of the graph is defined as a (connected) *component*, where a subgraph is any graph S formed from a subset of the vertices and edges of G. Each *components* of the graph is the maximal equivalence classes of the connection relation.

Consider G_k a connected component. Let $1_k \in \mathbb{R}^n$ be vector with 1's for each asset belonging to a connected component G_k and zero otherwise.

We first show that

$$1_k^T \Theta = 0 \text{ and } 1_k^T \Theta_a = 0.$$

for each conneted component.

(ii) The portfolio choice is

$$\omega^* (y) = \frac{1}{\gamma} (var (R|y))^{-1} (E (R|y) - r_f). \tag{27}$$

where the posterior mean $E (R|y)$ and the variance $var (R|y)$ are:

$$\begin{aligned} E(R|y) &= \left(\Theta + \Sigma_l^{-1}\right)^{-1} \left(\sum_{a=1}^{m} \Theta_a y_a \right) + \bar{R}, \\ var(R|y) &= \left(\Sigma_l^{-1} + \Theta\right)^{-1} + \Sigma_u \end{aligned} \tag{28}$$

where $\Sigma_u = (1 - \alpha) \Sigma$ and $\Sigma_l = \alpha \Sigma$. Note that

$$\begin{aligned} \Sigma_l^{-1} \Sigma_u &= \alpha^{-1} (1 - \alpha) I \\ \Sigma_l^{-1} &= \alpha^{-1} \Sigma^{-1} \end{aligned}$$

The total amount invested in assets beloging to a connected component G_k is given by $\sum_{i \in G_k} \omega_i (y) = 1_k^T \omega (y)$, where $1_k \in \mathbb{R}^n$ is defined in (i).

Consider

$$\begin{aligned} 1_k^T \omega (y) = 1_k^T \left(\left(\Sigma_l^{-1} + \Theta\right)^{-1} + \Sigma_u \right)^{-1} \left(\Theta + \Sigma_l^{-1}\right)^{-1} \left(\sum_{a=1}^{m} \Theta_a y_a \right) + \\ + 1_k^T \left(\left(\Sigma_l^{-1} + \Theta\right)^{-1} + \Sigma_u \right)^{-1} \left(\bar{R} - r_f \mathbf{1}\right) \end{aligned}$$

Thus

$$\begin{aligned} \left(\left(\Sigma_l^{-1} + \Theta\right)^{-1} + \Sigma_u \right)^{-1} &= \left(\Sigma_l \left(I + \Theta \Sigma_l\right)^{-1} + \Sigma_u\right)^{-1} \\ &= \left(\left(I + \Theta \Sigma_l\right)^{-1} + \Sigma_l^{-1} \Sigma_u\right)^{-1} \Sigma_l^{-1} \\ &= \left(\left(I + \Theta \Sigma_l\right)^{-1} + \Sigma_l^{-1} \Sigma_u\right)^{-1} \Sigma_l^{-1} \\ &= \left(\left(I + \Theta \Sigma_l\right)^{-1} + \alpha^{-1} (1 - \alpha) I\right)^{-1} \alpha^{-1} \Sigma^{-1} \end{aligned}$$

Therefore if 1_k^T is a left eigenvector with eigenvalue zero then

$$1_k^T \left(\left(I + \Theta \Sigma_l\right)^{-1} + \alpha^{-1} (1 - \alpha) I\right)^{-1} \alpha^{-1} \Sigma^{-1} = 1_k^T$$

because

$$\left(1+\alpha^{-1}\left(1-\alpha\right)\right)^{-1}\alpha^{-1}=1$$

:We now show that 1_k^T is a left eigevector with eigenvalue zero of the matrix

$$\left(\left(\Sigma_l^{-1}+\Theta\right)^{-1}+\Sigma_u\right)^{-1}\left(\Theta+\Sigma_l^{-1}\right)^{-1}=\left(I+\left(\Theta+\Sigma_l^{-1}\right)\Sigma_u\right)^{-1}$$
$$=\left(1+\left(\alpha^{-1}\left(1-\alpha\right)\right)I+\Theta\left(1-\alpha\right)\Sigma\right)^{-1}$$

which is true since 1_k^T is a left eigevector with eigenvalue zero of matrix Θ.

Since $1_k^T\Theta_a=0$ we conclude the proof.

Q.E.D.

Proof of Proposition 5: We obtain below the ex-ante expected utility for signals with precision matrix Θ for both the CARA and mean-variance investor. The ex-ante utility are explicitly given by:

$$U\left(\Theta\right)=-\frac{1}{\gamma}\log\left(E\left[\exp\left(-\frac{1}{2}\left(E\left(R|y\right)-r_f\mathbf{1}\right)'var\left(R|y\right)^{-1}\left(E\left(R|y\right)-r_f\mathbf{1}\right)\right)\right]\right):\text{CARA}$$
$$U\left(\Theta\right)=\frac{1}{2\gamma}E\left[\left(E\left(R|y\right)-r_f\mathbf{1}\right)'var\left(R|y\right)^{-1}\left(E\left(R|y\right)-r_f\mathbf{1}\right)\right]:\text{Mean-variance}$$

Define the conditional expectation by $X=E\left(R|y\right)-r_f\mathbf{1}$.

In order to take the expectation above we use the following result: if $X\sim N(\mu,\Sigma_X)$ is a n-dimensional random vector with mean μ and variance Σ_X then

$$E\left[\exp\left(-\frac{1}{2}X'\Omega X\right)\right]=\det\left(I+\Sigma_X\Omega\right)^{-1/2}\exp\left(-\frac{1}{2}\mu'\Omega\mu+\frac{1}{2}\mu'\Omega'\left(I+\Sigma_X\Omega\right)^{-1}\Sigma_X\Omega\mu\right)$$
$$E\left[X'\Omega X\right]=Tr(\Omega\Sigma_x)+\mu'\Omega\mu$$

Using the law of expectation,

$$\mu=E\left[X\right]=E\left[E\left(R|y\right)-r_f\mathbf{1}\right]=E\left(R\right)-r_f\mathbf{1}=\bar{R}-r_f\mathbf{1}.$$

We have seen before that

$$\Omega=var\left(R|y\right)^{-1}=\hat{\Sigma}^{-1}=\left(\Sigma_u+\left(\Sigma_l^{-1}+\Theta\right)^{-1}\right)^{-1}.$$

To obtain the conditional variance $\Sigma_X=Var\left[E\left(R|y\right)-r_f\mathbf{1}\right]=Var\left[E\left(R|y\right)\right]$, the key step is to use the law of total variance, instead of trying to compute it directly. By the total law of variance,

$$Var\left(R\right)=E\left[Var\left(R|y\right)\right]+Var\left[E\left(R|y\right)\right].$$

Taking into account that $R = r+\eta$, we have that $Var\left(R\right) = \Sigma = \Sigma_l+\Sigma_u$ and $E\left[Var\left(R|y\right)\right] = Var\left(R|y\right) = \hat{\Sigma}$ as given above. Thus

$$\Sigma_X = Var\left[E\left(R|y\right)\right] = \Sigma_l + \Sigma_u - \left(\Sigma_u + \left(\Sigma_l^{-1} + \Theta\right)^{-1}\right),$$
$$\Rightarrow \Sigma_X = \Sigma_l - \left(\Sigma_l^{-1} + \Theta\right)^{-1} = \Sigma - \hat{\Sigma}.$$

In addition,

$$\Sigma_X - \Sigma = \Sigma_X - \Sigma_l - \Sigma_u = -\Sigma_u - \left(\Sigma_l^{-1} + \Theta\right)^{-1},$$

and thus,

$$\left(\Sigma_X - \Sigma\right)\Omega = -I$$
$$\Rightarrow I + \Sigma_X\Omega = \Sigma\Omega$$

Taking into account that $\Omega' = \Omega$, and that $\Sigma_X\Omega = \Sigma\Omega - I$, we have that

$$\Omega'\left(I + \Sigma_X\Omega\right)^{-1}\Sigma_X\Omega = \Omega\left(\Sigma\Omega\right)^{-1}\left(\Sigma\Omega - I\right) = \Omega - \Sigma^{-1}.$$

Combining these results in the expression for ex-ante utility,

$$\det\left(I + \Sigma_X\Omega\right)^{-1/2}\exp\left(-\frac{1}{2}\mu'\Omega\mu + \frac{1}{2}\mu'\Omega'\left(I + \Sigma_X\Omega\right)^{-1}\Sigma_X\Omega\mu\right)$$
$$= \left[\det\left(\Sigma\Omega\right)\right]^{-1/2}\exp\left(-\frac{1}{2}\mu'\Omega\mu + \frac{1}{2}\mu'\left(\Omega - \Sigma^{-1}\right)\mu\right)$$
$$= \left[\det\left(\Sigma\Omega\right)\right]^{-1/2}\exp\left(-\frac{1}{2}\mu'\Sigma^{-1}\mu\right).$$

Thus, the certainty-equivalent utility of the CARA investor is

$$U\left(\Theta\right) = -\frac{1}{\gamma}\log\left(\left[\det\left(\Sigma\Omega\right)\right]^{-1/2}\exp\left(-\frac{1}{2}\mu'\Sigma^{-1}\mu\right)\right)$$
$$\Rightarrow U\left(\Theta\right) = \frac{1}{2\gamma}\left(\log\left(\det\left(\Sigma\hat{\Sigma}^{-1}\right)\right) + \mu'\Sigma^{-1}\mu\right).$$

For the mean variance investor, the utility is

$$U\left(\Theta\right) = \frac{1}{2\gamma}\left(Tr(\hat{\Sigma}^{-1}\Sigma_X) + \mu'\hat{\Sigma}^{-1}\mu\right)$$
$$= \frac{1}{2\gamma}\left(Tr(\Sigma\hat{\Sigma}^{-1} - I) + \mu'\hat{\Sigma}^{-1}\mu\right)$$
$$= \frac{1}{2\gamma}\left(Tr(\Sigma\hat{\Sigma}^{-1}) + \mu'\hat{\Sigma}^{-1}\mu - n\right).$$

It is straightfoward that in the full-learnable case, when $\Sigma_u = 0$ and $\Sigma_l = \Sigma$, then $\hat{\Sigma}^{-1} = \left(\left(\Sigma^{-1} + \Theta \right)^{-1} \right)^{-1} = \Sigma^{-1} + \Theta$, which implies the equation (18).

Q.E.D.

Proof of Lemma 1: Formally, the statement we will prove is: Let Θ_0 and Θ_1 be two positive semi-definite information matrices and let $\lambda \in [0,1]$ and $\Theta = \lambda\Theta_1 + (1-\lambda)\,\Theta_0$. The posterior precision mapping is (strictly) concave if

$$\hat{\Sigma}^{-1}(\Theta) \succeq \lambda \hat{\Sigma}^{-1}(\Theta_1) + (1-\lambda)\,\hat{\Sigma}^{-1}(\Theta_0) \tag{29}$$

(w/ inequality holding strictly for $\lambda \in (0,1)$ and $\Theta_0 \neq \Theta_1$).

The posterior precision mapping is (strictly) monotonic if for $\Theta \succeq \Theta^*$:

$$\hat{\Sigma}^{-1}(\Theta) \succeq \hat{\Sigma}^{-1}(\Theta^*) \tag{30}$$

(w/ inequality holding strictly for $\Theta \succ \Theta^*$).

(i.a) By the Woodbury identity

$$\hat{\Sigma}^{-1}(\Theta) = \left(\left(\Sigma_u^{-1}\right)^{-1} + \left(\Sigma_l^{-1} + \Theta\right)^{-1} \right)^{-1} = \Sigma_u^{-1} - \Sigma_u^{-1}\left(\Sigma_u^{-1} + \Sigma_l^{-1} + \Theta\right)^{-1}\Sigma_u^{-1}.$$

But the mapping $X \rightarrowtail -X^{-1}$ is operator monotonic and concave on the set of positive definite matrices (see also Bhatia (1991, Ch. 5), and the mapping $X \rightarrowtail \Sigma_u^{-1} - \Sigma_u^{-1}\left(\Sigma_u^{-1} + \Sigma_l^{-1} + X\right)^{-1}\Sigma_u^{-1}$ is strictly concave and monotonic on the set of positive semidefinite matrices (see also Bhatia (2007, Corollary 1.5.3).

(i.b) The mapping $X \rightarrowtail X^{-1}$ is (strictly) order-reversing over the set of positive definite matrices. Thus for $\Theta \succ \Theta^*$ then $\left(\Sigma_l^{-1} + \Theta\right)^{-1} \prec \left(\Sigma_l^{-1} + \Theta^*\right)^{-1}$ and thus $\Sigma_u + \left(\Sigma_l^{-1} + \Theta\right)^{-1} \prec \Sigma_u + \left(\Sigma_l^{-1} + \Theta^*\right)^{-1}$. Applying the mapping $X \rightarrowtail X^{-1}$ once again yields $\hat{\Sigma}^{-1}(\Theta) \succ \hat{\Sigma}^{-1}(\Theta^*)$.

In the full learning case, i.e., $\Sigma_u = 0$, then $\hat{\Sigma}^{-1}(\Theta) : \Theta \rightarrowtail \left(\Sigma_l^{-1} + \Theta\right)$ is linear in Θ and is trivially strictly monotonic.

(ii.a) *CARA preference case.* Let A and B be two positive semi-definite information matrices and let $\lambda \in [0,1]$ and $\Theta = \lambda A + (1-\lambda)\,B$.

Our goal is to prove the following two formal results:

$$\text{(i)}\ \log\det\left(\hat{\Sigma}^{-1}(\Theta)\,\Sigma\right) \geq \lambda \log\det\left(\hat{\Sigma}^{-1}(A)\,\Sigma\right) + (1-\lambda)\log\det\left(\hat{\Sigma}^{-1}(B)\,\Sigma\right), \tag{31}$$

(w/ inequality holding strictly for $\lambda \in (0,1)$ and $A \neq B$) and that

$$\text{(ii)}\ \log\det\left(\hat{\Sigma}^{-1}(\Theta)\,\Sigma\right) \geq \log\det\left(\hat{\Sigma}^{-1}(\Theta^*)\,\Sigma\right), \tag{32}$$

(w/ inequality holding strictly for $\Theta > \Theta^*$).

The log determinant mapping $X \rightarrowtail \log\det(X)$ is a strictly concave function on the convex set of positive definite matrices X (see Theorem 7.6.7 in Horn and Johnson (pg

466)). It is also a strictly monotonic function because $\log\det(X) = \sum_{i=1}^{n} \log(\lambda(X)_i)$, where $\lambda(X)_i$ are the eigenvalues of X (and if $X > X^*$ then $\lambda(X)_i > \lambda(X^*)_i$ for at least one eigenvalue).

We use below the following result, which follows directly from the Sylvester law of inertia.

Result HJ: *(Horn and Johnson (pg 223)) For any non-singular positive definite matrix Σ, if $\Theta \succeq Q$ then the matrix $\Sigma^{1/2}P\Sigma^{1/2} \succeq \Sigma^{1/2}Q\Sigma^{1/2}$, and if $P \succ Q$ then $\Sigma^{1/2}P\Sigma^{1/2} \succ \Sigma^{1/2}Q\Sigma^{1/2}$.*

Property (29) combined with the result above yields

$$\Sigma^{1/2}\hat{\Sigma}^{-1}(\Theta)\Sigma^{1/2} \succeq \lambda\Sigma^{1/2}\hat{\Sigma}^{-1}(A)\Sigma^{1/2} + (1-\lambda)\Sigma^{1/2}\hat{\Sigma}^{-1}(B)\Sigma^{1/2}.$$

Therefore, by the monotonicity property of the log determinant,

$$\log\det\left(\Sigma^{1/2}\hat{\Sigma}^{-1}(\Theta)\Sigma^{1/2}\right) \geq \log\det\left(\lambda\Sigma^{1/2}\hat{\Sigma}^{-1}(A)\Sigma^{1/2} + (1-\lambda)\Sigma^{1/2}\hat{\Sigma}^{-1}(B)\Sigma^{1/2}\right).$$

By the concavity property of the log determinant mapping we have

$$\begin{aligned}
&\log\det\left(\lambda\Sigma^{1/2}\hat{\Sigma}^{-1}(A)\Sigma^{1/2} + (1-\lambda)\Sigma^{1/2}\hat{\Sigma}^{-1}(B)\Sigma^{1/2}\right) \geq \\
&\lambda\log\det\left(\Sigma^{1/2}\hat{\Sigma}^{-1}(A)\Sigma^{1/2}\right) + (1-\lambda)\log\det\left(\Sigma^{1/2}\hat{\Sigma}^{-1}(B)\Sigma^{1/2}\right) = \\
&\lambda\log\det\left(\hat{\Sigma}^{-1}(A)\Sigma\right) + (1-\lambda)\log\det\left(\hat{\Sigma}^{-1}(B)\Sigma\right).
\end{aligned}$$

which combined implies property (31). Strict concavity follows from the strict concavity of the log determinant mapping and the Result HJ above.

The strict monotonicity also holds because of the strict monotonicity of the log determinant and the property (30) which we proved above.

(ii.b) *Mean-variance preference case.* The trace $X \rightarrowtail Tr(X\Sigma)$ is a linear and strictly monotonic mapping. Therefore, from property (29) and (30) we have

$$\begin{aligned}
Tr\left(\hat{\Sigma}^{-1}(\Theta)\Sigma\right) &\geq Tr\left(\lambda\hat{\Sigma}^{-1}(A)\Sigma + (1-\lambda)\hat{\Sigma}^{-1}(B)\Sigma\right) = \\
&\lambda Tr\left(\hat{\Sigma}^{-1}(A)\Sigma\right) + (1-\lambda)Tr\left(\hat{\Sigma}^{-1}(B)\Sigma\right).
\end{aligned}$$

Moreover, from property (29) we have:

$$\begin{aligned}
\mu'\hat{\Sigma}^{-1}(\Theta)\mu &\geq \mu'\left(\lambda\hat{\Sigma}^{-1}(A) + (1-\lambda)\hat{\Sigma}^{-1}(B)\right)\mu = \\
&\lambda\mu'\hat{\Sigma}^{-1}(A)\mu + (1-\lambda)\mu'\hat{\Sigma}^{-1}(B)\mu.
\end{aligned}$$

Combining both inequalities above, we obtain the monotonicity and concavity of the ex-ante utility (and strict monotonicity and concavity, whenever $\Sigma_u \neq 0$).

Q.E.D.

Proof of Lemma 2: The posterior variance is

$$\begin{aligned}
\hat{\Sigma}(\Theta) &= \left(\left((\alpha\Sigma)^{-1}+\Theta\right)^{-1}+(1-\alpha)\,\Sigma\right) = \\
&= \Sigma^{1/2}\Sigma^{-1/2}\left(\left((\alpha\Sigma)^{-1}+\Theta\right)^{-1}+(1-\alpha)\,\Sigma\right)\Sigma^{-1/2}\Sigma^{1/2} \\
&= \Sigma^{1/2}\left(\Sigma^{-1/2}\left((\alpha\Sigma)^{-1}+\Theta\right)^{-1}\Sigma^{-1/2}+(1-\alpha)\,I\right)\Sigma^{1/2} \\
&= \Sigma^{1/2}\left(\left(\alpha^{-1}I+\Sigma^{1/2}\Theta\Sigma^{1/2}\right)^{-1}+(1-\alpha)\,I\right)\Sigma^{1/2} \\
&= \Sigma^{1/2}\left(\left(\alpha^{-1}I+\Theta_\omega\right)^{-1}+(1-\alpha)\,I\right)\Sigma^{1/2} \\
&= \Sigma^{1/2}\left(\alpha\left(I+\alpha\Theta_\omega\right)^{-1}+(1-\alpha)\,I\right)\Sigma^{1/2}.
\end{aligned}$$

where $\Theta_\omega = \Sigma^{1/2}\Theta\Sigma^{1/2}$. Note that

$$\begin{aligned}
&\alpha\left(I+\alpha\Theta_\omega\right)^{-1}+(1-\alpha)\,I = \alpha\left(I+\alpha\Theta_\omega\right)^{-1}+(1-\alpha)\left(I+\alpha\Theta_\omega\right)^{-1}\left(I+\alpha\Theta_\omega\right) = \\
&\left(I+\alpha\Theta_\omega\right)^{-1}\left(\alpha I+(1-\alpha)\left(I+\alpha\Theta_\omega\right)\right) = \left(I+\alpha\Theta_\omega\right)^{-1}\left(\alpha I+(1-\alpha)\,I+\alpha(1-\alpha)\,\Theta_\omega\right) = \\
&= \left(I+\alpha\Theta_\omega\right)^{-1}\left(I+\alpha(1-\alpha)\,\Theta_\omega\right).
\end{aligned}$$

Thus, we have

$$\hat{\Sigma}(\Theta)^{-1} = \Sigma^{-1/2}\left(I+\alpha(1-\alpha)\,\Theta_\omega\right)^{-1}\left(I+\alpha\Theta_\omega\right)\Sigma^{-1/2} = \Sigma^{-1/2}f\left(\Theta_\omega\right)\Sigma^{-1/2}.$$

Therefore, the utility gain, for the CARA preference is

$$\log\det\left(\left(\hat{\Sigma}(\Theta)\right)^{-1}\Sigma\right) = \log\det f\left(\Theta_\omega\right) = \sum_{i=1}^{n}\log f(\lambda_i),$$

where $\lambda_i \geq 0$ are the eigenvalues of the weighted information matrix Θ_ω.

Observe that the eigenvalues of $\Theta_\omega = \Sigma^{1/2}\Theta\Sigma^{1/2}$ are the same as the eigenvalues of $\Theta\Sigma$. Indeed, the eigenvalues are the solutions of

$$\det(\lambda I-\Sigma^{1/2}\Theta\Sigma^{1/2}) = \det\left(\Sigma^{-1/2}(\lambda I-\Sigma^{1/2}\Theta\Sigma^{1/2})\Sigma^{1/2}\right) = \det(\lambda I-\Theta\Sigma) = 0.$$

Mean-variance preference case: *Let $\lambda_i \geq 0$ and ξ_i be, respectively, the eigenvalues and the eigenvectors of the weighted information matrix $\Theta_\omega = \Sigma^{1/2}\Theta\Sigma^{1/2}$ and $\Sigma_u = (1-\alpha)\,\Sigma$ and $\Sigma_l = \alpha\Sigma$, for some $\alpha \in [0,1]$. Define the strictly increasing and concave function f by*

$$f(x) = \frac{1+\alpha x}{1+\alpha(1-\alpha)\,x}. \tag{33}$$

In the mean-variance preference case, the ex-ante utility is

$$U(\Theta) = \frac{1}{2\gamma}\left(\sum_{i=1}^{n} f(\lambda_i)\left(1 + (\xi_i' s_r)^2\right)\right). \tag{34}$$

where $s_R = \Sigma^{-\frac{1}{2}}\left(\bar{R} - r_f\right)$.

Q.E.D.

Proof of Proposition 6: (i) We first show that

$$\mathcal{U} = \frac{1}{2\gamma}\left[\underbrace{\tau \sum_{i,j=1:i<j}^{q} \theta_i\theta_j\left[\left(var(r_i - r_j) + (\mu_i - \mu_j)^2\right)\right]}_{\mathcal{U}_R}\frac{\tau}{\phi+\tau} + \underbrace{\tau\left(\sum_{i=1}^{q}\theta_i\left(var(r_i) + \mu_i^2\right)\right)}_{\mathcal{U}_A}\frac{\phi}{\phi+\tau}\right].$$

Define the weighted incidence matrix $\nabla \in \mathbb{R}^{n\times n}$ as

$$\nabla(ij,k) = \begin{cases} -\sqrt{\theta_i\theta_j} & \text{if } k \text{ is the initial vertex of edge } e = ij \\ \sqrt{\theta_i\theta_j} & \text{if } k \text{ is the terminal vertex of edge } e = ij \\ 0 & \text{if } k \text{ not in edge } e = ij \end{cases}$$

(see for example Bapat (2010, ch 4). Note that

$$\Theta = diag(\theta) - \theta\theta' = \nabla^T\nabla$$

We now show that

$$Tr(\Theta\Sigma) = \frac{1}{2}\sum_{i=1}^{q}\sum_{j=1}^{q}\theta_i\theta_j var(r_i - r_j) =$$

Indeed,

$$\begin{aligned} Tr(\Theta\Sigma) &= Tr(\Sigma\Theta) = Tr(\nabla\Sigma\nabla^T) = \\ \left(\nabla\Sigma\nabla^T\right)_{ij,ij} &= \theta_i\theta_j\left(var(r_i) + var(r_j) - 2cov(r_i, r_j)\right) \\ &= \theta_i\theta_j var(r_i - r_j) \end{aligned}$$

$$Tr(\nabla\Sigma\nabla^T) = \sum_{i,j:i<j}\theta_i\theta_j var(r_i - r_j) = \frac{1}{2}\sum_{i,j=1}^{q}\theta_i\theta_j var(r_i - r_j)$$

The key property of the Laplacian is that:

$$
\begin{aligned}
\mu' \Theta \mu &= \mu' \left[diag(\theta) - \theta\theta' \right] \mu \\
&= \frac{1}{2} \sum_{i,j=1}^{q} \theta_i \theta_j \left(\mu_i - \mu_j \right)^2
\end{aligned}
$$

Finally,

$$
\Theta_A = \tau diag(\theta)
$$

Therefore,

$$
Tr\left(\Sigma\Theta_A\right) + \mu' \Theta_A \mu = \tau \left(\sum_{i=1}^{q} \theta_i \left(var(r_i) + \mu_i^2 \right) \right).
$$

which completes the proof.

Q.E.D.

Proof of Proposition 7: We proceed evaluating the utility gain

$$
\mathcal{U} = \frac{1}{2\gamma} \left(Tr\left(\Sigma\Theta\right) + \left(\mu 1\right)' \Theta \left(\mu 1\right) \right) - c(\tau, \phi),
$$

where the information matrix is

$$
\Theta = \left[\frac{\phi}{\tau + \phi} \left[diag(\tau\theta) \right] + \frac{\phi}{\tau + \phi} \left[\tau \left(diag(\theta) - \theta\theta' \right) \right] \right].
$$

Observe that

$$
\left(\mu 1\right)' \Theta \left(\mu 1\right) = \mu^2 1' \Theta 1 = \mu^2 \frac{\phi\tau}{\tau + \phi}
$$

and

$$
\begin{aligned}
Tr\left(\Sigma\Theta\right) &= \sigma^2 Tr\left(\left(I + \rho J \right) \Theta \right) \\
&= \sigma^2 Tr \left(\Theta + \frac{\rho\phi}{\tau + \phi} 11' \left[diag(\tau\theta) \right] \right) \\
&= \sigma^2 \left(Tr\Theta + \rho \frac{\phi\tau}{\tau + \phi} \right).
\end{aligned}
$$

Therefore,

$$
\mathcal{U} = \frac{1}{2\gamma} \left(\sigma^2 Tr\Theta + \left(\sigma^2 \rho + \mu^2 \right) \frac{\phi\tau}{\tau + \phi} \right) - c(\tau, \phi) \tag{35}
$$

Consider the problem of maximizing the trace of the information matrix $Tr\Theta$ subject to the constraints:

$$\begin{array}{cc} \max_\theta & Tr\left(\tau\left(diag(\theta)-\frac{\tau}{\tau+\phi}\theta\theta'\right)\right) \\ s.t. & \sum_{i=1}^m \theta_i = 1 \\ & \theta_i \geq 0, \forall i \\ & \#\{i\in N:\theta_i>0\}\leq q. \end{array}$$

Observe that

$$Tr\left(diag(\theta)-\frac{\tau}{\tau+\phi}\theta\theta'\right)=Tr\left(diag(\theta)\right)-\frac{\tau}{\tau+\phi}Tr\left(\theta\theta'\right),$$

and $Tr\left(diag(\theta)\right)=1$ (because $\sum_i\theta_i=1$) and $Tr\left(\theta\theta'\right)=\sum_{i=1}^m\theta_i^2$, therefore,

$$Tr\left(\tau\left(diag(\theta)-\frac{\tau}{\tau+\phi}\theta\theta'\right)\right)=\tau\left(1-\frac{\tau}{\tau+\phi}\sum_{i=1}^m\theta_i^2\right).$$

Thus the maximization problem is equivalent to

$$\begin{array}{cc} \min_\theta & \sum_{i=1}^m \theta_i^2 \\ s.t. & \sum_{i=1}^m \theta_i = 1 \\ & \theta_i \geq 0, \forall i \\ & \#\{i\in N:\theta_i>0\}\leq q. \end{array}$$

The solution to this problem is to spread the precision equally across all possible q assets so that $\theta_i=\frac{1}{q}$, and the minimum is $q\left(\frac{1}{q}\right)^2=\frac{1}{q}$, so that the maximum value of the problem is equal to $Tr\Theta=\tau\left(1-\frac{\tau}{(\tau+\phi)q}\right)$.

Replacing this value into (35) yields

$$\begin{aligned}\mathcal{U}(\tau,\phi)&=\frac{1}{2\gamma}\sigma^2\left(\tau\left(1-\frac{\tau}{(\tau+\phi)\,q}\right)+\left(\rho+\frac{\mu^2}{\sigma^2}\right)\frac{\phi\tau}{\tau+\phi}\right)-c(\tau,\phi)=\\ &=\frac{1}{2\gamma}\sigma^2\frac{\tau}{\tau+\phi}\left(\tau\left(\frac{q-1}{q}\right)+\left(\rho+\frac{\mu^2}{\sigma^2}+1\right)\phi\right)-c(\tau,\phi)\end{aligned}$$

The function $\mathcal{U}(\tau,\phi)$ is concave because the Hessian $D^2\mathcal{U}$ is negative semidefinite. Indeed, defining

$$F(\tau,\phi)=\frac{\tau}{\tau+\phi}\left(\tau\left(\frac{q-1}{q}\right)+\left(\rho+\frac{\mu^2}{\sigma^2}+1\right)\phi\right),$$

the second derivative is

$$D^2F(\tau,\phi)=\begin{bmatrix}\frac{\partial^2F(\tau,\phi)}{\partial\tau^2} & \frac{\partial^2F(\tau,\phi)}{\partial\tau\partial\phi}\\ \frac{\partial^2F(\tau,\phi)}{\partial\tau\partial\phi} & \frac{\partial^2F(\tau,\phi)}{\partial\phi^2}\end{bmatrix}=-\frac{2\left(\left(\rho+\frac{\mu^2}{\sigma^2}+1\right)-\left(\frac{q-1}{q}\right)\right)}{(\tau+\phi)^3}\begin{bmatrix}\phi^2 & -\tau\phi\\ -\tau\phi & \tau^2\end{bmatrix},$$

(and eigenvalues of the matrix above are $\tau^2 + \phi^2$ and 0), so all eigenvalues of the second derivative are non-positive. Combining with the convexity of $c(\tau, \phi)$ this completes the proof.

Q.E.D.

Proof of Proposition 8: Let Θ^* be the optimal balanced design with precision τ^* and ϕ^* with information matrix

$$\Theta^* = \frac{m}{n}\left[\frac{\tau^*\phi}{\tau^* + \phi^*}I + \frac{\tau^{*2}}{\tau^* + \phi^*}\frac{(q-1)}{(n-1)\,q}\,(nI - J)\right],$$

where τ^* and ϕ^* are the solution of equation (24) for the CARA preference and the corresponding equation for the mean-variance preference.

We will show that $\mathcal{U}(\Theta^*) \geq \mathcal{U}(\Theta)$ for any other feasible design Θ. The proof is inspired by Kiefer's (1975) original proof of universal optimality of balanced designs. The strategy of the proof is to show that there exist an (average) information matrix $\overline{P}$, obtained by averaging all permutations of Θ, such that (i) $\mathcal{U}\left(\bar{\Theta}\right) \geq \mathcal{U}(\Theta)$, and (ii) that $\mathcal{U}(\Theta^*) \geq \mathcal{U}\left(\bar{\Theta}\right)$.

(i) Let π be a permutation of the set of n assets, i.e., a bijection $\pi : \{1, ..., n\} \rightarrow \{1, ..., n\}$. Let T_π be the permutation matrix associated with a permutation π, which is the zero-one matrix with exactly one entry equal to 1 in each row i and column $\pi(i)$, for all rows, and all other entries equal to 0. Let Π be the set of all permutations (Π has $n!$ elements). Given any arbitrary matrix A, the product $T_\pi A T'_\pi$ is a matrix which permutes the rows and columns of A.

Consider now the average of the information matrix Θ,

$$\bar{\Theta} = \sum_{\pi\in\Pi}\frac{1}{n!}T_\pi\Theta T'_\pi.$$

where $\Theta = \sum_{a=1}^{m}\tau_a\left(diag(\theta_a) - \frac{\tau_a}{\tau_a+\phi_a}\theta_a\theta'_a\right)$.

The concavity property of the utility function implies that

$$\mathcal{U}\left(\bar{\Theta}\right) \geq \sum_{\pi\in\Pi}\frac{1}{n!}\mathcal{U}\left(T_\pi\Theta T'_\pi\right).$$

We now show that $\mathcal{U}\left(T_\pi\Theta T'_\pi\right) = \mathcal{U}(\Theta)$. Indeed $T_\pi T'_\pi = T'_\pi T_\pi = I$, thus $\Sigma^{1/2}T_\pi\Theta T'_\pi\Sigma^{1/2}$ and $\Sigma^{1/2}\Theta\Sigma^{1/2}$ have the same eigenvalues,

$$\begin{aligned}\det(\lambda I - \Sigma^{1/2}T_\pi\Theta T'_\pi\Sigma^{1/2}) &= \det\left[\Sigma(\lambda\Sigma^{-1} - T_\pi\Theta T'_\pi)\right]\\ &= \det\left[\Sigma(\lambda\Sigma^{-1} - \Theta)\right] = \det(\lambda I - \Sigma^{1/2}\Theta\Sigma^{1/2}).\end{aligned}$$

For the mean variance case in additional, note that $s_r = \bar{s}_r\mathbf{1}$ is symmetric then

$$s'_r f(T_\pi P_\omega T'_\pi)s_r = s'_r T_\pi f\left(P_\omega\right)T'_\pi s_r = s'_r f\left(P_\omega\right)s_r,$$

which completes the proof that $\mathcal{U}\left(T_\pi \Theta T_\pi'\right) = \mathcal{U}\left(\Theta\right).$

Combined this implies that

$$\mathcal{U}\left(\bar{\Theta}\right) \geq \mathcal{U}\left(\Theta\right).$$

(ii) We now show that $\mathcal{U}\left(\Theta^*\right) \geq \mathcal{U}\left(\bar{\Theta}\right).$

Start observing that given an arbitrary matrix A, taking the averages of all permutations of A, yields the average matrix,

$$\overline{A} = \sum_{\pi\in\Pi} \frac{1}{n!} T_\pi A T_\pi' = (w - z)\, I + zJ,$$

with all diagonal entries equal to w and off-diagonal entries equal to z given by

$$w = \sum_{i=1}^{n} \frac{a_{ii}}{n} = \frac{1}{n} Tr(A) \text{ and } z = \sum_{i=1}^{n} \sum_{j\neq i} \frac{a_{ij}}{n\,(n-1)} = \frac{1}{n\,(n-1)}\,(1'A1 - Tr(A))$$

Consider now the average of the information matrix Θ, that is,

$$\bar{\Theta} = \sum_{\pi\in\Pi} \frac{1}{n!} T_\pi \Theta T_\pi' = \sum_{a=1}^{m} \tau_a \left(\sum_{\pi\in\Pi} \frac{1}{n!} T_\pi diag(\theta_a) T_\pi' - \frac{\tau_a}{\tau_a + \phi_a} \sum_{\pi\in\Pi} \frac{1}{n!} T_\pi \theta_a \theta_a' T_\pi' \right).$$

Because $\theta_a 1' = 1$ for all a, then

$$\sum_{\pi\in\Pi} \frac{1}{n!} T_\pi diag(\theta_a) T_\pi' = \frac{1}{n} I.$$

Also because $1'\theta_a\theta_a'1 = 1$ then

$$\begin{aligned}
\sum_{\pi\in\Pi} \frac{1}{n!} T_\pi \theta_a \theta_a' T_\pi' &= \left(\frac{1}{n} Tr\left(\theta_a\theta_a'\right) - \frac{1}{n\,(n-1)}\left(1 - Tr\left(\theta_a\theta_a'\right)\right) \right) I + \frac{1}{n\,(n-1)}\left(1 - Tr\left(\theta_a\theta_a'\right)\right) J \\
&= \left(\frac{1}{n\,(n-1)}\left(nTr\left(\theta_a\theta_a'\right) - 1\right) \right) I + \frac{1}{n\,(n-1)}\left(1 - Tr\left(\theta_a\theta_a'\right)\right) J \\
&\left(\frac{1}{n\,(n-1)}\left(nTr\left(\theta_a\theta_a'\right) - n + (n-1)\right) \right) I + \frac{1}{n\,(n-1)}\left(1 - Tr\left(\theta_a\theta_a'\right)\right) J \\
&= \frac{1}{n} I - \frac{1}{n\,(n-1)}\left(1 - Tr\left(\theta_a\theta_a'\right)\right)(nI - J)
\end{aligned}$$

Combining both terms yields,

$$\begin{aligned}&\sum_{\pi\in\Pi}\frac{1}{n!}T_\pi diag(\theta_a)T'_\pi-\frac{\tau_a}{\tau_a+\phi_a}\sum_{\pi\in\Pi}\frac{1}{n!}T_\pi\theta_a\theta'_aT'_\pi=\\&\frac{1}{n}I-\frac{\tau_a}{\tau_a+\phi_a}\left(\frac{1}{n}I-\frac{1}{n\left(n-1\right)}\left(1-Tr\left(\theta_a\theta'_a\right)\right)\left(nI-J\right)\right)=\\&=\frac{1}{n}\frac{\phi_a}{\tau_a+\phi_a}I+\frac{\tau_a}{\tau_a+\phi_a}\frac{1}{n\left(n-1\right)}\left(1-Tr\left(\theta_a\theta'_a\right)\right)\left(nI-J\right).\end{aligned}$$

Thus the average information matrix $\bar{\Theta}$ is

$$\bar{\Theta}=\sum_{a=1}^m\tau_a\left(\frac{1}{n}\frac{\phi_a}{\tau_a+\phi_a}I+\frac{\tau_a}{\tau_a+\phi_a}\frac{1}{n\left(n-1\right)}\left(1-Tr\left(\theta_a\theta'_a\right)\right)\left(nI-J\right)\right).$$

By the Cauchy–Schwarz inequality the trace $Tr\left(\theta_a\theta'_a\right)=\sum_{i=1}^n\theta_{ai}^2\geq\frac{1}{q}$ since $1=\langle\theta_a,1_a\rangle^2\leq\langle\theta_a,\theta_a\rangle\langle 1_a,1_a\rangle=q\sum_{i=1}^n\theta_{ai}^2$ because $\sum_{i=1}^n\theta_{ai}=1$ and $\theta_{ai}\geq 0$ for at most q entries.

Therefore,

$$\begin{aligned}\bar{\Theta}\preceq\hat{\Theta}&:=\sum_{a=1}^m\tau_a\left(\frac{1}{n}\frac{\phi_a}{\tau_a+\phi_a}I+\frac{\tau_a}{\tau_a+\phi_a}\frac{q-1}{n\left(n-1\right)q}\left(nI-J\right)\right)\\&=\frac{m}{n}\left(\sum_{a=1}^m\frac{1}{m}\frac{\tau_a\phi_a}{\tau_a+\phi_a}\right)I+\left(\sum_{a=1}^m\frac{1}{m}\frac{\tau_a^2}{\tau_a+\phi_a}\right)\frac{q-1}{\left(n-1\right)q}\left(nI-J\right)\end{aligned}$$

and by the monotonicity property of the operator $\mathcal{U}$ thus $\mathcal{U}\left(\bar{\Theta}\right)\leq\mathcal{U}\left(\hat{\Theta}\right)$.

But we show in Proposition 9 that $\mathcal{U}\left(\hat{\Theta}\right)\leq\mathcal{U}\left(\Theta^*\right)$ which completes the proof.

Q.E.D.

Proof of Proposition 9: (i) We first obtain the information matrix Θ for the balanced design. Defining the $n\times m$ matrix $\theta=[\theta_1,...,\theta_m]\in\mathbb{R}^{n\times m}$ with columns equal to the time allocations of each analyst, we can express the information matrix Θ using Proposition 1 as

$$\Theta=\frac{\tau\phi}{\tau+\phi}diag(\theta 1_m)+\frac{\tau^2}{\tau+\phi}\left(diag(\theta 1_m)-\theta\theta'\right)$$

The agents are organized in a balanced design therefore

$$q^2\theta\theta'=\left(q\theta\right)\left(q\theta\right)'=\left(c-\lambda\right)I+\lambda J\Rightarrow\theta\theta'=\frac{1}{q^2}\left[\left(c-\lambda\right)I+\lambda J\right]\text{ and}$$

$$q\theta 1_m=c1_n\Rightarrow\theta 1_m=\frac{c}{q}1_n\Rightarrow diag(\theta 1_m)=\frac{c}{q}I.$$

Note that

$$\left(\frac{c}{q}-\frac{(c-\lambda)}{q^2}\right)=\frac{(c\,(q-1)+\lambda)}{q^2}$$

and taking into account that

$$\lambda\,(n-1)=c\,(q-1)$$

we have that

$$\frac{(c\,(q-1)+\lambda)}{q^2}=\frac{(\lambda\,(n-1)+\lambda)}{q^2}=\frac{\lambda n}{q^2}$$

This implies that

$$diag(\theta 1_m)-\theta\theta'=\frac{\lambda}{q^2}\,(nI-J)$$

and thus

$$\Theta=\frac{\tau\phi}{\tau+\phi}\frac{c}{q}I+\frac{\tau^2}{\tau+\phi}\frac{\lambda}{q^2}\,(nI-J)\,.$$

We can express the information matrix as

$$\Theta=wI+z\,(nI-J)$$

where

$$w=\frac{\tau\phi}{\tau+\phi}\frac{c}{q}\text{ and }z=\frac{\tau^2}{\tau+\phi}\frac{\lambda}{q^2}.$$

(ii) The the eigenvalues of Θ are

$$\begin{array}{cc}\rho_1=w+zn & \text{with multiplicity } n-1\\ \rho_2=w & \text{with multiplicity 1 associated with eigenvector } \mathbf{1}\end{array}$$

The weighted information matrix is $\Theta\Sigma$ where $\Sigma=\sigma^2\,(I+\rho J)$ is after multiplication equal to

$$\begin{aligned}\Theta\Sigma&=(wI+z\,(nI-J))\,\sigma^2\,(I+\rho J)\\&=\sigma^2wI+\sigma^2z\,(nI-J)+\sigma^2w\rho J\\&=\sigma^2w\,(I+\rho J)+\sigma^2z\,(nI-J)\end{aligned}$$

since $(nI-J)\,J=0$.

The eigenvalues of $\Theta\Sigma$ are equal to:

$$\begin{array}{cc}\lambda_1=\sigma^2\,(w+zn) & \text{with multiplicity } n-1\\ \lambda_2=\sigma^2\,(w+w\rho n) & \text{with multiplicity 1 associated with eigenvector } \mathbf{1}\end{array}$$

The eigenvalues are explicitly given by after replacing the expressions for w and z,

$$\lambda_1 = \sigma^2 \left(\tfrac{\tau\phi}{\tau+\phi} \tfrac{c}{q} + \tfrac{\tau^2}{\tau+\phi} \tfrac{\lambda}{q^2} n \right)$$
$$\lambda_2 = \sigma^2 \tfrac{\tau\phi}{\tau+\phi} \tfrac{c}{q} \left(1 + \rho n\right)$$

but since

$$\frac{c}{q} = \frac{m}{n}$$
$$\frac{\lambda}{q^2} = \frac{mq\left(q-1\right)}{nq^2\left(n-1\right)} = \frac{m\left(q-1\right)}{n\left(n-1\right)q}$$

then

$$\lambda_1 = \sigma^2 \frac{m}{n} \frac{\tau\phi}{\tau+\phi} \left(1 + \frac{\tau}{\phi} \frac{\left(q-1\right)}{\left(n-1\right)q} \right)$$
$$\lambda_2 = \sigma^2 \frac{\tau\phi}{\tau+\phi} \frac{m}{n} \left(1 + \rho n\right).$$

(iii) We show that $\mathcal{U}(\tau, \phi)$ is concave using the following two properties of concave functions: (a) non-negative weighted sum of concave functions is concave. (b) Moreover, the composition of two $f\left(g\left(\tau, \phi\right)\right)$ where f is a concave and increasing and g is convave is concave.

To establish that $\lambda_1\left(\tau, \phi\right)$ and $\lambda_2\left(\tau, \phi\right)$ are concave note that their Hessian matrix are negative semi-definite: Indeed:

$$\lambda_2\left(\tau, \phi\right) = \kappa \frac{\tau\phi}{\tau+\phi} \Rightarrow D^2\lambda_2 = \frac{1}{\left(\tau+\phi\right)^3} \begin{bmatrix} -2\phi^2 & 2\tau\phi \\ 2\tau\phi & -2\tau^2 \end{bmatrix}$$
$$\lambda_1\left(\tau, \phi\right) = \kappa \left(\frac{\tau\phi}{\tau+\phi} + b\frac{\tau^2}{\tau+\phi} \right) \Rightarrow D^2\lambda_1 = \frac{1-b}{\left(\tau+\phi\right)^3} \begin{bmatrix} -2\phi^2 & 2\tau\phi \\ 2\tau\phi & -2\tau^2 \end{bmatrix}$$

where $b = \frac{(q-1)}{(n-1)q} < 1$. Note that the matrix is negative semi-definite because all its eigenvalues are non-positive: $-2\tau^2 - 2\phi^2$ and 0.

For the comparative statics result observe that the solution (τ, ϕ) satisfies the first order conditions

$$\left(n-1\right) \frac{f'\left(\lambda_1\right)}{f\left(\lambda_1\right)} \frac{\partial \lambda_1}{\partial \tau} + \frac{f'\left(\lambda_2\right)}{f\left(\lambda_2\right)} \frac{\partial \lambda_2}{\partial \tau} = 2\gamma \frac{\partial c}{\partial \tau},$$
$$\left(n-1\right) \frac{f'\left(\lambda_1\right)}{f\left(\lambda_1\right)} \frac{\partial \lambda_1}{\partial \phi} + \frac{f'\left(\lambda_2\right)}{f\left(\lambda_2\right)} \frac{\partial \lambda_2}{\partial \phi} = 2\gamma \frac{\partial c}{\partial \phi}.$$

The derivatives above are

$$\begin{aligned}
\frac{\partial \lambda_1(\tau,\phi)}{\partial \tau} &= \frac{1}{(\tau+\phi)^2}\left(b\tau^2+2b\tau\phi+\phi^2\right),\\
\frac{\partial \lambda_1(\tau,\phi)}{\partial \phi} &= \frac{\tau^2(1-b)}{(\tau+\phi)^2},\\
\frac{\partial \lambda_2(\tau,\phi)}{\partial \tau} &= \frac{(1+\rho n)\phi^2}{(\tau+\phi)^2},\\
\frac{\partial \lambda_2(\tau,\phi)}{\partial \phi} &= \frac{(1+\rho n)\tau^2}{(\tau+\phi)^2},\\
\frac{f'(x)}{f(x)} &= \frac{\alpha^2}{(1+\alpha x)(1+\alpha(1-\alpha)x)}.
\end{aligned}$$

Replacing them into the first order condition, after simplification, yields

$$(n-1)\frac{f'(\lambda_1)}{f(\lambda_1)}\left(\frac{(q-1)}{(n-1)q}\left(\left(\frac{\tau}{\phi}\right)^2+2\frac{\tau}{\phi}\right)+1\right)+\frac{f'(\lambda_2)}{f(\lambda_2)}(1+\rho n)=\left(\frac{(\tau+\phi)}{\phi}\right)^2 2\gamma\frac{\partial c}{\partial \tau}$$

$$(n-1)\frac{f'(\lambda_1)}{f(\lambda_1)}\left(1-\frac{(q-1)}{(n-1)q}\right)+\frac{f'(\lambda_2)}{f(\lambda_2)}(1+\rho n)=\left(\frac{(\tau+\phi)}{\tau}\right)^2 2\gamma\frac{\partial c}{\partial \phi}$$

The analysis of the equation above shows that the solution is such that an increase in q leads to more weight on relative valuation $\partial\left(\frac{\tau}{\phi}\right)/\partial q \geq 0$, and such that an increase in asset correlations ρ leads to less weight on relative valuation $\partial\left(\frac{\tau}{\phi}\right)/\partial \rho \leq 0$.

Q.E.D.

Proof of Proposition 10: (i) Remind the strict monotonicity property of the utility function on information matrices shown in Lemma 1. Each analyst produces a diagonal information matrix equal to $\Theta_a = \tau(1-\kappa)\,diag(\theta_a)$ or $\Theta_a' = \tau(1-2\kappa)\,diag(\theta_a)$, respectively if all assets covered are in the same industry or not. Thus, any structure with cross-industry analysts can be dominated by a structure exclusively with industry specialists that is more informative $\sum_{a=1}^m \Theta_a$ due to the loss of precision of cross-industry analysts (e.g., by replacing two cross-industry analysts by two specialists following the same assets $\sum_{a=1}^m \Theta_a' \prec \sum_{a=1}^m \Theta_a$).

(ii) From Proposition 9 the structure with only industry specialists which creates the maximum utility must have all analysts organized in a balanced way in each industry. A balanced allocation with m_j analyst allocated to each industry j produces an information matrix, see equation 23,

$$\Theta = \begin{bmatrix} \Theta_1 & 0 \\ 0 & \Theta_1 \end{bmatrix}, \text{ where } \Theta_j = \frac{\tau(1-\kappa)m_j(q-1)}{q(n_j-1)}\left(I_j - \frac{1}{n_j}J_j\right).$$

The variance Σ is equal to

$$\Sigma = \begin{bmatrix} \sigma_1^2 \left(I_1 + \rho_1 J_1\right) & 0 \\ 0 & \sigma_2^2 \left(I_1 + \rho_2 J_2\right) \end{bmatrix} + \sigma_f^2 J$$

where J, J_1, J_2 are the matrices with all entries equal to one of dimensions $n = n_1 + n_2$, n_1, and n_2; and similary I_j are the industry identity matrices.

Similiarly to equation (25), the matrix $\Theta\Sigma$ has two zero eigenvalues and $n - 2$ non-zero eigenvalues equal to

$$\lambda_j = \sigma_j^2 \frac{\tau \left(1 - \kappa\right) m_j \left(q - 1\right)}{q \left(n_j - 1\right)} \text{ with multiplicity } n_j - 1.$$

From Lemma 2 the utility gain is

$$\mathcal{U}\left(\Theta\right) = \frac{1}{2\gamma} \sum_{j=1}^{2} \left(n_j - 1\right) \log f(\lambda_j),$$

with the number of total analysts $m = m_1 + m_2$ distributed across industries so that the marginal contribution of each analyst is equalized,

$$\frac{\partial \mathcal{U}\left(\Theta\right)}{\partial m_j} = \frac{\left(n_j - 1\right) f'(\lambda_j)}{f(\lambda_j)} \sigma_j^2 \frac{\tau \left(1 - \kappa\right) \left(q - 1\right)}{q \left(n_j - 1\right)},$$

which implies that the optimal allocation satisfies

$$\frac{f'(\lambda_1)}{f(\lambda_1)} \sigma_1^2 = \frac{f'(\lambda_2)}{f(\lambda_2)} \sigma_2^2.$$

Let $\mathcal{U}\left(m\right)$ denote the maximun utility gain without cross-industry analysts. Because $\frac{f'(x)}{f(x)}$ is monotonically decreasing in x and converges to zero as $x \to \infty$, the marginal utility $\mathcal{U}'\left(m\right)$ is decreasing in m and converging to zero as m converges to infinity.

We now show that the structure above is dominated, for m large enough, by the alternative structure with $m = m_1 + m_2 + m_c$, where there are m_c cross-industry analysts each of whom are allocating half of their time covering one asset in each industry.

Each cross-industry analyst contributes information matrix, by following assets i and i' with $\theta_i = \frac{1}{2}$ and $\theta_{i'} = \frac{1}{2}$, $\tau \left(1 - 2\kappa\right) \left(diag(\theta) - \theta\theta'\right)$, and thus m_c cross-industry analysts distributed in a balanced way, so that each of the $n_1 n_2$ cross-industry pair of assets are covered by the same number of analysts, produce information matrix

$$\Theta_c = \frac{m_c}{n_1 n_2} \times \frac{\tau \left(1 - 2\kappa\right)}{4} \begin{bmatrix} n_2 I_1 & -J_c \\ -J_c' & n_1 I_2 \end{bmatrix}$$

where J_c is the $n_1 \times n_2$ matrix of ones. The information matrix Θ becomes

$$\Theta = \begin{bmatrix} \Theta_1 & 0 \\ 0 & \Theta_1 \end{bmatrix} + \Theta_c.$$

The matrix $\Theta\Sigma$ has one zero, and $n_j - 1$ eigenvalues equal to

$$\lambda_j = \sigma_j^2 \left(\frac{\tau (1-\kappa) m_j (q-1)}{q(n_j - 1)} + \frac{m_c \tau (1-2\kappa)}{4n_j} \right),$$

and one eigenvalue equal to

$$\lambda_c = \frac{m_c \tau (1-2\kappa)}{4} \sum_{j=1}^{2} \left(\sigma_j^2 \rho_j + \frac{\sigma_j^2}{n_j} \right).$$

The utility gain is thus

$$\mathcal{U}(\Theta) = \frac{1}{2\gamma} \left(\sum_{j=1}^{2} (n_j - 1) \log f(\lambda_j) + \log f(\lambda_c) \right).$$

The marginal utility contribution of the first cross-industry analyst is at least

$$\begin{aligned} \frac{\partial \mathcal{U}(\Theta)}{\partial m_c}|_{m_c=0} &\geq \frac{1}{2\gamma} \frac{\partial \log f(\lambda_c(m_c))}{\partial m_c} = \frac{1}{2\gamma} \frac{f'(\lambda_c)}{f(\lambda_c)}|_{\lambda_c=0} \frac{\tau(1-2\kappa)}{4} \sum_{j=1}^{2} \left(\sigma_j^2 \rho_j + \frac{\sigma_j^2}{n_j} \right) \\ &= \frac{1}{2\gamma} \alpha^2 \frac{\tau(1-2\kappa)}{4} \sum_{j=1}^{2} \left(\sigma_j^2 \rho_j + \frac{\sigma_j^2}{n_j} \right) > 0 \end{aligned}$$

Therefore, the investor can obtain strictly higher utility utilizing some cross-industry analysts for all m large enough because we have seen before that $\mathcal{U}'(m)$ is decreasing in m and converging to zero as m converges to infinity.

Q.E.D.

B. Appendix B: Online Appendix

Proposition B.1 *(Asymmetric Assets - Mean-variance Utility) Consider a setting where* $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2) + \sigma_f^2 J$, *and the prior asset returns are* $\mu_i = E(R_i) - r_f$ *and mean-variance utility. Consider the optimal allocation of time of an analyst with precision* τ and ϕ. *Then unique global solution to the investor optimal design problem is for the analyst to allocate attention* θ_i *only to assets with values for*

$$\sigma_i^2 + \frac{\phi}{\tau + \phi}\mu_i^2 + \frac{\tau}{\tau + \phi}(\mu_i - \bar{\mu})^2 \geq \lambda$$

above a certain cut-off value λ. *The attention is given by*

$$\theta_i = \frac{1}{2\sigma_i^2}\left(1 + \frac{\phi}{\tau}\right)\left(\sigma_i^2 + \frac{\phi}{\tau + \phi}\mu_i^2 + \frac{\tau}{\tau + \phi}(\mu_i - \bar{\mu})^2 - \lambda\right)^+$$

where λ *and* $\bar{\mu}$ *are constants obtained by the solution of the two equations* $\sum_{i=1}^n \theta_i = 1$ *and* $\bar{\mu} = \sum_{i=1}^n \mu_i \theta_i$, *and the function* $x^+ := \max(x, 0)$.

Proposition B.2 *(Necessary condition-Colbourn and Dinitz (2006)) Given a triple* (n, m, q) *if* $\lambda-$*balanced design exists then the two necessary conditions must hold: (i)* $\lambda(n-1)$ *must be divisible by* $(q-1)$ *and (ii)* $\lambda n(n-1)$ *must be divisible by* $(q-1)$.
(Sufficient condition- Wilson (1970)) Given any triple (n, m, q), *there exists an integer* n_0, *such that for all* $n \geq n_0$ *there exists a* $\lambda-$*balanced design if the two necessary conditions above are satisfied.*

Proof of Proposition B.1: The investor utility gain in the case where all information is learnable is

$$\mathcal{U} = \frac{1}{2\gamma}\left(Tr(\Sigma\Theta) + \mu'\Theta\mu\right)$$

where $\mu = E[R] - r_f$. We have shown before that due to the linearity of the utility gain there is no interaction among the information production by different analysts so we focus on maximizing

$$\begin{array}{rc} \max & Tr(\Sigma\Theta) + \mu'\Theta\mu \\ \text{s.t.} & \Theta = \tau\left(diag(\theta) - \frac{\tau}{\tau+\phi}\theta\theta'\right) \\ & \theta_i \geq 0 \end{array}$$

To obtain $Tr(\Sigma\Theta)$ note that $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2) + \rho J$ thus

$$Tr(\Sigma\Theta) = Tr(diag(\sigma_1^2, ..., \sigma_n^2)\Theta) + Tr(\rho J\Theta).$$

The first part is

$$Tr(diag(\sigma_1^2, ..., \sigma_n^2)\Theta) = \sum_{i=1}^{n} \sigma_i^2 \Theta_{ii} = \tau m \sum_{i=1}^{n} \sigma_i^2 \theta_i \left(1 - \frac{\tau}{\tau + \phi}\theta_i\right)$$

since the diagonal terms are

$$\Theta_{ii} = \tau m \left(\theta_i - \frac{\tau}{\tau + \phi}\theta_i^2\right) = \tau m \theta_i \left(1 - \frac{\tau}{\tau + \phi}\theta_i\right).$$

To calculate the second term $Tr\left(\rho J \Theta\right)$, where $J = \mathbf{1}\mathbf{1}'$ note that the information matrix can be expressed as

$$\Theta = \tau \left(diag(\theta) - \frac{\tau}{\tau + \phi}\theta\theta'\right) = \tau \left(\frac{\phi}{\tau + \phi}\left[diag(\theta)\right] + \frac{\tau}{\tau + \phi}\left[diag(\theta) - \theta\theta'\right]\right).$$

Because $\left(diag(\theta) - \theta\theta'\right)\mathbf{1} = diag(\theta)\mathbf{1} - \theta\mathbf{1}$ =0 we have that $J\left(diag(\theta) - \theta\theta'\right) = 0.$ Thus

$$J\Theta = \frac{\tau\phi}{\tau + \phi}\mathbf{1}\mathbf{1}'diag(\theta) = \frac{\tau\phi}{\tau + \phi}\mathbf{1}\theta'$$

and thus

$$Tr\left(\rho J \Theta\right) = \frac{\rho\tau\phi}{\tau + \phi}\sum_{i=1}^{q}\theta_i = \frac{\rho\tau\phi}{\tau + \phi}.$$

Combining both terms we have

$$Tr(\Sigma\Theta) = \tau \left[\sum_{i=1}^{q} \sigma_i^2 \theta_i \left(1 - \frac{\tau}{\tau + \phi}\theta_i\right) + \frac{\rho\phi}{\tau + \phi}\right].$$

and thus the optimal allocation of time θ does not depend on the correlation term ρ.

To compute $\mu'\Theta\mu$, after factoring out the constant τ, it is equal to:

$$\begin{aligned}
&\mu'\left(diag(\theta) - \frac{\tau}{\tau + \phi}\theta\theta'\right)\mu \\
&= \mu' diag(\theta)\mu - \frac{\tau}{\tau + \phi}\mu'\theta\theta'\mu \\
&= \mu' diag(\theta)\mu - \frac{\tau}{\tau + \phi}\left(\theta'\mu\right)'\theta'\mu \\
&= \sum_{i=1}^{q}\mu_i^2\theta_i - \frac{\tau}{\tau + \phi}\left(\sum_{i=1}^{q}\mu_i\theta_i\right)^2
\end{aligned}$$

Hence the optimal design problem simplifies into solving the following concave constrained program which has a unique global solution characterized by the first order condition:

$$\begin{array}{lcr} \max & \tau\left[\sum_{i=1}^{q}\theta_i\left(\sigma_i^2+\mu_i^2-\frac{\tau}{\tau+\phi}\sigma_i^2\theta_i\right)-\frac{\tau}{\tau+\phi}\left(\sum_{i=1}^{n}\mu_i\theta_i\right)^2+\frac{\rho\phi}{\tau+\phi}\right] & \\ \text{s.t.} & \sum_{i=1}^{q}\theta_i=1 & \lambda \text{ multiplier} \\ & \theta_i\geq 0 & \lambda_i \text{ multiplier} \end{array}$$

Consider the Lagragian

$$\begin{aligned} \mathcal{L} = \tau &\left[\sum_{i=1}^{q}\theta_i\left(\sigma_i^2+\mu_i^2-\frac{\tau}{\tau+\phi}\sigma_i^2\theta_i\right)-\frac{\tau}{\tau+\phi}\left(\sum_{i=1}^{n}\mu_i\theta_i\right)^2+\frac{\rho\phi}{\tau+\phi}\right] \\ &+\sum_{i=1}^{n}\lambda_i\theta_i-\lambda\left(\sum_{i=1}^{n}\theta_i-1\right)-\delta\left(G\left(\tau,\phi\right)-\kappa\right) \end{aligned}$$

The first order condition, which is necessary and sufficient to characterize the unique optimal solution, is

$$\begin{gathered} \frac{\partial\mathcal{L}}{\partial\theta_i}=0,\forall i \\ \sum_{i=1}^{q}\theta_i=1 \text{ and } G\left(\tau,\phi\right)=\kappa \\ \lambda_i\geq 0,\ \ \theta_i\geq 0,\ \text{ and } \lambda_i\theta_i=0,\forall i \end{gathered}$$

Explicitly the first order condition above is equivalent to:

$$\begin{aligned} &\frac{\partial\mathcal{L}}{\partial\theta_i}=\sigma_i^2+\mu_i^2-\frac{2\tau}{\tau+\phi}\sigma_i^2\theta_i-\frac{2\tau}{\tau+\phi}\mu_i\left(\sum_{i=1}^{q}\mu_i\theta_i\right)+\lambda_i-\lambda=0 \\ &\sum_{i=1}^{q}\theta_i=1 \\ &\lambda_i\geq 0,\ \ \theta_i\geq 0,\ \text{ and } \lambda_i\theta_i=0,\forall i \end{aligned}$$

Let $\bar{\mu}:=\sum_{i=1}^{q}\mu_i\theta_i$ be the average value of excess returns, which appears in all equations above. The first equation becomes

$$\frac{2\tau}{\tau+\phi}\sigma_i^2\theta_i=\sigma_i^2+\mu_i^2-\frac{2\tau}{\tau+\phi}\mu_i\bar{\mu}+\lambda_i-\lambda$$

which determines θ_i. Note that for any asset i such that $\sigma_i^2+\mu_i^2-\frac{2\tau}{\tau+\phi}\mu_i\bar{\mu}-\lambda<0$ then

$\lambda_i > 0$ and thus $\theta_i = 0$. Moreover, if $\theta_i > 0$ then $\lambda_i = 0$. Therefore

$$\frac{2\tau}{\tau+\phi}\sigma_i^2\theta_i = \left(\sigma_i^2 + \mu_i^2 - \frac{2\tau}{\tau+\phi}\mu_i\bar{\mu} - \lambda\right)^+ ,$$

which is equivalent to

$$\theta_i = \frac{1}{2}\left(1+\frac{\phi}{\tau}\right)\left(1 - \frac{1}{\sigma_i^2}\left(\lambda - \mu_i\left(\mu_i - \frac{2\tau}{(\tau+\phi)}\bar{\mu}\right)\right)\right)^+ .$$

Q.E.D.

B.1 CARA preferences

An investor with Constant Absolute Risk Aversion evaluates lotteries according to $E_1[\exp(-\gamma W_3)]$, where W_3 is the investor date 3 wealth (and consumption). In our setting $W_3 = (1 + R_p)$ because the investor wealth is normalized to 1. It follows then that the ex-ante utility of a CARA investor can be represented (up to a strictly monotone transformation) as

$$U = -\frac{1}{\gamma}log(E[exp(-\gamma R_p)])$$

Note further that the portfolio choice of the CARA and mean-variance investors are the same because here we normalize their wealth and all asset prices to be equal to one. Therefore portfolio weights (in the mean-variance case) and asset quantities (in the CARA case) are one and the same in this special case.