Specialization in Investor Information and the Diversification Discount

Ronaldo Carpio*[1] and Meixin Guo†[2]

[1] *School of Banking & Finance, University of International Business & Economics*

[2] *Department of Economics, Tsinghua University*

March 10, 2017

Abstract

We propose a theory of the diversification discount and corporate spinoffs based on specialization in information on asset returns by investors. In the CARA-Gaussian framework, a discount arises when one investor has more precise information about a specific asset, compared to other investors. A discount exists in expectation in models of noisy rational expectations equilibrium with learning from prices. The discount (and hence the incentive for a spinoff) increases with the degree of information specialization among investors. We present results that apply to any number of investors and assets with any correlation structure.

Keywords: Diversification discount; Heterogeneous beliefs

JEL classification: G14; G34

*Corresponding author. 9th Floor, Boxue Bldg., 10 Huixindongjie, Beijing, 100029, China. Email: rncarpio@yahoo.com. Tel: +86 18811577940, Fax: +86 1064495059.

†Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, 100084. Email: guomx@sem.tsinghua.edu.cn. Tel: (8610) 6279-5839.

1 Introduction

In corporate finance, the "diversification discount" refers to the empirical regularity that a diversified firm seems to be valued less than a portfolio of comparable single-segment firms (Rajan, Servaes, and Zingales (2000), Scharfstein and Stein (2000), Martin and Sayrak (2003)). A closely related phenomenon is a corporate spin-off, in which one or more divisions are split off into a separate entity with own stock price, but is still owned by the shareholders of original firm. Managers and activist investors frequently justify spinoffs by arguing that components of a firm may be undervalued due to poor visibility; "pure play" stocks, on the other hand, are said to be rewarded by the market with higher valuations. For example, in 2014, the activist investor Carl Icahn invoked the diversification discount when arguing for the breakup of Ebay and PayPal (La Roche 2014):

> "We believe that the separation of the traditional eBay and PayPal businesses will: (1) *highlight the significant value of the disparate businesses currently shrouded by a conglomerate discount the market has afforded eBay;* (2) focus and empower independent management teams to most effectively build two very different business platforms, make economic decisions independent of each other and, most importantly, foster innovation; and (3) provide an even more valuable currency for future bolt-on acquisition opportunities..."

Empirically, spin-offs seem to generate positive abnormal returns (Veld and Veld-Merkoulova 2009). We present a theoretical explanation for this phenomenon based on heterogeneity of investor beliefs; specifically, in the CARA-Gaussian framework, we show that heterogeneity in the covariance matrix of investors' beliefs can cause a diversification discount. We interpret this heterogeneity as *specialization in information*; investors differ in the market sectors that they have expertise in, which corresponds to covariance matrices with reduced variances for different assets. This discount exists in expectation in models of noisy rational expectations equilibrium (REE), in which agents are allowed to learn from prices. In contrast to other models of the diversification discount, our result depends only on heterogeneity of beliefs, and does not assume "noise traders", overconfidence, or bounded rationality by investors. In some models of the diversification discount, a spinoff always results in increased value for the firm; if taken to its logical conclusion, this would imply that firms should split into very many parts. In contrast, in our model there is a natural limit to the types of spinoffs that increase market value; a spinoff when there is no heterogeneity in investor specialization does not result in an increase in market value. Furthermore, in contrast to most other results which are proved for a two-investor, two-asset case, we provide a more general result that applies to any number of investor types and assets and arbitrary correlations between assets (with some restrictions on investors' belief means). This is made possible by applying results from the analysis of positive definite matrices to investors' belief variances; as far as we are aware, these tools have not previously been applied to theoretical asset pricing.

2 Related Literature

Our paper is related to models in which the informativeness of stock prices about actual returns has an effect on firm valuation. This can arise either because investors give higher valuations to firms with more informative stock prices, or because managers make better decisions with a more informative stock price.

Habib, Johnsen, and Naik (1995) is the closest in spirit to our paper; they develop a two-asset model based on the Grossman and Stiglitz (1980) framework with an informed and uninformed investor (but without the choice of buying a signal that lowers uncertainty). In their model, when the assets are split into two separate stocks, uninformed investors receive two price signals, compared to one price signal when the assets are combined into a single stock. If these two price signals are less than perfectly correlated, this provides more information about asset returns than the single price signal, and after observing these signals, uninformed investors have a lower uncertainty about asset returns and hence will place a higher total value on the assets. As is common in these types of models, the noise in the price signal is interpreted as due to randomness in asset supply caused by ”noise traders”. However, the result is dependent on the assumptions made about the relative variances in the price signals before and after the split. If the variance due to noise traders after the split is large enough, then the combined market value may be lower than that of the original entity. Furthermore, if taken to its logical conclusion, this model implies that a firm should keep on splitting itself into more pieces as long as the pieces' payoffs are less than perfectly correlated; this will always increase total market value. In contrast, the model we present in this paper generates an increase in market value only if there is heterogeneity in the types of assets that investors specialize in. Liu and Qi (2008) also develop a model with two assets in which more price signals result in a higher market value, but with a different mechanism. In their model, a firm has an investment project whose return depends on the (unknown) returns of the two assets, and a quantity of investment that must be chosen by managers. Managers cannot observe the asset returns directly, and must rely on price signals to infer these returns; more informative price signals results in a better choice of investment and hence a higher firm value. As in Habib, Johnsen, and Naik (1995), noise in price signals is assumed to be due to noise traders; they show that a regime with two separate stocks is more informative than with one stock.

Cao, Wang, and Zhang (2005) develop a model with investors who are heterogeneous in their degree of *uncertainty aversion*: investors choose portfolios based on the worst-case possibility among a set of possible probability distributions. They show how investors with different levels of uncertainty aversion can generate a diversification discount and limited stock market participation in equilibrium, as the more uncertainty-averse investors do not invest at all when there are two separate stocks, but will invest when the two assets are combined. Hirshleifer and Teoh (2003) develop a model with two types of investors: *attentive* investors, who process all available financial information on a firm, and *inattentive* investors who do not. Specifically, if a firm is composed of many segments with possibly different growth rates, inattentive investors only pay attention to the average growth rate of the entire firm, and use this to extrapolate the future growth rate of the firm. This leads to undervaluation of a firm with a rapidly growing segment, as the segment with a higher growth rate will eventually dominate the firm over time. Therefore, if a firm has segments with different growth rates, the market will give a higher total value to the firm when these segments are split off.

Two game-theoretic models of the discount are presented in Chemmanur and Yan (2004), in which spinoffs are more likely to be taken over by more productive managers, and Nanda and Narayanan (1999), in which a spinoff enables investors to distinguish between divisions of differing quality in a separating equilibrium. Scheinkman and Xiong (2003) present a dynamic asset-pricing model in which two investors observe signals from two subsidiaries of the same firm; both investors are overconfident in that they believe their signal is more precise than it actually is. Finally, the decision of whether to create a spinoff or not can be seen as a type of security design problem. There are several papers examining this problem from different viewpoints; Allen and Gale (1988) use risk-allocation considerations, while Boot and Thakor (1993) use informational considerations. Note that the security design literature generally tries to explain how to maximize agents' welfare, while this paper and the other papers mentioned above try to explain how to maximize the firm's total market value, which is not necessarily the same thing.

The key heterogeneity in our model is the degree to which investors specialize in the precision of information they have on the returns of specific assets (which may be industries, firms, or risk factors). An investor with more precise information about an asset has a lower variance of beliefs about the return of that asset; specialization occurs when different investors have a lower variance for different assets. Our approach is inspired by recent research on investors with limited attention; in these models, investors have a finite capacity to reduce the variance of signals they receive, and must prioritize which signals to learn more about. Van Nieuwerburgh and Veldkamp (2009) incorporate limited attention into a Grossman and Stiglitz (1980)-type model of noisy rational expectations equilibrium. Under the assumption of mean-variance prefences, Gaussian returns, and a constraint on the product of perceived signal variances, investors will specialize in learning exclusively about a single asset type.

They show how this can explain the "home bias" puzzle where investors hold more of their home countries' equities than complete diversification would predict. Their paper develops a model with what they call *strategic substitutability* in information choice: the gain to learning about a specific asset decreases as more people learn about it. This provides a motive for specialization in the choice of which asset to learn about. Our paper does not explicitly incorporate a signal choice for investors, but we can interpret the exogenously given variances in our model as the result of such specialization.

The rest of the paper is organized as follows: Section 3 presents a simple two-investor, two-asset model with a diversification discount and shows how this is related to the degree of investor specialization. Section 4 extends this result to any number of investors and assets. Section 5 examines the configuration of investor beliefs that would result in the largest possible diversification discount. Section 6 discusses some of the empirical predictions of our model, and how they might be tested. Section 7 concludes.

3 The Model

We present a standard asset pricing model with heterogeneous investors, CARA utility, and normally distributed asset returns, as in Admati (1985), but with a finite number of investors. Initially, we will assume that investors' beliefs are given exogenously; later, we will allow investors to learn from prices and private signals.

There are I investors and two periods; investors trade in the first period and consume in the second. Each investor $i = 1, ..., I$ invests his initial wealth w_{0i} in a riskless asset and N risky assets. Agent i has population mass λ_i, where $\lambda_i > 0$ and $\sum_i \lambda_i = 1$. The riskless rate of return is assumed to be 1. Let $\tilde{F}$ denote the random vector of risky asset returns. Agent i's final wealth is

$$\tilde{w}_{1i} = w_{0i}R + D_i^T(\tilde{F} - PR) \tag{3.1}$$

P is the vector of market prices, and $D_i(P)$ is the vector of holdings of risky assets. Each investor i has exponential utility $u_i(w) = -\exp(w/\rho_i)$. Agent i's belief about the asset payoff $\tilde{F}$ is Gaussian, with mean μ^i and variance $\Sigma_i = E_i[(\tilde{F} - \mu^i)^T(\tilde{F} - \mu^i)]$. For now, we will take these parameters as exogenously given.

Given beliefs characterized by μ^i and Σ_i, optimal demand is given by

$$D_i(P) = \frac{1}{\rho_i}\Sigma_i^{-1}(\mu^i - P) \tag{3.2}$$

The aggregate supply of assets is given by the vector Z, which we will assume is a nonrandom vector of ones for now. Let $V_i = \frac{\rho_i}{\lambda_i}\Sigma_i$; this is the "effective" variance of investor i's belief, incorporating the investor's risk tolerance and population mass. In the rest of the paper, we will simply refer to V_i as the variance. In equilibrium, the market clearing condition is

$$\sum_{i=1}^{I} \lambda_i D_i(P) = \sum_{i=1}^{I} V_i^{-1}(\mu^i - P) = Z \tag{3.3}$$

Equilibrium prices are given by:

$$P = (\sum_{i=1}^{I} V_i^{-1})^{-1}(\sum_{i=1}^{I} V_i^{-1}\mu^i) - (\sum_{i=1}^{I} V_i^{-1})^{-1}Z \tag{3.4}$$

The first term in this expression, $(\sum_{i=1}^{I} V_i^{-1})^{-1}(\sum_{i=1}^{I} V_i^{-1}\mu^i)$ is equivalent to the Bayesian posterior mean that results from observing I normally distributed signals with realized values $\mu^1, ..., \mu^I$, and where the ith signal is known to have a variance of V_i. The quantity $(\sum_{i=1}^{I} V_i^{-1})^{-1}$ in the second term is the Bayesian posterior variance. Thus, we have:

Remark 3.1 (Markets as a Bayesian aggregator of information)**.** Equilibrium prices are identical to an economy with a representative investor who has "observed" the (scaled) beliefs of the I investors and combined them using Bayesian updating.

3.1 A Two-Asset Example

Suppose there are two assets $n = 1, 2$, and two investors $i = 1, 2$, each with the same size: $\lambda_1 = \lambda_2 = \frac{1}{2}$. We assume both types have the same risk aversion $\rho_1 = \rho_2 = \rho$. Investor i's belief has mean $\mu^i = (\mu^i_1, \mu^i_2)$ and variance Σ^i, where for $\beta \in (0, 1)$:

$$\Sigma_1 = \begin{bmatrix} \beta\sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \beta\sigma^2 \end{bmatrix}$$

We interpret this as specialization in information by investors; investor 1 has relatively more expertise about asset 1 than asset 2, and vice versa for investor 2. For example, the investor Warren Buffett is an expert on the financial industry, but by his own admission, knows little about the tech industry. Suppose that asset 1 models the "financial industry" while asset 2 models the "tech industry"; then an investor with greater expertise in the "financial industry" would have a belief covariance matrix with a smaller variance for asset 1. This specialization may arise if agent 1 has chosen to observe signals about the industry of asset 1 but not asset 2 (and vice-versa for agent 2). We will compare two regimes: a "combined-firm" regime where the two assets are held by a single firm with a single stock, and a "separate-stocks" regime where each asset has its own stock. In each case, we will assume that the supply of all stocks is normalized to 1.

First, consider the combined-firm regime. We assume there are no production efficiencies or inefficiencies caused by combining the two assets; the return of the single stock is simply the sum of the returns of the individual assets. The belief means of each agent for the combined firm are $\mu^1_1 + \mu^1_2$ and $\mu^2_1 + \mu^2_2$, respectively, and the variances of both agents are $(1 + \beta)\sigma^2$. The equilibrium price of the single stock is given by

$$P_{combined} = \left(\frac{1}{2\rho(1+\beta)\sigma^2} + \frac{1}{2\rho(1+\beta)\sigma^2}\right)^{-1} \left(\frac{1}{2\rho(1+\beta)\sigma^2}(\mu^1_1 + \mu^1_2) + \frac{1}{2\rho(1+\beta)\sigma^2}(\mu^2_1 + \mu^2_2) - 1\right) \tag{3.5}$$

$$= \frac{\mu^1_1 + \mu^1_2 + \mu^2_1 + \mu^2_2}{2} - \rho(1+\beta)\sigma^2 \tag{3.6}$$

Total market value is given by $P \cdot Z$, which is simply $P_{combined}$.

Now, consider the separate-stocks regime. The price vector is given by

$$P = \left(\frac{1}{2\rho}\Sigma_1^{-1} + \frac{1}{2\rho}\Sigma_2^{-1}\right)^{-1} \left(\frac{1}{2\rho}\Sigma_1^{-1}\begin{bmatrix}\mu_1^1 \\ \mu_2^1\end{bmatrix} + \frac{1}{2\rho}\Sigma_2^{-1}\begin{bmatrix}\mu_1^2 \\ \mu_2^2\end{bmatrix} - \begin{bmatrix}1 \\ 1\end{bmatrix}\right) \tag{3.7}$$

$$= \frac{1}{1+\beta}\begin{bmatrix}\mu_1^1 + \beta\mu_1^2 - 2\beta\rho\sigma^2 \\ \beta\mu_1^1 + \mu_2^2 - 2\beta\rho\sigma^2\end{bmatrix} \tag{3.8}$$

Combined market value is given by $P \cdot Z = p_1 + p_2$, which is equal to:

$$P_{separate} = \frac{\mu_1^1 + \beta\mu_1^2 + \beta\mu_1^1 + \mu_2^2 - 4\rho\beta\sigma^2}{1+\beta} \tag{3.9}$$

The difference between the two market values is:

$$P_{separate} - P_{combined} = \frac{(1-\beta)(\mu_1^2 - \mu_1^1 + \mu_2^1 - \mu_2^2 + 2(1-\beta)\rho\sigma^2)}{2(1+\beta)} \tag{3.10}$$

In general, if there are no restrictions on μ^1 and μ^2, the discount may be positive or negative.

3.1.1 Equal Belief Means

Consider the case where both agents have the same belief mean: $\mu^1 = \mu^2 = (\mu_1, \mu_2)$. Then we have

$$P_{combined} = (\mu_1 + \mu_2) - \rho(1+\beta)\sigma^2 \tag{3.11}$$

$$P_{separate} = (\mu_1 + \mu_2) - \rho\frac{4\beta\sigma^2}{1+\beta} \tag{3.12}$$

$$P_{separate} - P_{combined} = \rho\frac{(1-\beta)^2\sigma^2}{1+\beta} \tag{3.13}$$

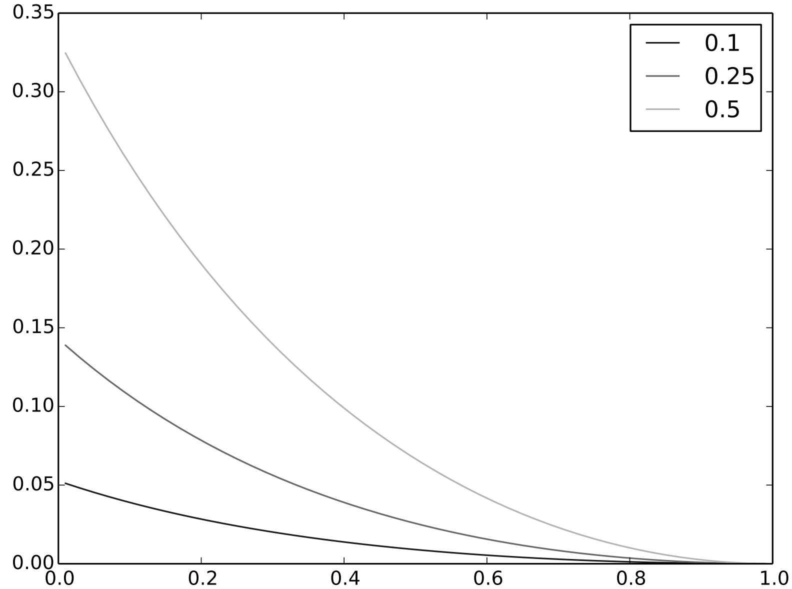

Figure 1: Discount as a fraction of the combined price for $\rho = 1, \frac{\sigma^2}{\mu} = \{0.1, 0.25, 0.5\}$

This expression is positive for $\beta \in (0, 1)$; its derivative with respect to β is $\rho\sigma^2 \left(1 - \frac{4}{(1+\beta)^2}\right)$, which is negative. The diversification discount increases with σ^2 and with a greater degree of specialization (that is, a lower β) among investors.

Suppose $\mu_1 = \mu_2 = \bar{\mu}$. We can express the discount (as a fraction of of $P_{combined}$) as a function of $\frac{\sigma^2}{\bar{\mu}}$:

$$\frac{P_{separate} - P_{combined}}{P_{combined}} = \left(\frac{1+\beta}{(1-\beta)^2}\left(\frac{2}{\rho\frac{\sigma^2}{\bar{\mu}}} - (1+\beta)\right)\right)^{-1} \tag{3.14}$$

Figure 1 plots the discount fraction for the parameter values $\rho = 1, \frac{\sigma^2}{\bar{\mu}} = \{0.1, 0.25, 0.5\}$. When the variance of payoffs is high compared to its mean and the degree of specialization is high, the discount can be economically significant.

Intuitively, we can explain this behavior as a type of market segmentation. Consider investor

i's demand for each individual asset in the separate-stocks regime:

$$D_i(P) = \frac{1}{\rho_i}\Sigma_i^{-1}(\mu^i - P) = \begin{bmatrix} D_{i,1} \\ D_{i,2} \end{bmatrix} = \frac{1}{\rho_i}\begin{bmatrix} (\beta\sigma^2)^{-1}(\mu_1^i - P_1) \\ (\sigma^2)^{-1}(\mu_2^i - P_2) \end{bmatrix} \tag{3.15}$$

The ratio of demand for asset 1 to asset 2 will depend on the ratio of the variance and the expected excess return for asset 1 compared to asset 2. Suppose $\mu_1^i = \mu_2^i$ and $P_1 = P_2$; then the ratio $D_{i,1}/D_{i,2}$ is equal to β (each agent has a greater demand for the asset in which he has more expertise). However, in the single-firm regime, all agents are constrained to own assets 1 and 2 in the same fixed proportion, resulting in lower expected utility for investors and hence lower demand. In the separate-firm regime, each investor could theoretically increase his relative holdings of the asset in which he has more expertise, resulting in a more preferred portfolio. The ability to actually achieve a more desirable ratio in equilibrium depends on the demand of other agents in the market; in our two-asset example, each agent prefers to hold relatively more of the asset that the other agent prefers to hold relatively less of, so the separate-stocks equilibrium does indeed achieve a more preferable ratio for both agents.

4 General Case

We can generalize this result to any number of assets and investor types, with investor beliefs that can have arbitrary correlations between asset returns, including non-diagonal covariance matrices. In order to do this, we will introduce tools drawn from the analysis of positive definite matrices; while this increases the technical burden for the reader, it allows us to give results that hold for any number of arbitrary covariance matrices. As far as we are aware, this is the first time that these tools have been applied to theoretical models of asset pricing.

4.1 Preliminaries

For any matrix A, let $\mathbf{sum}(A)$ denote the sum of all elements of A. Suppose we have an N-dimensional random vector $x = (x_1, ...x_N)^T$ with covariance matrix Σ; then the variance of $x_1 + ... + x_N$ is given by $\mathbf{sum}(\Sigma)$. Suppose an investor's belief covariance matrix for the separate-stocks regime is V; then, this investor's belief about returns for the combined firm must have variance $\mathbf{sum}(V)$. If V is diagonal, this is also equal to the *trace* of V, denoted $\mathbf{tr}(V)$. We will utilize the following definitions and results on positive definite matrices (proofs are given in the Appendix). In what follows, $A_1, ..., A_n, A'_1, ..., A'_n, A$, and B are arbitrary real-valued, symmetric, positive definite matrices of the same dimension.

Definition 4.1. We define the Loewner partial ordering: $A \geqslant_L (>_L) B$ whenever $A - B$ is positive semidefinite (positive definite).

The Loewner ordering has the following statistical interpretation (Horn (1990), p.141): suppose $\tilde{X}, \tilde{Y}$ are $\mathbb{R}^n$-valued random variables, with $Var(\tilde{X}) = A$ and $Var(\tilde{Y}) = B$. Then $A \geqslant_L (>_L) B$ iff for any nonzero $c \in \mathbb{R}^n$, $Var(c \cdot \tilde{X}) \geqslant (>) Var(c \cdot \tilde{Y})$.

Definition 4.2. The *parallel sum* of $A_1, ...A_n$ is denoted $A_1 : ... : A_n = (A_1^{-1} + ... + A_n^{-1})^{-1}$. The *harmonic mean* of $A_1, ..., A_n$ is $n(A_1 : ... : A_n)$.

The Bayesian posterior after observing normally distributed signals with realized values $\mu^1, ...\mu^I$ and known variances $V_1, ..., V_I$ has mean $(V_1 : ... : V_I)(\sum_{i=1}^I V_i^{-1}\mu^i)$ and variance $V_1 : ... : V_I$.

Lemma 4.1. *(monotonicity and joint concavity of parallel sum):* $A_1 : ... : A_n$ *is monotonically increasing in* $A_1, ..., A_n$*:*

$$A_1 \geqslant_L A'_1, ..., A_n \geqslant_L A'_n \Rightarrow A_1 : ... : A_n \geqslant_L A'_1 : ... : A'_n \tag{4.1}$$

and jointly concave *in* $A_1, ..., A_n$. *For* $t \in [0, 1]$*:*

$$(tA_1 + (1-t)A'_1) : ... : (tA_n + (1-t)A'_n) \geqslant_L t(A_1 : ... : A_n) + (1-t)(A'_1 : ... : A'_n) \tag{4.2}$$

Lemma 4.2. *(inequality between parallel sum of sum/trace and sum/trace of parallel sum):*

1. $\boldsymbol{sum}(A_1 : ... : A_n) \leqslant \boldsymbol{sum}(A_1) : ... : \boldsymbol{sum}(A_n)$.

2. $\boldsymbol{tr}(A_1 : ... : A_n) \leqslant \boldsymbol{tr}(A_1) : ... : \boldsymbol{tr}(A_n)$.

Lemma 4.3. *(joint concavity of sum/trace of parallel sum):* $\boldsymbol{sum}(A_1 : ... : A_n)$ *and* $\boldsymbol{tr}(A_1 : ... : A_n)$ *are jointly concave over* $A_1, ..., A_n$.

Lemma 4.4. *(equality of sum/trace of parallel sum): Suppose* $A_1, ..., A_n$ *are* simultaneously diagonalizable*: there exists an orthornormal* P *such that for each* $i = 1, ..., n$, $A_i = P^T D_i P$, *where* D_i *is the diagonal matrix containing the eigenvalues of* A_i*, and the rows of* P *are the shared eigenvectors of* $A_1, ..., A_n$*. If* $D_i = \alpha_i D$ *where* $\alpha_i \neq 0$ *for* $i = 1, ..., n$*, that is, each* D_i *matrix is some matrix* D *multiplied by a nonzero scalar, then* $\boldsymbol{tr}(A_1 : ... : A_n) = \boldsymbol{tr}(A_1) : ... : \boldsymbol{tr}(A_n)$ *and* $\boldsymbol{sum}(A_1 : ... : A_n) = \boldsymbol{sum}(A_1) : ... : \boldsymbol{sum}(A_n)$.

The economic interpretation of the conditions of Lemma 4.4 is that all investors agree on what the statistically independent risk factors affecting asset returns are (the eigenvectors), though they may disagree on how much each risk factor contributes to a specific asset's return (the eigenvalues). Furthermore, all investors agree on the relative contribution of each risk factor to every asset, though they may disagree on the absolute contribution.

Let I, the number of investor types, and N, the number of assets, be positive integers.

Assumption 4.1.

1. The risk-free interest rate R is 1.

2. The vector of asset supplies Z is a vector of ones.

3. Every investor's belief covariance matrix V_i is positive definite.

In the combined-firm regime, the equilibrium price is given by

$$P_{combined} = (\mathbf{sum}(V_1) : ... : \mathbf{sum}(V_I)) \left(\sum_i^I \mathbf{sum}(V_i)^{-1} \mathbf{sum}(\mu^i)) - 1 \right) \tag{4.3}$$

In the separate-stocks regime, the equilibrium price vector is given by

$$\begin{bmatrix} p_1 \\ \vdots \\ p_N \end{bmatrix} = (\sum_{i=1}^{I} V_i^{-1})^{-1}(\sum_{i=1}^{I} V_i^{-1}\mu^i) - (\sum_{i=1}^{I} V_i^{-1})^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{4.4}$$

The combined value of all stocks is given by

$$P_{separate} = \sum_{n=1}^{N} p_n = \mathbf{sum}\left((V_1 : ... : V_I)(\sum_{i=1}^{I} V_i^{-1}\mu^i) \right) - \mathbf{sum}(V_1 : ... : V_I) \tag{4.5}$$

Now, we can compare prices in the two regimes under different assumptions about investor beliefs.

Proposition 4.1. *(diversification discount under equal belief means): Suppose all investors have the same belief means: $\mu^i = \mu$ for $i = 1, ..., I$; then $P_{combined} \leqslant P_{separate}$.*

Proof.

$$P_{combined} = \mathbf{sum}(\mu) - [\mathbf{sum}(V_1) : ... : \mathbf{sum}(V_I)] \tag{4.6}$$

$$P_{separate} = \mathbf{sum}(\mu) - \mathbf{sum}(V_1 : ... : V_I) \tag{4.7}$$

By Lemma 4.2, $\mathbf{sum}(V_1 : ... : V_I) \leqslant \mathbf{sum}(V_1) : ... : \mathbf{sum}(V_I)$, therefore $P_{combined} \leqslant P_{separate}$.

□

Proposition 4.2. *(no discount with common eigenvectors and eigenvalues are equal up to scale factor): Suppose $V_1, ..., V_i$ satisfy the conditions in Lemma 4.4 (common eigenvectors and eigenvalues are equal up to a scale factor). Then $P_{combined} = P_{separate}$.*

Under the equal belief mean assumption, there is never a diversification premium. A zero discount holds if investors have homogeneous beliefs, but can also occur if all investors agree on the independent risk factors and their relative contributions to asset prices. Changes in the distribution of risk tolerance or population mass among investor types, therefore, do not lead to a discount. This also ensures that if a spinoff occurs when there is no heterogeneity of beliefs, there will be no diversification discount and hence no increase in value for the firm. This is in contrast to models where the informativeness of stock prices always increases with a spinoff.

If both belief means and variances are arbitrary, a discount, premium, or neither may exist. To see this, consider the equilibrium price equation 3.4 with $I = 2$. The first term is a Bayesian posterior mean after observing two signals (or alternatively, updating a prior with one signal); in the combined-firm regime and separate-stocks regime, it is given by the first term in Equations 4.3 and 4.5, respectively. It is known that in Bayesian inference with Gaussian signals, if only the prior and observation locations are known while the variances are free variables, then the posterior mean may lie virtually anywhere (Chamberlain and Leamer (1976), Theorem 1). Therefore, the element sum of the posterior mean in the separate-stocks regime may be arbitrarily higher or lower than in the combined-firm regime, as was demonstrated by Eqn 3.10 for the 2-investor, 2-asset case.

4.2 Private Signals and Learning from Prices

We have seen that a discount exists if belief means are equal, but may or may not exist if belief means are allowed to vary arbitrarily. In this section we consider the multi-asset, noisy REE model of Admati (1985). Noisy REE models, beginning with Grossman and Stiglitz (1980) and Hellwig (1980), relax the strong assumption of equal belief means in a structured way, by assuming agents share a common prior belief about (unobserved) asset returns, then endowing agents with a private signal that is correlated with asset returns. Furthermore, agents are allowed to learn from prices, which have dual roles of clearing markets while also serving as a signal that reveals some part of other agents' private information. Hellwig (1980) developed a multi-agent REE model with one risky asset; Admati (1985) extended this to the case with multiple risky assets.

Assume two time periods; at t=0, the economy receives a random shock $\tilde{y}$, which may consist of private signals to investors, as well as shocks to asset supplies. At t=1, investors trade and reach equilibrium. We assume that for any realization of $\tilde{y}$, the economy will reach an equilibrium in which:

- each investor type i has Gaussian beliefs with *random* mean $\mu^i(\tilde{y})$, and *deterministic* variance V_i.
- the asset supply vector is random, $Z(\tilde{y})$.

This implies that the equilibrium price will be in the form of Equation 3.4; i.e. it is equivalent to having a representative investor whose beliefs are a combination of the investors' beliefs via Bayesian updating. We assume that $E[Z(\tilde{y})]$ is a vector of ones. Taking time-0 expectations of Equations 3.4, 4.3 and 4.5, we get:

$$E[PR] = (\sum_{i=1}^{I} V_i^{-1})^{-1}(\sum_{i=1}^{I} V_i^{-1} E\left[\mu^i(\tilde{y})\right]) - (\sum_{i=1}^{I} V_i^{-1})^{-1} E[Z(\tilde{y})] \tag{4.8}$$

$$\begin{aligned} E[P_{combined}] = {} & (\mathbf{sum}(V_1) : ... : \mathbf{sum}(V_I))(\sum_{i}^{I} \mathbf{sum}(V_i)^{-1} \mathbf{sum}(E\left[\mu^i(\tilde{y})\right])) \\ & - (\mathbf{sum}(V_1) : ... : \mathbf{sum}(V_I)) \end{aligned} \tag{4.9}$$

$$E[P_{separate}] = \mathbf{sum}\left((V_1 : ... : V_I)(\sum_{i=1}^{I} V_i^{-1} E\left[\mu^i(\tilde{y})\right])\right) - \mathbf{sum}(V_1 : ... : V_I) \tag{4.10}$$

If the expected value of investors' belief means are equal, i.e. $E\left[\mu^i(\tilde{y})\right] = \bar{\mu}$ for $i = 1, ..., I$, then the proof of Proposition 4.1 goes through, applied to $E[P_{combined}]$ and $E[P_{separate}]$.

Proposition 4.3. *(diversification discount in expectation) Suppose that any realization of $\tilde{y}$ results in an equilibrium at t=1 in which all investors have Gaussian beliefs with mean $\mu^i(\tilde{y})$ and variance V_i, and that $E\left[\mu^i(\tilde{y})\right] = \bar{\mu}$ for $i = 1, ..., I$ and $E[Z(\tilde{y})]$ is a vector of ones. Then $E[P_{combined}] \leqslant E[P_{separate}]$. If $V_1, ..., V_I$ have common eigenvectors and their eigenvalues are equal up to a scale factor, then equality holds.*

If a discount holds in expectation, then a risk-neutral firm would optimally choose to break itself up into the separate-stocks regime rather than stay in the combined-firm regime. Note that we do not make any assumption on how equilibrium beliefs are reached, so long as they are Gaussian. We show that this result holds in the noisy REE model of Admati (1985), which we briefly summarize below. Let $\tilde{F}$ denote the unobserved vector of asset returns, and let $\tilde{Z}$ denote the asset supply; both are Gaussian random variables with means $\bar{F}, \bar{Z}$ respectively. Each investor i receives a private signal $\tilde{y}_i = \tilde{F} + \tilde{\epsilon}_i$, where $\tilde{\epsilon}_i$ is iid zero-mean Gaussian noise. If we assume that a linear equilibrium pricing function exists, then $(\tilde{F}, \tilde{P}, \tilde{Z}, \tilde{y}_1, ..., \tilde{y}_I)$ are jointly Gaussian with deterministic variances. In a noisy REE, each investor is assumed to know the correct joint distribution of $(\tilde{F}, \tilde{y}_i, \tilde{P})$; after observing his private signal $\tilde{y}_i$ and the equilibrium price $\tilde{P}$, his belief is the conditional distribution of $\tilde{F}|\tilde{y}_i, \tilde{P}$, which is also Gaussian. Let $\mu^i = E[\tilde{F}|\tilde{y}_i, \tilde{P}]$ denote the mean of this conditional distribution. By the law of iterated expectations, $E[\mu^i] = E\left[E[\tilde{F}|\tilde{y}_i, \tilde{P}]\right] = E[\tilde{F}] = \bar{F}$. Therefore, the condition that expected belief means are equal is satisfied, and Proposition 4.3 applies.

4.2.1 Infinite vs. Finite Agents

The REE literature typically assumes that there are infinitely many investors, each of who observes an i.i.d. Gaussian private signal. It is further assumed that a law of large numbers applies, and the *realized* average signal across investors is then assumed to be equal to the *expectation* of the signal distribution, a.s. This makes an analytic solution tractable, but also guarantees that the conditions for Lemma 4.1 hold a.s. To see this, suppose that there are a continuum of investors, indexed by $i \in [0, 1]$. After observing their private signal, each investor's belief mean is an i.i.d. Gaussian random variable: $\tilde{\mu}^i \sim N(\bar{\mu}, \Sigma_\mu)$. If we assume that the realized average belief across the investor population is equal to its expectation, a.s., then the first term in Eq 3.4 becomes (Admati (1985), Eq 16):

$$\left(\int_0^1 V_i^{-1}\right)^{-1} \int_0^1 V_i^{-1} \tilde{\mu}^i di = \left(\int_0^1 V_i^{-1}\right)^{-1} \int_0^1 V_i^{-1} \bar{\mu} = \bar{\mu} \qquad \text{a.s.} \tag{4.11}$$

This term is equal in the combined and separate stocks regime a.s., therefore Lemma 4.1 holds a.s. The assumption of infinite or finite agents results in different behavior at equilibrium; with infinite agents, a diversification premium will not occur as a result of agents' realized beliefs a.s., but may occur with finite agents.[1]

4.3 Extension to Dynamic Settings

We can extend this result to dynamic REE models in which investor beliefs are Gaussian in every period. A straightforward application of Prop. 4.3 implies that unconditionally, an expected discount will exist in every period, but when conditioning on past information, a discount may or may not exist, unless the equality conditions in Prop. 4.3 are satisfied. In theory, this should hold for both finite-horizon and stationary infinite-horizon models, but we are unaware of an existence proof of REE in the infinite-horizon case with a finite number of agents. Existence in the infinite-horizon model has been proven for infinite agents (Naik 1997), implying that a discount exists a.s. in every period.

4.3.1 The Model of Brennan & Cao (1997)

We present a finite-agent version of the model of Brennan and Cao (1997), which is a finite-horizon extension of Admati (1985). As before, $\tilde{F} \sim N(\bar{F}, \Sigma_F)$ is the vector of asset returns, which is realized at time $t = T$. Before time T, there are T trading periods $t = 0, ..., T-1$. There is no intermediate consumption; investors only care about their final wealth at $t = T$. The risk-free rate is assumed to be zero. Before each trading period, each investor i receives a private signal $\tilde{Y}_t^i = \tilde{F} + \tilde{\epsilon}_{Y,t}^i$, where $\tilde{\epsilon}_{Y,t}^i \sim N(0, S_{Y,t}^i)$, and a public signal, $\tilde{X}_t = \tilde{F} + \tilde{\epsilon}_{X,t}$, where $\tilde{\epsilon}_{X,t} \sim N(0, S_{X,t})$. Asset supply is a random vector: $\tilde{Z}_t = \bar{Z} + \tilde{\epsilon}_{Z,t}$, where $\tilde{\epsilon}_{Z,t} \sim N(0, U_{X,t})$. $U_{X,t}$ and $S_{Y,t}^i$ and $S_{X,t}$ for $i = 1, ..., I, t = 0, ...T-1$ are assumed to be positive definite,

[1] Hellwig (1980) proved existence of a linear REE in the single-asset case with infinite and finite agents; Admati (1985) only proved existence in the multi-asset case with infinite agents. Existence of equilibrium with finite agents was proven in Carpio and Guo (2016).

and $\tilde{F}$, $\tilde{\epsilon}^i_{Y,t}, \tilde{\epsilon}_{X,t}, \tilde{\epsilon}_{Z,t}$ for $i = 1, ..., I, t = 0, ..., T-1$ are assumed to be mutually independent. Investor i's information set at time t is denoted by $\Omega^i_t = \{\tilde{P}_j, \tilde{Y}^i_j, \tilde{X}_j | j = 0, ..., t\}$. Then in each period t, the following conditions hold (Brennan and Cao (1997), Eqns 1-2): (i) investor i's belief conditional on his information set is Gaussian, with mean $\tilde{\mu}^i_t = E[\tilde{F}|\Omega^i_t]$ and deterministic variance $\Sigma^i_t = Var(\tilde{F}|\Omega^i_t)$; (ii) optimal demand and the equilibrium price condition are of the same form as Eqns 3.2 and 3.4, respectively, using $\tilde{\mu}^i_t$ and Σ^i_t as the parameters of each investor's belief. Therefore, Remark 3.1 and Prop. 4.3 hold in each period. If expectations only condition on the common prior available before $t = 0$, then an expected diversification discount holds in each period, but may or may not exist when conditioning on all the information available at time t. As before, if there are an infinite number of investors, then a diversification discount will exist in each period a.s. Existence of equilibrium in each period with a finite number of investors can be established by using backwards induction. In the last period $T-1$, the situation is identical to the model of Admati (1985), and we can apply the result of Carpio and Guo (2016) to prove existence. Then, beliefs in the previous period will be Gaussian, and once again the situation reduces to the static case; by proceeding until we reach the first period 0, we prove that equilibrium exists in each period.

5 Beliefs That Maximize Combined Market Value

We can ask the question: what distribution of belief variances $V_1, ..., V_I$ among investors will result in a larger discount (and hence a larger incentive to create a spinoff)? The diversification discount is larger when $\mathbf{sum}(V_1 : ... : V_I)$ is smaller. The joint concavity of $\mathbf{sum}(V_1 : ... : V_I)$ (by Lemma 4.3) has two immediate implications. First, the discount is smaller when $V_1, ..., V_I$ is a convex combination of other possible collections of variances. Second, a concave function over a convex region is minimized (and therefore the discount is maximized) at an extreme point (i.e. a point that is not a convex combination of other points); if we can formulate an appropriate convex feasible region, we can find the variances that maximize the discount.

5.1 Simultaneously diagonalizable variances

A tractable case is to assume all investors' variances have the same eigenvectors: there exists an orthonormal P such that $V_i = P^T D_i P$, where D_i is a diagonal matrix; the rows of P are the shared eigenvectors and the diagonal entries of D_i are the associated eigenvalues of V_i. As before, we interpret this to mean that all investors agree on the independent risk factors underlying asset returns. The next results show that we can reduce the feasible set to the set of diagonal positive definite matrices.

Lemma 5.1. *(parallel sum for simultaneously diagonalizable matrices) Suppose $A_1, ..., A_n$ are positive definite, and $A_i = P^T D_i P$ for $i = 1, ..., n$, where P is orthogonal and $D_1, ..., D_n$ are diagonal. Then*

$$A_1 : ... : A_n = P^T(D_1 : ... : D_n)P \tag{5.1}$$

Lemma 5.2. *(joint concavity for simultaneously diagonalizable matrices) Let $D_1, ..., D_n$, $D'_1, ..., D'_n$ be diagonal, positive definite matrices, and let P be an orthogonal matrix of the same dimension. For $t \in [0, 1]$:*

$$\begin{aligned} P^T(tD_1 + (1-t)D'_1)P : ... : P^T(tD_n + (1-t)D'_n)P \geqslant \\ tP^T(D_1 : ... : D_n)P + (1-t)P^T(D'_1 : ... : D'_n)P \end{aligned} \tag{5.2}$$

and $\boldsymbol{sum}(P^T D_1 P : ... : P^T D_n P)$ *is jointly concave over $D_1, ..., D_n$.*

Therefore, we will assume that each V_i is diagonal, and we will seek an extreme point of a suitably defined convex subset of the positive definite diagonal matrices. Suppose that all investors have the same beliefs about the returns of the combined firm, which implies that the trace of each V_i is equal. We can set up an additive constraint on the diagonal entries by setting a lower bound on each diagonal entry of V_i; we interpret this as setting a minimum amount of uncertainty for each asset. Let $[V]_n$ denote the n-th diagonal entry of V.

Suppose $\mathbf{tr}(V_i) = N\underline{\sigma}^2 + T$ for $i = 1, ..., I$, where $\underline{\sigma}^2 \geqslant 0, T \geqslant 0$, and $[V_i]_n \geqslant \underline{\sigma}^2$ for $i = 1, ...I, n = 1, ...N$. We say that V is *specialized in variance* in asset n if all diagonal entries of V are equal to $\underline{\sigma}^2$, except for $[V]_n$ which is equal to $\underline{\sigma}^2 + T$.

Proposition 5.1. *(maximization of dividend discount under trace constraint): The dividend discount will be maximized when each V_i is specialized in variance for some asset, and each asset has at least $\lfloor N/I \rfloor$ V_is that specialize in it, and no asset has more than $\lfloor N/I \rfloor + 1$.*

These results imply that the incentive to generate spinoffs depends on the degree of specialization among the investor community. When there is less overlap of information among investor groups, the discount is larger.

6 Discussion and Empirical Implications

We have shown that a diversification discount exists when different investor types have have more precise information about different assets, and that the discount is maximized (and therefore the incentive for a spinoff is largest) when the population of investors ”specializes” in the sense that there is minimum overlap in the types of assets that each investor type knows more about. The results in this paper make predictions that can be tested empirically; we list some of these predictions and how they might be tested.

First, spinoff and merger patterns are affected by the interests of the investor/analyst community. If investors form into groups focusing on a specific sector, industry, country, etc, then a firm splitting up is more likely when it contains divisions that align with these groupings (and conversely, a merger is less likely). Furthermore, spinoffs whose businesses are aligned with this grouping will provide greater abnormal returns than those that do not. On the other hand, if there is no investor/analyst group following a particular sector, then there will be little incentive to spin off business segments in this sector, even if the segment is relatively independent of the rest of the firm.

Second, spinoffs of unrelated divisions should be more likely, and should return a higher abnormal return, since divisions in unrelated industries or market sectors will likely be the focus of different investor groups. Most other theoretical models of the diversification discount also make this prediction; one possible way to distinguish the predictions of this model from others is by comparing returns of spinoffs of unrelated divisions when there is little investor attention in a particular sector, to when it is high.

Third, in contrast with some other models of spinoffs, the incentive to split comes solely from investor behavior, with no necessary connection to management behavior or even firm productivity. Thus, our model predicts spinoffs of divisions or assets that seem to be currently adequately managed, if there is some investor group that would pay attention to it. For example, in recent years it has become popular for companies to spin off their real estate holdings into REITs, under the hypothesis that these assets were undervalued when held in the original firm. Furthermore, spinoffs may generate abnormal returns, even if there is no improvement in the actual productivity of the spinoff.

Fourth, if there is a time trend in the amount of investor interest in a specific sector, or if there is a trend towards more investor specialization in general, then there should be a corresponding trend in spinoffs. Interest in some specific sectors has varied over time (for example; Internet stocks, "social media" stocks, China stocks, etc). We can try to measure the degree of investor or analyst specialization into distinct sectors, and compare with the number and valuation of spinoffs in these sectors over time. Also, it seems that specialization of investors in general has been steadily increasing over time, due to improvements in information technology and financial sophistication. We could check if there has been a corresponding rise in spinoffs and valuations of all types over time.

7 Conclusion

We have shown that in the CARA-Gaussian framework, a diversification discount exists if investors have a specific type of heterogeneity in beliefs: different investors have more precise information about different assets. We interpret this as specialization in expertise or learning about specific firms or industries; this can explain the perception that the market gives a higher valuation to "pure plays". An expected discount exists in noisy rational expectations models that allow learning from prices. The discount (and therefore the incentive for a spinoff) increases with the degree of specialization among the population of investors. Our result depends only on heterogeneity of investor beliefs, and does not require "noise traders" or bounded rationality by investors. Furthermore, our results apply to any number of investors and assets with arbitrary correlations between assets, and introduces tools from the analysis of positive definite matrices into theoretical asset pricing.

For future research, we can ask what kind of financial structure will maximize the total market value for the firm, given its collection of assets and given the beliefs in the population. We can also examine the effect of investor heterogeneity in risk aversion and in the means of beliefs about asset returns. Finally, we can test the empirical implications of this model; the model predicts that spinoffs are more likely and should result in higher gains in total market value when they match the specialization pattern of the investing population.

8 Appendix: Proofs

Lemma 4.1 (monotonicity and joint concavity).

Proof. The case $n = 2$ is proved in Bhatia (2007), Theorem 4.1.1. We will prove the cases $n > 2$ by induction. Note that parallel sum is associative: $A_1 : ... : A_k : A_{k+1} = (A_1 : ... : A_k) : A_{k+1}$. Suppose that monotonicity holds for $n = k$, and $A_1 \geqslant_L A'_1, ..., A_{k+1} \geqslant_L A'_{k+1}$. Then

$$(A_1 : ... : A_k) : A_{k+1} \geqslant_L (A'_1 : ... : A'_k) : A'_{k+1} \Rightarrow \tag{8.1}$$

$$A_1 : ... : A_k : A_{k+1} \geqslant_L A'_1 : ... : A'_k : A'_{k+1} \tag{8.2}$$

Suppose that joint concavity holds for $n = k$. Then for $t \in [0, 1]$,

$$(tA_1 + (1-t)A'_1) : ... : (tA_k + (1-t)A'_k) : (tA_{k+1} + (1-t)A'_{k+1}) \tag{8.3}$$

$$\geqslant_L (t(A_1 : ... : A_k) + (1-t)(A'_1 : ... : A'_k)) : (tA_{k+1} + (1-t)A'_{k+1}) \tag{8.4}$$

$$\geqslant_L t(A_1 : ... : A_k) : A_{k+1} + (1-t)(A'_1 : ... : A'_k) : A'_{k+1} \tag{8.5}$$

$$= t(A_1 : ... : A_{k+1}) + (1-t)(A'_1 : ... : A'_{k+1}) \tag{8.6}$$

Therefore, monotonicity and joint concavity hold for all $n \geqslant 2$. □

Definition 8.1. Let $\mathbb{M}_n$ denote the set of all $n \times n$ real-valued matrices. A linear map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ is *positive* if $\Phi(A)$ is positive semidefinite whenever A is positive semidefinite.

Lemma 8.1. *(inequality of positive linear map of parallel sum): Let Φ be any positive linear map on $\mathbb{M}_n$, and let $A_1, ...A_n$ be positive semidefinite. Then $\Phi(A_1 : ...A_n) \leqslant_L \Phi(A_1) : ... : \Phi(A_n)$.*

Proof. The case $n = 2$ is proved in Bhatia (2007), Theorem 4.1.5. Suppose it holds for $n = k$. Then

$$\Phi(A_1) : ... : \Phi(A_k) \geqslant_L \Phi(A_1 : ... : A_k) \tag{8.7}$$

$$(\Phi(A_1) : ... : \Phi(A_k)) : \Phi(A_{k+1}) \geqslant_L \Phi(A_1 : ... : A_k) : \Phi(A_{k+1}) \tag{8.8}$$

$$\Phi(A_1) : ... : \Phi(A_k) : \Phi(A_{k+1}) \geqslant_L \Phi(A_1 : ... : A_k) : \Phi(A_{k+1}) \tag{8.9}$$

by associativity, and the right hand side is $\geqslant_L \Phi((A_1 : ... : A_k) : A_{k+1}) = \Phi(A_1 : ... : A_k : A_{k+1})$. Therefore, it holds for all $n \geqslant 2$. □

Lemma 4.2 (inequality for sum/trace).

Proof. **sum** and **tr** are positive linear maps (Bhatia (2007) Example 2.2.1). Therefore, Lemma 8.1 applies. □

Lemma 4.3 (joint concavity of sum/trace of parallel sum).

Proof. First, we show that **sum** and **tr** are monotone. For **tr**, see Bernstein (2009), Corollary 8.4.10. For **sum**, let e denote a column vector of ones; $\mathbf{sum}(A) = e^T Ae$. By the definition of positive semidefinite, $A \geqslant_L B \Rightarrow e^T(A - Be) \geqslant 0 \Rightarrow e^T Ae \geqslant e^T Be \Rightarrow \mathbf{sum}(A) \geqslant \mathbf{sum}(B)$.

By joint concavity of the parallel sum, for $t \in [0, 1]$:

$$(tA_1 + (1-t)A'_1) : ... : (tA_n + (1-t)A'_n) \geqslant_L t(A_1 : ... : A_n) + (1-t)(A'_1 : ... : A'_n) \quad (8.10)$$

$$\mathbf{tr}((tA_1 + (1-t)A'_1) : ... : (tA_n + (1-t)A'_n)) \geqslant_L \mathbf{tr}(t(A_1 : ... : A_n) + (1-t)(A'_1 : ... : A'_n)) \quad (8.11)$$

$$= t\mathbf{tr}(A_1 : ... : A_n) + (1-t)\mathbf{tr}(A'_1 : ... : A'_n) \quad (8.12)$$

The proof is identical for **sum**. □

Lemma 8.2 (variance of sum of variables)**.** *Suppose* $\tilde{X} = (\tilde{x}_1, ..., \tilde{x}_n)$ *is a n-dimensional random vector with variance* A*, diagonalized as* $A = P^T DP$*, with* P *an orthornormal matrix whose rows are the eigenvectors of* A *and* D *a diagonal matrix containing the eigenvalues of* A*. Then* $\boldsymbol{sum}(A) = Var(\sum_i \tilde{x}_i) = \sum_i(\sum_j P_{i,j})^2 D_{i,i}$.

Proof. Let $\tilde{Y} = (\tilde{y}_1, ..., \tilde{y}_n) \sim N(0, I_n)$. Then $\tilde{X} = P^{-1}D^{\frac{1}{2}}\tilde{Y}$ and $\tilde{x}_i = \sum_{j=1}^n P_{j,i}(D_{j,j})^{\frac{1}{2}}\tilde{y}_j$, and $\mathbf{sum}(A) = Var(\sum_i \tilde{x}_i) = \sum_i(\sum_j P_{i,j})^2 D_{i,i}$. □

Lemma 4.4 (equality of sum/trace of parallel sum).

Proof. For **tr**, applying Lemma 5.1:

$$\mathbf{tr}(A_1 : ... : A_n) = \mathbf{tr}(P^T D_1 P : ... : P^T D_n P) = \mathbf{tr}(P^T D_1 : ... : D_n P) \quad (8.13)$$

$$= \mathbf{tr}(D_1 : ... : D_n PP^T) = \mathbf{tr}(\alpha_1 D : ... : \alpha_n D) = (\alpha_1 : ... : \alpha_n)\mathbf{tr}(D) \quad (8.14)$$

$$\mathbf{tr}(A_1) : ... : \mathbf{tr}(A_n) = \mathbf{tr}(P^T D_1 P) : ... : \mathbf{tr}(P^T D_n P) = \mathbf{tr}(D_1 PP^T) : ... : \mathbf{tr}(D_n PP^T) \quad (8.15)$$

$$= \mathbf{tr}(D_1) : ... : \mathbf{tr}(D_n) = \mathbf{tr}(\alpha_1 D) : ... : \mathbf{tr}(\alpha_n D) = (\alpha_1 : ... : \alpha_n)\mathbf{tr}(D) \quad (8.16)$$

For **sum**, applying Lemma 8.2:

$$\mathbf{sum}(A_1 : ... : A_n) = \mathbf{sum}(P^T D_1 P : ... : P^T D_n P) = \mathbf{sum}(P^T D_1 : ... : D_n P) \quad (8.17)$$

$$= \sum_i (\sum_j P_{i,j})^2 (D_1 : ... : D_n)_{i,i} = \sum_i (\sum_j P_{i,j})^2 (\alpha_1 : ... : \alpha_n) D_{i,i} \tag{8.18}$$

$$\mathbf{sum}(A_1) : ... : \mathbf{sum}(A_n) = \mathbf{sum}(P^T D_1 P) : ... : \mathbf{sum}(P^T D_n P) \tag{8.19}$$

$$= \sum_i (\sum_j P_{i,j})^2 (D_1)_{i,i} : ... : \sum_i (\sum_j P_{i,j})^2 (D_n)_{i,i} \tag{8.20}$$

$$= \sum_i (\sum_j P_{i,j})^2 \alpha_1 D_{i,i} : ... : \sum_i (\sum_j P_{i,j})^2 \alpha_n D_{i,i} \tag{8.21}$$

$$= \sum_i (\sum_j P_{i,j})^2 (\alpha_1 : ... : \alpha_n) D_{i,i} \tag{8.22}$$

□

Proposition 4.2 (no discount with common eigenvectors and eigenvalues are equal up to scale factor).

Proof.

$$P_{combined} = (\mathbf{sum}(\alpha_1 P^T DP) : ... : \mathbf{sum}(\alpha_I P^T DP)) \left(\sum_i^I \mathbf{sum}(\alpha_i P^T DP)^{-1} \mathbf{sum}(\mu^i) - 1 \right) \tag{8.23}$$

$$= (\alpha_1 : ... : \alpha_I) \mathbf{sum}(P^T DP) \left(\mathbf{sum}(P^T DP)^{-1} \sum_i^I \alpha_i^{-1} \mathbf{sum}(\mu^i) - 1 \right) \tag{8.24}$$

$$= (\alpha_1 : ... : \alpha_I) (\sum_i^I \alpha_i^{-1} \mathbf{sum}(\mu^i)) - (\alpha_1 : ... : \alpha_I) \mathbf{sum}(P^T DP) \tag{8.25}$$

$$P_{separate} = \mathbf{sum} \left((\alpha_1 P^T DP : ... : \alpha_I P^T DP) \left(\sum_{i=1}^I \alpha_i P^T DP \mu^i \right) \right) - \mathbf{sum}(\alpha_1 P^T DP : ... : \alpha_I P^T DP) \tag{8.26}$$

$$= (\alpha_1 : ... : \alpha_I) \mathbf{sum} \left(\sum_{i=1}^I \alpha_i^{-1} \mu^i \right) - (\alpha_1 : ... : \alpha_I) \mathbf{sum}(P^T DP) \tag{8.27}$$

These two expressions are equal, since the first term in both expressions is simply $\sum_i^I \sum_n^N \alpha_i^{-1} \mu_n^i$. Therefore, there is no discount or premium. □

Lemma 5.1 (parallel sum for simultaneously diagonalizable matrices).

Proof. Since P is orthogonal, $P^T = P^{-1}$. The eigenvectors of A^{-1} are the same as A.

$$A_1 : ... : A_n = ((P^T D_1 P)^{-1} + ... + (P^T D_n P)^{-1})^{-1} \tag{8.28}$$

$$= (P^T D_1^{-1} P + ... + P^T D_1^{-1} P)^{-1} = (P^T (D_1^{-1} + ... + D_n^{-1}) P)^{-1} \tag{8.29}$$

$$= P^T (D_1^{-1} + ... + D_n^{-1})^{-1} P = P^T (D_1 : ... : D_n) P \tag{8.30}$$

□

Lemma 5.2 (joint concavity for simultaneously diagonalizable matrices).

Proof. The left hand side is equal to $tP^T D_1 P + (1-t) P^T D_1' P : ... : tP^T D_n P + (1-t) P^T D_n' P$. The right hand side is equal to $t(P^T D_1 P : ... : P^T D_n P) + (1-t)(P^T D_1' P : ... : P^T D_n' P)$. The results follow from Lemma 4.1 and Lemma 4.3. □

To prove Proposition 5.1 (maximization of dividend discount under trace constraint), we first establish two lemmas.

Lemma 8.3. *(convexity of weighted harmonic mean): Suppose $x, y \in \mathbb{R}$, $x > y > 0$. For $t \in [0,1]$, let $h(t; x, y)$ denote the weighted harmonic mean of x and y:*

$$h(t; x, y) = \left(\frac{1-t}{x} + \frac{t}{y} \right)^{-1} \tag{8.31}$$

Then $h(t; x, y)$ is strictly convex in t.

Proof. As t goes from 0 to 1, $h(t; x, y)$ goes from y to x. The first and second derivatives of $h(t; x, y)$ are:

$$\frac{\partial h}{\partial t} = \frac{x(x-y)y}{((1-t)x + ty)^2}, \frac{\partial^2 h}{\partial t^2} = \frac{2x(x-y)y}{((1-t)x + ty)^3} \tag{8.32}$$

The second derivative is always positive, so $h(t; x, y)$ is strictly convex. □

Lemma 8.4. *Suppose $x, y \in \mathbb{R}$, $x > y > 0$, and let $N \geqslant 2$ be an integer. For $n \in 0...N$, let $H(n; x, y)$ denote*

$$H(n; N, x, y) = \underbrace{x : ... : x}_{n \text{ times}} : \underbrace{y : ... : y}_{N-n \text{ times}} \tag{8.33}$$

$$= \left(\frac{n}{x} + \frac{N-n}{y} \right)^{-1} \tag{8.34}$$

Then $H(n+1; N, x, y) - H(n; N, x, y)$, for $n \in 0 \ldots N-1$, is increasing in n.

Proof. This follows from the fact that $H(n; N, x, y) = h(n/N; x, y)$ and strict convexity of $h(t; x, y)$. $\square$

Proposition 5.1 (maximization of dividend discount under trace constraint).

Proof. Let $v_n^i + \underline{\sigma}^2 = [V_i]_n$, the n-th diagonal entry of V_i. We wish to find the extreme points of the feasible set defined by the constraints

$$\sum_{n=1}^{N} (v_n^i + \underline{\sigma}^2) = N\underline{\sigma}^2 + T, v_n^i \geqslant 0 \qquad \text{for } i = 1, \ldots, I, n = 1, \ldots, N \tag{8.35}$$

For each i, this defines a N-dimensional simplex whose extreme points are where all v_n^i's are zero except for one, which is equal to T. Each of these extreme points corresponds to a diagonal matrix where all diagonal entries are $\underline{\sigma}^2$ except for one, which is equal to $\underline{\sigma}^2 + T$, i.e. a matrix specialized in variance for some asset.

The feasible set of the joint problem is an I-way Cartesian product of convex sets, therefore convex. Its extreme points are an I-way Cartesian product of the extreme points of the individual feasible sets. Suppose for asset $n = 1, \ldots, N$, there are k_n matrices specialized in that asset. Then

$$\mathbf{sum}(V_1 : \ldots : V_I) = \sum_{n=1}^{N} H(k_n; I, \underline{\sigma}^2 + T, \underline{\sigma}^2) \tag{8.36}$$

Suppose we reorder the assets such that the k_n's are in descending order, and suppose there exists an asset m such that $k_m < k_1 - 1$. If we transfer the specialization of one of the matrices from asset 1 to asset m, this sum changes by the amount

$$\begin{aligned} &H(k_m + 1; I, \underline{\sigma}^2 + T, \underline{\sigma}^2) - H(k_m; I, \underline{\sigma}^2 + T, \underline{\sigma}^2) \\ &- (H(k_1; I, \underline{\sigma}^2 + T, \underline{\sigma}^2) - H(k_1 - 1; I, \underline{\sigma}^2 + T, \underline{\sigma}^2)) \end{aligned} \tag{8.37}$$

which results in a decrease, by Lemma 8.4. Therefore, the sum is minimized when it is no longer possible to decrease the maximum k_n, and the proposition is proved. $\square$

References

Admati, A. (1985): "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Prices," *Econometrica*, 53(3), 629–658.

Allen, F., and D. Gale (1988): "Optimal Security Design," *Review of Financial Studies*, 1(3), 229–263.

Bernstein, D. S. (2009): *Matrix Mathematics - Theory, Facts, and Formulas, 2nd ed.* Princeton University Press.

Bhatia, R. (2007): *Positive definite matrices*. Princeton University Press.

Boot, A., and A. Thakor (1993): "Security Design," *Journal of Finance*, 48, 1349–1378.

Brennan, M., and H. Cao (1997): "International Portfolio Investment Flows," *Journal of Finance*, 52(5), 1851–1880.

Cao, H., T. Wang, and H. Zhang (2005): "Model Uncertainty, Limited Market Participation, and Asset Prices," *Review of Financial Studies*, 18(4), 1219–1251.

Carpio, R., and M. Guo (2016): "On Equilibrium Existence in a Finite-Agent, Multi-Asset Noisy Rational Expectations Economy," Discussion paper.

Chamberlain, G., and E. E. Leamer (1976): "Matrix Weighted Averages and Posterior Bounds," *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1), 73–84.

Chemmanur, T. J., and A. Yan (2004): "A theory of corporate spin-offs," *Journal of Financial Economics*, 72, 259–290.

Grossman, S., and J. Stiglitz (1980): "On the impossibility of informationally efficient markets," *American Economic Review*, 70, 393–408.

Habib, M. A., D. B. Johnsen, and N. Y. Naik (1995): "Spinoffs and Information," *Journal of Financial Intermediation*, 6, 153–176.

Hellwig, M. F. (1980): "On the Aggregation of Information in Competitive Markets," *Journal of Economic Theory*, 22, 477–498.

Hirshleifer, D., and S. H. Teoh (2003): "Limited attention, information disclosure, and financial reporting," *Journal of Accounting and Economics*, 36, 337–386.

Horn, R. A. (1990): "The Hadamard Product," in *Matrix Theory and Applications (Processdings of Symposia in Applied Mathematics 40)*, pp. 87–169.

La Roche, J. (2014): "Carl Icahn Shreds The Ebay Board, And Demands That Ebay And PayPal Split," http://www.businessinsider.com/carl-icahn-ebay-letter-2014-2, Accessed: 2015/10/04.

Liu, Q., and R. Qi (2008): "Stock trading and diversification discount," *Economics Letters*, 98, 35–40.

Martin, J., and A. Sayrak (2003): "Corporate diversification and shareholder value: a survey of recent literature," *Journal of Corporate Finance*, 9, 37–57.

Naik, N. Y. (1997): "On aggregation of information in competitive markets: The dynamic case," *Journal of Economic Dynamics and Control*, 21, 1199–1227.

Nanda, V., and M. Narayanan (1999): "Disentangling value: financing needs, firm scope, and divestitures," *Journal of Financial Intermediation*, 8, 174–204.

Rajan, R., H. Servaes, and L. Zingales (2000): "The cost of diversity: the diversification discount and inefficient investment," *Journal of Finance*, 55(1), 3580.

Scharfstein, D., and J. Stein (2000): "The dark side of internal capital markets: Divisional rent-seeking and inefficient investment," *Journal of Finance*, 55, 25372564.

Scheinkman, J., and W. Xiong (2003): "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111(6), 1183–1220.

Van Nieuwerburgh, S., and L. Veldkamp (2009): "Information Immobility and the Home Bias Puzzle," *Journal of Finance*, 64(3), 1187–1215.

Veld, C., and Y. Veld-Merkoulova (2009): "Value creation through spin-offs: A review of the empirical evidence," *International Journal of Management Reviews*, 11(4), 407420.