

Portfolio Insurance, Underdiversification, and Idiosyncratic Risks *

Hong Liu
Washington University in St. Louis

This version: October 8, 2009

*I thank Nick Barberis, Jonathan Berk, John Campbell, Jon Ingersoll, Ravi Jagannathan, Mark Loewenstein, Todd Milbourn, Stijn Van Nieuwerburgh, Valery Polkovnichenko, Paolo Sodini, Matthew Spiegel, Anjan Thakor, Harald Uhlig, Raman Uppal, Dimitri Vayanos, Laura Veldkamp, Guofu Zhou, and seminar participants at Ajou University, KAIST, Rutgers University, Singapore Management University, Washington University in St. Louis, Yale University, the 2007 WFA, the 2008 ASAP, and the 2009 AFA conferences for very helpful suggestions. I am indebted to Kerry Back and Phil Dybvig for numerous discussions that have significantly improved the paper. Email: liuh@wustl.edu.

Portfolio Insurance, Underdiversification, and Idiosyncratic Risks

Abstract

Contrary to the prediction of standard portfolio diversification theory, most investors place a large fraction of their stock investment in a small number of stocks. We show that underdiversification may be caused by risk control. The key assumption is that investors are portfolio insurers who require wealth above a nonnegative minimum level at which the marginal utility is finite (for investors whose marginal utility at zero is finite, this assumption can be replaced by the requirement of solvency). We show that investors underdiversify when wealth is low and that some investors may underdiversify no matter how wealthy they are. In addition, less wealthy and less risk-averse investors underdiversify more, and a less diversified portfolio (inclusive of the risk-free investment) is *less* risky than a more diversified one. Furthermore, investors select stocks solely by expected returns and covariances. Any higher moments (e.g., variance and skewness) are irrelevant for this selection. In an equilibrium setting, we show that a less diversified stock portfolio has a higher expected return, a higher volatility, and maybe also a higher skewness and a lower Sharpe ratio. Finally, underdiversification in equilibrium implies that no one holds the market portfolio in equilibrium and idiosyncratic risks should be priced.

Journal of Economic Literature Classification Numbers: D11, D91, G11, C61.

Keywords: Diversification, Underdiversification, Portfolio Selection, Portfolio Insurance.

In contrast to the theory of the celebrated Capital Market Pricing Model (CAPM), extensive empirical literature has shown that most investors underdiversify and idiosyncratic risks are priced. For example, among the households that hold individual stocks directly, the median number of stocks held is only about three.¹ In addition, less wealthy investors underdiversify more, and even wealthy investors may underdiversify. There is also a large literature on whether or not idiosyncratic risks are priced. For example, Bessembinder (1992) finds strong evidence that idiosyncratic risk was priced in the foreign currency and agricultural futures markets. Ang, Hodrick, Xing, and Zhang (2003) provide empirical evidence suggesting that individual stock returns are negatively related to idiosyncratic volatility. Obviously, the relevance of idiosyncratic risks for asset pricing can be a result of underdiversification. Some possible explanations such as trading costs, differential ambiguity aversion, psychological and behavioral factors, and “special” preferences have been proposed for underdiversification.² For example, Uppal and Wang (2003) show that with differential ambiguity aversion across stocks, investors may underdiversify relative to the market portfolio. However, it is unclear why investors do not invest almost exclusively in ETFs if trading cost is the main concern. In addition, the widely-documented wealth effect on underdiversification is absent in most of the models based on ambiguity aversion, psychological and behavioral factors, and special preferences.

In this article, we propose an alternative explanation: underdiversification and thus the importance of idiosyncratic risks for asset pricing can be caused by risk control. The main empirically testable predictions of our model are consistent with existing empirical evidence. Specifically, we consider an investor who can trade one

¹Blume and Friend (1975), Kelly (1995), Polkovnichenko (2005), Campbell (2006)

²Brennan (1975), Kraus and Litzenberger (1976), Huberman (2001), Merton (1987), Nieuwerburgh and Veldkamp (2005), Barberis and Huang (2005), Uppal and Wang (2003), and Mitton and Vorkink (2007).

risk-free asset and a finite number of risky stocks to maximize his expected utility from the terminal wealth in a one-period setting. Following the large literature on portfolio insurance,³ we assume that the investor is a portfolio insurer who requires the terminal wealth be above a nonnegative minimum level at which the marginal utility is finite.⁴ The nonnegative minimum wealth requirement implies that the investor must always stay solvent. If the marginal utility at zero wealth is finite, then this requirement can be replaced with the requirement of solvency. To ensure solvency, the investor can borrow or shortsell only a limited amount.⁵

As shown in the standard diversification theory, non-portfolio insurers who are not subject to the solvency constraint always fully diversify (defined as investing in all the available stocks), independent of their wealth levels and cross-sectional return distributions. In contrast, we show portfolio insurers always underdiversify when wealth is sufficiently low. Moreover, under certain conditions on cross-sectional return distributions, investors with certain preferences (e.g., CRRA or mean-variance preferences) always underdiversify, no matter how wealthy they are.

Furthermore, consistent with empirical findings, our model implies that less wealthy investors hold a smaller number of stocks. Surprisingly, investors select stocks solely by expected returns and covariances. Any higher moments, such as variance and

³For example, Leland (1980), Grossman and Vila (1989), Basak (1995), and Dybvig (1995).

⁴As pointed out by Basak (1995), this definition of portfolio insurance agrees with the description of portfolio insurance given in both the academic (Grossman and Vila (1989)) and professional (Luskin (1988, p. 77)) literature. For possible rational behind portfolio insurance, see Leland (1980) for example. In addition, portfolio insurance strategy can result from a utility function whose value below the insured level is *not* $-\infty$ (see Footnote 17). The finiteness of marginal utility at the minimum is important for our main results, but not new (e.g., Dybvig (1995) also assumes finite marginal utility at the insured consumption level). More discussion on this ensues.

⁵Although for expositional simplicity, we focus on a one-period static model in the main text, the results extend to a multi-period dynamic model. Accordingly, we can interpret a wealth change or a minimum wealth level change as either across investors or across time for the same investor. Therefore, the model can have both cross-sectional and time-series implications. For an extension to a continuous-time setting, one only needs to impose a limited borrowing and shortselling constraint (e.g., margin requirement) to produce the same qualitative results.

skewness (and thus Sharpe ratios), are irrelevant for this choice.⁶ In addition, given the same initial wealth, as the required minimum wealth increases, the investor holds a less diversified (defined as holding a smaller number of stocks) and less risky portfolio (defined as having a lower return volatility for the total portfolio that includes the riskfree asset). This implies that portfolio insurance does indeed help to reduce total risk.

If an portfolio insurer holds an illiquid asset (e.g., a house, a retirement portfolio) that is costly to liquidate or trade for daily consumption, then our model predicts that this investor underdiversifies in the directly held portfolio (i.e., outside the illiquid asset) when his liquid wealth is low. In addition, investors no longer always hold the same stocks with the highest expected returns, because the covariance with the illiquid asset also affects stock selection. Furthermore, if the illiquid asset has a high correlation with the stock market, then the directly held portfolio may consist of only a very small number of stocks even for investors with liquid wealth well above the required level.⁷

Next, we show that our main results hold in an equilibrium model that is solved in closed-form where only solvency is required. Since some investors underdiversify in equilibrium and thus no one holds the market portfolio, idiosyncratic risks are priced

⁶It is important to note that, while higher moments such as variance are irrelevant for stock selection, they do affect how much is invested in a stock *once the stock is selected*. The amount invested in a highly risky asset (e.g., an option) can be very small. Indeed, for any given asset (including an option), an investor in our model never holds more than what is predicted by the standard theory. Thus, small learning costs, such as those considered by Nieuwerburgh and Veldkamp (2005), would prevent him from holding it at all. This may reconcile our model's predictions with the fact that most investors do not hold options, despite the high expected returns, because options are highly risky and the associated learning costs may be nontrivial. In addition, low covariance with existing assets such as durable goods and retirement portfolios may be another reason for the low holdings of options.

⁷As an illustration of the economic significance of this model, suppose an investor holds a \$300,000 nontradable, equally weighted portfolio in the retirement account and he requires at least \$10,000 liquid wealth. Then we show that for reasonable parameter values, if the investor's liquid wealth is lower than \$61,500, then he holds at most two stocks directly. In fact, for this investor to hold 4 stocks, he needs to have at least \$299,000 liquid wealth.

in this economy. In addition, this equilibrium model implies that a less diversified *stock portfolio* (i.e., excluding the risk-free asset) has a higher expected return and a higher return volatility. This is consistent with empirical findings such as those in Calvet, Campbell, and Sodini (2007) (e.g., see their Table 5). Moreover, under some technical conditions, a less diversified stock portfolio has a higher skewness and a lower Sharpe ratio. Furthermore, less wealthy and less risk averse investors underdiversify more. All of these implications are consistent with empirical findings, such as those of Campbell, Lettau, Malkiel and Xu (2001), Goetzmann and Kumar (2005) and Mitton and Vorkink (2007). In contrast to the literature with skewness preferences (e.g., Kraus and Litzenberger (1976), Mitton and Vorkink (2007)), we show that a less diversified stock portfolio can have higher skewness even when skewness is irrelevant for stock selection.

To illustrate the essential intuition behind the above results (see Figure 1 for graphical explanations based on the mean-variance frontier), we consider the simplest case where interest rate is normalized to zero, stock payoffs are independent, unbounded above, and can get arbitrarily close to 0, and the investor has a mean-variance preference. Given discrete-time trading, the investor cannot borrow or short-sell in order to ensure solvency. We describe how the optimal portfolio composition changes as the initial wealth W_0 increases from the required minimum level $\underline{W} > 0$. Let $\mu_i > 0$ and $\sigma_i > 0$ be respectively the expected return and the return volatility of the i th stock and $\mu_1 > \mu_2 > \dots > \mu_n$. When $W_0 = \underline{W}$, obviously the investor can only invest in the risk-free asset. Suppose that wealth is slightly above the minimum level. Since the investor does not borrow or short-sell, the investor can invest only a small fraction of wealth, say $\eta > 0$, in stocks. Since the utility of investing η in Stock i is equal to $\mu_i\eta - \frac{1}{2}A\eta^2\sigma_i^2$, the marginal utility at $\eta = 0$ is μ_i .⁸ Therefore the

⁸ $A > 0$ denotes the risk aversion coefficient.

stock with the highest expected return (irrespective of variance) yields the greatest marginal utility when η is small enough. Thus the investor invests only in Stock 1 when wealth is low.⁹ This underdiversification result holds as long as all stocks do not yield the same marginal utility at zero investment. The result that the investor first chooses the stock with the highest expected return (irrespective of any higher moments) when wealth is low is the implication of the well-known local risk neutrality property of expected utility preferences.

As wealth increases, the investor invests more in the first stock, thereby increasing the portfolio risk. This reduces the marginal utility from investing more in this stock. When wealth increases to a critical level, at which the marginal utility of investing more in the first stock reduces to the marginal utility of investing a tiny bit in another stock, the investor adds a second stock.¹⁰ The same argument for adding the first stock implies that the second stock added must have the second highest expected return, as it provides the highest marginal utility among the rest of the stocks. This process continues with further increases in wealth. Given high enough wealth, the investor may invest in all stocks and thus fully diversify. However, in some cases (e.g., CRRA or mean-variance preferences), the marginal utility of investing more in some stocks may be always greater than that of investing even a small amount in a new stock, due to limited borrowing and shortselling. In these cases, the investor underdiversifies regardless of his wealth. If stocks are correlated, then both the expected return and the covariance affect the marginal utility of investing a small amount in a stock and thus also affect the choice of stocks.

If stock payoffs have finite support, then investors can have limited borrowing and

⁹If multiple stocks have the same highest expected returns, then the investor invests in all of them simultaneously.

¹⁰As we show later, in the presence of a well diversified illiquid (retirement) portfolio, the critical level for adding the second stock can be rather high.

shortselling. The main results remain valid in this case. This is because if wealth is low enough, the amount that can be invested in risky assets is still small as long as leverage is limited. The main arguments therefore still apply.¹¹

If an investor owns an illiquid asset (e.g., a retirement portfolio) that cannot be traded for current consumption, then the same argument as above implies that when the investor’s liquid wealth is just above the minimum required level, the investor only invests in the stock that yields the highest marginal utility determined by its expected return and covariance with the illiquid asset. Therefore the investor optimally underdiversifies in the directly held portfolio. If the illiquid asset is highly correlated with the stock market, then diversification benefit from adding more stocks is small, and thus the number of stocks in the directly held portfolio is small even for relatively high wealth level.

The main reason why a more diversified total portfolio (inclusive of the risk-free investment) is riskier than a less diversified one is simply that a more diversified portfolio implies that the investor invests more in risky assets, *cetera paribus*. Consider the extreme case, where an investor can only afford the minimum wealth. His portfolio risk is 0, since he can only invest in the risk-free asset. As wealth increases, the investor becomes less underdiversified and invests more in risky assets, thereby increasing portfolio risk. On the other hand, as the required minimum wealth increases, the investor decreases portfolio risk by investing less in risky assets.

¹¹Limited borrowing and shortselling can also be caused by margin requirement (see Cuoco and Liu (2000)). The margin requirement imposed by Regulation T takes the form $\rho_1^\top \max(w, 0) + \rho_2^\top \max(-w, 0) \leq 1$, where w is the portfolio weight vector, $\rho_1 \in \mathbb{R}_{++}^n$, and $\rho_2 \in \mathbb{R}_{++}^n$. In addition, ρ_1 and ρ_2 can be reinterpreted as risk-weights assigned by investors to a stock. The portfolio constraints become risk management constraints (similar to capital requirement constraints imposed by the Basel committee on financial institutions). With this generalized setup, we can show that, *cetera paribus*, investors only hold stocks with the highest expected return to risk-weight ratio. Therefore, to the extent that investors believe they know “local” stocks better and thus assign lower risk-weights or higher expected returns to them, our model can shed some light on the “local bias” or “home bias” phenomenon well-documented in the literature (e.g., French and Poterba (1991), Huberman (2001)).

In the equilibrium model, some investors only hold a subset of stocks while others fully diversify. Therefore the market portfolio overweighs the stocks that are held by everyone in the economy and underweighs the stocks that are held by only some (wealthy) investors. This implies that the market portfolio is inefficient and thus idiosyncratic risks should be priced in equilibrium.

A more risk-averse investor will hold a greater number of stocks in order to reduce risk. As wealth increases, the investor sequentially adds stocks with lower expected returns. Thus, a more diversified stock portfolio (exclusive of the risk-free asset) has a lower expected return. Due to the diversification benefit, the return on a more diversified stock portfolio has a lower volatility. Finally, our equilibrium model can produce the empirical skewness pattern found in Mitton and Vorkink (2007) (even though skewness is irrelevant for stock selection) because for some stock payoff distributions, a portfolio with a high expected return also has a high skewness.

The key assumption driving our main results is that the investor requires a non-negative wealth level at which the marginal utility is finite. One interpretation of this assumption is that the investor must achieve a minimum level of utility strictly above the subsistence utility in all states of nature. In other words, we assume that the investor can survive on the risk-free investment, even if he loses a little more than all of his risky investment.¹² Alternatively, similar to Dybvig (1995), one can assume that the investor can only tolerate a limited decline in his standard of living and interpret the minimum level as representing the lowest tolerable consumption level that is above the subsistence level. On the other hand, even this key assumption is

¹²This is supported by many empirical studies on household finance. For example, Bucks, Kennickell, and Moore (2006) report that the value of financial assets (including stocks, mutual funds, retirement accounts, checking and saving accounts, etc.) is only 35.7% of the total assets in 2004. The total stock holdings (directly or indirectly held) is 47.4% of the financial assets. This suggests that, overall, households only allocate a fraction of wealth to stock investment. Thus, they could survive even if they lose a little more than all their stock investment.

not always necessary for all cases. For example, consider an investor with CRRA or mean-variance preferences who is not a portfolio insurer. If the unconstrained portfolio requires high leverage, then the investor will underdiversify no matter how wealthy he is, due to limited borrowing and shortselling implied by the solvency requirement. We adopt this key assumption so that our main results apply to a broad class of preferences and asset return distributions.

While the main results are shown for risk-averse expected utility preferences, they also hold for many alternative preferences. First, as shown above, the result that investors underdiversify more when poor than when rich holds as long as all assets do not yield exactly the same marginal utility at zero investment. Therefore this result holds for many expected utility preferences with or without global risk aversion, as well as many non expected preferences. For expected utility preferences, for example, utility functions can be convex for a certain range of wealth, like the classic Friedman and Savage preferences (Friedman and Savage (1958)), as long as concavification exists. For non expected utility preferences, this result holds, for example, for many of the disappointment aversion preferences (Gul (2000)), the recursive preferences (Epstein and Zin (1989)), the loss aversion preferences (Kahneman and Tversky (1979)), and the Machina preferences (Machina (1982)), as long as one can rank assets by the marginal utility they can yield at zero investment. Intuitively, given the limited borrowing and shortselling constraint implied by the nonnegative minimum wealth requirement, an investor with low enough wealth will invest only in the asset(s) that can yield the highest marginal utility. As wealth increases, the marginal utility of investing more in the asset(s) may decrease and thus it may become optimal to add other assets. Second, the result that higher moments are irrelevant for stock choice critically depends on local risk neutrality. While this result still holds for quite general expected utility preferences as long as the utility function is differentiable, it does *not*

hold for some alternative preferences that do not exhibit local risk neutrality (e.g., due to an assignment of higher probabilities than actual ones to bad outcomes).

Portfolio insurance is related to the large literature on how habit-formation preferences affect portfolio selection (e.g., Constantinides (1990) and Dybvig (1995)). Similar to our model, an investor with a habit-formation preference invests in such a way to guarantee a minimum wealth level. The key difference from most of the literature is that, in our model, the marginal utility at the minimum level is finite and investors face limited borrowing and shortselling constraints implied by the solvency constraint.¹³ As discussed above, both of these features are important for our main results.

This article is clearly not the first to show that portfolio constraints can induce an investor to invest in only a subset of available stocks (e.g., Ross (1977), Dybvig (1984), and Cuoco and Liu (2006)). However, most previous work assumes either special preferences or special asset return distributions, and offers only partial equilibrium analyses. In addition, these works are done in different contexts and for different purposes. For example, Ross (1977) and Dybvig (1984) examine the shape of a mean-variance efficient frontier with short-sale constraints and there is no wealth effect on the optimal number of held stocks in their models. In considering the impact of the value-at-risk based capital requirements on the riskiness of a financial institution, Cuoco and Liu (2006) restrict their analysis to CRRA preferences, lognormally distributed stock prices, and a partial equilibrium setting.

The rest of the article is organized as follows. In Section I, we describe the model setup, state the main results, and explain the essential intuition. In Section II, we provide graphical examples to illustrate the main results. In Section III, we show

¹³The model of Dybvig (1995) also implies that marginal utility at the required living standard is finite. In contrast to our model, both Constantinides (1990) and Dybvig (1995) assume a financial market with a single risky asset.

that the main results hold in an equilibrium model solved in closed-form and derive some additional empirically testable predictions. Section IV concludes. We prove the main results in the Appendix.

I. The model

We consider a one-period discrete-time model, the simplest possible model to explain the main ideas. An investor with initial wealth W_0 can invest in one risk-free asset and $n \geq 1$ risky stocks and maximizes his expected utility from the end-of-period wealth \widetilde{W}_1 .¹⁴ The risk-free interest rate is normalized to 0. Let \tilde{P} denote the end-of-period gross return vector of the stocks. For expositional simplicity, we assume that the gross return \tilde{P}_i ($i = 1, 2, \dots, n$) is unbounded above, can get arbitrarily close to 0, and is *independent* of other returns. This assumption implies that, to ensure solvency at the end of the period, the investor must choose no borrowing or shortselling.¹⁵ Let $\tilde{z} = \tilde{P} - \bar{1}$ be the return vector, $\mu \equiv (\mu_1, \mu_2, \dots, \mu_n)^\top = E[\tilde{z}]$ be the expected return vector, and $\sigma\sigma^\top \equiv \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\} = E[(\tilde{z} - \mu)(\tilde{z} - \mu)^\top]$ be the diagonal variance-covariance matrix. In addition, we assume that¹⁶

$$\infty > \mu_1 > \mu_2 > \dots > \mu_n > 0.$$

¹⁴Although this model does not directly model labor income, the initial wealth may be viewed as the net present value of all current assets, future labor income, and future endowment. Alternatively, one can think of the minimum threshold wealth required from risk-free investment as the residual part that cannot be financed by labor income alone. The case with correlated returns will be discussed later (see footnote 27). The case where the investor owns an illiquid asset such as a retirement portfolio and a house is considered in Subsection A.

¹⁵No borrowing or shortselling is observationally close to what is found in individual trading behavior (e.g., Boehmer, Jones, and Zhang (2005)). As shown later, allowing limited borrowing and shortselling does not change our main results (see Page 15).

¹⁶The assumption of positive expected excess returns is only for expositional simplicity. Because of the short-sales constraints, the main results on underdiversification obviously hold if some expected excess returns are negative.

The investor derives utility from the end-of-period wealth \widetilde{W}_1 and the utility function $u(W)$ is twice continuously differentiable, strictly concave, and strictly increasing. Similar to the portfolio insurance literature (e.g., Leland (1980), Grossman and Vila (1989), Basak (1995)), we assume that¹⁷

Assumption 1 *The investor requires that terminal wealth be above an exogenously-given minimal floor $\underline{W} \geq 0$, at which the marginal utility is finite, i.e., $\widetilde{W}_1 \geq \underline{W}$ almost surely and the right derivative $u'(\underline{W}) < \infty$.*

One interpretation of this assumption is that the investor must achieve at least a minimum level of utility that is strictly above the subsistence utility in all states of nature. In other words, given the insurance of this minimum level, the investor can still survive even if the investor loses a little more than all of his risky investment. It is worth noting that if $u'(0) < \infty$, then \underline{W} can be set to 0, in which case this assumption can be replaced with the solvency requirement.

Let θ denote the column vector of the dollar amount invested in the stocks. For given $W_0 \geq \underline{W}$, the investor's problem is then:¹⁸

$$\max_{\theta} E \left[u(\widetilde{W}_1) \right] \quad (2)$$

¹⁷Alternatively, one can assume a utility function $\hat{u}(W)$ that is defined in \mathbb{R} , where

$$\hat{u}(W) = \begin{cases} u(W), & \text{if } W \geq \underline{W}, \\ u(W) - L, & \text{Otherwise,} \end{cases} \quad (1)$$

where $L \gg 0$ represents the large utility loss from not achieving the critical level \underline{W} . It can be shown that when L is large enough (not necessarily infinite), then an investor starting with a wealth $W_0 \geq \underline{W}$ always trades to ensure the end-of-period wealth \widetilde{W}_1 is never below \underline{W} . It can also be shown that with a preference like $\hat{u}(W)$, an investor whose wealth is slightly *below* \underline{W} would invest aggressively in a risky asset for the chance of getting above \underline{W} , even when the expected return is negative. This might help explain why poor people might also gamble (e.g., purchasing lotteries).

¹⁸In contrast to most of the literature on habit formation preferences, the required minimum wealth level is assumed to be above the subsistence level and thus the marginal utility at the minimum wealth level is finite. Obviously, one can still use the typical utility function form $u(W - \underline{W})$, as long as $u'(0) < \infty$ and the constraint $W \geq \underline{W}$ is imposed. It is worth noting that for modeling habit-formation investors with preferences such that $u'(0) < \infty$ (e.g., non-CRRA HARA utility functions), the constraint $W \geq \underline{W}$ is also necessary in addition to using $u(W - \underline{W})$.

subject to

$$\widetilde{W}_1 = W_0 + \theta^\top \widetilde{z} \geq \underline{W} \geq 0. \quad (3)$$

Given that \widetilde{z} is unbounded above and can be arbitrarily close to -1 , element by element, constraint (3) is equivalent to:

$$\theta^\top \bar{1} \leq W_0 - \underline{W} \quad \text{and} \quad \theta \geq 0, \quad (4)$$

where $\bar{1}$ is the column vector of 1's. Therefore, the investor's problem is equivalent to: For given $W_0 \geq \underline{W}$,

$$\max_{\theta} E \left[u(\widetilde{W}_1) \right], \quad \text{subject to (4)}.^{19} \quad (5)$$

Before we proceed to our main results, we provide two important definitions.

Definition 1 *An investor is said to underdiversify if he only invests in a proper subset of available stocks, i.e., only invests in $m < n$ stocks.*²⁰

Definition 2 *A portfolio is more diversified (or less underdiversified) than another if it holds a greater number of stocks.*

We collect our main analytical results in the following theorem.

Theorem 1. *We have:*

1. *For low enough initial wealth, the investor always underdiversifies. In addition, full diversification (i.e., investment in all available stocks) is optimal for all*

¹⁹There always exists a unique solution to the investor's problem, as the feasible set for θ is compact and the objective function is continuous and strictly concave.

²⁰This definition is commonly used in the literature (e.g., Goetzmann and Kumar (2005)). A broader definition may be that any portfolio different from the market portfolio is underdiversified. The definition we use is more conservative. Goetzmann and Kumar (2005) find that the correlations among commonly used underdiversification measures are high.

$W_0 > \underline{W}$ if and only if the expected returns are exactly the same across all stocks;

2. As $W_0 - \underline{W}$ increases, the investor invests a greater dollar amount in a greater number of stocks. In addition, both the expected value and the variance of the end-of-period wealth increase with the initial wealth W_0 and decrease with the floor level \underline{W} .
3. Which stocks the investor selects into the optimal portfolio depend only on their expected returns, but not on any higher moments (e.g. variance, skewness). Thus, Sharpe Ratios in particular are irrelevant for stock selection;
4. If the investor has a CRRA or mean-variance utility, then as the ratio W_0/\underline{W} of the initial wealth to the floor increases, both the expected return and the return volatility of the optimal portfolio (inclusive of the risk-free asset) increase;
5. If the investor has a CRRA or mean-variance utility, then under some conditions on the asset returns, it is optimal for the investor to always underdiversify, no matter how wealthy he is.

Proof: see Appendix.

Several observations are in order. First, Part 1 of Theorem 1 implies that it cannot be optimal for investors to fully diversify for *all* wealth levels. This is easy to see in the extreme case where $W_0 = \underline{W}$, since in this case the investor can only invest in the risk-free asset. In addition, even if the initial wealth W_0 is strictly above the floor level \underline{W} , full diversification is still rare, since this occurs only when the expected returns are exactly the same across all stocks.

Second, the proof of Theorem 1 reveals how the optimal portfolio compositions change as the initial wealth increases. As stated above, if $W_0 = \underline{W}$, then the investor

can only invest in the risk-free asset. When the initial wealth is slightly above the required floor, the investor adds the stock with the highest expected return (i.e., Stock 1) to his portfolio. He increases his investment in Stock 1 as the initial wealth increases until it reaches a critical level, say \hat{W}_2 , when he adds the stock with the second highest expected return (i.e., Stock 2). Then he increases his investment in both Stock 1 and Stock 2 as the initial wealth increases until it reaches the next critical level, say \hat{W}_3 , when he adds Stock 3. This process continues as initial wealth increases further. If the borrowing and short-sales constraints are not binding when the wealth is high enough, then the investor eventually invests in all of the n stocks and thus fully diversifies.²¹ The above process shows that, surprisingly, the order in which stocks are selected depends only on their expected returns, but not on any higher moments such as variance, skewness and hence not on Sharpe Ratio. To understand this result, suppose wealth increases slightly from the floor, such that the investor can only invest a small amount $\eta > 0$ in stocks. Since the final wealth \widetilde{W}_1 is affine in asset returns, the marginal utility at \underline{W} of investing a small amount in a stock is positively proportional to the stock's expected return and does not depend on any higher moments. In other words, when the dollar amount $\eta > 0$ that can be invested in additional stocks is small, the total risk of investing in a stock (e.g., in the order of $\eta^2 \sigma_i^2$) is very small no matter how large the stock volatility is. The investor therefore only cares about the expected return, a direct implication of the well-known local risk neutrality of expected utility and thus selects the stock with the next highest expected return irrespective of any higher moments. The same argument applies to the subsequent additions of other stocks. If stock returns were correlated, then covariances would also affect the stock selection. Still, higher moments do not

²¹Theorem 1 implies that an investor can add multiple stocks at the same time if these stocks have the same expected returns. See Figure 2 for an illustration.

matter. See footnote 27 for more details.

On the other hand, it is important to note that, while higher moments are irrelevant for stock selection, they may affect how much is invested in a stock *once the stock is selected*. For example, for a highly risky asset, the amount invested in the asset can be small, and thus a small learning cost (e.g., Merton (1987)) would prevent the investor from holding it at all. This may explain why most investors do not hold options, despite high expected returns, since options may also be highly risky with nontrivial learning costs.

Third, Parts 2 and 4 of Theorem 1 implies that, for a given initial wealth level, an increase in the required floor decreases the number of stocks that investor optimally holds and the riskiness of the optimal portfolio. Thus, portfolio insurance does indeed help to reduce portfolio risk. The reduction in risk stems from the reduction of the total dollar amount invested in the stocks. For the same reason, a less diversified portfolio (inclusive of the risk-free asset) is less risky than a more diversified one.

Finally, Part 5 of Theorem 1 suggests under some conditions on preferences and return distributions, no matter how wealthy an investor is, he never fully diversifies. To understand this result, consider the case of CRRA utility. In this case, for $1 \leq i \leq n$ the optimal fraction w_i of the initial wealth W_0 invested in stock i is independent of W_0 in the absence of portfolio insurance. If this portfolio requires leverage, i.e., $\sum_{i=1}^n w_i > 1$, then the investor can never hold this portfolio when he cannot borrow. In addition, if some of the expected returns are high enough, he will not hold stocks with the lowest expected returns since he cannot short sell these stocks. This result is consistent with the empirical evidence that even the rich may underdiversify.

In our baseline model, we assume that stock payoffs are unbounded above and can get arbitrarily close to 0, which implies no borrowing or shortselling to guarantee solvency in a discrete-time setting. We now discuss why our main results remain

valid even if we relax this assumption to allow the support of payoffs to be bounded above by $1 < \bar{P} \leq \infty$ and below by $1 > \underline{P} \geq 0$. In this generalized case, limited borrowing and shortselling is feasible. Specifically, suppose $W_0 = \underline{W} + \eta$, which allows the investor to invest $x \in [-\eta/(\bar{P} - 1), \eta/(1 - \underline{P})]$ in stocks. However, when η is small enough, x is also small. Thus, the local risk neutrality argument still applies and, as before, only expected returns and covariances affect stock selection. Minor modifications of the proof show that other main results still hold. The main difference is that, when limited shortselling is allowed, the investor shorts the stock with the lowest expected return in addition to the purchase of the stocks with the highest expected returns. Similar arguments imply that our main results extend to a continuous time setting as long as borrowing and shortselling are limited.

A. Investor with a nontradable asset: An extension

Many investors have illiquid assets such as retirement portfolios, houses, and other durable goods. These illiquid assets are typically too costly to trade for daily consumptions. In this subsection, we show that our main results still hold in the presence of illiquid assets and thus can help explain, for example, why investors with a diversified retirement portfolio underdiversify in the directly held portfolio.

To simplify analysis without losing any qualitative properties, we adopt the same setup as before but assume that an investor owns one unit of a nontradable asset, whose end-of-period payoff \tilde{N} is a nonnegative random variable that may be (highly) correlated with stocks. Since the nontradable asset cannot be traded for next period consumption, the investor requires the *tradable* wealth be above $\underline{W} \geq 0$. Let θ denote the column vector of the dollar amount invested in the stocks and W_0 be the initial

tradable wealth. For given $W_0 \geq \underline{W}$, the investor's problem is then:

$$\max_{\theta} E \left[u(\widetilde{W}_1) \right] \quad (6)$$

subject to

$$\widetilde{W}_1 = W_0 + \theta^\top \tilde{z} + \tilde{N} \quad (7)$$

and

$$W_0 + \theta^\top \tilde{z} \geq \underline{W}, \quad (8)$$

which as before is equivalent to

$$\theta^\top \bar{1} \leq W_0 - \underline{W}, \theta \geq 0. \quad (9)$$

If $W_0 = \underline{W}$, obviously the investor can only hold the riskfree asset in addition to the nontradable asset. Suppose $W_0 = \underline{W} + \eta$ with $\eta > 0$ and the investor invests η in Stock i . Then

$$\widetilde{W}_1 = \underline{W} + \eta \tilde{z}_i + \tilde{N}.$$

The marginal utility of investing in Stock i at $\eta = 0$ is thus

$$\lim_{\eta \downarrow 0} \frac{\partial E[u(\widetilde{W}_1)]}{\partial \eta} = E[u'(\underline{W} + \tilde{N})(\tilde{z}_i + 1)] = E[u'(\underline{W} + \tilde{N})](\mu_i + 1) + \text{Cov} \left(u'(\underline{W} + \tilde{N}), \tilde{z}_i \right). \quad (10)$$

Therefore the investor chooses to invest in only the stock with the combination of expected return and covariance with the nontradable asset that maximizes the marginal utility. In addition, higher moments such as variance and skewness do not affect the choice of the stock. Same arguments apply when the investor adds additional stocks as his tradable wealth increases. Thus as before, the investor underdiversifies when

tradable wealth is low and only expected return and covariance matter for stock selection. Moreover, investors with lower tradable wealth underdiversify more in the directly held portfolio. As shown below, if the nontradable asset is highly correlated with stocks, then the number of stocks directly held can be quite small. This may help explain why investors with a diversified retirement portfolio hold directly a relatively small number of stocks. Finally, in the absence of an illiquid asset, our basic model implies that all investors share the same (high expected return) risky assets. In this generalized model with an illiquid asset, investors may hold different stocks, because covariance with the illiquid asset also matters and investors' illiquid asset holding may be different.

II. Graphical examples

In this section, we provide graphical examples to illustrate the ideas behind our main results. For simplicity, we specialize to the mean-variance preference case, which provides the most transparent economic intuitions.²²

Suppose there are $n > 0$ risky assets and one risk-free asset that the investor can trade. The $(n + 1)^{st}$ risky asset is nontradable. The interest rate is normalized to 0. Let W_0 be the initial tradable wealth and W_N be the initial value of the nontradable asset. The $(n + 1) \times 1$ expected return vector is μ , the $(n + 1) \times (n + 1)$ variance-covariance matrix is σ . Let the $n \times 1$ vector w be the initial fraction of total wealth $(W_0 + W_N)$ invested in the tradable risky assets and $w_N \equiv W_N/(W_0 + W_N)$ be the initial fraction of total wealth invested in the nontradable asset. Then (9) is equivalent to

$$w\bar{1} \leq 1 - \frac{W + W_N}{W_0 + W_N}, \quad w \geq 0. \quad (11)$$

²²A detailed analysis with a closed-form solution is available from the author.

The mean-variance investor then solves the following problem:

$$\max_w \left[\left(\begin{array}{c} w \\ w_N \end{array} \right)^\top \mu - \frac{1}{2} A \left(\begin{array}{c} w \\ w_N \end{array} \right)^\top \sigma \sigma^\top \left(\begin{array}{c} w \\ w_N \end{array} \right) \right],$$

subject to (11), where $A > 0$ measures the investor's risk aversion.

Next we illustrate the implication of underdiversification on the validity of CAPM using the standard mean-variance framework (without any nontradable asset). We will consider a full-fledged equilibrium model in Section III. Consider two groups of investors: the rich with mass 1 and the poor with mass λ . The entire risky investment W_r of the rich is in the tangency portfolio on the mean-variance efficient frontier with the weight vector

$$w_r = \frac{(\sigma \sigma^\top)^{-1} \mu}{((\sigma \sigma^\top)^{-1} \mu)^\top \mathbf{1}}.$$

To simplify exposition, consider the simplest case where the entire risky investment W_p of the poor is put in the first stock only, i.e., the weight vector

$$w_p = (1, 0, 0, \dots, 0)^\top.$$

The weight vector for the market portfolio is thus

$$w_M = \frac{W_r}{W_r + \lambda W_p} w_r + \left(1 - \frac{W_r}{W_r + \lambda W_p} \right) w_p.$$

Therefore no one in this economy holds the market portfolio. Moreover, since the market portfolio is a convex combination of the tangency portfolio and Stock 1 which is inefficient, the market portfolio is itself inefficient. Intuitively, the market portfolio overweighs Stock 1 and underweighs other stocks relative to the tangency portfolio. Therefore CAPM does not hold and idiosyncratic risks should be priced in an economy

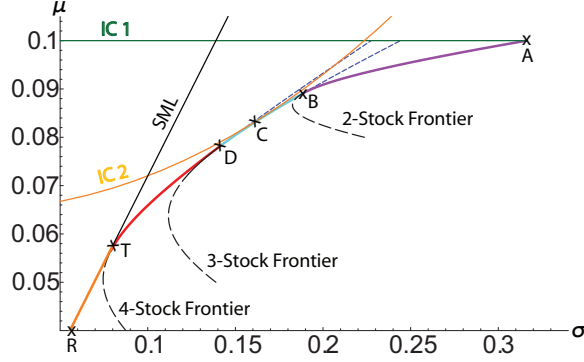


Figure 1: Mean-Variance Efficient Frontier.

with underdiversification.

Figure 1 plots the mean-variance frontier for four stocks with independent returns and illustrates the intuition behind our main results. Stocks 1–4 are sorted by expected returns, from the highest to the lowest. The security market line above the tangency point T is no longer relevant due to the borrowing constraint. The dotted segments are no longer achievable due to the short sales constraints. So the relevant frontier with the borrowing and short sales constraints is the curve $ABCDTR$. The investor's utility function is $u = \mu\eta - \frac{1}{2}\eta^2\sigma^2$ (without any nontradable asset) and thus the indifference curve at utility level u is

$$\mu = u/\eta + \frac{1}{2}A\eta\sigma^2. \quad (12)$$

IC 1 in Figure 1 is the indifference curve when the investor's wealth is just slightly above the floor level (i.e., η is small), which is almost horizontal by (12). To maximize utility, the investor chooses point A which represents investment only in Stock 1. As η increases, the indifference curve becomes more curved. When it becomes tangent to $ABCDTR$ at point A , the investor starts to add Stock 2. As η continues to increase,

the optimal portfolio moves from A to B, at which point the investor adds Stock 3. If η increases further, the investor invests more in each of the three stocks, as shown by point C implied by IC 2. If η is high enough, the tangency point between the indifference curve and ABCDTR moves to point D, where the investor adds Stock 4. As η further increases, the investor increases investment in every stock. After the tangency portfolio T is reached, the investor also increases the investment in the risk-free asset. This describes the case where the borrowing constraint is not binding for high enough wealth. If the tangency point between the indifference curve and the frontier ABCDTR never moves beyond point D for any wealth level due to the borrowing constraint, then even wealthy investors underdiversify.

Figure 2 plots the optimal portfolio weights as functions of the initial wealth-to-floor ratio W_0/\underline{W} . The figure shows that, as the ratio increases from 1, the investor first adds the stock with the highest expected return, i.e., Stock 1. As the ratio increases further, he increases the fraction of wealth invested in Stock 1. When the ratio reaches 1.07 (i.e., 7% above the floor), he adds Stocks 2 and 3 simultaneously, because they have the second highest expected returns. As the ratio increases beyond 1.07, the investor first increases the investment in Stocks 1, 2, and 3 and then adds Stock 4 when it reaches 1.35.

Figure 2 suggests that the order of the fraction of wealth invested in different stocks can be different for the rich and the poor. For example, a rich investor invests more in the second stock than in the first stock. However, the order is reversed for a poor investor, who may not invest in the second stock at all. This is because expected return determines the order in which a stock is selected, and the second stock has a lower expected return than the first.

Figure 3 plots the expected return and the return volatility of the optimal portfolio (inclusive of the risk free asset) as functions of the initial wealth-to-floor ratio W_0/\underline{W} .

It shows that, as the ratio decreases, both the expected return and volatility decrease. As shown in Figure 2, the number of stocks in the optimal portfolio decreases as the ratio decreases. In particular, this suggests that a less diversified portfolio is less risky than a more diversified one. The main reason for the decrease in the risk as wealth falls is that the constraints (4) become more binding, inducing the investor to invest less in stocks. This suggests that portfolio insurance can be effective in reducing portfolio risk.

Suppose there are 50 stocks and the investor holds an equally weighted, nontradable retirement portfolio (i.e., $w_N = 2\%$ in each stock), worth of $W_N = 30\underline{W}$. Figure 4 plots the optimal number of stocks held in the directly held portfolio against W_0/\underline{W} for two risk aversion levels. This figure shows that consistent with the empirical finding, the number of stocks an investor directly holds is quite small. For example, suppose $\underline{W} = \$10,000$ and thus the investor has 300,000 invested in the nontradable retirement portfolio. Figure 4 shows that if $A = 2$ and the investor's wealth outside the retirement account is smaller than \$61,500, then he holds at most two stocks in the directly held portfolio. In fact, for this investor to hold 4 stocks, he needs to have at least \$299,000 wealth outside the retirement account (not shown in the figure). This small number of stocks held in the directly held portfolio is a reflection of the small marginal benefit from additional diversification given an already diversified retirement portfolio.

In addition, Figure 4 suggests that, as the number of stocks increases, the marginal benefit of diversification decreases. Therefore, a greater wealth increase is needed to hold one additional stock in the portfolio. The model thus has another empirically testable implication: that a less diversified investor trades more often, since a small change of wealth would trigger adding or dropping an asset.

Moreover, the figure shows that as risk aversion increases, the number of stocks

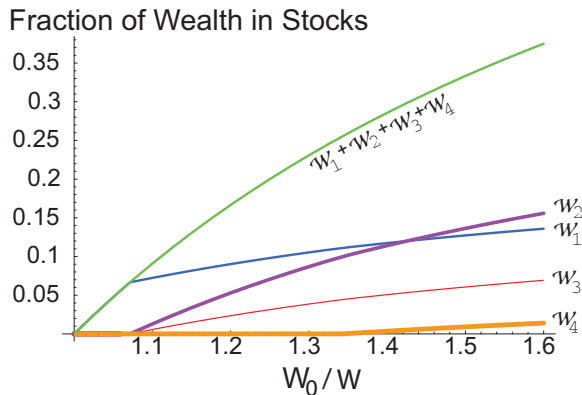


Figure 2: Portfolio weights against W_0/W for parameters $\mu_1 = 0.1$, $\sigma_1 = 0.3$, $\mu_2 = 0.07$, $\sigma_2 = 0.2$, $\mu_3 = 0.07$, $\sigma_3 = 0.3$, $\mu_4 = 0.05$, $\sigma_4 = 0.4$, and $A = 5$.

held in the optimal portfolio increases. So less risk averse investors underdiversify to a greater extent. This is due to the fact that the diversification benefit is more important for more risk averse investors.

III. An equilibrium model for underdiversification

In this section, we show that underdiversification can exist in equilibrium and derive some additional empirically-testable predictions. We again consider a one-period model where investors maximize the expected utility from the final wealth. For simplicity, we assume that investors have the same constant absolute risk aversion (CARA) preferences, i.e.,

$$u(W) = -e^{-AW},$$

where $A > 0$ is the CARA coefficient. There is one storable consumption good, the only risk-free asset in the economy. There are $n > 0$ risky stocks. We choose the consumption good as the numeraire and thus the interest rate is normalized to 0. For $1 \leq i \leq n$, there is a total supply of $\bar{w}_i > 0$ shares of stock i , whose payoff per share

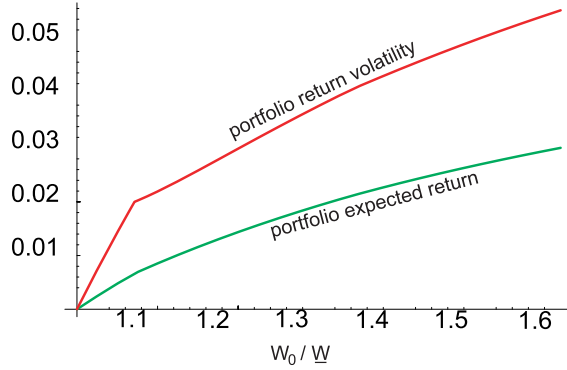


Figure 3: Portfolio expected return and variance against W_0/W for parameters $\mu_1 = 0.1$, $\sigma_1 = 0.3$, $\mu_2 = 0.07$, $\sigma_2 = 0.2$, $\mu_3 = 0.07$, $\sigma_3 = 0.3$, $\mu_4 = 0.05$, $\sigma_4 = 0.4$, and $A = 5$.

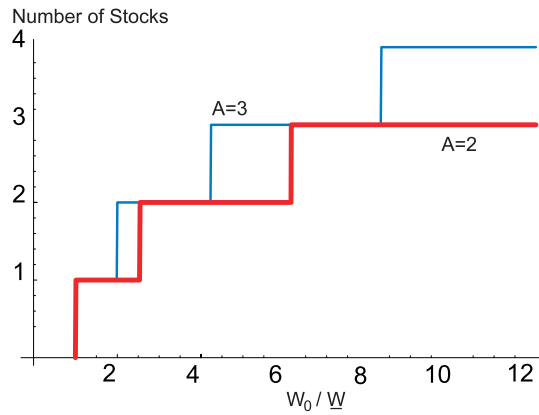


Figure 4: Number of stocks held against W_0/W given 50 independent stocks. For $i = 1, 2, \dots, 50$, $\mu_i = 0.2 - (0.2 - 0.01)(i - 1)/49$, $\sigma_i = 0.2 - (0.2 - 0.1)(i - 1)/49$, $W_N/W = 30$, and $p_N = 2\%$ in each stock

\tilde{P}_i is Gamma distributed with the probability density function

$$f_i(x) = \frac{x^{\alpha_i-1} e^{-x/\beta_i}}{\beta_i^{\alpha_i} \Gamma(\alpha_i)},$$

where $\alpha_i > 1$ and $\beta_i > 0$ such that its mean $\kappa_i = \alpha_i \beta_i < \infty$ and its variance $\varphi_i^2 = \alpha_i \beta_i^2 < \infty$.²³ The payoffs are independently distributed across all stocks. In addition, there are two groups of investors: the rich, with mass 1 and the poor, with mass $\lambda \geq 0$. Both types of investors are subject to the solvency constraint $\widetilde{W}_1 \geq 0$. The rich are endowed with $W_r \geq 0$ units of the consumption good and $\bar{\omega}_i$ shares of stock i , $1 \leq i \leq n$. The poor are endowed with $W_p \geq 0$ units of the consumption good, but no stocks. Since stock payoffs are unbounded above and can get arbitrarily close to 0, to ensure solvency, no one in the economy can borrow or short sell. Let p denote the initial equilibrium stock price vector. The rich solve

$$\max_{\omega} E \left[-e^{-A \widetilde{W}_1} \right], \quad (13)$$

s.t.,

$$\widetilde{W}_1 = W_r + (\bar{\omega}^\top - \omega^\top) p + \omega^\top \tilde{P} \geq 0, \quad (14)$$

where ω is the column vector of the number of shares held in stocks and \tilde{P} is the column vector of stock payoffs.

We first consider the case without the poor, i.e., $\lambda = 0$. In this case, since any equilibrium would be a no-trade equilibrium, the constraints are not binding. Then

²³CARA preferences and Gamma distributions are used only for tractability. Allowing different risk aversions is an straightforward extension. The main qualitative results in this section remain valid with other preferences and payoff distributions, although closed form solutions would become unlikely. Gamma distributions have similar properties to those of lognormal distributions, including the support set and moment characteristics.

(13) becomes

$$\max_{\omega} E \left[-e^{-A(W_r + \bar{\omega}^\top p + \omega^\top (\tilde{P} - p))} \right] = -e^{-A(W_r + \bar{\omega}^\top p)} \min_{\omega} \frac{e^{A\omega^\top p}}{\prod_{i=1}^n (1 + A\omega_i \beta_i)^{\alpha_i}}. \quad (15)$$

The first order conditions then imply that

$$\omega_i = \left(\frac{\kappa_i}{p_i} - 1 \right) \frac{\kappa_i}{A\varphi_i^2}, \quad (16)$$

which yields

$$p_i = \frac{\kappa_i^2}{\kappa_i + A\omega_i \varphi_i^2}. \quad (17)$$

The market clearing condition $\omega_i = \bar{\omega}_i$ then yields the equilibrium price

$$p_i = \frac{\kappa_i^2}{\kappa_i + A\bar{\omega}_i \varphi_i^2}, \quad (18)$$

which implies that the equilibrium expected return

$$\mu_i = \frac{\kappa_i}{p_i} - 1 = 2A\bar{\omega}_i \frac{\varphi_i}{s_i},$$

where $s_i = 2\kappa_i/\varphi_i$ denotes the skewness. In addition, the equilibrium return volatility

$$\sigma_i = \frac{\varphi_i}{p_i} = \frac{2}{s_i} (\mu_i + 1). \quad (19)$$

As expected, both the expected return and volatility are increasing in the payoff risk, the share supply, and risk aversion, but decreasing in payoff skewness. To facilitate

exposition, we assume the parameters are such that²⁴

$$\infty > \mu_1 > \mu_2 > \dots > \mu_n > 0. \quad (20)$$

We now consider the case with some poor investors in the economy, i.e., $\lambda > 0$. Suppose first that $W_p = \eta > 0$ is small. Since the marginal utility of investing in Stock i at $W_p = 0$ is

$$A(\mu_i + 1),$$

the poor will only buy Stock 1, which provides the greatest marginal utility. Since the poor will only buy Stock 1 and by (16), the rich's demand for a stock is independent of any other stocks, the equilibrium prices of other stocks will remain the same. Let \hat{p}_1 be the new equilibrium price of Stock 1. The market clearing condition for stock 1 becomes

$$\lambda \frac{\eta}{\hat{p}_1} + \left(\frac{\kappa_1}{\hat{p}_1} - 1 \right) \frac{\kappa_1}{A\varphi_1^2} = \bar{\omega}_1,$$

which implies that the new equilibrium price for Stock 1 is

$$\hat{p}_1 = \frac{\kappa_1^2 + \lambda A \varphi_1^2 \eta}{\kappa_1 + A \bar{\omega}_1 \varphi_1^2} = p_1 \left(1 + \lambda A \eta \frac{\varphi_1^2}{\kappa_1^2} \right). \quad (21)$$

By (20), there exists a small enough $\eta > 0$ such that (20) still holds with μ_1 replaced by $\hat{\mu}_1 \equiv \frac{\kappa_1}{\hat{p}_1} - 1$. This shows that when the poor is poor enough they buy only Stock 1 in equilibrium.

Now consider the scenario when a poor investor with a higher wealth is about to invest in the i th stock. Suppose it is at \hat{W}_i . Let δ_j be the dollar amount invested in Stock j , $1 \leq j \leq i$. Let \bar{p}_j be the new equilibrium price for stock j , $1 \leq j \leq i$. Since at \hat{W}_i , the poor investor has not bought stock i yet, the equilibrium price for stock i

²⁴Discussions about conditions under which expected returns become equal will be provided later.

remain unchanged from the case with $\lambda = 0$, i.e., $\bar{p}_i = p_i$. Let

$$V(\delta_1, \delta_2, \dots, \delta_i) = E \left[-\exp \left[-A \left(\sum_{j=1}^i \frac{\delta_j}{\bar{p}_j} \tilde{P}_j \right) \right] \right] \quad (22)$$

$$= -\prod_{j=1}^i \left(1 + \delta_j \frac{A\beta_j}{\bar{p}_j} \right)^{-\alpha_j} \quad (23)$$

be the value function. For investment in the i stocks to be optimal, we must have the marginal utility of investing in each is the same at $\delta_i = 0$. A little algebra implies that

$$\frac{\kappa_j}{\bar{p}_j + A\delta_j\beta_j} = \frac{\kappa_i}{p_i}, \quad j = 1, 2, \dots, i-1. \quad (24)$$

Given the investment of δ_j in stock j , a similar argument to that for (21) implies that the new equilibrium price of stock j is

$$\bar{p}_j = \frac{\kappa_j^2 + \lambda A \varphi_j^2 \delta_j}{\kappa_j + A \bar{\omega}_j \varphi_j^2}, \quad j = 1, 2, \dots, i-1. \quad (25)$$

Plugging (25) into (24) and simplifying, we have

$$\delta_j = \frac{\mu_j + 1}{\lambda + \mu_j + 1} \frac{\mu_j - \mu_i}{A\sigma_j^2} (> 0). \quad (26)$$

Equation (26) implies that the dollar amount invested in a stock is increasing in its risk premium, but decreasing in risk aversion and volatility. As the mass of the poor investor increases, the expected return is driven down and thus investment in the stock decreases. Equation (26) can be inserted into (25) to obtain the new equilibrium price:

$$\bar{p}_j = p_j \frac{\lambda/(\mu_i + 1) + 1}{\lambda/(\mu_j + 1) + 1}. \quad (27)$$

Equation (27) implies that $\bar{\mu}_1 > \bar{\mu}_2 > \dots \bar{\mu}_{i-1} > \mu_i > \dots > \mu_n$. So the poor will invest

in the first $i - 1$ stocks at \hat{W}_i . The threshold wealth level associated with adding stock i is thus

$$\hat{W}_i = \sum_{j=1}^{i-1} \delta_j.$$

Note that only the expected return μ_i , not σ_i for example, affects δ_j and the threshold wealth level \hat{W}_i . Since as i increases, μ_i decreases and thus δ_j increases, the threshold wealth level \hat{W}_i also increases as i increases. Since the above derivation applies to any $i = 2, 3, \dots, n$, we have shown that for $1 \leq i \leq n$, the poor holds the stocks with the highest $i - 1$ expected returns if and only if the initial wealth $W_0 \in (\hat{W}_{i-1}, \hat{W}_i]$.

Since the threshold wealth levels for holding additional stocks increase as the risk aversion coefficient decreases, our model predicts that less risk averse investors underdiversify more. To the extent that younger and male investors are less risk-averse, our model may help explain Mitton and Vorkink (2007)'s finding that younger and male investors tend to underdiversify more.

Equation (27) also implies that the expected return of stock j decreases from μ_j to μ_i as λ increases from 0 to ∞ , since the new equilibrium price increases from p_j to $p_j \frac{\mu_j + 1}{\mu_i + 1}$. This result shows that, for a finite mass of poor investors, the expected returns will always remain the same order as in the case with $\lambda = 0$.

The above equilibrium analysis shows that underdiversification is optimal when the initial wealth of the poor is low. In addition, as the initial wealth of the poor increases, they sequentially add the stocks with the highest expected returns. Higher moments do not impact this choice.

We next show that the expected value and the variance of the end-of-period wealth increase as the initial wealth of the poor increases.

First, for $W_0 = \eta < \delta_1$ (as defined by (26) with $i = 2$), (21) implies that the

expected end-of-period wealth is

$$\eta \frac{\kappa_1}{\hat{p}_1} = \eta \frac{\kappa_1(\kappa_1 + A\bar{\omega}_1\varphi_1^2)}{\kappa_1^2 + \lambda A\varphi_1^2\eta},$$

and the variance of the end-of-period wealth is

$$\left(\frac{\eta}{\hat{p}_1}\right)^2 \varphi_1^2 = \eta^2 \frac{\varphi_1^2(\kappa_1 + A\bar{\omega}_1\varphi_1^2)^2}{(\kappa_1^2 + \lambda A\varphi_1^2\eta)^2},$$

both of which increase as η increases. Now suppose $W_0 \in (\hat{W}_i, \hat{W}_{i+1}]$ and so the poor optimally hold i stocks for some $1 \leq i \leq n$. The marginal utilities of investing in each of these i stocks must be equal, which by (22) implies that

$$\frac{\kappa_j}{\check{p}_j + A\check{\delta}_j\beta_j} = k, \quad j = 1, 2, \dots, i, \quad (28)$$

for some $\mu_i + 1 > k > \mu_{i+1} + 1$, where \check{p}_j is the equilibrium price of stock j and $\check{\delta}_j$ is the dollar amount invested in stock j in equilibrium. Similar derivation to that for (26) shows that

$$\check{\delta}_j = \frac{\kappa_j^2(\mu_j + 1 - k)}{A k \varphi_j^2(\lambda + \mu_j + 1)} \quad (29)$$

and

$$\check{p}_j = p_j \frac{\lambda + k}{k(\lambda/(\mu_j + 1) + 1)}. \quad (30)$$

Since $W_0 = \sum_{j=1}^i \check{\delta}_j$, equation (29) implies that as W_0 increases, k decreases. The expected end-of-period wealth is

$$\sum_{j=1}^i \frac{\check{\delta}_j}{\check{p}_j} \kappa_j,$$

and the variance of the end-of-period wealth is

$$\sum_{j=1}^i \left(\frac{\check{\delta}_j}{\check{p}_j} \right)^2 \varphi_j^2.$$

The quantity $\frac{\check{\delta}_j}{\check{p}_j}$ increases as k decreases and as the initial wealth increases k decreases.

We therefore conclude that both the expected wealth and the wealth variance increase as the initial wealth increases.

This equilibrium model yields the following additional results.

Theorem 2. *In this equilibrium model,*

1. *Both the expected return and the return volatility of a less diversified stock portfolio are higher than those of a more diversified one;*
2. *If for $1 \leq i \leq n$,*

$$\frac{\sum_{j=1}^i C_{3j}(\mu_j + 1)}{\sum_{j=1}^i C_{3j}} \geq \frac{\sum_{j=1}^i a_j(\mu_j + 1)}{\sum_{j=1}^i a_j} \quad (31)$$

and

$$\frac{\sum_{j=1}^i C_{3j}(\mu_j + 1)^3}{\sum_{j=1}^i C_{3j}(\mu_j + 1)^2} + \frac{\sum_{j=1}^i C_{3j}(\mu_j + 1)}{\sum_{j=1}^i C_{3j}} \geq 2 \frac{\sum_{j=1}^i C_{3j}(\mu_j + 1)^2}{\sum_{j=1}^i C_{3j}(\mu_j + 1)}, \quad (32)$$

where $C_{3j} = \kappa_j^4 / \varphi_j^7$ and $a_j = \kappa_j^2 / ((\lambda + (\mu_j + 1))\varphi_j^2)$, then the return skewness of a less diversified stock portfolio is higher than that of a more diversified one;

3. *If for $1 \leq i \leq n$,*

$$\frac{\sum_{j=1}^i C_{1j}(\mu_j + 1)^2}{\sum_{j=1}^i C_{1j}(\mu_j + 1)} > \frac{\sum_{j=1}^i (C_{1j} - a_j)(\mu_j + 1)}{\sum_{j=1}^i (C_{1j} - a_j)} \quad (33)$$

and λ is small enough, the Sharpe ratio of a less diversified stock portfolio is lower than that of a more diversified one.

Our model therefore predicts that a less diversified stock portfolio has a greater

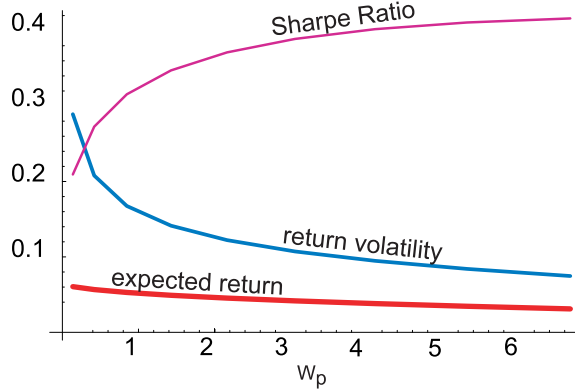


Figure 5: stock portfolio expected return, volatility, and Sharpe ratio against W_p for the following parameters: for $i = 1, 2, \dots, 10$, $\kappa_i = 1.2 - (1.2 - 1)(i - 1)/10$, $\varphi_i = 0.25 - (0.25 - 0.1)(i - 1)/10$, $\bar{\omega}_i = 1$, $W_r = 2$, $\underline{W} = 0$, $\lambda = 0.1$, and $A = 0.75$.

expected return and a higher volatility. In addition, under certain conditions such as (31), (32), and (33),²⁵ It also has a higher skewness and a lower Sharpe ratio. These predictions are supported by the findings of Mitton and Vorkink (2007).

Figure 5 confirms that a less diversified portfolio has a higher expected return and also a higher volatility. In addition, it shows that the Sharpe ratio of a less diversified portfolio is lower than that of a more diversified one. Figure 6 shows that the skewness of the portfolio return is also higher for a less diversified portfolio. All these patterns are also consistent with the findings in Mitton and Vorkink (2007). Figure 6 plots the volatility-scaled skewness (skewness/volatility) against the initial wealth level. This figure suggests that the scaled skewness is nonmonotonic in the number of the stocks held. Specifically, as the poor investor diversifies, it first decreases and then increases. The decreasing pattern is one of the main conclusions of Mitton and Vorkink (2007).

²⁵To see that these conditions can be satisfied, note that the terms in these inequalities are just different weighted averages of the individual expected gross returns $(\mu_j + 1)$ in the absence of the poor. To satisfy these conditions, one can change these weights by varying parameters such as κ_j , φ_j , and $\bar{\omega}_j$. In particular, these conditions are satisfied in Figures 5–6.

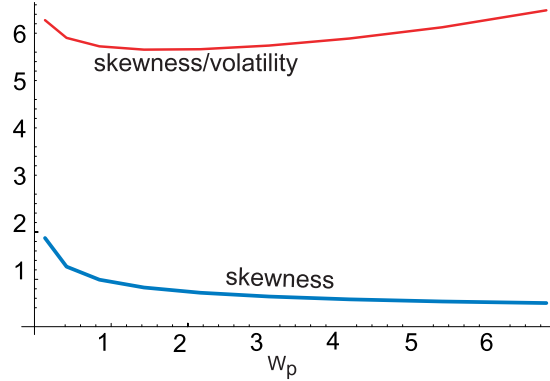


Figure 6: stock portfolio return skewness and volatility scaled skewness against wealth W_p for parameters: for $i = 1, 2, \dots, 10$, $\kappa_i = 1.2 - (1.2 - 1)/10$, $\varphi_i = 0.25 - (0.25 - 0.1)/10$, $\bar{\omega}_i = 1$, $W_r = 2$, $\underline{W} = 0$, $\lambda = 0.1$, and $A = 0.75$.

Our model suggests that the investors in the Mitton and Vorkink (2007) may have relatively low wealth. Although skewness does not impact stock selection in our model, the implied skewness pattern as shown in Figure 6 may appear to indicate that skewness is important for this choice. In this example, the skewness pattern is mainly driven by the fact that, for the Gamma distribution, skewness is proportional to mean for given variance.

IV. Concluding Remarks

We show that portfolio insurance behavior can help explain the well-documented underdiversification puzzle and why idiosyncratic risks may be priced. In particular, we demonstrate that investors who require a nonnegative minimum wealth always underdiversify when wealth is low and less wealthy investors underdiversify more. In addition, for investors with a well-diversified illiquid (retirement) portfolio, it is optimal to hold only a small number of stocks directly. These results hold for general asset return distributions and quite general preferences including many expected

utility and non expected utility preferences. In addition, if the marginal utility is finite at zero wealth (e.g., non-CRRA HARA preferences), then the portfolio insurance assumption can be replaced by the requirement of solvency. The generality of these results seems to suggest that underdiversification and thus the relevance of idiosyncratic risks for asset pricing should be the norm, not an exception. In addition, we show that if preferences exhibit local risk neutrality (e.g., differentiable expected utility preferences), then investors choose stocks solely by expected returns and covariances and any higher moments (e.g., variance and skewness) and thus Sharpe ratios are irrelevant for this choice.

Appendix

In this appendix, we provide the proofs for our main results.

PROOF OF THEOREM 1: Define the adjusted initial wealth $W_0^A \equiv W_0 - \underline{W}$ as the amount that is investable in risky assets. When $W_0^A = 0$, obviously, the investor can only invest in the risk-free asset. Now suppose his wealth increases to $W_0^A = \eta$, where $\eta > 0$ is small. Since he cannot borrow or short sell, the most he can invest in stocks is η . If he invests the entire amount η in stock j , then

$$\widetilde{W}_1 = \underline{W} + \eta(\tilde{z}_j + 1),$$

and the investor's marginal utility from investing only in Stock j at $\eta = 0$ is

$$\lim_{\eta \downarrow 0} \frac{\partial E[u(\widetilde{W}_1)]}{\partial \eta} = \lim_{\eta \downarrow 0} E[h(\eta, \tilde{z}_j)] = E[h(0, \tilde{z}_j)] = E[u'(\underline{W})(\tilde{z}_j + 1)] = u'(\underline{W})(\mu_j + 1), \quad (34)$$

where

$$h(\eta, \tilde{z}_j) \equiv u'(\underline{W} + \eta(\tilde{z}_j + 1))(\tilde{z}_j + 1),$$

the first equality in (34) follows from Leibnitz's rule, and the second equality follows from the Monotone Convergence Theorem (e.g., Williams 1991, p. 59) because $h(\eta, \tilde{z}_j)$ is strictly decreasing in η since $\partial h(\eta, \tilde{z}_j)/\partial \eta = u''(\underline{W} + \eta(\tilde{z}_j + 1))(\tilde{z}_j + 1)^2 < 0$ by the strict concavity of $u(\cdot)$. Therefore when η is small enough, since $\infty > u'(\underline{W}) > 0$, the marginal utility from investing in the stock with the highest expected return is strictly the highest. Thus, if wealth is slightly above the floor level \underline{W} , the investor invests \underline{W} in the risk-free asset and the rest in the stock with the highest expected return, i.e., Stock 1. Higher moments, such as variance, skewness, and kurtosis, do not impact this choice.

On the other hand, if multiple stocks have exactly the same expected returns, then the marginal utilities from investing in these stocks at $\eta = 0$ are the same across these stocks. It is therefore optimal to add them simultaneously when the initial wealth is slightly above the floor level \underline{W} .

Now suppose that, given the initial adjusted wealth W_0^A , the investor optimally invests \underline{W} in the risk-free asset, $\theta_i > 0$ in Stock i for $1 \leq i \leq k < n$, and 0 in the rest of the stocks. Since the investor is investing in only $k < n$ of n stocks, the constraint $\underline{W} + \sum_{i=1}^k \theta_i \leq W_0$ must be still binding. This is because the marginal utility of investing a bit in the rest of the stocks is always strictly greater than that of investing in the risk-free asset, since $\forall i = 1, 2, \dots, n, \mu_i > 0$ (the normalized interest rate). The end-of-period wealth is $\widetilde{W}_1 = \underline{W} + \sum_{i=1}^k \theta_i(\tilde{z}_i + 1)$. The Lagrangian is²⁶

$$V(W_0^A) = \max_{\theta} E \left[u \left(\underline{W} + \sum_{i=1}^k \theta_i(\tilde{z}_i + 1) \right) \right] + \nu \left(W_0^A - \sum_{i=1}^k \theta_i \right),$$

where ν is the Lagrangian multiplier. From the first order conditions, we have that the marginal utility of investing in Stock i for $1 \leq i \leq k$ is

$$\frac{\partial E \left[u \left(\widetilde{W}_1 \right) \right]}{\partial \theta_i} = E[u'(\widetilde{W}_1)(\tilde{z}_i + 1)] = \nu. \quad (35)$$

By the Envelop theorem, we have

$$\nu = \frac{\partial V(W_0^A)}{\partial W_0^A} > 0$$

since the constraint is still binding. Therefore, at the initial level W_0^A , the marginal utilities of investing in the first k stocks are equal, strictly positive, and strictly greater than the marginal utilities of investing in the risk-free asset and the other

²⁶The short sale constraint is not binding for these k stocks.

stocks (because it is optimal for the investor not to hold any of the other stocks).

Since

$$\frac{\partial E \left[u' \left(\underline{W} + \sum_{i=1}^k \theta_i (\tilde{z}_i + 1) \right) (\tilde{z}_i + 1) \right]}{\partial \theta_i} = E[u''(\widetilde{W}_1)(\tilde{z}_i + 1)^2] < 0,$$

the marginal utility of investing in a stock decreases as the investment in the stock increases. Thus, as the adjusted initial wealth increases from W_0^A , the investor will increase the investment in all these k stocks, i.e.,

$$\frac{\partial \theta_i}{\partial W_0^A} > 0, \quad i = 1, 2, \dots, k. \quad (36)$$

The marginal utility of investing in a new stock j at W_0^A is

$$E[u'(\widetilde{W}_1)(\tilde{z}_j + 1)] = E[u'(\widetilde{W}_1)](\mu_j + 1). \quad (37)$$

Therefore, if it becomes optimal to add another stock to the portfolio, the investor will add the stock with the next highest expected return, i.e., Stock $k + 1$, since $\infty > u'(\widetilde{W}_1) > 0$. Again, higher moments are irrelevant for this choice.²⁷

We next derive the critical adjusted wealth level \hat{W}_{k+1}^A at which it is optimal to add Stock $k + 1$. By (35) and (37) with $j = k + 1$, we must have that at this critical level, the marginal utilities of investing in these $k + 1$ stocks are exactly the same, i.e., for $i = 1, 2, \dots, k$, we have

$$E \left[u' \left(\underline{W} + \sum_{i=1}^k \theta_i (\tilde{z}_i + 1) \right) (\tilde{z}_i + 1) \right] = E \left[u' \left(\underline{W} + \sum_{i=1}^k \theta_i (\tilde{z}_i + 1) \right) \right] (\mu_{k+1} + 1), \quad (38)$$

²⁷If returns are correlated, then since the marginal utility $E[u'(\widetilde{W}_1)(\tilde{z}_j + 1)] = E[u'(\widetilde{W}_1)](\mu_j + 1) + \text{Cov} \left(u'(\widetilde{W}_1), \tilde{z}_j \right)$, only the expected return and the covariance matter. Therefore, the choice of stocks is still independent of higher moments (e.g., variance, skewness).

which can be simplified to

$$E \left[u' \left(\underline{W} + \sum_{i=1}^k \theta_i (\tilde{z}_i + 1) \right) (\tilde{z}_i - \mu_{k+1}) \right] = 0, i = 1, 2, \dots, k. \quad (39)$$

Therefore, we have $\hat{W}_{k+1}^A = \sum_{i=1}^k \theta_i$ where θ_i 's are the solution to (39). By (36), we have that as long as the constraint is binding, the critical wealth level \hat{W}_i^A at which it is optimal to add Stock i strictly increases with i . Combining these results shows that the number and the amount of stocks optimally held increases as the adjusted wealth W_0^A increases.

Since the expected end-of-period wealth is equal to

$$W_0 + \sum_{i=1}^k \theta_i \mu_i$$

and the end-of-period wealth variance is equal to

$$\sum_{i=1}^k \theta_i^2 \sigma_i^2,$$

both increase as the initial wealth W_0 increases for a given floor \underline{W} and decrease as the floor \underline{W} increases for a given initial wealth W_0 , by inequality (36).

If the investor has a CRRA utility, i.e.,

$$u(W) = \frac{W^{1-\gamma}}{1-\gamma},$$

then equation (39) is equivalent to

$$E \left[u' \left(1 + \sum_{i=1}^k w_i \tilde{z}_i \right) (\tilde{z}_i - \mu_{k+1}) \right] = 0, i = 1, 2, \dots, k. \quad (40)$$

where $w_i \equiv \theta_i/W_0$ represent the fraction of the initial wealth W_0 invested in stock i . And the constraints (4) become

$$\sum_{i=1}^k w_i \leq 1 - \frac{W}{W_0} \quad \text{and} \quad w_i \geq 0, i = 1, 2, \dots, n.$$

A similar argument for (36) shows that

$$\frac{\partial w_i}{\partial(W_0/W)} \geq 0, \quad i = 1, 2, \dots, k. \quad (41)$$

The expected return and the return variance of the optimal portfolio are, respectively,

$$\sum_{i=1}^k w_i \mu_i$$

and

$$\sum_{i=1}^k w_i^2 \sigma_i^2.$$

Both increase as W_0/W increases by inequality (41). A similar proof applies to mean-variance preferences. Since k is any integer number between 1 and n , this completes the proof for Parts 1-4.

Now we show Part 5. Due to the no borrowing constraint, if the w_i 's that solve (40) when $k < n$ are such that $\sum_{i=1}^k w_i > 1$, then the investor never holds more than k stocks, no matter how wealthy he is. To show this, suppose we have

$$E[u'(1 + w_1 \tilde{z}_1)(\tilde{z}_1 - \mu_2)] > 0 \quad (42)$$

for all $w_1 \in [0, 1]$, which can hold if μ_2 is low enough and μ_1 is high enough. Then it is never optimal for the investor to hold more than one stock. This is due to the fact that the marginal utility of investing in Stock 1 is strictly greater than that of

investing in Stock 2 (the stock with the next highest expected return) for all feasible w_1 . A similar proof applies for mean-variance preferences. \square

PROOF OF THEOREM 2: Suppose $W_0 \in (\hat{W}_i, \hat{W}_{i+1}]$ and thus the poor holds the i stocks with the highest expected returns. Define $\hat{\mu}_j = \mu_j + 1$ to be the gross return of j th stock for $j = 1, 2, \dots, n$. The m th root of the m th moment of the poor's stock portfolio return is

$$\xi_m = \left(\sum_{j=1}^i w_j^m \left(\frac{M_{mj}}{\check{p}_j} \right)^m \right)^{\frac{1}{m}}, \quad (43)$$

where the portfolio weight

$$w_j = \frac{\check{\delta}_j}{\sum_{j=1}^i \check{\delta}_j},$$

and M_{mj} is the m th moment of the payoff of the j th stock. Using (29) and (30), we have

$$\xi_m = \frac{k \left(\sum_{j=1}^i C_{mj} (\hat{\mu}_j - k)^m \right)^{1/m}}{(\lambda + k) \sum_{j=1}^i a_j (\hat{\mu}_j - k)}, \quad (44)$$

where $C_{mj} = M_{mj} \kappa_j^m / \varphi_j^{2m}$. As wealth decreases, poor investors become less diversified and k increases. We show that expected return, volatility, and skewness all increase with k . It is straightforward to show that ξ_m is strictly increasing in k if

$$f_m(k) = \sum_{j=1}^i a_j \hat{\mu}_j \sum_{j=1}^i C_{mj} (\hat{\mu}_j - k)^{m-1} - \sum_{j=1}^i a_j \sum_{j=1}^i C_{mj} \hat{\mu}_j (\hat{\mu}_j - k)^{m-1} \leq 0.$$

First, consider the expected return. Since $M_{1j} = \kappa_j$, we have $C_{1j} = \kappa_j^2 / \varphi_j^2 = a_j (\lambda + \hat{\mu}_j)$. Since

$$\frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j} = \frac{\sum_{j=1}^i C_{1j}}{\sum_{j=1}^i C_{1j} / (\lambda + \hat{\mu}_j)} - \lambda \leq \frac{\sum_{j=1}^i C_{1j} (\lambda + \hat{\mu}_j)}{\sum_{j=1}^i C_{1j}} - \lambda = \frac{\sum_{j=1}^i C_{1j} \hat{\mu}_j}{\sum_{j=1}^i C_{1j}},$$

where the inequality follows from Cauchy's inequality, we have

$$f_1(k) = \sum_{j=1}^i a_j \hat{\mu}_j \sum_{j=1}^i C_{1j} - \sum_{j=1}^i a_j \sum_{j=1}^i C_{1j} \hat{\mu}_j \leq 0.$$

The expected return therefore increases with k .

Next, consider the volatility. In this case, we have $C_{2j} = \kappa_j^2 / \varphi_j^4 \times M_{2j} = \kappa_j^2 / \varphi_j^2 = C_{1j}$. Since $k < \hat{\mu}_i$, which is the smallest one among the $\hat{\mu}_j$'s for $1 \leq j \leq i$, and

$$\frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j}{\sum_{j=1}^i C_{2j}}$$

is a weighted average of the $\hat{\mu}_j$'s, we must have

$$\frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j}{\sum_{j=1}^i C_{2j}} > k. \quad (45)$$

In addition, by Cauchy's inequality, we have

$$\frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j^2}{\sum_{j=1}^i C_{2j} \hat{\mu}_j} \geq \frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j}{\sum_{j=1}^i C_{2j}} \geq \frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j}. \quad (46)$$

Therefore

$$\begin{aligned} & \frac{f_2(k)}{\sum_{j=1}^i a_j \sum_{j=1}^i C_{2j}} \\ = & \frac{\sum_{j=1}^i a_j \hat{\mu}_j \sum_{j=1}^i C_{2j} (\hat{\mu}_j - k) - \sum_{j=1}^i a_j \sum_{j=1}^i C_{2j} \hat{\mu}_j (\hat{\mu}_j - k)}{\sum_{j=1}^i a_j \sum_{j=1}^i C_{2j}} \\ = & \left(\frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j} - \frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j^2}{\sum_{j=1}^i C_{2j} \hat{\mu}_j} \right) \frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j}{\sum_{j=1}^i C_{2j}} + \left(\frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j}{\sum_{j=1}^i C_{2j}} - \frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j} \right) k \\ < & \left(\frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j} - \frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j^2}{\sum_{j=1}^i C_{2j} \hat{\mu}_j} \right) k + \left(\frac{\sum_{j=1}^i C_{2j} \hat{\mu}_j}{\sum_{j=1}^i C_{2j}} - \frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j} \right) k \\ \leq & 0, \end{aligned}$$

where the inequalities follow from (45) and (46). Therefore, as poor investors become less diversified, the volatility of the optimal stock portfolio increases.

For the skewness, let

$$\begin{aligned}
h(k) &= \sum_{j=1}^i C_{3j} \sum_{j=1}^i C_{3j} (\hat{\mu}_j - k)^2 \hat{\mu}_j - \sum_{j=1}^i C_{3j} \hat{\mu}_j \sum_{j=1}^i C_{3j} (\hat{\mu}_j - k)^2 \\
&= -2 \left(\sum_{j=1}^i C_{3j} \sum_{j=1}^i C_{3j} \hat{\mu}_j^2 - \left(\sum_{j=1}^i C_{3j} \hat{\mu}_j \right)^2 \right) k \\
&\quad + \sum_{j=1}^i C_{3j} \sum_{j=1}^i C_{3j} \hat{\mu}_j^3 - \sum_{j=1}^i C_{3j} \hat{\mu}_j \sum_{j=1}^i C_{3j} \hat{\mu}_j^2.
\end{aligned}$$

It is easy to verify that

$$h'(k) = -2 \left(\sum_{j=1}^i C_{3j} \sum_{j=1}^i C_{3j} \hat{\mu}_j^2 - \left(\sum_{j=1}^i C_{3j} \hat{\mu}_j \right)^2 \right) \leq 0,$$

by Cauchy's inequality. In addition,

$$h(k^*) = 0,$$

where

$$\begin{aligned}
k^* &= \frac{\sum_{j=1}^i C_{3j} \sum_{j=1}^i C_{3j} \hat{\mu}_j^3 - \sum_{j=1}^i C_{3j} \hat{\mu}_j \sum_{j=1}^i C_{3j} \hat{\mu}_j^2}{2 \left(\sum_{j=1}^i C_{3j} \sum_{j=1}^i C_{3j} \hat{\mu}_j^2 - \left(\sum_{j=1}^i C_{3j} \hat{\mu}_j \right)^2 \right)} \\
&= \frac{\frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j^3}{\sum_{j=1}^i C_{3j} \hat{\mu}_j^2} - \frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j}{\sum_{j=1}^i C_{3j}}}{2 \left(\frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j^2}{\sum_{j=1}^i C_{3j} \hat{\mu}_j} - \frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j}{\sum_{j=1}^i C_{3j}} \right)} \frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j^2}{\sum_{j=1}^i C_{3j} \hat{\mu}_j} \\
&\geq \frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j^2}{\sum_{j=1}^i C_{3j} \hat{\mu}_j} \\
&> \mu_i.
\end{aligned} \tag{47}$$

The first inequality in (47) follows from (32) and the second inequality follows because (47) is a weighted average of the $\hat{\mu}_j$'s and $\hat{\mu}_i$ is smaller than any of the $\hat{\mu}_j$'s. Therefore, for all $k \leq \hat{\mu}_i$, we have

$$\frac{\sum_{j=1}^i C_{3j}(\hat{\mu}_j - k)^2 \hat{\mu}_j}{\sum_{j=1}^i C_{3j}(\hat{\mu}_j - k)^2} \geq \frac{\sum_{j=1}^i C_{3j} \hat{\mu}_j}{\sum_{j=1}^i C_{3j}}.$$

Finally, by (31), we have

$$\frac{\sum_{j=1}^i C_{3j}(\hat{\mu}_j - k)^2 \hat{\mu}_j}{\sum_{j=1}^i C_{3j}(\hat{\mu}_j - k)^2} \geq \frac{\sum_{j=1}^i a_j \hat{\mu}_j}{\sum_{j=1}^i a_j},$$

which implies that $f_3(k) \leq 0$. Thus, a less diversified stock portfolio has a higher skewness.

The Sharpe ratio of the stock portfolio is

$$\frac{\sum_{j=1}^i (C_{1j} - \frac{\lambda+k}{k} a_j)(\hat{\mu}_j - k)}{\sqrt{\sum_{j=1}^i C_{1j}(\hat{\mu}_j - k)^2}} = \frac{\sum_{j=1}^i (C_{1j} - a_j)(\hat{\mu}_j - k) - \sum_{j=1}^i \frac{\lambda}{k} a_j(\hat{\mu}_j - k)}{\sqrt{\sum_{j=1}^i C_{1j}(\hat{\mu}_j - k)^2}}.$$

So if λ is small enough and

$$g(k) \equiv \frac{\sum_{j=1}^i (C_{1j} - a_j)(\hat{\mu}_j - k)}{\sqrt{\sum_{j=1}^i C_{1j}(\hat{\mu}_j - k)^2}}$$

is strictly decreasing in k , then the Sharpe ratio will also be strictly decreasing in k and thus the claim holds. It can be easily verified that condition (33) directly implies that $g'(k) < 0$. □

References

- Anderson, J. M. 1999. The wealth of U.S. families: Analysis of recent census data, Working Paper, Bureau of Census.
- Benartzi, S. 2001. Excessive extrapolation and the allocation of 401(K) accounts to company stock, *Journal of Finance* 56, 1747–1764.
- Benartzi, S., and R. H. Thaler. 2001. Naive diversification strategies in retirement saving plans, *American Economic Review* 91, 79–98.
- Barberis, N. C., and M. Huang. 2007. Stocks as lotteries: The implications of probability weighting for security prices, Working Paper, Yale University.
- Berger, P. G., and E. Ofek. 1995. Diversification’s effect on firm value, *Journal of Financial Economics* 37, 39–65.
- Blume, M., and I. Friend. 1975. The asset structure of individual portfolios and some implications for utility functions, *Journal of Finance* 30, 585–603.
- Boehmer, E., C. M. Jones, and X. Zhang. 2005. Which shorts are informed? Working Paper, Columbia University.
- Brennan, M. J. 1975. The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: Theory and some empirical results, *Journal of Financial and Quantitative Analysis* 10, 483–496.
- Bucks, B. K., A. B. Kennickell, and K. B. Moore. 2006. Recent changes in U.S. family finances: Evidence from the 2001 and 2004 Survey of Consumer Finances, Federal Reserve Bulletin, 1–38.
- Calvet, L. E., Campbell, J. Y., P. Sodini, 2007, Down or out: Assessing the welfare costs of household investment mistakes, *Journal of Political Economy* 115, 707–747.
- Campbell, J. Y. 2006. Household finance, *Journal of Finance* 61, 239–244.
- Campbell, J. Y., L. Martin, B. G. Malkiel, and Y. Xu. 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of*

Finance 56, 1–43.

Constantinides, G. M. 1990. Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.

Cuoco, D., and H. Liu. 2000. A martingale characterization of consumption choices and hedging costs with margin requirements, *Mathematical Finance* 10, 355–385.

Cuoco, D., and H. Liu. 2006. An analysis of VaR-based capital requirements, *Journal of Financial Intermediation* 15, 362–394.

Dybvig, P. H. 1984. Short sales restrictions and kinks on the mean variance frontier, *Journal of Finance* 36, 239–244.

Dybvig, P. H. 1995. Duesenberry’s ratcheting of consumption: Optimal dynamic allocation and investment given intolerance for any decline in standard of living, *Review of Economic Studies* 62, 287–313.

Epstein, L. G., and S. E. Zin. 1989. Substitution, risk aversion and the temporal behavior of consumption and asset returns. *Econometrica* 57, 937-969.

French, K. and J. M. Poterba. 1991. Investor diversification and international equity markets, *American Economic Review* 81, 222–226.

Friedman, M., and L. J. Savage. 1948. The utility analysis of choices involving risk, *Journal of Political Economy* 56, 279–304.

Goetzmann, W. N., and A. Kumar. 2005. Why do individual investors hold under-diversified portfolios? Working Paper, Yale University.

Gul, F. 1991. A theory of disappointment aversion, *Econometrica* 59, 667-686.

Huberman, G. 2001. Familiarity breeds investment, *Review of Financial Studies* 14, 659–680.

Kraus, A., and R. Litzenberger. 1976. Skewness preference and the valuation of risk assets, *Journal of Finance* 31, 1085–1100.

- Leland, H. 1980. Who should buy portfolio insurance? *Journal of Finance* 35, 581–594.
- Machina, M. J. 1982. “Expected utility” analysis without the independence axiom, *Econometrica* 50, 277–323.
- Merton, R. C. 1987. A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.
- Mitton, T., and K. Vorkink. 2007. Equilibrium underdiversification and the preference for skewness, *Review of Financial Studies* 20, 1255–1288.
- Nieuwerburgh, S. V., and L. Veldkamp. 2005. Information acquisition and portfolio under-diversification, Working Paper, New York University.
- Polkovnichenko, V. 2005. Household portfolio diversification: A case for rank dependent preferences, *Review of Financial Studies* 18, 1467–1502.
- Ross, S. A. 1977. The capital asset pricing model (CAPM), short-sale restrictions and related issues, *Journal of Finance* 32, 177–183.
- Uppal, R., and T. Wang. 2003. Model misspecification and underdiversification, *Journal of Finance* 58, 2465–2486.
- Vissing-Jørgensen, A. 2002. Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structure, Working Paper, University of Chicago.
- Williams, D. 1991. Probability with Martingales, Cambridge, Cambridge University Press.