Solvency Constraint, Underdiversification, and Idiosyncratic Risks *

Hong Liu

Washington University in St. Louis and CAFR

This version: November, 2012

Forthcoming

Journal of Financial and Quantitative Analysis

*I thank an anonymous referee, Nick Barberis, Hendrik Bessembinder (the Editor), Jonathan Berk, John Campbell, Jon Ingersoll, Ravi Jagannathan, Li Liao, Mark Loewenstein, Todd Milbourn, Stijn Van Nieuwerburgh, Valery Polkovnichenko, Paolo Sodini, Matthew Spiegel, Anjan Thakor, Harald Uhlig, Raman Uppal, Dimitri Vayanos, Laura Veldkamp, Yajun Wang, Weiqiang Zhang, Guofu Zhou, Ning Zhu, and seminar participants at Ajou University, KAIST, Rutgers University, Singapore Management University, UC San Diego, 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, Phil Dybvig, and Pietro Veronesi for their constructive comments that have significantly improved the paper. Send correspondence to Hong Liu at Olin Business School, Washington University, Campus Box 1133, One Brookings Drive, St. Louis, MO 63130. Email: liuh@wustl.edu. Telephone: 314-935-5883. Fax: 314-935-6359.

Solvency Constraint, Underdiversification, and Idiosyncratic Risks

Abstract

Contrary to the prediction of the standard portfolio diversification theory, many investors place a large fraction of their stock investment in a small number of stocks. I show that underdiversification may be caused by solvency requirement. My model predicts that underdiversification decreases in the remaining wealth after committed consumption and variance and higher moments do not affect stock selection for quite general preferences and return distributions. In addition, a less diversified stock portfolio has a higher expected return, a higher volatility, and a higher skewness and idiosyncratic risks are priced. Many predictions of my model are consistent with empirical findings.

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

Keywords: Diversification, Underdiversification, Solvency Constraint, Idiosyncratic Risks.

I. Introduction

In contrast to the theory of the celebrated Capital Market Pricing Model (CAPM), extensive empirical literature has shown that many investors underdiversify and idiosyncratic risks are priced. For example, among the households that hold individual stocks directly, the median number of directly held stocks was two until 2001, when it increased to three (e.g., Campbell (2006)). The main empirical findings on underdiversification include (1) the number of stocks directly held by less wealthy investors is small and increases as investors' wealth increases; (2) even wealthy investors may hold a small number of stocks in the directly held portfolio; (3) many households simultaneously invest in well-diversified funds and in extremely underdiversified stock portfolios; (4) less diversified investors tend to hold stocks with higher expected returns; and (5) underdiversified portfolios can have higher expected returns, higher volatility, lower Sharpe ratios, and higher skewness.[1] 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 (2006) provide empirical evidence suggesting that idiosyncratic volatility affects stock return. Clearly, the relevance of idiosyncratic risks for asset pricing can be a result of underdiversification.

Some possible explanations for underdiversification such as trading costs, differential ambiguity aversion, psychological and behavioral factors, and "special" preferences, have

[1]See, for example, Polkovnichenko (2005), Ivkovich and Weisbenner (2005), Campbell (2006), Calvet, Campbell, and Sodini (2007), Kumar (2007), Mitton and Vorkink (2007), Goetzmann and Kumar (2008), Ivkovich, Sialm, and Weisbenner (2008). Even though less diversified investors tend to hold stocks with higher expected returns, some underdiversified individual investors may underperform net of transaction costs because of excessive trading due to factors outside this paper's main focus.

been proposed.[2] However, it is unlikely that these existing models can explain many of the above main findings. For example, if trading cost is the main concern, it is unclear why investors do not diversify through index funds or S&P 500 SPDR ETF. In most of the models based on ambiguity aversion, psychological and behavioral factors, and "special" preferences, the widely documented, prominent wealth effect on underdiversification is largely absent. Also, none of these models can explain why underdiversified portfolios tend to have higher expected returns and lower Sharpe ratios.

In an almost parallel literature on household consumption, it is widely documented that a large portion of the average household's budget is committed to ensure a certain critical level of consumption (e.g., Fratantoni (2001), Chetty and Szeidl (2007)). Committed consumption can be caused by sources such as housing and other durable goods consumption that is costly to adjust, habit formation, meeting fixed financial obligations (e.g., mortgage and tuition payments), and precautionary savings against unemployment or health shocks. Empirical findings also suggest that the committed consumption level is generally above the subsistence level and thus the marginal utility at the committed level is finite (e.g., Chetty and Szeidl (2007), Shore and Sinai (2010)). In support of the economic significance of consumption commitment, the existing literature also show that models with consumption commitment can outperform many alternative models. For example, they fit consumption data better than neoclassical models (e.g., Flavin and Nakagawa (2008)), can help explain the low stock ownership puzzle (e.g., Fratantoni (2001)), can explain why consumers insure risks and bunch uninsured risks together (e.g., Postlewaite, Silverman, and Samuelson (2008)), can help explain the discrepancy between moderate-stake and large-stake risk aversion and lottery playing by insurance buyers (e.g., Chetty and Szeidl (2007)), and can endogenize widely

[2]See, for example, Brennan (1975), Kraus and Litzenberger (1976), Huberman (2001), Merton (1987), Nieuwerburgh and Veldkamp (2005), Barberis and Huang (2007), Uppal and Wang (2003), Mitton and Vorkink (2007), and Boyle, Garlappi, Uppal and Wang (2009).

used reference-dependent preferences (e.g., Chetty and Szeidl (2010)).

In this article, I propose a new and simple explanation of underdiversification: It can be caused by solvency requirement in the presence of committed consumption. This paper's main results hold for quite general preferences and stock return distributions. The main assumption is that the investor commits to a certain level of terminal consumption at which the marginal utility is finite and must remain solvent after the committed consumption.

Different from the existing literature, my model can help explain *all* of the five main empirical findings listed above with unique predictions (4) and (5). In addition, the model can also help explain that (6) young or male investors underdiversify more (e.g., Mitton and Vorkink (2007)) and (7) idiosyncratic risks are priced (e.g., Bessembinder (1992)). Also different from the existing models, the model implies that for expected utility preferences, only expected return and covariance with already selected stocks affect stock selection. Other moments such as variance and skewness (and thus Sharpe ratio) are irrelevant for this choice.[3] In addition, in contrast to models based on psychological and behavioral factors and "special" preferences that use distorted probabilities, my model is fully rational without assuming any distortion of probabilities.

[3]It is important to note that, while other 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 small. Indeed, for any given asset (including an option), an investor in my 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 my 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.

To explain the essential intuitions for the main results in the simplest case, suppose an investor has a mean-variance preference and assume all stocks have different expected returns. Because of the solvency constraint and discrete-time trading, only limited borrowing (or shortselling) is feasible. This implies that when the investor's wealth is low, he can invest only a small amount in stocks, which implies that expected return has a first-order effect on utility, while risk has only a second-order effect, by local risk neutrality. Therefore when his wealth is low, the marginal benefit of diversification (i.e., reducing risk) is smaller than the marginal cost of diversification (i.e., lowering expected return), and thus he only invests in the stock with the highest expected return and does not diversify. As wealth increases, he invests more in the stock, his portfolio risk increases, and thus the marginal benefit of diversification increases. At a critical wealth level, the marginal benefit of diversification surpasses the marginal cost of diversification and thus the investor adds the second stock to his portfolio. Among all the return moments of a stock, only expected return affects the marginal cost of diversification and only the covariance with the first stock affects the marginal benefit of diversification. Therefore the selection of a stock as the second stock to be added to the portfolio only depends on the stock's expected return and its covariance with the first stock, but not on other moments such as variance and skewness. This process continues with further increases in wealth. Given high enough wealth, the investor may invest in all stocks and thus fully diversify. However, for some preferences (e.g., CRRA or mean-variance), because of limited borrowing and shortselling caused by the solvency constraint, the marginal benefit of full diversification may be always lower than the marginal cost of full diversification. In these cases, even wealthy investors underdiversify.

In equilibrium, while wealthy investors fully diversify and hold all the stocks, less wealthy investors only hold the stocks with the highest expected returns. Therefore, no one holds the market portfolio in equilibrium and idiosyncratic risks are priced. As less wealthy investors' wealth increases, they sequentially add stocks with next lower expected returns. Thus, a more diversified stock portfolio has a lower expected return. Due to the diversification effect, the return on a more diversified stock portfolio has lower volatility and may have a higher Sharpe ratio. In addition, because adding lower expected return stocks shifts the portfolio return distribution to the left, a more diversified portfolio also has lower skewness for some return distributions. Finally, more risk-averse investors diversify more because they are less risk tolerant. Combined with the empirical finding that younger investors and male investors are less risk averse, my model then predicts that younger or male investors underdiversify more.[4]

This paper is related to the large literature on how habit-formation preferences affect portfolio selection (e.g., Constantinides (1990)). The key difference from most of the literature is that, in my model, the marginal utility at the committed consumption level is finite and investors face limited borrowing and shortselling constraints implied by the solvency requirement.[5] Without the limited borrowing and shortselling constraints, a less wealthy investor would always borrow or shortsell to hold a fully diversified portfolio like a wealthy investor. If the marginal utility at the committed consumption level were infinite, then the marginal benefit of diversification would be high no matter how low an investor's wealth is, and therefore he would always fully diversify.

This paper is also related to the literature on portfolio selection with portfolio constraints (e.g., Ross (1977), Dybvig (1984), and Cuoco and Liu (2006)). Most of previous works are done in different contexts and for different purposes. In addition, they assume either special

[4]To the extent that borrowing constraints may be also more binding for young investors because of the lack of collateral or good credit history, my model also predicts younger investors underdiversify more.

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

preferences or special asset return distributions and offer only partial equilibrium analyses. For example, Ross (1977) and Dybvig (1984) examine the shape of a mean-variance efficient frontier with short-sale constraints. While short-sale constraints can prevent investors from shorting, they do not prohibit investors from borrowing to buy all the stocks with positive expected returns. Cuoco and Liu (2006) consider the impact of the Basel II capital requirements on the riskiness of a financial institution. Basel II capital requirements do not apply to individual investors and are much more stringent than the solvency constraint I assume. In addition, in contrast to my model, they restrict their analysis to CRRA preferences, lognormal stock prices, and portfolio allocation problems, without examining any equilibrium impact of the capital requirements.

The rest of the article is organized as follows. In Section II, I use some simple examples to illustrate the essential intuitions for the main results. In Section III, I describe a portfolio choice model setup to show that the main results hold in a quite general setting. In Section VI, I explicitly solve an equilibrium model to show that the main results can indeed hold in equilibrium. Section V concludes. We prove the main results in the online Appendix.

II. A Simple Example

In this section, I provide a simple example to explain the intuitions behind the main results that an investor underdiversifies when wealth is relatively low and invests in more stocks as his wealth increases. In a one-period setting, consider the simplest case where there are two independent stocks and a risk-free asset with interest rate r normalized to zero. Stock gross returns are unbounded above and can get arbitrarily close to 0. An investor has a mean-variance preference with a risk-aversion coefficient of A and a committed consumption $\underline{C} > 0$. Given discrete-time trading and the full support of stock gross returns, the investor cannot borrow or shortsell; otherwise, the committed consumption (and solvency) cannot be guaranteed. For $i = 1, 2$, let μ_i and σ_i be respectively the expected return and the return volatility of Stock i with $\mu_1 > \mu_2 > r = 0$. We now explain how the optimal portfolio composition changes as the initial wealth W_0 increases from $\underline{C}$. When $W_0 = \underline{C}$, the investor can only invest in the risk-free asset to guarantee the committed consumption. Suppose now wealth is slightly above the minimum level, i.e., $W_0 = \underline{C} + \eta$ for some small $\eta > 0$. Since the investor cannot borrow or shortsell and must invest at least $\underline{C}$ in the risk-free asset, the investor can invest at most the (small) fraction $w \equiv \eta/W_0$ of his wealth in stocks. The investor's problem is then

$$\max_{\{w_1,w_2\}} w_1\mu_1 + w_2\mu_2 - \frac{1}{2}Aw_1^2\sigma_1^2 - \frac{1}{2}Aw_2^2\sigma_2^2,$$

subject to the no-borrowing and no-shortselling constraint

$$w_1 + w_2 \leq w, w_1 \geq 0, w_2 \geq 0, \tag{1}$$

where w_i denotes the fraction of wealth W_0 (not η) invested in Stock i for $i = 1, 2$. The no-borrowing constraint (the first inequality in expression (1)) is binding for small enough w because stocks have higher expected returns than the risk-free asset. Thus the investor's problem becomes

$$\max_{0 \leq w_1 \leq w} V(w_1) \equiv w_1\mu_1 + (w - w_1)\mu_2 - \frac{1}{2}Aw_1^2\sigma_1^2 - \frac{1}{2}A(w - w_1)^2\sigma_2^2,$$

which implies that

$$V'(w_1) = (\mu_1 - \mu_2) - A\left(w_1\sigma_1^2 - (w - w_1)\sigma_2^2\right). \tag{2}$$

When w approaches zero, so does w_1. Therefore, if w is small enough, $V'(w_1)$ is always strictly positive because $\mu_1 > \mu_2$, which implies that the optimal $w_1 = w$, i.e., I have a corner solution. Thus, the investor invests all his disposable wealth η in the stock with the highest expected return (Stock 1) and does not diversify.

We can also show this result from comparing the marginal utilities from the two stocks. The marginal utility of investing $w \geq 0$ in Stock i is equal to $\mu_i - Aw\sigma_i^2$, which approaches μ_i as w approaches 0. Therefore when w is small, the stock with the highest expected return provides the greatest marginal utility, which makes the investor only invest in this stock. $V'(w_1)$ in equation (2) is exactly the difference in the marginal utilities from the two stocks (with $w_2 = w - w_1$), which approaches $\mu_1 - \mu_2$ as the investable amount w approaches 0.

Intuitively, a risk-neutral investor only invests in the stock with the highest expected return. A risk-averse investor may hold stocks with lower returns to reduce risk. The right-hand side of equation (2) is also the difference between the marginal cost and the marginal benefit of diversification. More specifically, the marginal cost of diversification is the reduction in the expected return $(\mu_1 - \mu_2)$, while the marginal benefit is the reduction in the risk (the second term in equation (2)). As w approaches 0, so does w_1, which implies that the marginal benefit of diversification goes to 0 too. Therefore, the marginal benefit of diversification is smaller than the marginal cost of diversification when wealth is low and less wealthy investors do not diversify.

As the investor's wealth further increases, he invests more in Stock 1, his portfolio risk increases, and thus the marginal benefit of diversification increases. When this marginal benefit surpasses the marginal cost of diversification, he adds the second stock. Suppose the critical wealth level beyond which the investor adds Stock 2 is $W_0 = \hat{W}$ and let $\hat{w} \equiv 1 - \underline{C}/\hat{W}$ be the fraction of wealth invested in Stock 1. Then I must have $V'(\hat{w}) = 0$, i.e.,

$$\mu_1 - \mu_2 = A\hat{w}\sigma_1^2, \tag{3}$$

where the right-hand side follows from setting $w_1 = w = \hat{w}$ in the second term of equation (2), and thus

$$\hat{w} = \frac{\mu_1 - \mu_2}{A\sigma_1^2}, \tag{4}$$

which implies that

$$\hat{W} = \frac{C}{1 - \frac{\mu_1 - \mu_2}{A\sigma_1^2}}. \tag{5}$$

Note that among all the moments of the second stock, only the expected return affects the marginal cost of diversification. Therefore, if there were other uncorrelated stocks, the second stock the investor adds would be the stock with the second highest expected return. If a stock is correlated with the first stock, then its covariance with the first stock affects the diversification effectiveness and thus the marginal benefit of diversification. In this case, both expected return and covariance affect stock selection. However, other moments such as variance and skewness do not affect this choice.

As wealth increases further, the investor invests more in both stocks. If it is optimal to have no leverage in the unconstrained case, i.e.,

$$w_1^* + w_2^* < 1$$

where

$$w_1^* = \frac{\mu_1}{A\sigma_1^2}, w_2^* = \frac{\mu_2}{A\sigma_2^2},$$

then when the wealth is high enough, the investor holds the tangency portfolio. However, if

$$\frac{\mu_1 - \mu_2}{A\sigma_1^2} > 1, \tag{6}$$

then because of the no-borrowing constraint ($w < 1$), the marginal benefit of diversification is always lower than the marginal cost, no matter how high W_0 is. Therefore, in this case, the investor never invests in Stock 2 and thus always underdiversifies regardless of his wealth level.

Example: Suppose $r = 0$, $A = 2$, $\mu_1 = 0.15$, $\mu_2 = 0.05$, $\sigma_1 = 0.3$, $\sigma_2 = 0.05$, $\underline{C} = \$10,000$. If the investor's wealth W_0 is equal to $10,000, then all $10,000 must be invested in the risk-free asset. If W_0 is above $10,000, then he starts to invest only in Stock 1 because it has the highest expected return, even though Stock 2 has a much greater Sharpe ratio. He will invest only in Stock 1 until his wealth gets above $\hat{W} = \$22,500$ by equation (5), i.e., he will invest up to 56% of his wealth in one stock that has a higher risk and lower Sharpe ratio. In addition, if $A = 1$, then by equation (6) the investor will never invest in Stock 2 no matter how wealthy he is.

Figure 1 plots the mean-variance frontier with three uncorrelated stocks. Stocks 1–3 are sorted by expected returns from the highest to the lowest. Because of the no-borrowing constraint, the security market line above the tangency point T is no longer relevant. Because of the short-sales constraints, the dotted segments are no longer achievable so the relevant frontier becomes the curve ABCTR. The investor's utility function is $u = \mu w - \frac{1}{2}w^2\sigma^2$ and thus the indifference curve at utility level u is

$$\mu = u/w + \frac{1}{2}Aw\sigma^2. \tag{7}$$

IC 1 in Figure 1 is the indifference curve when the investor's wealth is just slightly above the subsistence level (i.e., w is small), which is almost horizontal by equation (7), because the slope $\partial\mu/\partial\sigma$ is almost 0. To maximize utility, the investor chooses point A, which represents investment only in Stock 1. As w increases, the indifference curve becomes more curved. When it becomes tangent to ABCTR at point A, the investor starts to add Stock 2. As w

continues to increase, the tangent point moves from A to B, at which point the investor adds Stock 3. As w increases further, the investor invests more in each of the three stocks, as shown by point C implied by IC 2. After the tangency portfolio T is reached, the investor also increases the investment in the risk-free asset. If the tangency point between the indifference curve and the frontier ABCTR never moves beyond point B for any wealth level, then even wealthy investors will not hold Stock 3 and thus underdiversify.

III. The Model

In this section, I show that the main results that (1) underdiversification can be a result of solvency requirement, (2) less wealthy investors underdiversify more, and (3) stock selection does not depend on stock return variance or any higher moments hold for general expected utility preferences and general stock payoff distributions. Specifically, I consider a one-period discrete-time portfolio choice model in which an investor with initial wealth W_p can invest in one risk-free asset and $n \geq 1$ finitely many risky stocks and maximizes his expected utility from the end-of-period wealth $\widetilde{W_1}$. For expositional simplicity, the utility function $u(W)$ is assumed to be strictly increasing, strictly concave, and twice continuously differentiable.

According to the vast literature on consumption behavior, a large proportion of households commit to a critical consumption level and more than 50% of the average household's budget is devoted to ensure this level of consumption for moderate wealth shocks (e.g., Fratantoni (2001), Chetty and Szeidl (2007), Postlewaite, Silverman, and Samuelson (2008)). Models with consumption commitments are strongly supported by various empirical tests and have been shown to better fit consumption data and help explain various puzzles (e.g., Flavin and Nakagawa (2008), Chetty and Szeidl (2007)). Findings in this literature suggest that (1) households tend to commit to a certain consumption level and (2) the committed consumption level is generally above the subsistence level and thus the marginal utility at the committed level is finite. Consistent with these findings, I assume that[6]

Assumption 1 *The investor commits to an exogenous terminal consumption of* $\underline{C} \geq 0$ *at which the marginal utility is finite and must remain solvent after the committed consumption, i.e.,* $\widetilde{W}_1 \geq \underline{C} \geq 0$ *almost surely and the right derivative* $u'(\underline{C}) < \infty$.[7]

Assumption 1 is also related to the habit formation literature. The key difference from the standard habit formation literature is that the marginal utility at the "habit" level is not infinite, i.e., the "habit" level is above the subsistence level. In other words, even though the investor suffers a huge utility loss when his consumption falls below the "habit" level by only a small amount, he can still survive.[8]

[6]Although for expositional simplicity, I focus on a one-period static model in the main text, the results extend to a multi-period dynamic model. Accordingly, I can interpret an initial wealth change or a committed consumption 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.

[7]Assuming the committed consumption level $\underline{C}$ is exogenous is clearly a simplification. A reasonable model of the determination of the committed consumption level $\underline{C}$ should include at least historical consumption, income and wealth, future income expectation, cultural factors, peer group consumption, and health conditions. Such a model would divert this paper from its main focus and is unlikely essential for the main results, as the committed consumption level tends to be persistent and any effect of potential endogeneity issue is likely of a second order.

[8]For the main results, one can also use the typical habit formation utility function form $u(W - \underline{C})$, as long as $u'(0) < \infty$ and the constraint $W \geq \underline{C}$ 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 $W \geq \underline{C}$ is also necessary.

The risk-free interest rate is normalized to 0. Let $\tilde{P}$ denote the end-of-period gross return vector of the stocks. We assume that the gross return $\tilde{P}_i$ $(i = 1, 2, ..., n)$ is unbounded above and can get arbitrarily close to 0. To ensure solvency, the investor cannot borrow or shortsell. [9]

Let $\tilde{z} = \tilde{P} - \bar{1}$ be the return vector, $\mu \equiv (\mu_1, \mu_2, ..., \mu_n)^\top = E[\tilde{z}]$ be the expected return vector, and $\sigma\sigma^\top \equiv E[(\tilde{z} - \mu)(\tilde{z} - \mu)^\top]$ be the variance-covariance matrix. Without loss of generality, I assume that stock risk premia are all strictly positive.

Let θ denote the column vector of the dollar amount invested in the stocks. Given $W_p \geq \underline{C}$, the investor's problem is then

$$\max_{\theta} E\left[u(\widetilde{W_1})\right] \tag{8}$$

subject to

$$\widetilde{W_1} = W_p + \theta^\top \tilde{z} \geq \underline{C} \geq 0. \tag{9}$$

Because $\tilde{z}$ is unbounded above and can be arbitrarily close to -1, element by element, constraint (9) is equivalent to the no-borrowing and no-shortselling constraint:

$$\theta^\top \bar{1} \leq W_p - \underline{C} \;\text{ and }\; \theta \geq 0. \tag{10}$$

[9]No borrowing or shortselling is observationally close to what is found in individual trading behavior. For example, the results of Anderson (1999) and Boehmer, Jones, and Zhang (2005) imply that a vast majority of investors do not buy on margin and only about 1.5% of short sales come from individual investors. This seems to suggest that most individual investors is averse to risk solvency. As shown later, allowing limited borrowing and shortselling does not change my the main results.

Before I proceed, it is useful to clarify the meaning of underdiversification used in this paper.

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.*

Definition 2 *A portfolio is more underdiversified than another if it holds a smaller number of stocks.*[10]

The intuitions that drive the main results on underdiversification in the previous section still apply for general preferences and general payoff distributions. Specifically, with the no-borrowing and no-shortselling constraint implied by the solvency requirement, the amount of investment is limited by the initial wealth W_p. If W_p is small, then the investor only invests in the stocks that provide the highest marginal utility. Since the marginal utility at $\underline{C}$ is finite, investments in different stocks provide different levels of marginal utility in general. Therefore, the investor only invests in a small number of stocks when his wealth is low.[11] As

[10]The number of stocks in a portfolio is the most commonly used measure of diversification in the literature (e.g., Blume, Crockett, and Friend (1974), Vissing-Jorgensen (1999), Goetzmann and Kumar (2008)). Two other common measures of diversification are the volatility of a portfolio and the difference between the portfolio weights on stocks and the market portfolio stock weights (e.g., Goetzmann and Kumar (2008)). As shown later, in this paper, as wealth increases, the less wealthy invest in a greater number of stocks, the volatility of the portfolio decreases, and the portfolio weights get closer to the market portfolio weights. So all these three measures are highly positively correlated in my model. One advantage of using the number of stocks as a measure for underdiversification is that there is virtually no estimation error.

[11]As we show later, the number of stocks held can be small even for a relatively high wealth level if the number of stocks that have close-to-the-highest expected returns is small,

his wealth increases, he invests more in these stocks and risks increase, which drives down the marginal utility of investing any additional amount in these stocks. Beyond a critical wealth level, the marginal utility of investing more in the existing stocks becomes lower than investing in a new stock and thus the investor adds a new stock that provides the next highest marginal utility. In addition, for the choice of stocks, since the local risk neutrality still holds in this more general setting, higher moments such as variance and skewness are still irrelevant. However, different from the previous section, if stocks are correlated, then the covariance of the return of a stock with the current portfolio return affects the magnitude of the diversification benefit and thus also the marginal utility that this stock can provide. Therefore in addition to expected returns, covariances with the current portfolio also affect stock selection. With these intuitions in mind, I collect the main analytical results in the following theorem that is proven in the online Appendix.

Theorem 1 *We have:*

1. *For low enough initial wealth, the investor always underdiversifies;*
2. *As* $W_p - \underline{C}$ *increases, the investor invests a greater dollar amount in a greater number of stocks;*
3. *Whether a stock is selected into a portfolio or not depends only on its expected return and its covariance with the rest of the portfolio, but not on any other moments (e.g., variance and skewness). In particular, the Sharpe ratio is irrelevant for stock selection;*
4. *For some utility functions (e.g., CRRA or mean-variance) and some return distributions, it is optimal for the investor to always underdiversify no matter how wealthy he is;*

---

which can be justified by small fixed cost of trading a stock.

5. *If* $u'(\underline{C}) = \infty$*, then investors hold all the stocks as long as* $W_p > \underline{C}$*.*

Part 4 of Theorem 1 suggests under some conditions on preferences and return distributions, no matter how wealthy an investor is, he always underdiversifies. As explained in Section II, this is because the investor cannot borrow and for some preferences and return distributions, the marginal cost of diversification is still greater than the marginal benefit of diversification even when he invests 100% of his wealth in a proper subset of available stocks. This result is consistent with the empirical evidence that even the wealthy may underdiversify.

Part 5 and the main results on underdiversification show that the assumption of infinite marginal utility at $\underline{C}$ is critical for the standard diversification result that investors should diversify regardless of their wealth levels.

In the baseline model, to ensure solvency, an investor cannot borrow or shortsell in a discrete-time setting. However, even if an investor is allowed to borrow and shortsell and to trade continuously, as long as he can only borrow or shortsell a limited multiple of the initial wealth (e.g., with margin requirement, Cuoco and Liu (2000)), my results still hold. This is because when W_p is small enough, a limited multiple of W_p is also small and the investor still underdiversifies. In addition, the local risk neutrality argument still applies and thus, as before, only expected returns and covariances affect stock selection.

A. Additional Examples: Investor with a Nontradable Asset

Many investors have illiquid assets such as retirement portfolios, houses, and other durable goods. These illiquid assets are typically too costly to liquidate for daily consumptions. In this subsection, I provide some examples to show that the main results still hold and can even be stronger in the presence of illiquid assets. Thus, this paper can also help explain why investors with a diversified retirement portfolio underdiversify in the directly held portfolio (e.g., Goetzmann and Kumar (2008), Polkovnichenko (2005)).[12]

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. We assume that the nontradable asset is held for future consumption beyond the next period and so cannot be used for the next period's committed consumption $\underline{C} \geq 0$. Therefore, the investor requires the terminal *tradable* wealth be above $\underline{C} \geq 0$.

Given $W_p \geq \underline{C}$, the investor's problem is then

$$\max_{\theta} E\left[u(\widetilde{W_1})\right] \tag{11}$$

subject to

$$\widetilde{W_1} = W_p + \theta^{\top} \tilde{z} + \tilde{N} \tag{12}$$

and

$$W_p + \theta^{\top} \tilde{z} \geq \underline{C}, \tag{13}$$

[12]The "illiquidity" can be interpreted more broadly as the investor's unwillingness to change for whatever reasons. For example, an investor may allocate funds to different assets for different goals (e.g., a retirement portfolio is specifically for retirement, an education fund is specifically for college tuition, etc.) and thus is not willing to change these investments frequently. This unwillingness to change likely applies to mutual fund type assets held outside the retirement account by some investors if for some reasons (e.g., extra investment for retirement, paying for children's tuition), investors are unwilling to frequently change these investments.

which, as before, is equivalent to

$$\theta^{\top}\bar{1} \leq W_p - \underline{C} \text{ and } \theta \geq 0. \tag{14}$$

Suppose $W_p = \underline{C} + \eta$ with $\eta > 0$ and the investor invests η in Stock i. Then

$$\widetilde{W_1} = \underline{C} + \eta(\tilde{z}_i + 1) + \tilde{N}.$$

The marginal utility of investing in Stock i is

$$\frac{\partial E[u(\widetilde{W_1})]}{\partial \eta} = E[u'(\underline{C} + \eta(\tilde{z}_i + 1) + \tilde{N})(\tilde{z}_i + 1)], \tag{15}$$

which implies that, as η approaches 0, the marginal utility converges to

$$\lim_{\eta \downarrow 0} \frac{\partial E[u(\widetilde{W_1})]}{\partial \eta} = E[u'(\underline{C}+\tilde{N})(\tilde{z}_i+1)] = E[u'(\underline{C}+\tilde{N})](\mu_i+1)+\text{Cov}\left(u'\left(\underline{C}+\tilde{N}\right), \tilde{z}_i\right), \tag{16}$$

where the last equality follows from the covariance relation $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$ for any random variables X and Y. Therefore when wealth is low, the investor chooses to invest in only the stock with the combination of expected return and covariance with the nontradable asset that yields the highest marginal utility. Thus, as before, the investor underdiversifies when tradable wealth is low.

In addition, if a stock is positively correlated with the nontradable asset, then I have $\text{Cov}\left(u'\left(\underline{C}+\tilde{N}\right), \tilde{z}_i\right) < 0$ because the utility function is strictly concave. Therefore the marginal utility from this stock will be lowered by the positive correlation and thus the investor would be less willing to add this stock. As shown in the example below, if the nontradable asset is highly correlated with stocks, then the number of stocks directly held can be quite small, because the diversification benefit of additional stocks is small. This may help explain why investors with a diversified retirement portfolio directly hold a relatively small number of stocks.

Finally, while in the absence of an illiquid asset, the basic model implies that all investors share the same highest expected return risky assets, investors may hold different stocks in this generalized model with an illiquid asset. This is because covariance with the illiquid asset also matters and investors' illiquid asset holdings may be different.

Next, I graphically illustrate the effect of the ownership of a well-diversified retirement portfolio on underdiversification in the directly held portfolio. For simplicity, I again specialize to the mean-variance preference case.

Suppose there are $n > 1$ risky assets and one risk-free asset that the investor can trade. The $(n+1)^{st}$ risky asset is nontradable. Let W_p 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 and 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 $(W_p + W_N)$ invested in the tradable risky assets and $w_N \equiv W_N/(W_p + W_N)$ be the initial fraction of total wealth held in the nontradable asset. Then constraint (14) is equivalent to

$$w^\top \bar{1} \leq 1 - \frac{\underline{C} + W_N}{W_p + W_N}, \quad w \geq 0. \tag{17}$$

The mean-variance investor then solves the following problem:

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

subject to constraint (17), where $A > 0$ measures the investor's risk aversion.

Suppose there are 50 stocks, an investor's committed consumption is $\underline{C} > 0$, and he holds an equally weighted (i.e., 2% in each stock) nontradable retirement portfolio worth of $W_N = 30\underline{C}$. Figure 2 plots the optimal number of stocks held in the directly held portfolio against $W_p/\underline{C}$ for two risk-aversion levels. This figure shows that consistent with the empirical finding, the number of stocks an investor directly holds can be quite small. For example, suppose $\underline{C} = \$10,000$ and thus the investor has $\$300,000$ invested in the nontradable retirement portfolio. Figure 2 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 four stocks, he needs to have at least $\$299,000$ wealth outside the retirement account (not shown in the figure). The 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.

IV. An Equilibrium Model for Underdiversification

In this section, I show that underdiversification can arise in equilibrium and the equilibrium model can help explain many of the main empirical findings. Specifically, I consider a one-period model where investors maximize the expected utility from the final wealth on date 1. 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. We choose the consumption good as the numeraire and thus the interest rate is normalized to 0. There are $n > 0$ risky stocks with a positive total supply of $\bar{\omega}$ $(n \times 1)$ shares. The per-share payoffs $\tilde{P}$ $(n \times 1)$ are independently Gamma distributed with $n \times 1$ parameter vectors $\alpha > 1$ and $\beta > 0$.[13] The probability density function for Stock

[13]CARA preferences and Gamma distributions are used only for tractability. Allowing different risk aversions is an straightforward extension. The main qualitative results in this

j payoff $\tilde{P}_j$ is then

$$f_j(x) = \frac{x^{\alpha_j - 1} e^{-x/\beta_j}}{\beta_j^{\alpha_j} \Gamma(\alpha_j)},$$

with mean $\kappa_j = \alpha_j \beta_j$ and variance $\varphi_j^2 = \alpha_j \beta_j^2$, for $1 \le j \le n$.

There are two groups of investors: the wealthy, with mass 1, and the less wealthy, with mass $\lambda \ge 0$. Both types of investors are subject to the solvency constraint $\widetilde{W_1} \ge 0$. While committed consumption is not required because the marginal utility at zero wealth is finite for CARA preferences, in general "wealth" in this model should be interpreted as the remaining wealth after committed consumption when committed consumption is present.[14] The wealthy are endowed with $W_r \ge 0$ units of the consumption good and $\bar{\omega}$ shares of the stocks. The less wealthy are only endowed with $W_p \ge 0$ 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 shortsell. Let p denote the $n \times 1$ date 0 equilibrium stock price vector. The wealthy solve

$$\max_{\omega} E\left[-e^{-A\widetilde{W_1}}\right], \tag{18}$$

s.t.,

$$\widetilde{W_1} = W_r + (\bar{\omega}^\top - \omega^\top)p + \omega^\top \tilde{P} \ge 0, \tag{19}$$

where the $n \times 1$ vector ω denotes the number of shares held in stocks until date 1.

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.

[14]Assuming the wealthy has a greater dollar amount committed consumption (e.g., proportional to wealth) would strengthen the main results.

First, suppose all investors in the economy are wealthy, i.e., $\lambda = 0$. In this case, the solvency constraint is not binding for the wealthy, because $W_r \geq 0$ and the wealthy's stock endowment is positive. Then Problem (18) becomes

$$\max_{\omega} E\left[-e^{-A(W_r+\bar{\omega}^\top p+\omega^\top(\tilde{P}-p))}\right] = -e^{-A(W_r+\bar{\omega}^\top p)} \prod_{j=1}^{n} \min_{\omega_j} \frac{e^{Ap_j\omega_j}}{(1+A\omega_j\beta_j)^{\alpha_j}}. \tag{20}$$

The first-order conditions then imply that

$$\omega_j = \left(\frac{\kappa_j}{p_j} - 1\right)\frac{\kappa_j}{A\varphi_j^2}, \tag{21}$$

and thus

$$p_j = \frac{\kappa_j^2}{\kappa_j + A\omega_j\varphi_j^2}. \tag{22}$$

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

$$p_j = \frac{\kappa_j^2}{\kappa_j + A\bar{\omega}_j\varphi_j^2}, \tag{23}$$

which implies that the equilibrium expected return is

$$\mu_j = \frac{\kappa_j}{p_j} - 1 = A\bar{\omega}_j\frac{\varphi_j^2}{\kappa_j} \tag{24}$$

and the equilibrium return volatility is

$$\sigma_j = \frac{\varphi_j}{p_j} = \frac{\varphi_j}{\kappa_j}\left(\mu_j + 1\right). \tag{25}$$

To simplify notation, I label the risk-free asset as "Stock" $n+1$ with price $p_{n+1} = 1$, expected payoff $\kappa_{n+1} = 1$, payoff volatility $\varphi_{n+1} = 0$, expected return $\mu_{n+1} = 0$, and return volatility

$\sigma_{n+1} = 0$. In addition, I assume the parameters α and β are such that

$$\infty > \mu_1 > \mu_2 > ... > \mu_n > \mu_{n+1} = 0. \tag{26}$$

When there are some less wealthy investors in the economy, by the same arguments as illustrated in the previous section, these investors underdiversify when their wealth is low. For example, if their wealth is close to 0, then they only invest in the stock with the highest expected return (i.e., Stock 1). As their wealth increases, less wealthy investors first increase the investment in Stock 1, then add the stock with the second highest expected return (i.e., Stock 2), then increase the investment in both Stock 1 and Stock 2, then add the stock with the third highest expected return (i.e., Stock 3), and so on until they are wealthy enough to hold the same portfolio as the wealthy and thus become fully diversified.[15]

This result suggests that when the wealth of the less wealthy investors is low, they hold a different portfolio from the wealthy and therefore no one in economy holds the market portfolio and idiosyncratic risks are priced in equilibrium, consistent with the findings of Ang et al. (2006).

For given $2 \leq i \leq n+1$, define

$$\hat{W}_i = \sum_{j=1}^{i-1} \frac{\alpha_j(\mu_j - \mu_i)}{A(\mu_i + 1)(\lambda + \mu_j + 1)}, \tag{27}$$

[15]Different from examples in Section II, I need to take into account the price impact of the less wealthy as they invest more in the stocks when their wealth increases. As I show in the online Appendix, as long as the total wealth of the less wealthy is finite, the order of the equilibrium expected returns remains the same as the case without the less wealthy. So the order in which a stock is added is indeed from 1 to n.

$$\bar{p}_j = p_j \frac{\lambda/(k+1)+1}{\lambda/(\mu_j+1)+1}, \quad j = 1, 2, ..., i-1, \tag{28}$$

$$\bar{\mu}_j = \frac{\lambda}{\lambda+k+1} k + \frac{k+1}{\lambda+k+1} \mu_j, \quad j = 1, 2, ..., i-1, \tag{29}$$

$$\bar{\sigma}_j = \frac{\varphi_j}{\kappa_j} (\bar{\mu}_j + 1), \quad j = 1, 2, ..., i-1, \tag{30}$$

$$\delta_j = \frac{\alpha_j(\mu_j - k)}{A(k+1)(\lambda+\mu_j+1)} = \frac{\bar{\mu}_j+1}{k+1} \frac{\bar{\mu}_j - k}{A\bar{\sigma}_j^2}, \quad j = 1, 2, ..., i-1, \tag{31}$$

and

$$k = \frac{\sum_{j=1}^{i-1} \frac{\alpha_j \mu_j}{\lambda+\mu_j+1} - A \sum_{j=1}^{i-1} \delta_j}{\sum_{j=1}^{i-1} \frac{\alpha_j}{\lambda+\mu_j+1} + A \sum_{j=1}^{i-1} \delta_j}. \tag{32}$$

We now summarize the main results in the following theorem that is proven in the online Appendix.

Theorem 2 *Let μ_j be as defined in equation (24) such that inequality (26) holds and let $\hat{W}_j$ be as defined in equation (27) with $\hat{W}_1 \equiv 0$, for $j = 1, 2, ..., n+1$.*[16] *Then*

1. *if $W_p = 0$, then for $j = 1, 2, ..., n$, the equilibrium price p_j for Stock j is as stated in equation (23), the equilibrium expected return μ_j is as stated in equation (24), and the equilibrium return volatility σ_j is as stated in equation (25);*

2. *if $W_p \in (\hat{W}_{i-1}, \hat{W}_i]$ for some $2 \leq i \leq n+1$, then a less wealthy investor invests only in the first $i-1$ stocks. In addition, for $j = 1, 2, ..., i-1$, the equilibrium price $\bar{p}_j$, the equilibrium expected return $\bar{\mu}_j$, the equilibrium return volatility $\bar{\sigma}_j$, and the dollar*

[16] $\equiv$ should be read as "is defined as."

amount δ_j invested in Stock j are as stated in equations (28), (29), (30), and (31) respectively;

3. *as W_p approaches $\hat{W}_{n+1}$, the less wealthy investor's portfolio converges to that of the wealthy and thus everyone holds the market portfolio;*

4. *if $W_p \in (0, \hat{W}_{n+1})$, then no one holds the market portfolio, CAPM does not hold, and idiosyncratic risks are priced.*

Equation (31) implies that the amount an investor invests in a stock (once selected) decreases with volatility and risk aversion, which in particular implies that less risk-averse investors underdiversify more. Morin and Suarez (1983) and Palsson (1996) find that younger investors and male investors are less risk averse. Given these findings, this paper may help explain the empirical finding that younger or male investors tend to underdiversify more (Mitton and Vorkink (2007)).

Theorem 2 implies the following empirically testable predictions:

1. For low initial wealth, less wealthy investors invest only in a small number of stocks;

2. Less diversified investors tend to choose stocks with high expected returns regardless of risks;

3. Higher moments (e.g., variance and skewness) do not affect stock selection;

4. As the initial wealth of the less wealthy increases, the number of the stocks the less wealthy investors hold also increases;

5. The amount an investor invests in a stock *after* the stock is selected decreases with its volatility;

6. For the same initial wealth, less risk-averse investors invest in a smaller number of stocks;

7. In equilibrium, CAPM does not hold and idiosyncratic risks are priced.

There is an extensive literature (e.g., Calvet, Campbell, and Sodini (2007), Goetzmann and Kumar (2008), and Mitton and Vorkink (2007)) showing that a less diversified stock portfolio has a greater expected return, a higher volatility, a greater skewness, and a lower Sharpe ratio.[17] As far as I know, however, there are no existing models that can help explain all of these findings. In contrast, the following theorem shows that the model in this paper can.

Theorem 3 *In equilibrium, as W_p increases, the expected return, volatility, and skewness of a less wealthy investor's stock portfolio all decrease. In addition, if the investor holds more than one stock and λ is small, then as W_p increases, the Sharpe ratio also increases.*

As W_p increases, the less wealthy investor's portfolio becomes less underdiversified. Theorem 3 implies that consistent with the empirical evidence, a less underdiversified stock portfolio has a lower expected return, a lower volatility, and a lower skewness. When the total wealth of the less wealthy investors is small relative to that of the wealthy, a less underdiversified portfolio also has a higher Sharpe ratio. Intuitively, as wealth increases, the less wealthy investors invest in a greater number of stocks with lower expected returns, and thus a less underdiversified portfolio has a lower expected return and lower skewness, and, because of diversification, also a lower volatility. Whether the Sharpe ratio is lower or higher for a more diversified portfolio depends on the relative impact of diversification on the expected return and volatility. As the less wealthy investors buy more stocks, their price impact drives down both the volatility and the expected return. In addition to the price impact, volatility is also driven down by diversification. When the total wealth of the

[17] Although a less diversified portfolio tends to have a greater expected return, some less diversified investors may underperform net of transaction costs if they trade excessively.

less wealthy investors is small relative to that of the wealthy, the price impact of the less wealthy investors is small and thus expected return decreases by less than the volatility, and therefore the Sharpe ratio increases.

A. Graphical Illustrations

Next I provide some graphical illustrations of the main analytical results both qualitatively and quantitatively. We assume the relative risk-aversion coefficient of an investor with $\$20,000$ initial remaining wealth after committed consumption ("wealth" for short) is 2. This translates into an absolute risk aversion coefficient of $A = 10^{-4}$. We arbitrarily set the total number of shares for stocks at $\bar{\omega} = 10^4\bar{\mathbf{1}}$, where $\bar{\mathbf{1}}$ is a vector of 1's. For these illustrations, I then choose parameters α and β such that the top five stocks (Stocks 1-5) have expected returns evenly distributed from 25% to 5% and return volatilities evenly distributed from 56% to 25%. Figure 3 shows that when the wealth of an investor is low, he invests in a small number of stocks. For example, with $20,000, he invests only in four stocks, which is largely consistent with empirical evidence.[18] For example, Mitton and Vorkink (2007) find that in their database with the portfolios of 78,000 households the median portfolio value with only four stocks is $21,903.

Only when his wealth rises above $30,000 does he add the fifth stock. Also Figure 3 shows that a more risk-averse investor underdiversifies less. To help explain underdiversification shown in Figure 3, I plot the ratios of the marginal utility of investing in the five stocks to that of investing in Stock 5 in Figure 4. Figure 4 shows that when the wealth W_p is low enough (i.e., to the left of point B), the marginal utility of investing in Stock 1 is the highest.

[18]Note that since he chooses stocks in the order of their expected returns, the presence of additional stocks with expected returns lower than 5% will not affect his selection of these four stocks.

As W_p increases, the less wealthy investor invests more in Stock 1 and thus the marginal utility of investing in Stock 1 decreases. At point B, the marginal utility of investing more in Stock 1 becomes equal to that of investing a small amount in Stock 2 and the investor adds Stock 2. Between B and C, the investor increases investment in Stocks 1 and 2 so that the marginal utilities become lower but stay always the same across these two stocks as required by optimality. Beyond point C, the investor adds Stock 3. Between C and D, the investor increases investment in Stocks 1, 2 and 3 so that the marginal utilities are driven even lower but still always stay the same across the three stocks. Similarly, between D and E, the investor adds Stock 4 and invests more in Stocks 1–4 as his wealth further increases. Beyond point E, the investor also invests in Stock 5.

Figure 5 confirms that a more diversified portfolio has a lower expected return and also a lower volatility. Intuitively, as the wealth of the less wealthy increases, they sequentially add stocks with lower expected returns, which drives down their portfolio expected return. Due to diversification, the portfolio volatility also decreases. When the less wealthy's wealth increases beyond a critical level, they also fully diversify as the wealthy, and both the expected return and the volatility of their portfolio are driven down to the lowest. Figure 6 shows that the skewness of the portfolio return is also lower for a more diversified portfolio. In addition, it shows that the Sharpe ratio of a more diversified portfolio is higher and therefore diversification improves mean-variance efficiency. All these patterns are consistent with empirical findings such as those in Calvet, Campbell, and Sodini (2007), Goetzmann and Kumar (2008), and Mitton and Vorkink (2007). Although skewness does not impact stock selection in my model, the implied skewness pattern as shown in Figure 6 may appear to indicate that skewness is important for this choice.

V. Concluding Remarks

We show that solvency requirement in the presence of committed consumption can help explain many of the empirical findings on underdiversification and the relevance of idiosyncratic risks for asset pricing. In particular, I demonstrate that investors always underdiversify when wealth is low and less wealthy investors underdiversify more. In addition, I show that investors with expected utility preferences choose stocks solely by expected returns and covariances with already selected stocks, and any other moments (e.g., variance and skewness) are irrelevant for this choice. For investors with a well-diversified illiquid portfolio (e.g., a retirement portfolio), it can be optimal to hold only an even smaller number of stocks directly. In an equilibrium with underdiversification, no one holds the market portfolio and idiosyncratic risks are priced.

While the main results are shown for (quite general) risk-averse expected utility preferences, they also hold for many alternative preferences. For example, the main result that investors underdiversify more when wealth is low than when wealth is high 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-utility 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 disappointment aversion preferences (Gul (2000)), recursive preferences (Epstein and Zin (1989)), loss aversion preferences (Kahneman and Tversky (1979)), and Machina preferences (Machina (1982)), as long as the marginal utilities they can yield at zero investment are different.

Therefore the main results hold for quite general preferences and asset return distributions. Moreover, if the marginal utility is finite at zero wealth (e.g., non-CRRA HARA preferences), then the requirement of solvency itself is sufficient. 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.

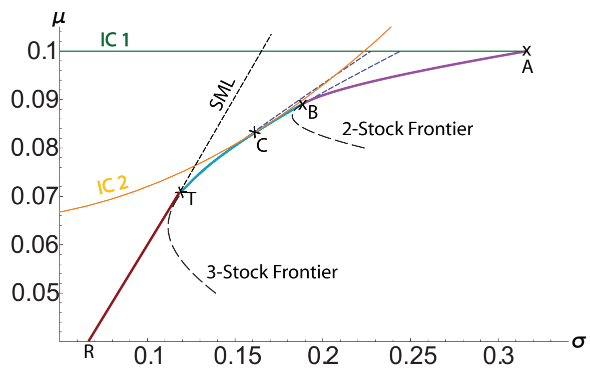

Figure 1: Mean-Variance Efficient Frontier.

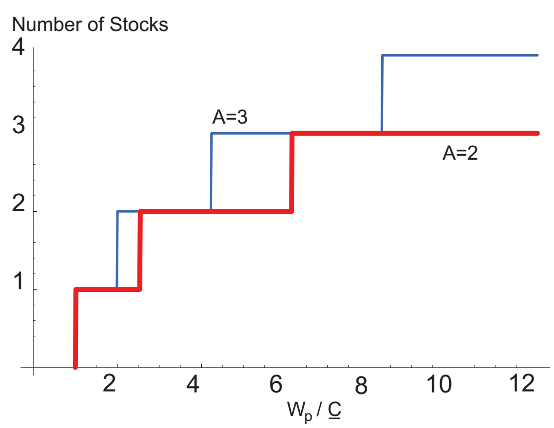

Figure 2: **Number of stocks directly held against $W_p/\underline{C}$ given 50 independent stocks.**

Parameter Values: for $i = 1, 2, ..., 50$, $\mu_i = 0.2 - (0.2 - 0.01)(i-1)/49$, $\sigma_i = 0.2 - (0.2 - 0.1)(i-1)/49$, $W_N/\underline{C} = 30$, and the retirement portfolio invests 2% in each stock.

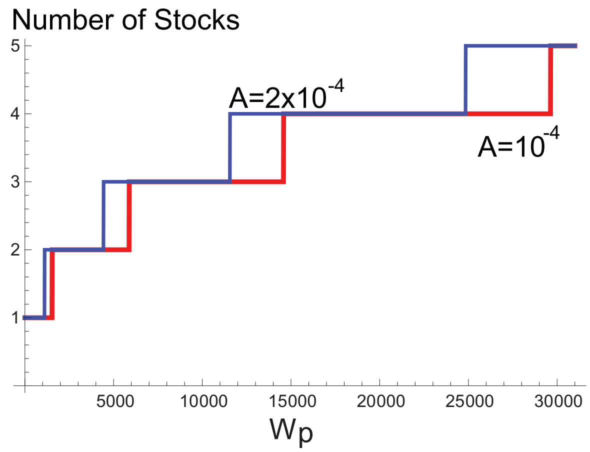

Figure 3: **The number of stocks in the less wealthy investor's stock portfolio against W_p.**

Parameter Values: for $i = 1, 2, ..., 5$, $\beta_i = 0.25 - (0.25 - 0.001)(i - 1)/5$, $\alpha_i = 20 - (20 - 5)(i - 1)/5$, $\bar{\omega}_i = 10^4$, and $\lambda = 0.1$.

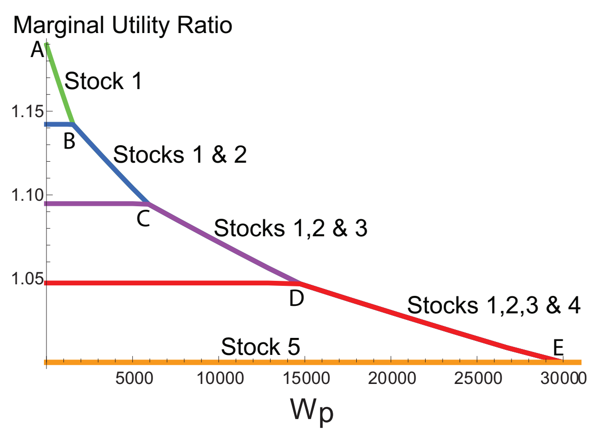

Figure 4: **The marginal utility ratios against W_p.**

Parameter Values: for $i = 1, 2, ..., 5$, $\beta_i = 0.25 - (0.25 - 0.001)(i - 1)/5$, $\alpha_i = 20 - (20 - 5)(i - 1)/5$, $\bar{\omega}_i = 10^4$, and $\lambda = 0.1$.

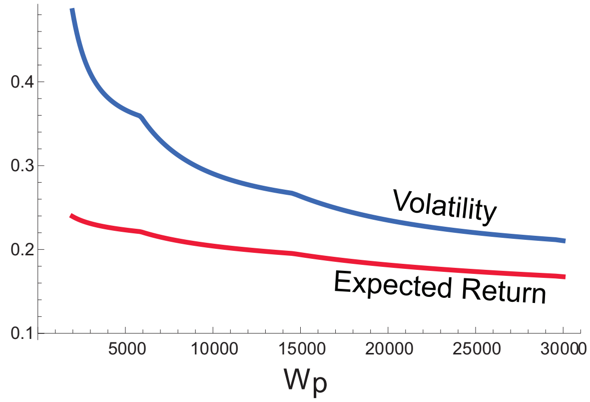

Figure 5: **Stock portfolio expected return and volatility against W_p.**

Parameter Values: for $i = 1, 2, ..., 5$, $\beta_i = 0.25 - (0.25 - 0.001)(i - 1)/5$, $\alpha_i = 20 - (20 - 5)(i - 1)/5$, $\bar{\omega}_i = 10^4$, $\lambda = 0.1$, and $A = 10^{-4}$.

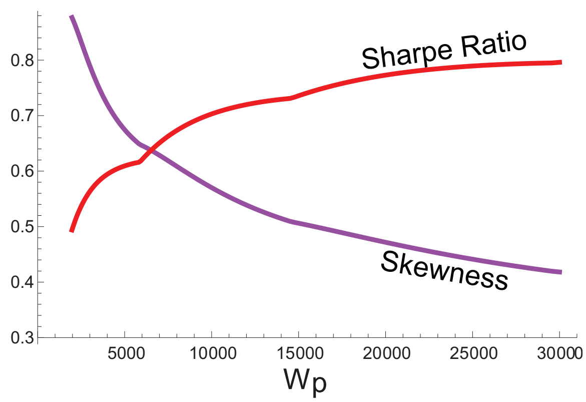

Figure 6: **Stock portfolio Sharpe ratio and skewness against wealth W_p.**

Parameter Values: for $i = 1, 2, ..., 5$, $\beta_i = 0.25 - (0.25 - 0.001)(i - 1)/5$, $\alpha_i = 20 - (20 - 5)(i - 1)/5$, $\bar{\omega}_i = 10^4$, $\lambda = 0.1$, and $A = 10^{-4}$.

References

Ang, A.; R. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61 (2006), 259–299.

Barberis, N. C., and M. Huang. "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices." Working Paper, Yale University (2007).

Bessembinder, H. "Systematic Risk, Hedging Pressure, and Risk Premiums in Futures Markets." *Review of Financial Studies*, 5 (1992), 637–667.

Blume, M. E.; J. Crockett; and I. Friend. "Stockholders in the United States: Characteristics and Trends." *Survey of Current Business* 54 (1974), 16–40.

Boehmer, E.; C. M. Jones; and X. Zhang. "Which Shorts Are Informed?" Working Paper, Columbia University (2005).

Boyle, P.; L. Garlappi; R. Uppal; and T. Wang. "Keynes Meets Markowitz: The Trade-off between Familiarity and Diversification." Working Paper, London Business School (2009).

Brennan, M. J. "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 (1975), 483–496.

Calvet, L. E.; Campbell, J. Y.; and P. Sodini. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* , 115 (2007), 707–747.

Campbell, J. Y. "Household Finance." *Journal of Finance*, 61 (2006), 239–244.

Chetty R., and Szeidl, A. "Consumption Commitments: A Foundation for Reference-Dependent Preferences and Habit Formation." Working Paper, Harvard University (2010).

Chetty R., and Szeidl, A. "Consumption Commitments and Risk Preferences," *Quarterly Journal of Economics*, 122 (2007), 831–877.

Constantinides, G. M. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy*, 98 (1990), 519–543.

Cuoco, D., and H. Liu. "A Martingale Characterization of Consumption Choices and Hedging

Costs with Margin Requirements." *Mathematical Finance*, 10 (2000), 355–385.

Cuoco, D., and H. Liu. "An Analysis of VaR-based Capital Requirements." *Journal of Financial Intermediation*, 15 (2006), 362–394.

Dybvig, P. H. "Short Sales Restrictions and Kinks on the Mean Variance Frontier." *Journal of Finance*, 36 (1984), 239–244.

Dybvig, P. H. "Duesenberry's Racheting of Consumption: Optimal Dynamic Allocation and Investment Given Intolerance for any Decline in Standard of Living." *Review of Economic Studies* 62 (1995), 287–313.

Epstein, L. G., and S. E. Zin. "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns." *Econometrica*, 57 (1989), 937–969.

Flavin, M., and S. Nakagawa. "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence." *American Economic Review*, 98 (2008), 474–495.

Fratantoni, M. C. "Homeownership, Committed Expenditure Risk, and the Stockholding Puzzle." *Oxford Economic Papers*, 53 (2001), 241-259.

Friedman, M., and L. J. Savage. "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy*, 56 (1948), 279–304.

Goetzmann, W. N., and A. Kumar. "Equity Portfolio Diversification." *Review of Finance*, 12 (2008), 433–463

Gul, F. "A Theory of Disappointment Aversion." *Econometrica*, 59 (1991), 667–686.

Huberman, G. "Familiarity Breeds Investment." *Review of Financial Studies*, 14 (2001), 659–680.

Ivkovich, Z.; C. Sialm; and S. Weisbenner. "Portfolio Concentration and the Performance of Individual Investors." *Journal of Financial and Quantitative Analysis*, 43 (2008), 613–656.

Ivkovich, Z., and S. Weisbenner. "Local Does as Local is: Information Content of the Geography of Individual Investors Common Stock Investments." *Journal of Finance*, 60 (2005), 267–306.

Kraus, A., and R. Litzenberger. "Skewness Preference and the Valuation of Risk Assets." *Journal of Finance*, 31 (1976), 1085–1100.

Kumar, A. "Do the Diversification Choices of Individual Investors Influence Stock Returns?" *Journal of Financial Markets*, 10 (2007), 362-390.

Machina, M. J. "'Expected Utility' Analysis without the Independence Axiom." *Econometrica*, 50 (1982), 277–323.

Merton, R. C. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance*, 42 (1987), 483–510.

Mitton, T., and K. Vorkink. "Equilibrium Underdiversification and the Preference for Skewness." *Review of Financial Studies*, 20 (2007), 1255–1288.

Morin, R. A., and F. Suarez. "Risk Aversion Revisited." *Journal of Finance*, 38 (1983), 1201–1216.

Nieuwerburgh, S. V., and L. Veldkamp. "Information Acquisition and Portfolio Underdiversification." Working Paper, New York University (2005).

Palsson, J. M. "Does the Degree of Relative Risk Aversion Vary with Household Characteristics?" *Journal of Economic Psychology*, 17 (1996), 771–787.

Polkovnichenko, V. "Household Portfolio Diversification: A Case for Rank Dependent Preferences." *Review of Financial Studies*, 18 (2005), 1467–1502.

Postlewaite, A.; D. Silverman; and L. Samuelson. "Consumption Commitments and Employment Contracts." *Review of Economic Studies*, 75 (2008), 559–578.

Ross, S. A. "The Capital Asset Pricing Model (CAPM), Short-sale Restrictions and Related Issues." *Journal of Finance*, 32 (1977), 177–183.

Shore, S., and T. Sinai, "Commitment, Risk, and Consumption: Do Birds of a Feather Have Bigger Nests?" *The Review of Economics and Statistics*, 92 (2010), 408-424.

Uppal, R., and T. Wang. "Model Misspecification and Underdiversification." *Journal of*

Finance, 58 (2003), 2465–2486.

Vissing-Jørgensen, A. “Comments on ‘Stock Prices and Fundamentals’ by John Heaton and Deborah Lucas.” *NBER Macroeconomics Annual* (1999).

Williams, D. “Probability with Martingales.” Cambridge, Cambridge University Press (1991).

Online Appendix for "Solvency Constraint, Underdiversification, and Idiosyncratic Risks" by Hong Liu

In this appendix, I provide the proofs for Theorems 1-3.

Proof of Theorem 1: Define the adjusted initial wealth as $W_p^A \equiv W_p - \underline{C}$. When $W_p^A = 0$, obviously, the investor can only invest in the risk-free asset. Now suppose his adjusted wealth increases to $W_p^A = \eta$, where $\eta > 0$ is small. Since he cannot borrow or shortsell, the most he can invest in stocks is η. Given this restriction, he will first invest in the stock that provides the highest marginal utility. To identify the stock that provides the highest marginal utility, suppose he invests $\underline{C}$ in the risk-free asset and η in Stock j with the end-of-period wealth

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

As η approaches 0, the investor's marginal utility from investing in Stock j converges to

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

where

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

the first equality in equation (A-1) 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 $\partial h(\eta, \tilde{z}_j)/\partial \eta = u''(\underline{C} + \eta(\tilde{z}_j + 1))(\tilde{z}_j + 1)^2 < 0$ by the strict concavity of $u(\cdot)$. Therefore when η is small enough, since $\infty > u'(\underline{C}) > 0$, the marginal utility from investing in the stock with the highest expected return is strictly the highest. Thus, if wealth W_p is slightly above the committed level $\underline{C}$, the investor invests $\underline{C}$ in the risk-free asset and the rest in the stock with the highest expected return, say, Stock 1. Higher moments, such as variance, skewness, and kurtosis, do not affect this choice.

We now prove by induction. Suppose that, given the initial adjusted wealth W_p^A, the investor optimally invests $\underline{C}$ in the risk-free asset, $\theta_i > 0$ in Stock i for $1 \le i \le m < n$, and 0 in the rest of the stocks. Since the investor is investing in only $m < n$ of n stocks, the constraint $\sum_{i=1}^m \theta_i \le W_p - \underline{C} = W_p^A$ must be still binding. This is because the marginal utility of investing a bit in the rest of the stocks is strictly greater than that of investing more in the risk-free asset, since the expected stock returns are strictly greater than the risk-free rate. The end-of-period wealth is $\widetilde{W}_1 = \underline{C} + \sum_{i=1}^m \theta_i(\tilde{z}_i + 1)$. The Lagrangian is

$$V(W_p^A) = \max_{\theta \ge 0} \left\{ E\left[u\left(\underline{C} + \sum_{i=1}^m \theta_i(\tilde{z}_i + 1) \right) \right] + \nu \left(W_p^A - \sum_{i=1}^m \theta_i \right) \right\},$$

where ν is the Lagrangian multiplier. From the first-order conditions, I have that the marginal utility of investing in Stock i for $1 \le i \le m$ is

$$\frac{\partial E\left[u\left(\widetilde{W}_1 \right) \right]}{\partial \theta_i} = E[u'(\widetilde{W}_1)(\tilde{z}_i + 1)] = \nu. \tag{A-2}$$

By the Envelop theorem, I have

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

since utility is strictly increasing in wealth. Therefore, at the initial level W_p^A, the marginal utilities of investing in these m stocks are equal, strictly positive, and strictly greater than the marginal utilities of investing more than $\underline{C}$ in the risk-free asset and the rest of the stocks (because it is optimal for the investor not to hold the rest of the stocks). Since

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

the marginal utility of investing in Stock i decreases as the investment in the stock increases. Thus, as the adjusted initial wealth increases from W_p^A, the investor will increase the investment in all these m stocks, i.e.,[19]

$$\text{(A-3)} \qquad \frac{\partial \theta_i}{\partial W_p^A} > 0, \quad i = 1, 2, ..., m.$$

Similar to equation (A-1), the marginal utility of investing a small amount in a new Stock j ($j > m$) at W_p^A is

$$\begin{aligned} E\left[u'\left(\underline{C} + \sum_{i=1}^{m} \theta_i(\tilde{z}_i + 1)\right)(\tilde{z}_j + 1)\right] &= E\left[u'\left(\underline{C} + \sum_{i=1}^{m} \theta_i(\tilde{z}_i + 1)\right)\right](\mu_j + 1) \\ \text{(A-4)} \qquad &+ \operatorname{Cov}\left(u'\left(\underline{C} + \sum_{i=1}^{m} \theta_i(\tilde{z}_i + 1)\right), \tilde{z}_j\right). \end{aligned}$$

Therefore, when it becomes optimal to add another stock to the portfolio, which stock to add only depends on its expected return and its covariance with the stocks in the current portfolio. Other moments of Stock j, such as variance and skewness, are irrelevant for this choice. If stocks are uncorrelated, then this choice only depends on the expected return of Stock j.

We next derive the critical adjusted wealth level $\hat{W}_{m+1}^A$ above which it is optimal to add another stock, say, Stock $m+1$. By equations (A-2) and (A-4) with $j = m+1$, I must have that at this critical level, the marginal utilities of investing in these $m+1$ stocks are exactly the same, i.e., for $i = 1, 2, ..., m$, I have

$$\text{(A-5)} \qquad E\left[u'\left(\underline{C} + \sum_{i=1}^{m} \theta_i(\tilde{z}_i + 1)\right)(\tilde{z}_i + 1)\right] = E\left[u'\left(\underline{C} + \sum_{i=1}^{m} \theta_i(\tilde{z}_i + 1)\right)(\tilde{z}_{m+1} + 1)\right],$$

[19] At W_p^A, the marginal utilities across all the m stocks are the same. When the wealth increases above W_p^A, if the investor increases investment only in some of the m stocks, then the marginal utilities from these stocks would be lowered and the marginal utilities from the rest of the m stocks would be strictly higher, which is a contradiction to optimality.

which can be simplified to

$$\text{(A-6)} \qquad E\left[u'\left(\underline{C}+\sum_{i=1}^{m}\theta_i(\tilde{z}_i+1)\right)(\tilde{z}_i-\tilde{z}_{m+1})\right]=0, \quad i=1,2,...,m.$$

Therefore, I have $\hat{W}^A_{m+1}=\sum_{i=1}^{m}\theta_i$ where θ_i's are the solution to equation (A-6). By equation (A-3), I have that as long as the constraint is binding, the critical wealth level $\hat{W}^A_i$ at which it is optimal to add Stock i strictly increases with i. Combining these results shows that both the number and the dollar amount of stocks optimally held increase as the adjusted wealth W^A_p increases. This completes the proof of Parts 1-3.

Now I show Part 4. If the investor has a CRRA utility, i.e.,

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

then equation (A-6) is equivalent to

$$\text{(A-7)} \qquad E\left[u'\left(1+\sum_{i=1}^{m}w_i\tilde{z}_i\right)(\tilde{z}_i-\tilde{z}_{m+1})\right]=0, \quad i=1,2,...,m,$$

where $w_i\equiv\theta_i/W_p$ represent the fraction of the initial wealth W_p invested in stock i. And the constraint (10) become

$$\sum_{i=1}^{n}w_i\leq 1-\frac{\underline{C}}{W_p} \quad \text{and} \quad w_i\geq 0, \quad i=1,2,...,n.$$

Due to the no-borrowing constraint, if the w_i's that solve equation (A-7) are such that $\sum_{i=1}^{m}w_i>1$ for some $m<n$, then the investor never holds more than m stocks, no matter how wealthy he is. To show that this can happen, suppose I have

$$\text{(A-8)} \qquad E\left[u'\left(1+w_1\tilde{z}_1\right)(\tilde{z}_1-\tilde{z}_j)\right]>0$$

for all $w_1 \in [0,1]$ and $j = 2, 3, ..., n$, which can hold if $\mu_1 - \mu_j$ is large enough, Stock 1's volatility is small enough, and Stock 1 is independent of other stocks. 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 any other stocks for all feasible w_1 (i.e., for $w_1 \in [0,1]$). A similar proof applies to mean-variance preferences.

For Part 5, if $u'(\underline{C})$ is infinite, then equation (A-1) implies that all stocks have the same (infinite) marginal utility at $\underline{C}$ irrespective of their expected returns and risks. It is therefore optimal to hold all the stocks even when the wealth of the less wealthy investor is just slightly above the committed level $\underline{C}$. As the wealth of the less wealthy increases, he will increase the investment in every stock such that the marginal utility from the investment in each stock stays the same. Therefore, if $u'(\underline{C})$ is infinite, an investor always invests in all the stocks as long as his wealth $W_p > \underline{C}$, as predicted in the standard portfolio selection theory. $\square$

Proof of Theorem 2: The case where $W_p = 0$ is the same as the case without the less wealthy, which is shown in the text. We now consider the case where $W_p = \eta > 0$. Suppose first η is small. Given the solvency constraint, the less wealthy cannot borrow or shortsell and thus the maximum amount that they can invest in any stock is η. Therefore, as long as investing in different stocks provides different marginal utilities, the investor will choose sequentially the stocks that provide the next highest marginal utility until his budget η is exhausted. We first examine the portfolio allocation problem of the less wealthy investors without taking into account the equilibrium price impact of their trades. The utility from investing a dollar amount $\eta > 0$ in Stock j is

$$U_j(\eta) = E\left[-e^{-A\frac{\eta}{p_j}\tilde{P}_j}\right] = -\left(1 + A\frac{\eta}{p_j}\beta_j\right)^{-\alpha_j}.$$

Accordingly, the marginal utility from investing η in Stock j is

(A-9) $$U_j'(\eta) = \frac{\frac{A\alpha_j\beta_j}{p_j}}{\left(1 + A\frac{\eta}{p_j}\beta_j\right)^{\alpha_j+1}},$$

which, as $\eta \downarrow 0$, converges to

(A-10) $$\lim_{\eta\downarrow 0} U_j'(\eta) = \frac{A\alpha_j\beta_j}{p_j} = A\frac{\kappa_j}{p_j} = A(\mu_j + 1),$$

where the last equality follows from equation (24). Therefore, irrespective of other moments (e.g., variance and skewness), the stock with the highest expected return yields the greatest marginal utility when the amount of investment η is small. By inequality (26) and the continuity of $U_j'(\eta)$, a less wealthy investor invests the entire amount η in Stock 1, when η is small enough.

We now derive the new equilibrium price for Stock 1, taking into account the equilibrium price impact of the less wealthy investor's purchase of Stock 1. Let $\hat{p}_1$ be the new equilibrium price of Stock 1. By equation (21), the market clearing condition for Stock 1 becomes

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

which implies that the new equilibrium price for Stock 1 is

(A-11) $$\hat{p}_1 = \frac{\kappa_1^2 + A\lambda\eta\varphi_1^2}{\kappa_1 + A\bar{\omega}_1\varphi_1^2} = p_1\left(1 + A\lambda\eta\frac{\varphi_1^2}{\kappa_1^2}\right),$$

and the new equilibrium expected return becomes

$$\hat{\mu}_1 = \frac{\kappa_1}{\hat{p}_1} - 1 = \mu_1 - \frac{A\lambda\eta\frac{\varphi_1^2}{\kappa_1^2}}{1 + A\lambda\eta\frac{\varphi_1^2}{\kappa_1^2}}(\mu_1 + 1).$$

Therefore, by inequality (26), there exists a small enough $\eta > 0$ such that inequality (26) still holds with μ_1 replaced by $\hat{\mu}_1$.[20] This shows that indeed when their wealth is low enough, less wealthy investors buy only Stock 1 in equilibrium and thus underdiversify.

We next show that as their wealth increases, less wealthy investors first increase the investment in Stock 1, then add the stock with the second highest expected return (i.e., Stock 2), then increase the investment in both Stock 1 and Stock 2, then add the stock with the third highest expected return (i.e., Stock 3), and so on until they are rich enough to hold the same portfolio as the wealthy and thus become fully diversified.

We show this by induction. Suppose at a higher wealth W_p a less wealthy investor invests only in the first $i-1$ stocks for $i \geq 2$. Let $\delta_j \geq 0$ denote the dollar amount invested in Stock j ($j = 1, 2, ..., n+1$) with $\delta_j > 0$ only for $j \leq i-1$ (recall that "Stock" $n+1$ is the risk free asset). Let $\bar{p}_j$ be the new equilibrium price for Stock j for $1 \leq j \leq n+1$. For $m \geq i$, since $\delta_m = 0$ and the wealthy's demand for a stock is independent of other stocks (as shown in equation (21)), the equilibrium price for Stock m remains the same as in the case with $\lambda = 0$, i.e., $\bar{p}_m = p_m$. Let

$$\begin{aligned} V(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1}) &= E\left[-\exp\left(-A\sum_{j=1}^{n}\frac{\delta_j}{\bar{p}_j}\tilde{P}_j - A\delta_{n+1}\right)\right] \\ &= -\prod_{j=1}^{n}\left(1 + A\frac{\delta_j}{\bar{p}_j}\beta_j\right)^{-\alpha_j} e^{-A\delta_{n+1}} \end{aligned} \quad \text{(A-12)}$$

be the value function of the less wealthy. Equation (A-12) implies that the marginal utility

[20]Since the less wealthy will only buy Stock 1 and by equation (21), the wealthy's demand for a stock is independent of any other stocks, the equilibrium prices and expected returns of other stocks remain the same.

from investing δ_j in Stock j is

$$\text{(A-13)} \qquad \frac{\partial V(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})}{\partial \delta_j} = A|V(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})| \frac{\kappa_j}{\bar{p}_j + A\delta_j \varphi_j^2 / \kappa_j},$$

which shows that for $m \geq i$, the marginal utility at $\delta_m = 0$ is

(A-14)
$$\frac{\partial V(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})}{\partial \delta_m} = A|V(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})| \frac{\kappa_m}{p_m} = A|V(\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})|(\mu_m + 1).$$

Similar to equation (A-10), equation (A-14) shows that the marginal utility from investing δ_m in Stock m at $\delta_m = 0$ only depends on its expected return μ_m, but not on any of its higher moments. Equations (A-14) and inequality (26) then imply that the next stock the less wealthy investor is going to add when his wealth increases enough beyond W_p will be Stock i, the stock with the next highest expected return and thus the highest marginal utility among the remaining stocks.

For investing δ_j in Stock j ($j = 1, 2, ..., n + 1$) at W_p to be optimal, the marginal utility from each of the first $i - 1$ stocks must be the same and must also be greater than the marginal utility from investing any positive amount in the rest of the stocks. By equations (A-13) and (A-14), these optimality conditions then imply that

$$\text{(A-15)} \qquad \frac{\kappa_j}{\bar{p}_j + A\delta_j \varphi_j^2 / \kappa_j} = k + 1, j = 1, 2, ..., i - 1,$$

for some $k \in [\mu_i, \mu_{i-1})$ that is to be determined later.[21]

Given the investment of δ_j in Stock j, a similar argument to that for equation (A-11)

[21] $k < \mu_{i-1}$ is because equation (A-15) shows that at $\delta_{i-1} = 0$, $k = \mu_{i-1}$, and the left-hand side of (A-15) is decreasing in δ_j as implied by (A-15) and (A-16).

implies that the new equilibrium price of Stock j is

$$\bar{p}_j = \frac{\kappa_j^2 + A\lambda\delta_j\varphi_j^2}{\kappa_j + A\bar{\omega}_j\varphi_j^2}, \quad j = 1, 2, ..., i-1. \tag{A-16}$$

Solving equations (A-16) and (A-15) for $\bar{p}_j$ and δ_j and simplifying, I have that the new equilibrium prices are

$$\bar{p}_j = p_j \frac{\lambda/(k+1)+1}{\lambda/(\mu_j+1)+1}, \quad j = 1, 2, ..., i-1, \tag{A-17}$$

which yields respectively the new expected returns and volatilities:

$$\bar{\mu}_j = \nu_k k + (1-\nu_k)\mu_j, \quad j = 1, 2, ..., i-1, \tag{A-18}$$

and

$$\bar{\sigma}_j = \frac{\varphi_j}{\kappa_j}(\bar{\mu}_j + 1), \quad j = 1, 2, ..., i-1, \tag{A-19}$$

where

$$\nu_k = \frac{\lambda}{\lambda + k + 1} \in [0, 1].$$

Thus, the new equilibrium expected returns are weighted averages of the original expected returns (μ_j's) and k. Since the weight ν_k is the same across all the first $i-1$ stocks, equation (A-18), inequality (26), and $k \in [\mu_i, \mu_{i-1})$ imply that $\bar{\mu}_1 > \bar{\mu}_2 > ... > \bar{\mu}_{i-1} > k \geq \mu_i > ... > \mu_n > \mu_{n+1}$. So the less wealthy will indeed invest only in the first $i-1$ stocks at W_p in equilibrium.

Plugging equation (A-17) back into equation (A-15) and simplifying, I have that the equilibrium dollar amount invested in Stock j is

$$\delta_j = \frac{\alpha_j(\mu_j - k)}{A(k+1)(\lambda + \mu_j + 1)} = \frac{\bar{\mu}_j + 1}{k+1}\frac{\bar{\mu}_j - k}{A\bar{\sigma}_j^2}, \quad j = 1, 2, ..., i-1. \tag{A-20}$$

Without borrowing or shortselling, I must have the following budget constraint

$$W_p = \sum_{j=1}^{i-1} \delta_j = \sum_{j=1}^{i-1} \frac{\alpha_j(\mu_j - k)}{A(k+1)(\lambda + \mu_j + 1)}, \tag{A-21}$$

which yields that

$$k = \frac{\sum_{j=1}^{i-1} \frac{\alpha_j \mu_j}{\lambda + \mu_j + 1} - A\,W_p}{\sum_{j=1}^{i-1} \frac{\alpha_j}{\lambda + \mu_j + 1} + A\,W_p}. \tag{A-22}$$

Equations (A-22), (A-20), and (A-13) show that as the wealth increases, k decreases and for all $j \leq i-1$, the dollar amount δ_j invested in Stock j increases, and the marginal utility of investing in Stock j decreases. When the wealth reaches a threshold level at which the marginal utility of investing more in each of the first $i-1$ stocks is equal to the marginal utility of investing a small amount in Stock i, the investor adds Stock i to his portfolio. By equations (A-13), (A-14), and (A-15), this threshold wealth level $\hat{W}_i$ above which the less wealthy investor holds Stock i must be such that $k = \mu_i$, which combined with equation (A-21) implies that

$$\hat{W}_i = \sum_{j=1}^{i-1} \frac{\alpha_j(\mu_j - \mu_i)}{A(\mu_i + 1)(\lambda + \mu_j + 1)}. \tag{A-23}$$

By equation (A-22), I have $k \in [\mu_i, \mu_{i-1})$ if and only if $W_p \in (\hat{W}_{i-1}, \hat{W}_i]$. As i increases, μ_i decreases, so the threshold wealth level $\hat{W}_i$ increases. Because the above derivation applies to any $i = 2, 3, ..., n$, I have shown that for $2 \leq i \leq n$, the less wealthy holds only the stocks with the highest $i-1$ expected returns if and only if the initial wealth $W_p \in (\hat{W}_{i-1}, \hat{W}_i]$, equivalently if and only if $k \in [\mu_i, \mu_{i-1})$.

Therefore, as wealth increases, the less wealthy investors sequentially add stocks with the next highest expected returns. Eventually, the less wealthy will start to invest in the risk-free asset (which has the lowest expected return) when the less wealthy's wealth increases to a critical level $\hat{W}_{n+1}$. Beyond this critical level, because investing more in the risk-free asset does not increase risk, equations (A-13) and (A-14) imply that the marginal utility of investing more in the risk-free asset decreases less than that of investing more in any of the risky stocks. Therefore, the investor optimally invests any amount above $\hat{W}_{n+1}$ in the risk-free asset and no additional amount in any of the risky stocks (a standard result for CARA preferences). This implies that $\hat{W}_{n+1}$ is also the critical wealth level at which the less wealthy hold the same unconstrained optimal stock portfolio as the wealthy. Therefore, to get $\hat{W}_{n+1}$, I can simply set $k = \mu_{n+1} = 0$ and $i = n+1$ in equation (A-23). This shows that as W_p approaches $\hat{W}_{n+1}$, the less wealthy's portfolio converges to that of the wealthy.

Finally, as shown above, as long as $W_p \in (0, \hat{W}_{n+1})$, the less wealthy and the wealthy hold different portfolios and thus no one in the economy holds the market portfolio in equilibrium. The market portfolio's weight on Stock j is

$$w_j^M = \frac{\bar{\omega}_j \bar{p}_j}{\sum_{i=1}^n \bar{\omega}_i \bar{p}_i}, j = 1, 2, ..., n.$$

Direct computation using equations (A-17), (A-18), and (A-19) shows that CAPM does not hold, i.e.,

$$\bar{\mu}_j - r \neq \boldsymbol{\beta}_{jM}(\bar{\mu}_M - r),$$

where $r = 0$,

$$\bar{\mu}_M = \sum_{i=1}^n w_i^M \bar{\mu}_i,$$

$$\boldsymbol{\beta}_{jM} = \frac{w_j^M \bar{\sigma}_j^2}{\sum_{i=1}^n (w_i^M)^2 \bar{\sigma}_i^2}.$$

Thus idiosyncratic risks are priced. What holds is a modified CAPM equation with a nonzero alpha term. Specifically, I have for $j = 1, 2, ..., n$,

$$\tilde{r}_j - r = \boldsymbol{\alpha}_j + \boldsymbol{\beta}_j(\tilde{r}_M - r) + \tilde{\varepsilon}_j, \tag{A-24}$$

where $\tilde{r}_j$ is the Stock j's return, $\tilde{r}_M$ is the market portfolio return, $\tilde{\varepsilon}_j$ is the mean-zero error term, and

$$\boldsymbol{\alpha}_j \equiv (\bar{\mu}_j - r) - \boldsymbol{\beta}_{jM}(\bar{\mu}_M - r). \tag{A-25}$$

□

The following lemma will be used repeatedly in the proof of Theorem 3.

Lemma 1 *Given positive integer i, if $x_j > 0$, $b_j > 0$, $\hat{b}_j > 0$, and both x_j and $b_j/\hat{b}_j$ decrease with j for $j = 1, 2, ..., i$, then*

$$\frac{\sum_{j=1}^i b_j \, x_j}{\sum_{j=1}^i b_j} \geq \frac{\sum_{j=1}^i \hat{b}_j \, x_j}{\sum_{j=1}^i \hat{b}_j}. \tag{A-26}$$

PROOF OF LEMMA 1: Let $\pi_j = \frac{b_j}{\sum_{l=1}^i b_l}$ and $\hat{\pi}_j = \frac{\hat{b}_j}{\sum_{l=1}^i \hat{b}_l}$. Then the left- (right-) hand side of inequality (A-26) can be viewed as the mean of a random variable $\tilde{x}$ with support $\{x_1, x_2, ..., x_i\}$ and a probability of π_j ($\hat{\pi}_j$, respectively) for x_j ($j = 1, 2, ..., i$). Next I show that the assumption that both x_j and $b_j/\hat{b}_j$ decrease with j implies that the probability distribution for the left-hand side of equation (A-26) first stochastic dominates that for the right-hand side and thus the mean on the left-hand side is greater than that on the right-hand side, and accordingly, inequality (A-26) holds. Since $x_j > 0$ decreases with j, for the first stochastic dominance I only need to show that for any $1 \leq m \leq i-1$,

(A-27)
$$\sum_{j=1}^{m} \pi_j \geq \sum_{j=1}^{m} \hat{\pi}_j,$$

which is equivalent to

(A-28)
$$\frac{\sum_{j=1}^{m} b_j}{\sum_{j=1}^{m} \hat{b}_j} \geq \frac{\sum_{j=m+1}^{i} b_j}{\sum_{j=m+1}^{i} \hat{b}_j}.$$

Since $b_j/\hat{b}_j > 0$ decreases with j, I have

(A-29)
$$\frac{\sum_{j=1}^{m} b_j}{\sum_{j=1}^{m} \hat{b}_j} = \frac{\sum_{j=1}^{m} \hat{b}_j \frac{b_j}{\hat{b}_j}}{\sum_{j=1}^{m} \hat{b}_j} \geq \frac{\sum_{j=1}^{m} \hat{b}_j \frac{b_m}{\hat{b}_m}}{\sum_{j=1}^{m} \hat{b}_j} = \frac{b_m}{\hat{b}_m} = \frac{\sum_{j=m+1}^{i} \hat{b}_j \frac{b_m}{\hat{b}_m}}{\sum_{j=m+1}^{i} \hat{b}_j} \geq \frac{\sum_{j=m+1}^{i} \hat{b}_j \frac{b_j}{\hat{b}_j}}{\sum_{j=m+1}^{i} \hat{b}_j} = \frac{\sum_{j=m+1}^{i} b_j}{\sum_{j=m+1}^{i} \hat{b}_j}.$$

Therefore inequality (A-28) indeed holds and thus inequality (A-26) also holds. □

Intuitively, Lemma 1 holds because b_j assigns higher weights to larger values of $\tilde{x}$ than $\hat{b}_j$. We are now ready to prove Theorem 3.

Proof of Theorem 3: Since as shown in equation (A-22), k decreases as W_p increases, I can equivalently show how the properties of the less wealthy's portfolio change with k. The mth central moment of the less wealthy's stock portfolio gross return is

(A-30)
$$\xi_m(k) = \sum_{j=1}^{i} w_j^m \frac{M_{mj}}{\bar{p}_j^m},$$

where the portfolio weight

$$w_j = \frac{\bar{\delta}_j}{\sum_{j=1}^{i} \bar{\delta}_j}, \quad j = 1, 2, ..., i,$$

and M_{mj} is the mth central moment of the payoff of the jth stock. Using equations (A-17) and (A-20), I have

$$\xi_m(k) = \frac{\sum_{j=1}^{i} C_{mj}(\mu_j - k)^m}{((\lambda/(k+1)+1)\sum_{j=1}^{i} a_j(\mu_j - k))^m}, \tag{A-31}$$

where $C_{mj} \equiv M_{mj}/\beta_j^m$ and $a_j \equiv \alpha_j/(\lambda + \mu_j + 1)$. Computing $\xi'_m(k)$ and rearranging yield that $\xi_m(k)$ is strictly increasing in k for all $\lambda > 0$ if and only if

$$\frac{\sum_{j=1}^{i} b_{mj}(\mu_j + 1)}{\sum_{j=1}^{i} b_{mj}} \geq \frac{\sum_{j=1}^{i} a_j(\mu_j + 1)}{\sum_{j=1}^{i} a_j}, \tag{A-32}$$

where

$$b_{mj} = C_{mj}(\mu_j - k)^{m-1}.$$

By Lemma 1, for inequality (A-32) to hold, I only need to show b_{mj}/a_j decreases with j. First, consider the expected return. Since $M_{1j} = \alpha_j\beta_j$, I have $b_{1j} = \alpha_j$. So

$$\frac{b_{1j}}{a_j} = \lambda + \mu_j + 1,$$

which indeed decreases with j by inequality (26) and thus the expected return decreases as k decreases. Next, consider the return volatility. Since $M_{2j} = \alpha_j\beta_j^2$, I have $b_{2j} = \alpha_j(\mu_j - k)$ and

$$\frac{b_{2j}}{a_j} = (\mu_j - k)(\lambda + \mu_j + 1),$$

which also decreases with j by inequality (26) and thus the volatility also decreases as k decreases.

The skewness of the stock portfolio is equal to

$$s(k) \equiv \frac{\xi_3(k)}{\xi_2(k)^{3/2}} = \frac{\sum_{j=1}^{i} C_{3j}(\mu_j - k)^3}{\left(\sum_{j=1}^{i} C_{2j}(\mu_j - k)^2\right)^{3/2}}.$$

It is easy to verify that $C_{3j} = 2\alpha_j$, $C_{2j} = \alpha_j$, and the skewness is increasing in k if and only if

$$\text{(A-33)} \qquad \frac{\sum_{j=1}^{i} b_j(\mu_j - k)}{\sum_{j=1}^{i} b_j} \geq \frac{\sum_{j=1}^{i} \hat{b}_j(\mu_j - k)}{\sum_{j=1}^{i} \hat{b}_j},$$

where

$$b_j = \alpha_j(\mu_j - k)^2, \quad \hat{b}_j = \alpha_j(\mu_j - k).$$

Since $b_j/\hat{b}_j = \mu_j - k$, which decreases with j, by Lemma 1, I have inequality (A-33) holds and therefore the skewness of the stock portfolio also decreases as k decreases.

For $W_p > \hat{W}_2$, the less wealthy investor holds at least two stocks, i.e., $i \geq 2$. The Sharpe ratio of the stock portfolio at $\lambda = 0$ is

$$SR(k) \equiv \frac{\xi_1(k) - 1}{\sqrt{\xi_2(k)}} = \frac{\sum_{j=1}^{i}(C_{1j} - a_j)(\mu_j - k)}{\sqrt{\sum_{j=1}^{i} C_{2j}(\mu_j - k)^2}}.$$

Computing $SR'(k)$ shows that the Sharpe ratio strictly *decreases* in k if and only if

$$\text{(A-34)} \qquad \frac{\sum_{j=1}^{i} d_j(\mu_j - k)}{\sum_{j=1}^{i} d_j} > \frac{\sum_{j=1}^{i} \hat{d}_j(\mu_j - k)}{\sum_{j=1}^{i} \hat{d}_j},$$

where

$$d_j = \alpha_j(\mu_j - k), \quad \hat{d}_j = \alpha_j \frac{\mu_j}{\mu_j + 1}.$$

Since $d_j/\hat{d}_j = \mu_j - k + 1 - k/\mu_j$, which strictly decreases with j, by Lemma 1, I have that inequality (A-34) holds and therefore by continuity the Sharpe ratio of the stock portfolio increases as k decreases when λ is small. □