The Effect of Capital Wealth on Optimal Diversification: Evidence from the Survey of Consumer Finances

Abstract

It is well known that the wealthier the household, the larger tends to be the proportion of its total capital portfolio allocated to publicly traded stock, and the larger tends to be the number of individual stock issues included in its portfolio. Using the "homogeneous securities" case of a mean-variance model originally proposed by Michael Brennan, explicit functional forms are obtained for both the optimal proportion of the portfolio allocated to stocks and the optimal number of individual stock issues in the portfolio. An empirical evaluation of these theoretical results, using a dataset derived from the 2004 Survey of Consumer Finances, lends substantial support to the model.

Journal of Economic Literature Classifications:
G11 (Portfolio Choice, Investment Decisions); D31 (Personal Income, Wealth, and Their Distributions)

Keywords:
capital wealth, portfolio choice, investment analysis, stock ownership, risk, distribution, Survey of Consumer Finances

Acknowledgement: The authors express their appreciation to Michael Brennan, and to referees for the *Quarterly Review of Economics and Finance*, for helpful comments on earlier drafts.

Contact Author:
James A. Yunker
Professor of Economics
Western Illinois University
Macomb, IL 61455
JA-Yunker@wiu.edu

Co-author:
Alla A. Melkumian
Assistant Professor of Economics
Western Illinois University
Macomb, Illinois 61455
AA-Melkumian@wiu.edu

The Effect of Capital Wealth on Optimal Diversification: Evidence from the Survey of Consumer Finances

1. Introduction

Relative to such fixed-return assets as corporate bonds, government bonds, Treasury notes and the like, publicly traded common stock instruments combine high expected rate of return with high variance of return. From the point of view of the private investor, the upside of common stock is high expected return, while the downside is high risk, as reflected in the high variance. According to the "gospel of diversification" preached by the large majority of investment counselors, an investor interested in common stock should hold a substantial number of different stocks as an offset to the high variance of individual stock returns. As expressed in the ancient wisdom of not putting all one's eggs in one basket, this strategy takes advantage of the so-called "law of averages," otherwise known as the "law of large numbers." From such classic studies as Markowitz (1952, 1959), Baumol (1962), Sharpe (1964), Lintner (1965) and Samuelson (1967), down to the practitioner literature of today, the diversification prescription is central.

It is generally believed that wealthier investors tend to hold a larger proportion of their assets in common stocks than do less wealthy investors, and that they diversify their stock portfolios to a greater extent. For a comprehensive survey of recognized empirical regularities on the issue, see Carroll (2000). Since the law of averages suggests that every investor in common stock, whatever his or her total wealth level, should hold a large number of different stocks, the fact that smaller investors tend to hold a relatively limited number of stocks suggests that they are influenced, to a larger extent than wealthier investors, by the transactions costs involved in purchasing different stocks. In a word, wealthier investors may be able to better afford the higher transactions costs involved in holding highly diversified portfolios.

The intention of this research is to develop a theoretical model of optimal diversification that generates hypotheses capable of being tested using existing, reliable data. Although a number of studies in the large diversification literature examine relationships between wealth and portfolio choice issues, to the authors' knowledge no prior study has developed mathematically explicit relationships, based on a relatively simple mean-variance model, between total capital wealth and both the proportion of total capital wealth devoted to stock and the number of individual stocks held by the investor, that are capable of being empirically tested using well known survey data.

We utilize the "homogeneous securities" case of a mean-variance model originally proposed by Michael Brennan (1975). Although Brennan did not himself develop explicit mathematical forms for the optimal proportion of the total portfolio to be invested in stocks, and the optimal number of stock issues to hold in the portfolio, it is straightforward to derive these formulae from the first-order maximization conditions of the model. These explicit theoretical predictions are tested using data from the 2004 Survey of Consumer Finances. The empirical analysis does in fact lend considerable support to the predictions.

The remainder of this article is organized as follows. Section 2 contains a brief survey of related literature. Section 3 sets forth the model. Section 4 describes the 2004 Survey of Consumer Finances and enumerates the variables utilized in this research. Section 5 presents the empirical results. Section 6 briefly reviews some of the more important caveats and qualifications. Section 7 concludes.

2. Related Literature

The cumulative literature on diversification and related issues has become very large. As of 2008, the EconLit database contained over 8,000 records with "diversification" as a keyword, about 3,000 records with "diversification" in the abstract, and over 1,600 records with "diversification" in the title. Contributions range over the spectrum from sophisticated mathematical exercises, exemplified by Carlos-Hatchondo (2008) and Bera and Park (2008), to commonsensical prescriptions from the practitioner literature, exemplified by Sparling (2008), Jaworski (2008), and Domian et al (2007). A substantial subset of the overall literature concerns factors that influence household investment patterns. Some illustrative examples follow.

Using Italian household portfolio data and times series data on financial assets and housing stock returns, Pelizzon and Weber (2008) determine that housing wealth plays a key role in determining whether or not portfolios chosen by households are efficient. Using Finnish data, Saarimaa (2008) shows that mortgage-encumbered investors hold a smaller share of stocks in a mean-variance efficient portfolio. Also using Finnish data, Kaustia and Knupfer (2008) find a strong positive link at the individual investor level between past IPO returns and future subscriptions, which is consistent with reinforcement learning, wherein personally experienced outcomes are over-weighted compared to rational Bayesian learning. Goetzmann and Kumar (2008) present evidence that U.S. individual investors hold under-diversified portfolios, with the degree of under-diversification greater among younger, low-income, less-educated and less-sophisticated investors. Berkowitz and Qui (2006) investigate how health status affects portfolio choice, and find that the diagnosis of a new disease causes a larger decrease in financial wealth than in non-financial wealth. Using data from Bank of Italy Surveys of Household Income and Wealth, Guiso and Jappelli (2005) document inadequate information about investment opportunities among investors, and find that the probability of information is positively correlated with education, household resources, long-term bank relations, and proxies for social interaction. Jianakoplas and Bernasek (2006) decompose the effects of chronological age, birth cohort, and calendar year on the age profile of household financial risk-taking, and find that the results support the conventional wisdom that risk-taking decreases with age.

Lundtofte (2006) shows that while investors without inside information will consistently invest less in stocks with higher variance, the relationship between stock variance and amount invested is ambiguous for investors with inside information. Massa and Simonov (2006), using a Swedish data set, show that investors tend to over-invest in stocks issued by companies that are sources of their personal wage and salary income. Garlappi and Huang (2006) show how the "pecking order" location rule emphasized in the retirement planning literature may become invalid when the household faces certain portfolio constraints. Pachamanova (2006) applies a recent mathematical technique known as "robust optimization" to determine the optimal portfolio under uncertainty regarding the means and variances of returns. Kole, Koedijk and Verbeek (2006) propose a novel approach incorporating the possibility of systemic crises to determine the optimal portfolio of investors in international equity markets. Cocco (2005) demonstrates a "crowding out" effect of housing investment: households owning a home tend to hold less stock, other things being equal. Berkelaar, Kouwenberg, and Post (2004) show that it is not possible to empirically disentangle the effects of loss aversion from those of theoretically distinct risk aversion in the determination of the household's optimal portfolio. Gomes and Michaelides (2003) show that the introduction of internal habit formation preferences into a life-cycle model of consumption and portfolio choice is not able to simultaneously explain two important stylized facts: a low stock market participation rate, and moderate equity holdings for those households that do invest in stocks.

Information on household wealth and portfolio choice has been used to estimate measures of household risk aversion, as in Vissing-Jorgensen and Attanasio (2003), and Brunnermeier and Nagel (2008). The effect of taxation on household portfolio choice has been investigated by King and Leape (1998), and Poterba and Samwick (2003). Using a bivariate binary-choice model and data from the Dutch Savings Survey 1993-1998, Alessie et al (2004) examine the ownership dynamics of stocks and mutual funds, and find that the negative relation between ownership of one type in one period and the other type in the next period is explained by correlated unobserved heterogeneity. Carroll (2000) documents that portfolios of wealthy investors are heavily skewed toward risky assets, particularly investments in their own privately held businesses.

Several researchers have applied the same data source utilized here, the Survey of Consumer Finances, to investigate issues in portfolio choice. Kelly (1995) adduces evidence suggesting that most U.S. household portfolios are inadequately diversified in terms of mean-variance efficiency. Poterba and Samwick (2003) find that the portfolio share invested in corporate stock, which is taxed less heavily than interest bearing assets, is increasing in the household's ordinary income tax rate. Hu (2004) finds evidence that homeowners facing more non-diversified and levered risks in housing invest their liquid assets more conservatively than those who have relatively less housing commitments. Bergstresser and Poterba (2004) examine household allocation patterns between taxable and tax-deferred accounts. Polkovnichenko (2005) shows that portfolio choice models with rank-dependent preferences are capable of explaining certain stylized facts inconsistent with expected utility maximization. Gutter and Saleem (2005) find that financial vulnerability, defined by the extent to which income and wealth are derived from the same source, is prevalent among small business owners, especially farmers.

The extent and complexity of the published research on portfolio choice and diversification suggests the impracticality of a major literature survey that exhaustively documents similarities and differences between the present research and related research. As mentioned, empirical evidence from the Survey of Consumer Finances provides substantial support for the theoretical hypotheses, and this fact indirectly supports the sensibility of the model. Some potential problems with the research, however, will be enumerated and briefly discussed in the penultimate section of the paper.

3. Optimal Diversification with Homogeneous Securities

Let the capital wealth of an individual be denoted assets a. The proportion of a held in stocks is ρ, and the proportion held in bonds is $1 - \rho$. The wealth constraint is:

$$s + b = \rho a + (1 - \rho)a = a \tag{1}$$

where s and b denote the holdings of stocks and bonds respectively. To simplify the analysis the variance of bond returns is set to zero. The rate of return on bonds, r_b, is then the risk-free rate of interest.

Using the capital asset pricing model, the rate of return on stock issue *i*, denoted $r_{s,i}$, is:

$$r_{s,i} = r_b + \beta_i (r_m - r_b) + \varepsilon_{s,i} \tag{2}$$

where β_i is the beta coefficient of stock issue *i*, r_m is the value-weighted market return, and $\varepsilon_{s,i}$ is the residual disturbance for stock issue *i*, with expected value zero and variance $\sigma_{s,i}^2$. Under the homogeneous securities assumption, $\beta_i = 1$ for all securities, and the mean and variance of the residual disturbance $\varepsilon_{s,i}$ for each stock issue are the same: $E(\varepsilon_{s,i}) = E(\varepsilon_s) = 0$ and $\sigma_{s,i}^2 = \sigma_s^2$. It is also assumed that the idiosyncratic components associated with different assets are orthogonal, and that the idiosyncratic components are orthogonal to market returns. Thus for all *i*, the random variable $r_{s,i} = r_m + \varepsilon_s$ has expected value $\overline{r_s} = \overline{r_m}$ and variance $\sigma_m^2 + \sigma_s^2$. Since *s* is a high return asset, we assume $\overline{r_s} > r_b$.

The first diversification decision variable of the capital owner is ρ, the proportion of total capital assets to be held in the form of stock issues. The capital owner then sub-divides the stock portfolio equally over *n* different stocks. The number of stocks held is the second decision variable. For analytical purposes, it will be taken to be a continuous variable, although in practice, of course, it must be a discrete variable taking only integer values. The model specifies that an equal amount is held in each stock. This is not efficient according to most portfolio choice models, but it is apparently descriptive of real-world practices among many investors, as discussed by a number of authors including Benartzi and Thaler (2001), Stevenson (2001), Windcliff and Boyle (2004) and McClatchey and Vandenhul (2005). This specification is consistent with the model assumption that all stocks are alike.

The cost of transacting in each security is assumed to be a fixed amount *c*. Then the rate of return on investment, net of the fixed costs of transacting, is:

$$r = \sum_{i=1}^{n} r_{s,i} \frac{\rho}{n} + r_b(1-\rho) - n(c/a)(1+r_b) \tag{3}$$

The expected value and variance of rate of return are:

$$E(r) = \overline{r}_s \rho + r_b(1-\rho) - n(c/a)(1+r_b) \tag{4}$$

$$V(r) = \rho^2 \left(\sigma_m^2 + \frac{\sigma_s^2}{n} \right) \tag{5}$$

Using a conventional mean-variance formulation, the criterion function to be maximized is:

$$L = \mathrm{E}(r) - \lambda \mathrm{V}(r) = \overline{r}_s \rho + r_b(1-\rho) - n(c/a)(1+r_b) - \lambda\left(\rho^2(\sigma_m^2 + (\sigma_s^2/n))\right) \tag{6}$$

where λ represents the marginal rate of transformation of risk for return, a standard measure of risk aversion. The derivation of the optimal values of ρ and n (denoted respectively ρ* and n*) is shown in the appendix. The appendix also obtains restrictions on the model parameters to enable economically sensible solutions. Finally, the appendix demonstrates that with this problem specification, the second-order conditions for a maximum are satisfied.

As appendix equations (A.5) and (A.6) are explicit formulae for optimal ρ* and n*, the comparative statics effects of the parameters on optimal diversification may be ascertained directly by inspection. Our particular interest is in the effect of total capital wealth a on optimal diversification. Isolating the a parameter, we have:

$$\rho^* = \phi_1 - \phi_2 \frac{1}{\sqrt{a}} \text{ where } \phi_1 = \frac{.5(\overline{r}_s - r_b)}{\lambda\sigma_m^2};\ \phi_2 = \frac{\sqrt{\sigma_s^2 c(1+r_b)}}{\sigma_m^2\sqrt{\lambda}};\text{ and} \tag{7}$$

$$n^* = \psi_1\sqrt{a} - \psi_2 \text{ where } \psi_1 = \frac{.5\sigma_s(\overline{r}_s - r_b)}{\sigma_m^2\sqrt{\lambda c(1+r_b)}};\ \psi_2 = \frac{\sigma_s^2}{\sigma_m^2}; \tag{8}$$

where the ϕ and ψ parameters are all positive. It is apparent that both ρ* and n* are concave increasing functions of total capital wealth a. Furthermore, whereas n* increases indefinitely with a, there is an asymptotic upper limit on ρ* at $\phi_1 = .5(\overline{r}_s - r_b)/\lambda\sigma_m^2$.

The marginal participation level of wealth (a^o) can be defined, for both stock proportion and number of stocks, as the wealth level below which respectively ρ* and n* are less than zero. By setting (7) equal to zero and solving for a, and then setting (8) equal to zero and solving for a, it may be determined that the marginal participation level of wealth is the same for both ρ and n:

$$a^o = \frac{\phi_2^2}{\phi_1^2} = \frac{\psi_2^2}{\psi_1^2} = \frac{\lambda c(1+r_b)}{.25(\overline{r}_s - r_b)^2}\sigma_s^2 \tag{9}$$

The other parameters have the intuitively expected effects. For optimal ρ*: (1) $d\rho^*/d\overline{r}_s > 0$; (2) $d\rho^*/d\sigma_s^2 < 0$ and $d\rho^*/d\sigma_m^2 < 0$; (3) $d\rho^*/dr_b < 0$; (4) $d\rho^*/d\lambda < 0$; and (5) $d\rho^*/dc < 0$. These are intuitively expected results because, respectively: (1) a higher expected rate of return on stocks makes them more attractive; (2) higher variance on stock return makes stocks riskier and hence less attractive; (3) a higher rate of return on bonds makes stocks relatively less attractive; (4) a higher level of risk aversion on the part of the capital owner, as reflected in a larger value of λ, makes risky stocks less attractive; and (5) higher transactions costs on stocks makes them less attractive. Because ρ* and n* are proportional to one another, the signs of the comparative statics derivatives for n* are the same as those for ρ*.

4. Survey of Consumer Finances Dataset

Although it is "common knowledge" that wealthier investors keep a larger proportion of their capital wealth in the form of common stock, formal statistical evidence of this fact is not overly abundant. Perhaps the most definitive early piece of evidence to this effect is reported in Table A 10 ("Composition of Portfolio of Liquid and Investment Assets, December 31, 1962") in the 1966 Projector-Weiss report on the Survey of Financial Characteristics of Consumers (SFCC) sponsored by the Federal Reserve Board. On the fifth page of this eight-page table (on p. 118 in the report), there is a size distribution by "size of portfolio" containing "mean investment assets" and "mean assets of publicly traded stock." For the low-wealth bracket of $500-$1,000, stock assets are 26.90 percent of total investment assets; for the medium wealth bracket of $50,000-$99,999, stock assets are 52.98 percent of total investment assets; while for the highest wealth bracket of $500,000 and over, stock assets are 71.30 percent of total investment assets.

There are only nine wealth brackets in the 1966 SFCC report. This high level of aggregation is repeated in other published sources of empirical information on capital wealth distribution. For example, the various articles documenting increasing financial inequality in the United States by Edward Wolff (1987, 1992, 1994) present size distributions containing five quintiles. The statistical results reported below are based on the entire dataset of 4,519 households contained in the 2004 Survey of Consumer Finances (SCF), and also on two aggregated datasets: the first consisting of 100 brackets each containing 45 households, and the second consisting of 25 brackets each containing 180 households. Descriptive statistics on the second of these are shown in Table 2 below. To the authors' knowledge, comparable information to that contained in this table has not previously appeared in a published source.

The Survey of Consumer Finances (SCF), described by the Federal Reserve Board as "a triennial survey of the balance sheet, pension, income, and other demographic characteristics of U.S. families," had its origins in the above-mentioned 1962 Survey of Financial Characteristics of Consumers and the 1963 Survey of Changes in Family Finances. The current triennial pattern was commenced in 1983. The SCF is remarkably comprehensive. The 2004 survey contains 2,834 data items (variables), and the full public dataset contains 4,519 households. Data obtained from the Survey of Consumer Finances have been utilized in numerous published studies on a wide variety of topics. Apart from those mentioned above, a few illustrative examples include Castronova and Hagstrom (2004) on the demand for and usage of credit cards, Ben-Gad (2004) on the welfare effects of the Reagan era deficits, Baek and Hong (2004) on the determinants of consumer indebtedness, Aizcorbe et al (2004) on household vehicle acquisition patterns, and Wu (2005) on the determinants of household saving behavior.

Table 1 lists the variables taken from the 2004 SCF dataset, as well as all constructed variables utilized in the research. The SCF variable X3914 is used directly as n (the number of stocks in the portfolio). Total capital wealth, the sum of variables 1 (X3721) through 11 (X3915), is the empirical analogue of wealth assets a. The proportion of capital wealth assets allocated to stock, ρ, is value of stock funds (X3822) plus value of stocks (X3915), divided by total capital wealth. Variable X3913 is a binary variable indicating whether or not the household owns some publicly traded stock. This variable is not used anywhere in the statistical analysis, but descriptive information on it is provided in Table 2, by way of general interest.

Variables 14 (X14) through 21 (X5901) are potential control variables for refining the estimated relationships between stock proportion and the capital wealth variable, and between number of stocks and the capital wealth variable. Certain of these variables are coded in a way inconsistent with the standard binary variable. For example, the variable X301 (expectations concerning the performance of the U.S. economy over the next five years relative to the last five years) are coded 1 for “better,” 2 for “worse” and 3 for “about the same.” These were re-coded to 1 for “better” and 0 for “otherwise.” This adjustment was also made for the two other analogous variables: X302 (expectations concerning interest rates) and X304 (expectations concerning household income).

The researchers’ expectation was that the large amount of random variation typically to be found in survey data would result in very low explanatory power of regressions of stock proportion and number of stocks on capital wealth, when the regressions are based on the entire dataset of 4,519 households. Not only is there considerable inaccuracy in responses, unintentional or otherwise, there will also be considerable unmeasured variation over households in the parameters of the model. For example, the parameter λ (marginal rate of substitution between risk and return), the indicator of household risk aversion, no doubt varies considerably over households, and there is no attempt to measure risk aversion in the SCF. The researchers’ expectation in this regard was indeed fulfilled. In order to cope with the random variation problem, the full-set regressions were supplemented by regressions using two additional datasets composed of aggregated data. The entire dataset was sorted in descending order on wealth and two aggregated datasets of smaller size were then computed. Although uncommon, a certain amount of precedent exists for this technique: Listokin (2008), and Gong and Lofgren (2007).

First, a dataset of 100 observations was constructed from the sorted data consisting of the mean values of the variables over 100 brackets, each containing 45 households. This method deletes the last 19 observations from the dataset, but this represents very little data loss from the full set of 4,519 observations. Second, a dataset of 25 observations was constructed from the sorted data consisting of the mean values of the variables over 25 brackets, each containing 180 households. Again, this loses data from the last 19 observations. As the statistical results shown below manifest, by suppressing random variation within brackets, the relationships between the variables of primary interest to this research become much stronger.

Sorting the full dataset of 4,519 households on capital wealth reveals that 2,424 households report positive capital wealth; the other 2,076 households, approximately 46 percent of the total, report zero capital wealth. The 100-bracket dataset shows the first 54 brackets having positive mean capital wealth, while the 25-bracket dataset shows the first 14 brackets having positive mean capital wealth. Table 2 shows bracket means for the 25-bracket dataset for the capital wealth-related variables. For bracket 1 (the wealthiest bracket), the top line of data shows that for the 180 households in this bracket, mean capital wealth is $44,080,431, the proportion of households reporting ownership of stock securities is .9444, the mean number of stock securities owned by the household is 48.27, mean stock wealth is $30,124,101, and mean stock wealth as a proportion of mean capital wealth is .6834.

5. Empirical Results

Positive heteroskedasticity in the data was found using the widely applied formal test of White (1980), and also informally through a regression of squared residuals on the independent variables of main interest in this research, namely the capital wealth variables: $1/\sqrt{a}$ for the ρ equation, and $\sqrt{a}$ for the n equation. The equations were therefore estimated using White's heteroskedasticity consistent covariance matrix estimator which provides corrected estimates of the coefficient covariances in the presence of heteroskedasticity of unknown form. The point estimates of the equation coefficients are the same as those obtained using ordinary least squares, but the standard errors of the estimated coefficients are larger for the capital wealth variables.

Hence the t-statistics corresponding to these variables are substantially lower in absolute value than their ordinary least squares counterparts. Nevertheless, the heteroskedasticity adjusted t-statistics of the capital wealth variables of primary interest are still sufficiently high to indicate significance at a high level of confidence.

Table 3 is based on the full SCF dataset of 4,519 households. Regression results for ρ (proportion of stock in portfolio) are on the left; those for *n* (number of stocks in portfolio) are on the right. In both cases there is a "sparse" formulation which omits the eight control variables and an "augmented" formulation which includes them. Note that for the ρ equation the number of observations is 2,424: the number of households with positive capital wealth. For the remaining households with zero capital wealth, both the ρ dependent variable and the $1/\sqrt{a}$ independent variable are undefined because of division by zero. This problem does not apply to the *n* equations, therefore they are based on all 4,519 observations.

Looking first at the ρ equations, the t-statistic on the $1/\sqrt{a}$ independent variable is –12.04 for the sparse formulation and –10.56 for the augmented formulation, both of which indicate the statistical significance of this variable at higher than the 99 percent confidence level. As for the control variables in the augmented formulation, some are significant and some are not, but as these variables are not of special concern to this research, interpretation of these results is left to the interested reader. The overall R-squared goodness-of-fit statistic, despite the high t-statistic of the $1/\sqrt{a}$ independent variable, is rather disappointing: 0.09 for the sparse formulation, rising only to 0.13 for the augmented formulation. Results for the *n* equation are basically analogous, except that the t-statistics on the $\sqrt{a}$ independent variable are quite high, and as a result the R-squared statistics are reasonably high (for cross-section data).

There is one important inconsistency between the econometric results and the theoretical model. According to equation (8): $n^* = \psi_1\sqrt{a} - \psi_2$. Thus a regression of *n* on $\sqrt{a}$ should show a positive slope coefficient and a negative intercept coefficient. But of the four estimated *n* equations shown in Tables 3 and 4, only the augmented equation for the full dataset (in Table 3) shows a negative estimated intercept. A possible explanation of this anomaly is if the control variables are left out of the estimation, the intercept estimate may be biased upwards.

A certain amount of experimentation was undertaken using parametric restrictions in the estimation of all the *n* equations other than the augmented formulation shown in the last column of Table 3. If the intercept is required to be less than or equal to zero, its estimate becomes zero, which suggests that the statistical best fitting methodology definitely "wants" the estimated intercept to be positive. According to theory, $\psi_2 = \sigma_s^2 / \sigma_m^2$: that is, the intercept of the relationship between number of stocks *n* and the square root of capital wealth $\sqrt{a}$ is the negative of the ratio of variance of an individual stock to the variance of the market average. Given the simple homogeneous securities model utilized here, this ratio should be unity. If the sparse formulation of the *n* equation shown in Table 3 is re-estimated using the restriction that the intercept is –1, the estimated coefficient of $\sqrt{a}$ changes from 0.007556 to 0.008148. There does not seem to be a consensus estimate of the ratio of individual stock variance to market average variance, although many investigators would likely lean toward this ratio factually being in excess of unity. The *n* equation was re-estimated for restrictions on the intercept ranging from –2 to –6. It was observed that the numerical value of the estimated regression coefficient of $\sqrt{a}$ rose slightly with the increasing absolute value of the required negative intercept, while its corresponding t-statistic also rose slightly, and at the same time the R-squared declined substantially. If the absolute value of the required negative intercept is set too high (e.g., –8), the computed R-squared becomes negative, suggesting that the theoretical basis of the estimation has been compromised.

Since in this case we really do not have any reliable prior information on the numerical value of $\psi_2 = \sigma_s^2 / \sigma_m^2$, and since there is also no guarantee that the theoretical methodology underlying the research is valid, it was decided to let the estimated result stand in reporting the research. Although this anomaly must be considered a weakness, we would emphasize that it pertains only to the intercept estimate of the sparse formulations of the equation for optimal number of stocks. With respect to estimates of the equation for optimal stock proportion, both the sparse and the augmented formulations produce statistically significant estimates that have the theoretically expected signs for both intercept and slope.

As expected, owing to the large amount of random variation in data obtained from a survey, the regression equations shown in Table 3 do not have a great deal of explanatory power, even for the *n* equation. Therefore regressions were also run on the aggregated datasets described above; results are presented in Table 4. The Table 4 regressions all pertain to sparse formulations that omit the eight control variables. One reason for this is to put less "strain" on the much smaller number of observations. Also there may be problems in interpreting the estimated regression coefficients of the control variables because the entire dataset was sorted on total capital wealth. Unless there are very strong correlations between total capital wealth and the various control variables, the within-bracket means of the control variables may be unrepresentative. Finally, from the results in Table 3, the control variables apparently do not have a substantive impact on the relationships of principal interest here: those between total capital wealth and stock proportion, and between total capital wealth and number of stock issues.

The left side of Table 4 pertains to dependent variable ρ and the right side to dependent variable *n*. For each dependent variable, results are shown for the 100-bracket dataset and the 25-bracket dataset. The ρ regressions are based on 54 observations from the 100-bracket dataset, and 14 observations from the 25-bracket dataset, because the remaining observations in these datasets are undefined in ρ and $1/\sqrt{a}$ (the observations for which *a* is zero). The *n* regressions are based on all observations: 100 from the 100-bracket dataset, and 25 from the 25-bracket dataset, since in this case all variables are defined for all observations. As expected, the explanatory power of the regressions are greater for the smaller datasets owing to the suppression of random variation within brackets. For example, the ρ regression equation has an R-squared of 0.80 for the 25-bracket dataset and 0.69 for the 100-bracket data, relative to 0.09 for the full dataset (sparse formulation).

Figures 1 and 2, based on the 100-bracket aggregated SCF dataset, are provided to illustrate visually the relatively good fit of the estimated equations to the SCF data. In these figures the horizontal axis represents not total capital wealth but rather the log of total capital wealth. If total capital wealth were used, the observations would be compressed too close to the left-hand vertical axis for the graph to be readable. Figure 1 pertains to ρ (proportion of stock in portfolio) and Figure 2 to *n* (number of stocks in portfolio). In both cases, a curve representing the estimated values of the dependent variables (respectively ρ and *n*), derived from the estimated equations, is superimposed over a scatter diagram of the actual values of these variables. Considering that these figures are based on notoriously variable survey data, the fits are fairly respectable.

6. Caveats and Qualifications

There are several legitimate questions that might be raised about the optimal diversification model utilized in this research, of which the following are examples. In the model, the parameters of the mean-variance function are constant over all households, including the degree of risk aversion as manifested in the λ parameter. But there have been some indications in the literature that risk aversion is not invariant, but rather varies according to the investor's age, health, personal attitudes, home ownership status, and other factors. In the model, transactions costs are treated as proportional to the number of stocks in the investor's portfolio at a certain point in time. But it is well known that some investors, particularly younger investors, tend to do a relatively large amount of trading over time, so that their incurred transactions costs, for the same number of stocks held at any one point in time, would be much higher than those of investors following the normally recommended “buy and hold” strategy. The model does not explicitly differentiate ordinary stock shares from mutual fund stock shares, but it could be argued that one mutual fund share represents as many stocks as are held by the fund. The model does not incorporate taxation, but some studies have shown that the tax status of the household affects its portfolio decisions. The model is based on the “homogeneous securities” concept, but it is well known that many if not most investors perceive major differences between any two stock securities. The model involves static optimization of a mean-variance utility function at a point in time, while some contributions to the portfolio choice literature specify dynamic optimization over a period of time using the concepts and methods of control theory. The model utilizes rate of return in the mean-variance criterion function rather than final wealth, and although in principle they amount to the same (final wealth being initial wealth multiplied by 1 plus rate of return), the fact remains that final wealth is more commonly utilized than rate of return in mean-variance criterion functions. Finally, not surprisingly given the long history of the mean-variance approach itself, some serious questions have in fact been raised against the fundamental sensibility of this approach.

In response to these kinds of problems, we can only point to the fact that a model that attempted to incorporate all or most of the numerous insights into the portfolio choice problem that have emerged from the very extensive literature in this area, would become impossibly complicated. A certain amount of simplification is essential if meaningful, tangible results are desired. We would also point out that this research goes beyond model specification and solution, in that the theoretical results are subjected to empirical testing using conventional econometric techniques. It is quite important that the empirical testing is generally supportive of the theoretical results. This is indirect evidence of the validity of the model as a simplified representation of reality.

The optimal diversification model in this case is an explicit-function model that produces mathematically explicit functions for the optimal proportion of total capital wealth to be devoted to stocks, and the optimal number of stocks to hold in the investor’s stock portfolio. Specifically, optimal proportion is linearly related to the reciprocal of the square root of total capital wealth, while optimal number is linearly related to the square root of total capital wealth. Some economists are skeptical of explicit-function models on the grounds that they are excessively specific. If a general-function model had been used instead, the comparative statics results would merely have been that the optimal proportion and the optimal number were some unknown function of total capital wealth. The standard econometric approach, when the function to be estimated is unknown, is to utilize a linear specification. As part of the research, we did in fact estimate linear versions of the various estimated equations. The linear specifications did not fit the data as well as the specifications actually used. For example, taking the case of the sparse estimation of the equation for optimal proportion ρ using the full dataset (shown in the first column of Table 3), the R-squared for the linear specification was 0.01, compared to an R-squared for the actual formulation of 0.09. Moreover, if it is mathematically possible to derive non-linear explicit-function comparative statics results from explicit-function optimization models, empirically testing these results constitutes a sharper test of hypothesis than applying the conventional linear specification to test general-function comparative statics results derived from general-function models. Thus the use of something other than the conventional linear specifications might be considered a strength of the present research.

With respect to the empirical data utilized to test the theoretical predictions of the optimal diversification model, further questions could be raised. Measuring the "true value" of capital wealth is notoriously difficult, and the present research does not pretend to make a contribution on this matter. A straightforward definition of capital wealth is utilized, based exclusively on data provided by a well-known government survey. Be that as it may, it is pure speculation that the measure applied herein is substantively and substantially different from what would be obtained from a more sophisticated approach. It is well-known that any sophisticated estimate of household capital wealth necessarily requires several questionable assumptions of its own. Moreover, quite likely there would a strong correlation between the "crude" capital wealth estimate used here and any given "sophisticated" estimate. Evidence to this effect lies in the likelihood that if the present capital wealth estimates were actually such poor measures of what they are supposed to be measuring, the good statistical results in Tables 3 and 4 would not have been obtained.

6. Conclusion

It is common knowledge that wealthier households hold a larger percentage of their total capital assets in the form of publicly traded stock, and that their stock capital portfolios are more diversified, than is the case with less wealthy households. The objective to this research has been to provide a relatively simple theoretical explanation for these facts that is conveniently testable with readily available, reliable survey data. On the basis of some strong assumptions, especially "homogeneous securities," our mean-variance model, based on Brennan (1975), produces mathematically explicit solutions for the optimal values of stock proportion ρ and number of stocks n. Direct inspection of these solutions indicates that the optimal stock proportion is a linear function of the reciprocal of the square root of total capital wealth, while the optimal number of stock issues is a linear function of the square root of total capital wealth. Data from the 2004 Survey of Consumer Finances (SCF) has been utilized in this research to evaluate these results.

The statistical analysis is supportive: the t-statistic of the $1/\sqrt{a}$ independent variable in the ρ equation (proportion of total portfolio allocated to stock) is strongly significant, both for the full SCF dataset and for the aggregated SCF datasets. While the R-squared goodness-of-fit statistic for the full SCF dataset is quite low, this statistic becomes respectably large for the aggregated SCF datasets. Results for the $\sqrt{a}$ variable in the n equation (number of stocks in the portfolio) are similar, except that even for the full SCF dataset, the R-squared goodness-of-fit statistic is fairly respectable given that the data is cross-sectional.

It goes without saying that some strong assumptions are necessary to obtain the mathematically explicit solutions and unambiguous comparative statics results forthcoming from the homogeneous securities model of optimal diversification. But it has been the universal experience of economic theoreticians that without strong assumptions, rather little of practical interest can be deduced. And strong assumptions are not necessarily invalid assumptions. Moreover, the fact that reasonably good fits to notoriously variable survey data are obtained using regression specifications indicated by the model, constitutes worthwhile evidence that the model, despite its relative simplicity, may in fact be a reasonable approximation to reality.

Appendix

The first-order conditions for the maximization of L as given by equation (6), with respect to ρ and n, are as follows:

$$\frac{\partial L}{\partial \rho} = L_\rho = (\overline{r}_s - r_b) - 2\lambda\rho\left(\sigma_m^2 + \frac{\sigma_s^2}{n}\right) = 0 \tag{A.1}$$

$$\frac{\partial L}{\partial n} = L_n = -(c/a)(1+r_b) + \lambda\rho^2\left(\frac{\sigma_s^2}{n^2}\right) = 0 \tag{A.2}$$

Solving (A.1) for n and (A.2) for n^2, we have respectively:

$$n = \frac{2\lambda\rho\sigma_s^2}{(\bar{r}_s - r_b) - 2\lambda\rho\sigma_m^2} \tag{A.3}$$

$$n^2 = \frac{\lambda\rho^2\sigma_s^2}{(c/a)(1+r_b)} \tag{A.4}$$

Equating n from (A.3) to n from (A.4), we have an expression that simplifies to a linear equation in ρ which may be solved for the optimal ρ*:

$$\rho^* = \frac{.5(\bar{r}_s - r_b)}{\lambda\sigma_m^2} - \frac{\sqrt{\sigma_s^2(c/a)(1+r_b)}}{\sigma_m^2\sqrt{\lambda}} \tag{A.5}$$

Substitution of ρ* into (A.3) above determines the optimal n, denoted by n*:

$$n^* = \frac{.5\sigma_s(\bar{r}_s - r_b)}{\sigma_m^2\sqrt{\lambda(c/a)(1+r_b)}} - \frac{\sigma_s^2}{\sigma_m^2} \tag{A.6}$$

For these solutions to be economically sensible, it is necessary that $0 < \rho^* \leq 1$ and $n^* > 0$. Using inequalities based on (A.5) and (A.6), we determine that the condition for $\rho^* > 0$ and $n^* > 0$ is the same:

$$\bar{r}_s > r_b + 2\sqrt{\lambda\sigma_s^2(c/a)(1+r^b)} \tag{A.7}$$

For $\rho^* \leq 1$, the condition is:

$$\bar{r}_s \leq r_b + 2\lambda\sigma_m^2 + 2\sqrt{\lambda\sigma_s^2(c/a)(1+r_b)} \tag{A.8}$$

The second-order conditions for a maximum in this problem are as follows:

$$L_{\rho\rho} < 0;\ L_{nn} < 0;\ D = L_{\rho\rho}L_{nn} - L_{\rho n}^2 > 0 \tag{A.9}$$

That these conditions are satisfied is shown in the following:

$$L_{\rho\rho} = -2\lambda\left(\sigma_m^2 + (\sigma_s^2/n)\right) < 0 \tag{A.10}$$

$$L_{nn} = -2\lambda\rho^2\sigma_s^2 n^{-3} < 0 \tag{A.11}$$

$$D = L_{\rho\rho}L_{nn} - L_{\rho n}^2 = 4\lambda^2\rho^2\sigma_s^2\sigma_m^s > 0 \tag{A.12}$$

References

Aizcorbe, Anna M., Martha Starr, and James T. Hickman. “Vehicle Purchases, Leasing, and Replacement Demand,” *Business Economics* 39(2): 7-17, April 2004.

Alessie, Rob, Stefan Hochguertel, and Arthur van Soest. “Ownership of Stocks and Mutual Funds: A Panel Data Analysis,” *Review of Economics and Statistics* 86(3): 783-796, August 2004.

Baek, Eunyoung, and Gong-Soog Hong. “Effects of Family Life-Cycle Stages on Consumer Debts,” *Journal of Family and Economic Issues* 25(3): 359-385, 2004.

Baumol, William J. “An Expected Gain-Confidence Limit Criterion for Portfolio Selection,” *Management Science* 10(1): 174-182, October 1963.

Benartzi, Shlomo, and Richard H. Thaler. “Naïve Diversification Strategies in Defined Contribution Savings Plans,” *American Economic Review* 91(1): 79-98, March 2001.

Ben-Gad, Michael. “The Welfare Effects of the Reagan Deficits: A Portfolio Choice Approach,” *Economic Inquiry* 42(3): 441-454, July 2004.

Bera, Anil K., and Sung Y. Park. “Optimal Portfolio Diversification Using the Maximum Entropy Principle,” Econometric Reviews 27(4-6): 484-512, July-December 2008.

Bergstresser, Daniel, and James Poterba. “Asset Allocation and Asset Location: Household Evidence from the Survey of Consumer Finances,” *Journal of Public Economics* 88(9-10): 1893-1915, August 2004.

Berkelaar, Arjan B., Roy Kouwenberg, and Thierry Post. “Optimal Portfolio Choice under Loss Aversion,” *Review of Economics and Statistics* 86(4): 973-987, November 2004.

Berkowitz, Michael K., and Jaiping Qiu. “A Further Look at Household Portfolio Choice and Health Status,” *Journal of Banking and Finance* 30(4): 1201-1217, April 2006.

Brennan, Michael 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(3): 483-496, September 1975.

Brunnermeier, Markus K., and Stefan Nagel. “Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-evidence on Individuals’ Asset Allocation,” *American Economic Review* 98(3): 713-736, June 2008.

Carlos-Hatchondo, Juan. “Asymmetric Information and the Lack of Portfolio Diversification,” *International Economic Review* 49(4): 1297-1330, November 2008.

Carroll, Christopher D. “Portfolios of the Rich.” NBER Working Paper No. W7826. August 2000.

Castronova, Edward, and Paul Hagstrom. “The Demand for Credit Cards: Evidence from the Survey of Consumer Finances,” *Economic Inquiry* 42(2): 304-318, April 2004.

Cocco, Joao F. “Portfolio Choice in the Presence of Housing,” *Review of Financial Studies* 18(2): 535-567, Summer 2005.

Domian, Dale L., David A. Louton, and Marie D. Racine. “Diversification in Portfolios of Individual Stocks: 100 Stocks Are Not Enough,” *Financial Review* 42(4): 557-570, November 2007.

Garlappi, Lorenzo, and Jennifer Huang. “Are Stocks Desirable in Tax-Deferred Accounts?” *Journal of Public Economics* 90(12): 2257-2283, December 2006.

Goetzmann, William N., and Alok Kumar. “Equity Portfolio Diversification.” *Review of Finance* 12(3): 433-463, 2008.

Gomes, Francisco, and Alexander Michaelides. “Portfolio Choice with Internal Habit Formation: A Life-Cycle Model with Uninsurable Labor Income Risk,” *Review of Economic Dynamics* 6(4): 729-766, October 2003.

Gong, Peichen, and Karl Gustaf Lofgren. “Market and Welfare Implications of the Reservation Price

Strategy for Forest Harvest Decisions," *Journal of Forest Economics* 13(4): 217-243, 2007.

Guiso, Luigi, and Tullio Jappelli. "Awareness and Stock Market Participation," *Review of Finance* 9(4): 537-567, 2005.

Gutter, Michael S., and Tabassum Saleem. "Financial Vulnerability of Small Business Owners," *Financial Services Review* 18(4): 1467-1502, Winter 2005.

Hu, Xiaoqing. "Households' Saving and Portfolios," Investment Management and Financial Innovations 3: 132-144, 2004.

Jaworski, Alexa. "In Down Economy, Vendors Preach Diversification and Differentiation," *Securities Industry News* 20(16): 1-19, April 21, 2008.

Jianakoplos, Nancy A., and Alexandra Bernasek. "Financial Risk Taking by Age and Birth Cohort," *Southern Economic Journal* 72(4): 291-1001, April 2006.

Kaustia, Markku, and Samuli Knupfer. "Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions," *Journal of Finance* 63(6): 2679-2702, December 2008.

Kelly, Morgan. "All Their Eggs in One Basket: Portfolio Diversification of U.S. Households," *Journal of Economic Behavior and Organization* 27(1): 87-96, June 1995.

King, Mervyn, and Jonathan Leape. "Wealth and Portfolio Composition: Theory and Evidence," *Journal of Public Economics* 69(2): 155-193, August 1998.

Kole, Erik, Kees Koedijk, and Marno Verbeek. "Portfolio Implications of Systemic Crises," *Journal of Banking and Finance* 30(8): 2347-2369, August 2006.

Lintner, John. "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance* 20(4): 587-615, December 1965.

Listokin, Yair. "Interpreting Empirical Estimates of the Effect of Corporate Governance," *American Law and Economics Review* 10(1): 90-109, Spring 2008.

Lundtofke, Frederik. "The Effect of Information Quality on Optimal Portfolio Choice," *Financial Review* 41(2): 157-185, May 2006.

Markowitz, Harry M. "Portfolio Selection," *Journal of Finance* 7(1): 77-91, March 1952.

Markowitz, Harry M. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley and Sons, 1959.

Massa, Massimo, and Andrei Simonov. "Hedging, Familiarity and Portfolio Choice," *Review of Financial Studies* 19(2): 633-985, Summer 2006.

McClatchey, Christine A., and Shawn P. Vandenhul. "The Efficacy of Optimization Modeling as a Retirement Strategy in the Presence of Estimation Error," *Financial Services Review* 14(4): 269-284, Winter 2005.

Pachamanova, Dessislava. "Handling Parameter Uncertainty in Portfolio Risk Minimization: The Robust Optimization Approach," *Journal of Portfolio Management* 32(4): 70-78, Summer 2006.

Pelizzon, Loriana, and Guglielmo Weber. "Are Household Portfolios Efficient? An Analysis Conditional on Housing," *Journal of Financial and Quantitative Analysis* 43(2): 401-431, June 2008.

Polkovnichenko, Valery. "Household Portfolio Diversification: A Case for Rank-Dependent Preferences," *Review of Financial Studies* 18(4): 1467-1502, Winter 2005.

Poterba, James M., and Andrew A. Samwick. "Taxation and Portfolio Composition: U.S. Evidence from the 1980s and 1990s." *Journal of Public Economics* 87(1): 5-38, January 2003.

Projector, Dorothy S., and Gertrude S. Weiss. *Survey of Financial Characteristics of Consumers*. Washington: Board of Governors of the Federal Reserve System, 1966.

Saarimaa, Tuukka. "Owner-Occupied Housing and Demand for Risky Financial Assets: Some Finnish

Evidence," *Finnish Economic Papers* 21(1): 22-38, Spring 2008.

Samuelson, Paul A. "General Proof that Diversification Pays," *Journal of Financial and Quantitative Analysis* 2(1): 1-13, March 1967.

Sharpe, William F. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* 19(3): 425-442, September 1964.

Sparling, Jeffry. "Increased Diversification Has Become Necessary in Managing Multiemployer Pension Plans," *Benefits & Compensation Digest* 45(6): 24-28, June 2008.

Stevenson, Simon. "Bayes-Stein Estimators and International Real Estate Asset Allocation," *Journal of Real Estate Research* 21(1-2): 89-103, January-April 2001.

Vissing-Jorgensen, Annette, and Orazio P. Attanasio. "Stock Market Participation, Intertemporal Substitution, and Risk-Aversion," *American Economic Review* 93(2): 383-391, May 2003.

White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica* 48(4): 817-838, May 1980.

Windcliff, Heath, and Phelim P. Boyle. "The 1/n Pension Investment Puzzle," *North American Actuarial Journal* 8(3): 32-45, July 2004.

Wolff, Edward N. "Changing Inequality of Wealth," *American Economic Review* 82(2): 552-558, May 1992.

Wolff, Edward N. "Estimates of Household Wealth Inequality in the U.S., 1962-1983," *Review of Income and Wealth* 33(3): 231-256, September 1987.

Wolff, Edward N. "Trends in Household Wealth in the United States, 1962-83 and 1983-89," *Review of Income and Wealth* 40(2): 143-174, June 1994.

Wu, Stephen. "Fatalistic Tendencies: An Explanation of Why People Don't Save," *B.E. Journals in Economic Analysis and Policy: Contributions to Economic Analysis and Policy* 4(1): 1-21, 2005.

Table 1
Variables Utilized in the Research

Survey of Consumer Finance (SCF) Variables:

#	Name	Explanation	Usage
1	X3721	TOTAL VALUE OF CDS	*a* component
2	X3822	TOTAL MKT VAL STOCK FUNDS	*a* component
3	X3824	TOT MKT VAL TAX FREE BONDS	*a* component
4	X3826	TOT MKT VAL GVMT BACK BOND	*a* component
5	X3828	TOTAL MKT VAL OTHER BONDS	*a* component
6	X3830	TOTAL MKT VAL COMBO FUNDS	*a* component
7	X3902	VALUE OF SAVINGS BONDS	*a* component
8	X3906	MORT_BONDS:FACE VALUE	*a* component
9	X3908	TREAS_BONDS:FACE VALUE	*a* component
10	X3910	MUNI/STATE_BONDS:FACE VALUE	*a* component
11	X3915	TOTAL MARKET VALUE OF STOCKS	*a* component
12	X3913	HAVE ANY PUBLIC TRADED STOCK?	1 = yes; 0 = no
13	X3914	NUMBER OF DIFFERENT STOCKS	*n*
14	X14	RESPONDENT'S RECONCILED AGE	Control variable
15	X101	NUM PEOPLE IN HH ACCORD TO HHL	Control variable
16	X301	EXPECTATIONS FOR ECONOMY	Control variable
17	X302	INTEREST RATES HGHR, LWR, SAME?	Control variable
18	X304	PAST 5 YEARS INC HGHR, LWR, SAME?	Control variable
19	X3103	OWN/SHARE OWNERSHIP ANY BUS?	Control variable
20	X5702	AMOUNT OF WAGE-SALARY INCOME	Control variable
21	X5901	RESPONDENT GRADE COMPLETED	Control variable

Constructed Variables:

a	= X3721 + X3822 + 3824 + X3826 + X3828 + X3830 + X3902 + X3906 + X3908 + X3910 + X3915
ρ	= (X3822 + X3915) / a
$1/\sqrt{a}$	reciprocal of square root of a
$\sqrt{a}$	square root of a
Log(a)	logarithm of a

Table 2
Bracket Means for the 25-Observation Dataset

Bracket	Value of Total Capital Wealth (*a*)	Proportion of Households Owning Some Stock (X3913)	Number of Stocks in Portfolio (*n* = X3914)	Value of Stocks in Portfolio (X3822 + X3915)	Value of Stocks as a Proportion of Value of Total Capital Wealth (ρ)
1	44,080,341	0.9444	48.27	30,124,101	0.6834
2	6,304,269	0.8389	25.44	3,817,593	0.6055
3	1,990,672	0.8722	20.61	1,338,079	0.6722
4	833,996	0.7889	12.91	556,062	0.6667
5	409,059	0.6667	9.49	270,766	0.6619
6	204,833	0.7111	5.59	144,152	0.7037
7	104,094	0.6222	3.56	66,863	0.6423
8	52,401	0.5167	2.84	32,138	0.6133
9	26,065	0.5611	2.12	15,467	0.5934
10	13,275	0.5000	1.58	7,229	0.5446
11	6,401	0.5278	1.15	3,681	0.5751
12	2,533	0.4056	0.71	1,306	0.5156
13	688	0.2056	0.31	186	0.2697
14	38	0.0278	0.04	4	0.1088
15	0	0	0	0	—
16	0	0	0	0	—
17	0	0	0	0	—
18	0	0	0	0	—
19	0	0	0	0	—
20	0	0	0	0	—
21	0	0	0	0	—
22	0	0	0	0	—
23	0	0	0	0	—
24	0	0	0	0	—
25	0	0	0	0	—

Table 3

Regression Equations for ρ (Proportion of Stock in Portfolio)
and n (Number of Stocks Held in Portfolio)

Full Survey of Consumer Finances (SCF) Dataset

Independent Variables	Estimated Regression Coefficients of Independent Variables (t-statistics in parentheses)			
	Dependent Variable ρ		Dependent Variable n	
	sparse	augmented	sparse	augmented
intercept	0.6356 (70.51)	0.3338 (3.98)	1.4808 (7.63)	–5.4915 (–4.82)
$1/\sqrt{a}$	–5.1055 (–12.04)	–4.8075 (–10.56)	—	—
$\sqrt{a}$	—	—	0.007556 (14.50)	0.006929 (12.27)
age of respondent	—	–0.0025 (–3.91)	—	0.0414 (3.62)
number in household	—	–0.0066 (–1.03)	—	–0.0864 (–0.65)
expect. better econ. perform.	—	0.0213 (1.34)	—	–0.4580 (–1.15)
expect. higher int. rates	—	0.0303 (1.18)	—	0.5006 (1.21)
higher income past 5 yrs	—	0.0466 (2.72)	—	0.7060 (1.30)
business ownership share	—	0.0162 (0.95)	—	2.4251 (3.81)
wage-salary income	—	–3.07E-10 (–0.13)	—	1.19E-07 (0.24)
years of education	—	0.0258 (6.07)	—	0.3049 (5.77)
R-squared	0.09	0.13	0.38	0.39
F-statistic	246.58	39.73	2716.62	316.03
observations	2424	2424	4519	4519

Table 4
Sparse Regression Equations for ρ and n
Using Aggregated SCF Datasets

Independent Variables	Regression Coefficients of Independent Variables (t-statistics in parentheses)			
	Dependent Variable ρ		Dependent Variable n	
	100 brackets	25 brackets	100 brackets	25 brackets
intercept	0.6364 (49.11)	0.6277 (33.07)	1.4057 (4.10)	1.1642 (2.65)
$1/\sqrt{a}$	–5.1607 (–4.94)	–3.5784 (–7.97)	—	—
$\sqrt{a}$	—	—	0.0077 (8.77)	0.0078 (11.45)
R-squared	0.69	0.80	0.87	0.94
F-statistic	117.77	48.46	641.08	388.61
observations	54	14	100	25

Figure 1
Actual and Estimated ρ (Proportion of Stocks in Portfolio)
54 Observations from 100-Household Aggregated SCF Dataset

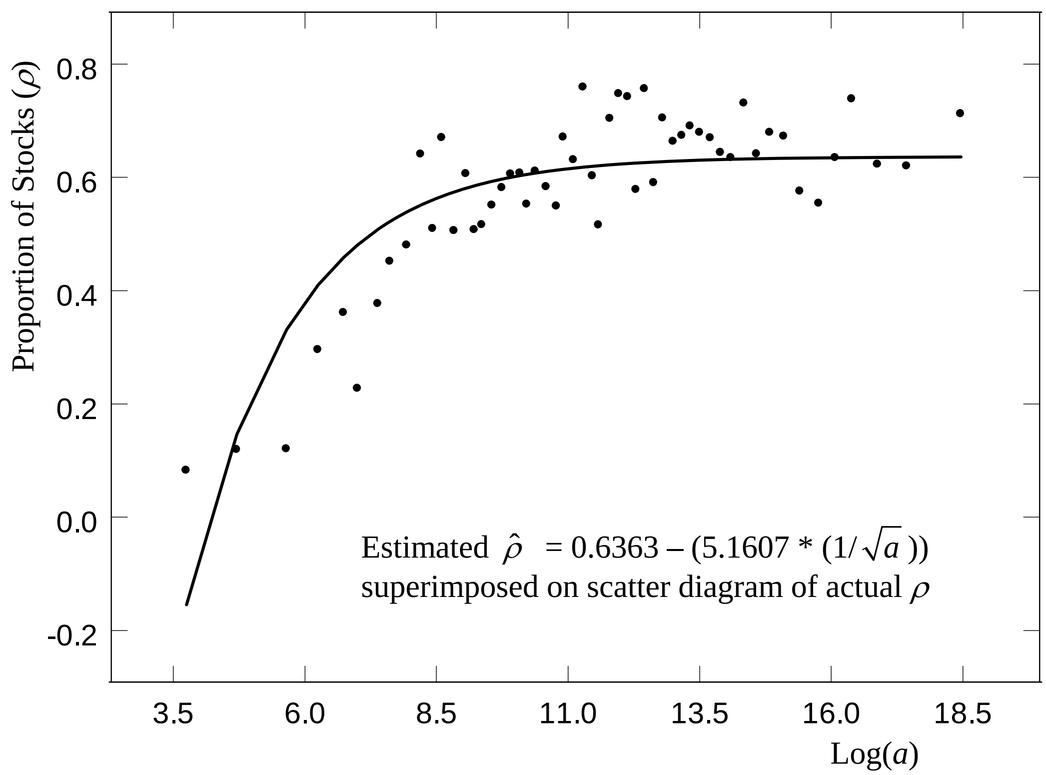

Figure 2
Actual and Estimated *n* (Number of Stocks in Portfolio)
100 Observations from 100-Household Aggregated SCF Dataset