Online Appendix
for
"Location Choice, Portfolio Choice"

Ioannis Branikas[a,*], Harrison Hong[b], Jiangmin Xu[c]

[a] *Department of Finance, Lundquist College of Business, University of Oregon*
[b] *Department of Economics, Columbia University*
[c] *Department of Finance, Guanghua School of Management, Peking University*

This Online Appendix consists of three sections. Section A lists the definitions of the MSA livability scores extracted from the 1993 edition of Places Rated Almanac (by Savageau and Boyer (1993)). Section B analyzes the expression of the control function. Section C contains the results for the additional robustness exercises.

*Corresponding author

A. Definition of the MSA Livability Scores

Transportation: This score is calculated based on the daily commute, public transportation, national highways, air service and passenger rail service. The higher the score of transportation, the better the transportation in the MSA.

Colleges: This score is based on the number of students enrolled in community or two-year colleges, the number of students enrolled in private four-year and graduate-level institutions and the number of students enrolled in public four-year and graduate level institutions. The higher the score of colleges, the better the colleges in the MSA.

Health Care: This score is based on the number of general/family practitioners per 100K population, the number of medical specialists per 100K population, the number of surgical specialists per 100K population and the number of hospitals approved for physician residency programs by the AMA. The higher the score of health care, the better the health care in the MSA.

Crime: This score is based on the violent crime rate and the property crime rate divided by 10. The lower the score of crime, the less the crime in the MSA.

Recreation: This score is based on the number of public golf courses, good restaurants, movie theater screens, zoos, aquariums, family theme parks, parimutuel betting attractions, professional sports, collegiate sports, miles of ocean or Great Lakes coastline, national forests, national parks and national wildlife refuges and state or provincial parks. The higher the score of recreation, the better the recreation in the MSA.

Climate: The score is based on the number of very hot and cold months, the seasonal temperature variation, the number of heating and cooling degree days, the number of freezing days, the number zero-degree days and the number of 90-degree days. The higher the score of climate, the better the climate in the MSA.

B. Control Function Expression

Given household i's decision to live in city c and the observed location utilities $\{V_{i,\ell}\}_{\ell=1}^{C}$, which are estimated from the location choice model in a first stage, the control function can be calculated as follows:

$$\begin{aligned}
\mathbb{E}\left(\epsilon_{i,c,h,j}\middle|v_{i,c}<0,\{V_{i,\ell}\}_{\ell=1}^{C}\right) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{0}\frac{\epsilon_{i,c,h,j}f\left(\epsilon_{i,c,h,j},v_{i,c}\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)}{\mathbb{P}\left(v_{i,c}<0\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)}dv_{i,c}d\epsilon_{i,j} \\
&= \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{0}\epsilon_{i,c,h,j}f\left(\epsilon_{i,c,h,j},v_{i,c}\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)dv_{i,c}d\epsilon_{i,j}}{\mathbb{P}\left(v_{i,c}<0\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)} \qquad (1) \\
&= \psi_{c,h}\left(\{V_{i,\ell}\}_{\ell=1}^{C}\right)
\end{aligned}$$

where $\psi_{c,h}(\cdot)$ is an *unknown* control function whose actual form depends on assumptions regarding the *joint* distribution of $\epsilon_{i,c,h,j}$ and $v_{i,c}$. There is a monotonic relationship between household i's location utilities, $\{V_{i,\ell}\}_{\ell=1}^{C}$, and location probabilities, $\{p_{i,\ell}\}_{\ell=1}^{C}$, i.e.

$$\psi_{c,h}\left(\{V_{i,\ell}\}_{\ell=1}^{C}\right) = \Psi_{c,h}\left(\{p_{i,\ell}\}_{\ell=1}^{C}\right) \qquad (2)$$

The value of the control function is in principle non-zero, unless $\epsilon_{i,c,h,j}$ and $v_{i,c}$ are independent. In that case, the numerator in the second line of Equation (2) becomes:

$$\begin{aligned}
\int_{-\infty}^{+\infty}\int_{-\infty}^{0}\epsilon_{i,c,h,j}f\left(\epsilon_{i,c},v_{i,c}\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)dv_{i,c}d\epsilon_{i,c,h,} &= \int_{-\infty}^{+\infty}\epsilon_{i,c,h,j}f\left(\epsilon_{i,c,h,j}\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)d\epsilon_{i,c,h,j}\int_{-\infty}^{0}f\left(v_{i,c}\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)dv_{i,c} \\
&= \mathbb{E}\left(\epsilon_{i,c,h,j}\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right)\mathbb{P}\left(v_{i,c}<0\middle|\{V_{i,\ell}\}_{\ell=1}^{C}\right) \\
&= 0 \qquad (3)
\end{aligned}$$

since the conditional mean of $\epsilon_{i,c,h,j}$, given the observables,is zero.

C. Robustness Checks and Extensions

In this section of the Online Appendix, we conduct a number of robustness checks and extensions.

First, in Appendix Table 1, we repeat the Tobit regression of household portfolio choice using linear and quadratic approximations of the control functions. In column 2, we show that with linear control functions, the estimated coefficient of *Distance* is decreased by 21% (i.e., from −0.014 to −0.011). In column 3, where the control functions are approximated quadratically, the location choice correction results in a decrease of 29% for the *Distance* coefficient (i.e., from −0.014 to −0.010).

Second, we consider alternative measures for the local bias, replacing the continuous distance variable measured in degrees with (i) indicators for whether the headquarters of a stock are more than a certain threshold of miles (e.g., 100 or 250 miles) away from a household's residence and (ii) the log of distance.

Specifically, in Panel A of Appendix Table 2, when we use the Tobit specification and the threshold of 100 miles, the reduction in the Away coefficient is 32% (i.e., from −0.634 in Column 1 to −0.432 in Column 2). When we use the more conservative threshold of 250 miles, the reduction in the local bias is 41% (from −0.505 in Column 3 to −0.3 in Column 4). The Tobit estimation results for the portfolio choice when we use the log of the distance are depicted in Panel A of Table 3. The coefficient of *LogDist* is decreased by 27% (from −0.166 in Column 1 to −0.122 in Column 2).

In Panels B of Appendix Tables 2 and 3, we present the 2SLS vs. OLS estimation results for the linear under-diversification model. For the Away 100 miles dummy variable, the decrease is 23% (from −7.537 in Column 1 to −5.819 in Column 2). For the Away 250 miles dummy variable, the reduction is 31% (from −4.866 in Column 3 to −3.343 in Column 4). For *LogDist*, the decrease is 32% (from −1.672 in Column 1 to −1.142 in Column 2).

Third, we extend our sample of stocks to the universe of Russell 3000 and our MSA sample to 80 MSAs whose population in the beginning of 1991 was at least 500,000.

In Panel A of Appendix Table 4, we depict the Tobit estimation results for households living in 80 MSAs and stocks that were members of Russell 3000 during the sample period. Without accounting for location choice, the (linear) distance coefficient is found to be -0.012. When we incorporate the control functions with the predicted location probabilities, the distance coefficient estimates decrease to -0.007. This change amounts to a 42% reduction.

In the same spirit, in Panel B of Appendix Table 4, we depict the 2SLS estimaton results of the extended model. The average OLS distance coefficient is -0.157 (in Column 1) with an average t-statistic of -6.8. The average 2SLS is is -0.099 (in Column 2) with an average t-statistic of -2.78, pointing to 37% lower local bias relative to the OLS model.

Lastly, instead of $LogRecreation \times LogAge$, we use as an instrument for *Distance* the interaction between a retirement age indicator variable for the household (that equals one when the household head is 55 to 74 years old) and the log recreation score of an MSA.

In Appendix Table 5, we show the new (first-stage) conditional logit estimation of household location choice. The results are similar to the ones obtained when using $LogRecreation \times LogAge$ in Table 3 of the main paper. In Appendix Table 6, we also depict the results from the second-stage estimation in selected subsamples of households whose length of residence is shorter than 10 years. In Panel A, where we use the Tobit specification, we obtain results that are virtually identical to the results in Table 6 of the main paper. In Panel B, where we present the 2SLS vs. OLS estimation results for the linear under-diversification model, we find a decrease in the *Distance* coefficient of 21% (from -0.112 in Column 1 to -0.088 in Column 2).

Appendix Table 1

Tobit Regressions of Household Portfolio Choice With Linear or Quadratic Control Functions of Location Choice Probabilities

This table presents Tobit regressions of household portfolio weights when the correction functions are linearly or quadratically approximated. The dependent variable is $w_{i,c,h,j}$, i.e. the portfolio weight of household i residing in MSA c on stock j headquartered in MSA h. The independent variable is *Distance*, i.e. the distance (in degrees) of stock j's headquarters' ZIP-Code from household i's address ZIP-Code. The instrument is the interaction between the (log) recreation score of MSA h of stock j's headquarters and the (log) age of household i. The controls are stock financial characteristics, household demographics and stock headquarters' MSA demographics — including the (log) recreation score. The interactions of all *other* MSA demographics with Household Basic Demographics are also included as controls. See Table 1 for a detailed description. In Column 1, there are no control functions, so location choice is not taken into account. In Column 2, the approximation of the control functions is linear. In Column 3, the approximation of the control functions is quadratic. The estimation is performed in a panel of households and stocks for every month separately. The table depicts the average monthly coefficient estimates and the average t-statistics [in brackets] based on standard errors clustered at the level of the household's MSA (for Column 1) and bootstrap standard errors (for Columns 2 and 3).

	(1)	(2)	(3)
	Uncor.	Linear Ψ's	Quadratic Ψ's
Distance	-0.014	-0.011	-0.01
	[-6.21]	[-3.33]	[-4.8]
LogSize	0.368	0.367	0.366
	[47.73]	[49.31]	[49.17]
BTM	0.05	0.049	0.049
	[6.34]	[5.31]	[5.27]
Turnover	0.579	0.568	0.561
	[12.17]	[10.8]	[10.22]
Momentum	-0.197	-0.197	-0.197
	[-8.66]	[-8.24]	[-8.43]
Volatility	2.897	2.9	2.888
	[15.21]	[14.3]	[14.11]
Profitability	-0.065	-0.069	-0.071
	[-1.61]	[-2.03]	[-2.2]
Investment	-0.094	-0.094	-0.094
	[-3.79]	[-3.19]	[-3.37]
Stock Industry FE	YES	YES	YES
Household Basic Demo's	YES	YES	YES
Household Prof. Demo's	YES	YES	YES
HH ZIP-Code Race Pct.	YES	YES	YES
LogRecreation	YES	YES	YES
Other MSA Demo's	YES	YES	YES
Other MSA Demo's $\times$ HH Basic Demo's	YES	YES	YES
Average Number of Households	4,339	4,339	4,339
Average Number of Stocks	988	988	988

Appendix Table 2

Regressions of Household Portfolio Weights on Distance Indicator Variables

This table presents the Tobit regressions of household portfolio weights and the linear regressions of household excess portfolio weights on distance indicator variables. In Panel A, the dependent variable is $w_{i,c,h,j}$, i.e. the portfolio weight of household i residing in MSA c on stock j headquartered in MSA h. The independent variable is $Away$, i.e. an indicator variable that equals one if the distance of stock j's headquartersâĂŹ ZIP Code from household i's address ZIP Code is greater than a specific threshold value. In Columns 1 and 2, the distance threshold is 100 miles. In Columns 3 and 4, the distance threshold is 250 miles. The instrument is the interaction between the (log) recreation score of MSA h of stocks j's headquarters and the (log) age of household i. The controls are stock financial characteristics, household demographics and stock headquarters' MSA demographics — including the (log) recreation score. The interactions of all *other* MSA demographics with Household Basic Demographics are included as controls. See Table 1 for a detailed description. Columns 1 and 3 do not correct for household location choice. Columns 2 and 4 correct for household location choice, approximating the control functions with fourth order polynomials. In Panel B, the dependent variable is $EW_{i,c,h,j} \equiv (w_{i,c,h,j} - w_j^{VW})/w_j^{VW}$, i.e. the excess portfolio weight of household i residing in city c on stock j headquartered in city h (w.r.t. to the market value-weighted portfolio weight on stock j). Columns 1 and 3 show the OLS, and Columns 2 and 4 the 2SLS. The F-statistic tests the hypothesis that the coefficient of the instrument is zero. The estimation is performed in a panel of households and stocks for every month separately. The table depicts the average monthly coefficient estimates and the average t-statistics [in brackets] based on standard errors clustered at the level of the household's MSA (for Columns 1 and 3 of Panel A and all columns of Panel B) and bootstrap standard errors (for Columns 2 and 4 of Panel A).

	(1)	(2)	(3)	(4)
	100 Miles Threshold		250 Miles Threshold	
	Panel A: Non-linear Portfolio Choice Model			
	Uncor.	Corrected	Uncor.	Corrected
Away	-0.634	-0.432	-0.505	-0.3
	[-4.96]	[-5.27]	[-5.32]	[-5.68]
Stock Financial Char's	YES	YES	YES	YES
HH Basic Demo's	YES	YES	YES	YES
HH Prof. Demo's	YES	YES	YES	YES
HH Occupation-Code FE	YES	YES	YES	YES
HH ZIP-Code Race Pct.	YES	YES	YES	YES
LogRecreation	YES	YES	YES	YES
Other MSA Demo's	YES	YES	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES	YES	YES
Average Number of Households	4,339	4,339	4,339	4,339
Average Number of Stocks	988	988	988	988
	Panel B: Linear Portfolio Under-Diversification Model			
	OLS	2SLS	OLS	2SLS
Away	-7.537	-5.819	-4.866	-3.343
	[-3.99]	[-2.1]	[-4.82]	[-2.24]
Stock Financial Char's	YES	YES	YES	YES
HH Basic Demo's	YES	YES	YES	YES
HH Prof. Demo's	YES	YES	YES	YES
HH Occupation-Code FE	YES	YES	YES	YES
HH ZIP-Code Race Pct.	YES	YES	YES	YES
LogRecreation	YES	YES	YES	YES
Other MSA Demo's	YES	YES	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES	YES	YES
First-stage F-statistic		17.19		16.7
Average Number of Households	4,339	4,339	4,339	4,339
Average Number of Stocks	988	988	988	988

Appendix Table 3

Regressions of Household Portfolio Weights on Log Distance

This table presents the Tobit regressions of household portfolio weights and the linear regressions of household excess portfolio weights on the natural logarithm of distance. In Panel A, the dependent variable is $w_{i,c,h,j}$, i.e. the portfolio weight of household i residing in MSA c on stock j headquartered in MSA h. The independent variable is *LogDist*, i.e. the log of distance of stock j's headquarters' ZIP Code from household i's address ZIP Code. The instrument is the interaction between the (log) recreation score of MSA h of stocks j's headquarters and the (log) age of household i. The controls are stock financial characteristics, household demographics and stock headquarters' MSA demographics — including the (log) recreation score. The interactions of all *other* MSA demographics with Household Basic Demographics are also included as controls. See Table 1 for a detailed description. Column 1 does not correct for household location choice. Column 2 corrects for household location choice, approximating the control functions with fourth order polynomials. In Panel B, the dependent variable is $EW_{i,c,h,j} \equiv (w_{i,c,h,j} - w_j^{VW})/w_j^{VW}$, i.e. the excess portfolio weight of household i residing in city c on stock j headquartered in city h (w.r.t. to the market value-weighted portfolio weight on stock j). Column 1 shows the OLS, and Columns 2 the 2SLS. The F-statistic tests the hypothesis that the coefficient of the instrument is zero. The estimation is performed in a panel of households and stocks for every month separately. The table depicts the average monthly coefficient estimates and the average t-statistics [in brackets] based on standard errors clustered at the level of the household's MSA (for Column 1 of Panel A and all columns of Panel B) and bootstrap standard errors (for Column 2 of Panel A).

	(1)	(2)
	Panel A: Non-linear Portfolio Choice Model	
	Uncor.	Corrected
LogDist	-0.166	-0.122
	[-6.99]	[-8.23]
Stock Financial Char's	YES	YES
Household Basic Demo's	YES	YES
Household Prof. Demo's	YES	YES
HH ZIP-Code Race Pct.	YES	YES
LogRecreation	YES	YES
Other MSA Demo's	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES
Average Number of Households	4,339	4,339
Average Number of Stocks	988	988
	Panel B: Linear Portfolio Under-Diversification Model	
	OLS	2SLS
LogDist	-1.672	-1.142
	[-5.2]	[-2.37]
Stock Financial Char's	YES	YES
Household Basic Demo's	YES	YES
Household Prof. Demo's	YES	YES
HH ZIP-Code Race Pct.	YES	YES
LogRecreation	YES	YES
Other MSA Demo's	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES
First-stage F-statistic		18.17
Average Number of Households	4,339	4,339
Average Number of Stocks	988	988

Appendix Table 4

Regressions of Household Portfolio Weights on Distance for Russell 3000 Stocks in 80 MSAs

This table presents the Tobit regressions of household portfolio weights and the linear regressions of household excess portfolio weights on distance, for households that reside in 80 MSAs with a population of at least 500K in the beginning of 1991, and when the investment universe consists of Russell 3000 stocks headquartered in these MSAs. In Panel A, the dependent variable is $w_{i,c,h,j}$, i.e. the portfolio weight of household i residing in MSA c on stock j headquartered in MSA h. The independent variable is $LogDist$, i.e. the log of distance of stock j's headquarters' ZIP Code from household i's address ZIP Code. The instrument is the interaction between the (log) recreation score of MSA h of stocks j's headquarters and the (log) age of household i. The controls are stock financial characteristics, household demographics and stock headquarters' MSA demographics — including the (log) recreation score. The interactions of all *other* MSA demographics with Household Basic Demographics are also included as controls. See Table 1 for a detailed description. Column 1 does not correct for household location choice. Column 2 corrects for household location choice, approximating the control functions with fourth order polynomials. In Panel B, the dependent variable is $EW_{i,c,h,j} \equiv (w_{i,c,h,j} - w_j^{VW})/w_j^{VW}$, i.e. the excess portfolio weight of household i residing in city c on stock j headquartered in city h (w.r.t. to the market value-weighted portfolio weight on stock j). Column 1 shows the OLS, and Columns 2 the 2SLS. The F-statistic tests the hypothesis that the coefficient of the instrument is zero. The estimation is performed in a panel of households and stocks for every month separately. The table depicts the average monthly coefficient estimates and the average t-statistics [in brackets] based on standard errors clustered at the level of the household's MSA (for Column 1 of Panel A and all columns of Panel B) and bootstrap standard errors (for Column 2 of Panel A).

	(1)	(2)
	Panel A: Non-linear Portfolio Choice Model	
	Uncor.	Corrected
Distance	-0.012	-0.007
	[-7.92]	[-9.38]
Stock Financial Char's	YES	YES
Household Basic Demo's	YES	YES
Household Prof. Demo's	YES	YES
HH ZIP-Code Race Pct.	YES	YES
LogRecreation	YES	YES
Other MSA Demo's	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES
Average Number of Households	5,524	5,524
Average Number of Stocks	3,517	3,517
	Panel B: Linear Portfolio Under-Diversification Model	
	OLS	2SLS
Distance	-0.157	-0.099
	[-6.8]	[-2.78]
Stock Financial Char's	YES	YES
Household Basic Demo's	YES	YES
Household Prof. Demo's	YES	YES
HH ZIP-Code Race Pct.	YES	YES
LogRecreation	YES	YES
Other MSA Demo's	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES
First-stage F-statistic		17.7
Average Number of Households	5,524	5,524
Average Number of Stocks	3,517	3,517

Appendix Table 5

Conditional Logistic Regressions of Household Location Choice on *LogRecreation* × *RetirementAge* Instead of *LogRecreation* × *LogAge*

This table presents the conditional logistic regressions of household location choice on *LogRecreation* × *RetirementAge* instead of *LogRecreation* × *LogAge*. The dependent variable is $r_{i,c}$, i.e. an indicator variable that equals one if household i resides in MSA c and zero otherwise. The independent variable is *LogRecreation* × *RetirementAge*, i.e. the interaction between the (log) recreation score of MSA h of stocks j's headquarters and the an indicator variable for the retirement age of household i that equals one if the household head is 55 to 74 years old. In Column 1, the (log) recreation score of the MSAs is the only control. In Column 2, other MSA Demographics are added as controls. In Column 3, the interactions of other MSA Demographics with the Household Basic Demographics and Retirement Age are also included as controls. A likelihood ratio test is performed in Columns 1 and 2 for restricted versions of the full model in Column 3. AIC is the Akaike information criterion. The table depicts the coefficient estimates and t-statistics [in brackets] based on standard errors clustered at the household level.

	(1)	(2)	(3)
LogRecreation × Retirement Age	0.307	0.271	0.218
	[4.19]	[4.16]	[3.54]
LogRecreation	0.972	0.382	0.442
	[23.01]	[9.05]	[9.99]
Other MSA Demo's	NO	YES	YES
Other MSA Demo's × HH Basic Demo's	NO	NO	YES
Other MSA Demo's × Retirement Age	NO	NO	YES
LR Test vs. Full Model	7,174	170.65	-
AIC	69,056	62,069	61,962
Number of HH	8,688	8,688	8,688
Number of MSAs	57	57	57

Appendix Table 6

Regressions of Household Portfolio Weights In Subsamples of Households with Residence Length Shorter than 10 Years Using as Instrument *LogRecreation* × *RetirementAge* Instead of *LogRecreation* × *LogAge*

This table presents the Tobit regressions of household portfolio weights and the linear regressions of household excess portfolio in selected subsamples of households whose whose length of residence is shorter than 10 years using as an alternative instrument *LogRecreation* × *RetirementAge* (instead of *LogRecreation* × *LogAge*). In Panel A, the dependent variable is $w_{i,c,h,j}$, i.e. the portfolio weight of household i residing in MSA c on stock j headquartered in MSA h. The independent variable is *Distance*, i.e. the distance of stock j's headquarters' ZIP Code from household i's address ZIP Code. The instrument is the interaction between the (log) recreation score of MSA h of stocks j's headquarters and the an indicator variable that equals one if the household head is 55 to 74 years old (i.e., has reached retirement age). The controls are stock financial characteristics, household demographics — including the retirement age indicator — and stock headquarters' MSA demographics — including the (log) recreation score. The interactions of all *other* MSA demographics with Household Basic Demographics and *RetirementAge* are also included as controls. See Table 1 for a detailed description. Column 1 does not correct for household location choice. Column 2 corrects for household location choice, approximating the control functions with fourth order polynomials. In Panel B, the dependent variable is $EW_{i,c,h,j} \equiv (w_{i,c,h,j} - w_j^{VW})/w_j^{VW}$, i.e. the excess portfolio weight of household i residing in city c on stock j headquartered in city h (w.r.t. to the market value-weighted portfolio weight on stock j). Column 1 shows the OLS, and Columns 2 the 2SLS. The first-stge F-statistic tests the hypothesis that the coefficient of the instrument is zero. The estimation is performed in a panel of households and stocks for every month separately. The table depicts the average monthly coefficient estimates and the average t-statistics [in brackets] based on standard errors clustered at the level of the household's MSA (for Column 1 of Panel A and all columns of Panel B) and bootstrap standard errors (for Column 2 of Panel A).

	(1)	(2)
	Panel A: Non-linear Portfolio Choice Model	
	Uncor.	Corrected
Distance	-0.014	-0.008
	[-7.78]	[-8.49]
Stock Financial Char's	YES	YES
Household Basic Demo's	YES	YES
Household Prof. Demo's	YES	YES
HH ZIP-Code Race Pct.	YES	YES
Retirement Age	YES	YES
LogRecreation	YES	YES
Other MSA Demo's	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES
Other MSA Demo's × Retirement Age	YES	YES
Average Number of Households	2,801	2,801
Average Number of Stocks	988	988
	Panel B: Linear Portfolio Under-Diversification Model	
	OLS	2SLS
Distance	-0.112	-0.088
	[-6.38]	[-2.41]
Stock Financial Char's	YES	YES
Household Basic Demo's	YES	YES
Household Prof. Demo's	YES	YES
HH ZIP-Code Race Pct.	YES	YES
Retirement Age	YES	YES
LogRecreation	YES	YES
Other MSA Demo's	YES	YES
Other MSA Demo's × HH Basic Demo's	YES	YES
Other MSA Demo's × Retirement Age	YES	YES
First-stage F-statistic		13.9
Average Number of Households	2,801	2,801
Average Number of Stocks	988	988

Appendix Figure 1

Geographical Distribution of Households in 57 MSAs and Russell 1000 Stocks

This figure depicts the ZIP-Code geographical coordinates of 8,688 households residing in 57 MSAs with a population of at least 750K at the end of the year 1996 and 1,193 publicly traded firms included in the Russell 1000 index during the sample period. The address ZIP-codes of households and the stocks' headquarters are converted to geographical coordinates based on the correspondence provided by the US Census Bureau. The horizontal axis is in longitude coordinates, while the vertical axis is in latitude coordinates. The blue circles indicate households, while the red squares indicate stocks. The sample period is from January 1991 to November 1996.