NBER WORKING PAPER SERIES

WALL STREET VS. MAIN STREET: AN EVALUATION OF PROBABILITIES

Robin L. Lumsdaine
Rogier J.D. Potter van Loon

Working Paper 19103
http://www.nber.org/papers/w19103

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2013

The authors are grateful to the RAND Corporation for assistance with the American Life Panel and for supplying the sampling weights used in this analysis and to Yacine Aït-Sahlia, Ron Anderson, Hector Calvo-Pardo, Maik Dierkes, Amos Golan, Anthony Hall, Nikolaus Hautsch, Peter Hudomiet, Miles Kimball, Olivia Mitchell, Anders Rahbek, Matthew Shapiro, Neil Shephard, Timo Teräsvirta, Martijn van den Assem, Peter Wakker, and Bob Willis, as well as seminar participants at American University, the University of Exeter, the University of Michigan, the University of Portsmouth, the ZEW Conference on "The Role of Expectations in Financial Markets", the 2012 Society for Financial Econometrics (SoFiE) annual meeting, and the 3rd Humboldt-Copenhagen Conference for comments on an earlier draft. All errors remain our own. Financial support from the Netherlands Organisation for Scientific Research (NWO) for the second author is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Wall Street vs. Main Street: An Evaluation of Probabilities
Robin L. Lumsdaine and Rogier J.D. Potter van Loon
NBER Working Paper No. 19103
June 2013
JEL No. C58,D84,G02,G13

ABSTRACT

This paper challenges recent conventional wisdom of a divide between Main Street (the average American consumer) and Wall Street (financial market participants). The views of survey respondents regarding the likelihood of stock index returns exceeding specific thresholds are compared to market views indicated by index options with strikes at analogous thresholds. The econometric specification explicitly addresses some important impediments to using elicited probabilities from survey data. We confirm that Main Street views track Wall Street views, although the association is not one-for one. We find a closer association for those demonstrating a better understanding of the laws of probability.

Robin L. Lumsdaine
Kogod School of Business
American University
4400 Massachusetts Avenue NW
Washington, DC 20016
and NBER
robin.lumsdaine@american.edu

Rogier J.D. Potter van Loon
Erasmus School of Economics
Erasmus University Rotterdam
Burgemeester Oudlaan 50
PO Box 1738, 3000 DR
Rotterdam, The Netherlands
pottervanloon@ese.eur.nl

1. Introduction

The phrase “Wall Street versus Main Street” became commonplace during the recent financial crisis as a shorthand means of describing opposing sides in a variety of contexts, from blame attribution to beneficiaries of government intervention to investor protection. In the aftermath of the crisis, this characterization has become synonymous with the divide between the “haves” and “have-nots”, manifested for example in the sentiment apparent in the Occupy Wall Street movement. It regularly appears in speeches of policymakers and politicians eager to address perceptions related to the faltering economy. In short, “Wall Street versus Main Street” distinguishes the views of financial insiders from those of the general population.

Yet beyond the rhetoric, is there really a divide between what Wall Street and Main Street think? Those that would argue against a divide might appeal either to early theories of information flow to argue that financial market transactions are merely the aggregate result of individual investor decisions or to standard finance theories that indicate management decisions of publicly-traded firms are a direct result of the desire to maximize shareholder value and that, therefore, investment decisions are a reflection of the views of Main Street citizens. In addition, both feedback and herding models in behavioral finance would suggest that Main Street decisions are influenced by what happens on Wall Street (e.g., Shiller 2003).

On the other side of the debate, those arguing that a divide does indeed exist might counter the above arguments by noting the low proportion of active investors in the Main Street population, citing evidence of low financial literacy rates (Lusardi and Mitchell, 2011), the fact that few Americans hold stocks outside of a retirement portfolio (Poterba and Samwick, 1995), and growing income inequality (Heathcote, Perri, and Violante, 2009). Even among the subset of the population that is active in financial markets, there is evidence that not all participants are informed (e.g., De Long, Shleifer, Summers, and Waldmann 1990) and that for a variety of reasons, returns of the two groups often differ (e.g., Barber and Odean 2002).

This paper attempts to measure the *magnitude* of the Wall Street/Main Street divide. We find a link between beliefs about future stock market returns elicited from surveys of a representative sample of the U.S. population and market beliefs as calculated through option prices using the Black-Scholes model. We also explore whether this link depends upon observable characteristics and find that evidence of greater probabilistic understanding is associated with a stronger Wall Street/Main Street link.

The approach taken in this paper draws on literature from finance, behavioral economics, econometrics, and survey methodology, specifically the following areas: (1) the formation of equity return expectations, (2) the information content of option prices, (3) the information content of subjective expectations, and (4) the tendency of survey respondents to report focal points (clustering around rounded numbers) when asked probabilistic questions.

Researchers using survey data find substantial heterogeneity across individuals with regards to their expectations about future equity returns (Brennan, Cao, Strong, and Xu, 2005; Dominitz and Manski, 2011; Hudomiet, Kézdi and Willis, 2011). Some of the interpersonal heterogeneity is attributable to differences in optimism among population subgroups: for example, women, African Americans and those in the lower education categories have less optimistic expectations relative to the overall population. This heterogeneity in expectations has in turn been used to explain heterogeneous equity investment decisions (Kézdi and Willis, 2003, 2011).

A separate literature gleans expectations of market participants from option price information, following the results from an early paper by Breeden and Litzenberger (1978) which shows that the price density of a state-contingent security can be derived from the prices of European options. A variety of functions for the price of an option are admissible in this context; the most often used one is the option pricing model described by Black and Scholes (1973) and Merton (1973) and we follow that convention in this paper.

A number of researchers consider the inclusion of survey expectations in models of economic behavior [see Manski (2004) for a survey of this literature] and demonstrate that including probabilistic expectations can improve inference about economic behavior relative to models using only data on economic choices (revealed preference models). Yet others show that survey responses do not exactly align with true expectations – for example, due to large clusters of responses occurring at focal points of the response distribution (e.g., Dominitz and Manski, 1997; Hurd, McFadden, and Gan, 1998; Kleinjans and Van Soest, 2010) – and argue that adjustments to survey data to account for such aspects are necessary to improve inference

(Bassett and Lumsdaine, 2000; Lillard and Willis, 2001). While early literature comparing expectations to economic behavior was in the context of analyzing consumer intentions (e.g., buying a new car, see Juster, 1966), more recent research considers expectations about equity markets and investors' behavior (e.g., Hurd and Rohwedder, 2012).

The paper proceeds as follows: Section 2 describes the data construction and descriptive statistics of the main variables. Section 3 describes the model used to examine the link between Wall Street and Main Street. Section 4 analyzes the regression results and discusses their implications. Section 5 examines whether the Wall Street/Main Street link varies by subgroup and specifically considers the information content of survey responses that are at odds with the laws of probability. The final section concludes. Supplemental material containing information on the sample construction, detailed descriptive statistics, the derivation of the likelihood function and additional analyses is available in the Appendix.

2. Data

The American Life Panel (ALP) provides a novel dataset for our analysis. An internet panel with about 3000 active panel members, the ALP contains more than 200 survey modules (and associated survey weights designed to ensure a nationally representative population) administered by the RAND Corporation. A more detailed description of the ALP sampling frame, survey population, interview length, response rate and participation incentives is in Appendix A1. Throughout the paper, sampling weights are used when reporting descriptive statistics and regression results.

While some of the survey modules are stand-alone, others belong to periodically-repeated series (waves) on the same topic. This paper uses responses obtained from modules designed by Michael Hurd and Susann Rohwedder to investigate the effects of the financial crisis on American households, gathered from November 5, 2008 until March 10, 2011, corresponding to 25 waves of information. Hurd and Rohwedder (2010) provide a detailed description of this series of modules; they are briefly summarized here. The first wave asks respondents about a wide range of topics such as labor force status, stock ownership, mortgage payments and expectations about the future. The second wave was conducted in February 2009 and subsequent waves have been conducted monthly from May 2009 onwards. Each module also contains demographic control variables such as age, race, gender, marital status, and education. The final sample (after adjustments for, e.g., missing observations) consists of 47,488 surveys from 2,652 respondents (94.9% of the total number of surveys and 98.3% of the total number of respondents) gathered over 364 survey days. The sample construction is further detailed in Appendix A1.

2.1 What Main Street thinks: survey expectations about stock market returns

As a proxy for the views of "Main Street", the ALP elicits expectations about the stock market from survey participants via a series of questions, the first of which is the following (hereafter referred to as the "*PositiveReturn*" question):

> *"We are interested in how well you think the economy will do in the future. On a scale from 0 percent to 100 percent where "0" means that you think there is absolutely no chance, and "100" means that you think the event is absolutely sure to happen, what are the chances that by next year at this time mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today?"*

Respondents can give an answer ranging from zero to one hundred (the answer need not be an integer) to indicate the percentage chance of the event happening, or they can leave the response blank. If a numerical response is not recorded to this question, the question is asked a second time, preceded by the additional instruction "You did not answer. Your answers are important to us. Please give us your best guess." To this second question, either (a) a response ranging from zero to one hundred, or (b) "don't know" is recorded.

The same structure is repeated for two additional questions, asking respondents to assess the chances of a greater than 20% return and a greater than -20% return.[1] For expositional ease, the questions referring to the probability of a positive return, a more than 20% return, and a more than -20% return will be referred to as *PositiveReturn, >Plus20,* and *>Minus20,* respectively.

---

[1] The exact wording of these questions is: "*By next year at this time, what are the chances that mutual fund shares invested in blue-chip stocks like those in the Dow Jones Industrial Average will have increased (fallen) in value by more than 20 percent compared to what they are worth today?"* Because the probability of a >20% decrease in value is equal to one minus the probability of a more than -20% return, the response is subtracted from one hundred percent to correspond to a greater than -20% return. This naming convention will be useful in the analysis when comparing the subjective response values to expectations inferred from option prices.

Using all three questions (when available) from the 47,488 surveys yields a total sample size for studying Main Street probabilities of 139,327 observations.

The phrasing of these questions may lead to differences in respondents' interpretation and hence the answers they give, since there is an implicit subjectivity associated with respondents' understanding of "mutual funds shares" or "blue chip stocks like those in the Dow Jones Industrial Average (DJIA)". For the purposes of this paper, however, it is necessary to assume that the responses given represent respondents' subjective probability that the nominal (not inflation-adjusted) level of the DJIA in one year will be greater (similarly, more than 20% greater, more than -20% greater) than the current level of the DJIA. For each respondent, the current level of the index is assumed to be the closing level on the most recent business date prior to the date of interview, so that the response is assured to chronologically follow the information on which the Wall Street probabilities are based.

FIGURE 1 HERE

Figure 1 shows a histogram of the frequency of specific responses to each of the three probabilistic questions individually, as well as of the responses combined ("Aggregate"). Most of the responses are integers -- only 41 out of 139,327 responses are non-integer. Further, responses appear to be clustered around certain focal points, a common occurrence in survey data that contains probabilistic subjective response questions such as these.[2] For the three questions in this paper, 93.8% of person-wave responses are a multiple of five and 68.0% of responses are a multiple of ten. A response of 50 occurs 19.9% of the time; 3.5% of the responses are zero and 3.1% are one hundred. In addition, 63.0% of the 8,701 responses that are not multiples of five are between zero and five or between 95 and one hundred. Due to the pileup of responses at 50 and the extreme nature of the values zero and one hundred, in the model section, the likelihood of these three responses is modeled explicitly. All subsequent references to "focal responses" in this paper will mean responses of these three values (zero, 50, and one hundred) only, following Lillard and Willis (2001) and Kézdi and Willis (2003). In

[2] See, for example, Hurd, McFadden, and Gan (1998), Bassett and Lumsdaine (2000), Lillard and Willis (2001), Hurd and McGarry (2002), Kézdi and Willis (2003), Manski (2004), Huynh and Jung (2010), Kleinjans and Van Soest (2010), and Dominitz and Manski (2011).

addition, a number of articles (e.g., Fischhoff and Bruine de Bruin, 1999) note that a response of "50" should be considered as distinct from other focal responses as it likely indicates uncertainty on the side of the respondent rather than a true subjective probability of 50%; this is also accounted for in the model. The mean response to each of the three questions, as well as the proportion of responses that are focal, varies by respondent characteristics (see Appendices A2 & A3 for more details), consistent with past studies that show heterogeneity across individuals with respect to their expectations about future equity returns.[3] This observation, also noted in Manski and Molinari (2010), motivates the decision to explicitly model the probability of giving a focal response as a function of observable covariates in the model, in addition to including demographic controls.

2.2 What Wall Street thinks: calculating option-implied probabilities

The three return thresholds given in the ALP questions (-20%, 0%, 20%) correspond precisely to strike price levels of a European call option, namely the 20% in-the-money, at-the-money, and 20% out-of-the-money thresholds. We therefore turn to the option-pricing literature to derive analogous Wall Street probabilities for comparison to those reported by Main Street respondents in the ALP. While we recognize that there are numerous ways to derive such probabilities, in this paper we adopt a fairly basic approach so as not to obscure the main question of interest (whether there is a relationship between Wall Street and Main Street beliefs).

In a risk neutral setting the probability that the price of a security will be above a strike price *K* at a future time *T* when the security trades at price S_t at time $t < T$ is given by (Hull, 1989, p. 251):

$$P(S_T > K|S_t) = \Phi(d_2) = \Phi\left(\frac{\ln\left({S_t}/{K}\right)+\left(r-q-\frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right) \quad (1)$$

with Φ the standard normal cumulative distribution function, *r* the (continuous) risk-free rate over the period *[t,T]*, *q* the (continuous) dividend rate over the same period and σ the volatility of the return on the security. To account for risk aversion, for computation of "Wall Street" probabilities we adjust the risk neutral setting by including an equity risk premium, ρ:

[3] An earlier version of the paper contained these results; in the interest of space they are omitted here but are available from the authors on request.

(2) $$P(S_T > K|S_t) = \Phi(d_2) = \Phi\left(\frac{\ln\left({S_t}/{K}\right)+\left(\rho+r-q-\frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right)$$

Although the price of the underlying security, interest rate and dividend yield (or estimates thereof) are available, the volatility cannot be directly observed. Since option prices are observable for the DJIA, however, the Black-Scholes equation can be used to solve for the volatility. A set of implied volatilities (with elements that vary according to day of interview and strike level) then can be used to derive the set of option-implied probabilities.

The Wall Street probability computation proceeds as follows, using the explicit formula for the price of an option as described in Black and Scholes (1973) and Merton (1973), adjusted for an equity risk premium ρ that is fixed at 6%, corresponding to the average annual risk premium over the period 1961-2011.[4] For each day that a survey was answered (364 days in total), the values of the parameters are extracted from Bloomberg® for the specific case of one-year options on the DJIA (a detailed description of how the parameters were obtained using Bloomberg® can be found in Appendix B). The interest rate (r) is the (continuous) U.S. dollar swap rate over the period $[t,T]$, the dividend rate (q) is initially set to zero since the DJIA is dividend-adjusted (sensitivity to these assumptions, as well as the choice of equity risk premium, is explored in the first section of Appendix E). The volatility (σ) for each specified strike price is the volatility implied by the option prices. In particular, implied volatilities for a time to expiration (T-t) of one year and strike prices (K) of 80%, 100%, and 120% of the level of the index at time t were constructed, consistent with the time horizon and return categories articulated in the ALP survey questions and corresponding to the questions *>Minus20*, *PositiveReturn*, and *>Plus20*, respectively.

2.3 Comparing Main Street to Wall Street

To compare Main Street probabilities to Wall Street, each of the 139,327 Main Street observations is first assigned a corresponding option-implied probability associated with the date before the day the interview was conducted (sensitivity to the choice of return date is examined in Appendix E3). Specifically, all individuals that were interviewed on a given day are assigned the same Wall Street probability, and the number of Wall Street probabilities assigned to a

[4] This is computed from the Fama-French "excess return on the market" factor, downloaded from Kenneth French's website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

specific person corresponds to the number of waves in which the person provided a Main Street probability. Table 1 contains summary statistics for the three Main Street and Wall Street probabilities, aggregated across all observations. Not surprisingly, the average probability associated with *>Plus20* is lower than the probability associated with *PositiveReturn,* which in turn is lower than the probability associated with *>Minus20.*

TABLE 1 HERE

There is a relatively close correspondence between the means of respondents' answers to the probabilistic questions and the mean of the option-implied probabilities associated with those same questions. This is especially the case with respect to the upper return threshold on the DJIA (the row labeled ">20%"); on average there is little difference between Wall Street (24.0%) and Main Street (27.1%) regarding the expectation that the one-year return on the DJIA will exceed 20%. There is more of a difference when considering the probability that the DJIA will increase (the row labeled "*>0%*"), with an average 40.4% Main Street probability of a positive return versus a Wall Street average of 57.1% from the corresponding option-implied probabilities. There is also a divide between the two measures when it comes to the probability of a more than 20% return in the DJIA (the row labeled "*>-20%*"), with the survey responses markedly more pessimistic (the mean of *>Minus20* is 75.8%) than the average option-implied probability (83.0%).[5]

Looking solely at the means across all three strike levels is of course not sufficient to draw conclusions as to whether survey respondents' and the market's beliefs coincide, despite similar patterns that show respondents assigning relatively higher probabilities to large changes in the level of the index. Comparing standard deviations, Main Street probabilities inherently have greater variation than can be explained by Wall Street probabilities alone, further motivating the need for a formal model that incorporates additional covariates. Standard deviations for Main Street (20.0%, 26.8%, and 21.6% for *>Minus20, PositiveReturn,* and *>Plus20*, respectively) are very large both as a proportion of the bounded range of 0-100% and in comparison to those for Wall Street (4.9%, 2.9%, and 4.3%, respectively). This is partly a

[5] This pattern (of a greater difference between probabilities of a return larger than -20% than between probabilities of a return larger than 20%) could be related to the volatility smirk that is often observed in index option data, namely that the Wall Street probability reflects higher demand for in-the-money call options (e.g., the ">-20%" option) than deep out-of-the-money call options (e.g., the ">20%" option).

result of the analytical design, since all participants on the same day are assigned the same Wall Street probability.

3. Model

We use a generalized linear model [see McCullagh and Nelder (1989) for an extensive description] to jointly model all three Main Street probabilities. Let X represent the matrix of data (covariates) available to the econometrician. Respondents' unobserved true belief $p*$ is assumed to be related to a linear combination of a subset of covariates $X_1 \in X$, through a 'link function' $f()$ such that

$$f(p^*) = X_1\beta_1 \quad (3)$$

where more generally for any i, a matrix X_i denotes a subset of the covariate matrix X and β_i is a vector of parameters corresponding to the columns of X_i.

There is evidence to suggest that respondents report their belief with error, however. As noted earlier, a substantial fraction of respondents give a focal response of zero, 50, or one hundred. Such a focal response might be the result of a lesser ability to express oneself in probabilistic terms. Similar to Hurd, McFadden and Gan (1998), these responses are modeled via a latent variable $w*$:

$$w^* = X_2\beta_2 + \eta \quad (4)$$

with η an error term. A non-focal answer is given if and only if $w* > 0$. Respondents report their true belief $p*$ with error. In the absence of a focal tendency ($w* > 0$), their response is a random variable, $\tilde{p}$, for which $E[\tilde{p}] = p^*$ holds. When the latent variable $w* \leq 0$, respondents instead give a focal response of zero, 50, or one hundred.

A further distinction is made between the focal response of 50 and a focal response of zero or one hundred. This distinction is motivated both by previous literature that suggests responses of 50 often indicate uncertainty on the part of the respondent (Fischhoff and Bruine de Bruin, 1999) and by the prevalence of responses of 50 in the sample (19.9% of all responses compared to 6.6% for zero and 100 combined).[6] To account for this possible uncertainty the model includes a third equation that describes an additional latent variable v^*:

$$v^* = X_3\beta_3 + \xi \tag{5}$$

with ξ an error term. Conditional on a focal response being given ($w^* \leq 0$), a response of 50 represents uncertainty and is given if and only if $v^* > 0$. When $v^* \leq 0$, respondents with a tendency to rely on focal responses feel certain and give a response of zero or one hundred percent (represented in the model as a probability of zero or one). In this case, the error with which the respondent reports his/her true belief is governed by an endogenously-determined cutoff value (ψ) that pushes the response to either of the two extreme endpoints, depending on where their belief lies relative to this constant threshold:

$$\begin{aligned} p &= 0 \quad \text{if} \quad \tilde{p} \leq \psi \\ p &= 1 \quad \text{if} \quad \tilde{p} > \psi \end{aligned} \tag{6}$$

The error terms η and ξ are assumed to be independently normally distributed with mean 0 and variance 1, as they are identified only up to scale.

The link function *f (.)*, that describes the relationship between $X_1\beta_1$ and the true beliefs p^*, must be chosen from the set of functions with range equal to the admissible values of $X_1\beta_1$ (i.e., the real line) and domain *[0,1]*. We use the inverse of the logistic function (the logit) in our model since it is the most commonly used function for binary data (see, e.g., Albert and Chib, 1993). With this link function, the true beliefs are given by:

$$p^* = f^{-1}(X_1\beta_1) = \frac{1}{1+\exp(-X_1\beta_1)} \tag{7}$$

Besides the linear part ($X_1\beta_1$) and the link function *f (.)*, the third part of any generalized linear model is a stochastic component. In the context of this paper, the stochastic component enters through the subjective responses $\tilde{p}$ (true belief p^* with error) that are assumed to come from a beta-distribution. The beta function is well suited for describing probabilities or proportions because it is defined on the unit interval, and has a flexible functional form that allows for a wide variety of shapes (e.g., Law and Kelton, 1982, pp.165-167). It has been used

[6] Large proportions of focal responses at 50 also have been documented in other surveys by Hurd, McFadden, and Gan (1998), Bruine de Bruin, Fischbek, Stiber and Fischhoff (2002), and Manski and Molinari (2010).

to model probabilistic responses in Bruine de Bruin et al. (2002) and earlier, by Winkler (1967). The probability density function is given by Mendenhall, Scheaffer and Wackerly (1981, p.632):

(8) $$f(p|\alpha_1, \alpha_2) = \frac{p^{\alpha_1-1}(1-p)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)}$$

for values $0 \le p \le 1$ and shape parameters α_1, $\alpha_2 > 0$. $B(\alpha_1,\alpha_2)$ (the beta function) normalizes the above density so that the cumulative density is equal to 1 at $p = 1$:

(9) $$B(\alpha_1, \alpha_2) = \int_{t=0}^{1} t^{\alpha_1-1}(1-t)^{\alpha_2-1}dt = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$$

with Γ the gamma function.

The mean and variance of the distribution are given by μ and υ, respectively:

(10) $$\mu = \frac{\alpha_1}{\alpha_1+\alpha_2} \qquad \upsilon = \frac{\mu(1-\mu)}{1+\alpha_1+\alpha_2}$$

Similar to the Bernoulli distribution, the variance is equal to the mean times one minus the mean, except it is additionally divided by ($1 + \alpha_1 + \alpha_2$). For ease of interpretation of the results, following Paolino (2001), the estimation considers a reparameterization of α_1 and α_2 in relation to the mean μ and a dispersion factor φ, defined as $\alpha_1 + \alpha_2$. The relationship between the parameters μ and φ and the underlying beta parameters α_1 and α_2, is the following:

(11) $$\alpha_1 = \mu\varphi \qquad \alpha_2 = (1-\mu)\varphi$$

As described earlier, the expected value of $\tilde{p}$ is equal to p^*, hence

(12) $$\mu = E[\tilde{p}] = p^* = \frac{1}{1+\exp(-X_1\beta_1)}$$

The complete model therefore consists of a system of three equations; we estimate it via maximum likelihood (details of the likelihood calculation are contained in Appendix D). Sensitivity to the v^* equation is examined in Appendix E2.

4. Results

The results from the estimation of all three equations are combined in Table 2. Because the model is highly nonlinear, the discussion of the results focuses on the marginal effects, reported in the third column of each group of estimates. In addition, unless otherwise noted, inferences are drawn with reference to statistical significance at the 5% level of significance (indicated in **bold** in the table). The first three columns of the table pertain to the subjective probability ("Main Street") assessment (equation (3)). The next three columns describe the likelihood that a respondent gives a focal response (equation (4)) and the final three columns describe the likelihood that a respondent gives a response of 50, conditional on giving a focal response (equation (5)). For the most part, the three equations include many of the same variables: demographic controls (i.e., gender, age, race, education, and marital status), dummy variables for whether the respondent owns a home, owns stocks, or has a retirement account (as a proxy for wealth and general financial wellbeing), measures of historical stock returns (i.e., over the past 30 days and over the past year) to capture possible adaptive expectations, proxies for stock market knowledge (i.e., a self-assessment of how closely the respondent follows the stock market and their understanding of it), dummy variables to distinguish responses across the three thresholds (>*Minus20* is the omitted category), as well as interactions between these and historical stock returns (to allow for the possibility that historical returns influence the different subjective probabilities differently) and wave dummy variables (not shown).

4.1. The subjective probability

In the subjective probability equation, the option-implied probability is the main variable of interest. The coefficient on this variable measures the extent to which Wall Street expectations (as measured by these probabilities) influence Main Street expectations (as proxied by the dependent variable, the subjective probabilities). The additional parameter ψ indicates the threshold value below which respondents are estimated to choose a response of zero rather than one hundred (when they respond with a focal answer and do not give a response of 50). The dispersion factor φ is inversely related to the variance of the fitted beta distribution describing people's responses.

As expected, the coefficients on the two dummy variables are negative, with marginal effects -0.366 for *PositiveReturn (*compared to *>Minus20)* and -0.450 for *>Plus20*. These differences are close to those between the average responses given in Table 1 (-0.354 and -0.487, respectively). Respondents who are female, older, Hispanic/Latino, are working, or are homeowners provide lower subjective probabilities while those who have higher educational attainment, own stocks or have a retirement account provide higher probabilities.

There is some evidence that, contrary to the familiar adage, past performance *is* an indicator of future expected returns; the 0.072 marginal effect of the past year's return implies that for each additional 10 percentage point return in the stock market over the past year, respondents' probabilities to the *>Minus20* question are on average 0.72 percentage points higher. Similarly, the marginal effect on the interaction of the past year's return with the *PositiveReturn* dummy suggests that respondents' probabilities are on average 0.39 percentage points higher (0.72 - 0.33) when the past year's return is 10 percentage points higher. There is no significant effect on the interaction of the past year's return with the *>Plus20* variable. More recent stock returns (over the past 30 days) do not appear to have a significant effect on the subjective probabilities.

TABLE 2 HERE

In addition, following or understanding the stock market appears to influence the subjective responses. The estimated probabilities of those that profess to have a good understanding of the stock market are on average 1.7 percentage points higher than for those who report only some understanding of the stock market, while they are 1.2 percentage points lower for those who admit to having a bad understanding. In addition, those that say they are not at all following the stock market are more pessimistic, with estimated probabilities on average 2.6 percentage points lower than those who are only somewhat following the stock market. Interestingly, those who claim to be closely following the stock market also are more pessimistic, perhaps reflecting the sample time frame (i.e., the aftermath of the financial crisis).

The coefficient on the option-implied probability is statistically significant, suggesting that the views of Main Street are indeed influenced by the views of Wall Street. The marginal effect of 0.114 implies that a ten percentage point increase in Wall Street's probability on average increases Main Street's probability by just over one percentage point. The effects may still vary substantially across individuals, for example, according to an individual's level of probabilistic understanding. We consider this possibility in Section 5.

4.2 The propensity to give a focal response

The second equation models the probability of a non-focal response (i.e., a negative coefficient indicates a higher probability of giving a focal response). It is assumed that any association between the financial controls and the probability of a focal response occurs through their correlation with the other controls, i.e., observed demographic factors such as gender, age, race, and educational attainment or the self-assessment regarding following/understanding the stock market. As a result, both the wealth/financial variables (e.g., homeownership, working for pay, stock ownership, and having a retirement account) and historical stock market returns are excluded from this equation.

Consistent with Figure 1, the results indicate that respondents are 7.5 percentage points more likely to give a focal response to the central *PositiveReturn* question than to the more extreme questions. In addition, the propensity to provide a focal response is higher for women and those with lower educational attainment and lower for those who are white, black, older or married. Among the demographic controls, education is the strongest predictor of the probability of a focal response; compared to those whose education did not go beyond the high school level, those with a bachelor's degree are 5.6 percentage points less likely, and those in the highest level (education beyond a bachelor's degree) are 9.7 percentage points less likely to give a focal response.

Consistent with intuition, those who admit to either not following the stock market or having a bad understanding of the stock market are substantially more likely to give a focal response than those in the omitted categories of somewhat following or having a moderate understanding (3.9 percentage points and 5.0 percentage points, respectively), although the marginal effects are lower than those associated with educational attainment.

4.3 Modeling uncertainty

The third equation (for *v**) models the probability that conditional on giving a focal response, the response is 50 (rather than zero or one hundred). A positive coefficient indicates a greater use of 50. Recall that a response of 50 could indicate extreme uncertainty (Bruine de Bruin et al., 2002). The covariates included in this equation are the same as those in the subjective probability equation, with one exception. Because implied volatility is more typically associated with market uncertainty than option-implied probability, it replaces the option-implied probability in the equation. Its coefficient then measures the extent to which Wall Street uncertainty is related to the Main Street uncertainty level of 50 (similar to how the option-implied probability corresponds to the Main Street probability in the first equation).[7]

Disappointingly, there appears to be no evidence that the implied volatility influences the propensity to use an uncertain focal response versus the extreme responses of zero or 100. This may be for a number of reasons, including the possibility that the implied volatility is not an appropriate proxy to use for measuring the kind of uncertainty that would manifest itself in a focal response of 50, e.g., perhaps implied volatility is an instantaneous (daily) measure and focal propensities are more static or the wording of the survey question leads to an interpretation that goes beyond the DJIA (on which the implied volatility is based).

Responses of 50 are 4.3 (4.4) percentage points more common in response to the *PositiveReturn* (*>Plus20*) question than to the *>Minus20* question, conditional on a focal response. This suggests there is greater certainty (or stronger views) about the *>Minus20* question. Women, those who are working, homeowners, stock owners and those with a retirement account are more likely to use focal responses to demonstrate uncertainty rather than certainty; in contrast, there is evidence of increasing certainty with age.

Recent stock market performance (the historical return during the past 30 days) is negatively associated with the propensity to rely on a focal response of 50 versus the extreme responses. In particular, a one percentage point *lower* stock market return over the past 30 days corresponds to an average 0.47 percentage point higher probability that a response of 50 is given

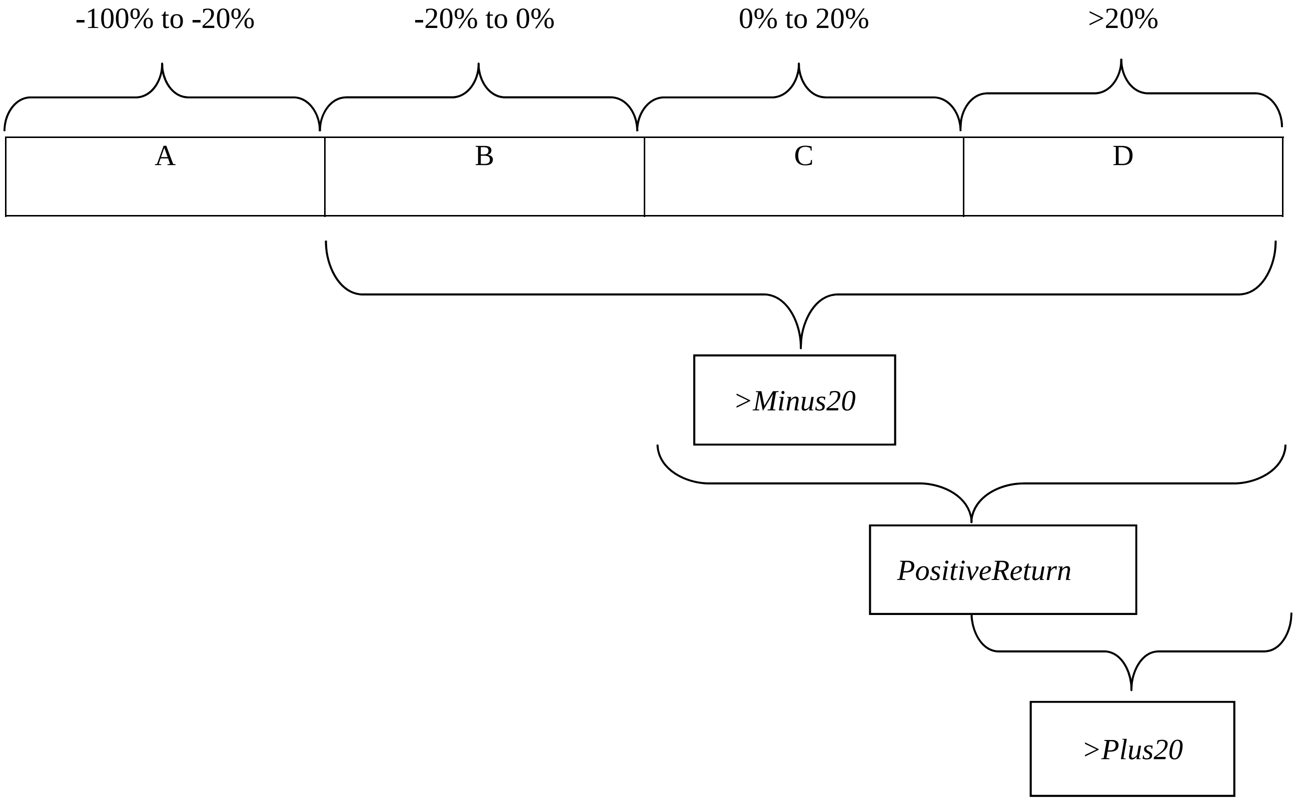
[7] Throughout this paper, we characterize the choice of 50 versus zero and one hundred as one of expressing uncertainty versus absolute certainty. Other characterizations are also possible, for example: lack of confidence versus complete confidence; indifference versus extreme optimism or pessimism; neutrality versus opinionated.

to the >*Minus20* question, rather than zero or 100. This result is consistent with the intuition that during the sample (post-financial crisis) period, stock market declines induced greater uncertainty (i.e., a greater likelihood that the focal point was 50) than stock market gains. It also corroborates findings of Hudomiet, Kézdi and Willis (2011), using data that mostly preceded our sample (from February 2008 to February 2009), that uncertainty increased temporarily following the 2008 stock market crash. In contrast, the probability of a response of 50 to the >*Plus20* question increases on average by only 0.19 percentage points (= -0.466 + 0.278) following a one percentage point lower stock market return.

The results also indicate that those who admit to not following the stock market at all are four percentage points less likely (than those who report somewhat following) to give a response of 50 when providing a focal response than a response of either zero or one hundred. Curiously, those who report having a good understanding of the stock market and those who report to follow the stock market closely are also less likely to reveal uncertainty (2.9 and 4.2 percentage points, respectively).

5. Inconsistent Responses

In addition to the challenge of focal responses, addressed via the model, a relatively large proportion of person-wave responses are inconsistent with the laws of probability. This naturally raises the question of the extent to which such responses represent a respondent's true beliefs and whether this affects the estimated Wall Street-Main Street connection. The following illustration helps explain the concept of (in)consistency with the laws of probability:



This figure shows the complete range of possible returns, from -100% to +∞, divided according to segments that correspond to the survey questions. When respondents answer >*Minus20*, they are being asked to state their probability of the return being outside of section A (i.e., in sections B, C, or D). Similarly, *>Plus20* refers to section D and *PositiveReturn* refers to the probability of the return being in the union of sections C and D. An individual's set of responses for a specific wave is inconsistent with the laws of probability if the answer to *>Plus20* is greater than that of *PositiveReturn* (since D is a subset of the union of C and D) or when their response to *PositiveReturn* is greater than *>Minus20*. Under this definition, inconsistent sets of responses were given in 17.6% of the surveys.[8]

In addition to the inconsistent person-wave sets, in 40.0% of the surveys at least one of the four line sections shown above was implicitly assigned a probability of zero (henceforth the

[8] This proportion is likely an underestimate of the true proportion of inconsistent responses since the survey design precludes the respondent from assigning a positive probability to the region where a zero probability is implied by their initial PositiveReturn response. Specifically, when the response to *PositiveReturn* was 0 or 100, only one of the other two subjective response questions was asked.

responses in these surveys will be called "near-inconsistent" sets since an individual assigns a probability of zero to a range with positive measure). For example, a respondent answered 60 to *PositiveReturn* and 60 to *>Plus20*, implying a probability of zero that the return will be in section C (0% to 20% return). Both the inconsistent and near-inconsistent sets appear to be related to giving a focal response: in 36.9% of the surveys in which an inconsistent set of responses was given, a focal answer was reported for at least one of the three questions; similarly at least one focal response was given in 72.5% of the near-inconsistent sets. In contrast, at least one focal response was given in only 26.7% of the consistent sets.

The model is re-estimated with controls included to account for the large proportion of inconsistent and near-inconsistent responses; the results are shown in Table 3. It is evident that controlling for potential inconsistencies strengthens the findings. For most covariates in the main (μ) equation, coefficients are more statistically significant and marginal effects are larger. The effects of past stock returns are attenuated but not qualitatively different.

Importantly, in the equation that specifies the subjective response, the effect of the option-implied probability on the survey response is greatest for those that provide consistent survey responses and almost double the effect estimated in the baseline model (i.e., without controlling for inconsistency); a ten percentage point increase in option-implied probability increases a consistent survey response by over two percentage points, compared to a one percentage point increase in the baseline model.

For those that give near-inconsistent responses, the effect is qualitatively similar to the baseline model, with a ten percent increase in option-implied probability corresponding to a 1.08 percentage point increase in subjective expectations. In contrast, for those that give inconsistent responses to the three questions, the effect of the option-implied probability is negative and significant, with a ten percent increase in option-implied probability corresponding to a 2.19 percentage point decline in subjective probability, on average. Therefore, although the baseline results indicate that across the whole sample, survey respondents' beliefs coincide well with the market's, there are some respondents whose stated beliefs represent significant departures from those of the market.

There is also evidence that those who do not fully understand the laws of probability are more likely to give a focal response (*w** equation). Those that give inconsistent responses are 11.8 percentage points less likely to give a non-focal response while those that give near-inconsistent responses are nearly 48 percentage points less likely, than those who provided consistent responses.[9] In only one case (the coefficient on bachelor's degree in the *w** equation) does a significant coefficient change sign relative to the baseline (Table 2) regression. Notably, nearly all of the race and education effects documented in the paper are attenuated once one controls for response inconsistency, suggesting that much of the variation in the propensity to give a focal response reflects lack of probabilistic understanding. In particular, once we control for inconsistent responses, there are no longer any significant differences by race in the propensity to respond focally. The attenuation is also evident in the variables that capture the respondents' self-assessment with respect to following or understanding the stock market.

While much of the inference regarding the likelihood of providing a response of 50, conditional on giving a focal response, is unchanged (*v** equation), those that give inconsistent responses are 26.5 percentage points more likely to give a response of 50, conditional on answering focally, than the other survey respondents, suggesting greater uncertainty among those that have a more limited probabilistic understanding. In addition, for those sets of responses that were inconsistent, the effect of Wall Street uncertainty (the implied volatility) on Main Street uncertainty is significantly less (nearly 6.9 percentage points for a 10% difference in the implied volatility) than it is for sets of responses that were not inconsistent.

6. Conclusion

Are Wall Street and Main Street beliefs at odds? A novel approach, comparing survey responses to probabilistic questions about future stock market performance with their corresponding option-implied probabilities, investigates one aspect of this question: whether Wall Street expectations have any influence on the views of Main Street. It would appear the answer is yes. We find a significant relationship between the probabilities extracted from option-prices and those elicited from longitudinal survey responses. The results further show

[9] This may be a result of the definitions of "inconsistent" and "near-inconsistent", respectively; recall that anyone who answers 0 or 100 to any of the three questions is automatically classified as "near-inconsistent" since then some segment of the above chart contains zero mass.

that while option-implied probabilities are linked to survey respondents' outlook, the association is far from one-to-one. Specifically, on average a ten percentage point increase in Wall Street's beliefs that future DJIA returns will exceed a given threshold leads to a more than one percentage point increase in Main Street's beliefs. This effect nearly doubles when controls for probabilistic consistency are included in the regression.

We find evidence that in the immediate aftermath of the financial crisis, respondents who purport to have a good understanding of the stock market or whose responses reflect a stronger understanding of probability display greater optimism; both subgroups on average report higher probabilities than others in the sample. In addition, there is evidence of adaptive expectations via the (statistically significantly) positive relationship between the return on the stock market in the past year and the subjective responses.

Despite an association between Wall Street and Main Street probabilities, no significant relationship is found between Wall Street uncertainty (as measured by implied volatility) and Main Street uncertainty (as measured by the likelihood of giving a response of 50% rather than of 0 or 100%, conditional on a focal response). In contrast, our results show that other stock market-related variables (i.e., returns over the past 30 days, lack of understanding of the stock market, and/or admitting to not following the stock market) do significantly influence the aspect of Main Street uncertainty defined by our metric.

The econometric model presented in this paper adjusts for a number of challenges often present in elicitations from surveys, including the pile-up at key focal points and whether a response of 50% should be interpreted as equal probabilities or complete uncertainty. The analysis demonstrates that subjective response elicitations are useful reflections of sentiment regarding the financial markets and are not necessarily at odds with the views of financial market participants as seen through option prices.

A further exploration considers the degree of probabilistic understanding in the set of responses that participants give. Controlling for variation in probabilistic understanding highlights the possibility that focal responses by survey participants reflect not just a greater degree of uncertainty about the *topic* of the question being asked (i.e., future stock returns) but also a lack of understanding about the concept of probability (i.e., uncertainty about the question framing or interpretation). While Wall Street and Main Street are linked, the link is stronger among those that exhibit probabilistic consistency. This suggests an avenue for future research – the association between probabilistic understanding and financial understanding. The results also demonstrate a possible way that observed inconsistencies in survey responses may provide useful information for inference — suggesting caution be exercised before imposing such consistency through the survey design.

References

Albert, James H., and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association* 88, 669-679.

Barber, Brad M., and Terrance Odean (2000), "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance* 55, 773-806.

Bassett, William F., and Robin L. Lumsdaine (2000), "Probability Limits: Are Subjective Assessments Adequately Accurate?" *Journal of Human Resources* 36, 327-363.

Black, Fischer, and Myron S. Scholes (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, 637-654.

Breeden, Douglas T., and Robert H. Litzenberger (1978), "Prices of State-Contingent Claims Implicit in Option Prices," *The Journal of Business* 51, 621-651.

Brennan, Michael J., Henry H. Cao, Norman Strong, and Xinzhong Xu (2005), "The Dynamics of International Equity Market Expectations," *Journal of Financial Economics* 77, 257-288.

Bruine de Bruin, Wändi, Paul S. Fischbeck, Neil A. Stiber, and Baruch Fischhoff (2002), "What Number is 'fifty-Fifty'?: Redistributing Excessive 50% Responses in Elicited Probabilities," *Risk Analysis* 22, 713-723.

Cui, Changrong, and David Frank (2011), "Equity Implied Volatility Surface Computation," Bloomberg Document No. 2056700.

De Long, James Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann (1990), "Noise Trader Risk in Financial Markets," *Journal of Political Economy* 98, 703-738.

Dickey, David A., and Wayne A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427-431.

Dominitz, Jeff, and Charles F. Manski (1997), "Perceptions of Economic Insecurity," *Public Opinion Quarterly* 61, 261-287.

Dominitz, Jeff, and Charles F. Manski (2011), "Measuring and Interpreting Expectations of Equity Returns," *Journal of Applied Econometrics* 26, 352-370.

Fischhoff, Baruch, and Wändi Bruine de Bruin (1999), "Fifty-Fifty = 50%," *Journal of Behavioral Decision Making* 12, 149-163.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante (2010), "Unequal we Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006," *Review of Economic Dynamics* 13, 15-51.

Hudomiet, Peter, Gábor Kézdi, and Robert J. Willis (2011), "Stock Market Crash and Expectations of American Households," *Journal of Applied Econometrics* 26, 393-415.

Hull, John C. (1989), *Options, Futures and Other Derivatives, 4th ed*. Upper Saddle, NJ: Prentice Hall.

Hurd, Michael D., Daniel L. McFadden, and Li Gan (1998), "Subjective Survival Curves and Life-Cycle Behavior," in *Inquiries in the Economics of Aging*, David A. Wise, ed. Chicago, IL: University of Chicago Press.

Hurd, Michael D., and Kathleen McGarry (2002), "The Predictive Validity of Subjective Probabilities of Survival," *The Economic Journal* 112, 966-985.

Hurd, Michael D., and Susann Rohwedder (2010), "Effects of the Financial Crisis and Great Recession on American Households," NBER Working Paper No. 16407.

Hurd, Michael D., and Susann Rohwedder (2012), "Stock Price Expectations and Stock Trading," NBER Working Paper No. 17973.

Huynh, Kim P., and Juergen Jung (2010), "Correcting Focal Point Biases in Subjective Health Expectations: An Application to the RAND-HRS Data," unpublished manuscript.

Juster, Thomas F. (1966), "Consumer Buying Intentions and Purchase Probabilities: an Experiment in Survey Design," *Journal of the American Statistical Association* 61, 658-696.

Kézdi, Gábor, and Robert J. Willis (2003), "Who Becomes a Stockholder? Expectations, Subjective Uncertainty and Asset Allocation," Michigan Retirement Research Center Working Paper No. 2003-039.

Kézdi, Gábor, and Robert J. Willis (2011). "Household Stock Market Beliefs and Learning," NBER Working Paper No. 17614.

Kleinjans, Kristin J., and Arthur van Soest (2010), "Nonresponse and Focal Point Answers to Subjective Probability Questions," Netspar Discussion Paper No. 10/2010-05.

Law, Averill, and David Kelton (1982), *Simulation Modeling and Analysis.* New York, NY: McGraw-Hill.

Lillard, Lee, and Robert J. Willis (2001), "Cognition and Wealth: The Importance of Probabilistic Thinking," Michigan Retirement Research Center Working Paper No. 2001-007.

Lusardi, Annamaria, and Olivia S. Mitchell (2011), "Financial Literacy and Retirement Planning in the United States," *Journal of Pension Economics and Finance* 10, 509-525.

Manski, Charles F., and Francesca Molinari (2010), "Rounding Probabilistic Expectations in Surveys," *Journal of Business and Economic Statistics* 28, 219-231.

Manski, Charles F. (2004), "Measuring Expectations," *Econometrica* 72, 1329-1376.

McCullagh, Peter, and John A. Nelder (1989), *Generalized Linear Models, 2nd ed.* New York: Chapman and Hall.

Mendenhall, William, Richard L. Scheaffer, and Dennis Wackerly (1981). *Mathematical Statistics with Applications, 2nd ed*. Boston: Duxbury Press, 1981.

Merton, Robert C. (1973), "Theory of Rational Option Pricing," *Bell Journal of Economics and Management* 4, 141-183.

Paolina, Philip (2001), "Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables," *Political Analysis* 9, 325-346.

Poterba, James M., and Andrew A. Samwick (1995), "Stock Ownership Patterns, Stock Market Fluctuations, and Consumption," *Brookings Papers on Economic Activity* 26, 295-372.

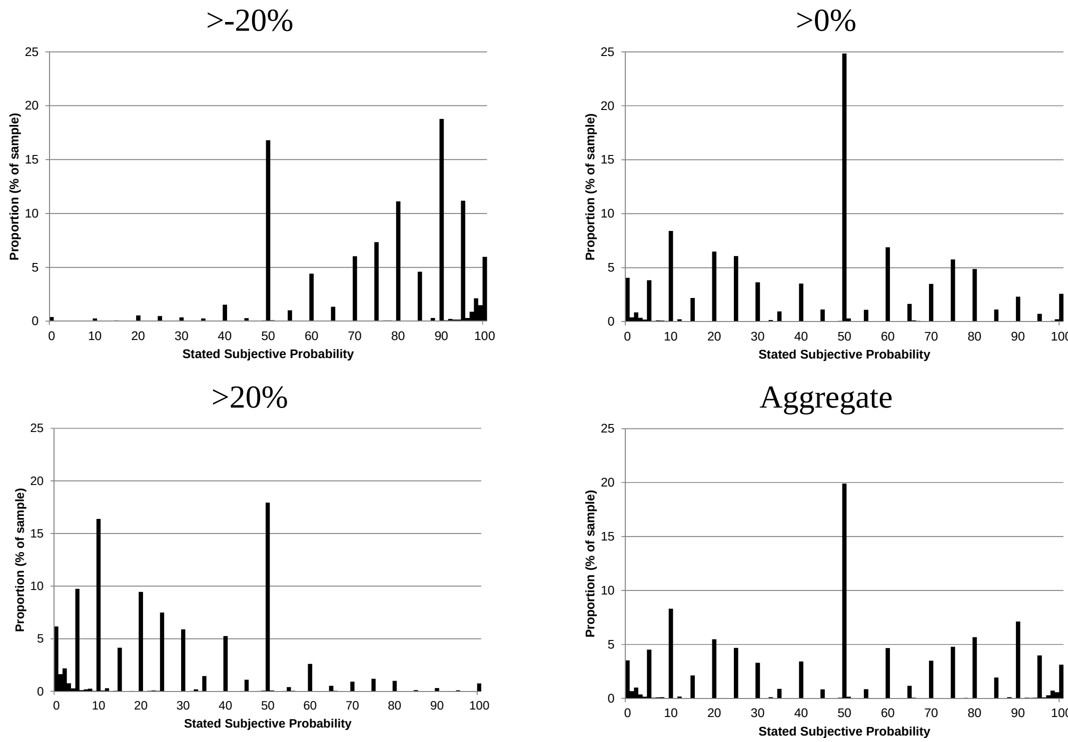
Schwarz, Gideon E. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics* 6, 461-464.

Shiller, Robert J. (2003), "From Efficient Markets Theory to Behavioral Finance," *Journal of Economic Perspectives* 17, 83-104.

Watts, Simon (2010), "OVME <GO> Userguide," Bloomberg Document No. 2052774.

Winkler, Robert L. (1967), "The Assessment of Prior Distribution in Bayesian Analysis," *Journal of the American Statistical Association* 62, 776-800.

Figure 1: *Frequency of responses to probabilistic questions*



Note: These figures contain histograms of the responses to the three questions that ask respondents to consider the probability of a more than -20% return, a positive return, and a more than 20% return, as well as for all three questions combined ("Aggregate"). The responses to the three questions are called *>Minus20, PositiveGain,* and *>Plus20,* respectively, and together comprise the dependent variable. The figures document the large pile-up of responses at focal points, particularly around the response of "50", motivating the econometric model. Observations in these figures are unweighted.

Table 1: Descriptive Statistics of Stated and Option-Implied Probabilities.

Summary statistics for the aggregate sample, computed over all person-wave observations. The rows show means and standard deviations of the three probabilities used: the probability of a more than -20% return (“*>-20%*”), a positive return (“*>0%*”), and a greater than 20% return (“*>20%”)*. The second column shows the number of person-wave observations, the third and fourth show the means and standard deviations of the Wall Street probabilities and the last two columns show these statistics for the Main Street probabilities.

Probability of Return	Observations	Wall Street Mean	Wall Street Std. Dev.	Main Street Mean	Main Street Std. Dev.
>-20%	46,232	0.830	0.049	0.758	0.200
>0%	47,438	0.571	0.029	0.404	0.268
>20%	45,657	0.240	0.043	0.271	0.216

Table 2: Baseline Regression Results

Maximum likelihood estimates of the model are presented in the table. For each variable, the estimated regression coefficient, corresponding standard error and the marginal effect evaluated at the variable means are reported. For dichotomous (binary) variables, the marginal effect is the difference in probability when evaluated at the value of one versus zero, *ceteris paribus.* The first three columns show the results of the equation pertaining to μ, the expected value of respondents' stated subjective probabilities. The second three columns refer to w^*, where $w^*>0$ corresponds to a non-focal response. The last three columns refer to v^*, where $v^*>0$ signifies that a respondent gives a response of 50, conditional on giving a focal response (of zero, 50 or one hundred). A complete set of wave dummies (not shown) is included in each of the three equations. Coefficients in **bold** represent significance at the 5% level.

Observations 139,327 Log likelihood -66,968 Average log likelihood -0.481

	μ			w^*			v^*		
	Coef.	Std. Err.	Marginal	Coef.	Std. Err.	Marginal	Coef.	Std. Err.	Marginal
Dummy (PositiveReturn)	**-1.594**	0.029	-0.366	**-0.220**	0.009	-0.075	**0.138**	0.026	0.043
Dummy (>Plus20)	**-2.055**	0.059	-0.450	-0.002	0.009	-0.001	**0.142**	0.042	0.044
<u>Demographic Characteristics</u>									
Female	**-0.115**	0.006	-0.029	**-0.028**	0.008	-0.009	**0.250**	0.015	0.079
Age	**-0.002**	0.000	0.000	**0.001**	0.000	0.000	**-0.011**	0.001	-0.003
Race									
Non-hispanic white	**0.095**	0.013	0.024	**0.067**	0.017	0.023	**-0.096**	0.034	-0.030
Non-hispanic black	**0.065**	0.016	0.016	**0.049**	0.019	0.016	**-0.148**	0.039	-0.049
Hispanic/Latino	**-0.042**	0.016	-0.010	0.020	0.020	0.007	**-0.171**	0.039	-0.057
Education									
Some college, no Bachelor	**0.085**	0.007	0.021	**-0.059**	0.009	-0.020	**0.078**	0.017	0.024
Bachelor's degree	**0.189**	0.008	0.047	**0.174**	0.010	0.056	**0.202**	0.022	0.061
>Bachelor's	**0.259**	0.010	0.065	**0.315**	0.014	0.097	0.024	0.029	0.007
Married	-0.005	0.006	-0.001	**0.077**	0.008	0.026	-0.014	0.016	-0.005
Working	**-0.046**	0.006	-0.011				**0.061**	0.015	0.019
Home owner	**-0.023**	0.008	-0.006				**0.094**	0.018	0.030
Stock owner	**0.088**	0.007	0.022				**0.164**	0.019	0.050
Have Retirement Account	**0.099**	0.007	0.025				**0.126**	0.017	0.040

Table 2: Baseline Regression Results (cont'd)

Financial Market Characteristics									
Return past 30 days	-0.050	0.189	-0.012				**-1.476**	0.407	-0.466
Return past year	**0.288**	0.110	0.072				0.323	0.278	0.102
Return past 30 days * PositiveReturn	0.043	0.134	0.011				0.515	0.294	0.163
Return past year * PositiveReturn	**-0.132**	0.033	-0.033				-0.054	0.066	-0.017
Return past 30 days * >Plus20	0.266	0.147	0.066				**0.880**	0.318	0.278
Return past year * >Plus20	**-0.320**	0.047	-0.080				-0.022	0.070	-0.007
Following the stock market									
Closely following	**-0.045**	0.014	-0.011	**-0.128**	0.018	-0.044	**-0.130**	0.036	-0.042
Not following	**-0.106**	0.007	-0.026	**-0.115**	0.009	-0.039	**-0.127**	0.018	-0.040
Understanding of stock market									
Good understanding	**0.070**	0.012	0.017	**0.087**	0.016	0.028	**-0.089**	0.034	-0.029
Bad understanding	**-0.050**	0.008	-0.012	**-0.149**	0.009	-0.050	-0.016	0.018	-0.005
Option-implied probability	**0.115**	0.022	0.114						
Implied volatility							0.521	0.579	0.165
Constant	**0.998**	0.044	0.248	**0.726**	0.030	0.242	**0.823**	0.159	0.260
Additional parameters									
ψ	0.463	0.004							
φ	3.909	0.016							

Notes to table: The option-implied probability is transformed using the inverse of the logistic function (logit) analogous to how the Main Street probabilities (dependent variable) are transformed, For ease of interpretation, however, the reported marginal effect corresponding to the option-implied probability is that of the untransformed Wall Street probability on the untransformed Main Street probability. Therefore, a 10 percentage point increase in the Wall Street probability results in a 1.14 percentage point increase in the Main Street probability, *ceteris paribus*. For the variables related to following the stock market and understanding of the stock market, the omitted category is "somewhat following" and "some understanding", respectively. Good understanding is a dummy variable equal to one if an individual rated their understanding as "very good" or "excellent" and zero otherwise while bad understanding is analogously constructed for responses of "poor" or "extremely poor". The parameter ψ determines the threshold at which a focal respondent selects either zero (if their perceived probability is below ψ) or 100 (if their perceived probability is above ψ). The parameter φ measures the dispersion in the beta distribution. A higher φ means a lower variance of the beta distribution.

Table 3: Regression Results Controlling for Inconsistent Responses

Maximum likelihood estimates of the model are presented in the table. For each variable, the estimated regression coefficient, corresponding standard error and the marginal effect evaluated at the variable means are reported. For dichotomous (binary) variables, the marginal effect is the difference in probability when evaluated at the value of one versus zero, *ceteris paribus.* The first three columns show the results of the equation pertaining to μ, the expected value of respondents' stated subjective probabilities. The second three columns refer to w^*, where $w^*>0$ corresponds to a non-focal response. The last three columns refer to v^*, where $v^*>0$ signifies that a respondent gives a response of 50, conditional on giving a focal response (of zero, 50 or one hundred). A complete set of wave dummies (not shown) is included in each of the three equations. Coefficients in **bold** represent significance at the 5% level.

Observations 139,327 Log likelihood -49,111 Average log likelihood -0.352

	μ			w^*			v^*		
	Coef.	Std. Err.	Marginal	Coef.	Std. Err.	Marginal	Coef.	Std. Err.	Marginal
Dummy (PositiveReturn)	**-1.625**	0.029	-0.373	**-0.285**	0.010	-0.090	**0.186**	0.027	0.055
Dummy (>Plus20)	**-2.096**	0.059	-0.457	**-0.034**	0.010	-0.010	**0.139**	0.042	0.041
Demographic Characteristics									
Female	**-0.114**	0.006	-0.028	**-0.040**	0.008	-0.012	**0.259**	0.015	0.078
Age	**-0.002**	0.000	0.000	**0.002**	0.000	0.001	**-0.011**	0.001	-0.003
Race									
Non-hispanic white	**0.095**	0.013	0.024	-0.034	0.018	-0.010	**-0.090**	0.034	-0.027
Non-hispanic black	**0.069**	0.015	0.017	-0.034	0.021	-0.011	**-0.148**	0.039	-0.046
Hispanic/Latino	**-0.039**	0.016	-0.010	0.018	0.022	0.005	**-0.171**	0.040	-0.054
Education									
Some college, no Bachelor	**0.086**	0.007	0.021	**-0.086**	0.010	-0.027	**0.086**	0.017	0.025
Bachelor's degree	**0.188**	0.008	0.047	**-0.037**	0.012	-0.012	**0.202**	0.022	0.058
>Bachelor's	**0.261**	0.010	0.065	**0.056**	0.015	0.017	0.014	0.030	0.004
Married	-0.005	0.006	-0.001	**0.032**	0.008	0.010	-0.023	0.016	-0.007
Working	**-0.051**	0.006	-0.013				**0.071**	0.015	0.021
Home owner	**-0.026**	0.008	-0.006				**0.103**	0.019	0.031
Stock owner	**0.087**	0.007	0.022				**0.158**	0.020	0.046
Have Retirement Account	**0.099**	0.007	0.025				**0.115**	0.018	0.035

Table 3: Regression Results Controlling for Inconsistent Responses (cont'd)

Financial Market Characteristics									
Return past 30 days	0.035	0.186	0.009				**-1.516**	0.410	-0.456
Return past year	**0.245**	0.108	0.061				0.431	0.281	0.130
Return past 30 days * PositiveReturn	-0.006	0.133	-0.001				**0.655**	0.295	0.197
Return past year * PositiveReturn	**-0.114**	0.032	-0.028				-0.035	0.066	-0.011
Return past 30 days * >Plus20	0.167	0.145	0.042				**0.891**	0.321	0.268
Return past year * >Plus20	**-0.283**	0.046	-0.070				0.007	0.071	0.002
Following the stock market									
Closely following	**-0.044**	0.014	-0.011	**-0.110**	0.020	-0.035	**-0.147**	0.037	-0.046
Not following	**-0.106**	0.007	-0.026	**-0.027**	0.010	-0.008	**-0.113**	0.019	-0.034
Understanding of stock market									
Good understanding	**0.073**	0.012	0.018	**0.045**	0.018	0.014	**-0.087**	0.034	-0.027
Bad understanding	**-0.054**	0.008	-0.013	**-0.045**	0.010	-0.014	-0.003	0.018	-0.001
Consistent *OIP	**0.212**	0.022	0.211						
Near-inconsistent * OIP	**0.109**	0.022	0.108						
Inconsistent *OIP	**-0.220**	0.022	-0.219						
Implied volatility							0.834	0.584	0.251
Inconsistent * Implied volatility							**-2.287**	0.401	-0.687
Constant	**1.047**	0.043	0.260	**1.473**	0.034	0.452	**0.649**	0.161	0.195
Inconsistent				**-0.357**	0.012	-0.118	**1.280**	0.109	0.265
Near-inconsistent				**-1.483**	0.009	-0.477			

Additional parameters	ψ	0.463	0.004	φ	4.084	0.017

Notes to table: See notes to Table 2. In addition dummy variables for inconsistency (=1 if an individual's set of survey responses is inconsistent with the laws of probability and zero otherwise) and near-inconsistency (=1 if an individual's set of survey responses implies zero probability over a measurable set of the probability space and zero otherwise) are included in the regression, as well as interacted with the main variable of interest, the option-implied probability (OIP in the table). An interaction between the inconsistency dummy and implied volatility is also included in the third equation of the model.

Appendix A: Construction and Description of Main Street Information

We use publicly available data from the RAND American Life Panel (ALP) (https://mmicdata.rand.org/alp). The Household information module[10] contains a number of demographic control variables such as age, gender, and race for the respondents to the two repeated surveys that we use, the "Monthly Survey" and the "Effects of the Financial Crisis". The latter survey is conducted every three months, with the "Monthly Survey" (a shorter version of the "Effects of the financial crisis" survey) conducted in the intervening two months. As of the writing of the paper, data from November, 2008 to March, 2011 (25 waves of data) was publicly available through the RAND website, with data through March, 2012 embargoed. Both of the repeated modules used in the paper are ongoing; as a result, new modules are added each month. After March, 2011 (wave 25), however, changes were made to the sample design that were beyond our control (i.e., the sample size was reduced and a portion of the reduced sample was not asked the subjective response questions that comprise our main variable of interest). For this reason, we have not used subsequent waves of the sample in our analysis.

The ALP website states the following about the duration of the interviews and incentives for respondents:

> "*Typically an interview will not take more than 30 minutes. Respondents are paid an incentive of about $20 per thirty minutes of interviewing (and proportionately less if an interview is shorter)*"

Sampling weights were constructed by RAND so as to match the distribution of the US adult population as reported in the Current Population Survey with regards to gender, age, race, income and education. Unless stated otherwise, sampling weights are used when reporting descriptive statistics and regression results.[11]

A1. Sample construction

The sample construction is detailed in Table A1. A total of 50,029 surveys were initiated by 2,699 respondents across 25 waves. For each wave, participants are given an approximately two-week window during which to complete the survey. Therefore not every calendar day during the sample period has survey responses associated with it. Of the 857 days between November 5, 2008 and March 10, 2011, inclusive, surveys were taken by these 2,699 respondents on 364 of these days. Out of this total number, 784 surveys were not fully completed and hence are omitted from our sample; an additional five surveys were omitted because the sampling weight was missing.

We also required surveys to contain responses to at least two of the three key variables of interest. The requirement of at least two responses rather than three reflects the sampling design: when respondents answer either zero or one hundred to the *PositiveReturn* question – which accounts for 4.1% and 2.6% of the person-wave responses to this question, respectively, only one of the two other questions is then asked, as a zero response to *PositiveReturn* implies a

---

[10] Accessible at: https://mmicdata.rand.org/alp/index.php?page=data&p=showsurvey&syid=90002

[11] The one exception is counts of observations – these are reported unweighted.

zero probability of a >20% increase and a 100 response implies a zero probability of a >20% decrease. In 660 surveys there were no responses to these three variables while in 94 of the surveys, only one of the three questions was answered.

Table A1: Sample Construction

	Persons	Dropped Surveys	Surveys Remaining	Persons Remaining
Start	0	0	50029	2699
Survey Design				
Haven't finished survey	519	784	49245	2689
Weights missing	4	5	49240	2688
Dependent variable				
Did not answer any key questions	138	660	48580	2685
Answered only 1 (of 3) key questions	73	94	48486	2681
Covariates				
Over 90 at first response	2	27	48459	2679
Gender, Ethnicity, Race, Birth year missing	6	25	48434	2678
Inconsistent	20	383	48051	2658
Gender	*9*	*165*		
Ethnicity	*5*	*99*		
Race	*3*	*45*		
Birth year	*3*	*74*		
Family income missing	9	100	47951	2656
"Holds stocks/stock mutual funds" missing	110	145	47806	2655
"Bought or sold stocks since [timeframe]" missing	67	83	47723	2655
Exact amount bought/sold (follow-up) missing	90	134	47589	2653
"Has retirement account" missing	97	101	47488	2652

Notes to the table: Surveys were dropped sequentially according to a series of filters, in the order that is indicated in the first column. The second column indicates how many persons had at least one survey in which the mentioned criterion is met. Note that since these individuals may have had admissible surveys in other waves, they may still remain in the sample; therefore, the number of persons *deleted* as a result of each filter (reflected in the fifth column) is generally lower than the number of persons listed in the second column. The third column indicates the number of surveys that were omitted as a result of the filter. The fourth and fifth column indicate how many surveys and persons, respectively, remained after applying the filter. An exception to this rule are the observations left out when a person was initially over 90 or gave inconsistent responses with regards to demographic variables; in these cases, all surveys answered by the person were dropped.

A modest age screen is necessary to minimize the small sample bias that could enter into the analysis by including individuals in the tail end of the age distribution. As a result twenty-seven surveys were omitted as the respondent was over 90 when first answering a survey. 588 surveys were excluded because key covariates were missing. Finally, 383 surveys answered by

20 respondents were excluded because of inconsistencies with regards to race, ethnicity, birth year or gender.[12]

After all sampling screens, our final sample contains 2,652 respondents (98.3% of the original respondents) over 364 days, for a total of 47,488 surveys (94.9% of the original surveys). Because in some surveys not all of the three questions were answered, we have a total of 47,438, 45,657, and 46,232 responses to the *PositiveReturn,>Plus20, and >Minus20* questions, respectively, leading to 139,327 responses to the three questions combined.

The questions "How would you rate your understanding of the stock market" and "How closely do you follow the stock market" are not asked in every wave, but in four waves of the "Effects of the Financial Crisis" module (including the first two waves). For the (subsequent) waves where this variable is unobserved, we assume that the respondent's answer remains the same until the next observable response.

[12] A respondent is considered inconsistent if *in more than one wave*, the reported gender, ethnicity, race or birth year is different from that most frequently given by that respondent – e.g., a person is reported to be male in 3 waves and female in the other 10 waves answered. The decision to restrict the exclusion criterion to more than one wave is in recognition of the possibility of an occasional error. Similarly, a difference of one year in reported birth years is not counted as being so different that the observation should be excluded, nor is the reporting of an "other" race when the most frequent response in that person's other surveys is a specific race and vice versa.

A2: Average Main Street Responses, by Various Characteristics

The mean response to each of the three main questions of interest is given in Table A2, for the overall sample and stratified by a variety of characteristics. Columns labeled "1" give the mean for respondents that are a member of the given group; "0" is the mean over the rest of the sample (i.e., those not in the group). Means for the characteristic are shown in **bold** when it is statistically significantly different from the mean of the rest of the sample at a 5% level of significance.

There is evidence of heterogeneity in the Main Street probabilities, according to observable characteristics. For example, those who own stocks, those with higher educational attainment and those who report a good understanding or close following of the stock market on average give more optimistic responses (higher mean responses) than the rest of the sample.

Table A2	*>Minus20*		*PositiveReturn*		*>Plus20*	
Overall	0.758		0.404		0.271	
Is in group (1=yes, 0 = no)	1	0	1	0	1	0
Female	**0.749**	**0.768**	**0.373**	**0.441**	**0.272**	**0.268**
Married	**0.762**	**0.752**	**0.416**	**0.387**	**0.263**	**0.282**
Homeowner	**0.768**	**0.742**	**0.418**	**0.382**	**0.259**	**0.289**
Owns Stocks	**0.775**	**0.751**	**0.478**	**0.373**	**0.275**	**0.269**
Have Retirement Account	**0.771**	**0.741**	**0.446**	**0.351**	**0.264**	**0.280**
Working for pay	**0.754**	**0.763**	**0.411**	**0.395**	**0.273**	**0.267**
Ethnicity						
Hispanic/Latino	0.755	0.758	**0.336**	**0.411**	0.270	0.271
Non-Hispanic White	**0.762**	**0.746**	**0.417**	**0.367**	**0.266**	**0.286**
Non-Hispanic Black	**0.743**	**0.760**	**0.364**	**0.409**	**0.300**	**0.267**
Other	**0.738**	**0.759**	**0.427**	**0.403**	0.283	0.270
Education						
High School	0.756	0.759	**0.334**	**0.450**	**0.260**	**0.277**
Some college, no degree	**0.738**	**0.765**	0.401	0.405	**0.284**	**0.265**
Bachelor's degree	**0.771**	**0.754**	**0.485**	**0.382**	0.273	0.270
Further education	**0.788**	**0.754**	**0.510**	**0.391**	0.268	0.271
Understanding stock market						
Extremely/Very good	0.769	0.758	**0.531**	**0.402**	**0.298**	**0.270**
Somewhat good/poor	0.760	0.758	**0.431**	**0.400**	0.276	0.270
Extremely/Very poor	0.752	0.758	**0.314**	**0.410**	**0.261**	**0.271**
Follow stock market						
Very closely	0.746	0.758	**0.499**	**0.403**	**0.306**	**0.270**
Somewhat closely	**0.765**	**0.757**	**0.464**	**0.398**	**0.283**	**0.269**
Not at all	0.753	0.758	**0.336**	**0.412**	**0.260**	**0.272**

A3. Focal Response Proportions

There are significant differences in the proportion of focal responses (responses of 0, 50, or 100 to the *>Minus20, PositiveReturn, and >Plus20* survey questions) by demographic characteristics (e.g., gender, marital status, work status, educational status) and also according to subjects' self-rated understanding and following of the stock market, as shown in Table A3. Those who report limited understanding/following of the stock market, have lower educational attainment, and non-Hispanic blacks provide the highest proportion of focal responses. The difference in focal proportions among population sub-groups (shown in **bold** when they are significantly different from the other observations at a 5% level of significance) motivates the decision to explicitly model the probability of giving a focal response as a function of observable covariates.

Table A3: Proportion of Focal Responses per Subgroup

Overall	0.281	
	Proportion focal	
Group	In group	Not in group
Female	**0.295**	**0.263**
Married	**0.263**	**0.306**
Houseowner	**0.265**	**0.305**
Owns stocks	**0.236**	**0.299**
Have Retirement Account	**0.242**	**0.329**
Working for pay	**0.276**	**0.286**
Ethnicity		
Hispanic/Latino	**0.307**	**0.278**
Non-hispanic White	**0.274**	**0.300**
Non-hispanic Black	**0.312**	**0.277**
Other	**0.266**	**0.282**
Education		
High School	**0.306**	**0.264**
Some college, Bachelor's degree	**0.322**	**0.264**
Bachelor's degree	**0.228**	**0.295**
Further education	**0.186**	**0.293**
Understanding stock market		
Extremely/Very good	**0.241**	**0.282**
Somewhat good/somewhat poor	0.280	0.281
Extremely/Very poor	**0.379**	**0.274**
Follow stock market		
Very closely	0.273	0.281
Somewhat closely	**0.266**	**0.282**
Not at all	**0.350**	**0.273**

Appendix B: Construction and Description of Wall Street Information

The price at time *t* of a European call option with a strike price of K and an expiry date of $T > t$ is given by:

$$\Phi(d_1)S_t e^{-q(T-t)} - \Phi(d_2)Ke^{-r(T-t)}$$

where

$$d_1 = \frac{\ln\left(S_t/_K\right)+\left(r-q+\frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \qquad d_2 = d_1 - \sigma\sqrt{T-t}$$

The interest rate (r) is the (continuous) U.S. dollar swap rate over the period *[t,T]* (which is the default rate for option price calculations in Bloomberg), the dividend rate (q) is the Bloomberg forecast for the DJIA dividend rate during the same period, the volatility (σ) is the implied volatility corresponding to each specified strike price (relative to the spot price S_t) and a time to expiration $T - t$.

The daily implied volatilities are determined by Bloomberg based on prices from out-of-the-money options. For those strike prices and times to expiration for which options on a particular asset are available (i.e., traded), corresponding implied volatilities can be derived. From these, implied volatilities for other combinations of strike prices and times to expiration can be estimated.[13] In particular, implied volatilities for a time to expiration *(T-t)* of one year and strike prices (K) of 80%, 100%, and 120% of the level of the index at time t were constructed, consistent with the time horizon and return categories articulated in the ALP survey questions and corresponding to the questions *>Minus20*, *PositiveReturn,* and *>Plus20*, respectively. Note that although Bloomberg uses a specific interest rate and dividend rate to calculate the implied volatility, its estimates for these values are based on market prices for observed options. The Black-Scholes model treats the three parameters (r, q and σ) as independently determined, namely a change in one of the three does not affect the other two.

<u>Obtaining parameters from Bloomberg</u>®

In order to obtain the interest rate and dividend rate for the DJIA corresponding to each day of the sample, Bloomberg®'s option pricing screen (OVME DIVA – see below)[14] was used: this screen calculates the price of an option with characteristics specified by the user, and also allows for the calculation of prices for days in the past (see example below). The user can put values in the highlighted sections specifying exactly the terms of the option s/he wants to price and the Bloomberg pricer then automatically inserts the market value of necessary parameters (such as the implied volatility) and calculates the price of the option.

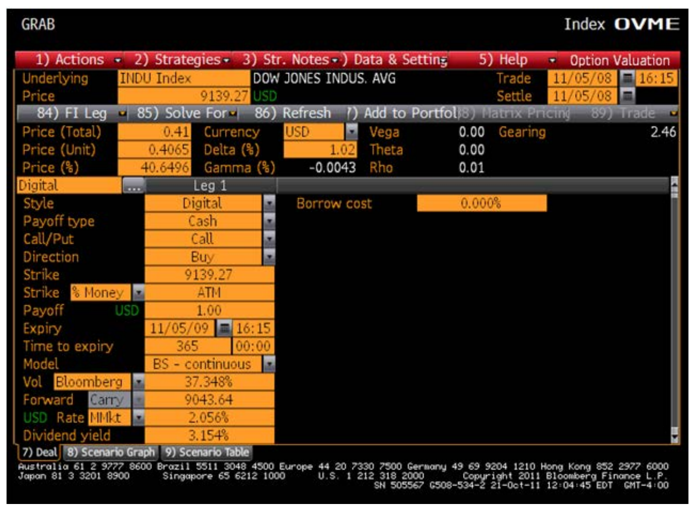
[13] For more information on the calculation of the implied volatility, see: Cui, C. and D. Frank 2011, "Equity Implied Volatility Surface Computation, version 3.6", *Bloomberg document,* 2056700, 1-10. The document can be found by typing DOCS 2056700 <GO> when logged in to a Bloomberg terminal.

[14] For more information on the option pricing screen (OVME), see the most recent user guide at the time of writing, Watts (2010), "OVME<GO> Userguide", *Bloomberg document 2052774, 1-21*. The document can be found by typing DOCS 2052774 <GO> when logged in to a Bloomberg® terminal.

Through the OVME screen, a user can also create a 'deal', whereby certain aspects of the options remain fixed. For simplicity, one such deal was created for the first of every period of 30 calendar days in the period corresponding to the ALP sample. The underlying security was set to DJIA and the expiration date (T) was set at 1 year from this first day. Once a deal is created, a function in the Bloomberg® Excel add-in (BDP) can then be used to download the interest and dividend rate for all 30 days in the interval. The OVME screen does not allow for keeping the *time* to expiration constant, only the expiration *date* can be kept constant. As such, the interest rate that is captured ranges from the (continuous) 1-year interest rate down to the 336-day (1 year minus 29 days) rate. As the latter is not appreciably different from the former, and each survey is available for less than 30 days (approximately 2 weeks), the effects of this simplification are minimal.

The 12-month implied volatility was gathered using Bloomberg®'s historical price add-in (BDH) for Microsoft Excel®, for DJIA options with a strike price of 80%, 100% and 120% of the level at closing for each day that surveys were answered.[15]

Figure B1: Example of the Bloomberg® OVME DIVA screen



The boxes in the two top rows show the underlying security and the price thereof, along with the day at which the price of the option is to be calculated. The boxes in the next three rows show the price of the option and other values characteristic to an option. The boxes below can be used to specify exactly the characteristics of the option for which the price is to be calculated. Bloomberg automatically fills in the (historical) market values for the implied volatility, interest rate and dividend rate (manual override of these values is possible but is not done in this study). The boxes are linked so that changing the value in one box may cause other values to change.

Two comments regarding the values of the interest rate and dividend series are in order:

(1) Although the choice of a 6% equity risk premium (ρ) reflects historical levels, it is admittedly arbitrary. The sensitivity of the results with respect to this decision to include an equity premium, as opposed to computing values under an assumption of risk neutrality, is considered as a robustness check in Appendix E1.

(2) The DJIA is dividend-adjusted, i.e., when a company in the DJIA pays its shareholders a dividend, the index is adjusted in such a way that the expected fall in share price as a result of the dividend payment (the value of the company decreases as cash flows out in the form of

[15] On November 26, 2010, the Friday after Thanksgiving, a 120% moneyness volatility was registered of nearly twice that of the trading days prior to and after that date. As such large movements are highly unusual, we have treated this as a mistake and carry over the value from the previous trading day.

dividends) is nullified. The index can therefore be seen as a non-dividend paying security; hence the value for the dividend yield q is set to zero when the probabilities are calculated. We consider instead using the value of the actual dividend yield as a robustness check in Appendix E1.

The time series of the interest rate and dividend yield are shown in Figure B2(a). For each of the three strike prices (corresponding to 80%, 100%, or 120% of the current DJIA spot price), the option-implied probability and implied volatility are shown in Figures B2(b) – B2(d). In each graph, the correlation between the *first differences* of the two series is reported in the upper left corner. The figures show an upward trend in the probability of a positive return and a greater than -20% return and a pronounced downward trend in the option-implied probabilities of a greater than 20% return.

Figure B2(a): *1-year interest rate & dividend yield used in pricing options on the DJIA*

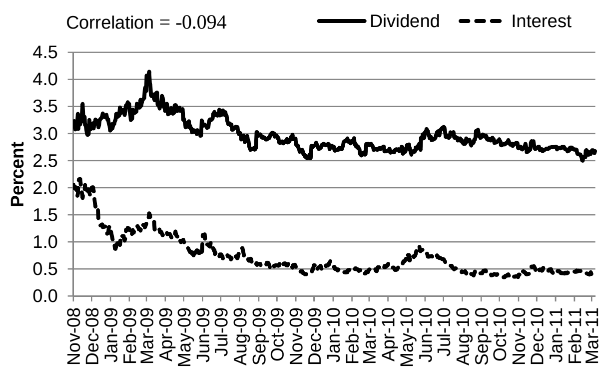


Figure B2(b): *Option-implied probability and volatility for a return of >-20% in one year for the DJIA*

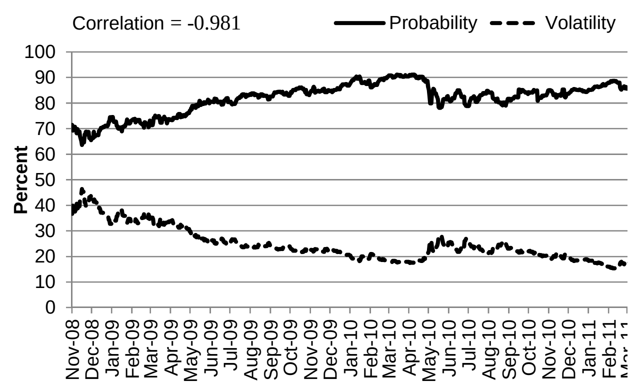


Figure B2(c): *Option-implied probability and volatility for a gain in one year for the DJIA*

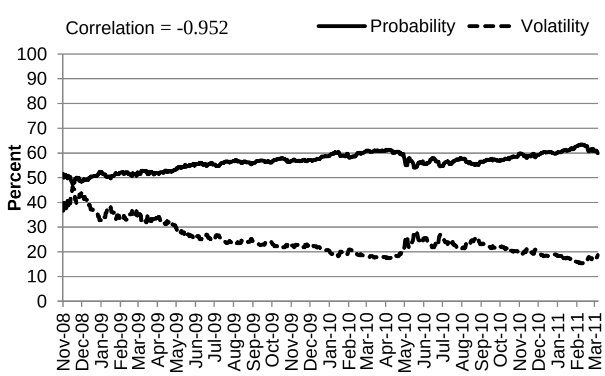
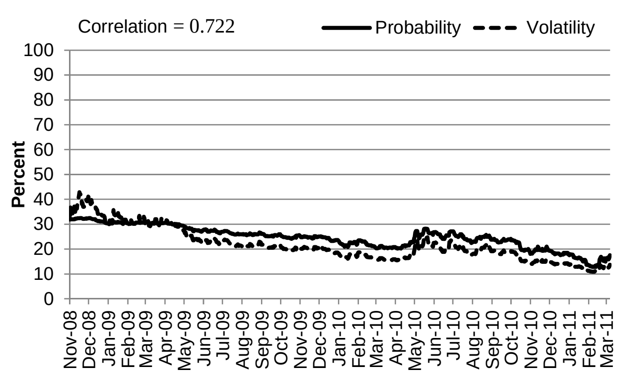


Figure B2(d): *Option-implied probability and volatility for a return of >20% in one year for the DJIA*



Source: Bloomberg

Summary statistics and correlations for both the levels and the first differences of these time series are provided in Table B. An augmented Dickey-Fuller test (Dickey and Fuller, 1979) was performed on all the option-implied probability and volatility time series in Figures B2(b)-(d) to test for a unit root including a drift and trend, with the number of included lags chosen to minimize the Schwarz (1978) Bayesian Information Criterion. Even at the 10% level of significance, the null hypothesis of a unit root was not rejected for any of the series.

Overall, the market expectations as measured by the statistics in Table B can be interpreted as somewhat negative (with the caveat that the assumption of risk neutrality implies that option-implied expectations are downward-biased under standard utility conditions). The average option-implied probability of a positive gain is below 50% and the average probability of a >20% decline (= one minus the average probability of a >-20% increase) in the stock market exceeds the average probability of a >20% increase by eight percentage points (26.1% versus 18.1%, respectively). This indication of negative sentiment is perhaps not surprising when considered in the historical context of the unfolding financial crisis during the sample period to which the data correspond.

The daily changes of the option-implied probabilities also display large fluctuations. As an example, consider the mean probability of a greater than -20% return in the stock index over the coming year (73.9%, row 1 column 1). The mean absolute daily change of 0.43 percentage points (row 1 column 6) for >*Minus20* means that each day the market's belief fluctuates by an average of 0.43 percentage points (e.g., increasing it from 73.9 to 74.33). The option-implied probabilities of >*Plus20* vary less on a daily basis (as seen by lower standard deviations, mean and median absolute deviations) and those of *PositiveReturn* by half as much as >*Plus20*. For all time series in the table, the large values of the minimum and maximum first difference compared to the standard deviation suggest fatter tails than a normal distribution would indicate; indeed, the kurtosis is between 7.58 and 10.21.

The Wall Street data exhibit evidence of volatility skew (implied volatility decreases with strike price): the mean implied volatility is highest for >*Minus20* (30.3%), lower for *PositiveReturn* (26.8%) and lowest for >*Plus20* (23.8%). With the exception of the standard deviation, the rest of the implied volatility statistics in levels display a pattern similar to the means. The comovement of these series is evident when considering the first differences of the implied volatilities (right hand block of the table), where all moments shown are quite similar. However, the magnitudes of the first differences of the implied volatilities suggest a remarkably large variation on a day-to-day basis.

Table B: Descriptive Statistics of the Time Series of Option-Implied Probabilities (computed over survey days only)

This table contains summary statistics of the daily time series of the Wall Street option-implied probabilities that are calculated using the parameters extracted from Bloomberg. The rows show summary statistics of these probabilities for a >-20% return, a positive (>0%) return, and a >20% return. Statistics are shown for both the levels and first differences (daily changes) of each time series. Because there are gaps in the survey days, only consecutive pairs of days are used for computation of the summary statistics of the first differences.

		Levels (n=364)					First differences (n= 323)				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Mean	Median	St.dev	Min	Max	Mean (abs)	Median (abs)	St.dev.	Min	Max
> -20%	Probability	79.9%	82.0%	7.1%	63.6%	91.2%	0.45%	0.23%	0.76%	-2.68%	3.99%
	Volatility	30.3%	27.5%	7.5%	19.9%	50.3%	0.46%	0.22%	0.84%	-4.34%	3.24%
> 0%	Probability	55.5%	56.1%	3.9%	47.5%	63.3%	0.24%	0.13%	0.40%	-1.55%	1.66%
	Volatility	26.8%	24.0%	7.8%	15.5%	46.3%	0.42%	0.19%	0.77%	-4.17%	3.25%
> 20%	Probability	25.7%	26.0%	4.9%	12.9%	32.5%	0.20%	0.11%	0.34%	-1.70%	1.87%
	Volatility	23.8%	21.3%	8.0%	11.0%	42.8%	0.43%	0.20%	0.79%	-3.84%	3.38%

Note: For the first differences, we report the mean and median of the absolute value because both mean and median are (practically) zero for all six of the first difference time series and these statistics would thus be uninformative. The mean and median of the absolute value provide additional information about the variation of the time series.

Appendix C: Who predicts better, Wall Street or Main Street?

A natural question to ask when comparing returns expectations of the two populations is which group does a better job of predicting future returns. To answer this question, we next use the probability thresholds corresponding to >*Minus20*, *PositiveReturn*, and >*Plus20* to approximate an empirical cumulative distribution function (cdf), assumed to be lognormal, corresponding to the distribution of expected one-year future returns on the DJIA, for each interview day in the sample. The option-implied probabilities are used to construct analogous Wall Street cdfs for each day. We then compare these daily cdfs to the corresponding ex-post realized one-year returns using a likelihood-based metric where we compute a corresponding probability density function (pdf), evaluate it at the realized return for each day, and cumulate the log of the result across all days. Hence outperformance corresponds to a higher likelihood. The results are in Table C. Surprisingly, using subjective response data elicited from a random small, yet representative sample of the US population results in only about a 30% greater likelihood than using probabilities inferred from option prices that reflect the views of thousands of market participants. Hence while Wall Street is better than Main Street at predicting subsequent future one-year returns, our results demonstrate sizable informational content in expectations elicited from survey response data.

Table C: Comparing Wall Street and Main Street accuracy of beliefs

A comparison of the accuracy of Wall Street and Main Street beliefs is shown, along with comparisons across a variety of subgroups. We report the mean log likelihood in the second and third columns, based on the (logarithm of the) probability density function (pdf) evaluated at the realized 1-year ahead return. To avoid the Main Street result being driven by extreme observations, we exclude days in which less than ten surveys were answered from the calculations (for each pairwise comparison, the number of survey-days that remain is shown). In addition to comparing Main Street to Wall Street, we similarly compare subsamples of the Main Street data to draw inference about which subgroups are more prescient; these are shown in the bottom half of the table. While men and those that are married outperform women and those that are single, respectively, the magnitude of the outperformance is not qualitatively very large. In contrast, the effects of education and being a homeowner or a stockowner double the outperformance, although still less than half of the Wall Street/Main Street outperformance.

	# Days	Wall Street	Main Street	Outperformance	p-value
Overall sample	281	**0.33**	0.06	30.7%	0.000
Subgroup		In Group	Not in Group		
Male	258	**0.13**	0.06	7.4%	0.010
Married	257	**0.10**	0.05	5.3%	0.060
Homeowner	231	**0.14**	0.01	14.2%	0.000
Stockowner	260	**0.16**	0.01	16.7%	0.000
>Bachelor’s	261	**0.17**	0.03	15.7%	0.000

Notes to table: The group with more accurate beliefs is shown in **bold.** Outperformance is measured by the ratio of the higher likelihood to the lower likelihood, where the likelihood is computed as the exponential of each mean log likelihood. P-values were obtained through bootstrapping 100,000 random draws of the number of observations for each group, with replacement.

Appendix D: Likelihood calculation

The derivation of the likelihood is provided below for the likelihood of an individual observation. Individual subscripts have been omitted for notational simplicity. The covariate matrices – which are row vectors in this case, as they refer to an individual observation – are represented with lower case instead of capital letters. The likelihood of a non-focal response is given by:

$$P(w^*>0) = P\left(x_2\beta_2 + \eta > 0\right) = 1 - P\left(x_2\beta_2 + \eta \leq 0\right) = 1 - P\left(\eta \leq -x_2\beta_2\right)$$

$$= 1 - \Phi\left(-x_2\beta_2\right) = \Phi(x_2\beta_2)$$

The likelihood of a focal response is then given by:

$$1 - \Phi\left(x_2\beta_2\right) = \Phi(-x_2\beta_2)$$

Where Φ is the standard normal cumulative distribution function. When there is a non-focal response *p*, the density, is given by the beta distribution with parameters α_1 and α_2 (Mendenhall, Scheaffer and Wackerly, 1981):

$$f(p|\alpha_1, \alpha_2) = \frac{p^{\alpha_1-1}(1-p)^{\alpha_2-1}}{B(\alpha_1\alpha_2)} = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}p^{\alpha_1-1}(1-p)^{\alpha_2-1}$$

with Γ the gamma function and α_1, α_2, and μ given by:

$$\alpha_1 = \mu\varphi \qquad \alpha_2 = (1-\mu)\varphi$$

$$\mu = \frac{1}{1+\exp(-X_1\beta_1)}$$

and φ a constant. When there is a focal response, the likelihood for a response of 50 (v*>0) is given by $\Phi\left(x_3\beta_3\right)$ and that of zero or 100 by $\Phi\left(-x_3\beta_3\right)$.

Conditional on a response of zero or 100 (v*≤0), the likelihood of a response of zero is given by:

$$P\left(\tilde{p} \leq \psi|\alpha_1, \alpha_{2)}\right) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_{t=0}^{\psi} t^{\alpha_1-1}(1-t)^{\alpha_2-1}dt$$

which is often referred to as the 'regularized incomplete beta function' $I_\psi(\alpha_1, \alpha_2)$.[16] The overall likelihood of any response is then given by the product of the three separate likelihoods (i.e., the likelihood of a response of zero or 100 given a focal response, the likelihood of a focal response, and the likelihood of the response conditional on a non-focal response):

p = 0

$$l = P(w^* \leq 0|x)P(v^* \leq 0|x, w^* \leq 0)P(\tilde{p} \leq \psi|x, w^* \leq 0, v^* \leq 0)$$

$$= \Phi\left(-x_2\beta_2\right)\Phi(-x_3\beta_3)I_\psi(\alpha_1, \alpha_2)$$

[16] Mendenhall, Scheaffer and Wackerly (1981), p. 147.

p = 1

$$l = \Phi\left(-x_2\beta_2\right)\Phi(-x_3\beta_3)(1 - I_\psi(\alpha_1, \alpha_2))$$

p = 0.5

$$l = \Phi\left(-x_2\beta_2\right)\Phi(x_3\beta_3)$$

p = other (non-focal)

$$l = \Phi\left(x_2\beta_2\right)\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}p^{\alpha_1 - 1}(1 - p)^{\alpha_2 - 1}$$

Appendix E: Robustness checks

A number of robustness checks were conducted; the results are summarized in this section.[17]

E1. Sensitivity to risk neutrality assumption, equity premium and dividend yield assumption

The baseline assumptions for the parameters r (the interest rate) and q (the dividend yield) reflect an adjustment to the assumption of risk neutrality (i.e., we use $r + \rho$ = the risk-free rate plus a 6% equity risk premium, in other words, the required rate of return) and a recognition that the DJIA is dividend-adjusted (i.e., $q = 0$), respectively. We consider in this section sensitivity to the assumption about the equity premium and/or estimates of the dividend yield. To evaluate the robustness of the results to the baseline assumptions, we therefore consider a number of alternative values for the parameters used to computing the option-implied probability: (1) $\rho = 0$, i.e., preferences are risk neutral, (2) $\rho = 0$ and q = the Bloomberg dividend forecast, and (3) $\rho = 6\%$ and q = the Bloomberg dividend forecast.

As would be expected, the assumption of risk neutrality shifts the option-implied probability distribution to the left, lowering the average, minimum, and maximum associated with the three questions. Figure D shows the time series of the option-implied probability to *>Minus20, PositiveReturn* and *>Plus20* with values for r and q as specified in the alternatives above, as well as the baseline values used in the paper. Not surprisingly, the pattern of the series is similar for the different calculations of the option-implied probability, although the difference between the series is not constant. The latter is a result of the non-linearity of the normal distribution function and the fact that the dividend and interest rate are not constant.

The inclusion of alternative shifted distributions of option-implied probabilities in the model causes only the coefficient on the option-implied probability and the constant to change; all other coefficients, standard errors, and marginal effects are practically unchanged. The estimated marginal effects on the option implied probability change depending on the model used: a ten percent increase in the option implied probability increases respondents' beliefs by 1.09 / 1.01 / 1.17 percentage points for models (1) to (3), respectively, compared to an increase of 1.14 percentage points for the baseline model (with r the required rate of return and $q = 0$). The coefficient on the option-implied probability is significant at the 5% level for all four models.

Taken together, these analyses demonstrate that the regression results in the paper are hardly sensitive to the assumptions about risk neutrality, the equity premium, and the dividend yield.

E2. Sensitivity to the meaning of 50% responses.

An attractive feature of the ALP survey design was that an additional elicitation question was included following a response of "50%" to the *PositiveReturn* question.[18] The response to this

[17] For each robustness check, the model was re-estimated according to the modification described. In most cases, due to space considerations, only a summary of the findings is included. A complete set of results is available from the authors on request.

follow-up question (hereafter referred to as *50%Followup*) enables the possibility of separately distinguishing between (1) people for whom 50% conveys complete uncertainty, versus (2) those who view the probabilities of an increase or a decrease in the stock market over the coming year as being equal, and testing whether the two groups demonstrate different Wall Street/Main Street linkages. Of the subjects that answered 50 to *PositiveReturn*, a slight majority (52%) noted (via the follow-up elicitation question) that they were "just unsure", rather than truly thinking there was an equally likely chance that the DJIA would rise versus fall. That such a large proportion of the respondents use this focal point to indicate their uncertainty motivated our decision to model the choice of 50 as a separate behavior (following Hurd, McFadden, and Gan, 1998 and Bruine de Bruin et al., 2002).

As a starting point, the proportion of "unsure" responses is computed, stratified according to how closely respondents are following the stock market, their self-assessed understanding of the stock market, and their understanding of the laws of probability (as measured by our "inconsistent" and "near-inconsistent" definitions); the results are in Table E1. Overall, about 42% of those responding to the two stock market questions reported that their focal response of 50% meant there was an equal probability of an increase or decrease and 58% reported that it meant they were unsure what the probability was. Yet for both questions, the proportions vary significantly according to how closely individuals follow and how well they understand the stock market. The proportion of 50% responses that are reported to have meant "equal probability" is highest among those claiming to follow the stock market very closely or to have an excellent or very good understanding of the stock market (58% and 70%, respectively) and lowest for those who claim to not follow the stock market at all or to have an extremely poor or very poor understanding (33% and 25%, respectively).[19]

There are also statistically significant differences in the proportions that use 50% to mean equal probabilities according to how well respondents seem to understand the laws of probability. Among consistent sets of responses, a response of 50% to the *PositiveReturn* question means equal probability of an increase and decrease significantly more often as it means the respondent is unsure about the exact probability (62% versus 38%). Among inconsistent sets of responses, the proportion is nearly 11 percentage points lower than the proportion of consistent responses. Near-inconsistent sets of responses are least likely to use the focal response of 50% to indicate equal probabilities; not only is the proportion nearly 24 percentage points lower in these sets than among the consistent sets, it is also significantly less than the proportion of unsure responses (39% versus 61%).

As a result of these findings, the model was re-estimated to take these differences into account (Table E2). A dummy variable equal to one if the respondent answered 50% to the *PositiveReturn* question **and** subsequently reported that their response indicated they were "unsure" and zero otherwise is included in both the μ and w* equations; an interaction between

[18] The exact wording of the question was, "Do you think it is equally likely the shares will be worth more in a year as it is they will be worth less or are you just unsure about the chances?" The sample contains 11,810 person-wave responses to this question.

[19] The proportions reporting "unsure" are 100% minus the proportions reporting "equal probability". The differences in these proportions are statistically significantly different at a 1% level of significance.

this "unsure" dummy variable and the option-implied probability variable is also included in the first equation.[20] The coefficient on the dummy variable in the μ equation is not statistically significant, and the qualitative results are for the most part unchanged. In the w* equation that models the likelihood of a focal response, the inclusion of the "unsure" dummy variable renders the effect of being female and non-Hispanic white no longer statistically significant. The coefficient for age changes sign from positive to negative; older people are now significantly more likely to give a non-focal response.

Most importantly, the coefficient on the interaction term between the "unsure" dummy and the option-implied probability variable is significantly negative, indicating that those who are unsure report probabilities that bear less resemblance to the option-implied probability. When those that are "unsure" and those that believe the probability is equal are included separately in the estimation, the marginal effect of the option-implied probability for the "unsure" group is -0.044 (though with p-value 0.066), while the marginal effect for the other observations (those in the "equal probability" group plus those that gave a non-50 response to *PositiveReturn)* is larger than in the baseline regression, estimated to be 0.122. A possible explanation for this finding could be that those who admit to being unsure have little understanding of the market and are thus less likely to know the 'true' probabilities.

E3. <u>Sensitivity to choice of return date</u>

The decision to compute returns through the close of business of the day before the survey was administered was made in order to keep the explanatory variables chronologically prior to the dependent variable, thus facilitating an approximate causal interpretation (i.e., how do Wall Street beliefs influence Main Street beliefs).[21] An argument can be made, however, for including returns computed through the close of business of the day of the interview, enabling a more contemporaneous determination of views (i.e., how similar are Wall Street and Main Street beliefs).

The model is re-estimated using returns, option-implied probabilities, and implied volatilities computed as of the close of the day of interview, rather than as of the close of the day prior. The results are qualitatively the same, indicating little sensitivity of the results to the causally-prior modeling choice. Including the contemporaneous day's information, the return in the past 30 days is no longer statistically significant at the 5% level in the third (v^*) equation, while the return in the past year is. In addition, the implied volatility is statistically significant at the 10% level.

[20] It is not possible to include the dummy in the third (v^*) equation since the follow-up elicitation is only asked following a response of 50%, where the latter corresponds to $v^*>0$ in all cases. In the interest of space only a summary of the key findings is provided here; the full set of results is available from the authors on request.
[21] A Granger-causality test was performed as a robustness check to consider the possibility that causality instead goes from Main Street beliefs to Wall Street beliefs. Wall Street beliefs were found to Granger-cause Main Street beliefs, but not vice versa.

Table E1: Contingency table of responses to the follow-up question after a response of 50 to *PositiveReturn*.

Numbers represent the (weighted) proportion of "Equal probability" and "Unsure" responses to the follow-up question for different subgroups. T-values corresponding to a test for a difference in proportions are reported and shown in bold when significant at a 5% level. Note that the numbers shown in the rows "Number of Obs." are not weighted, as these are count data. The number of observations in panels (a) and (b) is less than in panel (c) as a result of the question not being asked in every survey. The number of observations in panel (c) is less than the overall sample because the *50%Followup* question was only asked of those who gave a response of 50 to *PositiveReturn*.

(a) Following stock market

	Very Close	Somewhat	Not at All	Total
Equal probability	0.58	0.51	0.33	0.42
Unsure	0.42	0.49	0.67	0.58
Number of Obs.	166	1,111	1,086	2,363
t-value difference	1.83	0.81	**-11.79**	**-7.58**

(b) Understanding stock market

	Extremely/Very Good	Somewhat	Extremely/Very Poor	Total
Equal probability	0.70	0.49	0.25	0.42
Unsure	0.30	0.51	0.75	0.58
Number of obs.	234	1,451	681	2,366
t-value difference	**5.29**	-0.88	**-14.21**	**-7.55**

(c) Consistency

	Inconsistent	Near-inconsistent	Consistent	Total
Equal probability	0.51	0.39	0.62	0.48
Unsure	0.49	0.61	0.38	0.52
Number of obs.	735	6,229	4,776	11,810
t-value difference	0.74	**-19.08**	**15.65**	**-5.10**

Table E2: Regression Results Accounting for the Interpretation of a Response of 50

Maximum likelihood estimates of the model are presented in the table. For each variable, the estimated regression coefficient, corresponding standard error and the marginal effect evaluated at the variable means are reported. For dichotomous (binary) variables, the marginal effect is the difference in probability when evaluated at the value of one versus zero, *ceteris paribus.* The first three columns show the results of the equation pertaining to μ, the expected value of respondents' stated subjective probabilities. The second three columns refer to w^*, where $w^*>0$ corresponds to a non-focal response. The last three columns refer to v^*, where $v^*>0$ signifies that a respondent gives a response of 50, conditional on giving a focal response (of zero, 50 or one hundred). A complete set of wave dummies (not shown) is included in each of the three equations. Coefficients in **bold** represent significance at the 5% level.

Observations 139,327 | Log likelihood -57,075 | Average log likelihood -0.410

	μ			w^*			v^*		
	Coef.	Std. Err.	Marginal	Coef.	Std.	Marginal	Coef.	Std. Err.	Marginal
Dummy (PositiveReturn)	**-1.600**	0.029	-0.367	**-0.256**	0.009	-0.086	**0.138**	0.026	0.043
Dummy (>Plus20)	**-2.065**	0.059	-0.452	0.012	0.009	0.004	**0.142**	0.042	0.044
Demographic Characteristics									
Female	**-0.115**	0.006	-0.029	0.004	0.008	0.001	**0.250**	0.015	0.079
Age	**-0.002**	0.000	0.000	**-0.001**	0.000	0.000	**-0.011**	0.001	-0.003
Race									
Non-hispanic white	**0.096**	0.013	0.024	0.033	0.018	0.011	**-0.096**	0.034	-0.030
Non-hispanic black	**0.067**	0.016	0.017	**0.090**	0.021	0.029	**-0.148**	0.039	-0.049
Hispanic/Latino	**-0.041**	0.016	-0.010	0.006	0.021	0.002	**-0.171**	0.039	-0.057
Education									
Some college, no Bachelor	**0.085**	0.007	0.021	**-0.055**	0.009	-0.018	**0.078**	0.017	0.024
Bachelor's degree	**0.189**	0.008	0.047	**0.164**	0.011	0.052	**0.202**	0.022	0.061
>Bachelor's	**0.259**	0.010	0.065	**0.275**	0.014	0.084	0.024	0.029	0.007
Married	-0.005	0.006	-0.001	**0.041**	0.008	0.013	-0.014	0.016	-0.005
Working	**-0.046**	0.006	-0.011				**0.061**	0.015	0.019
Home owner	**-0.024**	0.008	-0.006				**0.094**	0.018	0.030
Stock owner	**0.088**	0.007	0.022				**0.164**	0.019	0.050
Have Retirement Account	**0.099**	0.007	0.024				**0.126**	0.017	0.040

Financial Market Characteristics									
Return past 30 days	-0.051	0.189	-0.013				**-1.476**	0.407	-0.466
Return past year	**0.287**	0.110	0.071				0.323	0.278	0.102
Return past 30 days * PositiveReturn	0.035	0.134	0.009				0.515	0.294	0.163
Return past year * PositiveReturn	**-0.130**	0.033	-0.032				-0.054	0.066	-0.017
Return past 30 days * >Plus20	0.270	0.147	0.067				**0.880**	0.318	0.278
Return past year * >Plus20	**-0.308**	0.047	-0.077				-0.022	0.070	-0.007
Following the stock market									
Closely following	**-0.046**	0.014	-0.011	**-0.087**	0.019	-0.029	**-0.130**	0.036	-0.042
Not following	**-0.106**	0.007	-0.026	**-0.096**	0.010	-0.032	**-0.127**	0.018	-0.040
Understanding of stock market									
Good understanding	**0.071**	0.012	0.018	0.015	0.017	0.005	**-0.089**	0.034	-0.029
Bad understanding	**-0.050**	0.008	-0.012	**-0.063**	0.010	-0.021	-0.016	0.018	-0.005
Option-implied probability	**0.122**	0.022	0.122						
Unsure * OIP	**-0.167**	0.010	-0.166						
Implied volatility							0.521	0.579	0.165
Unsure	0.026	0.014	0.006	**-1.479**	0.011	-0.539			
Constant	**1.001**	0.044	0.248	**0.992**	0.032	0.326	**0.823**	0.159	0.260
Additional parameters									
ψ	0.463	0.004							
φ	3.918	0.016							

Notes to table: See notes to Table 2. In addition a dummy variables for Unsure (=1 if an individual answered 50 to *PositiveReturn* and then indicated he was unsure about the probability) is added to the μ and w^* equation, and an interaction between this dummy and the the option-implied probability (OIP in the table) is added to the μ equation.