Modeling Non-Standard Financial Decision Making

ISBN: 978 90 361 0416 6

Cover illustration: “Bron van de Rotte”, Ernst de Vogel, 1995.

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **601** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Modeling Non-Standard Financial Decision Making

Het modelleren van atypisch financieel keuzegedrag

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 11 december 2014 om 11:30 uur

door

Rogier Jan Dave Potter van Loon

geboren te 's-Gravenhage

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren: Prof.dr. H. Bleichrodt

Prof.dr. P.P. Wakker

Overige leden: Prof.dr. A. Baillon

Prof.dr. M. Dierkes

dr. M.J. van den Assem

Ter nagedachtenis aan mijn oudtante

Sophia (Fiete) van Herwijnen

Acknowledgments

The last three years could not have been as pleasant without all the people around me, and I would hereby like to take the opportunity to thank those who have helped me during my PhD.

First of all, I thank my supervisors, Peter Wakker and Han Bleichrodt. They had sparked my interest in behavioral economics during lectures in my Bachelors, and it has been truly an honor to end up pursuing a PhD under their supervision. Professor Bleichrodt lent me an ear whenever I required this and was very pleasant to work with on joint papers. I admire the ease with which he is able to deal with technical writing and general issues. Professor Wakker's way of educating his students has helped me a lot. Although it is often considered non-orthodox, I truly believe that there is no better way of bringing the message across. My most memorable experience was receiving my lowest grade ever, a '3', on a two page homework assignment that was wholly correct except for one word: "if". I was shocked and annoyed, but it is unlikely that I will ever again mess up a conditional statement. His organizing of the Thursday meetings is paramount in bringing the group together, both socially and with regards to academic thoughts.

The Behavioral Economics group is truly a wonderful working environment. I was lucky to be approached by Martijn van den Assem in the final stage of my MPhil to work on a project with him. He gave me a 'kick-start' into research and has taught me a lot on how to present my work. My daily supervisor, Aurélien Baillon, has been a catalyst for the group and I admire how he helps the PhD students. Kirsten Rohde is able to create technically challenging work while never losing sight of human intuition/feelings.

Jan Brinkhuis has given me great guidance and advice ever since I became his student assistant seven years ago. He has helped me especially by introducing me to Robin Lumsdaine in my first year. It has been a great experience to work abroad with her immediately at the start of my PhD. I am truly grateful to have been given the opportunity to work with someone so knowledgeable and hard-working; I have really benefited from the work ethic that she has instilled in me.

In my final year, I had the privilege of working on a project with Drazen Prelec. It is truly amazing to see the ease with which he comes up with elegant ideas about difficult technical topics.

Dennie has been the best office mate I could wish for; we had great times together and he has sharpened my mind a lot through our sparring sessions. Jan was always good for a killer joke when he sneaked into our office. Ning & Yu are the loveliest couple I know, and I'm perpetually amazed by their altruism. Chen & Vitalie are the smartest couple I know; it was always a pleasure to discuss technical stuff with them. It has been a pleasure to cooperate with Zhenxing and Tong on projects. My thanks also go to Arthur, Julia, Amit, Umut, Asli, Zhihua, Sofie, Uyanga, Ilke, Ivan, Sander, Jonathan, Justinas and Saeedeh for the pleasant interactions over the years.

Financial support from the Netherlands Organization for Scientific Research and the Tinbergen Institute is gratefully acknowledged.

Finally, I would like to thank my family for their support. My parents have stimulated me during two decades of education and have always shown me the path to follow. I admire especially their altruism, my mother's cheerfulness and my father's dedication. My sister has helped me with her knowledge of psychology, both for research and non-research related topics. Marrying Doré is without any doubt the best decision I have made in the last three years. She is always there for me, cheers with me after an accomplishment, lifts me up after a setback, helps with new ideas and motivates me. She makes me want to be a better man.

Contents

List of Tables

List of Figures

Chapter 1 | Introduction

1.1 Games and Financial Decision Making

The first financial decision that was preceded by a probability calculation was a dice game. In 1654, Antoine Gombaud asked for the chance of winning a game with multiple throws of the dice. Blaise Pascal and Pierre de Fermat took on the challenge and calculated the probability distribution. This laid the foundation for probability theory. If one is able to calculate the probabilities of outcomes, it is a simple second step to calculate the expected value of various choices by taking the probability-weighted sum of the outcomes. By choosing the financial options that give the greatest expected value, one ends up with the highest result in the long term.

This does not, however, take into account that people are generally risk-averse. The Saint Petersburg paradox, for example, describes a game of chance with infinite expected value that is generally valued at less than €20. To account for people's risk aversion, Bernoulli (1738) introduced the concept of utility. When making financial choices, he stated, one should not look at expected value, but rather at expected *utility*. As an example, he mentions the merchant Caius, who is considering insuring his cargo on a ship from Amsterdam to Petersburg. Bernoulli argues that Caius should not look at the arithmetic average of the outcomes (expected value), but at the geometric average, which is equivalent to behaving according to expected utility with a logarithmic utility function.

It is generally accepted that it is rational to maximize expected utility, both when probabilities are known (Von Neumann & Morgenstern 1944) as well as when they are unknown (Savage 1954). The *normative* method of making financial decisions (how people ought to choose) has thus been known for over 275 years. In practice, however, we see that expected utility is not an accurate *descriptive* model of how people make financial decisions. They are often influenced by psychological factors that have no place in the standard financial model.

This phenomenon is visible in various aspects of financial markets. Perhaps most notably, the return on stocks is much higher than can be expected based on risk preferences and historical volatility. Investors also give disproportionate attention to probabilities of familiar events: most individual investors invest only in stocks of their home country, whereas diversification across countries can greatly reduce portfolio risk. Further, people give disproportionate attention to short-term gains and losses, which is against their own (long-term) interest.

In response to these observations, a new field titled Behavioral Finance has developed. Here, the assumption of a rational, optimizing decision maker has been dropped and psychological phenomena have been introduced in the financial world. DeBondt & Thaler (1985) gave a first start by writing about under- and overreaction to news on financial markets. The field uses much of the insights gathered by Amos Tversky and Daniel Kahneman, who, among other things, introduced "prospect theory" in Tversky & Kahneman (1979), a descriptive model of decision making that comprehends multiple irrationalities. Thaler (1980) already suggests prospect theory as a model that is suitable for explaining how humans make decisions, instead of how they *should* choose. It can explain, for example, why stock returns are so high (Benartzi & Thaler 1995) and why individual private investors wait too long to sell bad performing stocks and are too quick to sell stocks that perform well (Barber & Odean 2000).

This thesis analyzes fundamental aspects of financial decision making by modeling human decision making. Chapter 2 describes a model to explain choice behavior under 'ambiguity': situations in which people have some idea about probabilities but do not know the exact probability distribution. For practically all financial decisions, no one can determine exactly the risks. It is not possible to exactly calculate, for example, the probability that, the Dow Jones index will rise over 20% in the next year. Chapter 3 analyzes how people assess the probabilities of situations like the one mentioned above, and whether these probabilities correspond to those of the market (homogeneous expectations are often assumed in financial models). Chapters 4 and 5 are about intertemporal choices (e.g., €100 now or €110 in one year); financial decisions are often about receiving something in the future. Chapter 4 suggests an improvement of an often used model and Chapter 5 argues in favor of a relatively new model. The last Chapter returns to where it all started by looking at a game: it analyzes the role of skill in the game of poker.

1.2 Thesis Outline

Chapter 2 discusses multiplier preferences, a model that describes behavior under ambiguity, where people have subjective beliefs but do not know exact probabilities. The subjective expected utility model (Savage 1954) prescribes that ambiguity should not affect a person's preferences (only subjective beliefs matter), but numerous studies (most notably Ellsberg 1961) have found that people do behave differently when ambiguity is involved.
Multiplier preferences are widely used in macroeconomic models, but they not been applied in microeconomic settings, largely because they do not permit behavior in which people are ambiguity seeking. We give a preference foundation for an extension of the model, such that it allows for ambiguity seeking as well as ambiguity aversion. Also, we propose a simple method to measure multiplier preferences, thereby giving a straightforward measure of ambiguity attitude. This allows for a broad application of the model in analyzing individuals' dealing with ambiguity.
We give a first example of such an application by analyzing the ambiguity preferences of a large representative sample of the Dutch and of the American population. We thereby obtain the first micro-economic estimates of multiplier preferences. Nearly one third of the respondents is found to be ambiguity seeking, illustrating the need for the extension of the model. Contrary to the predictions from the theoretical literature, ambiguity aversion is not (negatively) correlated with stock market participation, but the deviation from ambiguity neutrality is.

Chapter 3 analyzes people's subjective beliefs about movements in the stock market, and how these subjective beliefs compare to the beliefs implicit in the prices of financial securities. Many theoretical models assume homogeneous beliefs, implying that the two sets of beliefs should be the same (at least in expectation). On the other hand, many would argue that 'regular people' know little about the stock market as they are often not directly involved, and as such will have beliefs that are not at all aligned with those of the financial market.

This research is the first to highlight the correspondence between the wording of subjective response questions in surveys and the design of index options. A large representative sample of the US population was asked for their beliefs that the index return would be above a certain threshold, which corresponds to a probability that can be extracted from option prices. Market beliefs and general population beliefs can thus be compared. We find that there is a relationship between the two sources of views, although the association is far from one-for-one. We find a closer association for those demonstrating a better understanding of the laws of probability.

Chapter 4 dicusses the quasi-hyperbolic or (β, δ)discounting model. This is the most widely used model to explain a phenomenon called decreasing impatience. People are generally more impatient for outcomes in the short run than for those in the long run (normatively one should display constant impatience). This can cause dynamic inconsistencies, in which people reverse their preferences over time.
In the quasi-hyperbolic model, β causes this decreasing impatience, and is therefore commonly interpreted as an index of dynamic inconsistency. We show that this interpretation is problematic because β captures other components of intertemporal attitudes and interacts with the discount factor δ. Instead, $\tau = \ln(\beta) / \ln(\delta)$ is a proper index of dynamic inconsistency, as we prove by a preference axiomatization. It leads to a rewriting of quasi-hyperbolic discounting ($\beta\delta^t$) as $\delta^{\tau+t}$.
The index τ has a natural interpretation as a perceived time penalty for any delay beyond the present and thus as the time period over which the decision maker is vulnerable to dynamic inconsistencies. We give an empirical illustration of the use of $\boldsymbol{\tau}$ by reanalyzing the data from Tanaka et al. (2010).

The quasi-hyperbolic model is useful because of its simplicity and tractability. To measure peoples' true discount functions, however, more advanced models are necessary. It is my belief that the Constant Relative Decreasing Impatience (or Unit Invariance) function suggested in Bleichrodt, Rohde & Wakker (2009) is among the most suitable for this task. Chapter 5 is a comment on Doyle's (2013) survey of discount functions, in which this family is criticized based on incorrect assumptions. We show his mistakes and make a case in favor of the two families of functions.

The final Chapter analyzes the skill-component in online poker. A major issue in the widespread controversy about the legality of poker and the appropriate taxation of winnings is whether poker should be considered a game of chance. Many behavioral theories would suggest skill has a substantial influence on poker, but legally, it is often considered a game of chance.

We present an analysis into the role of skill in the performance of online poker players, using a large database with hundreds of millions of player-hand observations from real money ring games at three different stakes levels. We find that players whose earlier profitability was in the top (bottom) deciles perform better (worse) and are substantially more likely to end up in the top (bottom) performance deciles of the following time period. Regression analyses of performance on historical performance and other skill-related proxies provide further evidence for persistence and predictability. Simulations point out that skill dominates chance when performance is measured over 1,500 or more hands of play.

Chapter 2 | Robustness: Extending Multiplier Preferences

Multiplier or robust preferences are widely used in macroeconomics to capture model uncertainty. At the micro level, they have not been applied yet, because they do not permit ambiguity seeking, which is usually observed for a substantial proportion of subjects. This chapter makes three contributions. First, we give a preference foundation for (extended) multiplier preferences that allows for both ambiguity aversion and ambiguity seeking. Second, we propose a simple method to measure multiplier preferences, which gives an axiomatically founded measure of ambiguity attitude. Third, we apply this method to two large representative samples (one Dutch and one American) and obtain the first micro estimates of multiplier preferences. We find that nearly one third of the respondents is ambiguity seeking, illustrating the need for extended multiplier preferences. Contrary to predictions from the theoretical literature, ambiguity aversion is not (negatively) correlated with stock market participation but the deviation from ambiguity neutrality is.

This chapter is based on the paper "Robustness: (Extended) Multiplier Preferences for the American and the Dutch Population", co-authored by Aurélien Baillon, Han Bleichrodt and Zhenxing Huang. The authors are grateful to Roy Kouwenberg, Tomasz Strzalecki and Peter P. Wakker for helpful comments.

2.1 Introduction

Dissatisfaction with the dominant rational expectation hypothesis has led to new models in macroeconomics and finance, which permit that decision makers' beliefs about economic phenomena are non-unique. One of the most influential of these new models was proposed by Hansen and Sargent (2001). In their *multiplier preferences* model, decision makers rank payoff profiles f according to the criterion:

$$V(f) = \min_{p} \int u(f)\, dp + \frac{1}{\sigma} R(p||q), \tag{2.1}$$

where u is a utility function, q is a subjective probability distribution on the states of the world, σ is a behavioral parameter, and $R(p||q)$ is the relative entropy of any probability distribution p with respect to q. The intuition underlying Eq. (1) is that the decision maker has some best guess q of the probability distribution, but he does not have full confidence in his guess and also considers other probability distributions p. The plausibility of these other distributions decreases with their distance from q, as measured by the relative entropy R. The parameter $\frac{1}{\sigma}$ captures the degree to which the decision maker takes alternative probability distributions into account. The lower is σ, the more the decision maker trusts that q is the correct distribution. In the limit, if σ goes to zero, Eq. (2.1) becomes subjective expected utility.

The lack of trust decision makers have in their beliefs may result from ambiguity (Hansen and Sargent 2001). In empirical studies, most subjects are not neutral towards ambiguity, as assumed by expected utility, but are ambiguity averse. Multiplier preferences capture ambiguity aversion while remaining analytically convenient and easy to incorporate in economic models of aggregate behavior. However, they do not accommodate ambiguity seeking, which limits their applicability at the micro level where a wide range of ambiguity attitudes is typically observed and a substantial proportion of respondents is ambiguity seeking.

This chapter makes three contributions. First, we extend multiplier preferences to accommodate both ambiguity aversion and ambiguity seeking. We give a preference foundation of this extended model that complements Strzalecki (2011) and that makes multiplier preferences suitable for microeconomic applications.

Second, we present a simple method to measure extended multiplier preferences. Our method is easy to apply and measures multiplier preferences at the individual subject level. Hence, we obtain an axiomatically founded measure of ambiguity aversion that can easily be used in empirical research and that captures the heterogeneity in individual ambiguity attitudes.

Third, we provide the first micro estimates of the (extended) multiplier preferences for two large representative samples of the Dutch and the US population involving over 5,000 subjects in total. Most subjects were moderately ambiguity averse, but between 25% (Dutch sample) and 35% (US sample) were ambiguity seeking. In both samples, we observed that better educated respondents deviated less from ambiguity neutrality. On the other hand, income and gender had little to no effect on ambiguity attitudes. The data also allowed us to explore a prediction from the theoretical literature that ambiguity aversion leads to less stock market participation (Bossaerts et al. 2010, Cao, Wang, and Zhang 2005, Dow and Werlang 1992, Easley and O'Hara 2009, Epstein and Schneider 2010). We found no evidence for this prediction, but we did observe that respondents who deviated less from ambiguity neutrality (subjective expected utility) were more likely to participate in the stock market.

2.2 Extended Multiplier Preferences

We use the Anscombe-Aumann setting. Let S be the *state space,* i.e. the set of all possible *states of the world* s. S can be finite or infinite. One state s will occur but the decision maker does not know which one. Σ denotes a sigma-algebra on S. Its elements are called *events* and are typically denoted E. The set of all countably additive probability measures on (S, Σ) is denoted by $\Delta(S)$ and is endowed with the weak* topology. A probability measure $p \in \Delta(S)$ is *absolutely continuous* with respect to $q \in \Delta(S)$ if for all $E \in \Sigma$, $q(E) = 0$ implies $p(E) = 0$. Let $\Delta(q)$ denote the set of all countably additive probability measures that are absolutely continuous with respect to q. For any $p, q \in \Delta(S)$, the *relative entropy* of p with respect to q is given by $R(p||q) = \int_S \log\left(\frac{dp}{dq}\right) dp$ if $p \in \Delta(q)$ and $R(p||q) = \infty$ otherwise.

We denote the *outcome set* by Z. $\Delta(Z)$ is the set of all simple lotteries on Z. Elements of $\Delta(Z)$ are denoted x or y. The decision maker chooses between *acts*, finite-valued mappings from S to $\Delta(Z)$, which are Σ-measurable. Acts are usually denoted f or g. For event E, $f_E g$ denotes the act that gives $f(s)$ if $s \in E$ and $g(s)$ if $s \in E^c$ with E^c the complement of E. The set of all acts is $\mathcal{F}$. Acts have two stages: the first stage corresponds to the uncertainty modeled by S and the second stage to the risks modeled by $\Delta(Z)$. The *mixture act* $\alpha f + (1-\alpha)g$ for $\alpha \in [0,1]$ is the act that assigns the lottery $\alpha f(s) + (1-\alpha)g(s)$ to state s for all $s \in S$. The decision maker's preferences over acts in $\mathcal{F}$ are denoted by $\succcurlyeq$ (with $\sim$, $\succ$, $\preccurlyeq$, and $\prec$ defined as usual). A functional V *represents* $\succcurlyeq$ if $V: \mathcal{F} \to \mathbb{R}$ is such that $f \succcurlyeq g \Leftrightarrow V(f) \geq V(g)$.

Definition 2.1: We call $\succcurlyeq$ *extended multiplier preferences* if $\succcurlyeq$ can be represented by

$$V(f) = \begin{cases} \min\limits_{p\in\Delta(S)} \int_S u\big(f(s)\big)dp(s) + \dfrac{1}{\sigma}R(p||q) & if\ \sigma > 0 \\ \int_S u\big(f(s)\big)dp(s) & if\ \sigma = 0 \\ \max\limits_{p\in\Delta(S)} \int_S u\big(f(s)\big)dp(s) + \dfrac{1}{\sigma}R(p||q) & if\ \sigma < 0 \end{cases}$$

where u is a nonconstant expected utility functional, $q \in \Delta(S)$, and $\sigma \in \mathbb{R}$. We call these preferences robust if $\sigma \geq 0$ and opportunity seeking if $\sigma \leq 0$.

A decision maker whose preferences are opportunity seeking chooses the probabilities that will maximize his expected utility minus a cost, which depends on the distance between these probabilities and his best guess. A decision maker with robust preferences tries to find options that are maximally insensitive to remaining uncertainties. By contrast, an opportunity seeking decision maker is looking for possibilities to improve his expected utility and he values options for which the remaining uncertainties can lead to high expected utilities.

An alternative interpretation of extended multiplier preferences approach comes from a comparison with $\int u(f)\,dp + \theta[R(p||q) - \eta]$, the Lagrange function deduced from minimizing (in the robust approach) or maximizing (in the opportunity seeking approach) $\int u(f)\,dp$ such that the relative entropy does not exceed a threshold ($R(p||q) < \eta$). This comparison shows that the multiplier parameter $\theta = \frac{1}{\sigma}$ is the Lagrange multiplier of the optimization problem and can be interpreted as the shadow price of relaxing the constraint imposed on the relative entropy (Hansen and Sargent, 2001).

There is a third interpretation of the multiplier parameter as an index of ambiguity aversion. Lemma 2.1 in the Appendix shows that extended multiplier preferences are ordinally equivalent to Neilson's (2010) second-order expected utility (SOEU) $V(f) = \int_S \varphi_\sigma\left(u\big(f(s)\big)\right) dq(s)$ with $\varphi_\sigma(t) = \begin{cases} -e^{-\sigma t} & if\ \sigma > 0 \\ t & if\ \sigma = 0 \\ e^{-\sigma t} & if\ \sigma < 0 \end{cases}$ with u, q, and σ the same as in Definition 1. We know from Pratt (1964) that under expected utility the exponential utility function is equivalent to constant absolute risk aversion. This implies that adding an amount c to all outcomes of the lotteries under comparison does not change the preferences between these lotteries. For the exponential function, the Arrow-Pratt index of risk attitude $-\frac{u''}{u'}$ is constant and equal to the exponential parameter. Under SOEU, we can give a similar interpretation to the exponential φ_σ function in terms of utility: adding the same (expected) utility to each state of the acts under comparison does not change the preferences between these acts. Grant and Polak (2013) coin the term constant absolute uncertainty aversion to describe this property. The index $-\frac{\varphi''}{\varphi'} = \sigma$ is then an Arrow-Pratt index of ambiguity attitude.

2.3 Axiomatization

Strzalecki (2011) axiomatized extended multiplier preferences for $\sigma \geq 0$, i.e. for decision makers with robust preferences. We will characterize extended multiplier preferences, i.e. including the case of opportunity seeking ($\sigma \leq 0$). We do so by dropping uncertainty aversion (his A.5) from Strzalecki's set of axioms and by replacing results in his proof that depend on this axiom by other results that do not depend on it.

We impose the following conditions on $\succcurlyeq$:

1. *Weak order:* $\succcurlyeq$ is complete and transitive.
2. *Weak certainty independence:* for all $f, g \in \mathcal{F}$, for all $x, y \in \Delta(Z)$, and for all $\alpha \in (0,1)$, $\alpha f + (1-\alpha)x \succcurlyeq \alpha g + (1-\alpha)x \Rightarrow \alpha f + (1-\alpha)\mathrm{y} \succcurlyeq \alpha g + (1-\alpha)y$.
3. *Continuity:* for all $f, g, h \in, \mathcal{F}$, the sets $\{\alpha \in [0,1]: \alpha f + (1-\alpha)\mathrm{g} \succcurlyeq h\}$ and $\{\alpha \in [0,1]: \alpha f + (1-\alpha)\mathrm{g} \preccurlyeq h\}$ are closed.
4. *Monotonicity:* for all $f, g \in \mathcal{F}$ if $f(s) \succcurlyeq g(s)$ for all $s \in S$ then $f \succcurlyeq g$.
5. *Nondegeneracy*: there exist acts $f, g \in \mathcal{F}$ such that $f \succ g$.
6. *Weak monotone continuity:* for all $f, g \in \mathcal{F}$, for all $x \in \Delta(Z)$, and for all $\{E_n\}_{n\geq 1} \in \Sigma$ with $E_1 \supseteq E_2$ …. and $\cap_{n\geq 1} E_n = \emptyset$, $f \succ g$ implies that there exists an n_0 such that $x_{E_{n_0}} f \succ g$.
7. *Sure thing principle:* for all $E \in \Sigma$ and for all $f, g, h, h' \in \mathcal{F}$, $f_E h \succcurlyeq g_E h \Rightarrow f_E h' \succcurlyeq g_E h'$.

An event is *essential* if there exist $f, g, h \in \mathcal{F}$ such that $f_E h \succ g_E h$.

Theorem 2.1: If S has at least three disjoint essential events[1] then the following two statements are equivalent:

$\succcurlyeq$ is a continuous, nondegenerate weak order that satisfies weak certainty independence, monotonicity, weak monotone continuity and the sure thing principle.

$\succcurlyeq$ has an extended multiplier representation.

[1] If only one event is essential then the Theorem also holds but the uniqueness properties are different. If exactly two disjoint events are essential then the sure thing principle should be strengthened to the hexagon condition (Wakker 1989).

Observation 2.1: Two triples (σ, u, q) and (σ', u', q') represent the same extended multiplier preference if and only if q and q' are identical and there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u' = \alpha u + \beta$ and $\sigma' = \sigma/\alpha$.

We can distinguish the robust and the opportunity seeking approaches using Schmeidler's (1989) condition of ambiguity aversion and its counterpart of ambiguity seeking.

Definition 2.2: *Ambiguity aversion* (*seeking*) holds if for all acts f, g in $\mathcal{F}$and for all α in (0,1), $f \sim g \Rightarrow \alpha f + (1-\alpha)g \succcurlyeq (\preccurlyeq) f$.

Theorem 2.2: Under extended multiplier preferences, ambiguity aversion is equivalent to robust preferences and ambiguity seeking is equivalent to opportunity seeking preferences.

All proofs are in Appendix 2.

2.4 Measuring Extended Multiplier Preferences

Method

Strzalecki (2011, Example 3) explained how the multiplier parameter σ could be measured under the assumption that utility u is a power function. We describe an alternative method that makes no assumptions about utility and requires fewer questions. Because extended multiplier preferences are ordinally equivalent to second-order expected utility (SOEU) and our method is easier to understand under SOEU, we will use SOEU in what follows.

Suppose that a ball will be drawn from an urn with an unknown number of yellow and purple balls. Let $S = \{Y, P\}$ where Y stands for "the ball is yellow" and P for "the ball is purple". The decision maker can win either \$15 or nothing, depending on the color of the ball. Hence, $Z = \{0,15\}$. The act f_Y pays \$15 if the ball is yellow and nothing otherwise and the act f_P pays \$15 if the ball is purple and nothing otherwise. Each lottery from $\Delta(Z)$ can be written as 15_r0, where r is the probability to get 15. We scale utility so that $u(0) = 0$ and $u(15) = 15$. Then $u(15_r0) = r * 15 + (1 - r) * 0 = 15r$.

Assume that $f_Y \sim f_P \sim 15_r0$ for some probability r. We call this probability r a *matching probability* of the acts f_Y and f_P. Under SOEU, we obtain from $f_Y \sim f_P$ that $q(Y) = q(P) = ½$. The second indifference, $f_P \sim 15_r0$, then implies $\varphi_\sigma(15r) = ½\, \varphi_\sigma(15) + ½\, \varphi_\sigma(0)$. We prove in the Appendix that this equation has a unique solution σ for each value of $r \in (0,1)$. If $r = ½$, then $\sigma = 0$ and the decision maker is indifferent between an objective and a subjective probability of ½ . If $r < ½$ then $\sigma > 0$and the decision maker prefers an objective probability of ½ to a subjective probability of ½.. This corresponds to ambiguity aversion. Similarly, $r > ½$ implies ambiguity seeking ($\sigma < 0$). If $r \to 0$, preferences are extremely robust (ambiguity averse) and $\sigma \to +\infty$. If $r \to 1$, preferences are extremely opportunity seeking and $\sigma \to -\infty$.

Calibration

Observation 1 shows that the sign of the multiplier parameter does not depend on the scaling of the utility function, but its magnitude does. In the empirical study reported in Section 2.5, we scale utility such that the utility of initial wealth W (the wealth before making the decision) is 0 and that of $W + 15$ is 15. For any utility function v, the multiplier parameter σ_v can be computed from the σ that we report below using $\sigma_v = \frac{15\sigma}{v(W+15)-v(W)}$. For moderate utility curvature (which is plausible for the small increases in final wealth in our surveys), we get the approximation $\sigma_v \approx \frac{\sigma}{v'(W)}$.

2.5 Extended Multiplier Preferences in the Dutch Population

Data

The data for the first analysis come from the Dutch Longitudinal Internet Study for the Social Sciences (LISS), a representative household survey conducted by CentERdata at Tilburg University. Respondents answer survey modules, the results of which are publicly available. LISS also contains information about many demographic characteristics and economic background variables. Summary statistics of the variables that we used are in Table 5 in the Appendix.

In January 2010, 1,933 subjects participated in a survey about ambiguity preferences. The survey was designed by Dimmock, Kouwenberg and Wakker (2013). Subjects had to choose between two urns: a known urn K and an ambiguous urn A. Urn K contained 100 yellow and purple balls in known proportions. Urn A contained 100 yellow and purple balls in unknown proportions. By default, purple was the winning color, but subjects could change the winning color to yellow. Only 1% of all subjects did so, which indicates that most subjects were not suspicious and had no preference between the two winning colors. This implies that $f_Y \sim f_P$.

The survey measured the matching probability r for which subjects were indifferent between urn A and urn K with $r * 100$ balls of their winning color. Subjects made a series of choices between urn A and urn K, where urn A remained the same while the proportion of winning balls in urn K changed depending on previous choices.

At the end of the experiment, half of the sample played one randomly selected choice for real (for the other half all choices were hypothetical). A ball was drawn from the urn that the subject preferred in that choice. The subject received €15 euro if the ball had his winning color and nothing otherwise.

Results

Figure 1 shows the estimated distribution of σ using a kernel density estimate. The median value of σ was equal to 0.05, which corresponds with a matching probability of 40.6%. The distribution was centered to the right of zero and was concentrated in the ambiguity averse domain. Sixty seven percent of the subjects were ambiguity averse and 22.5% were ambiguity seeking. The boxes at the far right and left of the distribution show that for 9.6% (6.2%) of the subjects, the matching probability was in the interval [0,0.06] ([0.94,1]) which corresponds with a value of σ that exceeds 0.8 (-0.8).

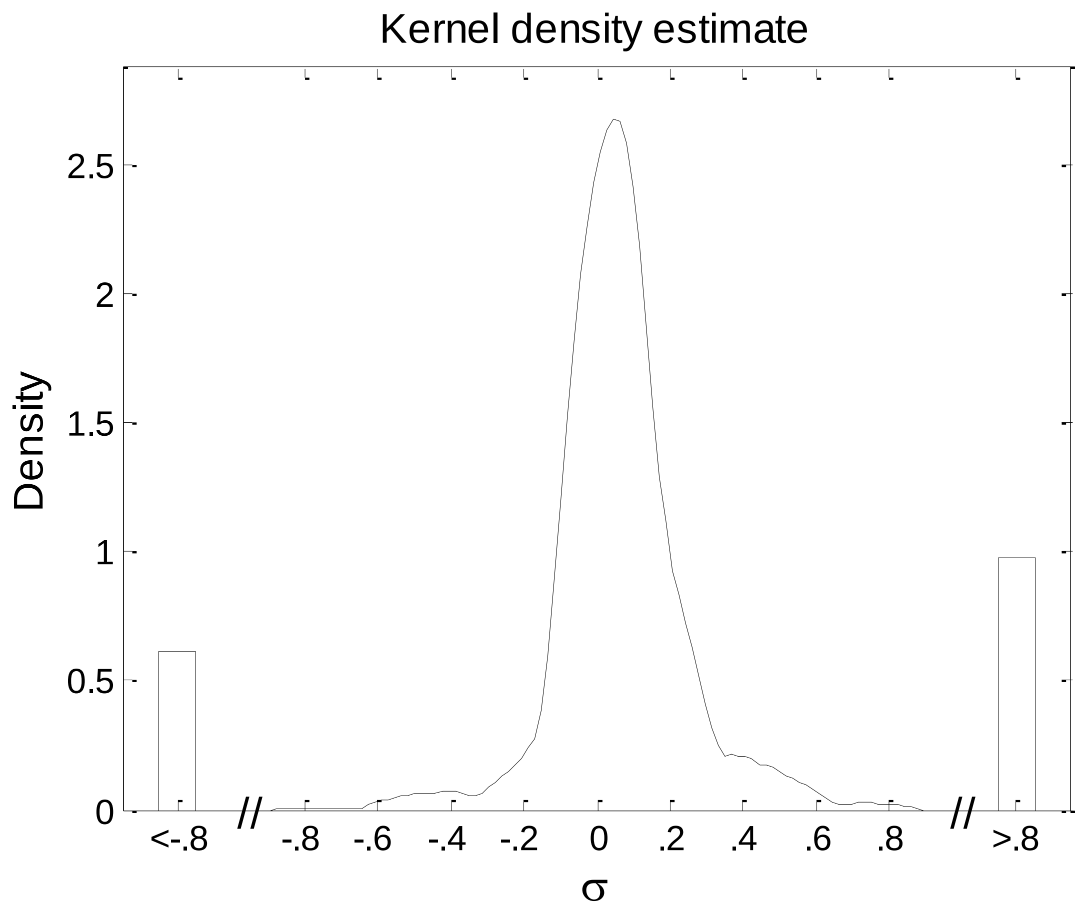


Figure 1: Kernel density estimate of subjects' σ values (NL)

The Epanechnikov function was used, with a kernel width of 0.07. The boxes at the upper and lower end indicate the proportion of subjects with σ values of greater (less) than .8 (–.8).

Table 1 answers the question whether ambiguity attitudes are related to demographic variables. In regression (1) σ is the dependent variable, while in regression (2) the absolute value of σ is the dependent variable. We also used the absolute value of σ because some effects may be correlated with the deviation from ambiguity-neutrality, which is often seen as the rational model of choice under uncertainty (Wakker 2010, p.326), rather than with the degree of ambiguity aversion.

Table 1: Regression of σ and |σ| on demographic variables(NL)

Coefficients are reported in percentage points , robust standard errors in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

	(1)	(2)
Incentivized	5.23*	0.07
	(2.81)	(2.38)
Number of children	1.19	0.24
	(1.47)	(1.24)
Female	1.61	-2.38
	(3.08)	(2.61)
Age	-0.26***	0.44***
	(0.10)	(0.08)
Married/Living together	-0.85	-0.32
	(3.48)	(2.95)
Very urban	2.05	-1.13
	(2.93)	(2.48)
High income	3.23	-5.88**
	(3.34)	(2.83)
Education		
Vocational	-1.50	-7.57***
	(3.12)	(2.65)
University	-0.90	-12.03**
	(5.61)	(4.76)
Constant	17.83***	19.95***
	(6.76)	(5.73)
Adjusted R^2	0.01	0.03
N	1,821	1,821

The only variable that had an effect in both regressions was age. Older respondents were more ambiguity seeking and also deviated more from ambiguity neutrality. This suggests that they had more extreme ambiguity attitudes. Similarly, better educated respondents deviated less from ambiguity neutrality, which seems consistent with the finding that people with higher cognitive abilities deviate less from models of rational choice (Frederick 2005, Dohmen et al. 2010). Education had no effect on ambiguity aversion. Real incentives led to marginally more ambiguity aversion (p = 0.06) but they had no effect on deviations from ambiguity neutrality. The absence of a gender effect is perhaps surprising given that women are usually found to be more risk averse than men (Croson and Gneezy 2009). Our findings suggest that this gender effect for risk attitudes does not carry over to ambiguity attitudes, which is consistent with Sutter et al. (2013) who found no relation between ambiguity attitudes and risk attitudes.

Table 2: Probit regression of NL market participation on σ and demographic variables

In regressions (1) and (2), σ is used as independent variable; |σ| is used in regressions (3) and (4). Mean marginal effects in percentage points are reported; robust standard errors are in parentheses.
* significant at 10% ** significant at 5% *** significant at 1%.

	(1)	(2)	(3)	(4)
σ or $\|\sigma\|$	-1.30	-0.78	-4.77**	-3.89**
	(1.53)	(1.49)	(1.86)	(1.81)
Incentivized		1.21		1.17
		(1.76)		(1.76)
Number of children		2.18**		2.18**
		(0.92)		(0.92)
Female		-6.28***		-6.45***
		(1.93)		(1.93)
Age		0.30***		0.32***
		(0.06)		(0.06)
Married/Living together		1.39		1.33
		(2.22)		(2.22)
Very urban		-0.28		-0.39
		(1.83)		(1.84)
High income		7.37***		7.01***
		(2.06)		(2.07)
Education				
Vocational		6.20***		5.95***
		(1.96)		(1.97)
University		18.51***		18.05***
		(3.15)		(3.16)
Pseudo R^2	0.001	0.07	0.004	0.08
N	1,821	1,821	1,821	1,821

We also explored (using probit regressions) whether stock market participation was related to ambiguity aversion. Table 2 shows the results. Regressions (1) and (2) use σ as an independent variable, and regressions (3) and (4) use |σ|. Contrary to the theoretical literature, which predicts that ambiguity aversion leads to less stock market participation, ambiguity aversion had no effect on participation. However, the deviation from ambiguity-neutrality did have a negative effect on stock market participation: subjects who were more ambiguity averse *or* ambiguity seeking were less likely to participate in the stock market. This effect remained significant after we controlled for demographic variables. The Table further shows that males, older respondents, those with children and those with high income and education were more likely to participate in the stock market.

2.6 Extended Multiplier Preferences in the US

Data

The U.S. data come from the American Life Panel (ALP), a household survey conducted by the RAND Corporation. It includes demographic characteristics and data about stock market participation.

In March and April 2012, 3,290 subjects answered a survey about ambiguity preferences designed by Dimmock, Kouwenberg, Mitchell and Peijnenburg (2013). The survey was in many respects similar to the LISS survey except that the balls were orange and purple instead of yellow and purple, that subjects could not choose their winning color, and that all subjects played out one randomly selected question for real.

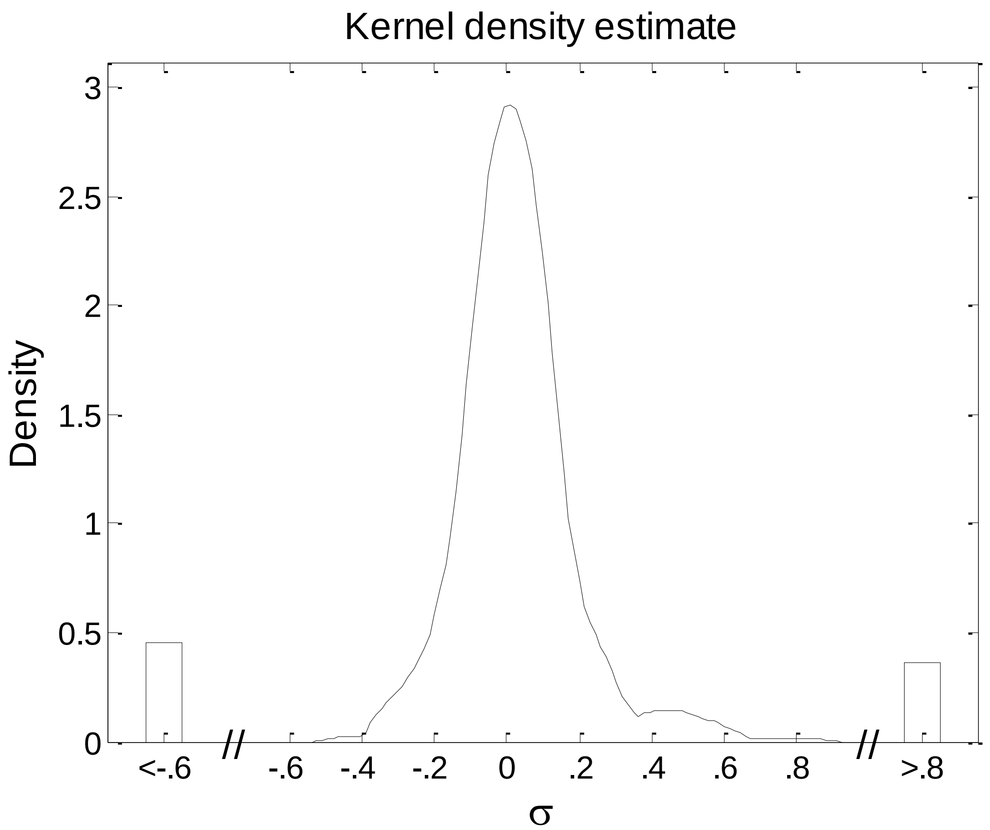


Figure 2: Kernel density estimate of subjects' σ values (US)

The Epanechnikov function was used, with a kernel width of 0.7. The boxes at the upper and lower indicate the proportion of subjects with σ values greater (less) than .8 (–.6).

Results

Figure 2 shows the estimated distribution of σ for the U.S. sample. The median value of σ was equal to 0.02, which corresponds with a matching probability of 47.0%. A slight majority (52.2%) of the subjects was ambiguity averse and more than one third (35.9%) was ambiguity seeking. These values indicate that there was less ambiguity aversion in the US sample than in the Dutch sample. The proportion of subjects with extreme ambiguity attitudes was also lower than in the Dutch survey: 3.6% had σ greater than .8 and 4.5% had σ smaller than –.6.[2]

Table 3 shows the results of regressions of σ (regression (1)) and |σ| (regression (2)) on the demographic variables. As in the Dutch sample, better educated respondents deviated less from ambiguity neutrality and income had no effect on ambiguity

[2] These thresholds are not of the same absolute value due to an asymmetry in the question design of Dimmock et al. (2013).

attitudes. In addition, we observed a significant effect of ethnicity on ambiguity attitudes: Hispanics were more ambiguity averse and farther from ambiguity neutrality than whites.

Table 3: Regression of σ and $|\sigma|$ on demographic variables (US)

Coefficients are reported in percentage points, robust standard errors are in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

	(1)	(2)
Number of children	0.80*	0.55
	(0.46)	(0.41)
Female	-4.15***	0.29
	(1.32)	(1.16)
Age	-0.04	0.09*
	(0.05)	(0.05)
Married/Living together	0.39	-1.35
	(1.41)	(1.24)
Ethnicity		
Non-hispanic black	3.09	2.74
	(2.17)	(1.91)
Hispanic & other	4.26**	4.97***
	(1.73)	(1.52)
High income	1.31	-1.56
	(1.46)	(1.28)
Education		
College, no degree	1.12	-2.41
	(1.87)	(1.65)
College degree	3.22*	-4.22***
	(1.72)	(1.51)
Constant	4.09	16.13***
	(3.42)	(3.00)
Adjusted R^2	0.01	0.01
N	3,217	3,217

There were also some differences between the US and the Dutch samples: in the US sample, there was no age-effect and there was some effect of gender with men being more ambiguity averse than women. The higher ambiguity aversion for men than for women confirms the conclusion drawn from the Dutch data that the widely-observed higher risk aversion for women does not imply that women are also more ambiguity averse.

Table 4 shows the results from the probit regressions of ambiguity attitude on stock market participation. As in the Dutch sample, ambiguity aversion has no effect on stock market participation, but the deviation from ambiguity-neutrality does with people who are more ambiguity averse or more ambiguity seeking less likely to participate in the stock market. This effect remains when we include demographic variables. Table 4 further shows that stock market participation increased with age, income and education, and it was lower for blacks and Hispanics.

Table 4: Probit regression of US market participation on *σ* and demographic variables

Regressions (1) and (2) use σ as a covariate, regressions (3) and (4) use |σ|. Mean marginal effects in percentage points are reported; robust standard errors are in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

	(1)	(2)	(3)	(4)
σ / |σ|	-0.05	-0.10	-8.87***	-4.12*
	(2.11)	(2.10)	(2.54)	(2.44)
Number of children		-0.69		-0.68
		(0.58)		(0.58)
Gender		-1.47		-1.41
		(1.48)		(1.47)
Age		0.40***		0.40***
		(0.06)		(0.06)
Married & Living together		2.55		2.52
		(1.65)		(1.64)
Ethnicity				
Non-Hispanic black		-21.33***		-21.31***
		(3.11)		(3.12)
Hispanic & other		-15.50***		-15.28***
		(2.22)		(2.22)
High income		15.87***		15.79***
		(1.62)		(1.62)
Education				
College, no degree		3.91		3.84
		(2.43)		(2.43)
College degree		15.64***		15.47***
		(2.11)		(2.11)
Pseudo R^2	<0.001	0.16	0.004	0.16
N	3,217	3,217	3,217	3,217

2.7 Concluding Remarks

Multiplier preferences, proposed by Hansen and Sargent (2001), are a popular model in macroeconomics and finance. In its original form, multiplier preferences only capture ambiguity aversion, which make them less suitable for applications at the micro level where substantial ambiguity seeking has also been observed. This chapter extends multiplier preferences to include ambiguity seeking and it gives a preference foundation for these extended multiplier preferences. We also show how extended multiplier preferences can be measured and thereby obtain an axiomatically-founded measure of ambiguity aversion that can easily be applied in empirical studies and that captures the substantial heterogeneity in ambiguity attitudes that typically exists in micro data.

We applied our method to two large scale representative surveys, one from the Netherlands and one from the US. In both samples a substantial fraction of the respondents was ambiguity seeking, which illustrates the desirability of our extension of multiplier preferences. More educated respondents deviated less from ambiguity neutrality. In contrast with the theoretical literature, we observed that stock market participation decreased not only with ambiguity aversion, but also with ambiguity seeking. Perhaps deviations from ambiguity neutrality signal deviations from rationality with more irrational respondents less likely to participate in the stock market.

Appendix 2: Proofs and Summary Statistics

Lemma 2.1: Preferences $\succcurlyeq$ are extended multiplier preferences if and only if there exists $\sigma \in \mathbb{R}$ such that $\succcurlyeq$ can be represented by SOEU with $q \in \Delta(S)$ and $\varphi = \varphi_\sigma$.

Proof:

The equivalence between robust preferences and $\varphi(t) = -e^{-\sigma t}$ has been shown by Strzalecki (2011). It is based on Proposition 1.4.2 of Dupuis and Ellis (1997) stating that for all countably additive probability measures $q \in \Delta(S)$ and for all Σ-measurable functions v:

$\min_{p\in\Delta(S)} \int_S \ v(s)dp(s) + \frac{1}{\lambda}R(p||q) = \varphi_\lambda^{-1}\left(\int_S \ \varphi_\lambda\big(v(s)\big)dq(s)\right)$.

For $\sigma < 0$, we apply this formula to $v = -u \circ f$ and $\lambda = -\sigma$ and we obtain:

$$\begin{aligned}\max_{p\in\Delta(S)} \int_S \ u\big(f(s)\big)dp(s) + \frac{1}{\sigma}R(p||q) \quad &= \quad -\left[\min_{p\in\Delta(S)} \int_S \ v(s)dp(s) + \frac{1}{\lambda}R(p||q)\right] \\ &= \quad -\varphi_\lambda^{-1}\left(\int_S \ \varphi_\lambda\big(v(s)\big)dq(s)\right) \\ &= \quad \varphi_\sigma^{-1}\left(\int_S \ \varphi_\sigma\left(u\big(f(s)\big)\right)dq(s)\right).\end{aligned}$$

The last equality follows from $\varphi_\sigma^{-1}(t) = -\frac{\ln(t)}{\sigma} = \frac{\ln(t)}{\lambda} = -\varphi_\lambda^{-1}(-t)$ and

$\varphi_\lambda\big(v(s)\big) = -e^{-\lambda v(s)} = -e^{-\sigma u(f(s))} = -\varphi_\sigma\left(u\big(f(s)\big)\right)$.

Hence, both robust and opportunity seeking preferences are equivalent to SOEU with an exponential φ function. □

Proof of Theorem 2.1:

(ii) $\Rightarrow$ (i). Because (ii) is a normalized niveloid that represents $\succcurlyeq$ and u is nonconstant and affine, Lemma 28 in Maccheroni, Marinacci and Rustichini (2006) implies that $\succcurlyeq$ is a constinuous, nondegenerate weak order that satisfies weak certainty independence and monotonicity. Because q is countably additive, $\succcurlyeq$ satisfies uniform continuity by Theorem 5.4 in Krantz et al. (1971). Finally, by Proposition 1.4.2 in Dupuis and Ellis (1997), (ii) is equivalent to a second order expected utility representation. Consequently, the sure thing principle must hold.

We show that (i) $\Rightarrow$ (ii) by closely following Strzalecki's proof without imposing uncertainty aversion. First we introduce some new notation. Let $B_0(\Sigma)$ denote the set of all real-valued Σ-measurable simple functions[3] and let $B_0(\Sigma, \mathrm{K})$ denote the set of functions in $B_0(\Sigma)$ that take values in a convex set $K \subseteq \mathbb{R}$. Let $\Phi_3$denote the set of finite partitions of S that contain at least three essential events. For all $G \in \Phi_3$, let $\mathcal{A}(G)$ be the algebra generated by G and let $\mathcal{F}_G$ denote the set of acts in $\mathcal{F}$ that are measurable with respect to $\mathcal{A}(G)$.

[3] A function is simple if it takes no more than countably many distinct values.

By Lemmas 25 and 28 of Maccheroni et al. , there exist a real-valued nonconstant affine function u on $\Delta(Z)$ and a normalized real-valued functional $I: B_0(\Sigma, \mathcal{U}) \to \mathbb{R}$ where $\mathcal{U}$ is the range of $u(\Delta(Z))$ and such that for all acts $f, g \in \mathcal{F}$, $f \succcurlyeq g$ iff $I(u \circ f) \geq I(u \circ g)$ and $I(\alpha\psi + (1-\alpha)k) = I(\alpha\psi) + (1-\alpha)k$ for all $\psi \in B_0(\Sigma, \mathcal{U})$, $k \in \mathcal{U}$ and $\alpha \in (0,1)$.

Theorem 1 in Grant, Polak, and Strzalecki (2009) ensures that for finite S $\succcurlyeq$ can be represented by $f \longmapsto \sum_{s\in S} v_s(u(f(s)))$ with u nonconstant and affine and with range $\mathcal{U}$ and v_s continuous, nondecreasing, and with at least three v_s nonconstant. Weak certainty independence then ensures that indifference curves in the utility space are parallel and have common supporting hyperplanes at the set of constant vectors in $\mathcal{U}^S$. By the proof of Theorem 3 in Grant et al. it follows that for all $G \in \Phi_3$the restriction of $\succcurlyeq$ to $\mathcal{F}_G$ can be represented by $f \longmapsto \sum_{s\in S} p_G(s)\varphi_G(u_G(f_s))$ with u_G nonconstant and affine, φ_G continuous and strictly increasing, and measure $p_G: \mathcal{A}(\mathrm{G}) \to [0,1]$ such that at least three events in G are nonzero. In applying Theorem 3, we replace uncertainty aversion and their Axiom A.7 by weak certainty independence. Uncertainty aversion is used in the application of Theorem 3 in Debreu and Koopmans (1982) to derive differentiability of the functions v_s. However, as noted by Grant et al. and Maccheroni et al. (p.1475, 1491), weak certainty independence implies Lipschitz continuity and hence differentiability. By Theorem 4 in Strzalecki (2011), the proof of which does not use uncertainty aversion, $\succcurlyeq$ can be represented by second order expected utility $f \longmapsto \int_S \varphi(u(f_s))dq(s)$ with $q \in \Delta(Z)$ and φ continuous and strictly increasing. q is countably additive by uniform continuity (Villegas 1964, Theorem 1). Moreover, if (u, φ, q) and (u', φ', q') both represent $\succcurlyeq$ then there exist $\alpha, A > 0, \beta, B \in \mathbb{R}$ such that $q^{'} = q, u^{'} = \alpha u + \beta, \varphi^{'}(\alpha r + \beta) = A\varphi(r) + B$ for all r in $\mathcal{U}$.

I represents $\succcurlyeq$ and is translation invariant, i.e. for all $f, g \in \mathcal{F}$ and k such that $f(s) + k, g(s) + k \in \mathcal{U}$ for all $s \in S$, $I(u \circ f) \geq I(u \circ f)$ iff $I(u \circ f + k) = I(u \circ f) + k \geq I(u \circ f) + k = I(u \circ f + k)$. It then follows that for all acts $f, g \in \mathcal{F}$ and k such that $f(s) + k, g(s) + k \in \mathcal{U}$ for all $s \in S$, $\int_S \varphi(u(f(s)))dq(s) \geq \int_S \varphi(u(g(s)))dq(s)$ iff $\int_S \varphi(u(f(s) + k)dq(s) \geq \int_S \varphi(u(g(s) + k)dq(s)$.

Hence, (u, φ, q) and (u, φ_k, q) defined by $\varphi_k(l) = \varphi(l + k)\ \forall l, l + k \in \mathcal{U}$ are both SOEU representations of $\succcurlyeq$. Consequently, $\varphi(l + k) = A(k)\varphi(l) + B(k)$. Because φ is nonconstant, if $\mathcal{U}$ is unbounded, it follows from Corollary 1 in Aczél (1966, Section 3.1.3) that φ equals φ_σ. If $\mathcal{U}$ is bounded then because φ is nonconstant Theorem 4 in Aczél (2005) implies $\varphi = \varphi_\sigma$ on the interior of $\mathcal{U}$. Because φ is continuous, the extension to all of $\mathcal{U}$ follows.

By Proposition 1.4.2 in Dupuis and Ellis (1997) and Lemma 2.1, we then obtain the extended multiplier representation. □

Proof of Observation 2.1:

The proof of Theorem 2.1 already showed that the probability measure q is unique and that the utility function u is unique up to positive affine transformations. We also know that for $A > 0$ and $B \in \mathbb{R}$, $\varphi' = A\varphi + B$. Because $e^{-\sigma' u'} = e^{-\sigma'(\alpha u + \beta)} = e^{-\beta} e^{-\alpha\sigma' u}$, it follows from the uniqueness properties of φ that $\sigma' = \frac{1}{\alpha}\sigma$. □

Proof of Theorem 2.2:

Ambiguity aversion states that preferences are convex. Hence it is equivalent to a concave representation. Since u is linear with respect to mixture of lotteries, ambiguity aversion is equivalent to the SOEU with φ concave, which means $\sigma \geq 0$. The opposite reasoning applies to ambiguity seeking. □

Proof that there is a unique solution σ for each value of r.

$f_Y \sim f_P$ and $f_P \sim 15_r 0$ jointly imply $\varphi_\sigma(15r) = ½\varphi_\sigma(15) + ½\varphi_\sigma(0)$, which is equivalent to $15r = ½(15) + ½(0)$ if $\sigma = 0$ and to
$\exp(-15\sigma r) = ½\exp(-15\sigma) + ½\exp(0)$ otherwise. Hence,

$r = ½$ if $\sigma = 0$

$r = -\frac{\ln(½\exp(-15\sigma) + ½)}{15\sigma}$ if $\sigma \neq 0$

r is continuous as a function of σ for $\sigma \neq 0$. For notational convenience and without loss of generality, we write $\tilde{\sigma} = 15\sigma$. Because

$$\lim_{\tilde{\sigma}\to 0^-} r(\tilde{\sigma}) = \lim_{\tilde{\sigma}\to 0^+} r(\tilde{\sigma}) = \lim_{\tilde{\sigma}\to 0^+} -\frac{\ln\left(\frac{1+\exp(-\tilde{\sigma})}{2}\right)}{\tilde{\sigma}}$$

$$= \lim_{\tilde{\sigma}\to 0^+} -\frac{\frac{2}{1+\exp(-\tilde{\sigma})}\cdot -½ \cdot \exp(-\tilde{\sigma})}{1} = ½ = r(0)$$

r is also continuous at zero.

Differentiating r with respect to $\tilde{\sigma}$ gives after some rewriting

$$\frac{dr}{d\tilde{\sigma}} = \tilde{\sigma}^{-2}\left[\left(\frac{\tilde{\sigma}}{\exp(\tilde{\sigma})+1}\right) + \ln\left(\frac{1+\exp(-\tilde{\sigma})}{2}\right)\right]. \tag{2A.1}$$

Because $\tilde{\sigma}^{-2} > 0$, the sign of (2A.1) depends on the sign of

$$D(\tilde{\sigma}) = \left(\frac{\tilde{\sigma}}{\exp(\tilde{\sigma})+1}\right) + \ln\left(\frac{1+\exp(-\tilde{\sigma})}{2}\right). \tag{2A.2}$$

Now,

$$D(0) = \left(\frac{0}{\exp(0)+1}\right) + \ln\left(\frac{1+\exp(-0)}{2}\right) = 0 + \ln\left(\frac{2}{2}\right) = 0$$

and,

$$\frac{dD(\tilde{\sigma})}{d\tilde{\sigma}} = -\frac{\tilde{\sigma}\exp(\tilde{\sigma})}{(\exp(\tilde{\sigma})+1)^2}. \tag{2A.3}$$

Because (2A.3) is negative for $\tilde{\sigma} > 0$ and positive for $\tilde{\sigma} < 0$, $D(\tilde{\sigma})$ is negative everywhere except at $\tilde{\sigma} = 0$. Thus r is a continuous and strictly decreasing function of $\tilde{\sigma}$ and the function $r(\tilde{\sigma})$ is a one-to-one function, which shows that any solution that we find is unique.

In the limits, $r(\tilde{\sigma})$ goes to 0 and 1:

$$\lim_{\tilde{\sigma}\to +\infty} r(\tilde{\sigma}) = \lim_{\tilde{\sigma}\to +\infty} -\frac{\ln\left(\frac{1+\exp(-\tilde{\sigma})}{2}\right)}{\tilde{\sigma}} = \lim_{\tilde{\sigma}\to +\infty} -\frac{\ln(½)}{\tilde{\sigma}} = 0$$

$$\lim_{\tilde{\sigma}\to -\infty} r(\tilde{\sigma}) = \lim_{\tilde{\sigma}\to -\infty} -\frac{\ln\left(\frac{1+\exp(-\tilde{\sigma})}{2}\right)}{\tilde{\sigma}} = \lim_{\tilde{\sigma}\to -\infty} -\frac{\frac{2}{1+\exp(-\tilde{\sigma})}\cdot -½ \cdot \exp(-\tilde{\sigma})}{1}$$

$$= \lim_{\tilde{\sigma}\to -\infty} \frac{\exp(-\tilde{\sigma})}{1+\exp(-\tilde{\sigma})} = 1.$$

Consequently, r is a continuous and strictly decreasing function from $\mathbb{R}$ to $(0,1)$. By the intermediate value theorem, there is a unique solution $\tilde{\sigma}$ for each $r \in (0,1)$. □

Table 5: Descriptive statistics of the Dutch and US dataset

Apart from σ and age, all variables are dummy variables that can take on values 0 and 1. The two datasets do not have exactly the same variables: ethnicity is not available in LISS and the dummy for someone living in an urban area is not available in ALP. Real incentives are used for all subjects in ALP, so a dummy for incentives is not necessary. The variable "High income" is equal to one if a respondent's income is above the sample median.

	The Netherlands		**United States**	
	Mean	SD	Mean	SD
Ambiguity attitude σ	0.10	0.60	0.05	0.36
Market participation	0.18	0.38	0.25	0.43
Incentivized	0.48	0.50		
Demographic Variables				
Female	0.53	0.50	0.60	0.49
Age	48.51	16.95	47.34	13.55
Married/Living together	0.77	0.42	0.60	0.49
Very urban	0.39	0.49		
High income	0.49	0.50	0.49	0.50
Number of children	0.80	1.12	1.21	1.53
Education				
Secondary & lower	0.49	0.50	0.22	0.42
Vocational	0.43	0.50	0.26	0.44
University	0.08	0.27	0.52	0.50
Ethnicity				
Non-Hispanic white			0.69	0.46
Non-Hispanic black			0.11	0.31
Hispanic & other			0.21	0.40

Chapter 3 | Wall Street vs. Main Street: an Evaluation of Probabilities

This Chapter considers whether the views of survey respondents regarding the likelihood of stock index return exceeding specific thresholds are comparable to market views indicated by index options with strikes at analogous thresholds. This study is the first to highlight the correspondence between the wording of subjective response questions found in surveys about expected future returns and the design of financial options. While we do find a relationship between the two sources of views, the association is not one-for-one. We find a closer association for those demonstrating a better understanding of the laws of probability, suggesting that numeracy affects the accuracy of an elicited response.

This Chapter is based on "Wall Street vs. Main Street: an Evaluation of Probabilities", co-authored by Robin Lumsdaine. 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, Andrei Shleifer, Timo Teräsvirta, Martijn van den Assem, Peter Wakker, and Bob Willis, as well as seminar participants at American University, Erasmus University, the University of Bournemouth, 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.

3.1 Introduction

This Chapter compares people's beliefs about future stock market returns as elicited through surveys with those derived from option prices. In performing this comparison, we contribute to the evidence that suggests expectations of professionals and the general population differ (e.g., Mankiw, Reis, and Wolfers 2004) and that links expectations of private investors to financial decisions (e.g., Hoffmann, Post, and Pennings 2014, forthcoming). Using the results from Breeden and Litzenberger (1978), we obtain option-implied probabilities that match the characteristics specified in questions that survey respondents are asked about future stock market returns. We specify an econometric model to link these two sets of beliefs ("Wall Street" and "Main Street", respectively), controlling for a variety of observable characteristics. We find that survey expectations do seem to be related to the beliefs of Wall Street and that this link is stronger for those with greater probabilistic understanding.

Our results add to the literature in the following areas: (1) the expectations of the general population versus those of professional forecasters/traders, (2) the information content of subjective (survey) expectations, (3) the tendency of survey respondents to report focal points (clustering around rounded numbers) when asked probabilistic questions, and (4) whether increased financial literacy improves financial expectations and decisions. Although these areas have all been studied extensively, no study (to the best of our knowledge) has drawn the connection between subjective expectations of a specified return threshold and corresponding option strike levels, as we do in this Chapter.

The Chapter proceeds as follows: Section 3.2 contains a brief review of literature related to the four areas above. Section 3.3 describes the data construction and descriptive statistics of the main variables. Section 3.4 describes the model used to examine the link between Wall Street and Main Street. Section 3.5 analyzes the regression results and discusses their implications. Section 3.6 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. We consider the robustness of our results to various assumptions in Section 3.7. The final Section concludes. Supplemental material containing information on the sample construction, the construction of Wall Street probabilities, detailed descriptive statistics, the derivation of the likelihood function and additional analyses is provided in a series of Appendices.

3.2 A Brief Review of Related Literature

We are interested in whether beyond the popular rhetoric there is really a divide between what Wall Street and Main Street think. Those that would argue against a divide might appeal to theories of information flow to argue that financial market fluctuations are merely the aggregate result of individual investor decisions. In addition, standard finance theories indicate management decisions of publicly-traded firms are a direct result of the desire to maximize shareholder value and therefore investment decisions are a reflection of the views of Main Street citizens. Even beyond an efficient markets framework, feedback and herding models in behavioral finance would suggest that Main Street beliefs are influenced by what happens on Wall Street (e.g., Hirshleifer 2001, 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). Those who have documented a link between expectations and returns emphasize the importance of eliciting expectations from financial market participants rather than the general population (e.g., Bacchetta, Mertens, and van Wincoop 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 subgroups of investors often differ systematically (e.g., Barber and Odean 2000, Coval, Hirshleifer, and Shumway 2005). Thus it is not at all apparent ex ante whether survey expectations from Main Street would reflect Wall Street views, and if so, to what extent.

Researchers using survey data to elicit expectations about future equity returns find substantial heterogeneity across individuals (Brennan, Cao, Strong, and Xu, 2005; Ben-David, Graham, and Harvey 2010; Dominitz and Manski, 2011; Hudomiet, Kézdi and Willis, 2011), whether those surveys cover professional forecasters or members of the general population. This heterogeneity has been linked to a variety of demographic characteristics (race, gender, education), financial knowledge or experience, behavioral biases such as the disposition effect (Odean 1998), optimism, as well as many other explanations, and has in turn been used to explain heterogeneous equity investment decisions (Kézdi and Willis, 2003, 2011). A number of researchers also explore whether this heterogeneity is related to, or reflected in, market measures of uncertainty (Anderson, Ghysels and Juergens 2005), such as the VIX volatility index (Graham and Harvey 2001, 2007), and/or disagreement (Rich Song and Tracy 2012). There is also substantial evidence in non-finance contexts 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 that adjustments to survey data to account for such aspects are necessary to improve inference (Bassett and Lumsdaine, 2000; Lillard and Willis, 2001).

Beyond studying the relationship between survey expectations and subsequent realizations, 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). In the context of equity returns, most research using survey expectations has focused on the views of "informed" investors, i.e., those that are active in the financial markets. Two important exceptions are Greenwood and Shleifer (2014), who use the University of Michigan survey of consumers, along with a number of other investor-based surveys, and Hurd and Rohwedder 2012, who use the same data we consider to identify correlations between survey expectations and subsequent equity returns. Rather than taking a stand on whether informed investors or survey respondents representative of the general population are a more appropriate subsample, we consider both by examining in detail the link between the Wall Street and Main Street views.

3.3 Data

We use the American Life Panel (ALP) for our analysis. An internet panel with about 6000 panel members, the ALP contains more than 300 survey modules administered by the RAND Corporation. Responses to each of these survey modules are publicly available. The Household information module contains a number of demographic characteristics of respondents such as age, race, gender, marital status, and education. Sampling weights are assigned such that the weighted distribution is representative of the U.S. population with regards to socio-demographic variables.[4] Throughout the Chapter, 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 Chapter 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. 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 3A.

[4] Weights are determined by the RAND corporation via an iterative (raking) process until the weighted distribution is sufficiently close to the target distribution (i.e., the Current Population Survey).

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.

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.[5] 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. 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 Chapter, however, it is necessary to assume that the responses given represent respondents' subjective probability that the nominal (not inflation-adjusted) level of

[5] 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.

the DJIA in one year will have increased (similarly, will have increased by more than 20% or more than -20%) relative to 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.

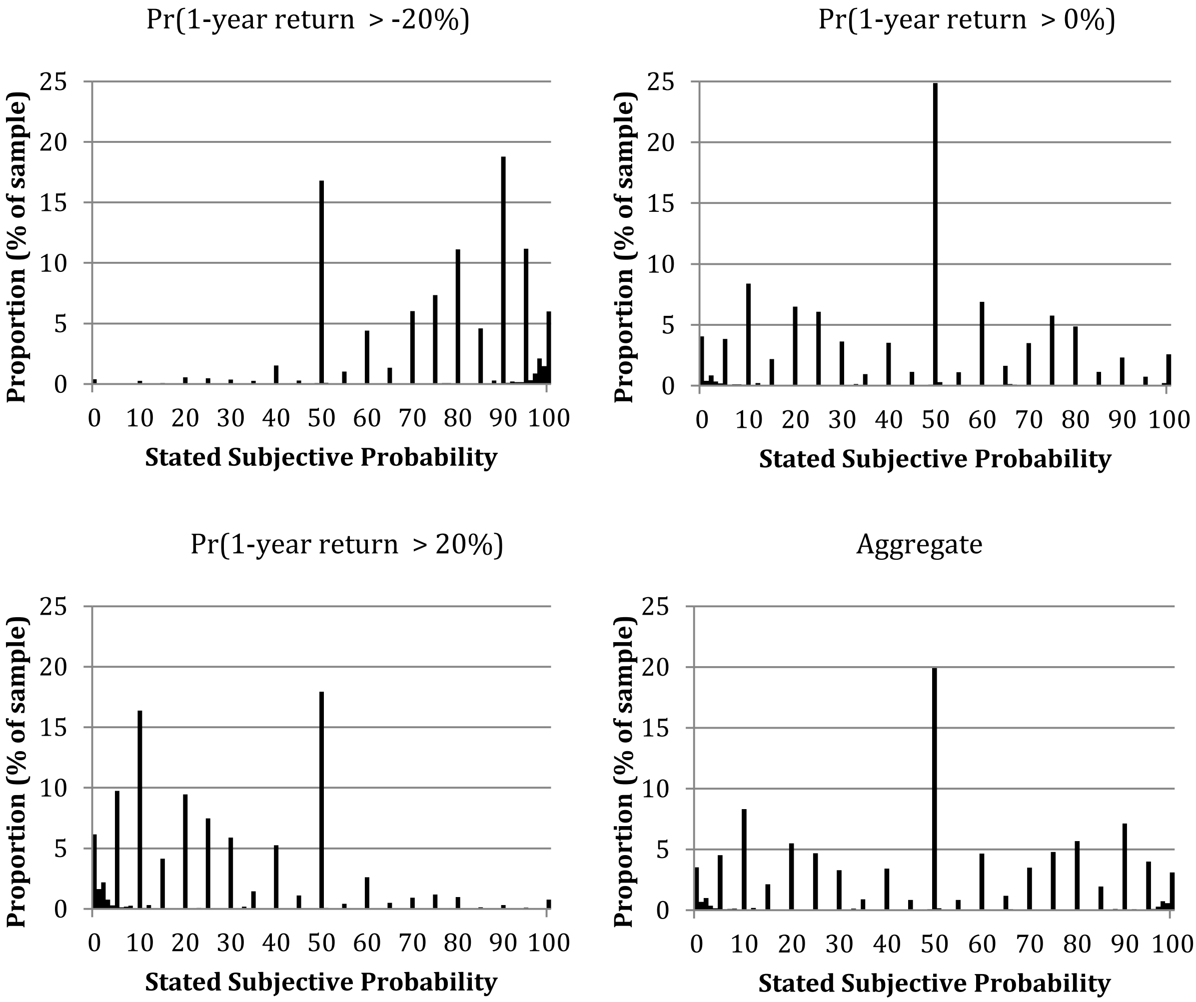


Figure 3: Frequency of responses to probabilistic questions

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.

Figure 3 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.[6] For the three questions in this Chapter, 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.

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 Chapter we adopt a fairly basic approach so as not to obscure the main question of interest (the degree of relationship between Wall Street and Main Street beliefs). For expositional purposes, we suppress the 't' subscript unless necessary for clarity but emphasize that *all* parameters in the computation of option-implied probabilities are time-varying.

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):

[6] 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), Manski and Molinari (2010), and Dominitz and Manski (2011).

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

with Φ the standard normal cumulative distribution function, r the (continuous) time-varying risk-free rate over the period *[t,T]*, q the (continuous) time-varying dividend rate over the same period and σ the time-varying 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:

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

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.[7] 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 3B). 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

[7] 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.

risk premium, is explored in Appendix 3D). The volatility (σ) for each specified strike price is the volatility implied by the option prices. In particular, for each survey day t, 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.

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. Specifically, for a given option threshold 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 specific person corresponds to the number of waves in which the person provided a Main Street probability at that threshold. Table 6 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 6: 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

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%).[8]

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 the 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 result of the analytical design, since all participants on the same day are assigned the same Wall Street probability (which eliminates any intraday Wall Street volatility).

[8] 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).

3.4 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

$$f(p^*) = X_1\beta_1 \tag{3.1}$$

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. For example, as noted earlier, nearly 20% of respondents in our sample give a focal response of 50; a number of articles (e.g., Fischhoff and Bruine de Bruin, 1999) note that such a response should be considered as distinct from other responses as it likely indicates uncertainty on the side of the respondent rather than a true subjective probability of 50%. In addition, responses of zero or one hundred cannot reflect the true probability (since the true probability distribution lies in the open unit interval) and contribute zero mass to the likelihood calculation. Such focal responses therefore might be the result of a lesser ability to express oneself in probabilistic terms. More generally, the propensity to give one of these three responses (zero, 50, or one hundred, hereafter referred to as "focal responses", following Lillard and Willis 2001) varies by observable demographic characteristics. This observation motivates the decision to explicitly model the probability of giving a focal response as a function of observable covariates in the model. Similar to Hurd, McFadden and Gan (1998), these focal responses are modeled via a latent variable w^*:

$$w^* = X_2\beta_2 + \eta \tag{3.2}$$

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 $w^* \le 0$, 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).[9] To account for this possible uncertainty the model includes a third equation that describes an additional latent variable v^*:

$$v^* = X_3\beta_3 + \xi \tag{3.3}$$

with ξ an error term. Conditional on a focal response being given ($w^* \le 0$), a response of 50 represents uncertainty and is given if and only if $v^* > 0$. When $v^* \le 0$, 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:

$p = 0 \quad \text{if} \quad \tilde{p} \le \psi$

$p = 1 \quad \text{if} \quad \tilde{p} > \psi$

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

[9] 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).

binary data (see, e.g., Albert and Chib, 1993). With this link function, the true beliefs are given by:

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

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 Chapter, the stochastic component enters through the subjective responses $\tilde{p}$ (true belief p^* with error) that are assumed to come from a beta distribution. This distribution 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 to model probabilistic responses in, e.g., 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):

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

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

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

with Γ the gamma function.

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

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

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:

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

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

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

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 3C).

3.5 Results

The results from the estimation of all three equations are combined in Table 7. 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 3.1). The next three columns describe the likelihood that a respondent gives a focal response (equation 3.2) and the final three columns describe the likelihood that a respondent gives a response of 50, conditional on giving a focal response (equation 3.3). 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).

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 6 (-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 expectations of future 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.[10] 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.

[10] These findings complement those found in Greenwood and Shleifer (2014). In particular, in their study a 20% return during the past year corresponds to an approximately 2.1 percentage point increase in expectations of the next year's return using their measure from the Gallup survey of investors. The time period of our sample covers the gap in the Gallup Survey due to it being temporarily discontinued between November 2009 and February 2011.

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 our main variable of interest, 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. That Main Street probabilities respond only partially to a change in Wall Street's probability is consistent with a variety of behavioral theories on the partial updating of beliefs (see Hirshleifer 2001 for a discussion of these in the context of financial markets). 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 3.7.

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.

Table 7: 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. * Significant at 5%.

Observations 139,327 | Log likelihood -66,968 | Average log likelihood -0.48

	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 7: Baseline regression results (continued)

<u>Financial Market Characteristics</u>									
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
<u>Additional parameters</u>									
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.

Consistent with Figure 3, 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.

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).[11]

[11] Throughout this Chapter, 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.

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., 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). We consider this possibility by exploring in Section 3.7 the robustness of our results to the inclusion of the v* equation, and to the inclusion of zero and 100 in our definition of "focal"; the results are unchanged.

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 to the *>Minus20* question, rather than zero or 100. This result is consistent with the intuition that during the (post-financial crisis) sample 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).

3.6 Inconsistent Responses

In addition to the econometric challenge of focal responses, addressed via the model, a relatively large proportion of person-wave responses are inconsistent with the laws of probability, perhaps suggesting that some respondents did not fully understand the questions. Ben-David, Graham, and Harvey (2010) document miscalibration of survey respondents with respect to probability distributions in their sample of Chief Financial Officers; such miscalibration is likely to be more severe among the general Main Street population. This naturally raises the question of the extent to which the survey responses we use 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:

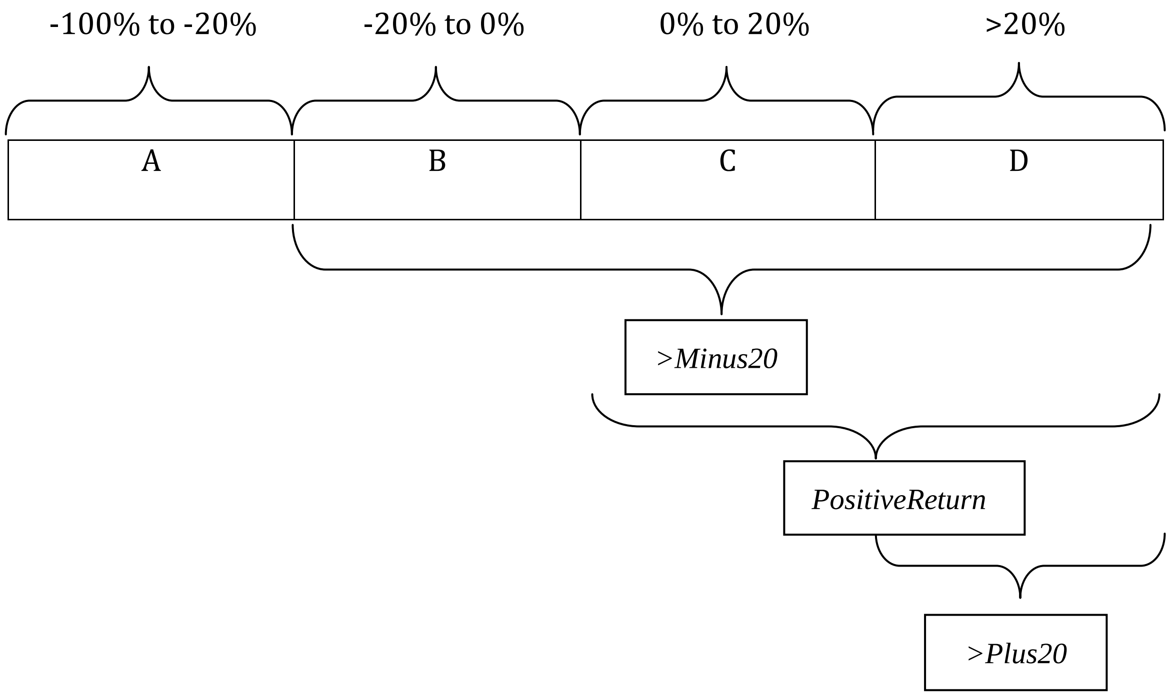


Figure 4: Range of possible returns segmented according to survey questions

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 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 surveys.[12]

In addition to the inconsistent person-wave sets, in 40.0% of the surveys at least one of the four line sections was implicitly assigned a probability of zero (henceforth referred to as "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 of the return being in section C (0- 20% return). The (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 8. Most coefficients do not change much in the new specification. Importantly, though, 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.

[12] 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.

Table 8: 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. * Significant at 5%.

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
<u>Demographic Characteristics</u>									
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 8: Regression results controlling for inconsistent responses (continued)

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 7. 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.

For those that give near-inconsistent responses, the effect is qualitatively similar to the baseline model, with a ten percentage point 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 percentage point 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.[13] 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 7) regression. Notably, nearly all of the race and education effects documented in the Chapter 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

[13] The relatively large magnitude of the "near-inconsistent" effect may be a result of the definition; 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 probability space contains zero mass.

effect of Wall Street uncertainty (the implied volatility) on Main Street uncertainty is significantly less (nearly 6.9 percentage points for a 10 percentage point difference in the implied volatility) than it is for sets of responses that were not inconsistent.

3.7 Sensitivity analysis

To test the robustness of the results to some of the assumptions, sensitivity analyses are performed.[14]

Sensitivity to the expected return assumptions[15]

Changes to these assumptions effectively shift the option-implied probability distribution. The estimated marginal effect of the option-implied probability ranges between 0.101 and 0.117 for three alternative models, compared to 0.114 for the baseline model. Taken together, these analyses demonstrate that the regression results in the Chapter are hardly sensitive to the assumptions about risk neutrality, the equity premium, and the dividend yield. More details can be found in Appendix 3D.

Sensitivity to the focal response specification

We consider robustness with respect to our inclusion of responses of zero and one hundred as "focal" responses by re-estimating the model by defining only responses of 50 as focal responses. In this case, the v^* equation (which separates responses of 50 and 0/100) is not necessary. The resulting specification is similar to that considered in Hurd, McFadden, and Gan (1998). There is very little change to the coefficients and marginal effects in the two remaining equations (μ and w^*). The signs of all coefficients remain the same, and only one coefficient changes significance: Hispanics/Latinos are significantly less likely to give a focal response of 50 in this alternative specification. The estimated marginal effect of the option-implied probability remains unchanged at 0.114.

---

[14] Tables corresponding to the results mentioned in this Section are omitted here in the interest of space.

[15] We consider this alternative because, some (e.g., Cochrane 2011) might argue that respondents "report not their true beliefs, but instead their 'risk neutral' equivalents".

Sensitivity to the number of days with which the option-implied probability is lagged

To allow for the possibility that individuals update their beliefs with some delay, we re-estimate the baseline specification, using a lag on the option-implied probability from 1 to 30 days. As the lag increases, the marginal effect of the option-implied probability declines gradually (albeit non-monotonically), reaching a low of 0.031 at 22 days (p-value 0.126) but is significant for any lag length out to 10 days.

3.8 Conclusion

Are subjective probability responses from surveys at odds with probabilities derived from financial markets data? A novel approach, comparing such responses to probabilistic questions about future stock market performance with their corresponding option-implied probabilities, investigates one aspect of this question: whether financial market probabilities have any influence on the views of survey respondents. 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 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 the option-implied probability that future DJIA returns will exceed a given threshold leads to a 1.14 percentage point increase in the average beliefs of the survey respondents. This effect nearly doubles when controls for probabilistic consistency are included in the regression. When considered in the context of the large literature documenting that a higher degree of financial literacy leads to better financial forecasts and decisions (e.g., Bernheim and Garrett 2003, Lusardi and Mitchel 2011), our results provide further evidence of a link between mathematical skill and financial literacy (Lusardi 2012).

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 Chapter 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 an assignment of equal probability 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 (e.g., 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.

Appendix 3A: 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[16] 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 two repeated modules 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 these reasons, we do not use subsequent waves of the sample in our analysis. The data we use cover the period November, 2008 to March 2011.

Sample construction

The sample construction is detailed in Table 9. 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 the total number of surveys, 784 were not fully completed and hence are omitted from our sample; an additional five surveys were omitted because the sampling weight was missing. In 660 surveys there were no responses to the three dependent variables of interest, while in 94 of the surveys, only one of the three questions was answered; these surveys are also excluded from our sample.

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 while at the same time including an age control in the regressions. 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 they gave inconsistent

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

responses to questions with regards to race, ethnicity, birth year or gender (e.g., respondents indicated being female in some waves and male in other waves).

The questions "How would you rate your understanding of the stock market" and "How closely do you follow the stock market" are asked only in waves 1, 2, 11, 14 and 24. For the waves prior to the first observation of this variable, we assume that the respondent's answer is the same as that first observed response. For subsequent waves, we assume the respondent's answer remains the same until the next observable response.

Table 9: Main Street sample construction

Surveys were dropped sequentially according to a series of filters, in the order that is indicated in the first column. The second column indicates the number of surveys that were omitted as a result of the filter. The third and fourth columns indicate how many surveys and persons, respectively, remained after applying the filter.

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

Appendix 3B: 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:

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

where

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

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.[17] 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.

[17] 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.

Obtaining parameters from Bloomberg

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)[18] 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.

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.[19]

[18] 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.

[19] 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.

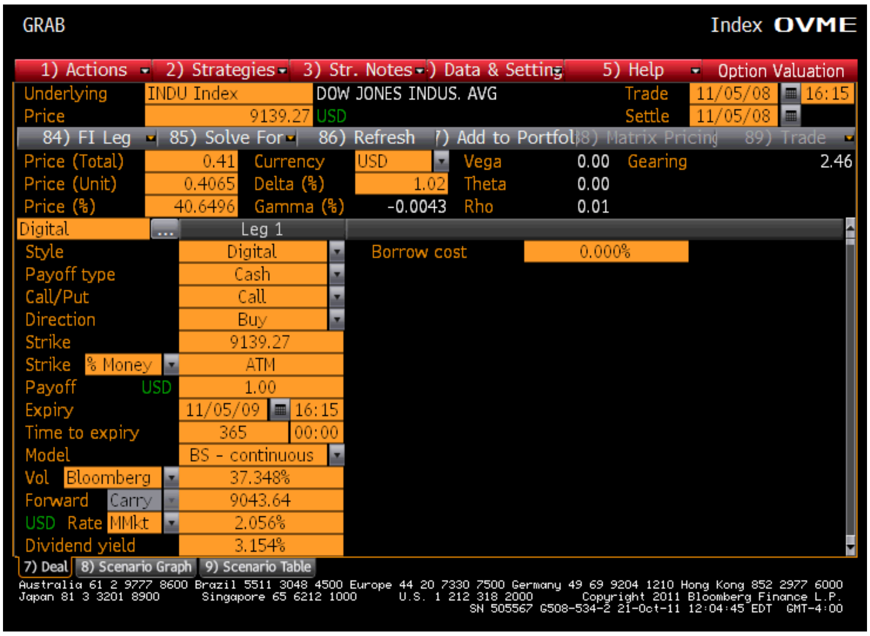


Figure 5: Example of the Bloomberg ® OVME DIVA screen

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 3D.

(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 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 (robustness to this assumption also is considered in Appendix 3D).

The time series of the interest rate and dividend yield are shown in Figure 6a. 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 6b – 6d. 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 6a: *1-year interest rate & dividend yield used in pricing options on the DJIA*

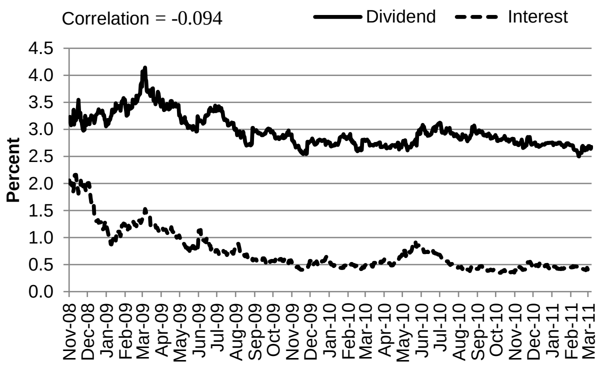


Figure 6b: *Option-implied probability and volatility for a return of >-20% in one year*

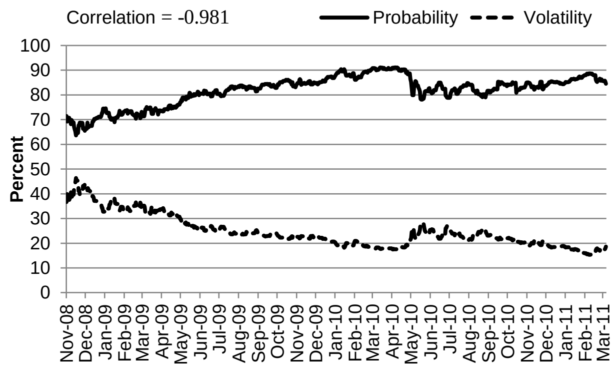


Figure 6c: *Option-implied probability and volatility for a gain in one year*

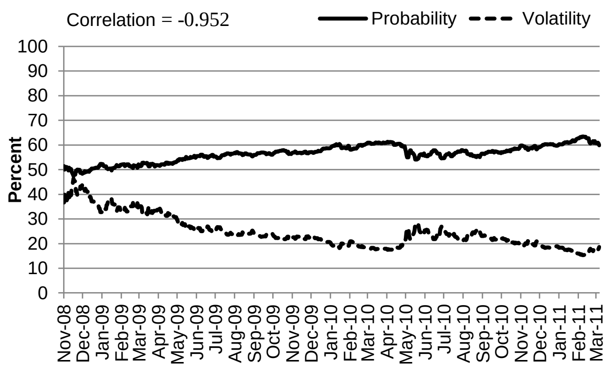
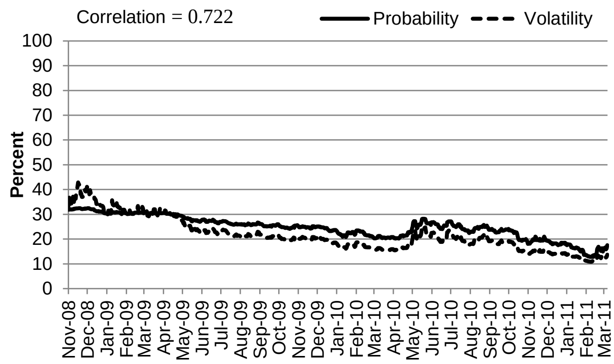


Figure 6d: *Option-implied probability and volatility for a return of >20% in one year*



Source: Bloomberg

Figure 6: Time series of rates, probabilities and volatilities

Summary statistics and correlations for both the levels and the first differences of these time series are provided in Table 10. An augmented Dickey-Fuller test (Dickey and Fuller, 1979) was performed on all the option-implied probability and volatility time series in Figures 6b-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.

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.

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 (79.9%, row 1 column 1). The mean absolute daily change of 0.45 percentage points (row 1 column 6) for *>Minus20* means that each day the market's belief fluctuates by an average of 0.45 percentage points (e.g., increasing it from 79.9 to 80.35). The option-implied probabilities of *>Plus20* and *PositiveReturn* vary less on a daily basis (as seen by lower standard deviations, mean and median absolute deviations). 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 7.54 and 11.42.

Table 10: Descriptive statistics of the time series of option-implied probabilities

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. Probabilities are only calculated over survey days. 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 3C: 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:

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

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

The likelihood of a focal response is then given by:

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

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:

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

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

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

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

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

which is often referred to as the 'regularized incomplete beta function' $I_{\psi}(\alpha_1, \alpha_2)$. 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):

$$l = \begin{cases} \Phi(-x_2\beta_2)\Phi(-x_3\beta_3)I_{\psi}(\alpha_1, \alpha_2) & if\ p = 0 \\ \Phi(-x_2\beta_2)\Phi(x_3\beta_3) & if\ p = 0.5 \\ \Phi(-x_2\beta_2)\Phi(-x_3\beta_3)\left(1 - I_{\psi}(\alpha_1, \alpha_2)\right) & if\ p = 1 \\ \Phi(x_2\beta_2)\dfrac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}p^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1} & if\ p = other\ (non - focal) \end{cases}$$

Appendix 3D: Sensitivity to risk neutrality, equity premium and dividend yield assumptions

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. Yet some might argue that respondents "report not their true beliefs, but instead their 'risk neutral' equivalents" (Cochrane 2011). We consider in this Section sensitivity to these assumptions 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 in 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. 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 for all four models.

Chapter 4 | A New Parameterization of Quasi-hyperbolic Discounting

In the quasi-hyperbolic or (β, δ) discounting model, the parameter $\beta \leq 1$ is commonly interpreted as an index of time-consistent preferences, and thus an (inverse) comparative measure of how much individuals may be vulnerable to self-control problems when future outcomes are at stake. Such an interpretation is problematic because time-consistency is jointly determined by both model parameters, increasing in β but also decreasing in δ. We prove that the ratio, $\tau = log\,\beta / log\,\delta$ is a proper index of time inconsistency in preference, in that the number of future selves that might prefer a different inter-temporal outcome stream than the one preferred by the current self is exactly τ (assuming one self per time period). This suggests a reformulation of $\beta\delta^t$ as $\delta^{\tau+t}$, and an empirical prediction that individuals with greater τ will face more problems of self-control. We illustrate the usefulness of the (δ, τ) parametrization by reanalyzing the dataset in Tanaka et al. (2010).

This Chapter is based on "Beta-Delte or Tau-Delta? A Reformulation of Quasi-hyperbolic Discounting", co-authored by Han Bleichrodt, Drazen Prelec and Peter Wakker.

4.1 Introduction

Because of its simplicity and tractability, the quasi-hyperbolic β-δ model is the most widely used representation of decreasingly impatient ("present biased") time preferences. The model takes as starting point the standard exponential (compound) discounting equation, with discount factor δ. It then assigns an additional discount β to all future time periods, yielding:

$$U(x_0, t_0; \dots; x_n, t_n) = u(x_0) + \beta \sum_{i=1}^{n} u(x_i)\delta^{t_i} \qquad (4.1)$$

where $u(x_i)$ is the utility of outcome received at time t_i, and $t_0 \equiv 0$. The β-parameter reflects (inversely) the additional weight assigned to immediate consequences, creating a wedge between the preferences of the current self and the future selves. Decreasing impatience (DI) can lead to time-inconsistent preferences and costly pre-commitment strategies by 'sophisticated' agents, or actual plan reversals and money-pumping of 'naïve' ones.

In empirical work, the quasi-hyperbolic model often serves as a diagnostic instrument, revealing that one or other group is more deviant relative to the exponential, time-consistent norm. The standard view in the literature is that β provides the appropriate measure of deviance, so that someone with a smaller β is less time consistent and therefore more vulnerable to self-control problems. For example, in an influential review, DellaVigna (2009, p.318) refers to β as "*…capturing the self-control problems.*" Here, we show that this view is incomplete at best, and one should consider instead the ratio, $\log\beta / \log\delta$, or, equivalently, the rewriting of β in Eq. 4.1 as δ^{τ}. $\tau = \log\beta / \log\delta$ is present-bias converted into time units, as a "virtual extra delay" assessed for outcomes that are not immediate. As we show, it is also the maximum temporal "window of vulnerability," during which an option that is disliked by the current self would be chosen if made available to a future self.

The formal argument supporting the τ-measure is based on revealed preferences and is related to a definition of DI proposed earlier (Prelec, 1989). According to this definition, two sets of preferences fall in the same DI class if and only if the associated discount functions are related by a power transformation, which for quasihyperbolics implies that they have the same ratio $\log\beta/\log\delta$ (Prelec, 2004). Essentially, a high discount rate δ dilutes the impact of β, reducing the range of decisions that trigger conflict. To be strongly vulnerable, a person not only has to exhibit bias for immediate outcomes but must also care significantly about future outcomes. It is an ironic, and as yet largely unrecognized implication that a desirable trait — concern about the future — exacerbates problems of self-control.

The question of vulnerability is important in drawing policy inferences from laboratory or econometric parameter estimates of Eq. 4.1. Such estimates could shed light on the causes of self-harming behavior and focus attention on specific remedies. For example, if cigarette smokers care little about the future then their choices might be consistent with the exponential model, and in that narrow sense might be rational. However, if smokers are time inconsistent then other interpretations of their behavior become available, such as sophisticated fatalism ("I believe I cannot stop smoking hence I might as well smoke now") or naïve optimism leading to procrastination ("I believe I will quit tomorrow and therefore I can smoke now"). Because the preferences of different temporal selves are already in conflict, the policy maker may feel justified in acting paternalistically on behalf of one self against another, e.g., by imposing penalties or banning certain goods altogether.

The paper begins with a geometric derivation of our index τ (Figure 7) followed by representation theorems for τ. The potential usefulness of the new index is illustrated by a re-analysis of inter-temporal choice data from Tanaka et al. (2010). We find that τ exhibits less correlation with impatience δ than does β, and also shows a different relationship with demographic variables.

4.2 A Visual Argument

Informally, one could say that person A is more vulnerable to self-control problems than person B, if A disagrees in preference with a greater number of his future selves. In the (β, δ) model, disagreement is promoted by lower values of β, and, perhaps less obviously, by higher values of δ.

To reinforce intuition about the latter claim, concerning δ, consider the following example, where we allow decisions at discrete timepoints only, e.g., the beginning of each year: Suppose A and B have the same $\beta_A = \beta_B = 0.5$, but different annual δ's, $\delta_A = 1$ and $\delta_B = 0.5$. Here A has no time preference apart from an extra weight given to the immediate outcome. Suppose also that there is a prospect of a 'higher reward' yielding 10 (utils) in year 2020, and an alternative, 'immediate lower reward' yielding 6 that will become available sometime between 2015 and 2019. Given these assumptions, A and B, that is, their annual self-incarnations in years 2015 through 2019 will consistently choose the lower reward at the time when it becomes available. If the present moment is 2014, then the lower reward is still a future reward, and is a bad outcome for A, as he currently prefers 10 in 2020 to 6 in all earlier periods, 2015-2019. However, from the current standpoint of B things are less dire, as choosing 6 is precisely the decision he would now, in 2014, want implemented in years 2015-2018. The only feared scenario is if the lower reward becomes available in the penultimate year, 2019, when he would now prefer his future self to wait one extra year for the higher reward, but in 2019 will prefer the immediate smaller reward. Consequently, while A disagrees in preference with *all* of her future selves, B has to worry only about *one*.

Indeed, it is a corollary of Theorem 4.1 below that even if completely general outcome streams are involved, B would never fear the decisions of more than one future annual self, because the aforementioned vulnerability interval $\tau = \log\beta/\log\delta$ equals exactly one year. However, before stating any formal results, we first present a diagram that conveys the gist of the argument.

Figure 7 displays an inter-temporal dilemma in the Ainslie (1975) representation. Let (x, t) stand for "receiving outcome $x \geq 0$ at time t and nothing at all other time points." Time refers to calendar time relative to an arbitrary starting date, which for convenience we set at zero. In the Figure, calendar time is on the x-axis, now in months, and present value of various options on the y-axis. We plot present value in logarithmic coordinates, so that the present value of a reward increases linearly with slope $-\log\delta$ until the moment of reward, when it exhibits a vertical jump of $-\log\beta$. The diagram visually answers the question: "At which time points will my future self choose an immediate reward that I would prefer to avoid in favour of a larger, later reward?" If the date of the later reward is held fixed (in the Figure it is 1 year from now), then the intervening time divides into three distinct intervals:

(1) an initial interval of un-conflicted impatience , where the future self prefers the lower reward but this is OK with the current self,

(2) a *vulnerable* interval of conflicted impatience, where the future self prefers the lower reward, but this is not OK with the current self, and,

(3) a final interval of un-conflicted patience, where both the present and the future self prefers to wait for the later reward.

These three intervals need not always arise (one could have just one, or any consecutive pair of intervals), but if a vulnerable interval is present its duration will be exactly $\tau = \log\beta / \log\delta$ (except for some degenerate cases). This result is evident from the geometry in the Figure, where the horizontal segment of the bold triangle equals the ratio of $\log\beta$ to $\log\delta$.

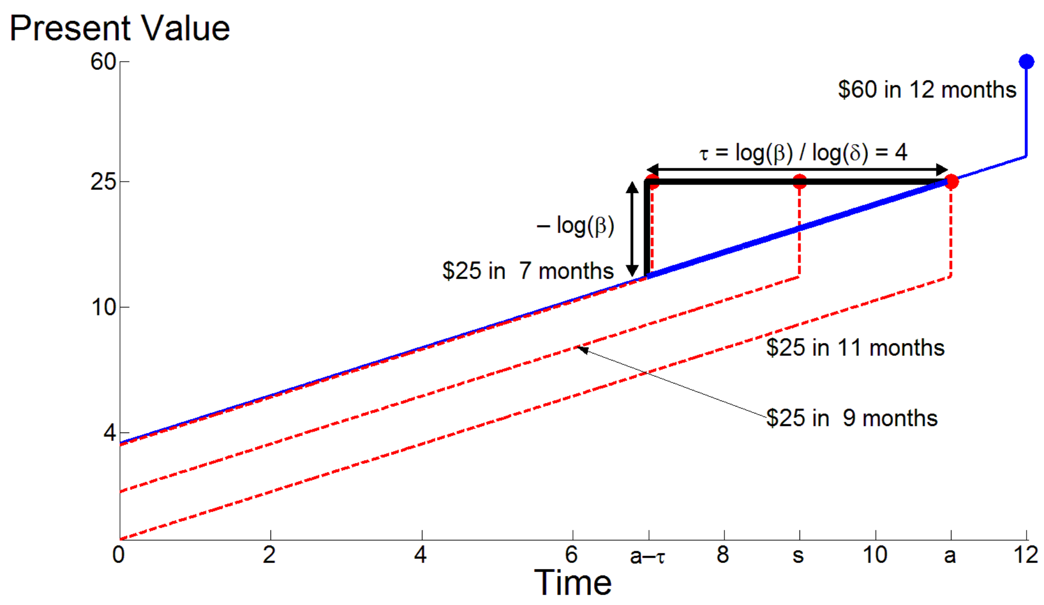


Figure 7: Present value of timed outcomes

The present value of timed outcomes (60,12), (25,7), (25,9), and (25,11) is shown, using Ainslie's (1975) representation and assuming linear utility. Decision times t are shown on the x-axis, and the present value of various options at the considered decision time is shown on the y-axis with logarithmic scaling. From the vantage point of time zero, \$60 at month 12 is preferred to \$25 at month 9; however \$25 is preferred if offered immediately at month 9. Hence the preferences of the month zero self are inconsistent with the preferences of month 9 self, and are also inconsistent with any other self between months 7 and 11. Month 7 marks the transition from unconflicted impatience to vulnerability, and Month 11 the transition from vulnerability to unconflicted patience. The triangle in bold shows that the vulnerable interval equals exactly the height $\log\beta$ divided by $\log\delta$, the slope of the line.

4.3 A Measure of Decreasing Impatience for Timed Outcomes

The next two theorems make our claims mathematically precise. The first focuses on choices between two single outcomes as in Figure 7, and the second focuses on choices between general outcome streams.

Nontrivial choices are always between timed outcomes (l, s) and (h, r) with $l < h$ low and high outcomes, and $s < r$ soon and remote times of consumption. The preference relation is subscripted by decision time, that is: $(l, s) \succ_t (h, r)$ indicates that at calendar time $t \leq s$ the person would choose (l, s)over (h, r).

With quasi-hyperbolic discounting, time-inconsistency can only arise if the lower outcome is immediate $(t = s)$ and the time-zero self fears that the time-*s* self will choose the lower reward:

$$(l, s) \prec_0 (h, r) \text{ and } (l, s) \succ_s (h, r) \text{ with } 0 < s. \tag{4.2}$$

For ease of presentation, in the main text we assume non-degenerate preferences that allow to exactly measure the degree of impatience:

$$(l, 0) \prec_0 (h, \varepsilon) \text{ for some } 0 < \varepsilon \text{ and } (l, b) \succ_0 (h, r) \text{ for some } b > 0. \tag{4.3}$$

That is, h is sufficiently big to compete with an immediate l if h comes soon enough, and r is sufficiently remote that a sufficiently early l without the immediacy effect can still compete with (h, r). This way the three intervals in Section 4.2 are nonempty. The Appendix shows how to handle cases where Eq. 4.3 does not hold.

We quantify the degree of decreasing impatience by inspecting which early times t besides s are vulnerable to inconsistencies. Formally, (assuming Eqs. 4.2 and 4.3), we call time point t *vulnerable* if Eq. 4.2 also holds with t instead of s. The proof of the following theorem will show that under Eq. 4.3 there exists a unique time point a such that $a \geq s, \tau$ and

$$(l, a) \sim_a (h, r). \tag{4.4}$$

Theorem 4.1: Assume Eqs. 4.2 and 4.3. Then the set of vulnerable time points is $(a - \tau, a)$ with a as in Eq. 3.3.

Proof: By continuity and impatience a exists, and is between s and $r-\varepsilon$ with ε as in Eq. 4.3. For $t > a$, $(l,t) \prec_{s'} (h,r)$ irrespective of the decision time s' (including $s' = t$ and $s' = 0$) and no dynamic inconsistency arises. The high reward is too attractive to be affected by the immediacy effect.

Eq. 3.3 implies $u(l) = \delta^{r-a+\tau}u(h)$, $\beta\delta^{a-\tau}u(l) = \beta\delta^{r}u(h)$, and, hence, $(l, a-\tau) \sim_0 (h,r)$, where $a - \tau \geq b \geq 0$ (with b defined in Eq. 4.3). Hence, for $t < a - \tau$ we have $(l,t) \succ_{s'} (h,r)$ for all s' (including $s' = t$ and $s' = 0$). The late reward h is too far away and is less preferred even when the immediacy effect plays no role. For $a - \tau < t < a$, we have the preferences in Eq. 4.2 with t instead of s and these t are vulnerable. □

Theorem 4.1. shows that τ has a natural interpretation as the length of the vulnerable period. The larger τ, the more a decision maker is prone to dynamic inconsistencies.

4.4 Streams of outcomes

Section 4.3 considered preferences between two timed outcomes; this Section extends our result to the general setting of outcome streams. The key observation is that moving from single outcomes to streams of outcomes cannot increase the length of the vulnerability interval in the quasi-hyperbolic model; in this sense, timed outcomes are the worst-case scenario for self-control.

Let $x = (x_1, t_1, \ldots, x_n, t_n)$ denote an income stream that gives money amount $x_j > 0$ at timepoint t_j, $j = 1, \ldots, n$, and nothing otherwise.[20] Implicit is $t_1 < \ldots < t_n$. For $\varepsilon \in \mathbb{R}$, $x^{\uparrow\varepsilon}$ denotes the shift $(x_1, t_1 + \varepsilon, \ldots, x_n, t_n + \varepsilon)$ of x, where $t_1 + \varepsilon \geq 0$. We again analyze a general preference reversal:

$(x_1, s_1, \ldots, x_n, s_n) \succ_t (y_1, r_1, \ldots, y_m, r_m)$ and $(x_1, s_1, \ldots, x_n, s_n) \prec_{t'} (y_1, r_1, \ldots, y_m, r_m)$.

Under quasi-hyperbolic discounting, the extra weight on the immediate outcome must be the cause of the preference reversal and we assume, without loss of

[20] Deviating from Eq. 4.2, we do not write the outcome at time 0 unless it is positive.

generality, that this favors x. Hence $t = s_1$, and either $s_1 < r_1$ and $x_1 > 0$, or $s_1 = r_1$ and $x_1 > y_1$. We may again assume $t' = 0$, which gives

$$\begin{aligned}&(x_1, s_1, \dots, x_n, s_n) \succ_t (y_1, r_1, \dots, y_m, r_m) \text{ and}\\ &(x_1, s_1, \dots, x_n, s_n) \prec_0 (y_1, r_1, \dots, y_m, r_m)\end{aligned} \tag{4.5}$$

We investigate the degree of time inconsistency by considering which shifts $x^{\uparrow\varepsilon}$ preserve the preference reversal:

$$\begin{aligned}&(x_1, s_1, \dots, x_n, s_n)^{\uparrow\varepsilon} \succ_{s_1+\varepsilon} (y_1, r_1, \dots, y_m, r_m) \text{ and}\\ &(x_1, s_1, \dots, x_n, s_n)^{\uparrow\varepsilon} \prec_0 (y_1, r_1, \dots, y_m, r_m).\end{aligned} \tag{4.6}$$

In keeping with Section 4.3, we call such ε *vulnerable*.

Theorem 4.2: Under Eq. 4.5, τ is the maximum length of sets of vulnerable ε.

Proof: We write $QH(y)$ for the quasi-hyperbolic discounted utility of y at time s_1.

$(x_1, s_1, \dots, x_n, s_n)^{\uparrow\varepsilon} \succ_{s_1+\varepsilon} (y_1, r_1, \dots, y_m, r_m)$ implies

$U(x_1) + \beta \sum_{j=2}^{n} \delta^{s_j - s_1} U(x_j) > \delta^{-\varepsilon}\, QH(y)$. Because $\beta < 1$, it is also true that,

$U(x_1) + \sum_{j=2}^{n} \delta^{s_j - s_1} U(x_j) > \delta^{-\varepsilon}\, QH(y)$. Because $\beta\delta^{-\tau} = 1$,

$\beta\delta^{-\tau} U(x_1) + \beta \sum_{j=2}^{n} \delta^{s_j - s_1 - \tau} U(x_j) > \delta^{-\varepsilon}\, QH(y)$, or (multiplying with $\delta^{s_1+\varepsilon}$)

$\beta\delta^{s_1+\varepsilon-\tau} U(x_1) + \beta \sum_{j=2}^{n} \delta^{s_j+\varepsilon-\tau} U(x_j) > \delta^{s_1}\, QH(y)$.

The last inequality concerns preference $\succ_0$ and the shift $x^{\uparrow\varepsilon-\tau}$. It gives the opposite preference to the second one in Eq. 4.6 for the shift ε. Hence Eq. 4.6 cannot hold for both a shift by ε and a shift by $\varepsilon + \tau$. For shifts that exceed $\varepsilon + \tau$, the above inequalities become stronger and favor x more. This implies that the set of vulnerable shifts cannot contain shifts that are further than τ apart. Numerical examples show that the length of the set of vulnerable time periods can indeed be less than τ for some x, y. We saw in Section 4.3 that for one nonzero outcome sets of vulnerable shifts have length τ and, hence, the maximum length is τ. □

Theorem 4.2 shows that for preferences over n outcomes, vulnerable periods can be shorter than τ, but not longer. As we already saw in Theorem 4.1, the maximum length τ is reached, for instance, in choices between timed outcomes. That is, the maximum length is exactly τ.

4.5 Empirical Illustration

To illustrate how τ affects our interpretation of individual differences, we reanalyze the data from Tanaka, Camerer, and Nguyen (2010). They combined data from an experimental study and a household survey in Vietnam to estimate individual (risk and) time preferences and related those to demographic variables. 181 subjects answered 15 time preference questions by choosing between a money amount now and a larger amount in the future (3 days to 3 months), with real incentives (average payment was about 6-9 days' wage).

Table 11 gives the correlation matrix of parameter estimates when computed for each individual on the basis of the 15 questions (we display Spearman rank correlations to reduce the impact of outliers). The estimates of β and δ are strongly correlated, suggesting that the two parameters are tapping a common individual difference variable — impatience. In contrast, estimates of δ and τ are less related.

Table 11: Spearman correlation matrix

The correlation between δ and τ is significantly smaller (in absolute value) than the correlation between β and δ (z =2.98, p = 0.003).

	β	δ	τ
β	1.00		
δ	-0.45	1.00	
τ	-0.47	-0.17	1.00

We then repeated the authors' group-level analysis with non-linear least squares regression (see Tanaka et al. for details), on (β, δ) and on (δ, τ). To ensure both impatience and decreasing impatience, as characteristic of the (β, δ) model, we truncated $0 \leq \beta \leq 1$, $0 < \delta < 1$, and $\tau \geq 0$.

Table 12 shows the estimation results, with conventional labeling of variables. The dummy variable "Trusted agent" is equal to 1 for subjects who stored the money earned in the experiments, and the variable "Risk payment" is equal to the amount of money the subject received in the elicitation of the risk preferences. All variables (including dummy) are standardized.

The average β is .65, smaller than 1 ($p < 0.001$); β is unrelated to the demographic variables. The average *daily* discount factor δ is equal to .992, and is higher (more patience) for subjects with higher age, education, income, and money won in the risk part.

The third and fourth columns show the estimates for the (τ, δ) model. The p-values are usually lower for τ than for β, meaning that τ delivers more statistical power than β. Whereas β is unrelated to any of the demographic variables, τ is associated with two variables: people in the South of Vietnam and people who received more money in the first part of the experiment have a higher τ (more decreasing impatience).

A comparison between the second and fourth columns shows that the standard errors for δ are higher in the (β, δ) framework and that we have more statistical power in the (τ, δ) framework. Again, p-values are generally lower in the (τ, δ) framework. This may be the result of collinearity between β and δ: both measure impatience (β the short run and δ the long run impatience), and will thus be collinear when estimated jointly.

Table 12: Regression results for the $(\boldsymbol{\beta}, \boldsymbol{\delta})$ and the $(\boldsymbol{\tau}, \boldsymbol{\delta})$ model

Data from Tanaka et al. (2010) are used. For each variable the Table shows (from top to bottom) the coefficient, the standard error, and the p-value. Asterisks indicate significance at the 5% level.

	Tanaka et al. (β, δ)		New framework (τ, δ)	
	β (%)	δ (%)	τ	δ (%)
Constant $(\beta_0, \delta_0, \tau_0)$	64.85	99.20	85.10	99.29
	1.88	0.07	15.87	0.07
	0.00	0.00	0.00	0.00
Chinese	-0.71	0.04	14.51	0.03
	1.67	0.07	9.38	0.02
	0.67	0.58	0.12	0.19
Trusted agent	-0.71	0.03	-1.58	-0.03
	1.32	0.04	2.21	0.02
	0.59	0.47	0.48	0.12
Age	1.18	0.18*	7.86	0.12*
	1.94	0.07	4.82	0.04
	0.54	0.02	0.10	0.00
Female	0.64	0.05	-3.30	0.00
	1.87	0.07	4.22	0.03
	0.73	0.44	0.44	0.93
Education	-3.33	0.15*	4.49	0.01
	2.03	0.07	5.36	0.03
	0.10	0.03	0.40	0.75
Income	1.08	0.10*	1.95	0.04*
	1.18	0.03	6.19	0.02
	0.36	0.00	0.75	0.02
Distance to market	2.49	0.02	-0.75	0.04
	2.13	0.07	9.19	0.04
	0.25	0.82	0.94	0.20
South Vietnam	-2.67	0.08	30.84*	0.18*
	2.28	0.08	15.18	0.09
	0.25	0.30	0.04	0.05
Risk payment	-1.75	0.15*	23.23*	0.14*
	2.14	0.08	10.94	0.06
	0.42	0.05	0.04	0.02
# Observations	5,340		5,340	

4.6 Conclusion

Decreasing impatience is the key characteristic of hyperbolic preferences. We have shown that decreasing impatience for the popular quasi-hyperbolic discounting is better captured by writing the model as $f(t) = \delta^{t+\tau}$ rather than by the common $f(t) = \beta\delta^t$ (for $t > 0$, with $f(0) = 1$). This new writing permits a proper separation between impatience (measured by δ) and decreasing impatience (measured by τ).

The index τ has natural interpretations because it is expressed in time units as the perceived time penalty of any delay beyond the present. Additionally, τ is the period of vulnerability to dynamic inconsistencies and, hence, to self-control problems and book making.

An empirical illustration showed that compared to the (β, δ) model, there is less correlation among the parameters and a stronger association of the parameters with demographic variables in the (τ, δ) model. Of course, whether impatience and self-control are distinct psychological dimensions is an empirical rather than a modeling question. It is certainly possible that individuals who care little about the future as measured by their δ will also exhibit more time inconsistent preferences, as measured by τ. In that case, the coefficient β, which merges impatience and time inconsistency into a single index of 'intertemporal misbehavior' may be useful in empirical work, just as combining, say, verbal and mathematical ability into a single summary cognitive aptitude index may be useful in certain applications. The theoretical point we underline here is that such aggregation of two conceptually distinct dimensions into a single number should be done with eyes open. In contrast, the traditional approach of estimating β and δ and then interpreting β as self control may suggest a relationship between impatience and self-control when no such relationship exists.

Appendix 4: Proofs

A measure of decreasing impatience

Lemma 4.1: In the quasi-hyperbolic model, $\succcurlyeq^*$ exhibits more decreasing impatience than $\succcurlyeq$, and $ln(d^*)$ is more convex than $\ln(d)$, if and only $\tau^* = \frac{\ln(\beta^*)}{\ln(\delta^*)} \geq \tau = \frac{\ln(\beta)}{\ln(\delta)}$.

Proof: All functions considered take value 0 at $t = 0$, and we describe only their values at $t \neq 0$. Substitution shows that

$$\ln\big(d^*(t)\big) = \ln(\beta^*) - \frac{\ln(\delta^*)}{\ln(\delta)}\ln(\beta) + \frac{\ln(\delta^*)}{\ln(\delta)}\ln\big(d(t)\big).$$

Given value 0 at 0, this transformation of $\ln\big(d(t)\big)$ is convex if and only if $\ln(\beta^*) - \frac{\ln(\delta^*)}{\ln(\delta)}\ln(\beta) \leq 0$, which holds if and only if $\frac{\ln(\beta^*)}{\ln(\delta^*)} \geq \frac{\ln(\beta)}{\ln(\delta)}$. □

Prelec (2004) showed that two discount functions with the same τ are related through a power transformation, which suggests τ may serve as an index of decreasing impatience. No derivations were given, however; in particular, the discontinuity at $t = 0$ was not handled.

Theorem 4.1 without Eq. (4.3)

If the second condition of Eq. 4.3 is dropped, then $a < \tau$ is possible and the time interval $(a - \tau, a)$ may contain negative time points. We then should either intersect the interval with $[0, \infty)$ or allow for negative time points, which for calendar time is possible. If the first condition of Eq. 4.3 is dropped, then the discontinuity of discounting at $t = 0$ complicates the analysis. Then there may not exist an a that gives indifference in Eq. 4.4. and the set of vulnerable time points is of the form $(\bar{t} - \tau', \bar{t})$ where $\bar{t}$ is the supremum of time points t with $(l, t) \succ_t (h, r)$, and $\tau' \leq \tau$. If $a > r$ solves $\delta^a U(l) = \beta\delta^r U(h)$, then $\bar{t} - \tau' = a - \tau$. □

Chapter 5 | A Criticism of Doyle's Survey of Time Preferences

Doyle's (2013) theoretical survey of discount functions criticizes two parametric families abbreviates as CRDI and CADI families. We show that Doyle's criticisms are based on a mathematical mistake and are incorrect.

This Chapter is based on "A Criticism of Doyle's Survey of Time Preferences: a Correction Regarding the CRDI and CADI Families", co-authored by Han Bleichrodt, Kirsten Rohde and Peter Wakker, and published in the Judgment and Decision Making (Bleichrodt et al., 2012).

5.1 Background and Analysis

Doyle (2013) provides a useful theoretical survey of the most popular discount functions for intertemporal choice. Unfortunately, his analyses of the CRDI (constant relative decreasing impatience) and CADI (constant absolute decreasing impatience) families of Bleichrodt, Rohde and Wakker (2009; BRW henceforth) are incorrect. Let D denote the discount function and let T denote time. The CRDI family is defined as follows, with $\beta > 0$, $\rho > 0$, and $\psi \in \mathbb{R}$ denoting parameters:

If $\Psi > 0$, then $D(T) = \beta \exp(-\rho T^{\Psi})$ for $T \geq 0$; (5.1)

If $\Psi = 0$, then $D(T) = \beta T^{-\rho}$ for $T > 0$; D is not defined for $T = 0$ (5.2)

If $\Psi < 0$, then $D(T) = \beta \exp(\rho T^{\Psi})$ for $T > 0$; D is not defined for $T = 0$ (5.3)

The CADI family is defined by replacing T by $\exp(T)$ in the right-hand sides of the equalities in Eqs. 1-3 (the Chapter number 5 is left out of equations for ease of reading), but we focus on CRDI here. Doyle wrongly assumes the normalization

$$D(0) = 1 \tag{5.4}$$

for all families he considers. This normalization has often been assumed in the literature, but it should not be used in Eqs. 1-3, as pointed out by BRW (p.29 l.2 ff). For Eq. 1, $D(0) = \beta$ immediately follows from substitution, invalidating Eq. 4. For Eqs. 2 and 3 we have $\lim_{T\downarrow 0} D(T) = \infty$ (BRW 2009 p. 32), again invalidating the extension in Eq 4. The normalization in Eq. 4 can be obtained in Eq. 1 by setting $\beta = 1$. But then, obviously, $\beta = 1$ should be consistently set for all T, including all $T > 0$. Such a consistent normalization does not affect preference. The normalization of Eq. 4 cannot be obtained from Eqs. 2 and 3.

Doyle assumes Eq. 4 for Eq. 1, but inconsistently does not set $\beta = 1$ for $T > 0$. He also erroneously assumes Eq. 4 for Eqs. 2 and 3. He does not state his assumed Eq4 explicitly, but all his results and conclusions about the CRDI and CADI families essentially use it (see Appendices) and, hence, are invalid. For example, contrary to Doyle (p. 127 following Eq. 31), the β parameter in CRDI is not the β parameter in the

β-δ model, but is just a normalization parameter with no empirical meaning.[21] Further, contrary to Doyle's implicit assumption throughout, present values do not exist for Eqs. 2 and 3.

The CRDI family D is simply the transform $D = \exp(-U)$ of the CRRA utility family U for expected utility, which is one natural family (Wakker 2008) rather than "three distinct sub-models" (Doyle's end of Section 3.6.2.). Doyle's mistake in setting $D(0) = 1$in Eqs. 2 and 3, for instance, is the same as wrongly assuming $U(0) = 0$ for the negative-power $U(x) = -x^{\Psi}$ for $\Psi < 0$ or for the logarithmic $U(x) = \ln(x)$. It should be $U(0) = -\infty$, because otherwise monotonicity (and continuity) are violated.

BRW explained that their CRDI family has the flexibility to capture all possible degrees of increasing and decreasing impatience, unlike any other currently popular discount family. Hence, their CRDI family can accommodate the full range of individual time preferences, including subjects whose deviations from constant discounting are extreme. It can therefore serve well to fit data at the individual level. We hope that the family will be used despite the incorrect criticisms by Doyle.

[21] See BRW (p. 31 last line and Definition 4.3), who use k to denote Doyle's β. Doyle (p. 127 bottom), strangely enough, cites this text of BRW, but still maintains his incorrect interpretation of β.

Appendix 5A: Reproducing Doyle's Results for the CRDI family

Doyle does not explicitly state his use of the false Eq. 4, but it can be inferred from his incorrect Eqs. 31-36. It also explains his incorrect interpretation of the β parameter. Because multiplying all discounted utilities by the same factor $\beta > 0$ does not affect preference, the readers can immediately see that β is just a normalization factor with no empirical meaning in Eqs. 1-3.

We first provide an algebraic derivation of our claim about Doyle's implicit assumption for the most complex case, Doyle's Eq. 33. We explained in the main text that neither Doyle's Eq. 33 nor any close analog can be derived, because present values do not exist. We next show that with the incorrect Eq. 4 and thus, in particular, with the incorrect assumption that present values exist, Doyle's Eq. 33 readily follows. This reveals that Doyle's analysis indeed incorrectly assumed Eq. 4. For the other cases, the analysis is similar but simpler. Doyle's Eq. 33 concerns our Eq. 3 ($\Psi < 0$), and is reproduced next:

$$\rho = \frac{\ln(\beta F/P)}{-T^{\Psi}} \tag{5A.1}$$

In all the analyses relevant here, Doyle assumes that utility is the identity function ($U(x) = x$). He assumes, with $(T{:}\,F)$ denoting receiving T at time F, and P denoting present value (wrongly assumed to exist, as an implication of Eq. 4):

$$(0{:}\,P) \sim (T{:}\,F) \tag{5A.2}$$

If we apply Eq. 4, $D(0) = 1$ to the left-hand side but Eq. 3 to the right-hand side, then we get:

$$P = F\beta\exp(\rho T^{\Psi}) \tag{5A.3}$$

Lemma 5.1

For Eq. 3 with Eq. 4 ($D(0) = 1$), Eq. 6 can be satisfied, and it implies Eq. A.1 (Doyle's Eq. 33).

Proof: Consider the following rewritings of Eq. A.3:

$$\frac{\beta F}{P} = \frac{1}{\exp(\rho T^{\Psi})} = \exp(-\rho T^{\Psi})\,;$$

$$\ln\left(\frac{\beta F}{P}\right) = -\rho T^{\Psi};$$

$\frac{\ln(\beta F/P)}{-T^{\Psi}} = \rho.$ □

We next consider Doyle's Eqs. 31 and 32.

Lemma 5.2

For Eq. 1 with Eq. 4, Eq. A.2 implies Doyle's Eq. 31: $\rho = \frac{\ln(\beta F/P)}{T^{\Psi}}$

Proof: Eqs. 1, 4 & A.2 jointly imply $P = F\beta \exp(-\rho T^{\Psi})$

Rewritings give $\frac{\beta F}{P} = \frac{1}{\exp(-\rho T^{\Psi})} = \exp(\rho T^{\Psi})$; $\ln(\beta F/P) = \rho T^{\Psi}$; $\frac{\ln(\beta F/P)}{T^{\Psi}} = \rho.$ □

Lemma 5.3

For Eq. 2 with Eq. 4, Eq. A.2 implies Doyle's Eq. 32: $\rho = \frac{\ln(\beta F/P)}{\ln(T)}$

Proof: Eqs. 2, 4 & A.2 jointly imply $P = F\beta T^{-\rho}$

Rewritings give $\frac{\beta F}{P} = \frac{1}{T^{-\rho}} = T^{\rho}$; $\ln(\beta F/P) = \rho\ln(T)$; $\frac{\ln(\beta F/P)}{\ln(T)} = \rho.$ □

Appendix 5B: Reproducing Doyle's Results for the CADI family

The CADI family is defined, with parameters $\beta > 0, \sigma > 0$, and η, by[22]

If $\eta > 0$, then $D(T) = \beta \exp(-\sigma e^{\eta T})$ (5B.1)

If $\eta = 0$, then $D(T) = \beta \exp(-\sigma T)$ (5B.2)

If $\eta < 0$, then $D(T) = \beta \exp(\sigma e^{\eta T})$ (5B.3)

Unlike with CRDI, for all its parameters, CADI is defined for all for $T \in \mathbb{R}$.

We show how Doyle's analysis in his §3.6.3 essentially uses the incorrect Eq. 4. We again assume the present value P of Eq. A.2. Unlike with the CRDI family, for the CADI family a present value P always exists, and Eq. A.2 can be satisfied for each of the Eqs. B.1-3, given that $U(x) = x$ is unbounded. We consider the three cases of η separately.

Lemma 5.4

For Eq. B.1 with Eq. 4, Eq. A.2 implies Doyle's Eq. 34: $\sigma = \frac{\ln(\beta F/P)}{e^{\eta T}}$

Proof: Eqs. 4, A.2 & B.1 jointly imply $P = F\beta \exp(-\sigma e^{\eta T})$

Rewritings give $\frac{\beta F}{P} = \frac{1}{\exp(-\sigma e^{\eta T})} = \exp(\sigma e^{\eta T})$; $\ln(\beta F/P) = \sigma e^{\eta T}$; $\frac{\ln(\beta F/P)}{e^{\eta T}} = \sigma$. □

Lemma 5.5

For Eq. B.2 with Eq. 4, Eq. A.2 implies Doyle's Eq. 35: $\sigma = \frac{\ln(\beta F/P)}{T}$

Proof: Eqs. 4, A.2 & B.2 jointly imply $P = F\beta \exp(-\sigma T)$

Rewritings give $\frac{\beta F}{P} = \frac{1}{\exp(-\sigma T)} = \exp(\sigma T)$; $\ln(\beta F/P) = \sigma T$; $\frac{\ln(\beta F/P)}{T} = \sigma$. □

Lemma 5.6

For Eq. B.3 with Eq. 4, Eq. A.2 implies Doyle's Eq. 36: $\sigma = \frac{\ln(\beta F/P)}{-e^{\eta T}}$

Proof: Eqs. 4, A.2 & B.3 jointly imply $P = F\beta \exp(\sigma e^{\eta T})$

Rewritings give $\frac{\beta F}{P} = \frac{1}{\exp(\sigma e^{\eta T})} = \exp(-\sigma e^{\eta T})$; $\ln(\beta F/P) = -\sigma e^{\eta T}$; $\frac{\ln(\beta F/P)}{-e^{\eta T}} = \sigma$. □

[22] BRW, p. 31, use the following notation: $\varphi = D, t = T, k = \beta$, and $c = -\eta$

Chapter 6 | Beyond Chance? The Persistence of Performance in Online Poker

A major issue in the widespread controversy about the legality of poker and the appropriate taxation of winnings is whether poker should be considered a game of skill or a game of chance. To inform this debate we present an analysis into the role of skill in the performance of online poker players, using a large database with hundreds of millions of player-hand observations from real money ring games at three different stakes levels. We find that players whose earlier profitability was in the top (bottom) deciles perform better (worse) and are substantially more likely to end up in the top (bottom) performance deciles of the following time period. Regression analyses of performance on historical performance and other skill-related proxies provide further evidence for persistence and predictability. Simulations point out that skill dominates chance when performance is measured over 1,500 or more hands of play.

This chapter is based on the paper "Beyond Chance? The Persistence of Performance in Online Poker", co-authored by Martijn van den Assem and Dennie van Dolder.

6.1 Introduction

Poker is the most popular card game in the world. Every day, hundreds of thousands of people play poker for real money on the Internet (Online Poker Traffic Reports). In 2013, online poker rooms generated approximately €2.8 billion in gross win (H2 Gambling Capital). The popularity of the game is also evidenced by the many TV reports of major poker tournaments and the number of participants in these tournaments. In 2014, for example, 6,683 people paid $10,000 to participate in the most renowned poker tournament, the Main Event of the World Series of Poker in Las Vegas.

At the same time, there is a widespread controversy about the legality of poker and the appropriate taxation of winnings. A key issue in the debate is whether poker is to be considered a game of chance or a game of skill. Unlike with games of skill, organizing or playing a game of chance is prohibited or restricted in many countries. Also, many countries have a separate gam(bl)ing tax for games of chance, while money won in a game of skill is generally subject to regular income tax. Kelly, Dhar and Verbiest (2007) map legislation and case law on poker for various countries, and show that there is great variation. US regulation even differs across states. Over recent years, several law papers have argued that poker is a skill game and should be recognized as such (Cabot and Hannum, 2005; Grohman, 2006; Tselnik, 2007).

Authorities often have a less permissive stance towards online poker than towards live poker. In the US, for example, the Unlawful Internet Gambling Enforcement Act (UIGEA) that was adopted in 2006 had a major impact: although the Act did not forbid online gambling, it did prohibit the transfer of funds to and from online gambling businesses. As depositing money is necessary for playing online poker, this Act effectively declared online poker illegal. If poker would be considered as a game of skill the UIGEA would probably no longer affect the online poker business.

Behavioral research hints that there are important skill elements in poker. For optimal play in the long run, poker players need to apply the expected value criterion to every decision in the game. There is considerable evidence, however, that they will systematically deviate from this ideal and that they will do so in different degrees.

One issue is that people tend to evaluate their decisions one at a time and thereby weigh losses more heavily than gains, a phenomenon labeled myopic loss aversion (Benartzi and Thaler, 1995). In addition, people are typically risk averse for gains and risk seeking for losses (the reflection effect; Kahneman and Tversky, 1979). Loss aversion and reflection lead to path-dependent risk preferences if people do not perfectly update their reference point after prior gains or losses, a phenomenon indeed found in various studies (Thaler and Johnson, 1990, Kameda and Davis, 1990; Post *et al*., 2008). People's nonlinear sensitivity to probabilities is another reason why behavior diverges from the expected value criterion (Allais, 1953; Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Tversky and Fox, 1995).

Perhaps even more important, probabilities in poker are rarely precise as they depend on the strategic choices of other players. As a result, players are likely to go wrong systematically. Typically, people are overly optimistic about their own chances when they have a sense of control (Langer, 1975; Langer and Roth, 1975; Weinstein, 1980; Svenson, 1981), overly confident about the precision of their probability judgments (Alpert and Raiffa, 1982; Fischhoff *et al*., 1977), and averse to ambiguity in a degree that depends on how competent they feel in evaluating the gamble (Ellsberg, 1961; Heath and Tversky, 1991). Also, instead of using Bayesian statistics people often judge probabilities by the subjective ease with which relevant instances come to mind and by relying on representativeness heuristics (Kahneman and Tversky, 1972, 1973; Tversky and Kahneman, 1971, 1973; Rabin, 2002).

A number of behavioral studies have directly examined the behavior of poker players on some of these issues. Park and Santos-Pinto (2010) survey poker players and report that they overestimate their own expected performance. Smith *et al*. (2009) find that online players display path-dependent risk attitudes by playing more adventurously after big losses. Similarly, Eil and Lien (2014) find that players who have lost tend to play longer and less conservatively.

Differences between players in the degree to which they are prone to these deficiencies lead to differences in their expected performance. Whereas a great deal of the heterogeneity in proneness is attributable to unobservable characteristics, behavioral research has also uncovered factors that explain systematic differences between people. For example, Cesarini *et al*. (2012) and Cronqvist and Siegel (2014) find survey-based evidence from twins that genetic differences are a source of heterogeneity in anomalous choice behavior, and Stanovich and West (1998) and Grinblatt, Keloharju and Linnainmaa (2012) report a relation with cognitive ability.

Simultaneously, players are likely to have different levels of strategic sophistication. At the first level, a player considers her own cards only. A second-level player understands that opponents base their decisions on their own cards likewise, and she updates the relative strength of her cards according to other players' actions. At the third level, a player also bears in mind that opponents are influenced by her actions, entailing awareness of opportunities to bluff. In theory, there are infinitely many such levels.[23] Heterogeneity in strategic sophistication is illustrated, for example, in a field experiment by Palacios-Huerta and Volij (2009): in a game called the centipede game, experienced chess players chose the (rational) action dictated by backward induction more often than college students did. Other examples are Stahl and Wilson (1994), Haruvy, Stahl and Wilson (2001), and Camerer, Ho and Chong (2004). Altogether, these behavioral insights suggest that a player's performance in the game of poker is determined by more than chance alone.

Two different research tracks have examined the skill component in poker. One track focuses on developing and calculating measures of skill in games, and can be traced back to Kadane (1986). Borm and van der Genugten (2001), Dreef, Borm and van der Genugten (2003, 2004a, 2004b), and Hendrickx *et al.* (2008) propose measures that compare the performances of different types of players, including an informed hypothetical player who knows exactly the cards that will be drawn. The use of their approach is, however, limited to relatively simple games. Because of the virtually infinite number of possible game situations that result from the many different choice (betting) options that players have and because of the importance of players' hidden higher-order beliefs, the approach cannot be accurately implemented for the most popular form of poker, No Limit Texas Hold'em. Nevertheless, even for simple poker variants, the different studies report a substantial degree of skill. Heubeck (2008) reviews the various kinds of proposed skill measures.

[23] The analysis of optimal strategies in poker has captivated many game theorists. See, for example, von Neumann and Morgenstern (1944), Kuhn (1950), Nash and Shapley (1950), and Friedman (1971).

The second track of studies takes a more empirically oriented approach. Likewise, these papers suggest that poker involves a skill component. Larkey *et al*. (1997) and Cabot and Hannum (2005) ran large-scale simulations with different pre-defined playing strategies and find that their more sophisticated strategies perform better. DeDonno and Detterman (2008) carried out experiments with student-subjects and demonstrate that the group of players who received strategic instructions during the session outperformed the control group. Siler (2010) analyzes online poker data and establishes that performance is related to playing style, and that style and performance differences between players decrease with the level of the stakes.

In the same spirit as some of the analyses in the present chapter, Croson, Fishman and Pope (2008) and Levitt and Miles (2014) examine whether there is persistence in the performance of poker players. Croson, Fishman and Pope analyze how well players who have finished in the top 18 of a high-stakes tournament fare when they are among the final 18 players in a subsequent major tournament, and they compare their results with those from a similar analysis for professional golf. They find that previous finishes predict current finishes, and that the skill differences across the poker players in their sample are similar to those across the golfers. Levitt and Miles analyze a data set that comprises the complete rankings of all players who entered a 2010 World Series of Poker tournament. They report that players who were *a priori* classified as being especially skilled indeed outperformed the other players.

In what follows we analyze the role of skill in the performance of online poker players, using a large database with 456 million player-hand observations from real money ring games at three different stakes levels. Online poker seems to be the most obvious data source, because the chance-skill debate is especially oriented towards issues regarding the legality of internet poker and the taxation of winnings from online play. Moreover, the vast amount of data that is available allows for powerful analyses.

We define skill as anything that affects a player's performance other than chance. In a pure game of chance, each player's expected winnings are zero (in absence of costs) and there is no persistence or positive autocorrelation in their performance: players' performance over a given period is independent of that over any other period. If performance is predictable, the game involves elements of skill.

Our results indicate that skill is an important factor. When we split our sample into subperiods, we find that players whose performance was in the top (bottom) deciles of the previous period perform better (worse) and are more likely to end up in the top (bottom) deciles of the current period. Regression analyses of performance on past performance and other skill proxies reinforce this evidence of persistence in performance.

From a legal viewpoint, the key question is whether skill dominates chance, that is, whether poker is more a game of skill than a game of chance. The answer to this question heavily depends on the duration and intensity of play, as the effect of chance diminishes with the number of hands and eventually cancels out in the long run.[24] Our simulations point out that skill predominates after approximately 1,500 hands.

The chapter proceeds as follows: Section 6.2 discusses our data and presents descriptive statistics, Section 6.3 analyzes the persistence of performance using decile analyses, Section 6.4 reports on our regression analyses, Section 6.5 presents our simulations, and Section 6.6 concludes.

6.2 Data and Descriptive Statistics

For our analysis we use data on real money ring games ("cash games") played at one of the major online poker sites. We consider No Limit (NL) Texas Hold'em only because this variant is by far the most popular form of poker worldwide. Our data is acquired through an online service called HHDealer. In recent years, several companies have specialized in gathering and trading so-called "hand" histories from online poker rooms. With software applications they continuously collect information on hands played at online poker tables. Many players buy these data to have information on the playing styles of others. Because of limited resources, hand history providers are unable to store data on every hand that is played online. Out of the websites that responded to our inquiries, HHDealer was able to provide the largest

[24] A "hand" is the game that is played between two subsequent shuffles of the deck: dealing of cards, betting, and awarding of the pot (in another context, the term can also refer to the cards dealt to a player). With "playing a hand" we mean to be dealt in. Whenever possible, we avoid the use of poker terminology. We believe that reading this Chapter does not require the reader's understanding of the game.

number of hands for an uninterrupted period of twelve months. We purchased all available data for the games that had been played at three particular stakes levels in the period October 2009 – September 2010.[25] In poker, stakes levels are distinguished by the size of the small and the big blind bet. To ground our analysis on distinct stakes levels, we selected data from so-called “low”, “medium” and “high stakes” games, with big blind sizes of $0.25, $2 and $10, respectively.

The resulting raw data set contains a total of 76.9 million different hands. The average number of players participating in a hand is 5.9, yielding 456.1 million different player-hand observations. Of these, 190.6 million (41.8%) are from the low stakes games, 229.1 million (50.2%) are from the medium stakes, and 36.4 million (8.0%) are from the high stakes. The smallest number of observations in a month has been recorded in February 2010 (17.3 million, or 3.8%), and relates to a software change that temporarily made data mining more difficult. The peak was in January 2010 (57.9 million, or 12.7%).

Table 13 summarizes the data. Our sample contains over 600,000 different players.[26] About 457,000 of them played at least one hand at our low stakes level ($0.25 big blind), 230,000 played in the medium stakes game ($2 big blind) and 34,000 played in the high stakes game ($10 big blind). They rarely switched between these three levels: nearly all hands (96%) were played at the stakes level at which the player played most frequently. A small minority (16%) was active at more than one of the three levels, but even these players still played 90 percent of their hands at their most favorite level.

[25] For the middle stakes level, the data that we received also contained hands played in September 2009. We treat these as if they were played in October 2009.

[26] We interpret each account as a separate player. People are not allowed to have multiple accounts.

Table 13: Summary statistics

The Table shows descriptive statistics for our full sample of 456.1 million player-hand observations from real money ring games No Limit Texas Hold'em at three different stakes levels. For each stakes level and for the three levels combined, the first two rows show the number of players and the number of hands at the aggregate level. The other rows provide statistics at the player level on the number of hands played, the total winnings expressed as the number of big blinds won, and the average number of big blinds won per hundred hands. Profitability statistics are shown with and without correction for rake (the commission taken by the operator). Winnings are corrected for rake by adding back rake in proportion to players' contribution to the pot. The small, medium and high stakes games have big blinds of $0.25, $2 and $10, respectively.

		Small stakes	Medium stakes	High stakes	All stakes
Total players		457,063	230,098	33,572	611,487
Total hands (in millions)		190.6	229.1	36.4	456.1
Hands	Mean	417	996	1,085	746
	Minimum	1	1	1	1
	Median	52	82	70	71
	Maximum	341,498	763,791	461,743	764,890
	Stdev	2,648	8,261	7,003	5,909
Big blinds won (raked)	Mean	-39	-51	-20	-49
	Minimum	-13,134	-10,461	-6,886	-13,135
	Median	-19	-29	-21	-25
	Maximum	11,641	26,516	31,348	30,501
	Stdev	203	374	405	307
Big blinds won (not raked)	Mean	0	0	0	0
	Minimum	-8,748	-7,605	-6,438	-8,749
	Median	-9	-21	-19	-15
	Maximum	26,209	44,110	38,631	44,832
	Stdev	255	573	467	433
Big blinds won per 100 hands (raked)	Mean	-99	-104	-106	-104
	Minimum	-21,500	-15,740	-12,673	-20,000
	Median	-28	-31	-24	-30
	Maximum	11,200	15,000	10,030	15,000
	Stdev	494	455	588	461
Big blinds won per 100 hands (not raked)	Mean	-83	-95	-103	-88
	Minimum	-21,000	-15,666	-12,670	-19,400
	Median	-16	-24	-22	-20
	Maximum	11,600	15,000	10,040	15,000
	Stdev	474	451	587	448

Players who participated in the high stakes game played, on average, 1,085 hands at that particular level. For the medium and small stakes this number is 996 and 417, respectively. The average number of hands played at the three levels combined is 746. There is much variation across players in the number of hands that they played at the stakes we have selected. One exceptional player was involved in approximately

765,000 hands (0.17% of our sample), while 58.9 percent of all players participated in less than one hundred hands. The degree of concentration is high: the one percent most active players played 58.5 percent of all hands, and 12.0 percent played 90 percent.

Table 13 also shows statistics on players' winnings, both before and after commission taken by the operator. This commission is known as "rake". To compare and combine performance statistics across stakes levels, winnings are scaled by the size of the big blind. For example, a profit of 5 big blinds corresponds to a profit of $50 at the high stakes, and $1.25 at the low stakes. To also account for differences in the number of hands played, performance is expressed as the number of big blinds won per 100 hands (bb/100). For example, a player who has won $20 at a big blind of $2 after playing 400 hands has realized a performance of 2.5 bb/100.

On average, players lost 104 bb/100 after charging of rake. The average win rate is much worse than the ratio of the average total number of big blinds lost (49) and hands played (746), or 6.6 bb/100. This difference is explained by a positive relation between a player's profitability and the number of hands that she played. This relation may reflect the effect of experience, but can also be a consequence of liquidity constraints becoming an obstacle after losses. Only 32 percent of all players in our sample achieved a positive overall result after the deduction of rake.

Rake substantially affects players' winnings. If a hand is not finished in the first betting round ("pre-flop"), the operator takes a fixed percentage (5% for our data) from the pot with a fixed nominal cap that depends on the number of players at the table. To correct players' winnings for rake, we add back the rake in proportion to their contributions to the pot. On average, rake reduces players' performance by 16 bb/100 in our sample. As a result of the fixed nominal cap, the effect of rake on players' win rates is larger for games with smaller stakes. In the absence of rake, 37.5 percent of all players would have made a profit. The extreme values for the best and worst win rate in the Table were recorded for lucky and unlucky players who played only one or two hands.

For our analysis of the role of skill in performance, we measure performance as the win rate in big blinds won per 100 hands before the deduction of rake. Later on, we introduce a second performance measure. We control for rake, because we do not want our findings to be conditional on the rake structure that is employed by the operator. Rake is not an intrinsic element of the game, and the percentages and caps differ across sites. Moreover, the amount of rake that a player effectively pays is not observable. Players can easily participate in reward schemes and receive deposit bonuses that partly make up for it, and via affiliates of the operator they can enter into so-called "rakeback deals" that reimburse 25 percent or more of the amount initially collected.

6.3 Decile Analyses

Under the null hypothesis that poker is a game of chance alone, there is no relation between a player's performance scores across different subperiods. Alternatively, if skill plays a material role in the game of poker, we would expect a player's performance in one particular subperiod to be indicative of her performance in later subperiods. In this Section we subdivide players into deciles based on their performance in the first six months of our sample period and examine how the players in these deciles fared in the last six months. In the next Section we look at the persistence and predictability of performance through regression analysis.

Our sample period covers twelve consecutive months. We split up our data into the subsamples October 2009 – March 2010 and April – September 2010, and rank the different players into deciles according to the average number of big blinds they have won per hand across the first period (the "ranking" period). Because small collections of hands are likely to yield very noisy indicators of performance, we filter out players who have played less than 1,000 hands during this ranking period. This leaves a sample of 17,257 players for the small stakes, 16,435 for the medium stakes, and 2,725 for the high stakes. A total of 36,570 players participated in 1,000 or more hands at the three levels combined. On average, they played 5,706 hands each (median: 2,245). Next, we examine the average performance of the various deciles of players over the second period of six months (the "measurement" period). To prevent selection effects, we impose no restriction on the number of hands in this measurement period. As explained in the previous Section, hand outcomes are corrected for rake and scaled by the size of the big blind.

Table 14 shows the results for the three individual stakes levels (Panel A, B and C) and for the three levels combined (Panel D). The left part of the Table includes the average performance (in bb/100) for each decile over the ranking period (Period 1), while the right part displays how well each decile fared in the measurement period (Period 2).[27] In a nutshell, the results indicate that there is substantial and significant persistence in performance: deciles of players that performed relatively well in the first period on average continued to do so in the second period. The findings for individual stakes levels are generally similar to those for the three levels combined, and our discussion below therefore mainly concentrates on the aggregated sample.

[27] Note that the deciles comprise less players in the measurement period than in the ranking period. Players either ceased to play at some point, moved up or down in stakes, or were simply not covered in our hand histories. For the three stakes levels combined, out of the 36,570 players who played at least 1,000 hands during the first six months, a subgroup of 20,632 were also active during the subsequent six months. On average, the players in this subgroup played 4,814 hands in this second period (median: 717) and 7,038 (median: 2,526) in the first.

Table 14: Standard performance measure deciles

The Table ranks all players who played 1,000 hands or more over the first six months of our sample period into deciles by their performance over these six months, where performance is measured as the average number of big blinds won per hundred hands after correction for rake. For each decile, the first columns show the number of included players (*N*) and their average performance (bb/100) for this ranking period (Period 1). The next columns show the number of players from the decile who played at least one hand in the last six months of our sample period, as well as their average performance for this measurement period (Period 2) and how they rank on average relative to all other players who played at least one hand (Rank). Average decile performance (bb/100) is expressed both as a straight average (unweighted) and as a weighted average, where the weights are either the square roots of players' number of hands (weighted by √n) or players' number of hands (weighted by n). Panel A (B/C) shows the results for observations from the small (medium/high) stakes level separately. Panel D shows the results for all stakes levels combined. For each panel, the Table shows the Spearman rank correlation between the two periods for the average performance at the decile level (for each weighting method) and for performance at the player level (in Rank column).

	Period 1				Period 2				
		bb/100				bb/100			
Decile	*N*	un-weighted	weighted by √n	weighted by n	*N*	un-weighted	weighted by √n	weighted by n	Rank
Panel A: Small stakes									
1	1,726	36.6	35.8	34.7	826	-12.9	3.5	6.8	4,034
2	1,725	19.7	19.6	19.5	862	-12.4	4.7	7.5	3,985
3	1,726	13.5	13.5	13.4	866	1.4	5.2	6.7	3,895
4	1,726	9.2	9.2	9.2	842	-2.4	4.7	6.5	3,956
5	1,726	5.9	5.9	5.9	837	-2.3	4.8	6.4	3,925
6	1,725	3.0	3.0	3.0	775	-8.3	3.4	4.9	4,068
7	1,726	-0.1	0.0	0.0	759	-7.5	3.3	5.1	4,141
8	1,726	-4.0	-3.9	-3.8	771	-11.1	1.4	3.8	4,181
9	1,725	-10.2	-10.1	-10.0	790	-26.7	-0.1	4.7	4,320
10	1,726	-30.0	-29.3	-28.4	890	-20.2	-7.5	-1.3	4,587
Correlation						0.370	0.818	0.964	0.074
(*p*-value)						(0.296)	(0.007)	(0.000)	(0.000)
Panel B: Medium stakes									
1	1,644	34.0	33.0	31.6	946	-25.6	-3.5	3.8	4,398
2	1,643	16.0	15.7	15.3	983	-14.7	0.9	6.0	4,229
3	1,644	9.8	9.7	9.6	937	-22.0	1.6	5.9	4,253
4	1,643	6.0	6.0	5.9	923	-14.2	3.0	5.3	3,972
5	1,644	3.0	3.0	3.1	894	-10.2	0.9	3.8	4,132
6	1,643	-0.2	-0.2	0.0	833	-36.2	-2.9	2.3	4,485
7	1,644	-4.4	-4.3	-4.2	821	-18.8	-4.2	1.8	4,627
8	1,643	-9.7	-9.6	-9.5	831	-29.6	-5.8	1.4	4,673
9	1,644	-18.2	-18.2	-18.0	862	-40.5	-10.3	-2.7	4,806
10	1,643	-41.6	-40.8	-39.9	905	-48.8	-18.0	-5.2	5,190
Correlation						0.612	0.733	0.927	0.104
(*p*-value)						(0.066)	(0.021)	(0.000)	(0.000)
Panel C: High stakes									
1	273	36.8	35.7	33.9	177	-4.7	3.5	5.8	783
2	272	16.3	16.1	15.8	171	-12.7	-0.8	2.6	862
3	273	9.7	9.6	9.5	181	4.3	2.8	2.9	797
4	272	5.8	5.8	5.8	191	-0.3	2.2	3.9	793
5	273	2.8	2.8	2.8	182	-3.8	2.1	4.1	817

6	272	-0.1	0.0	0.2	179	-17.0	0.6	3.5	887
7	273	-4.0	-3.9	-3.7	159	-5.6	-1.9	-1.1	832
8	272	-9.1	-9.0	-8.8	148	-7.6	-4.0	-0.8	891
9	273	-16.5	-16.4	-16.1	158	-32.1	-9.5	-4.1	950
10	272	-40.2	-39.2	-38.0	150	-21.5	-9.6	-1.9	906
Correlation						0.636	0.879	0.770	0.088
(p-value)						(0.054)	(0.002)	(0.014)	(0.000)
Panel D: All stakes									
1	3,657	34.7	33.8	32.7	2,082	-23.0	-0.2	5.1	10,135
2	3,657	17.5	17.4	17.2	2,126	-13.1	2.2	6.1	9,810
3	3,657	11.3	11.3	11.1	2,140	-12.6	2.7	6.0	9,736
4	3,657	7.3	7.2	7.1	2,105	-6.0	3.8	5.9	9,442
5	3,657	4.1	4.1	4.1	2,099	-13.7	2.3	4.5	9,850
6	3,657	1.1	1.1	1.2	2,004	-13.3	1.3	3.7	9,907
7	3,657	-2.5	-2.5	-2.4	1,939	-16.0	-1.0	3.0	10,382
8	3,657	-7.4	-7.3	-7.3	1,948	-22.0	-3.5	1.7	10,764
9	3,657	-14.9	-14.8	-14.7	2,047	-25.7	-6.0	0.0	11,062
10	3,657	-37.7	-37.0	-36.1	2,142	-42.5	-16.5	-5.9	12,098
Correlation						0.600	0.733	0.927	0.101
(p-value)						(0.073)	(0.021)	(0.000)	(0.000)

We first discuss the results where measurement period decile performance is calculated as the unweighted average performance across players. In general, players from higher-ranked deciles outperform players from lower-ranked deciles. For example, the average player from the top decile for the three stakes levels combined lost 23.0 bb/100, while the average player from the bottom decile lost 42.5 bb/100; the difference of 19.5 bb/100 is statistically significant (t = 3.12; p = 0.002). However, across all deciles, the Spearman rank correlation between the average decile performances in the ranking period and those in the measurement period is only marginally significant ($\rho = 0.600; p = 0.073$). At the individual stakes levels, the correlation coefficient is not significant for the small stakes, and marginally significant for both the medium stakes and the high stakes.

The unweighted average in period two is negative for all ten deciles. This result is related to the equal weight assigned to every player in calculating decile performance. There is much variation across players in the number of hands they played in Period 2: this number ranges from 1 to 622,936. Because liquidity constraints can force players to stop playing when losses accumulate, a negative average result from a bad sequence of hands is less likely to be cancelled out or diluted by subsequent hands than a positive result after a streak of luck. Consequently, (extremely) negative average performances at the player level are more likely to occur than (extremely) positive average performances. Indeed, players who played relatively few hands in Period 2 have lower scores: those who played less than 100 hands (18.9% of all active players) recorded a score of -79.9 bb/100, while the others (81.1%) recorded -4.6 bb/100 on average.

The substantial share of players who played relatively few hands in the measurement period may also explain why the decile-level correlation between the average performances in the ranking and measurement period is only marginally significant. Performance measurements for infrequent players are relatively noisy, and their widely varying scores consequently distort the strength of the correlation. In fact, players who played only a few hands are given a questionably large weight when decile performance is expressed as a straight average across players. Using a weighted average with players' number of hands as weights would avoid this problem, and we therefore propose this measure as an alternative indicator. This weighted average is identical to the average profitability per hand across all hands played by the players in a decile combined. Because players who played only infrequently are hardly reflected in this alternative measure, we also consider a compromise weighting method that uses the square roots of players' numbers of hands as weights.

Evaluating on the basis of the two weighted average performance measures strengthens the pattern observed. Players from higher-ranked deciles again outperform players from lower-ranked deciles in Period 2. For example, hands played by players in the top decile yielded a profit of 5.1 bb/100 across all stakes levels, while hands played by bottom decile players lead to a loss of 5.9 bb/100 (difference: 11.0 bb/100; t = 12.36; p < 0.001). The Spearman rank correlations across the deciles between performances in the two periods are higher with weighted than with unweighted average scores and always statistically significant – both for the three individual stakes levels and for the three levels combined (for all stakes and weightings: $\rho \geq 0.733, p \leq 0.021$). Note that the measurement period performance is positive for most deciles when players' numbers of hands are used as weights. This is striking, because, by definition, the average winnings per hand are zero across all hands in our unfiltered sample. Apparently, players who played 1,000 or more hands in the prior six months (and thus satisfied our selection criterion) played more profitably than others. This, in itself, might indicate that experience pays off in this game.

The persistence of performance also appears from how players in a given decile rank relative to all other players in Period 2. The last column of Table 14 shows that players from higher-ranked deciles generally rank higher than players from lower-ranked deciles do. For example, for all stakes levels combined, the average rank of top-decile players is 10,135 (out of 20,632), while that of bottom-decile players is 12,098 ($t = 9.77$; $p < 0.001$).

At the individual player level, the strength of the correlation between players' Period 1 and Period 2 ranks is rather moderate. The correlation coefficient ranges between 0.074 (for the small stakes) and 0.104 (for the medium stakes). The relatively low degree of correlation as compared to the correlation coefficients at the decile level reflects the relevance of variance in performance at the individual level – in particular of the variance for players who played only few hands in Period 2. Statistically, however, the rank correlation at the individual player level is highly significant for every (sub)sample (all $p < 0.001$).

As a robustness check, we have also run similar analyses that use three instead of six months as the ranking and measurement period, where we divided our one-year sample period into four non-overlapping quarters (Q1 = October – December 2009, Q2 = January – March 2010, Q3 = April – June 2010, and Q4 = July – September 2010). Regardless of the pair of successive quarters that we use for ranking and measuring, we observe the same pattern of persistence as before: higher-ranked deciles generally outperform lower-ranked deciles. Again, the correlations are stronger when we reduce the influence of relatively infrequent players by calculating performance as a weighted average, and at the individual player level the rank correlation is always highly significant.

Thus far we have ranked players on the basis of their average performance in big blinds. Though simple and natural, this approach ignores the importance of differences between players in the number of hands that they played. Few would share the view that a player who has won 500 big blinds over 1,000 hands (50 bb/100) is to be considered a better performing player than someone who has won

40,000 big blinds over 100,000 hands (40 bb/100). One of the drawbacks of the previous approach is that it does not account for the basic statistical rule that the sampling distribution of the mean depends on the sample size (n_i): the greater the number of observations, the less likely that the mean takes an extreme value. For example, if we consider two players with equal ability from a larger population, the player who participates in a smaller number of hands is more likely to be classified in one of the top or bottom deciles if players are ranked by their average winnings per hand. Similarly, the previous approach does not account for differences in playing style or the standard deviation of winnings (s_i): when two players are equally profitable, the more adventurous player is more likely to end up in one of the two extremes of the ranking.

We therefore propose an alternative measure to rank players that accounts for the number of hands and playing style of an individual player (i):

$$PRM_i = \frac{win\ rate_i}{std(win\ rate_i)} = \frac{BB_i/n_i}{s_i/\sqrt{n_i}} = \frac{BB_i}{s_i\sqrt{n_i}} \tag{6.1}$$

where BB_i is the sum of big blinds won (before deduction of rake), s_i is the standard deviation of big blinds won, and n_i is the number of hands played. We label this measure the "performance robustness measure" (PRM). In fact, PRM_i equals the t-value of a test of a player's observed performance against the null-hypothesis of zero expected performance.

Table 15 presents the new results. Accounting for playing style and number of hands in the ranking period strengthens the previous evidence for performance persistence. Deciles of players who rank higher by their PRM_i generally fare better than lower-ranked deciles. The new ranking method turns out to be more accurate: in many cases, the performance of a decile in Period 2 is now perfectly or almost perfectly monotonically increasing with the rank of a decile for Period 1. For example, for the aggregate data, the rank correlation is perfect when Period 2 decile performance is measured with players' numbers of hands as weights. For each stakes level, the rank correlation of performance at the individual player level is stronger as well. The new coefficients are about two to four percentage points larger, and range between 0.091

(small stakes) and 0.148 (medium stakes). Additional analyses with three-month periods yielded similar results.

Table 15: Performance robustness measure deciles

The Table ranks all players who played 1,000 hands or more over the first six months of our sample period into deciles by their performance over these six months. Here, the performance measure that is used to rank players is the performance robustness measure, which is defined as the average number of big blinds won per hand after correction for rake divided by its estimated standard error. The estimated standard error is the sample standard deviation of the rake-corrected winnings per hand divided by the square root of the number of hands. The statistics shown for each resulting decile are defined as in Table 14. The various panels and correlation coefficients are also identically defined.

	Period 1				Period 2				
		bb/100				bb/100			
Decile	N	un-weighted	weighted by √n	weighted by n	N	un-weighted	weighted by √n	weighted by n	Rank
Panel A: Small stakes									
1	1,726	26.9	21.1	16.0	908	-1.3	6.2	7.4	3,777
2	1,725	21.0	17.4	13.9	822	1.4	6.1	6.8	3,843
3	1,726	16.1	13.5	10.7	856	-6.0	3.9	5.5	4,008
4	1,726	11.9	10.1	8.2	822	-10.8	3.0	5.4	4,089
5	1,726	7.8	6.7	5.5	819	-8.4	3.7	6.0	4,030
6	1,725	4.1	3.5	2.9	773	-10.9	2.8	5.2	4,094
7	1,726	-0.1	0.0	0.0	784	-7.9	2.6	5.3	4,180
8	1,726	-5.0	-4.4	-3.8	794	-19.0	0.6	3.6	4,236
9	1,725	-11.5	-10.3	-8.8	807	-23.2	0.2	4.6	4,294
10	1,726	-27.6	-24.9	-21.6	833	-17.3	-6.6	-0.6	4,587
Correlation						0.867	0.988	0.938	0.091
(p-value)						(0.003)	(0.000)	(0.000)	(0.000)
Panel B: Medium stakes									
1	1,644	20.9	12.6	8.7	1,084	-6.4	4.7	6.0	3,709
2	1,643	19.4	13.7	8.9	946	-8.7	2.0	5.2	4,015
3	1,644	14.4	10.9	7.2	920	-20.5	-0.4	4.3	4,223
4	1,643	9.5	7.7	5.6	867	-23.8	-4.1	2.4	4,600
5	1,644	4.4	3.6	2.6	850	-34.7	-4.2	2.5	4,541
6	1,643	-0.3	-0.2	-0.2	856	-32.7	-5.6	1.1	4,602
7	1,644	-5.6	-4.9	-3.9	827	-26.5	-5.5	2.1	4,670
8	1,643	-11.6	-10.0	-7.9	873	-40.2	-6.5	1.7	4,741
9	1,644	-20.2	-17.7	-14.6	836	-29.1	-8.4	-1.0	4,777
10	1,643	-36.3	-32.1	-26.4	876	-42.6	-14.3	-4.8	5,063
Correlation						0.855	0.988	0.952	0.148
(p-value)						(0.004)	(0.000)	(0.000)	(0.000)
Panel C: High stakes									
1	273	24.0	15.1	10.1	204	9.4	4.3	4.1	705
2	272	20.6	14.7	9.1	176	-0.1	2.1	3.6	807
3	273	13.5	9.7	5.9	185	-11.5	0.2	2.1	854
4	272	8.8	6.8	4.6	169	-12.1	-0.6	3.1	863
5	273	4.3	3.3	2.3	173	-5.8	2.0	5.7	832
6	272	-0.1	-0.1	0.0	170	-16.4	0.3	2.1	863
7	273	-5.7	-4.6	-3.6	155	-4.6	-2.6	1.2	868
8	272	-11.1	-9.1	-6.9	159	-8.6	-4.0	-0.7	865
9	273	-18.3	-15.1	-11.3	150	-35.9	-6.5	-2.1	920

10	272	-34.6	-29.1	-22.6	155	-17.9	-10.1	-2.7	959
Correlation						0.697	0.903	0.867	0.118
(p-value)						(0.031)	(0.001)	(0.003)	(0.000)
Panel D: All stakes									
1	3,657	23.6	15.6	10.2	2,324	-7.0	5.2	6.2	8,885
2	3,657	19.8	15.1	10.4	2,112	-5.3	3.4	5.3	9,486
3	3,657	14.8	11.7	8.4	2,076	-10.7	1.4	4.8	9,966
4	3,657	10.3	8.4	6.2	2,038	-20.3	0.4	4.4	10,113
5	3,657	6.0	5.0	3.7	2,006	-21.6	0.0	4.2	10,248
6	3,657	1.5	1.3	1.0	2,005	-19.5	-1.4	2.9	10,398
7	3,657	-3.3	-2.8	-2.3	1,970	-23.8	-2.7	2.7	10,654
8	3,657	-9.0	-7.8	-6.3	2,002	-26.1	-4.5	1.5	10,902
9	3,657	-16.7	-14.7	-12.1	2,017	-24.2	-5.1	1.2	10,892
10	3,657	-33.5	-29.8	-24.8	2,082	-31.8	-12.2	-3.9	11,852
Correlation						0.939	1.000	1.000	0.131
(p-value)						(0.000)	(0.000)	(0.000)	(0.000)

Another way to look at the persistence of performance is through transition probabilities. Table 16 shows transition probabilities across performance deciles for players who played 1,000 hands or more over the first six months of our sample period. These players are ranked on the basis of their performance twice: for Period 1 and for Period 2. The probabilities in the Table indicate the empirical probability of transitioning from a given decile in the first half-year period to a given decile in the second half-year period. Players for whom we have no observations for the second period are not included in the ranking for the second period, so essentially the probabilities are conditional on participation in the second six months.

In Panel A, the performance measure that is used to rank players is the standard performance measure (bb/100) after correction for rake. The fraction of players in the top decile of Period 1 who end up in the top decile in Period 2 is 13.6 percent; players who are in the worst decile end up in the worst decile 19.6 percent of the time. These empirical probabilities are substantially greater than the value of 10 percent that would be expected under the null hypothesis of no performance persistence (all $p < 0.001$). At the same time, however, there is some evidence that the likelihood of ending up at the opposite extreme is also greater than 10 percent. For example, the chance of transitioning from the very best (worst) category to the very worst (best) is 12.3 (11.2) percent. This pattern is symptomatic of the inadequacy of the ranking measure used here: players with a higher variance of their average winnings due to adventurous or infrequent play are more likely to end up in the extreme win rate categories. Ranking players on the basis of our alternative performance robustness measure controls for this variance effect.

In Panel B, players are ranked on the basis of their PRM_i.[28] The results are compelling: players from the top decile reappear in this decile 20.7 percent of the time, and with a probability of 5.4 percent they end up in the bottom decile relatively infrequently. Similarly, losers are unlikely to become winners: the worst ten percent of players rank among the best ten percent in the next six months only 5.2 percent of the time and among the worst ten percent 18.5 percent of the time. The empirical probabilities are even more telling when we look at percentiles (not tabulated): the very best one percent of players in Period 1 rank among the very best one percent in Period 2 11.4 percent of the time, and among the best ten percent 32.8 percent of the time (11.4 and 3.3 times the base rate). They are among the worst ten percent only 3.4 percent of the time. Similarly, the least successful players from Period 1 often keep performing badly: the worst percentile stay in that category 10.2 percent of the time, and belong to the worst decile in 32.0 percent of the cases. They rarely outperform: the best decile is reached only 2.7 percent of the time.

[28] If $n_{i,2}$ < 1,000 we substitute $s_{i,2}$ by $\tilde{s}_{i,2} = \sqrt{((1{,}000 - n_{i,2})s_{i,1}^2 + n_2 s_{i,2}^2)/1000}$, where $n_{i,t}$ is the number of hands played by player i in period t, and $s_{i,t}$ is the standard deviation of big blinds won by player i in period t. This approach avoids the use of an unreliable standard deviation estimate for players who participated in a relatively small number of hands during Period 2, and assumes that σ_i is stable through time. In the extreme case where $n_{i,2} \leq 2$ we set $\tilde{s}_{i,2} = s_{i,1}$.

Table 16: Transition probabilities

The Table shows the transition probabilities across performance deciles for players who played 1,000 hands or more over the first six months of our sample period. Each probability indicates the empirical probability of transitioning from a given performance decile in the first half-year period (Period 1) to a given performance decile in the second half-year period (Period 2). In Panel A, the performance measure that is used to rank players in Period 1 and Period 2 is the standard performance measure, where performance is measured as the average number of big blinds won per hundred hands after correction for rake (bb/100). In Panel B, the performance measure that is used for Period 1 and Period 2 is the performance robustness measure, which is defined as the average number of big blinds won per hand after correction for rake divided by the estimated standard error. Players for whom we have no observations for Period 2 are not included in the Period 2 ranking.

Period 1 decile	Period 2 decile 1	2	3	4	5	6	7	8	9	10
Panel A: Ranking by Standard Performance Measure										
1	0.136	0.116	0.098	0.089	0.063	0.078	0.095	0.094	0.108	0.123
2	0.108	0.108	0.131	0.097	0.094	0.086	0.099	0.096	0.089	0.091
3	0.087	0.114	0.124	0.119	0.115	0.099	0.094	0.083	0.079	0.086
4	0.083	0.103	0.124	0.143	0.140	0.103	0.093	0.090	0.069	0.053
5	0.076	0.091	0.118	0.133	0.131	0.126	0.099	0.083	0.071	0.072
6	0.100	0.087	0.093	0.119	0.143	0.119	0.100	0.087	0.079	0.072
7	0.101	0.093	0.094	0.097	0.099	0.121	0.092	0.109	0.104	0.091
8	0.094	0.105	0.079	0.079	0.088	0.109	0.118	0.120	0.105	0.103
9	0.103	0.089	0.081	0.071	0.077	0.097	0.118	0.124	0.130	0.110
10	0.112	0.092	0.056	0.051	0.051	0.065	0.094	0.116	0.166	0.196
Panel B: Ranking by Performance Robustness Measure										
1	0.207	0.133	0.118	0.107	0.091	0.078	0.083	0.067	0.060	0.054
2	0.126	0.117	0.107	0.105	0.108	0.108	0.102	0.091	0.076	0.060
3	0.101	0.110	0.116	0.104	0.106	0.090	0.095	0.104	0.088	0.086
4	0.105	0.104	0.110	0.099	0.093	0.099	0.095	0.112	0.094	0.090
5	0.101	0.090	0.099	0.107	0.104	0.113	0.095	0.100	0.098	0.093
6	0.087	0.092	0.092	0.103	0.112	0.101	0.105	0.102	0.111	0.095
7	0.072	0.108	0.091	0.090	0.099	0.107	0.106	0.102	0.121	0.105
8	0.067	0.089	0.093	0.094	0.106	0.103	0.101	0.111	0.119	0.115
9	0.063	0.078	0.090	0.104	0.100	0.113	0.114	0.100	0.116	0.122
10	0.052	0.074	0.080	0.087	0.081	0.091	0.107	0.116	0.125	0.185

6.4 Regression Analyses

To further analyze the role of skill we regress performance over the final six months of our sample on performance over the first six months, and on other measures that may serve as skill proxies. We consider the following variables:

SPM: the standard performance measure or “win rate”, defined as the average number of big blinds won per hundred hands after correction for rake.

PRM: the performance robustness measure, defined as the average number of big blinds won per hand after correction for rake divided by the estimated standard error.

The estimated standard error is the sample standard deviation of the rake-corrected winnings per hand divided by the square root of the number of hands.

Hands (log): the natural logarithm of the number of hands played. This variable is a proxy for the experience of players and thus a possible indicator of skill.

Tightness: one minus the proportion of hands in which a player voluntarily wagered money in the first betting round ("called or raised before the flop"). The degree of tightness is one of the two simple measures that are typically used to broadly categorize players' playing styles. Generally, tighter play is thought to be indicative of a better player. Common mistakes in poker are to impatiently look for "action" and to overestimate the profitability of playing a given hand.

Aggressiveness: the number of times a player led the betting ("bet" or "raised") as a proportion of the total number of times the player voluntarily wagered money ("bet", "called" or "raised"). This factor is the other of the two simple playing style measures. Aggressive play is generally thought to yield a higher expected performance than passive play, because increasing the cost of playing at the right times can pressure other players to give up stronger cards or to wager more with weaker ones.

Tournaments: a player's tournament ability rating according to SharkScope, a website that collects virtually all online poker tournament results. The worst possible rating is 50 and the best possible rating is 100. The exact calculation is not disclosed by SharkScope. Tournament performance is a possible indicator of skill, because of the many similarities between tournament and cash game play.

The last three variables are standardized such that they have a mean of zero and a standard deviation of one. To avoid endogeneity issues, all six explanatory variables are solely based on data from before Period 2: the first five cover the prior six months (Period 1), and the tournament ability rating is determined over the prior twelve months. The tournament ability rating was available for 79 percent of the players who played 1,000 or more hands in Period 1.

We run two sets of regressions, one for the standard performance measure and the other for our performance robustness measure. In the former case, we face the issue of heteroskedasticity: the variance of the error term is proportional to the sample variance of the number of big blinds won (s_i^2) and inversely proportional to the number of hands played (n_i) in Period 2. We therefore apply weighted least squares (WLS) to estimate these regression models, where the weighing factor is the inverse of the variance of the error term (n_i / s_i^2). When our performance robustness measure is the dependent variable we use ordinary least squares (OLS), because the errors have constant variance by construction.

Panel A of Table 17 presents the WLS results for the standard performance measure. In each univariate regression, performance is significantly related to the skill proxy from the previous period (all $p < 0.001$). Not only the historical performance measure (Model 1), but also the number of hands played (Model 2), the two style measures (Models 3 and 4) and the tournament ability variable (Model 5) predict performance to a modest but statistically significant extent. Players who participated in more hands in the previous period perform better, as do players who adopted a tight or aggressive playing style and players who did well in tournaments. Combined, the measures explain 3.3 percent of the variance in performance. The smaller-than-unity coefficient in Model 1 indicates that there is regression in players' performance over time.

We obtain qualitatively similar results when we use our performance robustness measure (Panel B), but the explanatory power is higher now. The percentage of variance explained by the joint skill proxies is 8.1 percent, which is about 2.5 times as high as the empirical fit of the previous multivariate specification. We have also performed the regression analyses for the three stakes levels separately. The results and conclusions are all similar to the results for the aggregate sample.

Table 17: Regression results

The Table displays the regression results for our subsample of players who played 1,000 hands or more over the first six months of our sample period (Period 1) and at least 1 hand over the second six months (Period 2). The dependent variable is either the standard performance measure (Panel A) or the performance robustness measure (Panel B). The standard performance measure is defined as the average number of big blinds won per hundred hands after correction for rake (bb/100). The performance robustness measure is the average number of big blinds won after correction for rake divided by its estimated standard error. All explanatory variables are calculated using data from Period 1 only. The results reported in Panel A are weighted least squares regression results with the ratio of a player's number of hands and her sample variance of the number of big blinds won in Period 2 as weight. Panel B presents ordinary least squares results. The *p*-values are in parentheses.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Panel A: Standard Performance Measure (WLS)							
Constant	3.225	-2.536	2.433	3.435	3.984	-0.057	-1.204
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.930)	(0.031)
SPM	0.167					0.141	0.148
	(0.000)					(0.000)	(0.000)
Hands (log)		0.687				0.199	0.338
		(0.000)				(0.003)	(0.000)
Tightness			2.048			1.594	1.368
			(0.000)			(0.000)	(0.000)
Aggressiveness				1.101		0.668	0.532
				(0.000)		(0.000)	(0.000)
Tournaments					0.412	0.509	
					(0.000)	(0.000)	
R^2	0.022	0.007	0.012	0.007	0.001	0.033	0.035
N	20,632	20,632	20,632	20,632	16,368	16,368	20,632
Panel B: Performance Robustness Measure (OLS)							
Constant	-0.055	-2.267	0.010	0.010	0.009	-1.289	-1.333
	(0.000)	(0.000)	(0.269)	(0.273)	(0.366)	(0.000)	(0.000)
PRM	0.229					0.123	0.142
	(0.000)					(0.000)	(0.000)
Hands (log)		0.281				0.155	0.161
		(0.000)				(0.000)	(0.000)
Tightness			0.242			0.131	0.125
			(0.000)			(0.000)	(0.000)
Aggressiveness				0.187		0.080	0.085
				(0.000)		(0.000)	(0.000)
Tournaments					0.078	0.041	
					(0.000)	(0.000)	
R^2	0.049	0.049	0.035	0.021	0.004	0.081	0.086
N	20,632	20,632	20,632	20,632	16,368	16,368	20,632

Although these results reinforce our earlier findings of performance persistence and the role of skill in poker, the major part of variance in performance remains unexplained by the models and appears to be attributable to chance. An issue that we have not yet explicitly addressed so far is the problem of errors in variables. In the ideal situation we would know every player's precise skill level, but given the lack of this information we have to use noisy proxies. When explanatory variables are mismeasured, coefficients estimated via standard regression methods are biased towards zero and the true explanatory power is underestimated. The low empirical fit of the regression models indicates that measurement error is a serious issue for the historical performance measures: if a random factor explains much of the variation in performance, any measurement of previous performance is likely to be subject to a large degree of randomness as well.[29]

The bias of an estimated coefficient towards zero as a consequence of measurement error is known as *attenuation* or *regression dilution*. Although measurement error is not a problem for predictive modeling, it can give an unjust impression of the size of the effect of skill on performance here and may falsely suggest that a player's skill is not a stable quality over time. To account for error in both the dependent and the independent variable we therefore also run a so-called Deming regression (methodological details are in Appendix 1). The results indicate that the standard regression understates the size of the effect of skill on performance to a considerable extent: when we regress the win rate from Period 2 on the win rate from Period 1, we obtain a coefficient of 1.392 ($p < 0.001$). When the performance robustness measure is used for the dependent and for the independent variable, the coefficient is 1.156 ($p < 0.001$). These new coefficients are not only substantially higher and closer to unity than the values of 0.167 and 0.229 reported before, but also significantly greater than unity ($p < 0.001$). Taken at face value, the coefficients indicate that the disparity in performance between players increases over time.

The underestimation of the true explanatory power of skill as a consequence of measurement error decreases with the number of hands used to calculate the proxy for skill. With more observations, measurement error becomes relatively less important: the ratio of the variance of the measurement error and the variance of the true explanatory variable decreases with the number of hands. This holds for each of our two historical performance measures. For the standard performance measure the variance of the measurement error decreases as the number of hands increases. For the performance robustness measure the variance of the measurement error of is

[29] The playing style variables are measured with relatively little error because they are based on a large number of draws from a binomial distribution. Their relatively poor predictive power appears to be especially related to their more indirect reflection of skill.

constant (at unity), but for this measure an increase in the number of hands leads to more distinctive differences between players with a different expected win rate, reducing the relative size of measurement errors.

To illustrate the effect of the number of observations per player on the empirical fit, we run regressions for the pooled results of "teams" of players. More precisely, we first rank players on the basis of their performance in Period 1. We group the players into percentiles, where the best one percent of players forms a group, the second best one percent form another group, et cetera. Next, for each percentile we calculate Period 1 and Period 2 performance across all hands of the players in the group combined. Last, we regress the pooled Period 2 performance on the pooled Period 1 performance. The average hypothetical "player" has now played about 2.1 million hands in Period 1 (instead of 7,038) and 1.0 million in Period 2 (instead of 4,814). The results are remarkable. When performance is expressed as the win rate the R^2 is 66.7 percent, and when the performance robustness measure is used the R^2 is 80.1 percent.

We conclude this Section with a robustness analysis. To make sure that the documented persistence of performance truly reflects the role of skill, we need to verify that the results are not driven by differences in liquidity constraints between players. As already explained in the Section 6.2, a liquidity constraint can force a player to stop playing when losses accumulate, and, consequently, a negative performance is less likely to be cancelled out or diluted by subsequent hands than a positive performance. The stronger a player's liquidity constraint at the start of a given period, the greater the likelihood that she needs to stop early after losses, and the lower her expected average performance over this period. Differences in liquidity constraints across players can be both exogenously and endogenously determined: some players may simply have smaller fixed budgets for playing than others in each period, and players who have lost in a previous period have less funds available in their accounts than players who have won. In both cases, the contemporaneous relation between the strength of a liquidity constraint and performance can lead to spurious correlation in players' performance through time.

To avoid the possible influence of liquidity constraints, we use hand samples of a fixed size for every player. For n = 1,000, 5,000 and 10,000, we select all players who have played at least 2n hands over our entire sample period, and test whether performance over the first n hands is predictive of performance over the following n hands.

The regression results are in Table 18, and point out that the persistence of performance is robust to this alternative specification. Regardless of n and regardless of which of the two performance measures is being used, performance over the second n hands is significantly related to performance over the first n hands (all $p < 0.001$).

Note that there are also two downsides to this alternative approach. First, there is more measurement error because in many cases fewer hands are being used to proxy for skill (especially when n = 1,000). Raising n solves this issue, but comes at the cost of the inclusion of fewer players. Second, if losing players play less (because of liquidity constraints or a lost appetite to play), the selection criteria lead to a more homogeneous set of players in terms of their performance. Together with the exclusion of a possible spurious effect of liquidity constraints, these effects may explain why the R^2 values for the two present Model 1 specifications (0.5% for SPM and 0.6% for PRM) are remarkably lower than before (2.2% and 4.9%, respectively; see Table 17). Increased measurement error also explains why the regression coefficients are smaller than before.[30]

To account for measurement error we also estimate the six univariate models for the same fixed-size hand samples using Deming regression. Interestingly, all six resulting coefficients are qualitatively close to unity (between 0.86 and 1.11) and only two are statistically significantly different from unity. This suggests that the Deming coefficients found before were larger than unity due to a spurious effect from liquidity constraints that is now eliminated, and, more importantly, this result points out that skill differences between poker players are close to constant over time.

[30] Note that the R^2 values for the different fixed numbers of hands suggest that decreasing measurement errors are more important than increasing homogeneity when n increases from 1,000 to 5,000, and that the two effects are limited or (more or less) cancel out with a further increase to 10,000. Furthermore, the larger regression coefficients for n = 5,000 than for n = 1,000 again underline the nature and role of measurement error, and the larger constants for larger n confirm the selection effect that occurs here.

Table 18: Regression results for fixed number of hands

The Table displays the regression results for subsamples of players who played at least 2*n* hands during our entire sample period, with *n* = 1,000, 5,000 or 10,000. The dependent variable is the player's performance over the second *n* hands, as measured by either the standard performance measure (Panel A) or the performance robustness measure (Panel B). The explanatory variables are calculated over the first *n* hands.

	Model 1 ($n_1 = n_2 = 1{,}000$)	Model 2 ($n_1 = n_2 = 5{,}000$)	Model 3 ($n_1 = n_2 =$
Panel A: Standard Performance Measure (WLS)			
Constant	2.348	3.656	3.798
	(0.000)	(0.000)	(0.000)
SPM	0.066	0.142	0.152
	(0.000)	(0.000)	(0.000)
R^2	0.005	0.021	0.025
N	31,991	7,340	3,464
Panel B: Performance Robustness Measure (OLS)			
Constant	0.147	0.465	0.705
	(0.000)	(0.000)	(0.000)
PRM	0.074	0.136	0.125
	(0.000)	(0.000)	(0.000)
R^2	0.006	0.018	0.016
N	31,991	7,340	3,464

6.5 Simulations

The previous analyses demonstrate that there is persistence in the performance of poker players. Based on the results, we can confidently rule out that we are dealing with a game of pure chance. Skill is a factor of importance, but the key question left unanswered is whether skill also *dominates* chance, that is, whether poker is more a game of skill than a game of chance. The answer to this question critically depends on the number of hands "the game of poker" is supposed to constitute. The role of chance diminishes with the number of hands, and when the number of hands grows large enough it eventually cancels out.

In the present Section we use simulations to approximate the number of hands above which skill predominates. More specifically, in accordance with the approach in Section 6.3, we first rank all players who have played 1,000 hands or more over the first six months of our sample period according to their performance during that first subperiod. Next, for the best performers, we randomly draw (with replacement) a given number of *h* hands from their combined sample of hands recorded for the second six months and we compute their total winnings (in big blinds) across these h hands.[31] We do the same for the worst performers, and then compare the total winnings of the two player categories. For each different value of h, we repeat this procedure 25,000 times and determine the proportion of times that the supposedly more skilled players do better than their supposedly less skilled counterparts. A similar approach was used by Randal D. Heeb in his expert report for a U.S. Federal court case in New York in 2012.[32]

Skill predominates in this comparison when the proportion is greater than 0.75. This threshold follows from a simple model where we define the skill factor, π_h, as the probability that skill determines the more profitable player across h different hands. Accordingly, 1 - π_h is the chance factor, or the probability that chance determines the more profitable player. When skill alone determines the winner (π_h = 1), the more skilled player always wins; when chance alone determines the winner (π_h = 0), the more skilled player wins half the time. More generally, the overall probability that the more skilled player is ahead after h hands is equal to $p_h = \pi_h \cdot 1 + (1 - \pi_h) \cdot 0.5$. Skill predominates when $\pi_h > 0.5$, implying $p_h > 0.75$.

The accuracy of our simulation approach depends on the formation of two distinct groups of players. With each draw of hands, we want to simulate and compare the winnings of a relatively skilled and a relatively unskilled player. Every time, the former is thus assumed to be the better player, with a higher expected performance than the latter. Because we cannot observe a player's true skill and need to rely on an imperfect proxy, we cannot exclude that our simulations sometimes confuse the two

[31] To circumvent technical limitations, we draw from a representative subset of one million hands when the actual sample size is greater than one million. We have verified that the results are insensitive to this approximation.

[32] Case 1:11-cr-00414-JBW. For one half of the players in his sample, Heeb estimates a regression model that links performance to hundreds of playing style characteristics, including many different variants of *Tightness* and *Aggressiveness* (see Section 6.4). For the other half, he employs the obtained regression coefficients to compute players' predicted performance. Heeb's simulations point out that players who rank high according to this self-constructed skill measure are ahead of lower-ranked players more than 75 percent of the time after only a few hundreds of hands. A weakness of Heeb's analysis is that he measures players' characteristics and their performance over the same set of hands. This is likely to lead to spurious correlation between skill and performance scores, because both scores are contemporaneously co-determined by the same chance elements. For example, players who are dealt a greater fraction of strong hands or hands that connect well with the community cards are more likely to score high on the dimension of aggressiveness (and thus relatively high on skill) and to record a strong performance. Consequently, his analysis is likely to produce an underestimation of the critical number of hands above which skill predominates.

types. To limit this risk, we draw from the hands of players who ended up in either the very best or the very worst performance percentile of the first six months. We use our performance robustness measure to rank players, given our earlier evidence that this measure is more accurate than the standard performance measure. For the sake of completeness, we also run the simulations with the standard performance measure and with the top and bottom deciles of players instead of the top and bottom percentiles.

Figure 8 displays the results. Across a selection of a few hands, the game is hardly different from a pure game of chance: the higher-ranked players perform better only slightly more than half the time. The proportion steeply increases with the number of hands, at a decreasing marginal rate.[33] As indicated by the solid black line, the critical point where the best percentile of players (according to the performance robustness measure) is ahead 75 percent of the times is reached after 1,471 hands. As expected, this number is larger when the best and worst percentiles are being selected on the basis of the standard performance measure (2,139), and even larger when deciles are being used instead of percentiles (6,512 and 7,293 for PRM and SPM, respectively).

[33] In fact, as h increases, the empirical win proportion converges to the win proportion that would result from performance being normally distributed (just as the central limit theorem predicts).

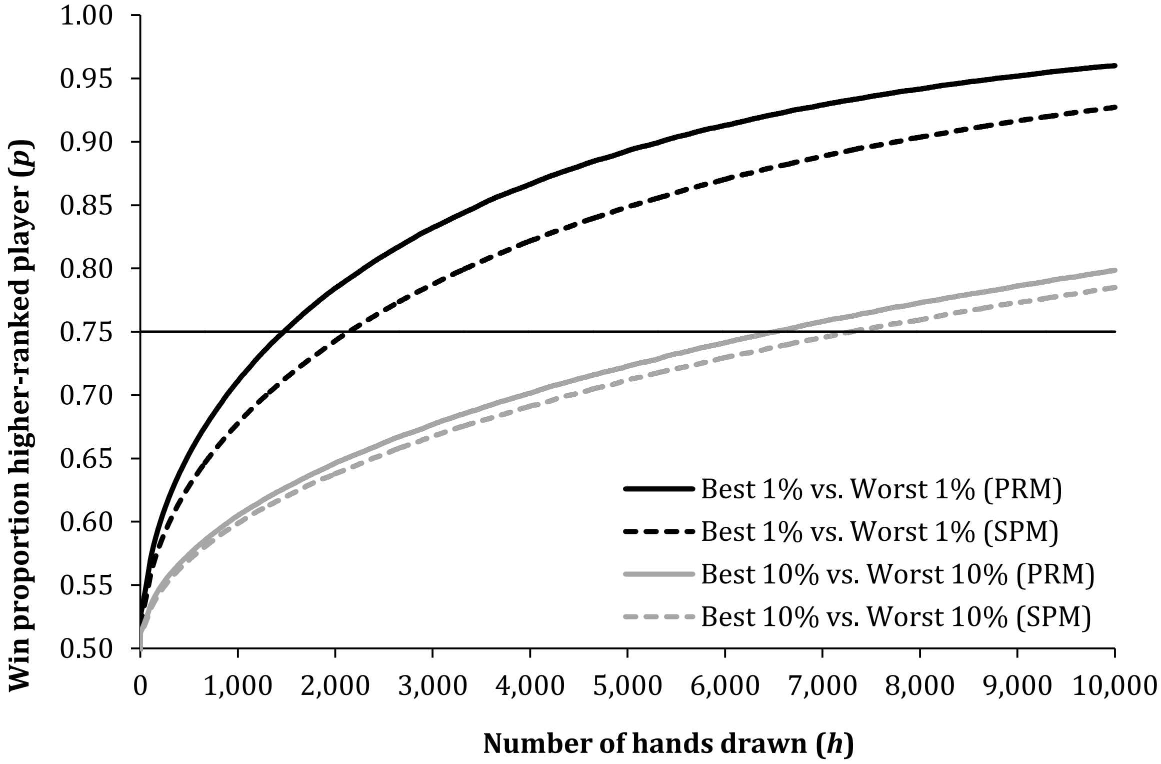


Figure 8: Simulation results

The Figure displays the proportion of times that a selection of *h* randomly drawn hand outcomes for players who were among the best performing players in the past do better than a similar-size selection of hand outcomes for players who were among the worst performing players in the past. Hand outcomes are randomly drawn from the subsample of hands from the second six months of our sample period for players who ranked among the best or worst performing percentiles (black lines) and for players who ranked among the best or worst performing deciles (grey lines) over the first six months of our sample period. Players are ranked according to the performance robustness measure (solid lines) or the standard performance measure (dashed lines). The lines are smoothed, with each point representing the moving average proportion across the simulation outcomes available for *h*-100 up to and including *h*+100.

Figure 9 zooms in on the simulation results for h = 10, 100, 1,000 and 10,000, and shows histograms for the distribution of the difference in win rate (number of big blinds won per hundred hands) between the two groups. While the previous Figure only shows the proportion of times this difference is positive at a given h, the histograms also show the magnitude of the difference in profitability between the higher-ranked and lower-ranked players. Upon visual inspection, the distribution is widely but symmetrically distributed around zero when h = 10. The distribution gradually becomes more centered around its mean when h increases, and, consequently, with a greater h the positive mean win-rate difference of 21.2 bb/100 becomes more apparent (note the different scales for the horizontal axes). At h = 10, 100, 1,000 and 10,000, the fractions where the higher-ranked players are behind amount to 47, 43, 29 and 4 percent, respectively. Their chances of underperforming by more than 10 bb/100 shrink from 44 to 39, 21 and 0.5 percent, respectively. At the critical number of approximately 1,500 hands where skill dominates chance, the frequency of underperformance by more than 10 bb/100 amounts to 16 percent.

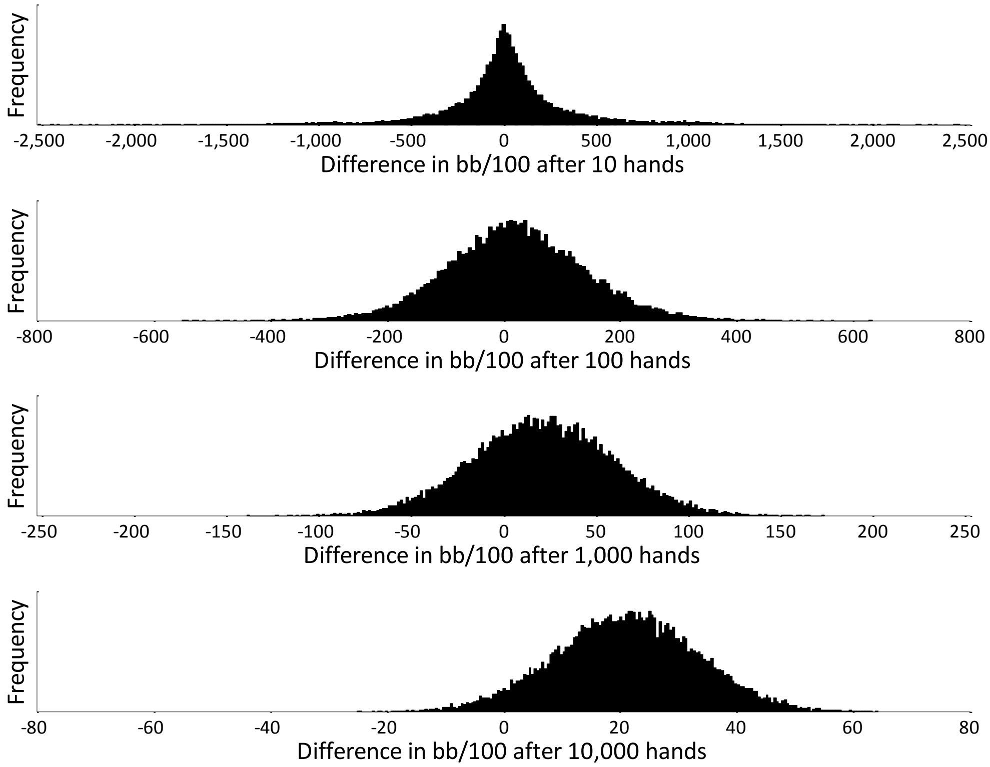


Figure 9: Difference in win rate after 10, 100, 1,000 and 10,000 hands

The histograms display the simulated distributions of the difference in win rate (number of big blinds won per hundred hands) between players who were in the very best and players who were in the very worst performance percentile of the first six months of our sample period. Players are ranked according to the performance robustness measure. For each percentile, h = 10, 100, 1,000 or 10,000 outcomes are randomly drawn from their subsample of hands from the second six months of our sample period.

6.6 Discussion and Conclusions

Our study shows that there is a significant skill factor in online ring game poker, and that this factor dominates the luck factor after a moderate duration of play. In Section 6.3 and Section 6.4 we have examined whether possible skill differences between online poker players explain differences in their performance. The results in these Sections provide strong evidence against the hypothesis that poker is a game of pure chance. For a game of pure chance there would be no correlation in the winnings of players across successive time intervals.

The decile analyses demonstrate that players who rank higher (lower) in profitability over the previous six months generally continue to perform better (worse) than others during the present six months. For example, players from the best decile earn about 20 to 25 big blinds per 100 hands more during the subsequent six months than players from the worst decile. When we rank players on the basis of how well they did according to our alternative performance robustness measure, we find that top ten percent players rank among the top ten percent of the next six months approximately twice as often as others, and among the worst ten percent approximately half as often. The results are even more pronounced if we look at the best one percent. Similarly, those who perform the worst hardly ever end up in the top category.

Our regression results reinforce these findings, and show that current performance is not only related to historical performance but also in some extent to simple measures of playing style. Players who are characterized by a tight or aggressive style generally perform better than their loose or passive opponents. Performance is also related to the number of hands that players have played over the previous period: more frequent or experienced players achieve better results. This finding can indicate that better players choose to play more and that players learn from playing. Both interpretations conflict with the pure-chance hypothesis.

Given these results we believe that we can legitimately conclude that skill is an important factor in online ring game poker.[34] However, most jurisdictions do not ask whether a game involves an *important* degree of skill, but, more specifically, whether skill *predominates*. At the same time, no government seems to prescribe how this should be tested. The key complication is that the extent to which skill differences

[34] Behavioral research also hints that there are important skill elements in poker. A large literature from psychology and behavioral economics finds that people systematically deviate from rational norms when they make decisions under uncertainty, and that there are differences between people in the degree to which they are prone to deficiencies. Appendix 2 highlights some of these insights in the context of poker.

explain differences in performance depends on the number of hands over which performance is measured. If sufficiently many hands are played, skill explains practically all variation in performance. This is nicely illustrated by the high explanatory power of our regressions with the pooled performance scores of percentiles of players, and by the decile analyses, where the (rank) correlation between the past and the current performance of large groups is near-perfect or even perfect. A definite answer to the predominance question thus calls for a definition of the relevant measurement interval. The possible extremes are a single hand and a player's life time, and intermediate options include an average session, a month, and a (fiscal or calendar) year.

Instead of predetermining one particular interval ourselves, we have employed simulations to estimate the number of hands where skill and chance are equally important. In Section 6.5, we basically run horse races of different durations between a relatively skilled player and a relatively unskilled player who are playing independently from each other. These simulations point out that skill dominates chance when performance is measured over 1,500 or more hands of play. To put this number into perspective: at a rate of 60-80 hands per hour per table, playing 1,500 hands takes people who play only one table at a time about 19 to 25 hours (four to six evenings) of play. Participating on multiple tables simultaneously – which is what many experienced players do – effectively reduces this duration to one or two sessions.

As with any empirical estimate, the exact outcome depends on the specific approach. Our estimate that skill predominates after 1,500 hands should be seen in this light. However, we believe that we have taken a conservative approach in testing whether skill predominates, because the two types of players in our simulations were not playing the game against each other, and because of two selection effects.

With few exceptions, the series of hand outcomes that we compare consist of hands that have been played at tables where the relatively skilled and the relatively unskilled players (virtually) sat down with *other* players than their counterparts in our comparison. The hands were played at different tables and at different moments in time. We are thus not analyzing how well a selection of strong players fare against a selection of weak ones, but comparing how well they did against a cross-section of others, including players from their own category. Would higher-ability players be directly playing against lower-ability players, skill should be expected to predominate substantially quicker.

To avoid using extremely noisy historical skill estimates, we have required a minimum data history per player. Unintentionally, this approach is likely to have generated a selection effect: because intensity of play and experience are (almost tautologically) related, relatively inexperienced players are underrepresented in our analyses. Every player in our final sample will normally be well informed about the rules of the game and master some basic strategic concepts. If we would also observe complete beginners playing the game, the differences in performance across players would presumably be greater, and the critical number of hands where skill starts to predominate would consequently be more quickly reached.

The relative homogeneity of our sample is probably strengthened by players' self-selection into stakes levels on the basis of their perception of their skill level. Better players are more likely to play for larger stakes, while worse or beginning players may feel more comfortable at smaller stakes. This self-selection into the game is not unique for poker. In many games, people play against opponents of relatively similar ability. When self-selection happens it takes a longer series of events before skill differences materialize – even with sports and with games like chess and bridge.

We conclude with a brief discussion of the generalizability of our findings. Our study has been limited to online play. Due to a lack of readily available data, it is practically impossible to execute an analogous, large-scale analysis for offline play. Nevertheless, given that skill is important in the online variant, we conjecture that it is likely to be even more important for brick-and-mortar play. One reason is that offline play also involves body language and other subtle forms of communication. Players are sitting face-to-face and need to carefully control their behavior to not reveal the strength of their cards, and by observing others they can sometimes discover useful "tells" about their play. At the same time, body language can also be used to deliberately mislead opponents. Furthermore, players' patience is put to the test more in live play than in online play because fewer hands are dealt per hour. In live poker, skill will probably dominate chance at fewer hands, but because of the slower pace of play and the impossibility to play on multiple tables it may take more hours to reach the critical number.

Another limitation is that we have looked at cash game play only, while both online and offline poker are also frequently played in tournament form. This focus was deliberate, because the value of a given amount of chips wagered in a tournament hand depends on the phase of the tournament and on players' chip stack size relative to the chip stacks of their opponents. This issue greatly complicates the analysis of performance using hand-level data. A more straightforward approach for tournaments would be to analyze players' finishes, which is precisely what Croson, Fishman and Pope (2008) and Levitt and Miles (2014) do for major live events. It is hard to tell whether tournament poker requires more or less skill than the cash game variant, but we believe that a substantial difference is not very likely. In the early phases, tournaments are very similar to cash games, and so will be the roles of chance and skill. At later stages, chance increases in importance because the blind bets become larger relative to players' chip stacks, which effectively reduces the opportunities for strategic betting. However, for the same reason, meticulous hand selection (which dealt hands to play and which not) then becomes even more consequential. Furthermore, especially at later stages, players also need to factor in the prize money structure in their decisions. Future work could exploit the large amount of available tournament data and see if our speculation that skill similarly predominates after a few sessions of play indeed holds true.

Appendix 6: Deming Regression

Deming regression was first introduced by Adcock (1878). Kummell (1879) extended the method by allowing for the errors in the dependent and in the independent variable to have different variances (although the former was still assumed to be proportional to the latter for all observations). The method is named after the statistician W. Edwards Deming who propagated it (Deming, 1943). For our regression, we use the following model:

$$y_{i,2}^{*} = \alpha + \varphi y_{i,i}^{*} \tag{6A.1}$$

$$y_{i,t} = y_{i,t}^{*} + \varepsilon_{i,t} \tag{6A.2}$$

where $y_{i,t}^{*}$ are the true values and $y_{i,t}$ are the observed values for the performance of player i in period t ($t = 1{,}2$) and where $\varepsilon_{i,t}$is the measurement error that is normally distributed with a mean of zero.

While standard regression approaches minimize the sum of squares of residuals for the dependent variable ($e_{i,2}$) only, we here minimize the sum of squares of the standardized residuals for both the dependent and the independent variable:

$$SSR = \sum_{i=1}^{N} \left(\frac{e_{i,2}^{2}}{var(\varepsilon_{i,2})} + \frac{e_{i,i}^{2}}{var(\varepsilon_{i,1})} \right) \tag{6A.3}$$

This leads to the following objective function:

$$\min_{\widehat{\alpha,\varphi,\hat{y}_{1,1}^{*},\ldots,\hat{y}_{N,1}^{*}}} \sum_{i=1}^{N} \left[\frac{\left(y_{i,2} - \hat{\alpha} - \hat{\varphi}\hat{y}_{i,1}^{*}\right)^{2}}{var(\varepsilon_{i,2})} + \frac{\left(y_{i,1} - \hat{y}_{i,1}^{*}\right)^{2}}{var(\varepsilon_{i,1})} \right] \tag{6A.4}$$

As show in York (1966), the solution is given by:

$$\hat{\varphi} = \frac{\sum_{i=1}^{N} w_i z_i \tilde{y}_{i,2}}{\sum_{i=1}^{N} w_i z_i \tilde{y}_{i,1}} \tag{6A.5}$$

and

$$\hat{\alpha} = \bar{y}_{2}^{w} - \hat{\varphi}\bar{y}_{1}^{w} \tag{6A.6}$$

where

$$\tilde{y}_{i,t} = y_{i,t} - \bar{y}_t^w \tag{6A.7}$$

$$\bar{y}_t^w = \frac{\sum_{i=1}^{N} w_i y_{i,t}}{\sum_{i=1}^{N} w_i} \tag{6A.8}$$

$$w_i = \left(v_{i,2} + \hat{\varphi} v_{i,1}\right)^{-1} \tag{6A.9}$$

$$v_{i,t} = var\left(\varepsilon_{i,t}\right) \tag{6A.10}$$

$$z_i = w_i\left(v_{i,2} y_{i,1} + \hat{\varphi} v_{i,1} y_{i,2}\right) \tag{6A.11}$$

In most applications, the variance of the measurement error is not known at the level of individual observations. In order to solve the optimization problem it is then usually assumed that the ratio $var\left(\varepsilon_{i,2}\right)/var\left(\varepsilon_{i,1}\right)$ is the same for all i. We are in the unique situation that we do have accurate estimates, which allows us to obtain unbiased estimates of φ and α. When the performance of poker players is measured as the average number of big blinds won, the variance of the measurement error $var\left(\varepsilon_{i,t}\right)$ for any specific observation is approximated by $s_{i,t}^2/n_{i,t}$, or the ratio of the underlying sample variance of the number of big blinds won ($s_{i,t}^2$) and the number of hands ($n_{i,t}$). When our performance robustness measure is used, the measurement error variance is always equal to unity.

Chapter 7 | Conclusions

This thesis examines non-standard models of financial decision making through theoretical and empirical analysis.

Chapter 2 analyzes multiplier preferences, a popular model in macroeconomics and finance proposed by Hansen and Sargent (2001). This model allows for a deviation from expected utility (the standard model) due to a different treatment of 'ambiguous' events. People have a guess for the likelihood of these events occurring, but do not know the exact probability distribution.

In its original form, multiplier preferences only capture ambiguity aversion, where people prefer known probabilities to unknown ones. This is not a problem on a macroeconomic level, but on a micro level, a substantial proportion of people is often ambiguity seeking: they prefer unknown probabilities to known ones. We give a preference foundation for an extension of multiplier preferences, such that it can be used to explain ambiguity-seeking behavior.

We also show how extended multiplier preferences can be measured and thereby obtain a measure of ambiguity aversion that can easily be applied in empirical studies. A first application of this method on two large scale representative surveys (Netherlands & US) showed that a substantial fraction (around one third) of the population was indeed ambiguity seeking.

Chapter 3 looks at people's subjective beliefs with regards to the stock market, and how these compare to market beliefs. Many standard theoretical models assume homogeneous beliefs, in which case the two sets of beliefs would be the same (in expectation). By linking subjective response questions in a longitudinal survey to the probability distribution implicit in option prices, we are the first to be able to see how the people's beliefs co-move with those of the market. We find that there is a relationship between the two views, although the association is far from one-for-one. A closer association is found for those who demonstrate a better understanding of the laws of probability, which suggests that numeracy affects the accuracy of elicited responses.

Chapters 4 and 5 discuss the modeling of intertemporal preferences. It is generally considered normative to make decisions according to exponential or constant discounting, i.e. to display constant impatience over time. In practice, people are often found to be 'decreasingly impatient': people are very impatient in the short run, but less so as time progresses.

Chapter 4 discusses the most widely used model to explain this phenomenon: the quasi-hyperbolic or (β, δ) discounting model. Here, β describes a 'present bias', the factor with which people discount all future payments in addition to the regular discounting through the δ parameter. Since β is the cause of the decreasing impatience in this model, it is often used as a measure of the degree of decreasing impatience or dynamic inconsistency (changing one's mind over time). We show that this interpretation is incorrect, and that one should look β and δ together to get to such an index; more specifically, $\tau = \frac{\ln(\beta)}{\ln(\delta)}$ is a proper index. This calls for a rewriting of the model, from $(\beta\delta^t)$ to $\delta^{\tau+t}$.

The new index τ has a natural interpretation as a time penalty for any delay beyond the present. Because of this, it also indicates the period over which the decision maker is vulnerable to dynamic inconsistencies. A first illustration of the use of τ on an existing dataset shows that τ has a stronger correlation with demographic variables and less interference with δ than does β.

Chapter 5 comments on Doyle's (2013) survey of time preferences, in which he criticizes the Constant Relative Decreasing Impatience (CRDI) and Constant Absolute Decreasing Impatience families of Bleichrodt, Rohde & Wakker (2009). We show that his comments are based on incorrect assumptions. As explained by Bleichrodt et al., the CRDI family has the flexibility to capture all possible degrees of increasing and decreasing impatience. This sets it apart from any other currently popular discount family and warrants its widespread use.

Finally, Chapter 6 discusses the skill factor in online poker. Although the cards drawn depend purely on chance, behavioral research suggests that skill may vary well affect the outcomes of the game through the (re)actions of the players. By analyzing the outcomes of over 450 million games, we find that there is a significant skill factor and that this factor dominates the luck factor after a moderate duration of play. We find that players who performed well in an earlier period continue to do so in a latter period. By simulating the performance of a high- and low-skilled player, we find that the skill factor is greater than the luck factor in less than 1,500 games, or about 20 hours.

Samenvatting

Dit proefschrift beschrijft theoretisch en empirisch onderzoek naar atypisch financieel keuzegedrag.

Hoofdstuk 2 analyseert 'vermenigvuldingsvoorkeuren' (multiplier preferences), een veelgebruikt model in macro- en financieel-economisch onderzoek om gedrag bij meerduidigheid te verklaren. Hierbij kunnen mensen weliswaar een inschatting maken van de kansen op uitkomsten, maar weten ze niet de exacte kansverdeling. Theoretisch gezien zou alleen iemands inschatting van de kansverdeling moeten uitmaken, maar in de praktijk blijkt men vaak een voorkeur te hebben voor bekende kansen – situaties waarbij er minder meerduidigheid is. Zo hebben particuliere beleggers vaak een sterke voorkeur om te beleggen in eigen land, waar het vanuit het oogpunt van diversificatie juist beter is om in buitenlandse aandelen te beleggen.

Uit experimenteel onderzoek is echter gebleken dat een substantieel gedeelte van de mensen juist een voorkeur heeft voor meerduidigheid. Dit wordt niet toegestaan in het originele vermenigvuldingsvoorkeuren model van Hansen en Sargent (2001), wat voornamelijk gericht is op macro-economisch modelleren (daar vormt dit geen probleem). Via een nieuwe axiomatisering breiden wij het model dusdanig uit dat dit soort preferenties (vóór meerduidigheid) wel mogelijk is. Tevens laten wij zien hoe het model op een simpele en efficiënte wijze op individueel niveau kan worden geschat, en geven we de eerste micro-economische schatting van het model voor de Nederlandse en Amerikaanse bevolking. In beide groepen heeft een substantiële proportie een voorkeur voor meerduidigheid.

Hoofdstuk 3 bekijkt verwachtingen van mensen ten aanzien van bewegingen in de aandelenmarkt, en hoe deze verwachtingen zich verhouden tot de verwachtingen in de markt. In vele standaard theoretische modellen wordt uitgegaan van homogene verwachtingen, waardoor de twee (gemiddeld genomen) gelijk zouden moeten zijn. Doordat we een verbinding leggen tussen antwoorden op bepaalde vragen in een grote longitudinale enquête en de kansverdeling die uit optieprijzen kan worden gehaald, kunnen we als eerste kijken hoe deze twee (gezamenlijk) bewegen. Het blijkt dat er inderdaad een verband is tussen de twee types verwachtingen, maar de verhouding is verre van een-op-een. Voorts vinden we een sterker verband bij mensen met een goed begrip van kansen, wat suggereert dat gecijferdheid invloed heeft op de nauwkeurigheid van de gegeven antwoorden.

Hoofdstukken 4 en 5 gaan over tijdspreferenties. Bij optimaal gedrag moet men constant verdisconteren, d.w.z. de mate van ongeduld moet gelijk zijn over de tijd. In de praktijk zien we echter dat mensen vaak op de korte termijn erg ongeduldig zijn en op de langere termijn relatief geduldiger. We noemen dit ‘afnemend ongeduld’.

Hoofdstuk 4 gaat over het vaakst gebruikte model om dit fenomeen te beschrijven: het quasi-hyperbolische ofwel (β, δ) model. In dit model zit een ‘heden-effect’ doordat alle uitkomsten in de toekomst (maar niet die in het heden) naast het reguliere verdisconteren (via δ) extra worden verdisconteerd met een factor β. Omdat β de oorzaak is van het afnemend ongeduld wordt deze parameter vaak gezien als een maat hiervoor. Men vergeet dan echter dat kijken naar β niet voldoende is om de mate van afnemend ongeduld in te schatten, maar dat men moet kijken naar een combinatie van beta en delta. Wij tonen aan dat er gekeken moet worden naar $\tau = \frac{\ln(\beta)}{\ln(\delta)}$ voor de mate van afnemend ongeduld. Het model kan dan worden herschreven van $(\beta\delta^t)$ naar $\delta^{\tau+t}$. De nieuwe parameter τ heeft een natuurlijke interpretatie als een denkbeeldige vertraging die aan de wachttijd wordt toegevoegd. Hieruit blijkt ook dat de nieuwe parameter geschikt is om te zien in wat voor situaties er inconsistenties kunnen optreden. Een eerste empirische toepassing toont aan dat τ een sterkere correlatie heeft met demografische variabelen dan β, en dat τ minder gecorreleerd is met δ.

Hoofdstuk 5 geeft commentaar op Doyle’s (2013) overzicht van modellen van tijdspreferenties. In zijn overzicht geeft hij kritiek op de “Constant Relative Decreasing Impatience” (CRDI) en “Constant Absolute Decreasing Impatience” modellen van Bleichrodt, Rohde & Wakker (2009). Wij tonen aan dat zijn commentaar gebaseerd is op onjuiste aannames, en geven de voordelen aan van de besproken families. De CRDI familie kan – in tegenstelling tot de meeste bestaande modellen – alle maten van afnemend alsook toenemend ongeduld aan.

Tot slot bespreekt Hoofdstuk 6 de vaardigheidscomponent in online poker. Welke kaarten worden getrokken berust volledig op kans, maar gedragseconomisch onderzoek suggereert dat vaardigheid zijn intrede kan doen bij de strategische beslissingen die spelers maken. Door te kijken naar de consistentie van spelers' prestaties kunnen we de vaardigheidscomponent identificeren. Als een deel van de spelers consistent wint en een ander deel consistent verliest, kan dit alleen het gevolg zijn van verschillen in vaardigheid (systematische prestaties kunnen niet het gevolg zijn van kans). Wij analyseren ruim 450 miljoen spellen en vinden een substantiële vaardigheidscomponent. Spelers die het goed deden in een periode bleven bovengemiddeld presteren in een latere periode en verliezers bleven ondermaats presteren in opvolgende periodes. Uit simulaties blijkt dat de vaardigheidscomponent groter is dan de kanscomponent binnen 1.500 handen, ofwel circa 20 uur.

Bibliography

Aczél, János. 2005. "Utility of Extension of Functional equations—when Possible" *Journal of Mathematical Psychology*, 49(6): 445-449.

Aczél, Joseph. 1966. *Lectures on Functional Equations and their Applications*. New York, NY: Academic Press.

Adcock, Robert J. 1878. "A Problem in Least Squares" *The Analyst*, 5(2): 53-54.

Ainslie, George. 1975. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control" *Psychological Bulletin*, 82(4): 463-496.

Albert, James H. and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data" *Journal of the American Statistical Association*, 88(322): 669-679.

Allais, Maurice. 1953. "Le Comportement De l'Homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes De l'École Américaine" *Econometrica*, 21(4): 503-546.

Alpert, Marc and Howard Raiffa. 1982. "A Progress Report on the Training of Probability Assessors." In *Judgment Under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky, 294-305. New York: Cambridge University Press.

Anderson, Evan W., Eric Ghysels, and Jennifer L. Juergens. 2005. "Do Heterogeneous Beliefs Matter for Asset Pricing?" *Review of Financial Studies*, 18(3): 875-924.

Bacchetta, Philippe, Elmar Mertens, and Eric Van Wincoop. 2009. "Predictability in Financial Markets: What do Survey Expectations Tell Us?" *Journal of International Money and Finance*, 28(3): 406-426.

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(2): 773-806.

Bassett, William F. and Robin L. Lumsdaine. 2000. "Probability Limits: are Subjective Assessments Adequately Accurate?" *Journal of Human Resources*, 36(2): 327-363.

Benartzi, Shlomo and Richard H. Thaler. 1993. "Myopic Loss Aversion and the Equity Premium Puzzle" *Quarterly Journal of Economics*, 110(1): 75-92.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. 2013. "Managerial Miscalibration" *The Quarterly Journal of Economics*, 128(4): 1547-1584.

Bernanke, Ben. "Dedication of the New Federal Reserve Bank of Kansas City Building" *Speech Delivered June 12th 2008*, available at http://www.federalreserve.gov/newsevents/speech/bernanke20080612a.htm, accessed 12/30/2011.

Bernheim, B. D. and Daniel M. Garrett. 2003. "The Effects of Financial Education in the Workplace: Evidence from a Survey of Households" *Journal of Public Economics*, 87(7–8): 1487-1519.

Black, Fischer and Myron S. Scholes. 1973. "The Pricing of Options and Corporate Liabilities" *Journal of Political Economy*, 81(3): 637-654.

Bleichrodt, Han, Rogier J.D. Potter van Loon, Kirsten I. M. Rohde, and Peter P. Wakker. 2013. "A Criticism of Doyle's Survey of Time Preference: A Correction regarding the CRDI and CADI Families" *Judgment and Decision Making*, 8(5): 630-631.

Bleichrodt, Han, Kirsten I. Rohde, and Peter P. Wakker. 2009. "Non-Hyperbolic Time Inconsistency" *Games and Economic Behavior*, 66(1): 27-38.

Borm, Peter and Ben van der Genugten. 2001. "On a Relative Measure of Skill for Games with Chance Elements" *Top*, 9(1): 91-114.

Bossaerts, Peter, Paolo Ghirardato, Serena Guarnaschelli, and William R. Zame. 2010. "Ambiguity in Asset Markets: Theory and Experiment" *Review of Financial Studies*, 23(4): 1325-1359.

Breeden, Douglas T. and Robert H. Litzenberger. 1978. "Prices of State-Contingent Claims Implicit in Option Prices" *The Journal of Business*, 51(4): 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(2): 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(4): 713-723.

Cabot, Anthony and Robert Hannum. 2005. "Poker: Public Policy, Law, Mathematics, and the Future of an American Tradition" *Thomas M. Cooley Law Review*, 22(2): 443-514.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games" *Quarterly Journal of Economics*, 119(3): 861-898.

Cao, H. H., Tan Wang, and Harold H. Zhang. 2005. "Model Uncertainty, Limited Market Participation, and Asset Prices" *Review of Financial Studies*, 18(4): 1219-1251.

Cesarini, David, Magnus Johannesson, Patrik K. Magnusson, and Björn Wallace. 2012. "The Behavioral Genetics of Behavioral Anomalies" *Management Science*, 58(1): 21-34.

Cochrane, John H. 2011. "Presidential Address: Discount Rates" *Journal of Finance*, 66(4): 1047-1108.

Coval, Joshua D., David A. Hirshleifer, and Tyler Shumway. 2005. "Can Individual Investors Beat the Market?" *Harvard University Working Paper no. 02-45*.

Cronqvist, Henrik and Stephan Siegel. 2014. "The Genetics of Investment Biases" *Journal of Financial Economics*, 113(2): 215-234.

Croson, Rachel, Peter Fishman, and Devin G. Pope. 2008. "Poker Superstars: Skill Or Luck?" *Chance*, 21(4): 25-28.

Croson, Rachel and Uri Gneezy. 2009. "Gender Differences in Preferences" *Journal of Economic Literature*, 47(2): 448-474.

Cui, Changrong and David Frank. 2011. "Equity Implied Volatility Surface Computation" *Bloomberg Document no. 2056700*.

De Long, James B., Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets" *Journal of Political Economy*, 98(4): 703-738.

Debreu, Gerard and Tjalling C. Koopmans. 1982. "Additively Decomposed Quasiconvex Functions" *Mathematical Programming*, 24(1): 1-38.

DeDonno, Michael A. and Douglas K. Detterman. 2008. "Poker is a Skill" *Gaming Law Review*, 12(1): 31-36.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field" *Journal of Economic Literature*, 47(2): 315-372.

Deming, William E. 1944. *Statistical Adjustment of Data*. New York: Wiley & Sons.

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(366): 427-431.

Dimmock, Stephen G., Roy Kouwenberg, and Peter P. Wakker. 2013. "Ambiguity Attitudes in a Large Representative Sample" *Mimeo*.

Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2013. "Ambiguity Aversion and Household Portfolio Choice: Empirical Evidence" *National Bureau of Economic Research Working Paper Series no. 18743*.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100(3): 1238-1260.

Dominitz, Jeff and Charles F. Manski. 1997. "Perceptions of Economic Insecurity" *Public Opinion Quarterly*, 61(2): 261-287.

Dominitz, Jeff and Charles F. Manski. 2011. "Measuring and Interpreting Expectations of Equity Returns" *Journal of Applied Econometrics*, 26(3): 352-370.

Dow, James and S. R. d. C. Werlang. 1992. "Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio" *Econometrica*, 60(1): 197-204.

Doyle, John R. 2013. "Survey of Time Preference, Delay Discounting Models" *Judgment and Decision Making*, 8(2): 116-135.

Dreef, Marcel, Peter Borm, and Ben Van Der Genugten. 2003. "On Strategy and Relative Skill in Poker" *International Game Theory Review*, 5(2): 83-103.

Dreef, Marcel, Peter Borm, and Ben Van der Genugten. 2004. "Measuring Skill in Games: Several Approaches Discussed" *Mathematical Methods of Operations Research*, 59(3): 375-391.

Dreef, Marcel, Peter Borm, and Ben van der Genugten. 2004. "A New Relative Skill Measure for Games with Chance Elements" *Managerial and Decision Economics*, 25(5): 255-264.

Dupuis, Paul and Richard S. Ellis. 2011. *A Weak Convergence Approach to the Theory of Large Deviations*. New York: John Wiley & Sons.

Easley, David and Maureen O'Hara. 2009. "Ambiguity and Nonparticipation: The Role of Regulation" *Review of Financial Studies*, 22(5): 1817-1843.

Eil, David and Jaimie W. Lien. 2014. "Staying Ahead and Getting Even: Risk Attitudes of Experienced Poker Players" *Games and Economic Behavior*, 87: 50-69.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms" *Quarterly Journal of Economics*, 75(4): 643-669.

Epstein, Larry G. and Martin Schneider. 2010. "Ambiguity and Asset Markets" *Annual Review of Financial Economics*, 2: 315-346.

Fischhoff, Baruch and Wändi Bruine de Bruin. 1999. "Fifty-Fifty = 50%" *Journal of Behavioral Decision Making*, 12(2): 149-163.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein. 1977. "Knowing with Certainty: The Appropriateness of Extreme Confidence" *Journal of Experimental Psychology: Human Perception and Performance*, 3(4): 552-564.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making" *Journal of Economic Perspectives*, 19(4): 25-42.

Friedman, Lawrence. 1971. "Optimal Bluffing Strategies in Poker" *Management Science*, 17(12): 764-771.

Graham, John R. and Campbell R. Harvey. 2001. "Expectations of Equity Risk Premia, Volatility and Asymmetry from a Corporate Finance Perspective" *National Bureau of Economic Research Working Paper Series no. 8678*.

Graham, John R. and Campbell R. Harvey. 2007. "The Equity Risk Premium in January 2007: Evidence from the Global CFO Survey" *ICFAI Journal of Financial Risk Management*, 4(2): 46-61.

Grant, Simon and Ben Polak. 2013. "Mean-Dispersion Preferences and Constant Absolute Uncertainty Aversion" *Journal of Economic Theory*, 148(4): 1361-1398.

Grant, Simon, Ben Polak, and Tomasz Strzalecki. 2009. "Second-Order Expected Utility" *Mimeo*.

Greenwood, Robin and Andrei Shleifer. 2014. "Expectations of Returns and Expected Returns" *Review of Financial Studies*, 27(3): 714-746.

Grinblatt, Mark, Matti Keloharju, and Juhani T. Linnainmaa. 2012. "IQ, Trading Behavior, and Performance" *Journal of Financial Economics*, 104(2): 339-362.

Grohman, Christopher. 2006. "Reconsidering Regulation: A Historical View of the Legality of Internet Poker and Discussion of the Internet Gambling Ban of 2006" *Journal of Legal Technology Risk Management*, 1(1): 34-74.

Hansen, Lars P. and Thomas J. Sargent. 2001. "Robust Control and Model Uncertainty" *American Economic Review*, 91(2): 60-66.

Haruvy, Ernan, Dale O. Stahl, and Paul W. Wilson. 2001. "Modeling and Testing for Heterogeneity in Observed Strategic Behavior" *Review of Economics and Statistics*, 83(1): 146-157.

Heath, Chip and Amos Tversky. 1991. "Preference and Belief: Ambiguity and Competence in Choice Under Uncertainty" *Journal of Risk and Uncertainty*, 4(1): 5-28.

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(1): 15-51.

Hendrickx, Ruud, Peter Borm, Ben Van der Genugten, and Pim Hilbers. 2008. "Measuring Skill in More-Person Games with Applications to Poker" *CentER Discussion Paper Series no. 2008-106*.

Heubeck, Steven. 2008. "Measuring Skill in Games: A Critical Review of Methodologies" *Gaming Law Review and Economics*, 12(3): 231-238.

Hirshleifer, David. 2001. "Investor Psychology and Asset Pricing" *The Journal of Finance*, 56(4): 1533-1597.

Hoffmann, Arvid O., Thomas Post, and Joost M. Pennings. 2013. "Individual Investor Perceptions and Behavior during the Financial Crisis" *Journal of Banking & Finance*, 37(1): 60-74.

Hoffmann, Arvid O., Thomas Post, and Joost M. Pennings. 2014. "How Investor Perceptions Drive Actual Trading and Risk-Taking Behavior" *Journal of Behavioral Finance, Forthcoming*.

Hudomiet, Peter, Gábor Kézdi, and Robert J. Willis. 2011. "Stock Market Crash and Expectations of American Households" *Journal of Applied Econometrics*, 26(3): 393-415.

Hull, John C. 1989. *Options, Futures and Other Derivatives*, 4th ed. Upper Saddle, New Jersey: 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*, ed. David A. Wise, 259-309. Chicago: University of Chicago Press.

Hurd, Michael D. and Kathleen McGarry. 2002. "The Predictive Validity of Subjective Probabilities of Survival" *The Economic Journal*, 112(482): 966-985.

Hurd, Michael D. and Susann Rohwedder. 2010. "Effects of the Financial Crisis and Great Recession on American Households" *National Bureau of Economic Research Working Paper Series no. 16407*.

Hurd, Michael D. and Susann Rohwedder. 2012. "Stock Price Expectations and Stock Trading" *National Bureau of Economic Research Working Paper Series no. 17973*.

Huynh, Kim P. and Juergen Jung. 2010. "Correcting Focal Point Biases in Subjective Health Expectations: An Application to the RAND-HRS Data" *Towson University Department of Economics Working Paper no. 2010-08*.

Kadane, Joseph B. 1986. "Does Electronic Draw Poker Require Skill to Play?" In *Statistics and the Law*, ed. Morris H. DeGroot, Stephen E. Fienberg, and Joseph B. Kadane, 333-343. New York: John Wiley & Sons.

Kahneman, Daniel and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness" *Cognitive Psychology*, 3(3): 25-48.

Kahneman, Daniel and Amos Tversky. 1973. "On the Psychology of Prediction." *Psychological Review*, 80(4): 237-251.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk" *Econometrica*, 47(2): 263-292.

Kameda, Tatsuya and James H. Davis. 1990. "The Function of the Reference Point in Individual and Group Risk Decision Making" *Organizational Behavior and Human Decision Processes*, 46(1): 55-76.

Kelly, Joseph M., Zeeshan Dhar, and Thibault Verbiest. 2007. "Poker and the Law: Is it a Game of Skill Or Chance and Legally does it Matter?" *Gaming Law Review*, 11(3): 190-202.

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

Kézdi, Gábor and Robert J. Willis. 2011. "Household Stock Market Beliefs and Learning" *National Bureau of Economic Research Working Paper Series no. 17614*.

Kleinjans, Kristin J. and Arthur van Soest. 2014. "Rounding Focal Point Answers and Nonresponse to Subjective Probability Quesitons" *Journal of Applied Econometrics*, 29(4): 567-585.

Krantz, David H., R. D. Luce, Patrick Suppes, and Amos Tversky. 1971. *Foundations of Measurement: Vol. 1: Additive and Polynomial Representations*. New York: Academic Press.

Kuhn, Harold W. 1950. "A Simplified Two-Person Poker." In *Contributions to the Theory of Games*, ed. Harold W. Kuhn and Albert W. Tucker. Princeton: Princeton University Press.

Kummell, Charles H. 1879. "Reduction of Observation Equations which Contain More than One Observed Quantity" *The Analyst*, 6(4): 97-105.

Langer, Ellen J. 1975. "The Illusion of Control." *Journal of Personality and Social Psychology*, 32(2): 311-328.

Langer, Ellen and Jane Roth. 1975. "Heads I Win, Tails it's Chance: The Illusion of Control as a Function of the Sequence of Outcomes in a Purely Chance Task" *Journal of Personality and Social Psychology*, 32(6): 951-955.

Larkey, Patrick, Joseph B. Kadane, Robert Austin, and Shmuel Zamir. 1997. "Skill in Games" *Management Science*, 43(5): 596-609.

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

Levitt, Steven D. and Thomas J. Miles. 2014. "The Role of Skill Versus Luck in Poker Evidence from the World Series of Poker" *Journal of Sports Economics*, 15(1): 31-44.

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. 2012. "Numeracy, Financial Literacy and Financial Decision-Making" *Numeracy*, 5(1).

Lusardi, Annamaria and Olivia S. Mitchell. 2011. "Financial Literacy and Retirement Planning in the United States" *Journal of Pension Economics and Finance*, 10(4): 509-525.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006. "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences" *Econometrica*, 74(6): 1447-1498.

Mankiw, N. G., Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." In *NBER Macroeconomics Annual 2003, Volume 18*, ed. Mark Gertler and Kenneth Rogoff, 209-270. Cambrige, Massachussetts: The MIT Press.

Manski, Charles F. and Francesca Molinari. 2010. "Rounding Probabilistic Expectations in Surveys" *Journal of Business and Economic Statistics*, 28(2): 219-231.

Manski, Charles F. 2004. "Measuring Expectations" *Econometrica*, 72(5): 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.

Merton, Robert C. 1973. "Theory of Rational Option Pricing" *Bell Journal of Economics and Management*, 4(1): 141-183.

Nash, John F. and Lloyd S. Shapley. 1996. "A Simple Three-Person Poker Game." In *Contributions to the Theory of Games*, ed. Harold W. Kuhn and Albert W. Tucker. Princeton: Princeton University Press.

Neilson, William S. 2010. "A Simplified Axiomatic Approach to Ambiguity Aversion" *Journal of Risk and Uncertainty*, 41(2): 113-124.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize their Losses?" *The Journal of Finance*, 53(5): 1775-1798.

Palacios-Huerta, Ignacio and Oscar Volij. 2009. "Field Centipedes" *American Economic Review*, 99(4): 1619-1635.

Paolina, Philip. 2001. "Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables" *Political Analysis*, 9(4): 325-346.

Park, Young J. and Luís Santos-Pinto. 2010. "Overconfidence in Tournaments: Evidence from the Field" *Theory and Decision*, 69(1): 143-166.

Post, Thierry, Van den Assem, Martijn J, Guido Baltussen, and Richard H. Thaler. 2008. "Deal Or no Deal? Decision Making Under Risk in a Large-Payoff Game show" *American Economic Review*, 98(1): 38-71.

Poterba, James M. and Andrew A. Samwick. 1995. "Stock Ownership Patterns, Stock Market Fluctuations, and Consumption" *Brookings Papers on Economic Activity*, 26(2): 295-372.

Pratt, John W. 1964. "Risk Aversion in the Small and in the Large" *Econometrica*, 32(1-2): 122-136.

Prelec, Drazen. 1989. "Decreasing Impatience: Definition and Consequences" *Harvard Business School Working Paper*.

Prelec, Drazen. 2004. "Decreasing Impatience: A Criterion for Non-stationary Time Preference and “Hyperbolic” Discounting" *The Scandinavian Journal of Economics*, 106(3): 511-532.

Rabin, Matthew. 2002. "Inference by Believers in the Law of Small Numbers" *Quarterly Journal of Economics*, 117(3): 775-816.

Rich, Robert, Joseph Song, and Joseph Tracy. 2012. "The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters" *Federal Reserve Bank of New York Staff Report no. 588*.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York, NY: Wiley.

Schmeidler, David. 1989. "Subjective Probability and Expected Utility without Additivity" *Econometrica*, 57(3): 571-587.

Schwarz, Gideon E. 1978. "Estimating the Dimension of a Model" *The Annals of Statistics*, 6(2): 461-464.

Shiller, Robert J. 2003. "From Efficient Markets Theory to Behavioral Finance" *Journal of Economic Perspectives*, 17(1): 83-104.

Siler, Kyle. 2010. "Social and Psychological Challenges of Poker" *Journal of Gambling Studies*, 26(3): 401-420.

Smith, Gary, Michael Levere, and Robert Kurtzman. 2009. "Poker Player Behavior After Big Wins and Big Losses" *Management Science*, 55(9): 1547-1555.

Stahl, Dale O. and Paul W. Wilson. 1994. "Experimental Evidence on Players' Models of Other Players" *Journal of Economic Behavior & Organization*, 25(3): 309-327.

Stanovich, Keith E. and Richard F. West. 1998. "Cognitive Ability and Variation in Selection Task Performance" *Thinking & Reasoning*, 4(3): 193-230.

Strzalecki, Tomasz. 2011. "Axiomatic Foundations of Multiplier Preferences" *Econometrica*, 79(1): 47-73.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rüetzler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior" *American Economic Review*, 103(1): 510-531.

Svenson, Ola. 1981. "Are we all Less Risky and More Skillful than our Fellow Drivers?" *Acta Psychologica*, 47(2): 143-148.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam" *American Economic Review*, 100(1): 557-571.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice" *Journal of Economic Behavior & Organization*, 1(1): 39-60.

Thaler, Richard H. and Eric J. Johnson. 1990. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice" *Management Science*, 36(6): 643-660.

Tselnik, Michael A. 2006. "Check, Raise, Or Fold: Poker and the Unlawful Internet Gambling Enforcement Act" *Hofstra Law Review*, 35(3): 1617-1670.

Tversky, Amos and Craig R. Fox. 1995. "Weighing Risk and Uncertainty" *Psychological Review*, 102(2): 269-283.

Tversky, Amos and Daniel Kahneman. 1971. "Belief in the Law of Small Numbers" *Psychological Bulletin*, 76(2): 105-110.

Tversky, Amos and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability" *Cognitive Psychology*, 5(2): 207-232.

Tversky, Amos and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty" *Journal of Risk and Uncertainty*, 5(4): 297-323.

Villegas, Carlos. 1964. "On Qualitative Probability σ-Algebras" *Annals of Mathematical Statistics*, 35(4): 1787-1796.

Von Neumann, John and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Wakker, Peter P. 2008. "Explaining the Characteristics of the Power (CRRA) Utility Family" *Health Economics*, 17(12): 1329-1344.

Wakker, Peter P. 1989. *Additive Representations of Preferences, a New Foundation of Decision Analysis*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Wakker, Peter P. 2010. *Prospect Theory for Risk and Ambiguity*. Cambridge, UK: Cambridge University Press.

Watts, Simon. 2010. "OVME <GO> Userguide" *Bloomberg Document no. 2052774*.

Weinstein, Neil D. 1980. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology*, 39(5): 806-820.

Winkler, Robert L. 1967. "The Assessment of Prior Distribution in Bayesian Analysis" *Journal of the American Statistical Association*, 62(319): 776-800.

York, Derek. 1966. "Least-Squares Fitting of a Straight Line" *Canadian Journal of Physics*, 44(5): 1079-1086.

About the author

Rogier Jan Dave Potter van Loon (1987) holds a (cum laude) BSc in Econometrics and Operations Research from Erasmus University Rotterdam, and an MPhil in Economics from the Tinbergen Institute. He worked on his PhD thesis under the supervision of Han Bleichrodt and Peter Wakker. As part of his PhD training, Rogier spent time as a visiting researcher at the American University in Washington, D.C. Currently, he works as an Assistant Professor at the Econometric Institute in Rotterdam, where he studies topics on Behavioral Finance.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

551 X. ZHANG, *Modeling Time Variation in Systemic Risk*
552 H.R.A. KOSTER, *The internal structure of cities: the economics of agglomeration, amenities and accessibility.*
553 S.P.T. GROOT, *Agglomeration, globalization and regional labor markets: micro evidence for the Netherlands.*
554 J.L. MÖHLMANN, *Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products*
555 S.M. HOOGENDOORN, *Diversity and Team Performance: A Series of Field Experiments*
556 C.L. BEHRENS, *Product differentiation in aviation passenger markets: The impact of demand heterogeneity on competition*
557 G. SMRKOLJ, *Dynamic Models of Research and Development*
558 S. PEER, *The economics of trip scheduling, travel time variability and traffic information*
559 V. SPINU, *Nonadditive Beliefs: From Measurement to Extensions*
560 S.P. KASTORYANO, *Essays in Applied Dynamic Microeconometrics*
561 M. VAN DUIJN, *Location, choice, cultural heritage and house prices*
562 T. SALIMANS, *Essays in Likelihood-Based Computational Econometrics*
563 P. SUN, *Tail Risk of Equity Returns*
564 C.G.J. KARSTEN, *The Law and Finance of M&A Contracts*
565 C. OZGEN, *Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth*
566 R.S. SCHOLTE, *The interplay between early-life conditions, major events and health later in life*
567 B.N.KRAMER, *Why don't they take a card? Essays on the demand for micro health insurance*
568 M. KILIÇ, *Fundamental Insights in Power Futures Prices*
569 A.G.B. DE VRIES, *Venture Capital: Relations with the Economy and Intellectual Property*
570 E.M.F. VAN DEN BROEK, *Keeping up Appearances*
571 K.T. MOORE, *A Tale of Risk: Essays on Financial Extremes*
572 F.T. ZOUTMAN, *A Symphony of Redistributive Instruments*
573 M.J. GERRITSE, *Policy Competition and the Spatial Economy*
574 A. OPSCHOOR, *Understanding Financial Market Volatility*
575 R.R. VAN LOON, *Tourism and the Economic Valuation of Cultural Heritage*
576 I.L. LYUBIMOV, *Essays on Political Economy and Economic Development*
577 A.A.F. GERRITSEN, *Essays in Optimal Government Policy*
578 M.L. SCHOLTUS, *The Impact of High-Frequency Trading on Financial Markets*

579 E. RAVIV, *Forecasting Financial and Macroeconomic Variables: Shrinkage, Dimension reduction, and Aggregation*
580 J. TICHEM, *Altruism, Conformism, and Incentives in the Workplace*
581 E.S. HENDRIKS, *Essays in Law and Economics*
582 X. SHEN, *Essays on Empirical Asset Pricing*
583 L.T. GATAREK, *Econometric Contributions to Financial Trading, Hedging and Risk Measurement*
584 X. LI, *Temporary Price Deviation, Limited Attention and Information Acquisition in the Stock Market*
585 Y. DAI, *Efficiency in Corporate Takeovers*
586 S.L. VAN DER STER, *Approximate feasibility in real-time scheduling: Speeding up in order to meet deadlines*
587 A. SELIM, *An Examination of Uncertainty from a Psychological and Economic Viewpoint*
588 B.Z. YUESHEN, *Frictions in Modern Financial Markets and the Implications for Market Quality*
589 D. VAN DOLDER, *Game Shows, Gambles, and Economic Behavior*
590 S.P. CEYHAN, *Essays on Bayesian Analysis of Time Varying Economic Patterns*
591 S. RENES, *Never the Single Measure*
592 D.L. IN 'T VELD, *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*
593 Y.YANG, *Laboratory Tests of Theories of Strategic Interaction*
594 M.P. WOJTOWICZ, *Pricing Credits Derivatives and Credit Securitization*
595 R.S. SAYAG, *Communication and Learning in Decision Making*
596 S.L. BLAUW, *Well-to-do or doing well? Empirical studies of wellbeing and development*
597 T.A. MAKAREWICZ, *Learning to Forecast: Genetic Algorithms and Experiments*
598 P. ROBALO, *Understanding Political Behavior: Essays in Experimental Political Economy*
599 R. ZOUTENBIER, *Work Motivation and Incentives in the Public Sector*
600 M.B.W. KOBUS, *Economic Studies on Public Facility use*