Overview

• from the 1950s to the 1990s, finance research was dominated by the “traditional” finance paradigm

• this framework assumes that:
  – individuals have rational beliefs (update their beliefs according to Bayes’ rule when new information arrives)
  – and make decisions according to Expected Utility (with an increasing, concave utility function defined over consumption outcomes)

• starting in the 1990s, a new paradigm emerged: behavioral finance

• this field tries to make sense of the behavior of investors, markets, and firms using models that are psychologically more realistic than their predecessors

• behavioral finance models aim for psychological realism along three dimensions
  – allow for less than fully rational beliefs
  – use more realistic preferences
  – take account of cognitive limits
Overview, ctd.

• the emergence of behavioral finance in the 1990s was primarily due to three factors
  – a growing sense that many important facts were not easily understood in the traditional framework
  – a response to the “arbitrage critique”
  – major developments in an area of psychology known as “judgment and decision-making”

• the field is ambitious in scope
  – offers a new way of thinking about many fundamental topics in finance

• in this part of the course, we discuss applications to investor behavior and asset prices
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Course structure

I. Introduction
  • overview (LN 1)

II. Background
  • empirical facts (LN 2)
  • limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Roadmap

Asset prices

- aggregate stock market
- cross-section of stock returns
- other asset classes
- bubbles

Investor trading and portfolio choice

- individual investor behavior
Aggregate stock market

The volatility puzzle

• it is challenging to explain stock market volatility in a model with fully rational investors
  – e.g. in a model with rationally-varying forecasts of future cash flows (Shiller, 1981), interest rates, or risk

• rational approaches include:
  – habit preferences, long-run risk, rare disasters, and learning
Aggregate stock market, ctd.

The predictability puzzle

- excess aggregate stock market returns are predictable in the time series
  - e.g. by the price-dividend or price-earnings ratio

- this is hard to explain based on rationally-varying forecasts of interest rates or risk

- rational approaches include:
  - habit preferences, long-run risk, rare disasters, and learning
Aggregate stock market, ctd.

*The equity premium puzzle*

- the historical equity premium is much higher than predicted by a simple rational, frictionless model with power utility preferences
  - Mehra and Prescott (1985)

- rational approaches include:
  - habit preferences, rare disasters
The cross-section of stock returns

• evidence that firm characteristics predict stock returns in the cross-section
  – e.g. stocks with low values of characteristic $F$ have higher average returns than stocks with high values of characteristic $F$

• in a rational, frictionless model, the main approach to understanding this evidence is based on risk
  – e.g. beta

• evidence below known as “anomalies” because it is not explained by beta
Some important return predictors:

- past return
  - long-term past return (-)
  - medium-term past return (+)
- price-to-fundamentals ratio (-)
- issuance (-)
- earnings surprise (+)
- idiosyncratic volatility (-)
Other asset classes

Note:

• we have described the aggregate and cross-sectional patterns in the context of the stock market

• an important finding of recent years is that many of these patterns are present in other asset classes as well

• the excess volatility and time-series predictability in the aggregate stock market are also present in other major asset classes
  – real estate, long-term bonds

• several of the empirical patterns in the cross-section of stock returns also hold in other asset classes
  – e.g. momentum, long-term reversals, volatility

• this suggests a common mechanism that applies across asset classes
  – potentially good news for behavioral finance
Bubbles

One definition:

• a bubble is an episode in which an asset becomes significantly overvalued for some period of time
  – its price is higher than a reasonable present value of its future cash flows
  – or, its price is higher that it would be in an economy with fully rational investors

• this definition is conceptually sound, but can be hard to work with

Another, empirically-based definition:

• a bubble is an episode in which:
  – the price of an asset rises sharply over some period of time and then collapses
  – during the price rise, there is much talk of overvaluation in the media and among investors

• also, some of the following are observed:
  – very high trading volume
  – extrapolative expectations
  – sophisticated investors “riding the bubble”
  – good fundamental news near the start of the price rise (Kindleberger, 1978)
Bubbles, ctd.

Motivation:

- bubbles tend to be accompanied by very high trading volume (Hong and Stein, 2007)
- and sophisticated traders often ride the bubble (Brunnermeier and Nagel, 2004)
- understanding bubbles is an important challenge — their collapse can trigger economic downturns

Rational approach: “rational bubbles”
Investor trading and portfolio choice

- we focus primarily on the behavior of individual investors
  - we know more about them, and behavioral finance ideas may be more relevant to them

Individual investor behavior

- non-participation
- buying high / selling low
- under-diversification
  - home bias, local bias, concentrated holdings, own-company stock holdings
- preference for active management
- poor stock-picking performance
- selling behavior: the disposition effect
- buying behavior: buying of long-term past winner stocks
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Overview

• behavioral finance applications to asset prices often posit that irrational investors affect prices

• there is a classic critique of this idea
  – the “arbitrage critique”

• according to this critique, irrational investors cannot affect prices for any significant amount of time
  – as soon as irrational investors move prices, this creates an attractive opportunity for rational investors
  – the rational investors trade against the mispricing, quickly correcting it (“arbitrage”)

• a major achievement of behavioral finance is to push back against the arbitrage critique
  – i.e. to show that there are “limits to arbitrage”
Overview, ctd.

Terminology:

- the “fundamental value” of an asset is its price in an economy with rational investors and no frictions
  - the price that properly reflects all available public information
  - the efficient markets price
- in an economy with frictions, or where some people are not fully rational, an asset’s price may depart from fundamental value
  - this is a “mispricing”
  - or an “inefficiency”
- rational investors are sometimes referred to as “arbitrageurs”
- less than fully rational investors are sometimes referred to as “noise traders”
Limits to arbitrage: Theory

What is the response to the arbitrage critique?

• the critique says that it will be easy for rational investors to correct a mispricing

• in reality, however, it is not easy
  – there are risks and costs that limit arbitrageurs’ ability to correct a mispricing
  – this allows irrational investors to affect prices significantly and for a long time

• specific limits to arbitrage
  – risks: fundamental risk, noise trader risk
  – costs: trading costs, implementation costs
Limits to arbitrage: Theory, ctd.

Fundamental risk

- the risk that there will be adverse news about the fundamental value of the mispriced asset

Noise trader risk (De Long et al., 1990; Shleifer and Vishny, 1997)

- the risk that, as a result of the mispricing worsening in the short run, the arbitrageur is forced to close out his trade at a loss
- this risk arises because real-world arbitrageurs manage other people’s money
  - if the mispricing worsens in the short run, nervous clients may withdraw from the arbitrageur’s fund, forcing a liquidation
- the use of leverage amplifies this problem
  - if the mispricing worsens in the short run, banks may call their loans, again forcing a liquidation
Limits to arbitrage: Theory, ctd.

Costs

- general trading costs, but also:
  - short-selling costs
  - the cost of detecting, understanding, and exploiting a mispricing

Note:

- we have learnt a lot by studying specific empirical phenomena that are widely viewed as mispricings
  - twin shares
  - equity carve-outs (Mitchell, Pulvino, Stafford, 2002)
  - index inclusions (Shleifer, 1986)

- these demonstrate that there are limits to arbitrage
  - and help us understand which limits are more relevant in which settings
Summary

• the research on limits to arbitrage has been influential
  – there is now wide agreement among academics (and practitioners) that arbitrage is limited
  – albeit some disagreement as to how limited it is
• this success was one reason why behavioral finance “took off” in the 1990s
• still, we should not be complacent
  – whenever we argue that irrational investors affect prices, we should ask what the limits to arbitrage are
Course structure

I. Introduction
   • overview (LN 1)

II. Background
   • empirical facts (LN 2)
   • limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Overview

• behavioral finance models aim for psychological realism along three dimensions
  – allow for less than fully rational beliefs
  – use more realistic preferences
  – take account of cognitive limits

• in Lecture Notes 4 and 5, we focus on the first dimension: investor beliefs
  – Lecture Note 4: extrapolative beliefs
  – Lecture Note 5: overconfidence and other belief biases
Overview, ctd.

• over-extrapolation of fundamentals or returns is one of the most important and widely-applied ideas in behavioral finance
  – the idea that, when people form beliefs about future returns or cash-flow growth, they put too much weight on recent past returns or cash-flow growth

Roadmap

• return extrapolation
  – intuition
  – application: aggregate stock market
  – application: bubbles
  – sources of return extrapolation

• cash-flow extrapolation

• experience effects
Return extrapolation

• we start with *return* extrapolation
  – the idea that some investors form beliefs about the future returns of an asset, asset class, or fund by extrapolating its past returns

History

• several references in classic qualitative accounts
  – Bagehot (1873), Galbraith (1954)

• first wave of research on return extrapolation
  – Cutler, Poterba, Summers (1990), De Long et al. (1990), Hong and Stein (1999), Barberis and Shleifer (2003)

• new wave of research on return extrapolation
Return extrapolation, ctd.

- a catalyst for the new wave of research is the survey data on the expectations of real-world investors about future asset returns
- investor expectations of future stock market returns are a positive function of past returns

- the data point to over-extrapolation
  - investor expectations are *negatively* correlated with subsequent realized returns
Return extrapolation, ctd.

Several important applications:

- aggregate stock market
  - excess volatility, predictability
- bubbles
  - high prices *and* high volume
- cross-section of stock returns
  - momentum, long-term reversals, value premium

Note:

- the above patterns are present in many asset classes, suggesting a single underlying mechanism
  - return extrapolation is a simple candidate mechanism
Return extrapolation: Intuition

• suppose that some investors in the economy form beliefs as follows:

\[ E_t^e(P_{t+1} - P_t) = (1 - \theta)((P_{t-1} - P_{t-2}) + \theta(P_{t-2} - P_{t-3}) + \theta^2(P_{t-3} - P_{t-4}) + \theta^3(P_{t-4} - P_{t-5}) + \ldots) \]

where \(0 < \theta < 1\)

• in an economy with such investors, we are likely to observe:
  – excess volatility, predictability
  – momentum, long-term reversals, a value premium
  – bubbles

• extrapolators have reasonable returns at some points in the cycle, but do poorly overall
Return extrapolation: Aggregate market

- Barberis, Greenwood, Jin, Shleifer (2015) study a model in which some investors form beliefs about future returns by extrapolating past returns, while other investors have fully rational beliefs.

- The model captures important facts about prices:
  - Excess volatility in returns
  - Predictability of returns from the P/D ratio
  - Return autocorrelations
  - Persistence of P/D ratio

- But is also consistent with the survey evidence.
Return extrapolation: Aggregate market, ctd.

Assets

• an economy with two assets
  – a risk-free asset with a constant interest rate $r$
  – a risky asset, the aggregate stock market, with a fixed per-capita supply $Q$

• the risky asset is a claim to a continuous dividend stream whose level per unit time evolves as an arithmetic Brownian motion

\[ dD_t = g_D dt + \sigma_D d\omega \]

• the value (price) of the stock market at time $t$ is denoted as $P_t$, and is determined in equilibrium
Return extrapolation: Aggregate market, ctd.

**Traders**

- two types of traders: “extrapolators” and “rational traders”
- there is a continuum of each type
- rational traders make up a fraction $\mu$ of the investor population, and extrapolators, a fraction $1 - \mu$

**Belief structure**

- we introduce a “sentiment” variable
  \[
  S_t = \beta \int_{-\infty}^{t} e^{-\beta(t-s)}dP_{s-dt}, \quad \beta > 0
  \]
  - an average of past price changes, with exponentially-declining weights, governed by $\beta$
- the extrapolator’s expected price change, per unit time, is
  \[
  E^e_t[dP_t]/dt = \lambda_0 + \lambda_1 S_t,
  \]
  - where $\lambda_0$ and $\lambda_1$ are constants, with $\lambda_1 > 0$
- the rational traders, on the other hand, have *correct* beliefs about the evolution of future stock prices
Return extrapolation: Aggregate market, ctd.

**Information sets**

- both extrapolators and rational traders observe $D_t$ and $P_t$ on a continuous basis
- they both know the values of $\mu$ and $Q$
- traders of one type understand how traders of the other type form beliefs about the future

**Preferences**

- both types of traders have constant absolute risk aversion (CARA) preferences with risk aversion $\gamma$ and time discount factor $\delta$
- they each maximize lifetime consumption utility subject to their budget constraints

**Market clearing**

$$\mu N_t^r + (1 - \mu) N_t^e = Q$$
Return extrapolation: Aggregate market, ctd.

Equilibrium price of the risky asset

\[ P_t = A + BS_t + \frac{D_t}{r}, \]

where, in equilibrium, \( B > 0 \) and \( S_t \) is mean-reverting

- the model generates the key facts about stock market prices
  - excess volatility
  - predictability
  - return autocorrelations
  - persistence of P/D ratio
- but is also consistent with the survey evidence
  - an important contrast to other models of the aggregate stock market
  - e.g. habit preferences, long-run risk, rare disasters, gain/loss utility
Return extrapolation: Bubbles

• a bubble is an episode in which:
  – the price of an asset rises sharply over some period of time and then collapses
  – during the price rise, there is much talk of overvaluation in the media and among investors

• also, some of the following are observed:
  – very high trading volume
  – extrapolative expectations
  – sophisticated investors “riding the bubble”
  – good fundamental news near the start of the bubble

• we now present a model of return extrapolation that can generate such episodes
  – mechanism for high prices is the usual one
  – mechanism for high volume is novel, and based on a concept called “wavering”
Return extrapolation: Bubbles, ctd.

Timing

- $t = 0, 1, \ldots, T$

Assets

- riskless asset, constant return of zero
- risky asset
  - fixed supply of $Q$ shares
  - claim to a final cash flow $\tilde{D}_T$

$$\tilde{D}_T = D_0 + \tilde{\varepsilon}_1 + \ldots + \tilde{\varepsilon}_T$$

$$\tilde{\varepsilon}_t \sim N(0, \sigma_\varepsilon^2) \text{ i.i.d.}$$

Investors

- two types
  - fundamental traders
  - extrapolators
Return extrapolation: Bubbles, ctd.

**Fundamental traders**

- arbitrageurs, with time $t$ demand
  \[
  \frac{D_t - \gamma \sigma^2 \varepsilon (T - t - 1) Q - P_t}{\gamma \sigma^2 \varepsilon}
  \]
  - the “fundamental value” of the asset is the price that would obtain if all investors were fundamental traders

**Extrapolators**

- $I$ types of extrapolators
- initial specification of demand:
  \[
  \frac{X_t}{\gamma \sigma^2 \varepsilon}, \quad \text{where}
  \]
  \[
  X_t = (1 - \theta) \sum_{k=1}^{t-1} \theta^{k-1} (P_{t-k} - P_{t-k-1}) + \theta^{t-1} X_1
  \]
  with $0 < \theta < 1$
  - demand of an investor with CARA preferences over next period’s wealth
    - and who expects the price change over the next period to be a weighted average of past price changes
Return extrapolation: Bubbles, ctd.

Extrapolators, ctd.

- make two modifications to the traditional extrapolation specification
- first, extrapolators pay some attention to how price compares to fundamental value

\[
    w_i \left( \frac{D_t - \gamma \sigma_\xi^2 (T - t - 1) Q - P_t}{\gamma \sigma_\xi^2} \right) + (1 - w_i) \frac{X_t}{\gamma \sigma_\xi^2}
\]

where \( w_i \) takes a low value (\( \approx 0.1 \))

- refer to the two components of demand as signals
  - a “value” signal and a “growth” signal, which often point in opposite directions
Return extrapolation: Bubbles, ctd.

Extrapolators, ctd.

• in addition, the relative weight an extrapolator puts on the two signals varies *slightly* over time
  – independently across extrapolators and over time

• extrapolator $i$’s demand becomes:

$$w_{i,t} \left( \frac{D_t - \gamma \sigma^2 \varepsilon (T - t - 1)Q - P_t}{\gamma \sigma^2 \varepsilon} \right) + (1 - w_{i,t}) \frac{X_t}{\gamma \sigma^2 \varepsilon}$$

\[ w_{i,t} = \bar{w}_i + u_{i,t} \]
\[ u_{i,t} \sim N(0, \sigma^2_u) \text{ i.i.d.} \]

• we call this “wavering”
  – may stem from small fluctuations in the relative attention extrapolators pay to the two signals

• also impose short-sale constraints on both fundamental traders and extrapolators
  – but only the wavering assumption is critical
Return extrapolation: Bubbles, ctd.

Parameter values

• investor-level parameters
  – 30% of investors are fundamental traders, 70% extrapolators
  – 50 types of extrapolator
  – extrapolator base weight $\bar{w}_i$ on the value signal is 0.1
  – degree of risk aversion $\gamma$ is 0.1
  – extrapolation parameter $\theta$ is 0.9
  – degree of wavering $\sigma_u$ is 0.03

• asset-level parameters
  – initial expected dividend $D_0$ is 100
  – asset supply $Q$ is 1
  – fundamental risk $\sigma_\varepsilon$ is 3
  – number of periods $T$ is 50
  – length of each period is one quarter
Return extrapolation: Bubbles, ctd.

Prices

- model can generate the most basic feature of a bubble, a large overvaluation
- look at the asset’s price and fundamental value for a specific sequence of cash-flow shocks

\[
\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_{10}\} = \{0, \ldots, 0\} \\
\{\tilde{\varepsilon}_{11}, \ldots, \tilde{\varepsilon}_{14}\} = \{2, 4, 6, 6\} \\
\{\tilde{\varepsilon}_{15}, \ldots, \tilde{\varepsilon}_{50}\} = \{0, \ldots, 0\}
\]
Return extrapolation: Bubbles, ctd.

Prices, ctd.

• the bubble evolves over three stages
  – Stage 1: both fundamental traders and extrapolators hold the asset
  – Stage 2: only extrapolators hold the asset
  – Stage 3: fundamental traders re-enter
Return extrapolation: Bubbles, ctd.

**Volume**

- can the model help us understand why volume is high during bubbles?

- plot share demands of fundamental traders \((N_t^F)\) and extrapolators \((N_t^{E,i})\) for original sequence of cash-flow shocks

\[
\{\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_{10}\} = \{0, \ldots, 0\} \\
\{\bar{\varepsilon}_{11}, \ldots, \bar{\varepsilon}_{14}\} = \{2, 4, 6, 6\} \\
\{\bar{\varepsilon}_{15}, \ldots, \bar{\varepsilon}_{50}\} = \{0, \ldots, 0\}
\]
Return extrapolation: Bubbles, ctd.

*Volume*, ctd.

- also plot trading volume at each date
  
  - both total trading volume (solid line) and trading volume within the set of extrapolators (dashed line)
Return extrapolation: Bubbles, ctd.

*Volume*, ctd.

- the model predicts high volume during a bubble
  - its source varies by bubble stage

*First stage*

- volume is substantial
  - consists of extrapolators buying from fundamental traders

*Third stage*

- volume is again substantial
  - the fundamental traders buy from extrapolators
Return extrapolation: Bubbles, ctd.

Second stage

• even though asset is held and traded only by extrapolators, volume is very high

• source of volume in this stage is wavering

Key idea:

• even though degree of wavering is constant over time, it endogenously generates much higher volume during the bubble

• extrapolator i’s demand is \( w_{i,t}V_t + (1 - w_{i,t})G_t \)

  \[ \Rightarrow \text{a 0.01 shift in } w \text{ leads to a change in share demand of } 0.01(|V - G|) \]

• during “normal” times, the value and growth signals have small magnitudes, e.g. \( V = -2 \) and \( G = 2 \)

  \[ \Rightarrow \text{a 0.01 shift in } w \text{ leads to a change in the extrapolator’s share demand of } 0.01(4) = 0.04 \]

• but during a bubble, the value and growth signals are very large, e.g. \( V = -20 \) and \( G = 20 \)

  \[ \Rightarrow \text{a 0.01 shift in } w \text{ leads to a large change in share demand of } 0.01(40) = 0.4 \]
Return extrapolation: Bubbles, ctd.

• analysis of volume points to a testable prediction
  – during a bubble, volume will be strongly positively related to the asset’s past return

• test and confirm the prediction in four bubble episodes
  – the correlations between monthly turnover and past annual returns are high (0.67, 0.71, 0.84, 0.83)
Return extrapolation: Bubbles, ctd.

Summary

- a bubble is an episode in which:
  - the price of an asset rises sharply over some period of time and then collapses
  - during the price rise, there is much talk of overvaluation in the media and among investors

- also, some of the following are observed:
  - very high trading volume
  - extrapolative expectations
  - sophisticated investors “riding the bubble”
  - good fundamental news near the start of the price rise

- the model generates episodes with most of these features
Return extrapolation: Sources

• the most commonly-cited source of return extrapolation is the representativeness heuristic
  – Kahneman and Tversky (1974)

Representativeness

• consider questions such as:
  – what is the probability that object A comes from class B?
  – what is the probability that event A was generated by process B?

• Kahneman and Tversky (1974) argue that people often answer by using the representativeness heuristic
  – evaluate the probability by the extent to which A is representative of B
  – i.e. degree to which A reflects the essential characteristics of B

• this is often reasonable, but can lead to serious biases
  – base-rate neglect, sample-size neglect
Return extrapolation: Sources, ctd.

Representativeness: Base-rate neglect

• consider the following description

“Steve is very shy and withdrawn, invariably helpful, but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.”

• is Steve more likely to be a librarian or a lawyer?

\[
p(\text{lib}|\text{data}) = \frac{p(\text{data}|\text{lib})p(\text{lib})}{p(\text{data})}
\]
Return extrapolation: Sources, ctd.

- return extrapolation can be motivated by base-rate neglect

Other sources of return extrapolation:

- past returns are a signal of changes in fundamentals that are hard to observe directly
  - Hong and Stein (1999), Glaeser and Nathanson (2016)
- a belief that the true mean stock market return is time-varying
Return extrapolation, ctd.

• can make a case for return extrapolation as one of the most useful concepts in behavioral finance

Broad range of important applications:

• aggregate stock market
  – excess volatility, predictability

• bubbles
  – high prices and high volume

• cross-section of stock returns
  – momentum, long-term reversals, value premium
Cash-flow extrapolation

• we now turn to over-extrapolation of *fundamentals*

• this can address some of the same applications as return extrapolation
  – excess volatility and predictability in aggregate asset classes
  – momentum, long-run reversals, and the value premium in the cross-section

• however, it may not capture the survey evidence on return expectations

• the possible sources of cash-flow extrapolation are similar to those for return extrapolation
  – e.g. representativeness
  – but also: underestimation of competitive pressure (Greenwood and Hanson, 2015)

• some references
Over-extrapolation: Summary

- *over-extrapolation* of fundamentals or returns is one of the most important and widely-applied ideas in behavioral finance
  - the idea that, when people form beliefs about future returns or cash-flow growth, they put too much weight on recent past returns or cash-flow growth

Roadmap

- return extrapolation
  - intuition
  - application: aggregate stock market
  - application: bubbles
  - sources of return extrapolation

- cash-flow extrapolation

- experience effects
Experience effects

• research on “experience effects” posits that people form beliefs about future returns or cash flows as a weighted average of returns or cash flows they have observed *in their lifetimes*
  
  – with more weight on more recent observations

• we can think of this as introducing a form of heterogeneity in extrapolative beliefs

• such beliefs can help explain stock market participation and stock market risk exposure
  
  – Malmendier and Nagel (2011)
BEHAVIORAL FINANCE
Nicholas Barberis, AEA 2017
Lecture Note 5: Overconfidence and Other Belief Biases

© Nicholas C. Barberis
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Overview

• behavioral finance models aim for psychological realism along three dimensions
  – allow for less than fully rational beliefs
  – use more realistic preferences
  – take account of cognitive limits

• in Lecture Notes 4 and 5, we focus on the first dimension: investor beliefs
  – Lecture Note 4: extrapolative beliefs
  – Lecture Note 5: overconfidence and other belief biases
Overconfidence

Overconfidence is a robust phenomenon, and manifests itself in at least two forms:

Overprecision

• people are too confident in the accuracy of their beliefs
  – 90% confidence intervals contain the correct answer around 50% of the time

Overplacement

• people have overly rosy views of their abilities relative to other people
Overconfidence, ctd.

• the main motivation for invoking overconfidence in finance is the very high *trading volume* in financial markets

• non-speculative motives for trade are unlikely to explain much of this

• speculative motives are a more plausible driver
  – i.e. differing beliefs about the future price change of an asset

• overconfidence is a promising way of generating differences in beliefs and trading volume
Overconfidence and disagreement

• two individuals who have the same prior beliefs, observe the same information, and are both rational, will have the same posterior beliefs

• disagreement can therefore stem from one of three sources
  – different priors
  – different information
  – departures from rationality

• economists have explored all three channels as possible sources of trading volume
  – the three channels make different predictions
Overconfidence and disagreement, ctd.

• a key insight from the 1980s is that models where rational investors observe different information may not generate much trading volume

• each investor infers others’ signals from prices, or from their willingness to trade
  – this reduces her own willingness to trade

• overconfidence offers a way out of this logjam
  – here, use “overconfidence” to mean overestimation of the precision of one’s own information signals
  – and “dismissiveness” to mean underestimation of the precision of others’ signals

• both overconfidence and dismissiveness can generate significant trading volume

• see Morris (1996) for an analysis of disagreement and trading volume based on non-common priors
Overconfidence and disagreement, ctd.

• an intuitive prediction is that more overconfident people will trade more

Empirical tests:

*Grinblatt and Keloharju (2009)*

• use data from Finland to show that more overconfident people trade more
  
  – overconfidence is self-reported confidence minus how confident the individual should be based on performance on aptitude tests

*Barber and Odean (2001)*

• argue that, since men tend to be more overconfident than women, they will trade more and have worse returns
  
  – confirm this using brokerage data
Disagreement with short-sale constraints

- an important framework in finance couples overconfidence-based disagreement with short-sale constraints (SSC)
  - this offers an appealing way of thinking about overpricing and bubbles
- overconfidence-based disagreement and short-sale constraints can generate overpricing through two distinct channels

*Static argument* (Miller, 1977)

- if investors disagree about an asset’s future prospects, the optimists buy the asset while the pessimists stay out of the market
  \[ \Rightarrow \text{the asset becomes overpriced} \]
Disagreement with SSC, ctd.

Dynamic argument (Harrison and Kreps, 1978)

- if investors disagree, each is willing to pay more than her estimate of the present value of future cash flows
  - when information is released, there is a chance that she will be able to sell to someone more optimistic

- Scheinkman and Xiong (2003) build on this idea
  - put in an explicit source of disagreement, namely overconfidence
  - make predictions not only about prices, but about volume and volatility as well
  - put in a trading cost
Disagreement with SSC, ctd.

*Scheinkman and Xiong (2003)*

- single risky asset in finite supply, paying a dividend with unobserved drift

\[
\begin{align*}
    dD_t &= f_t dt + \sigma_D dZ^D_t \\
    df &= -\lambda(f - \bar{f}) dt + \sigma_f dZ^f_t
\end{align*}
\]

- two sets of risk-neutral agents, A and B

- two signals, observed by both sets of agents

\[
\begin{align*}
    ds^A_t &= f_t dt + \sigma_S dZ^A_t \\
    ds^B_t &= f_t dt + \sigma_S dZ^B_t
\end{align*}
\]

\(Z^D, Z^f, Z^A, Z^B\) are all independent

- group A *thinks* that \(dZ^A\) is correlated with \(dZ^f\), to an extent determined by a parameter \(\phi\)

  - group B *thinks* that \(dZ^B\) is correlated with \(dZ^f\)

- a trading cost \(c\) is paid by sellers
Disagreement with SSC, ctd.

*Scheinkman and Xiong (2003), ctd.*

- the model predicts
  \[ \text{price} = \text{fundamental value} + \text{resale value} \]
- i.e. it predicts overpricing and high volume
  - price and volume move together as we vary the exogenous parameters
- the bubble is largest when the trading cost \( c = 0 \)
  - as \( c \) increases, volume drops quickly
  - prices also drop, but less quickly
Disagreement with SSC, ctd.

- models of disagreement with SSC are popular because they not only explain overpricing, but also another important empirical fact:
  - the coincidence of high valuations and heavy trading
- evidence (Hong and Stein, 2007)
  - value stocks vs. growth stocks
  - technology stocks in the late 1990s
  - shares at the center of famous bubble episodes (South Sea bubble)

Turnover in Value and Glamour Stocks, 1986–2005
Overconfidence

• Daniel, Hirshleifer, Subrahmanyam (1998) present a model of misvaluation based on a different implementation of overconfidence
  – apply it to the cross-section of stock returns

• a risk-neutral, representative investor is overconfident about the private information he gathers
  – this leads to long-term return reversals and a value premium

• also add in “self-attribution” bias
  – when public information confirms the private signal, the investor becomes even more confident in the private signal
  – when public information disconfirms the private signal, he does not lose much confidence in the private signal

• this leads to momentum in addition to a value premium
Other belief assumptions

• the most useful assumptions about investor beliefs are:
  – extrapolation of the past
  – overconfidence

• but other belief assumptions have been explored as well
  – belief perseverance, confirmation bias
  – the availability heuristic
  – the effect of “feelings” on beliefs
Belief perseverance

• once we have formed an opinion, we are often too slow to change it on receipt of new evidence
  – “belief perseverance”

• we don’t look for evidence that would falsify our beliefs
  – and ignore evidence that goes against us

• more extreme version is “confirmation bias”
  – we misread evidence that goes against us as actually being in our favor
  – e.g. capital punishment studies (Lord, Ross, Lepper, 1979)

• belief perseverance offers an explanation of post-earnings announcement drift and momentum based on slow updating of beliefs
Availability

• when we judge the likelihood of an event, we often do so based on how easy it is to recall instances of the event
  – Kahneman and Tversky (1974)

• however, there are biases in recall
  – more recent events and more salient events are recalled more easily
Availability, ctd.

• Jin (2014) considers a model in which “extrapolators” and “long-term investors” trade a riskless asset and a risky asset

• the risky asset is subject to occasional crashes in fundamentals that occur with constant likelihood
  – however, extrapolators think that a crash is less likely, the fewer such crashes have recently been observed
  – such beliefs can be motivated by the availability heuristic

• after a long quiet period, extrapolators under-estimate the likelihood of a crash
  – as a result, they take an excessively levered position in the risky asset

• when a crash in fundamentals occurs, the drop in prices is even larger
  – the extrapolators delever, and update their beliefs
The effect of feelings

• an improvement in mood due to an exogenous stimulus leads to more positive judgments about unrelated events
  – Johnson and Tversky (1983)

Example: *Soccer*

• when the national soccer team loses a World Cup match, the national stock market falls the next day

Example: *Sun*

• the stock market has higher returns on sunnier days
  – Hirshleifer and Shumway (2003)
BEHAVIORAL FINANCE
Nicholas Barberis, AEA 2017
Lecture Note 6: Prospect Theory
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

- extrapolation (LN 4)
- overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

- prospect theory (LN 6)
- ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

- bounded rationality (LN 8)

IV. Conclusion

- summary and conclusion (LN 9)
Overview

• Behavioral finance models aim for psychological realism along three dimensions
  – allow for less than fully rational beliefs
  – use more realistic preferences
  – take account of cognitive limits

• In Lecture Notes 6 and 7, we focus on the second dimension: investor preferences
  – Lecture Note 6: prospect theory
  – Lecture Note 7: ambiguity aversion and other preference hypotheses
Overview, ctd.

• most models of financial markets assume that investors evaluate risk according to Expected Utility (EU)

• however, a large body of work shows that, at least in experimental settings, EU is not an accurate description of risk attitudes

• there are now many non-EU models that try to capture these departures from EU

  – prospect theory, due to Kahneman and Tversky (1979, 1992), is the best known

• research question: can we make progress by incorporating ideas from prospect theory into our models of financial markets?

Note:

• while prospect theory is the non-EU model that has been most widely applied in finance, others have also been explored

  – disappointment aversion (Gul, 1991)
  – rank-dependent utility (Quiggin, 1982, 1983; Yaari, 1987)
  – salience theory (Bordalo, Gennaioli, Shleifer, 2012, 2013)
Prospect Theory

The original version (Kahneman and Tversky, 1979)

Consider the gamble \((x, p; y, q)\)

- under EU, it is assigned the value
  \[ pU(W + x) + qU(W + y) \]

- under prospect theory, it is assigned the value
  \[ w(p)v(x) + w(q)v(y) \]
Prospect Theory, ctd.

Four key features:

Reference dependence

• the carriers of value are gains and losses, not final wealth levels

  – experimental evidence
  – consistent with perception of other attributes

Loss aversion

• \( v(\cdot) \) has a kink at the origin

  – captures a greater sensitivity to losses (even small losses) than to gains of the same magnitude
  – inferred from aversion to \((\$110, \frac{1}{2}; -\$100, \frac{1}{2})\)

Diminishing sensitivity

• \( v(\cdot) \) is concave over gains, convex over losses

  – inferred from \((\$500, 1) \succ (\$1000, \frac{1}{2})\) and \((-\$500, 1) \prec (-\$1000, \frac{1}{2})\)
Prospect Theory, ctd.

**Probability weighting**

- transform probabilities with a weighting function $w(\cdot)$ that overweights low probabilities
  - inferred from our simultaneous liking of lotteries and insurance, e.g. ($5, 1) \prec (5000, 0.001)$ and $(-5, 1) \succ (-5000, 0.001)$

Note:

- transformed probabilities should not be thought of as beliefs, but as decision weights
- it is interesting to think about the psychological foundations of probability weighting
  - diminishing sensitivity (Tversky and Kahneman, 1992)
  - evolutionary interpretation
  - affect (Rottenstreich and Hsee, 2001)
Prospect Theory, ctd.

Cumulative prospect theory

• proposed by Tversky and Kahneman (1992)
  – addresses some limitations of the original prospect theory

• applies the probability weighting function to the cumulative distribution function:

\[(x_{-m}, p_{-m}; \ldots; x_{-1}, p_{-1}; x_0, p_0; x_1, p_1; \ldots; x_n, p_n),\]

where \(x_i < x_j\) for \(i < j\) and \(x_0 = 0\), is assigned the value

\[
\sum_{i=-m}^{n} \pi_i v(x_i)
\]

\[
\pi_i = \begin{cases} 
  w(p_i + \ldots + p_n) - w(p_{i+1} + \ldots + p_n) & \text{for } 0 \leq i \leq n \\
  w(p_{-m} + \ldots + p_i) - w(p_{-m} + \ldots + p_{i-1}) & \text{for } -m \leq i < 0
\end{cases}
\]

• the individual now overweights the tails of a probability distribution
  – this preserves a preference for lottery-like gambles

• one possible foundation for the overweighting of tails is salience (Bordalo, Gennaioli, Shleifer, 2012)
Prospect Theory, ctd.

- Tversky and Kahneman (1992) also suggest functional forms for $v(\cdot)$ and $w(\cdot)$ and calibrate them to experimental evidence

$$v(x) = \begin{cases} 
  x^\alpha & \text{for } x \geq 0 \\
  -\lambda(-x)^\alpha & \text{for } x < 0 
\end{cases}$$

$$w(P) = \frac{P^\delta}{(P^\delta + (1 - P)^\delta)^{1/\delta}}$$

with

$$\alpha = 0.88, \lambda = 2.25, \delta = 0.65$$
Prospect Theory, ctd.
Prospect theory, ctd.

• prospect theory is often implemented in conjunction with “narrow framing”

• in a traditional model where utility is defined over wealth or consumption, an individual evaluates a new risk by combining it with pre-existing risks and checking if the combination is an improvement

• but in experimental settings, people often seem to evaluate a new risk *in isolation*, separately from other concurrent risks
  
  – narrow framing

• e.g. the widespread rejection of the gamble \((110, \frac{1}{2}; -100, \frac{1}{2})\) is not only evidence of loss aversion, but of narrow framing as well
  
  – Barberis, Huang, Thaler (2006)

• implications for finance
  
  – we will sometimes take the “gains and losses” of prospect theory to be gains and losses in specific components of wealth
  
  – e.g. gains and losses in stock market wealth or gains and losses in specific stocks
Prospect theory applications

[1]  
• the cross-section of stock returns  
  – one-period models  
  – new prediction: the pricing of skewness  
  – probability weighting plays the most critical role

[2]  
• the aggregate stock market  
  – intertemporal representative-agent models  
  – address the equity premium, non-participation, volatility, and predictability puzzles  
  – loss aversion plays a key role; but probability weighting also matters

[3]  
• trading behavior  
  – multi-period models  
  – address the disposition effect and other trading phenomena  
  – all aspects of prospect theory play a role
Prospect theory applications, ctd.

Note:

• a fundamental challenge in applying prospect theory is defining the “gains” and “losses”
  – gains and losses in total wealth, financial wealth, stock market holdings, individual stocks?
  – annual gains and losses?
  – is a gain a return that exceeds zero, or one that exceeds the risk-free rate or the investor’s expectation?

• we typically take the gains and losses to be annual gains and losses in financial wealth
  – where a gain is measured relative to the risk-free rate
The cross-section

Barberis and Huang (2008), “Stocks as Lotteries...”

- single period model; a risk-free asset and $J$ risky assets with multivariate Normal payoffs
- investors have identical expectations about security payoffs
- investors have identical CPT preferences
  - defined over gains/losses in wealth (i.e. no narrow framing)
  - reference point is initial wealth scaled up by the risk-free rate, so utility defined over $\hat{W} = \tilde{W}_1 - W_0 R_f$
  - full specification is:
    $$V(\hat{W}) = \int_{-\infty}^{0} v(W) \, dw(P(W)) - \int_{0}^{\infty} v(W) \, dw(1-P(W))$$
    (continuous distribution version of Tversky and Kahneman, 1992)

Then:

- the CAPM holds!
  - i.e. prospect theory gives the same prediction as the EU model
  - see also De Giorgi, Hens, Levy (2011)
The cross-section, ctd.

• to make more interesting predictions, break away from the multivariate Normal assumption
  – introduce a small, independent, positively skewed security into the economy
• obtain a novel prediction: the new security earns a negative excess return
  – skewness itself is priced, in contrast to concave EU model where only coskewness with market matters
• equilibrium involves heterogeneous holdings
  (assume short-sale constraints for now)
  – some investors hold a large, undiversified position in the new security
  – others hold no position in it at all
  – heterogeneous holdings arise from non-unique global optima, not from heterogeneous preferences

• since the new security contributes skewness to the portfolios of some investors, it is valuable, and so earns a low average return
Figure 3. A Heterogeneous Holdings Equilibrium

Notes: The figure shows the utility that an investor with cumulative prospect theory preferences derives from adding a position in a positively skewed security to his current holdings of a Normally distributed market portfolio. The skewed security is highly skewed. The variable $x$ is the fraction of wealth allocated to the skewed security relative to the fraction of wealth allocated to the market portfolio. The two lines correspond to different mean returns on the skewed security.
Applications

• low average return on IPOs
  – IPO returns are highly positively skewed
  – Green and Hwang (2012) show that IPOs predicted to be more positively skewed have lower long-term returns

• low average return of distressed stocks, bankrupt stocks, OTC stocks (Eraker and Ready, 2015)

• “overpricing” of out-of-the-money options on individual stocks
  – Boyer and Vorkink (2014) find that stock options predicted to be more positively skewed have lower returns

• low average return on stocks with high idiosyncratic volatility (Ang et al., 2006; Boyer, Mitton, Vorkink, 2010)
The cross-section, ctd.

Applications, ctd.

• under-diversification
  – Mitton and Vorkink (2007) find that undiversified individuals hold stocks that are more positively skewed than the average stock

• several papers test the model’s basic prediction that skewness is priced in the cross-section
  – Boyer, Mitton, Vorkink (2010) use a regression model to predict future skewness
  – Conrad, Dittmar, Ghysels (2013) use option prices to infer the perceived distribution of the underlying stock
  – Bali, Cakici, Whitelaw (2011) use the maximum daily return in the past month as a skewness proxy

• all three studies find evidence in line with the prediction
The cross-section, ctd.

Note:

• an example of how psychology can lead to interesting new predictions

• probability weighting plays a central role

Alternative framing assumptions?

• models with *stock*-level framing are also being explored
  – Barberis and Huang (2001), Barberis, Mukherjee, Wang (2016)

• such models will likely continue to predict the pricing of (idiosyncratic) skewness
The aggregate stock market

Can prospect theory help us understand the properties of, and attitudes to, the aggregate stock market?

- Benartzi and Thaler (1995) argue that a model in which investors are loss averse over annual changes in the value of their stock market holdings predicts a large equity premium.

- three elements:
  - loss aversion
  - annual evaluation
  - narrow framing

- Benartzi and Thaler (1995) emphasize the first two elements
  - “myopic loss aversion”
The aggregate stock market, ctd.

Subsequent developments:

• formalizing the argument
• studying the role of probability weighting
• trying to address the volatility puzzle as well
The aggregate stock market, ctd.

Formalizing the argument

• to fill out the argument, we need to embed it in the setting where the equity premium is usually studied
  – an intertemporal, representative agent model where consumption plays a non-trivial role
  – e.g. where preferences include a utility of consumption term alongside the prospect theory term

• two possible ways of doing this:
  – Barberis, Huang, and Santos (2001)
  – Barberis and Huang (2009)

• for other formalizations, see Andries (2013) and Pagel (2015)
The aggregate stock market, ctd.

*Formalizing the argument*, ctd.

Barberis, Huang, and Santos (2001)

- intertemporal model; three assets: risk-free \(R_{f,t}\), stock market \(R_{S,t+1}\), another risky asset \(R_{N,t+1}\)
- representative agent maximizes:

\[
E_0 \sum_{t=0}^{\infty} \left[ \rho^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_0 \rho^{t+1} C_{t+1}^{-\gamma} v(G_{S,t+1}) \right]
\]

\[
G_{S,t+1} = \theta_{S,t}(W_t - C_t)(R_{S,t+1} - R_{f,t})
\]

\[
v(x) = \begin{cases} 
  x & \text{for } x \geq 0 \\
  \lambda x & \text{for } x < 0, \lambda > 1 
\end{cases}
\]

- this assumes narrow framing
- and that the reference point is the risk-free rate
- \(v(\cdot)\) captures loss aversion
- we ignore concavity/convexity and probability weighting for now

- for “reasonable” parameters, get a substantial equity premium, although not as large as in Benartzi and Thaler (1995)
The aggregate stock market, ctd.

The role of probability weighting

• De Giorgi and Legg (2012) build on a framework of Barberis and Huang (2009) to also incorporate probability weighting and concavity/convexity
  – they show that probability weighting can significantly increase the equity premium
  – because the aggregate market is \textit{negatively} skewed

The volatility and predictability puzzles

• Barberis, Huang, and Santos (2001) also build in dynamic aspects of loss aversion
  – based on evidence in Thaler and Johnson (1990), posit that loss aversion rises (falls) after past gains (losses)
  – can be interpreted in terms of “capacity for dealing with bad news”
  – generates excess volatility and predictability in addition to a high equity premium
The aggregate stock market, ctd.

Note:

• we are using frameworks in which investors derive utility from fluctuations in financial wealth, not just consumption

• we can justify this in terms of “mental accounting”
  – to try to ensure good future consumption outcomes, investors track wealth fluctuations on a regular basis
  – an increase in wealth is “good news” and becomes associated with a positive utility burst
  – a decrease in wealth is “bad news” and becomes associated with a negative utility burst
Trading behavior

• can prospect theory help us understand how individuals trade financial assets over time?

• a particular target of interest is the “disposition effect”
  – individual investors’ greater propensity to sell stocks trading at a gain relative to purchase price, rather than at a loss

• at first sight, prospect theory, in combination with stock-level narrow framing, appears to be a promising approach

• but it turns out that we need to be careful how we implement prospect theory
  – prospect theory defined over annual stock-level trading profits does not necessarily generate a disposition effect
  – Barberis and Xiong (2009), “What Drives the Disposition Effect?...”
Trading behavior, ctd.

• consider a simple portfolio choice setting
  
  – $T + 1$ dates: $t = 0, 1, \ldots, T$
  – a risk-free asset, gross return $R_f$ each period
  – a risky asset with an i.i.d binomial distribution across periods:

  $$R_{t,t+1} = \begin{cases} 
  R_u > R_f \text{ with probability } \frac{1}{2}, \\
  R_d < R_f \text{ with probability } \frac{1}{2}, 
  \end{cases} \text{, i.i.d.}$$

• the investor has prospect theory preferences defined over his “gain/loss”
  
  – simplest definition of gain/loss is trading profit between 0 and $T$, i.e. $W_T - W_0$
  – we use $W_T - W_0 R_f^T$
Trading behavior, ctd.

The investor therefore solves

$$\max_{x_0, x_1, \ldots, x_{T-1}} E[v(\Delta W_T)] = E[v(W_T - W_0 R_f^T)]$$

where

$$v(x) = \begin{cases} x^\alpha & \text{for } x \geq 0 \\ -\lambda(-x)^\alpha & \text{for } x < 0 \end{cases},$$

subject to

$$W_t = (W_{t-1} - x_{t-1} P_{t-1}) R_f + x_{t-1} P_{t-1} R_{t-1,t}$$

$$W_T \geq 0$$

- note that we are assuming stock-level narrow framing
  - and are ignoring probability weighting
- we can derive an analytical solution for any number of trading periods
Trading behavior, ctd.

Results

- the investor often exhibits the opposite of the disposition effect
- for $T = 2$ and for the Tversky and Kahneman (1992) parameterization, he always exhibits the opposite of the disposition effect

Intuition

- loss aversion generates the opposite of the disposition effect
  - the expected return has to be high for the investor to buy the stock at all
  - after a gain, he is therefore further from the kink
- the concavity/convexity estimated by Tversky and Kahneman (1992) is too weak to overcome this
- for stronger concavity / convexity, the model does generate a disposition effect
  - more recent estimates of $\alpha$ suggest that this may be empirically relevant
Trading behavior, ctd.

- an alternative gain-loss utility approach can generate the disposition effect more reliably
  - one based on “realization utility”
  - the idea that investors derive utility directly from realized gains and losses (Shefrin and Statman, 1985)
- e.g. if you buy a stock at $40 and sell it at $60
  - you get a jolt of positive utility *at the moment of sale*, based on the size of the realized gain
- what is the source of realization utility?
  - people often think about their investing history as a series of investing episodes
  - and view selling a stock at a gain as a “good” episode
  ⇒ when an investor sells an asset at a gain, he feels a burst of pleasure because he is creating a positive new investing episode
Trading behavior, ctd.

• Barberis and Xiong (2012), “Realization Utility,” study linear realization utility, coupled with time discounting

Assets

• a risk-free asset, with net return of zero

• $N$ risky assets, “stocks”; stock $i$ has price process

\[
\frac{dS_{i,t}}{S_{i,t}} = \mu dt + \sigma dZ_{i,t}
\]

– $\mu$ and $\sigma$ are the same for all stocks

The investor

• at each time $t$, he either allocates all of his wealth to the risk-free asset, or all of his wealth to one of the $N$ stocks

– time $t$ wealth is $W_t$

• if he is holding stock at time $t$, let $B_t$ be the cost basis of the position

• if he sells stock at time $t$, he pays a transaction cost $kW_t$
Trading behavior, ctd.

Key assumption:

• if, at time $t$, the investor switches his wealth from a stock to the risk-free asset or to another stock, he receives realization utility of

$$u((1 - k)W_t - B_t)$$

• he also faces the possibility of a random liquidity shock
  – when a shock hits, the investor sells his asset holdings and exits the asset markets

• the investor maximizes the discounted sum of expected future realization utility flows
  – $\delta$ is the time discount rate
  – we take $u(x) = x$
Trading behavior, ctd.

Solution

• if the expected return on stocks is low, the individual invests in the risk-free asset forever

• if the expected return on stocks is high enough, he buys a stock at time 0
  – and sells it only if its value rises a certain percentage amount above purchase price
  – he then immediately reinvests in another stock, and so on
Trading behavior, ctd.

- applications:
  - the disposition effect

but also:

- “excessive trading”
- the underperformance of individual investors even before transaction costs
- the greater turnover in bull markets
- the greater selling propensity above historical highs
- the negative premium to volatility in the cross-section
- the fact that overpriced assets are also heavily traded

Note:

- realization utility can also be coupled with an S-shaped utility function, rather than with time discounting (Ingersoll and Jin, 2013)
  - generates a disposition effect
  - but also, voluntary selling of loser stocks
Trading behavior, ctd.

• realization utility is an unusual assumption
  – can we provide more evidence for it?

• Frydman, Barberis, Camerer, Bossaerts, and Rangel (2014) use neural data to test the realization utility hypothesis
  – fMRI brain scan data collected while participants trade stocks in an experimental stock market

• by design, the stocks have positively autocorrelated price changes
  – the optimal strategy is therefore the opposite of the disposition effect

• however, participants exhibit a strong disposition effect
  – we test the realization utility hypothesis for this behavior
Trading behavior, ctd.

• realization utility says that investors experience a burst of utility when they sell a stock at a gain

• the ventral striatum (vSt) is thought to encode hedonic value, or subjective feelings of pleasure

• therefore, under realization utility, activity in the vSt should spike up when a participant issues a command to sell a stock at a gain

  – as compared to when he issues a command to hold a stock with a similar embedded gain

• we plot a time series of activity in the vSt around the moment at which a participant issues a command to sell a stock at a gain

  – and compare it to the time series of activity in the vSt around the moment at which a participant issues a command to hold a stock with a gain
Figure 8. Direct tests of the realization utility hypothesis. Panel A: Yellow voxels are those in our a priori region of interest in the vSt. Red voxels are those that exhibit greater activity when subjects realize capital gains compared to when they hold capital gains (shown at $p < 0.001$ uncorrected with a 15-voxel extent threshold, for illustrative purposes only). Orange voxels are those that are in the intersection of the two groups. The $y = 6$ coordinate indicates which two-dimensional plane is shown in the brain map. Panel B: The figure shows the time-course of activity in the vSt, averaged over the a priori region of interest, during trials when subjects are offered the opportunity to sell a stock trading at a gain. The blue line plots the average activity in trials where subjects decide to realize the gain, while the red line plots the average activity in trials where subjects instead decide not to realize the gain. ** denotes $p < 0.01$, * denotes $p < 0.05$ (paired $t$-test). $t = 0$ corresponds to the instant at which the subject enters his trading decision on a hand-held device.
Trading behavior, ctd.

Summary

- a model in which the investor derives prospect theory utility from *annual trading profits* only generates a disposition effect for some parameter values

- a model in which the investor derives utility from *re- alized gains and losses* delivers a disposition effect more reliably
  - realization utility can be coupled with either time discounting or an S-shaped utility function

Note:

- the trading models we have seen ignore probability weighting
  - in dynamic settings, probability weighting leads to a *time inconsistency* that may be important in some trading contexts
  - but also in casinos (see Barberis, 2012)
Summary

[1]
• the cross-section of stock returns
  – one-period models
  – new prediction: the pricing of skewness
  – probability weighting plays the most critical role

[2]
• the aggregate stock market
  – intertemporal representative-agent models
  – address the equity premium, non-participation, volatility, and predictability puzzles
  – loss aversion plays a key role; but probability weighting also matters

[3]
• trading behavior
  – multi-period models
  – address the disposition effect and other trading phenomena
  – all aspects of prospect theory play a role
Summary, ctd.

• prospect theory is helpful for thinking about asset prices and trading behavior
• probability weighting plays an important role
  – perhaps more important than loss aversion
• while there has been progress, crucial questions remain unanswered
  – what are the principles that should guide the way prospect theory is implemented?
  – how should “gains” and “losses” be defined in any given context?
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Overview

• behavioral finance models aim for psychological realism along three dimensions
  – allow for less than fully rational beliefs
  – use more realistic preferences
  – take account of cognitive limits

• in Lecture Notes 6 and 7, we focus on the second dimension: investor preferences
  – Lecture Note 6: prospect theory
  – Lecture Note 7: ambiguity aversion and other preference hypotheses
Ambiguity aversion

• an important motivation for research on ambiguity aversion is the Ellsberg paradox

The Ellsberg paradox, 1961

• two urns
  – Urn C: 100 balls, 50 Red, 50 Black
  – Urn U: 100 balls, all either Red or Black, unknown distribution of colors

• choose between:
  – bet R1: draw a ball from urn C, get $20 if Red
  – bet R2: draw a ball from urn U, get $20 if Red

• choose between:
  – bet B1: draw a ball from urn C, get $20 if Black
  – bet B2: draw a ball from urn U, get $20 if Black

• the most prominent hypothesis for thinking about the observed behavior is the ambiguity aversion hypothesis
  – people dislike situations where they feel uncertain about the probability distribution of outcomes, i.e. situations of “ambiguity”
**Ambiguity aversion: Models**

There is a large (and complex) literature on models of ambiguity aversion

**I. Multiple priors** (Gilboa and Schmeidler, 1989)

\[
\max_{\text{action}} \min_{\text{models}} E_\pi U(\tilde{X})
\]

- the individual has in mind many possible probability distributions (i.e. models) for future outcomes
  - she chooses an action that maximizes the worst expected utility she can get under any of these probability distributions

**II. Smooth ambiguity** (Klibanoff et al., 2005)

\[
\max_{\text{action}} E_\mu \phi(E_\pi U(\tilde{X}))
\]

- \(\mu\) is a probability distribution over models
  - \(\pi\) is a probability distribution over outcomes, for some model
- when \(\phi(\cdot)\) is concave (convex), get ambiguity aversion (ambiguity seeking)
Ambiguity aversion: Models, ctd.

III. Robust control (Hansen and Sargent, 2007)

• start with a reference model $q$

$$\max_{\text{action}} \min_{\text{models}} (E_{\pi} U(\bar{X}) + \theta R(\pi, q))$$

where $R(\cdot)$ is a distance measure
Ambiguity aversion: Applications

- stock market non-participation
- under-diversification
  - home bias, local bias, own-company stock holdings
  - e.g. see Uppal and Wang (2003)
- the equity premium
- amplification of crises
  - Krishnamurthy (2010)

Note:

- the natural cross-sectional prediction of ambiguity aversion does not seem to hold
  - more “uncertain” stocks do not have higher average returns (Diether, Malloy, Scherbina, 2003)
- we may broaden the scope of ambiguity aversion applications by exploiting research in psychology
Other preference hypotheses

• a large body of evidence suggests that people have a *preference for the familiar*

• a useful version of this is the “mere exposure effect”
  – mere exposure to something makes us like it more than justified based on informational considerations alone

• the mere exposure effect offers a way of thinking about the lack of diversification in household portfolios
  – home bias
  – local bias
  – holdings of own-company stock
BEHAVIORAL FINANCE
Nicholas Barberis, AEA 2017
Lecture Note 8: Bounded Rationality

© Nicholas C. Barberis
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Overview

• behavioral finance models aim for psychological realism along three dimensions
  – allow for less than fully rational beliefs
  – use more realistic preferences
  – take account of cognitive limits
• in this lecture note, we focus on the third dimension
  – frameworks that incorporate “bounded rationality”
Overview, ctd.

Broad theme:

• people have limited capacity for gathering and processing information

Specific topics:

• limited attention
  – underreaction to news
  – buying vs. selling decisions of individual investors

• the $\frac{1}{n}$ heuristic

• nominal illusion

• category-based thinking
Limited attention I

Theme:

- because of limited attention, investors may underreact to news

Underreaction to earnings news

- limited attention may explain post-earnings announcement drift (PEAD)
- two tests:
  - PEAD is stronger for firms that announce earnings at the same time as many other firms (Hirshleifer, Lim, Teoh, 2009)
  - PEAD is stronger for firms that announce earnings on Friday (Della Vigna and Pollett, 2009)
Limited attention I, ctd.

Underreaction to other news: Customer links (Cohen and Frazzini, 2008)

- firms are required to report their major customers
  - investors are slow to recognize that good news for a customer is good news for the associated firm

Underreaction to other news: Demographic news (Della Vigna and Pollett, 2008)

- investors are slow to recognize the impact of demographic shifts on the future profitability of firms with age-sensitive products
Limited attention II

Theme:

- attention is more important for individual investors’ buying decisions than for their selling decisions – why?

- Barber and Odean (2008) show that there is indeed stronger buying interest than selling interest, among individual investors, for attention-grabbing stocks – stocks with extreme returns, high volume, or news announcements
The $\frac{1}{n}$ heuristic

- Benartzi and Thaler (2001) hypothesize that people diversify in a naive way
  - given $n$ investment options, they invest $\frac{1}{n}$ in each investment option
- in a laboratory setting, ask people to allocate between:
  - a bond fund and a stock fund
  - a bond fund and a balanced fund
  - a stock fund and a balanced fund
- find very different equity allocations: 54%, 35%, and 73%, respectively
The $\frac{1}{n}$ heuristic, ctd.

- Benartzi and Thaler (2001) predict that, in 401(k) plans that offer more stock funds, we should see greater allocation to stocks
  
  – they confirm this in a sample of 170 large plans

<table>
<thead>
<tr>
<th>Relative number of equity-type investment options</th>
<th>Mean relative number of equity investment options</th>
<th>Mean allocation to equities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>54</td>
<td>0.37</td>
</tr>
<tr>
<td>Medium</td>
<td>54</td>
<td>0.65</td>
</tr>
<tr>
<td>High</td>
<td>54</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>p-value (ANOVA test)</strong></td>
<td><strong>p-value (ANOVA test)</strong></td>
<td><strong>p-value (ANOVA test)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 6—The Relative Number of Equity-Type Investment Options and Asset Allocation Using the MMD Sample of 401(k) Plans (as of 6/30/96)
Nominal illusion

• when valuing an asset we should:
  – discount nominal cash flows using a nominal discount rate
  – or, discount real cash flows using a real discount rate

• the nominal illusion hypothesis says that investors make the mistake of discounting real cash flows at a nominal rate
  – may help to explain stock market movements in the 1970s and late 1990s
Category-based thinking

• a fundamental feature of human thought is that we put things into categories
  – and form beliefs at the category level

• Barberis and Shleifer (2003) use this idea as the basis for a behavioral theory of *comovement*
  – one that can make sense of some puzzling empirical facts
Comovement

The traditional view

• in an economy with rational investors and without frictions, price equals fundamental value

\[ P_1 = FV_1 = \frac{E(CF_1)}{1 + r_1} \]
\[ r_1 = r_f + (\text{risk aversion})(\text{risk}_1) \]
\[ P_2 = FV_2 = \frac{E(CF_2)}{1 + r_2} \]
\[ r_2 = r_f + (\text{risk aversion})(\text{risk}_2) \]

Therefore prices can comove (i.e. price changes can be positively correlated) because:

• there is a common factor in news about future cash flows

• there is a common factor in discount rate changes
  – changes in \( r_f \)
  – changes in risk aversion
  – a common factor in news about risk
Numerous facts about comovement are hard to square with the traditional view.

Example: *Value stocks and small stocks*

- there are strong common factors in the returns of small stocks and value stocks.
- the simplest traditional explanation attributes this to common factors in news about earnings of small stocks and value stocks.
- such common factors in earnings exist, but are only weakly linked to return factors. 
  – Fama and French (1995)

Example: *Commodities*

- commodity prices move together strongly even when their fundamentals are unrelated.
  – Pindyck and Rotemberg (1990)
Category-based thinking, ctd.

Barberis and Shleifer (2003)

• two assumptions:
  – investors group stocks into *categories* (e.g. value, growth, small-cap, large-cap, etc.)
  – investors’ beliefs about the future return on a category is a weighted average of its past returns

• obtain numerous predictions
  – e.g. excessive comovement within a category

• applications
  – comovement of value stocks, small stocks
  – comovement of commodities
Category-based thinking, ctd.

*Barberis, Shleifer, Wurgler (2005)*

- predict, based on Barberis and Shleifer (2003), that stocks added to the S&P 500 will comove more with the index after addition than before
  - find confirming evidence

![Graph 1: 1988-2000 additions with matches (N = 153)](image1)

![Graph 2: 1988-2000 matching firms](image2)
Course structure

I. Introduction

• overview (LN 1)

II. Background

• empirical facts (LN 2)
• limits to arbitrage (LN 3)
Course structure, ctd.

III. Models and applications

IIIA. Models of investor beliefs

• extrapolation (LN 4)
• overconfidence and other belief biases (LN 5)

IIIB. Models of investor preferences

• prospect theory (LN 6)
• ambiguity aversion and other preference specifications (LN 7)

IIIC. Models of bounded rationality

• bounded rationality (LN 8)

IV. Conclusion

• summary and conclusion (LN 9)
Progress in behavioral finance

- Behavioral finance tries to make sense of the behavior of investors, firms, and markets using frameworks that are *psychologically more realistic* than their predecessors.

- The field has been successful on some dimensions:
  - Shows that a small number of simple, intuitive ideas can explain a wide range of facts.
  - And makes predictions that have found support in the data.

- It is ambitious in its applications:
  - Addresses fundamental topics in finance.
  - E.g. asset market fluctuations, bubbles, cross-section of returns, volume, security issuance, M&A, ...

- Behavioral assumptions that appear particularly helpful:
  - Extrapolation
  - Overconfidence
  - Prospect theory
Progress, ctd.

Extrapolation

- excess volatility and predictability in aggregate asset classes
- momentum and reversals
- bubbles

Overconfidence

- trading volume; misvaluation
- in conjunction with short-sale constraints, overvaluation together with heavy trading

Prospect theory

- high average return on the aggregate stock market, low average return on positively-skewed assets (e.g. IPOs)
- disposition effect, momentum
Progress, ctd.

• in the 1990s, researchers worried about a “lack of discipline” in behavioral finance
  – they feared a profusion of psychological ideas, each designed to “explain” one fact

• this concern has proven unfounded
  – in the 1990s, the center of gravity in behavioral finance was in three ideas: extrapolation, overconfidence, and prospect theory
  – the field’s center of gravity remains in these three ideas today
The future

• thus far, behavioral finance has engaged with the “judgment and decision-making” portion of psychology
  – but this is only a small part of psychology

• in future years, may see broader engagement with other areas of psychology
  – and with areas of neuroscience

• behavioral finance has made progress by developing models based on its ideas
  – showing that these models explain a range of facts
  – and that they make testable predictions

• expect this process to continue, with, hopefully, continued success for the field
REFERENCE LIST for:

BEHAVIORAL FINANCE
Asset Prices and Investor Behavior
AEA, January 2017

Nicholas Barberis
Yale University

Books


Survey papers


Lecture Note 2: Empirical Facts


**Lecture Note 3: Limits to Arbitrage**


**Lecture Note 4: Extrapolation**


**Lecture Note 5: Overconfidence**


**Lecture Note 6: Prospect theory**


**Lecture Note 8: Bounded rationality**


