On Kernels and Sentiment

Hersh Shefrin
Mario L. Belotti Professor of Finance
Leavey School of Business
Santa Clara University
Santa Clara, CA 95053
Email: hshefrin@scu.edu

This version: July 2000

[1],

[1] I thank Roni Michaely, Oded Sarig, Simon Benninga, Jacob Boudoukh, Eugene Kandel, Zvi Weiner, Itzhak Venezia, David Hirshleifer, Bhaskaran Swaminathan, Terry Odean, Ming Huang, Peter Carr, Dilip Madan, Frank Milne, Joseph Langsam, Peter Cotton, and Arturo González. I am also grateful to participants at seminars presented at Queen's University, the University of Michigan, Tel Aviv University, Hebrew University, and Stanford University. Financial support from Santa Clara University and the Dean Witter Foundation is gratefully acknowledged.

Abstract

Heterogeneous beliefs are ubiquitous. Heterogeneity can, but need not, affect the pricing of assets. In this paper I develop a measure of sentiment and show that the log-pricing kernel can be decomposed into two stochastic processes, one pertaining to fundamentals and the other to sentiment. Hence, prices are efficient if and only sentiment is uniformly zero. When sentiment is nonzero, I demonstrate that heterogeneity can lead to "smile" effects both in the graph of the kernel and in option prices, and "frown" effects in mean-variance portfolios. Nonzero sentiment distorts the mean-variance frontier from its "efficient" position, thereby giving rise to behavioral betas. In addition, nonzero sentiment interferes with the expectations hypothesis of the term structure, and can affect the volatility of the return to the market portfolio, depending on traders' risk tolerance spectrum. I also argue that heterogeneity can cause the representative trader to have different characteristics than the individual traders.

1 Introduction

A common assumption in asset pricing is that traders are homogeneous. Of course in the real world traders differ from one another in respect to beliefs, risk tolerance, and patience.[2] In this paper I use a *general equilbrium* framework to analyze the impact of heterogeneity on asset pricing. I demonstrate that there are conditions where asset pricing under heterogeneity is dramatically different from asset pricing under homogeneity. I also demonstrate that there are (other) conditions where asset pricing is the same under heterogeneity as under homogeneity. In this respect, I discuss why the former conditions are more robust than the latter.

Heterogeneity is particularly important when it comes to market efficiency. For this reason, I develop a new variable to measure the degree of market sentiment. I selected "On Kernels and Sentiment" as the title for this paper because the pricing kernel and sentiment are the core concepts for understanding how asset pricing is impacted by heterogeneity. The pricing kernel serves as the basis for the pricing of *all* assets. Sentiment measures the aggregate error (in the market) that drives a wedge between price and fundamental value. The central result in the paper concerns the relationship between the two concepts. The result states that the log-kernel can be decomposed into two stochastic processes, one pertaining to fundamentals and the other to sentiment. Note that the result implies that prices are efficient if and only if sentiment is always zero.[3] In other words, heterogeneity is compatible with market efficiency as long as the heterogeneity is associated with zero sentiment.

When sentiment is nonzero, heterogeneity affects the term structure of interest rates, the pricing of options, and the returns to both mean-variance efficient portfolios and the market portfolio. I show why nonzero sentiment:

- induces stochastic volatility into the process governing interest rates, and interferes with the expectations hypothesis of the term structure;
- induces volatility "smiles" and stochastic volatility into option pricing, thereby preventing options from being priced by Black-Scholes in equilibrium;
- induces a "frown" effect into the return to mean-variance efficient portfolios, mirroring

[2] An interesting illustration of the heterogeneous beliefs involves the contrasting views of Glassman and Hassett (1999) who wrote *Dow 36,000*, and Shiller (2000) who wrote *Irrational Exuberance*.

[3] That is, the probability that sentiment is nonzero at some date is zero.

"smile" effects in option pricing, thereby affecting the character of beta; and

- may or may not affect the return distribution to the market portfolio, depending on traders' risk tolerances.

Much of the approach in the paper is new, especially: the formal treatment of sentiment, the decomposition of the log-kernel, and the associated implications for option pricing, term structure of interest rates, mean-variance efficient frontier, and return on the market portfolio. I note that some of my results serve to extend the analysis in Shefrin-Statman (1994), that is based on log-utility.

Shefrin-Statman (1994) show that heterogeneity does not always lead to inefficient prices because instead of distorting prices, traders' errors can be self-canceling. They establish an efficiency condition that involves two terms: (1) the mean discounted trader error; and (2) the covariance between discounted errors and wealth. In their model, prices are efficient if and only if the wealth-scaled mean error[4] and error-wealth covariance sum to zero. Consider two of the implications attached to the Shefrin-Statman efficiency condition. First, when individual errors (1) average to zero across the trader population; and (2) are uncorrelated with wealth, then prices are efficient. One of the strongest empirical implications stemming from the behavioral decision literature is that individual errors are *systematic*, meaning that the mean trader error is nonzero. Second, the Shefrin-Statman efficiency condition tells us that prices can be inefficient even when errors are nonsystematic. This occurs when the mean error is zero, but the error-wealth covariance is nonzero. Note that a nonzero error-wealth covariance implies that the errors of wealthier traders count for more than the errors of less wealthy traders.

In most of the paper I intentionally leave unspecified the exact errors commited by individual traders. Instead, I focus on effects stemming from heterogeneity alone. There are two reasons for doing so. First, as I indicated in the preceding paragraph, heterogeneity itself may be the cause of mispricing. Second, in the behavioral decision studies that document particular errors and biases, those errors and biases are group averages. They do not afflict each and every individual. Many subjects in these studies do not even commit the particular error under investigation. In other words, behavioral studies feature considerable heterogeneity in subjects' responses. I contend that it is important to have a theory general enough to accommodate individual differences in errors. The main results in the paper[5] hold

[4]The mean error is multiplied by market wealth.

[5]Such as those pertaining to the decomposition of the kernel.

irrespective of the precise errors individuals make. And these results are hardly nullified by sharpening the assumptions to accommodate specific types of errors. Indeed, at the end of the paper, I discuss the effect of incorporating specific behavioral features into the model.

There is a message in this paper for those who use representative trader models of asset pricing, whether rationally-based pricing models or behaviorally-based models. Those who use a rationally-based representative trader approach tend to assume a representative trader whose features resemble some unspecified average individual trader. Pertinent examples include Lucas (1978), Mehra-Prescott (1985), and Whitelaw (2000). Of particular note is Bates' (1996) survey article about option pricing. Bates points out that when both volatility and interest rates are stochastic, options *cannot* generally be priced using the arbitrage method that underlies Black-Scholes. Rather, an equilbrium method must instead be used, where the method Bates describes centers on a representative trader: See section 2 of his article. Another example is the option pricing model developed by David and Veronesi (1999). Their model features a representative trader who learns about the true value of the drift term associated with the fundamental uncertainty governing the system.

My point is that in a heterogeneous trader model, the representative trader does not typically resemble any of the individual traders. The aggregation process is more complex than commonly assumed. For example, aggregating traders with constant relative risk aversion utility functions does not always lead the representative trader to exhibit constant relative risk aversion. And aggregating traders whose beliefs are log-normal does not always lead the representative trader to hold log-normal beliefs. Hence, researchers who rely on representative trader models may succumb to the behavioral bias known as *representativeness*.

Those who use a behaviorally based representative trader approach tend to assume a representative trader who commits one or more of the standard errors identified in the behavioral decision literature. The articles by Barberis, Shleifer, and Vishny (1998) and Daniel, Hirshleifer, Subrahmanyan (1998) are typical in their focus on the mean error. But in a world of heterogeneous errors, that being the world we live in, the collective error in the market need not conform with any particular error identified in the behavioral decision literature. In this respect, my treatment of sentiment is different from the approach taken in these articles.

I have organized the paper as follows. In Section 2, I set the stage with a short discussion about heterogenous beliefs. I briefly discuss instances of heterogeneus beliefs in both behavioral studies and in practice, following which I review the asset pricing literature on heterogenity. My strategy for the remainder of the paper is twofold, first highlighting the main features in the Shefrin-Statman's (1994) log-utility framework, and second, generalizing their results.

In sections 3 through 5, I use a standard binomial option pricing model to describe the main results in Shefrin-Statman (1994). By invoking a limit argument to achieve a continuous time diffusion process, I am able to show when and why Black-Scholes breaks down: by this I mean when and why heterogeneity causes a difference between equilibrium option prices and their corresponding Black-Scholes values.

Sections 6 through 13 are the heart of the paper, where I describe the general framework (section 6), formally define sentiment (sections 7 and 8), establish the log-kernel decomposition theorem (section 8), discuss robustness issues involving my choice of utility function (section 9), and develop the implications of nonzero sentiment for the returns to mean-variance efficient portfolios (section 10), interest rates (section 11), option prices (section 12), and the market portfolio (section 13).

In section 14, I discuss specific issues associated with phenomena described in the behavioral decision literature, and provide a short conclusion in section 15.

2 Heterogeneous Beliefs

One of the most striking features in psychological studies of prediction in the heuristics and biases literature is the wide dispersion of beliefs. A good example is the work of De Bondt (1993) on trend extrapolation and overconfidence in predicting the future value of the S&P 500. One way of measuring the extent of heterogeneity is by measuring the coefficient of variation across forecasts, that is the standard deviation of the different forecasts, divided by the mean forecast of the group. I have replicated De Bondt's study with two separate groups of MBA students and one group of investment professionals.[6] For students the coefficient of variation is about 12.8 percent, and for investment professionals it is a little lower at 11.1 percent. Hence there is considerable heterogeneity within groups, but little difference in the degree of heterogeneity between groups.

[6]The investment professionals were portfolio managers, security analysts, and administrative staff members at a hedge fund. De Bondt does not report the coefficient of variation, though the degree of heterogeneity in his original results can be seen from the t- statistics he reports.

Consider an instance of heterogeneity from a well-known setting that does not stem from an academic study. Starting in 1983, the television program Wall $treet Week with Louis Rukeyser has been collecting twelve-month forecasts from panelists for the year-end closing value of the Dow Jones Industrial Average. I have found that the mean annual coefficient of variation over the period 1983 through 1999 is 8.9 percent.[7] The intertemporal standard deviation, which indicates the extent of variation in the dispersion of forecasts over time, is 2.6 percent.[8] This is especially interesting, given that the composition of panelists is quite stable from year to year. The highest value for the coefficient of variation is 16 percent, and occurred in connection with the forecast for year-end 1988. In other words, disagreement was highest for the prediction made two months after the 1987 crash. Depending on the year, 17 to 27 panelists participated in the Wall $treet Week survey.[9]

Theoretical work on heterogeneity goes back several years and includes Rubinstein (1973), Jaffee and Winkler (1976), Figlewski (1978, 1983), Feiger (1978), Mayshar (1983), Shefrin (1984), Dumas (1989), Harris and Raviv (1993), Benninga and Mayshar (1993, 1997), Detemple and Murthy (1994), Shefrin and Statman (1994), Huang (1996), Basak (2000), Kurz (1997), and Carr and Madan (1997). The articles by Jaffee-Winkler, Figlewski, Feiger, and Shefrin study partial equilibrium models where traders hold heterogeneous beliefs, but share the same tolerance for risk. Mayshar (1983) and Dumas (1989) focus instead on a two-trader general equilibrium model where the traders hold the same

[7]The data is available from Wall $treet Week.

[8]Here is another example. In December 1998, BusinessWeek published the forecasts of 49 Wall Street strategists and analysts for the year-end 1999 values of the Dow Jones Industrial Average, S&P 500, and Nasdaq Composite. The coefficient of variation across the 49 forecasts, expressed in percentage terms, are 17.0, 13.5, and 17.8 respectively. This dataset also includes six-month forecasts. Interestingly, the coefficient of variation is smaller for the shorter forecast horizon (10 percent). A similar feature holds in the Livingston data set that was initially compiled by Philadelphia Enquirer columnist Joseph Livingston, and is now maintained by the Federal Reserve Bank of Philadelphia. The Livingston data set contains forecasts for several variables, such as interest rates and the CPI. The CPI is a less volatile than a stock index, and its forecasts feature smaller coefficients of variation. It is interesting that the coefficient of variation seems to be similar across the range of cases, De Bondt's study, BusinessWeek, and Wall $treet Week with Louis Rukeyser.

[9]In replicating the Wall $treet Week prediction exercises with MBA students, using the actual percentage change in the Dow, I obtained a mean coefficient of variation of 12 percent, with a standard deviation over time of 3.3 percent, not that different from the 9 percent actual mean and 2.5 percent actual standard deviation. Again, there is considerable heterogeneity within groups, but the differences between groups are small.

beliefs, but have different tolerances for risk. Benninga and Mayshar (1993, 1997) extend the Dumas approach to many traders, and focus on how the representative trader serves to aggregate the risk tolerance parameters of the individual traders. Their 1997 paper applies their methodology to analyze the impact of heterogeneity on option prices. Detemple and Murthy develop a continuous time, incomplete market, logarithmic utility model in which traders hold differential beliefs. They analyze how a representative trader aggregates the beliefs of the individual traders. Shefrin and Statman (1994) analyze a model similar to that of Detemple and Murthy, but where markets are complete and time is discrete. There is a parallel literature in noisy rational expectations models that also features heterogeneity. As I discuss below, because these models assume constant absolute risk aversion (CARA-utility), there is no germane role for the error-wealth covariance. However, the noisy rational expectations model has provided important insights into the effects stemming from overconfidence (Odean, 1998).[10]

3 A General Equilbrium Binomial Example

In this section I introduce the basic structure of my model through a familiar framework, the binomial option pricing model. This will enable me to highlight the main results in Shefrin and Statman (1994) upon whose analysis I build.

Consider a binomial option pricing model, where the price of a security Z unfolds over successive dates, $t = 0, 1, 2, \cdots, T$. Let $q_{z,t}$ be the price of Z at date t. At date $t + 1$, the price of Z will be either $uq_{z,t}$, where $u > 1$, or $dq_{z,t}$, where $d = 1/u < 1$. The probability attached to an up-move u is denoted P_u. In the standard binomial option pricing model developed by Cox, Ross, and Rubinstein (1979), the interest rate i is given, and the price of a European call option on a non-dividend paying stock is independent of the value taken

[10] All models make simplifying assumptions and therefore feature weaknesses. For example, the model I use features complete markets rather than the incomplete markets studied by Detemple and Murthy. I restrict attention to a specific family of utility functions rather than deal with the general formulation used by Basak (2000) in his work on representative traders. I assume that every trader's utility function features constant relative risk aversion (CRRA), though there may be differences in the degree of risk aversion across traders. This leads the representative trader's utility function to have a generalized power function form, a feature that builds on a result in Rubinstein (1976) that the representative trader has a CRRA utility function. And I impose a finite time horizon instead of allowing for an infinite horizon. These restrictions bring some cost in terms of generality. But the restrictions also lead to a model where sentiment can be clearly defined, is easily understood, and enters the pricing kernel in a straightforward manner.

by P_u. This is because the state prices $Q_u$and Q_d associated with date 1 are determined by u, d, and i according to:

$$Q_u = [(i - d)/(u - d)]/i$$

$$Q_d = [(u - i)/(u - d)]/i$$

The state prices for subsequent dates have the form $C(t,n)Q_u^n Q_d^{t-n}$, where n is the number of up-moves that occur in t periods, and $C(t,n)$ is the associated binomial coefficient.

To fix ideas, let $u = 1.05$, and $P_u = 0.7$. Let the interest rate be 1.87% so that the gross interest rate is given by $i = 1.0187$. Consider a non-dividend paying stock whose price at date 0 is 4.00, and a European call option on this stock that expires at $t = 2$ and has an exercise price $K = 3.80$. Figure 1 shows that the date 0 price of this option is equal to 0.355.

In the binomial option pricing model, option prices are independent of P_u, the actual probability, or more correctly traders' beliefs about the actual probability. This property carries over in the limit, as the binomial option pricing formula converges to the Black-Scholes formula. See Cox, Ross, and Rubinstein (1979). This raises the question of whether belief heterogeneity is even relevant for option pricing.

The standard binomial option pricing model is a partial equilbrium framework, because the interest rate is given. In contrast, the model in Shefrin-Statman (1994) is a general equilbrium framework where the interest rate is endogenously determined as a function of both u and P_u. This dependence is important, because it provides a link from P_u to option prices. That is, unlike the partial equilbrium framework where option prices are independent of P_u, in the general equilibrium framework option prices vary with P_u, through their impact on the interest rate.

To show how the Shefrin-Statman (1994) framework relates to the standard binomial option pricing model, consider a special case of Shefrin-Statman (1994) where $u = 1.05$, $P_u = 0.7$, $i = 1.0187$, and $T = 5$.[11] Assume that there is a single physical asset that produces a single consumption good at each date. The amount of the consumption available for consumption at date 0 is 1 unit. Thereafter, the amount of consumption will grow stochastically from date to date, either at rate u (with probability P_u) or at rate d (with probability $1 - P_u$.) The market portfolio is a security that pays the value of aggregate consumption at each date.

[11] Figure 1 displays the first four dates.

There are two traders in the model. These two traders initially hold portfolios consisting exclusively of the market portfolio. By assumption, each trader initially holds one half of the market portfolio. There is also a risk-free security available for trade at each date. Because of the binomial character of uncertainty, these two securities will be sufficient to complete the market.

Both traders are assumed to have additively separable preferences, logarithmic utility, and discount factors equal to unity (zero impatience). They also hold beliefs about the branch probability in the binomial tree. Trader 1 assumes that the value of the branch probability P_u is $P_{1,u}$, while trader 2 believes the value to be $P_{2,u}$. Each trader seeks to maximize subjective expected utility subject to the condition that the present value of lifetime consumption be equal to initial wealth. The single budget constraint here stems from markets being complete.

For the moment, consider the case of homogenous beliefs, where $P_{1,u} = P_{2,u} = P_u = 0.7$. There are two key equations from Shefrin-Statman (1994) that are central here, equations (4) and (7).[12] Equation (4) implies that in equilibrium, $Q_u = P_u/u$, and equation (7) implies that

$$i = [P_u/u + (1 - P_u)/d]^{-1}$$

Given the values of u and P_u assumed above, it is easily verified that $i = 1.0187$. To obtain the counterpart of the non-dividend paying stock, consider a security Z that is defined so that it has the same (dividend) payoff as the market portfolio for the four dates 2 through 5 inclusive, but pays no dividend prior to date 2. By constructing the state prices from Q_u, it is easily verified that this security has a price of 4.00 at date 0, and that its price either grows by a factor of u or d in every period before the option expiration date. Therefore, options in the general equilbrium version of the example can be priced using the standard binomial method described above.

4 Heterogeneous Beliefs, Efficient Prices, and Interest Rate Volatility

Shefrin and Statman (1994) develop their framework to explore the implications of heterogeneous beliefs on equilibrium prices. On pages 326 and 327 of their (1994) article, they

[12] I generalize these equations below, in theorems 1 and 4 of this paper.

establish that equilbrium prices can be characterized through the beliefs of a representative trader R, whose tree probabilities are a convex combination of the tree probabilities of the individual traders, where the weights are given by relative wealth. Because the two traders in this example have the same wealth at date 0, the representative trader attaches probability

$$P_{R,u} = (P_{1,u} + P_{2,u})/2$$

to the occurrence of an up-move at the end of date 0. The probability that the representative trader attaches at date 0 to two successive up moves, occuring at the end of date 0 and the end of date 1 respectively, is:

$$P_{R,u}(2) = (P_{1,u}^2 + P_{2,u}^2)/2$$

which is the (relative wealth-weighted) average of the two traders' binomial probabilities attached to the node in question. Note that the representative trader does *not* attach a probability to this node using the equation:

$$P_{R,u}^2 = ((P_{1,u} + P_{2,u})/2)^2 = P_{R,u}(2)/2 + P_{1,u}P_{2,u}/4$$

In order to illustrate the result just described, let $P_{1,u} = 0.8$ and $P_{2,u} = 0.6$, but maintain the assumption that the objective probability $P_u = 0.7$. In this case it will turn out that the one-step probability $P_{R,u}$, the simple average of 0.8 and 0.6, will actually equal the objective probability $P_u = 0.7$. However, such equality will not occur for the case of the $n-$step probabilities, for $n > 1$. For example the representative trader will attach probability 0.50 to the occurrence of two successive up-moves beginning at date 0, whereas the objective probability of this event is actually $0.7^2 = 0.49$. Indeed, the difference in value between these values implies a fat tail in the representative trader's probability density function. See figure 2 which provides an illustration of the tree.

Lying at the heart of all the issues I discuss in this paper is the question of when heterogeneity causes prices to be inefficient. Shefrin-Statman (1994) establish a necessary and sufficient condition for efficient prices in their model, (theorem 2, p. 329.) In the context of the current example, their theorem stipulates that prices are efficient if and only if the relative wealth-weighted average of the trader probability errors, a dot product, is zero at every node.[13] In the previous example with heterogeneous beliefs, it is straightforward to see that this weighted average is zero for the date 1 nodes. The error in the probability that

[13] Since the probability density function of the representative trader is a wealth weighted convex combi-

trader 1 attaches to an up-move at date 1 is $0.8 - 0.7 = 0.1$. Similarly, the corresponding error that trader 2 attaches to the same event is $0.6 - 0.7 = -0.1$. Since both traders have relative wealth levels equal to 1/2, the relative wealth-weighed error in the market is $(0.1 - 0.1)/2$, which equals zero. However, the weighted average is nonzero for nodes that occur later in the tree. For instance, for two successive up-moves leading to date 2, trader 1's error is 0.025, whereas trader 2's error is -0.022, so the weighted sum is 0.0015.

Theorem 2 from Shefrin-Statman (1994) is fundamental for understanding how heterogeneous beliefs impact asset pricing. The theorem also provides an alternative (equivalent) necessary and sufficient condition for efficient prices, one that yields additional insight. This alternative condition is based on two terms. The first term is the average trader error. When the average trader error is zero, then traders are not subject to systematic biases, even though individual traders do commit errors. The second term is the covariance between traders' errors and their wealth levels. When this covariance is zero, then trader errors are distributed uniformly across the trading population, rather than being concentrated. Notice that this is the case at date 0: since the two traders have equal intial wealth, there is no variation in wealth, and hence zero covariation with errors.

Theorem 2 in Shefrin-Statman (1994) implies that there are two main reasons why prices can be inefficient, at least in their model. First, traders may commit systematic errors, thereby leading the average trader error to be nonzero. Second, prices can be inefficient even though the average error may be nonzero, because errors are concentrated. The error of a wealthy trader can exert a greater impact on market prices than less wealthy traders. Formally, what theorem 2 states is that prices are efficient if and only if the sum of the error-wealth covariance, and the wealth-normalized mean trader error is zero.[14]

One of the most pronounced differences between the character of asset pricing in the homogeneous beliefs binomial model, and the heterogenous beliefs binomial model, concerns interest rates. Both Shefrin-Statman (1994) and Detemple-Murthy (1994) analyze the impact that heterogeneity exerts on the short-term interest rate. Consider the case when

nation of the density functions of the individual traders, the aggregate error is a wealth weighted convex combination of the individual errors. That is why market efficiency corresponds to the dot products being zero, node by node. Theorem 2 is actually stated in terms of the dot product of trader errors and absolute wealth, rather than relative wealth. The former dot product is equal to the product of the latter dot product and total market wealth.

[14]The theorem is actually stated in terms of discounted errors; however, discounting is not germane in this example. The mean trader error is normalized by multiplying it by total market wealth.

homogenous beliefs lead the short-term rate to be constant over time. Shefrin-Statman emphasize that when their dot product efficiency condition fails, heterogeneity causes the short-term interest rate to be stochastic. To see why this is so, consider four cases. In three of these cases, the two traders agree about the value of P_u. In the first case, both correctly believe its value to be 0.7. In the second case, both believe its value to be 0.8. In the third case, both believe its value to be 0.6. And in the fourth case, trader 1 believes its value to be 0.8 while trader 2 believes its value to be 0.6. Some computation shows that the interest rate in case 1 is a constant[15] 1.87%, in case 2 it is a constant 2.9%, while in case 3 it is a constant 0.87%.

And what will the interest rate be in case 4, where the two traders disagree? To answer this question, compute the discount factors associated with each of the interest rates above. For 2.9%, the one period discount factor (bond price) is $1/1.029 = 0.9719$. For 0.87%, the discount rate is 0.9914. The discount rate in case 4 will be a convex combination of the discount factors 0.9719 and 0.9914, with weights given by relative wealth.

Why? Recall that the probability beliefs of the representative trader are a relative wealth-weighted convex combination of the beliefs of the individual trader. In addition, equation (4) of Shefrin-Statman (1994) indicates that the equilibrium value of a state price is a function of a ratio, the ratio of the representative trader's probability to the (gross cumulative) consumption growth rate. In combination, these two features imply that in equilbrium, asset prices generally can be expressed as weighted sums of asset prices derived from corresponding homogeneous cases.

At date 0, the relative wealth levels are 0.5, so the equilibrium one-period interest rate is 1.87%, the same value as in case 1. However, because the traders disagree about the value of P_u, they bet against each other on the date 0 market. Trader 1 is more optimistic than trader 2. As a result, trader 1 bets more aggressively on the occurrence of an up-move leading to date 1 than trader 2. If an up-move does occur in the first period, relative wealth will shift from trader 2 to trader 1. As a result, trader 1's beliefs will exert more of an impact on pricing on the date 1 market, and the interest rate will climb above 1.87% (in the direction of 2.89%). In this specific example, an up-move in the first period results in trader 1 holding 57% of overall wealth, and trader 2 holding the residual. In consequence, the one-period interest rate at date 1 rises from 1.87% to 2.01%. I note that if we condition on an up-move at the end of date 0, then the conditional error-wealth covariance terms will

[15] Constant means nonstochastic and time-invariant.

no longer be uniformly zero along the tree.[16]

In the above discussion, I have focused on a case where heterogeneity interferes with market efficiency. Consider next a case where this is not so, where prices are efficient, even though traders hold heterogeneous beliefs. Keep in mind that prices are efficient when the relative wealth-weighted sum of trader errors is zero. In addition, if the objective stochastic process is I.I.D.,[17] interest rates will be constant over time. To demonstrate a case with these features, consider a modification to the previous example.[18]

At date 0, let trader 2 attach a probability of 0.34 to two successive up-moves, instead of 0.36. With this modification, the two traders' errors will cancel themselves, at least for this node. Indeed, for any node, we can always solve for the value of trader 2's error that would cancel that of trader 1. Such a solution will lead to efficient prices. Shefrin-Statman call this a knife-edge case that occurs when the error-wealth covariance has the same absolute value, but opposite sign, as the product of the average error and total wealth.[19] Readers should also be aware that the knife-edge case can feature perverse probability revisions. For instance, the probability that trader 2 attaches to the second up-move, conditional on the first, is 57%, *less* than the value of 60% associated with the prior transition.[20] Indeed, along the branch of successive up-moves in the modified example, trader 2 will continually reduce the conditional probability he attaches to the next up-move.[21]

---

[16]Uniformly zero means zero at every node in the tree.

[17]Indendently and identically distributed.

[18]The modification is based on the discussion on page 335, Shefrin-Statman (1994).

[19]In the numerical example described in this paragraph, the average error and error-wealth covariance are both zero at date 0. But if we condition on nodes at later dates, both become nonzero, with equal absolute values and opposite signs.

[20]This occurs because trader 2's error needs to be even more pronounced in order to offset the wealth shift that occurs in the wake of the realization at the end of date 0. In addition, the branch probabilities associated with self-cancelling errors are highly node-dependent. Hence in general, they cannot be determined independently of wealth.

[21]That is, the direction of trader 2's associated probability revision is opposite to the frequency with which the up-move state occurs, conforming with the error known as "gambler's fallacy."

5 Heterogeneous Beliefs and Black-Scholes

In a general equilibrium framework, option prices reflect the beliefs of traders. Consider whether this feature has any implications for whether the Black-Scholes formula coincides with equilibrium option prices. Specifically, does heterogeneity cause the equilibrium values of options to differ from their corresponding Black-Scholes values? The answer to this question is yes, but only under the particular conditions I describe in the next paragraph.

When the efficiency condition in theorem 2 of Shefrin-Statman (1994) holds, heterogeneity has the same pricing implications as homogeneity. Hence, if equilibrium option prices conform with Black-Scholes under homogeneous beliefs, they will also conform with Black-Scholes under heterogeneous beliefs. However, consider what happens in the binomial framework when the Shefrin-Statman efficiency condition fails, no matter what objective value of P_u is selected. In this case, no set of homogeneous beliefs lead to the same prices as occur under heterogeneus beliefs. As I show below, this has important implications for Black-Scholes.

In the previous section, I explained why heterogeneity causes the short-term interest rate to be volatile. In turn, the volatility of short-term interest rates implies that the one-period conditional state prices do not remain invariant over time. Notably, this disrupts the usual limiting argument developed by Cox, Ross, and Rubinstein (1979), where the Black-Scholes pricing equation is achieved as a limiting case of the binomial option pricing formula. Put another way, heterogeity can prevent the conditions necessary for Black-Scholes pricing from holding.[22]

In this section, I extend the example discussed in the preceding section to discuss how options are priced as the discrete model converges to a continuous time diffusion process. I demonstrate that in the example, equilibium option prices are not generally given by the Black-Scholes formula. I also show that equilibrium option prices display volatility "smile" effects. Moreover, the smile pattern for call options need not be the same as the smile

[22]Heterogeneous beliefs do not prevent the option from being priced by arbitrage. However, the binomial distribution for state prices that gets used in the standard binomial option pricing model does not apply. The binomial property will fail because the state prices in the standard framework stay the same over time, but under heterogeneous beliefs, they vary. And remember, Black-Scholes emerges from the binomial framework because by the central limit theorem the binomial distribution converges to the normal. Heterogeneous beliefs will stand in the way of that argument when we seek to apply the central limit theorem in the manner of Cox, Ross, and Rubinstein (1979): see the middle of page 252 of their article.

pattern for put options.

The Black-Scholes formula C_{BS} for the price of a call option is:

$$C_{BS}(S, K, \sigma, t, r) = SN(d_1) - Ke^{-rt}N(d_2) \tag{1}$$

where

$$d_1 = [\ln(S/K) + (r + \sigma^2/2)t]/\sigma\sqrt{t}$$

$$d_2 = d_1 - \sigma\sqrt{t}$$

In the usual notation, S stands for the intial price of the asset underlying the call option, and that is how I use it in this section.[23] K is the strike price, σ denotes the return standard deviation of the underlying asset, t is the time to expiration, and r is the continuous compounding rate of interest.

Consider what happens when we follow the standard procedure of achieving a continuous time framework as a limit of the discrete binomial model. I begin with the standard parameterization of u, d, the (gross) interest rate i and its associated continuously compounded rate r, and branch probability $p = P_u$.

$$u = e^{\sigma\sqrt{t}}$$

$$d = 1/u$$

$$i = e^{r\Delta t}$$

$$p = [e^{\mu\Delta t} - d]/(u - d)$$

n, the number of stages in the binomial process, is set equal to $1/\Delta t$, where Δt is the length of time in a single stage of the binomial process. Notice that $u, d, i,$ and p are implicitly functions of n, through their dependence on Δt.

The argument presented by Cox, Ross, and Rubinstein (1979) implies that for fixed μ, σ and r, if we set $\Delta t = 1/n$, and let n go to infinity, then the return distribution associated with the limiting binomial process converges to a lognormal, and the binomial call option pricing formula converges to C_{BS}, the Black-Scholes formula. In my example, heterogeneous beliefs lead the short-term rate i to be stochastic, rather than fixed. This is why the limiting argument that leads to Black-Scholes breaks down.

[23]Elsewhere in the paper, I use q_z for the initial price, and use S as a symbol for the number of states that can occur at any date.

There is still a limiting argument to be made. The model developed by Detemple and Murthy (1994) is a continuous time counterpart to the discrete time version in Shefrin-Statman (1994). However, the continuous time version need not feature options being priced by Black-Scholes. Recall that in the previous section, I mentioned that one can price all assets by taking a convex combination of two components. The first component is the price that would emerge if both agents accepted $P_{1,u}$ as the branch probability. The second component is the option price that would emerge if both agents accepted $P_{2,u}$ as the branch probability. This implies that in continuous time, a European option on a non-dividend paying stock, can be priced as a convex combination of Black-Scholes functions. And that generally produces a different value from value taken by C_{BS}.

How does the argument work? Begin with two values for μ, μ_1 and μ_2. Then define two branch probabilities

$$p_1 = [e^{\mu_1 \Delta t} - d]/(u - d)$$

$$p_2 = [e^{\mu_2 \Delta t} - d]/(u - d)$$

Develop two limiting processes, one corresponding to μ_1 and the other corresponding to μ_2. These will be limits of general equilibrium binomial frameworks where the interest rates will be endogenous. This will give rise to two different instantaneous interest rates, r_1 and r_2, one to be used in a Black-Scholes homogeneous belief economy corresponding to μ_1 and the other in a homogeneous belief Black-Scholes economy corresponding to μ_2. At $t = 0$, in the limit, the option can be priced as a weighted average of the two Black-Scholes values.

The preceding argument implies that in the limiting form of the previous binomial example, the price of a call option at the initial date converges to:

$$C_{eq} = [C_{BS}(S, K, \sigma, t, r_1) + C_{BS}(S, K, \sigma, t, r_2)]/2$$

In general, the components of the above expression will be weighted by relative wealth. However, in my example, initial wealth levels are assumed equal for the two traders.[24]

A convex combination of Black-Scholes functions, where each function uses a different interest rate, will not typically lead to Black-Scholes pricing based on the equilibrium

[24] I want to emphasize that this pricing feature is part of both the discrete framework developed by Shefrin-Statman (1994) and the continuous time framework developed by Detemple-Murthy (1994). Neither article makes the preceding point about option prices explicitly, but the result follows directly from both.

instantaneous interest rate in the heterogeneous beliefs economy. That is, stochastic interest rates are as much a part of the continuous time framework as the discrete time framework.

Consider whether the equilibrium price for options will be Black-Scholes. To address this question, focus on a stock that pays no dividends, whose return is lognormally distributed, and whose return standard deviation is 9.76%. Let the current value of the stock be 4.00. Suppose that we focus attention upon call options on this stock that expire in one year. Suppose further that there are two traders: trader 1 believes that $\mu_1 = 5.95\%$, and trader 2 believes that $\mu_2 = 1.95\%$.

Were the beliefs of trader 1 to be held by both traders, the equilibrium value of r would be 5%. This can be seen by computing the branch probability p, and solving out the binomial model to obtain i, from which the continuous compounding rate r can be inferred. In this case, meaning homogeneous beliefs where $\mu = 5.95\%$, Black-Scholes can be used to price all European options on the stock. For example, the call option with an exercise price of 3.75 would be priced at $C_{BS}(4, 3.75, .0976, 1, .05) = 0.455$.

However, traders disagree about the expected return on the stock. If both traders held the beliefs of trader 2, the equilibrium value of r would be 1%. In this case, the call option with exercise price 3.75 would have a Black-Scholes value of $C_{BS}(4, 3.75, .0976, 1, .01) = 0.355$.

In view of the preceding argument, the instantaneous interest rate at $t = 0$ will be obtained as follows. To find the equilibrium value of r, set e^{-rdt} equal to the average of the discount terms $e^{-r_1 dt}$ and $e^{-r_2 dt}$ and let dt approach zero. By invoking the Maclaurin series $e^x = 1 + x + x^2/2! + \cdots$, we can conclude that $r = (r_1 + r_2)/2$. This implies that the equilibrium value of r will be 3%.

The equilibrium value of the call option will be $C_{eq} = 0.395$, the average of $C_{BS}(S, K, \sigma, t, r_1)$ and $C_{BS}(S, K, \sigma, t, r_2)$.

We can certainly evaluate the Black-Scholes value C_{BS} for the above call option, since we know the values of its arguments, those being the stock price, exercise price, return standard deviation, expiration period, and r (3%). If we do this, we will obtain a Black-Scholes value of $C_{BS}(4, 3.75, .0976, 1, .03) = 0.394$–close to, but not the same as, the equilibrium value of 0.395. Note that this difference does not present an arbitrage opportunity because it is the equilibrium price that is consistent with the absence of arbitrage opportunities, not the Black-Scholes formula that is predicated on spanning under time invariant interest rates.

Consider an example with more extreme values than those discussed above. Specifically, let $\mu_1 = 59\%$ and $\mu_2 = -41\%$. These values lead $r_1 = 50\%$, $r_2 = -50\%$, and $r_{eq} = 0\%$. In addition, let $\sigma = 30\%$.[25]

Figure 3a shows how the four call option prices discussed in this example vary as a function of K. The top curve in figure 3a pertains to the case $r_1 = 50\%$, while the bottom curve pertains to the case $r_2 = -50\%$. The curves in the middle are for the equilibrium option prices (solid curve), and Black- Scholes prices (dashed curve). Figure 3b provides another view of how the difference between the equilibrium call option price and Black-Scholes price varies as a function of the exercise price K. Notice that the pattern is cyclical, and is negative for low values of K.

The Black-Scholes formula for the price of a put option is:

$$P_{BS}(S, K, \sigma, t, r) = Ke^{-rt}N(-d_2) - SN(-d_1)$$

The equilibrium price of a put option can be obtained in the same manner as a call option, with an analogous expression

$$P_{eq} = [P_{BS}(S, K, \sigma, t, r_1) + P_{BS}(S, K, \sigma, t, r_2)]/2$$

Figures 4a and 4b are the counterparts to figures 3a and 3b.

Consider what happens when, for an interval of exercise prices, we infer the implied Black-Scholes volatilities from the equilibrium prices of options. To do so, we solve

$$C_{BS}(4.00, K, \sigma, 1, .03) = C_{eq}$$

and

$$P_{BS}(4.00, K, \sigma, 1, .03) = P_{eq}$$

for σ as implicit functions of K. Figure 5 illustrates the nature of the volatility patterns associated with these implicit functions.

Notice several features about the volatility patterns. First, the implied volatilities are different for calls than for puts. Second, neither pattern is flat.[26] In a world where

[25] The choice of extreme values only serves to make the underlying relationships more salient. Moreover, although it is not usual to discuss negative interest rates in option pricing models, negative interest rates are consistent with the Black-Scholes framework.

[26] Although Figure 5 may not illustrate the "U-shape" that led to the term "smile," a smile is now generally understood to mean not-flat.

Black-Scholes holds, both curves would coincide with one another and be flat. Third, the implied volatility lies above the actual volatility for most of the range, including the case when options are at-the-money. Fourth, the implied volatility may be undefined at low exercise prices, particularly in the case of call options.

The final topic I discuss in this section is stochastic volatility. In the general equilibrium binomial example above, heterogeneity alters the the representative trader's probability density, which in turn alters the return standard deviation of the asset underlying the option. Note that this is the only variable that causes the return standard deviation to vary because the price of the asset at any node is invariant to traders' probability beliefs. At the same time, let me note that in the above binomial limit, volatility becomes constant. This is best seen by considering what happens in the limiting process that leads the binomial process to converge to a continuous time diffusion process. The analysis in Cox, Ross, and Rubinstein (1979) implies that the branch probabilities P_u and Q_u both converge to 1/2 in the limit (p. 249). This implies that there is little room for the representative trader's branch probabilities to move during a short interval. Hence in the limit, u alone determines volatility. But since the value of u is fixed at each n, volatility is virtually constant for large n. Hence volatility is constant in the limit as the binomial process converges to a diffusion process. This means that in my example, the failure of Black-Scholes stems entirely from stochastic interest rates. However, in the general multinomial framework I develop below, the limiting diffusion process can feature stochastic volatility.

6 The General Framework

The model in Shefrin and Statman (1994) is based on log-utility. In the discrete time example discussed in the previous sections, a further restriction was added, namely a binomial structure. In this section, I present the model to be used in the remainder of the paper. This model generalizes the log-utility assumption to the more general condition of CRRA-utility, and generalizes the binomial condition described earlier to any finite discrete stochastic process. In the discussion below, heterogeneity refers not only to beliefs, but to risk tolerance and impatience as well.

Consider a financial market with H individual traders. Time is discrete, with a set of dates indexed $t = 0, 1, 2, \cdots, T$. At the beginning of each date, new information s is revealed. Call s a state, and assume that it belongs to a finite set $S = \{s_i\}$. The binomial case occurs when S has two elements. Let $s^t \in S$ denote the state revealed at date t. The public information at the beginning of t is denoted by the *trajectory* $x_t = (s^0, s^1, \cdots, s^t)$. That is, uncertainty unfolds according to a tree whose nodes are the date event pairs $\{x_t\}$. Let $\Pi(x_T)$ denote the objective probability attached to the occurrence of x_T. I assume that the probability attached to a trajectory x_t is derived from the terminal node density $\{\Pi(x_T)\}$ as follows: $\Pi(x_t) = \sum \Pi(x_T)$ where a terminal node x_T is in the summation if and only if x_t is an ancestor node of x_T.

In this section, I assume that all information is held in common, and trading is costless. At the outset of date 0, trader h holds an initial portfolio ω_h. If h holds ω_h through date t, and date-event pair x_t materializes, then h receives dividend $\omega_h(x_t)$ during date t. The symbol $\omega = \sum \omega_h$ denotes the unlevered *market portfolio*.[27] In equilibrium, the consumption growth rate is $\omega(x_{t+1})/\omega(x_t)$.

A financial security is represented as a vector $Z = [Z(x_t)]$ where $Z(x_t)$ is the amount which one unit of the security pays its owner at x_t. Assume that the following securities are available for trade at every date t: (1) Zero coupon, risk-free bonds: these bonds underlie the term structure of interest rates. Assume that a zero coupon bond maturing at any date t is available for trade at any date before t. (2) The market portfolio: this security is denoted by Z_ω, and is a scalar multiple of ω. (3) European put and call options on the market portfolio. I assume that we can guarantee markets to be complete by allowing enough variation in the option exercise prices. A call option issued at x_t has an exercise price of K, expires at date $t+j$, and pays $max\{q_\omega(x_{t+j}) - K, 0\}$, where $q_\omega(x_{t+j})$ denotes the price of the market portfolio on the x_{t+j}-market. A put option is analogous to a call option, but returns $max\{0, K - q_\omega(x_{t+j})\}$.

Since markets are assumed to be complete, there are state prices that underlie security prices. Let $v(x_t)$ denote the price of an x_t-state contingent claim, and $v = [v(x_t)]$. I take x_0 as numeraire: that is, $v(x_0) = 1$. On the date 0 market, the price $q_z(x_0)$ of security $Z = [Z(x_t)]$ is $r \bullet Z$. On the x_t market, the price $q_z(x_t)$ of Z is the v-value of the Z-payoffs from date t on, divided by $v(x_t)$.[28]

A trader's wealth at the beginning of t consists of the market value of his x_{t-1} portfolio, including dividends paid in x_t. The trader then divides his x_t-wealth into a

[27] The levered market portfolio, as a portfolio of levered securities is a combination of call options.

[28] Define the vector $v'(x_t)$ as follows: The x_j-th component of $v'(x_t)$ is $v(x_j)$ for all successor nodes x_j to x_t, and the x_j-th component is zero otherwise. Then $q_z(x_t) = v'(x_t) \bullet Z/v(x_t)$.

portion to be consumed at t, and a portion to be saved. The saved portion is invested in the securities which comprise his x_t-portfolio. Denote trader h's net trade of the x_t-contingent commodity by $z_h(x_t)$. Then the consumption vector $c_h = [c_h(x_t)]$ is given by $c_h = \omega_h + z_h$.

I assume that each trader has a utility function featuring constant relative risk aversion. That is:

$$u_h(c) = \frac{c^{1-\theta_h}}{1-\theta_h} \tag{2}$$

where $c = c_h(x_t)$ and $1/\theta_h$ is h's risk tolerance parameter. Furthermore, h's preferences are additively separable over time and trajectories. Hence preferences are representable as the sum of weighted utilities, with weights $D_h(x_t)$, where $D_h(x_t)$ takes the form of a discounted probability $\delta_h^t P_h(x_t)$. Here δ_h is a discount factor satisfying $0 < \delta_h < 1$. $P_h(x_t)$ is nonnegative, and sums to unity for each t. Moreover, for each t, $P_h(x_t)$ is determined by conditioning on the probabilities $P_h(x_T)$ which attach to the terminal date T.

Every trader is assumed to choose his consumption plan c_h by maximizing the sum of weighted utilities

$$\sum_{t=1}^{T} \sum_{x_t} D_h(x_t) u_h(c_h(x_t)) \tag{3}$$

subject to the lifetime budget constraint $v \bullet z_h \leq 0$. Denote trader h's x_0-wealth by $W_h = v \bullet \omega_h$. Then h's demand function is:

$$c_h(x_t) = \frac{(D_h(x_t)/v(x_t))^{1/\theta_h} W_h}{\sum_\tau v(x_\tau)(D_h(x_\tau)/v(x_\tau))^{1/\theta_h}} \tag{4}$$

Note that in (4) the pattern of the consumption profile is keyed from wealth W_h, in that (4) specifies the fraction of wealth W_h which is to be consumed in each date-event pair x_t. In the discussion below, it will be useful to consider the consumption profile as being keyed to initial consumption $c_h(x_0)$ rather than to W_h. Note that $v(x_0) = 1$ since x_0 is taken as numeraire. Hence the denominator of (4) is equal to $W_h/c_h(x_0)$, so that by substitution, h's consumption growth rate is given by:

$$c_h(x_t)/c_h(x_0) = (D_h(x_t)/v(x_t))^{1/\theta_h} \tag{5}$$

7 A Representative Trader Characterization

The equilibrium state prices v are defined by the condition $\sum_h c_h(v) = \sum_h \omega_h$, with numeraire x_0. These prices underlie the way that all securities are valued in the market. An important feature about the model is that the equilibrium v can be characterized as if there were a single representative trader in the market. Indeed, imagine that there is only one trader, having risk aversion parameter θ_R and discounted probability weights $\delta_R^t P_R(x_t)$. Define the *cumulative* growth rate of aggregate consumption $g(x_t)$ as:

$$g(x_t) = \omega(x_t)/\omega(x_0)$$

In this case, (4) together with the equilibrium condition imply that $v(x_t)$ takes the form:

$$v(x_t) = \delta_R^t P_R(x_t) g(x_t)^{-\theta_R} \tag{6}$$

Because log-utility corresponds to the case $\theta_R = 1$, this last equality generalizes equation (4) in Shefrin-Statman (1994) to the general case of CRRA- utility; and Theorem 1 generalizes the representative trader characterization described on pp. 326-327.

I characterize sentiment through the beliefs of a representative trader.[29] Theorem 1 below shows how the parameters of a representative trader can be defined in terms of the parameters of the individual traders, market portfolio, and equilbrium prices. I emphasize that this theorem is utilitarian, in that its purpose is to provide a vehicle for characterizing the pricing kernel. The utilitarian point is important because I establish that the representative trader characterization is not unique, and show how different representative traders relate to one another. In particular, any representative trader can be chosen for the purpose of characterizing the kernel. The pricing kernel in no way depends on which particular representative trader is selected. However, I would point out that the choice of representative trader affects the ease with which the structure of the kernel can be elucidated. Only in this sense is the choice of representative trader critical.

I note that the representative traders' beliefs and preferences are not as well behaved as those of the individual traders. For example, the risk aversion parameter of the representative trader may depend on x_t, rather than being constant. Likewise, the discount term $\delta_{R,t}$ may vary with t. This variability occurs when risk tolerance or the discount factor is not uniform across the trading population. In this case, $\theta_R(x_t)$ reflects the curvature of the utility function of x_t-consumption, but loses its interpretation as a measure of the degree of relative risk aversion. In addition, the representative trader's probability beliefs may violate the principle of conditional probability.[30]

[29] The literature establishing the existence of a representative trader is quite extensive. See Basak (2000).

[30] Theorem 1 describes a constructive approach for tying the beliefs of the representative trader to those of

Investor sentiment is generally understood to mean the collective error in the market. By definition, sentiment is absent or zero when all traders hold objectively correct beliefs; that is, when $P_h = \Pi$ for all h. This case, where all traders hold homogeneous beliefs, $P_h = \Pi$, serves as a benchmark in the development of theorem 1 below. The interpretation of Π as an objective distribution is important for issues that involve market efficiency. However, I would emphasize that the formal results in the paper do not require that Π be objective. An alternative interpretation of Π is simply a benchmark case where traders hold homogeneous beliefs.

Let v_π be the equilibrium price vector v, and $c_{h,\pi}$ be the equilibrium value of c_h that occur when $P_h = \Pi$ for all h.[31] These are benchmark values. To set the stage for the theorem, I define two variables used in the construction of the representative trader. The first variable plays a key role in the approach that Benninga-Mayshar (1993) develop to define the representative trader's coefficient θ_R.

$$\alpha_h(x_t) = \frac{c_{h,\pi}(x_0)}{\omega(x_t)}[\delta_h^t \Pi(x_t)/v_\pi(x_t)]^{1/\theta_h} \tag{7}$$

The second variable is:

$$\gamma(x_t) = \sum_{h=1}^{H} \frac{c_h(x_0)(D_h(x_t))^{1/\theta_h}}{\sum_{j=1}^{H} c_j(x_0)}$$

Theorem 1 *Let v be an equilibrium state price vector.*

(1) v satisfies

$$v(x_t) = \delta_{R,t}^t P_R(x_t) g(x_t)^{-\theta_R(x_t)} \tag{8}$$

where θ_R, δ_R, and P_R have the structure described below:

$$1/\theta_R(x_t) = \sum_h \alpha_h(x_t)(1/\theta_h) \tag{9}$$

$$\delta_{R,t}^t = \sum_{x_t} \gamma(x_t)^{\theta_R(x_t)} \tag{10}$$

---

the individual traders, the risk tolerance of the representative trader to those of the individual traders, and the discount factor of the representative trader to those of the individual traders. This requires some care because beliefs, risk tolerance, and time discounting do not aggregate in separable fashion. The beliefs of the representative trader are impacted by the risk tolerances and discount factors of the individual traders, not just their beliefs. Theorem 1 provides a characterization that stresses beliefs in the beliefs-aggregation, risk tolerances in the risk tolerance-aggregation, and discount factors in the discount factor-aggregation. The nature of this emphasis is best seen in the aggregation conditions that appear in the proof of theorem 1.

[31] Because of the gross substitute condition for CRRA-demand functions, equilbrium is unique.

where the summation in (10) is over all traders and x_t-events at date t, and

$$P_R(x_t) = \frac{\gamma(x_t)^{\theta_R(x_t)}}{\delta_{R,t}^t} \tag{11}$$

(2) The representative trader is not unique. Equation (6) implies that any two representative traders, denoted $R,1$ and $R,2$, are related together through the expression:

$$\frac{\delta_{R,1}^t P_{R,1}}{\delta_{R,2}^t P_{R,2}} = g^{\theta_{R,2}/\theta_{R,1}} \tag{12}$$

(3) An alternative definition for $\theta_R(x_t)$, one that is not defined in terms of the objective probabilities Π, is:

$$1/\theta_R(x_t) = \frac{\ln(\gamma(x_t)) - \ln(g(x_t))}{\ln(v(x_t))} \tag{13}$$

In the remainder of the paper, I illustrate the nature of the insights that my results provide about the impact of heterogeneity on asset pricing. In order to maintain continuity between the examples discussed in previous examples, and the interconnections among all the results in the paper, I have based the subsequent illustrations on a common example that features log-utility, $T = 1$, a set S containing 65 states, and two traders with equal initial wealth. Each state is associated with a rate g of consumption growth, with g ranging from 0.80 to 1.44, in increments of 0.01. I note that because $\theta_h = 1$ for all h, (9) implies that $\theta_R = 1$. The log-utility assumption is not restrictive, at least for the points I illustrate with the example. Nevertheless, the log-utility assumption is special in certain respects, and I indicate where this is so at various spots in the paper.[32] The example features a market populated by two traders, whose beliefs are (approximately) lognormal, but featuring different means. One trader is overly optimistic about future consumption growth (bullish) and the other is overly pessimistic (bearish). The bull believes that the mean value of log-consumption growth ($\ln(g)$) is 12.01%, while the bear believes it to be -1.00%. The true mean is 5.28%. Both believe the standard deviation to be 3.76%, and I assume this to be the objectively correct value.

The first item to illustrate with the example is theorem 1. Being approximately lognormal, the density functions of the individual traders are single peaked. But because

[32] I considered including an additional example that features values for θ_h different from 1, but decided against it for reasons of length. The general features of that example are the same as the one I present in the paper. I chose $\theta_h = 1$ for all h, in the example I do use, to make clear how sentiment generalizes the efficiency condition in the Shefrin-Statman (1994) framework.

of the divergence in their viewpoints, the modes (peaks) are widely separated. Recall that in Shefrin-Statman (1994), the representative trader's beliefs are a convex combination of the individual traders' beliefs, this being a special case of theorem 1. Figure 6 displays the impact of these divergent viewpoints on the shape of the representative trader's density function. The figure shows the objective density function, the individual traders' density functions, and the representative trader's density function. Compared to the objective density, notice that representative trader's density is bimodal, and fat-tailed.[33]

8 Pricing Kernels and Sentiment

In this section, I propose a formal definition of sentiment, and relate this definition to the pricing kernel. The definition of sentiment generalizes the condition delveloped by Shefrin-Statman (1994) that combines the error-wealth covariance and average trader error. The sentiment variable I develop below has two advantages over the Shefrin-Statman (1994) condition. First, it relates directly to the structure of the pricing kernel. Second, the error-wealth condition in theorem 2 of Shefrin-Statman (1994) is specific to log-utility. Rather than focusing on whether or not the error-wealth dot product is zero, the more general condition focuses on whether or not sentiment is zero.[34]

I consider this section to be the core of the paper. Sentiment lies at the heart of the debate about the extent to which asset pricing fails to correctly reflect fundamental value. And the pricing kernel is the basis for pricing *all* assets. Theorem 2 below, decomposes the pricing kernel into two components, one pertaining to fundamentals and the other to

[33] When there are more than two traders, the representative trader's probability density is a convex combination of lognormal densities. This typically produces a lumpy, fat-tailed density, just as in the case of two traders.

[34] The dot product condition is special because heterogeneity involves implicit probability weighting. Equation (4) shows how a trader's return distribution varies as a function of θ_h. This equation enables us to compare how the value of θ_h affects the manner in which a trader weights probabilities relative to a log-utility trader (for whom $\theta_h = 1$). Note from (4) that the power function $P^{1/\theta}$ lies at the heart of probability weighting. For $\theta > 1$ and $0 < P < 1$, the power function is strictly concave, its first derivative goes to infinity as P approaches zero, and is unity when $P = 1$. This implies that small probabilities are implicitly overweighted relative to the case of log-utility, a feature shared by prospect theory (Kahneman-Tversky, 1979). Therefore, the dot product of relative wealth and probability error is not sufficient to analyze market efficiency in the general case, because probability weighting is ignored. Of course, we could normalize probability weights, produce associated probabilities, and compute the value of the dot product function. This would provide an efficiency measure relative to a log-utility model, but not one relative to the true spectrum of preferences.

sentiment. This result is straightforward to prove, and lies at the core of how heterogeneity impacts asset pricing.

Let $r(Z)$ denote the (gross) return vector for security Z. In general, a pricing kernel M_t is a stochastic process that satisfies $E_t(M_{t+1}r_{t+1}(Z)) = 1$. The state price vector v provides the present value, at date 0, of a contingent claim to one x_t-dollar. In a discrete time, discrete state model, a pricing kernel M_t restates this present value, in terms of per unit probability, v/Π. Now $M(x_t)$ is more correctly written $M(x_t|x_0)$. To obtain the stochastic process for the kernel, define $M_t \equiv M(x_{t+1}|x_t)/v(x_t)$, where M_t is a random variable. For this reason, I focus on $M(x_1)$ below, as the prototypical case. Using (8), obtain:

$$M_1 \equiv M(x_1) = \delta_R(P_R(x_1)/\Pi(x_1))g(x_1)^{-\theta_R} \tag{14}$$

where I have suppressed the notation indicating that both δ_R and θ_R are time and state-dependent respectively.

Note the likelihood ratio $\Lambda(x_1) = P_R(x_1)/\Pi(x_1)$ that appears in (14). Shefrin and Statman (1994) establish that this ratio captures the effect of noise traders on prices. Define sentiment by the variable $\lambda \equiv \ln(\Lambda)$.[35] In figure 2, the market sentiment attached to both date 1 nodes is zero, because $P_R(x_1) = \Pi(x_1)$ for those nodes. However, note that for the date 2 node that features two successive up-moves, the log-likelihood ratio is positive: $\ln(.50/.49) = 0.02 > 0$. The wealth shift from the bearish trader 2 to the bullish trader 1, that results from the first up-move, will lead the value of λ, conditioned on the first up-move, to become positive after the second up-move.[36]

The impact of sentiment on the pricing kernel is most easily summarized through the log-kernel process $m \equiv \ln(M)$ and sentiment process λ. Equation (14) implies:

Theorem 2

$$m = \lambda - \theta_R \ln(g) + \ln(\delta_R) \tag{15}$$

where m, λ, θ_R, and g are functions of x_1.

[35] λ is defined relative to Π, the objective distribution. This enables me to discuss how asset prices are diffent relative to the case of market efficiency. However, we may instead want to compare how asset prices are different relative to some other situation, such as when traders falsely believe the underlying process to be I.I.D. In this case, the theorems hold, as long as we are careful to reinterpret the role of Π, a point I emphasized in the previous section.

[36] Some additional computation shows that the value of λ at this particular date 2 node, conditional on an up-move at date 1, will be 0.019.

Theorem 2 states that the log-kernel is the sum of two stochastic processes, a sentiment process and a fundamental process based on aggregate consumption growth. Note that prices are efficient when the sentiment variable λ is uniformly zero, meaning its value is zero at *every* node in the tree. Hence, in an efficient market there is no aggregate belief distortion, in which case there is only one effective driver in (15), the fundamental process.

Figure 7 consists of three panels that illustrate the main concepts discussed in this section. The top panel depicts the graph of the dot product of relative wealth and trader errors against gross consumption growth rate g. Notice that the dot product is negative in the range 1.01 to 1.09, and nonnegative elsewhere. This pattern corresponds to the relationship between the objective probability density and representative trader's probability density in figure 6. Notice that in that figure, the representative trader attaches too low a probability to growth rates between 1.01 and 1.09.

The middle panel in figure 7 depicts the graph of sentiment λ against g. Note that in this example, the graph takes the shape of a smile.[37] Note too that λ takes values that are negative between 1.01 and 1.09, and positive elsewhere. Hence the dot product function and sentiment functions convey identical information. However, notice that the dot product function is defined in absolute terms, and approaches zero at the extreme ends of the range. However, because sentiment is defined as a log- likelihood ratio, it is expressed in relative terms, and is actually furthest away from zero at the extremes. In relative terms, the extremes are where the errors are most severe.

The bottom panel of figure 7 shows the graph of the log-kernel and its two components, as functions of g. Theorem 2 tells us that when sentiment is zero, the graph of the log-kernel (equal to $-\ln(g)$) is downward sloping. But the theorem also shows the sentiment smile being transmitted to the graph of the log-kernel.

[37] The smile is a special case. Here are three other cases. First, in an efficient market, $\lambda = 0$ for all g, so the graph of λ against g is flat. This is because when $P_h = \Pi$ for all h, the value of P_R in theorem 1 turns to be Π. Second, when the market is dominated by bulls, the graph of λ is positively sloped. This is because bulls attach too high a probability to high consumption growth states, and too low a probability to low consumption growth states. Therefore, λ is negative for low consumption growth states, and positive for high consumption growth states. Third, if the market is dominated by bears, then the graph is negatively sloped. The smile feature in the sentiment function occurs when the graph of λ is negatively sloped for low consumption growth states, and positively sloped for high consumption growth states. That is, although sentiment reflects the beliefs of both bulls and bears, the beliefs of the bulls dominate for high consumption growth states, and the beliefs of bears dominate for low consumption growth states.

There are three side issues associated with the characterization of sentiment and market efficiency used in the paper. The first issue concerns nonuniqueness of the representative trader. Recall that I defined sentiment in terms of the representative trader's probability beliefs as obtained from theorem 1. However, as I noted in the statement of this theorem, the characterization of the representative trader is not unique.

Consider the merits of the particular representative trader described in theorem 1. If all traders hold objectively correct beliefs, meaning $P_h = \Pi$ for all h, then the value of P_R will be Π, a property that is easily verified.[38] If all traders have the same coefficient of risk aversion, then the representative trader in theorem 1 will share that same value, across all x_t. This too is easily verified. Hence the particular representative trader described in theorem 1 is structured so that efficient prices serve as the base case, with sentiment λ reflecting departures from the base case.

The pricing kernel does not change if we focus on other representative traders than the one described in theorem 1. But the use of other representative traders may be less useful for understanding the impact of traders' beliefs on the structure of the kernel. For example, one can always define the representative trader's beliefs by the objectively correct distribution Π, even when $P_R \neq \Pi$ for the P_R of theorem 1. In the notation of equation (12), $P_{R,1} = P_R$ and $P_{R,2} = \Pi$. But this comes at a cost, in that sentiment λ gets transferred into the associated coefficient of relative risk aversion $\theta_\Pi = \theta_{R,2}$. Specifically (12) implies that

$$\theta_\Pi = \theta_R - \frac{\lambda + t \ln(\delta_R/\delta_\Pi)}{\ln(g)}$$

where δ_Π is the discount factor $\delta_{R,2}$. Unlike the representative trader of theorem 1, the $\Pi-$based representative trader's beliefs are defined independently of the beliefs of the individual traders, which in turn causes λ to drive a wedge between θ_Π and the individual traders' coefficients of risk aversion. As discussed above, the kernel might feature a smile effect stemming from the differential beliefs of bulls and bears, but its analysis through θ_Π will be less illuminating, and less intuitive.

The second side issue concerns the choice of probability used to normalize the state prices v, in order to obtain a kernel. In the preceding discussion, I chose the objective

[38] Benninga-Mayshar (1993) characterize θ_R for the case when all traders know the objective probabilities. See the proof of theorem 1 in the appendix.

distribution Π. An alternative is to choose P_R, and use (8) to obtain a different kernel,

$$M'(x_t) = \delta_R^t g(x_t)^{-\theta_R} \tag{16}$$

This alternative appears to make the kernel depend only on fundamentals, not sentiment. But remember that the kernel is a stochastic process, and under this definition the underlying probability distribution is P_R. In this respect, I note that sentiment λ describes how P_R is derived from Π. Moreover, P_R may not conform with the laws of conditional probability, thereby violating Bayes rule.

The final side issue deals with the impact of sentiment on the consumption growth profiles of the individual traders. Combining equations (8) and (5), obtain:

$$c_h(x_t)/c_h(x_0) = (D_h(x_t)/D_R(x_t))^{1/\theta_h} g(x_t)^{\theta_R/\theta_h} \tag{17}$$

Equation (17) expresses h's consumption growth in terms of aggregate consumption growth, h's risk tolerance relative to that of the representative trader, and h's subjective probability of the realized event relative to that of the representative trader. The case when sentiment $\lambda = 0$ provides an important benchmark, since the representative trader's consumption always grows at rate $g(x_t)$. Equation (17) implies that the consumption of a trader who is more (less) risk averse than the representative trader grows as a concave (convex) function of $g(x_t)$. Notably, h's consumption growth is an increasing function of the probability he attaches to the occurrence of x_t. Observe that a higher tolerance for risk amplifies the impact of belief heterogeneity.

9 Issues Involving Robustness

There are two main issues about robustness in the paper, robustness in respect to CRRA-utility, and robustness in respect to my choice of illustrative examples. I begin with the utility function issue.

In the previous section, I defined sentiment λ as a log-likelihood ratio function, where the numerator is the representative trader's probability density and the denominator is its objective counterpart.[39] Under this definition, market efficiency and uniformly zero sentiment are synonymous. This approach is not specific to CRRA-utility, because the

[39] The denominator need not be the objective distribution. It can actually be any distribution we wish to study for asset pricing purposes.

existence of a representative trader is not unique to CRRA-utility. See Basak (2000), who analyzes the character of representative traders in a model that features general utility functions. However, the exact expressions in my theorems are based on CRRA-utility.[40] In particular, the log-kernel decomposition result in theorem 2 is specific to CRRA-utility. This raises the question of how the results in the paper about trade-induced wealth redistribution are affected when traders possess other utility functions.

One of the most widely used utility functions in financial economics features constant absolute risk aversion (CARA). Therefore, consider how the character of my results change when all traders have CARA-utility, instead of CRRA-utility. The CARA-utility function has the form $u_h(c) = e^{A_h c}$, where A_h denotes trader h's Arrow-Pratt risk aversion measure $(-u''/u')$.

By maximizing expected CARA-utility subject to the budget constraint $v \bullet z_h \leq 0$, and noting that $P_h(x_0) = v(x_0) = 1$, we can solve for h's demand function:

$$c_h(x_t) = c_h(x_0) - \frac{\ln(v(x_t)/P_h(x_t))}{A_h} \tag{18}$$

and

$$c_h(x_0) = \frac{W_h}{\sum_t v(x_t)} + \frac{\sum_{t>0} v(x_t) \ln(v(x_t)/P_h(x_t))}{A_h \sum_t v(x_t)} \tag{19}$$

where the notation $\sum_t$ means summation over all nodes in the tree.

There are two crucial issues associated with the impact of heterogeneity on asset pricing. The first concerns self-cancellation of individual trader errors; i.e., uniformly zero sentiment. In general, sentiment is defined in relation to the the level sets associated with trader beliefs in the excess aggregate demand function, $\sum_h (c_h - \omega_h)$. The first issue is: under what conditions do changes in traders' beliefs leave the value of aggregate demand, and hence equilibrium prices v, unchanged? The second issue concerns the extent to which aggregate demand and prices are affected by dynamic wealth transfers.

To address the first issue, notice that by (18), the aggregate demand function involves the sum:

$$\sum_h^H \frac{\ln(v(x_t)/P_h(x_t))}{A_h}$$

Consider the special case when traders share the same CARA-coefficient, $A_h = A$. Notice

[40] As with the discussion concerning theorem 1, care must be taken to deal with issues involving nonuniqueness.

that this term can be written as a log-product

$$A^{-1}\ln(v(x_t)^H\prod_h P_h(x_t))$$

In combination with (19), this implies that when traders share the same CARA-coefficients, equilibrium prices are invariant to shifts in traders' beliefs, as long as the shift preserves the value of the probability-products. This is a different condition than the one embodied within the CRRA-based aggregate demand function.[41] The point here is that the level sets associated with the aggregate demand function are different for different utility functions.[42]

Volatility in asset prices stems from a combination of factors: the nature of traders' beliefs, the extent of trader disagreement, and the stochastic nature of wealth redistribution induced by trading. As I discussed in sections 3 through 5, stochastic wealth redistribution is an important source of volatility in my model. But it is not a feature of all models. I note that the issue has been well studied in aggregation theory: See Heineke and Shefrin (1988). Notice that (18) and (19) feature the property for which CARA-utility is well known:

$$\partial c_h(x_t)/\partial W_h = 1/\sum_t v(x_t) \tag{20}$$

Condition (20) stipulates that each trader allocates every marginal dollar of portfolio wealth to the risk-free security. But this means that wealth distribution plays no role in determining equilibrium prices, since the aggregate excess demand function $\sum_h(c_h(v)-\omega_h)$ is invariant to wealth redistribution. Observe that equations (18) and (19) imply that the CARA-demand function satisfies the Gorman polar form:[43]

$$c_h(x_t) = J(v,h) + G(v)W_h$$

[41] Hence, the structure of state prices is different. For example, when traders have the same coefficient of absolute risk aversion, A, then v is a ratio of the representative trader's discounted probability to the following ratio: $e^{A\omega(x_t)}/e^{A\omega(x_0)}$. Contrast this with (6).

[42] In (18), the term $\ln(v(x_t)/P_h(x_t))$ is actually $\ln((v(x_t)/P_h(x_t))/(v(x_0)/P_h(x_0))$, where $v(x_0) = P_h(x_0) = 1$ because date 0 consumption is the numeraire and there is no uncertainty at date 0. In a one date binomial setting where there is no date 0 consumption, we would choose one of the two date 1 nodes as numeraire. In this case the equal product condition described in the paragraph is a product of likelihood ratios. Suppose there are just two traders with the same wealth, and the objective probability of both states is 1/2. This is an interesting special case because maintaining the value of the product is the same as maintaining the arithmetic average. The product of the likelihood ratio remains at 1 as beliefs are altered to maintain a constant artithmetic average. This means that for this special case, the CARA efficiency condition is the same as the CRRA efficiency condition.

[43] There are two features of CARA-utility that strike me as problematic. First, the fact that each trader allocates 100% of every marginal dollar of portfolio wealth to the risk-free security is extremely unrealistic.

$$\sum_h (c_h(x_t) - \omega_h(x_t)) = \sum_h J(v,h) + G(v) \sum_h W_h \tag{21}$$

where in the case of CARA-utility,

$$G(v) = 1/\sum_t v(x_t)$$

Gorman (1953) established that the aggregate demand function is invariant to the wealth (income) distribution, if and only if individual demand functions satisfy the polar form.[44]

I note that the Gorman polar form also underlies the two-fund separation property in the capital asset pricing model (CAPM). When traders have quadratic utility and share the same beliefs, but not necessarily the same risk tolerance parameters, their demand functions will satisfy the Gorman polar form. See the next section.

To recapitulate, there are two issues involving the impact of heterogeneity on equilibrium prices. The first issue concerns the case when traders' errors are self-cancelling. This situation corresponds to sentiment λ being uniformly zero. I note that there is nothing specific to CRRA-utility in the approach. Nevertheless, the precise equations described in my theorems are specific to CRRA, which takes us to the second issue. An important feature of my model is that stochastic wealth redistribution injects volatility into asset pricing. This feature is not universal. Specifically, when the individual utility functions satisfy the Gorman polar form, trading-induced wealth redistribution stemming from trader heterogeneity does not impact the volatility of asset prices.

The final robustness issue concerns the illustrative examples I present in the paper. Some examples feature nonzero sentiment, and other examples feature zero sentiment. In my model, the representative trader's probability density function underlies asset pricing, and can feature fat tails and/or multi-modality. Think about whether this feature is

Second, the Gorman polar form does not constrain consumption to be nonnegative. Indeed, if $J(v,h) \neq 0$, then some components of $J(v,h)$ must be negative. This is because the budget constraint requires that $v \bullet J(v,h) = 0$.

[44] CARA-utility is a feature of both noisy rational expectations models, and some of the behaviorally-based asset pricing models featuring sentiment. An important feature of noisy rational expectations models is that traders condition their inferences on price and their private signals, without having to take the joint distribution of wealth and signals into account. This is a strength when it comes to solving the model, and a weakness when it comes to applying the model. Notably, the Gorman polar is the underlying reason why the wealth distribution can be ignored in the problem. In addition, models that use CARA-utility tend to feature just two assets, one risk-free and the other risky. This strikes me as restrictive as far as drawing general conclusions about asset pricing is concerned.

common (robust), or whether it is instead a contrived artifact. Consider the process of constructing an example, beginning with a collection of different probabilty distributions for the individual traders. Theorem 1 provides the associated representative trader's beliefs. In constructing the example, we are free to choose the objective density Π. In this regard, we can choose to make Π equal to the representative trader's density P_R. If we do so, we will have produced an example that features zero sentiment. However, because the aggregation process in my model involves weighted averages of individual density functions, the resulting objective density will typically be multi-modal and/or fat-tailed.[45] On the other hand, if we begin with a unimodal objective density, we can certainly construct individual beliefs that will aggregate to the objective distribution. However, as I showed in section 4, the resulting individual densities are tightly constrained, constitute a knife-edge case, and because they are contrived to produce zero sentiment, may behave perversely.[46]

10 The Mean-Variance Efficient Frontier

In theory, what is the nature of beta in a world where returns reflect both compensation for risk and mispricing stemming from noise trader errors? That is one of the questions posed by Shefrin-Statman (1994). In this section, I extend their analysis to address the formal connection between the mean-variance efficient frontier and the pricing kernel, and the effect of nonzero sentiment upon the frontier.

The risk premium on any security Z is determined by the covariance of its return with the kernel. In fact the risk premium is just $-i_1 cov(r(Z), M)$. Of course, the risk premium can also be expressed in CAPM-like terms, by means of a mean-variance efficient benchmark portfolio and a beta. Beta is just the covariance between $r(Z)$ and the return to the benchmark, divided by the variance of the benchmark return.

Below I provide a general characterization of the mean-variance efficient frontier in terms of the pricing kernel.[47] Identifying the benchmark involves maximizing the expected quadratic utility $E_\Pi\{r(x_t) - \nu r(x_t)^2\}$ of the return r to a one dollar investment. This

[45] When $\theta_h \neq 1$ for some h, this statement holds approximately.

[46] In the example I presented in section 4, trader 2 adjusts the probability of a given state downwards, as the state occurs repeatedly along a realization.

[47] My focus is general and structural. In contrast, Daniel, Hirshleifer, and Subrahmanyan (1998) focus on more specific issues, and propose a theory that explains the relationship between returns and valuation measures such as book-to-market in the presence of mispricing stemming from overconfidence.

maximization underlies the next theorem.

Theorem 3 *The return $r_{MV}(x_1)$ to a mean-variance efficient portfolio is:*

$$r_{MV}(x_1) = \nu^{-1}(1 - [M(x_1)E_\Pi\{M(y_1) - \nu\}[E_\Pi\{M(y_1)^2\}]^{-1}]) \tag{22}$$

where

$$E_\Pi\{M(y_1) - \nu\}[E_\Pi\{M(y_1)^2\}]^{-1} = E_\Pi\{\delta_R\Lambda(y_1)g(y_1)^{-\theta_R} - \nu\}[E_\Pi\{(\delta_R\Lambda(y_1)g(y_1)^{-\theta_R})^2\}]^{-1}$$

and ν is a nonnegative parameter whose variation generates the mean-variance efficient frontier.

Theorem 3 expresses the mean-variance return in terms of the pricing kernel. Notice that the return r_{MV} is linear in the kernel $M(x_1)$, and has a negative coefficient.[48] Hence, the return is low in a state that bears a high price per unit probability.

Sentiment can alter the shape of the relationship between the mean-variance return r_{MV} and aggregate consumption growth g. In an efficient market, $\lambda \equiv 0$, in which case r_{MV} is a monotone increasing, concave function of g, for a suitably low value of ν.[49] This follows from (8) and the proof of theorem 3 (see the appendix). In an efficient market, a mean-variance portfolio earns very low returns in low-consumption growth rate states. Indeed mean-variance returns can fall below one hundred percent: there is no limited liability attached to a mean-variance efficient portfolio.[50]

However, sentiment can distort the shape of the above relationship between r_{MV} and aggregate consumption growth g, by introducing local extrema.[51] To see how a smile pattern in the sentiment function can affect the shape of the r_{MV} function, consider figure 8, which involves the same example that I presented in sections 7 and 8. The top panel displays the graph of the return to a mean-variance efficient portfolio as a function of g, when sentiment is zero. Here the function is concave and monotone increasing. The bottom panel display the graph of the same relationship for the case when the sentiment function

[48] The coefficient is time-varying, and stochastic. Note that equation (22) is general whereas the equation that follows it is specific to CRRA-utility.

[49] As $\nu \to 0$, the mean-variance utility function approaches risk-neutrality.

[50] Hence, gross returns can be negative, which is an important feature of benchmark portfolios used in the calculation of correct betas. Concavity implies that the marginal return to consumption growth is declining. For very high consumption growth rates, the return peaks, and can decline.

[51] In the relevant range.

exhibits the smile pattern displayed in figure 7b. Here the r_{MV}-function achieves a local maximum at $g = 1.05$. That is, the graph exhibits a "frown" pattern.

To understand what drives the shape of the r_{MV}-function in this example, observe that in figure 6, the mode of Π lies in the valley between the local maxima of P_R. This implies underpricing in the valley (where $\Pi > P_R$), and overpricing at the extremes (where $\Pi < P_R$). That is, the bulls cause the the highest consumption growth rate states to be overpriced, and the bears cause the lowest consumption growth rate states to be overpriced. A mean-variance efficient portfolio, constructed using objectively correct probabilities, will respond to the attendant mispricing by tilting towards underpriced states, and away from overpriced states; hence, the "frown".

What does a "frown" pattern, meaning the existence of a local maximum in the mean-variance efficient graph, imply for the pricing of securities, like the market portfolio, that feature high payoffs in the highest consumption growth states? It implies that these securities have negative "up-betas," and low overall betas.[52]

11 The Term Structure of Interest Rates

Economists have long been puzzled by the fact that the expectations hypothesis of the yield curve fails to hold. See Campbell (1995).[53] In this section, I discuss the effect of sentiment on the term structure of interest rates, and the expectations hypothesis. Theorem 4 below describes the relationship between the term structure and the representative trader's parameters.

Theorem 4 *Let i_t^t denote the gross return to a risk-free investment in which one real dollar is invested at date 0 and pays off t periods later. The discount factors which are based upon*

[52] Theorem 3 pertains to one-period returns. There is a counterpart result for $t-$period returns. But the benchmark portfolio used to price risk for $t-$period returns is not equivalent to compounded one-period mean-variance returns. In other words, from a theoretical perspective, betas based on monthly returns should not be used to price annual returns.

[53] Campbell points out that many of the term structure studies during the 1960s did not impose rational expectations, and therefore allowed systematic profit opportunities to exist. Although that may be the case in my model, I note that these opportunities are not riskless. There may well be traders who do have objectively correct beliefs, and yet refrain from seizing these opportunities because of the risk involved. See footnote 17, Campbell (1995).

(6) and define the term structure of interest rates have the form:

$$(1/i_t)^t = \delta_{R,t}^t E_R\{g(x_t)^{-\theta_R(x_t)}|x_0\} \tag{23}$$

where E_R is the expectation under the representative trader's probability distribution.

Equation (23) follows directly from (8) and the fact that the term structure is based on securities that offer a fixed payoff across all date $t-$states. This equation makes explicit the connection between the yield curve and the beliefs of the representative trader.[54] The equation captures how interest rates evolve in terms of the discount factor $\delta_{R,t}$, the parameter θ_R, and the expectations of the representative trader E_R.

The expectations hypothesis of the term structure predicts that subject to a risk premium, the expected return to holding short term Treasury securities is the same as the return to holding long term Treasury securities. Therefore when the current slope of the yield curve is steep, future yields on short-term bonds must rise to compensate for the current shortfall, and yields on long bonds must rise to generate capital losses to holders of long bonds. Summaries of the evidence concerning the expectations hypothesis can be found in Shiller (1990), Campbell and Shiller (1991), Campbell (1995) and Campbell, Lo, and MacKinlay (1997). This evidence is mixed. It appears that when the current slope of the yield curve is steep, future short-term rates do rise. However, they do not rise across all maturities, to the degree predicted by the expectations hypothesis. Moreover, future yields on long bonds do not rise on average: they fall.

Now the representative trader holds the market portfolio and consumes its dividends. Notably, a representative trader for whom $c_R(x_0) = 1$, consumes the cumulative dividend growth rate $g_t = g(x_t)$. The expectations hypothesis is driven by the fact that at the margin, the representative trader is indifferent to substituting bonds with long term maturities with those of shorter maturities in his portfolio. For example, if we consider $t = 2$ as the long term and $t = 1$ as the short term, then indifference at the margin implies that:

$$E_R\{g_2^{-\theta_R}\}i_2^2 = E_R\{g_2^{-\theta_R} i_1(x_1)\}i_1 = 1 \tag{24}$$

That is, the marginal utility of a dollar invested in either the short-term bond or the long-term bond is equal to the marginal utility of a dollar, which is unity. For ease of notation

[54]This equation treats x_0 as the current event. The expression is easily generalized when the current event is x_t.

the x_2 argument in θ_R is suppressed, as is the x_0 associated with the conditional expectation in (24). Define the date 2 forward rate by:

$$f_2 = \frac{(i_2)^2}{i_1} \tag{25}$$

I note that (24) can be rewritten to obtain a condition that relates the spot and forward interest rates, a relationship often used to test the expectations hypothesis. When equation (26) below holds, the representative trader is indifferent to substituting a long bond for a short bond in his portfolio.[55] This condition is derived using (2), (23), and the fact that the representative trader consumes the cumulative growth rate of the market portfolio. We have:

Theorem 5

$$f_2 - E_R\{i_1(x_1)\} = \frac{cov_R[g(x_2)^{-\theta_R}, i_1(x_1)]}{E_R\{g_2^{-\theta_R}\}} \tag{26}$$

Equation (26) implies that there are three impediments to the validity of the expectations hypothesis. The first is a nonzero risk premium that interferes with the pure expectations hypothesis, which states that the forward rate equal the expected spot rate. In general, the right-hand-side of (26) is nonzero.

As for the version of the expectations hypothesis that requires a nonzero risk premium but requires it to be time invariant, there are two additional impediments. First, the expectations hypothesis requires that the expectation in equation (26) be taken with respect to the objective distribution Π. However, in (26), the expected spot rate is computed relative to the representative trader's beliefs P_R, not the objective process Π. The point is that when $\lambda \neq 0$, the representative trader holds erroneous beliefs.

The final impediment concerns the stability of the risk premium, given the covariance term in (26). Let Z_2 be the long bond. Since the risk premium is given by well known expression $-i_1 cov_\Pi(r(Z_2), M)$, the expectations hypothesis requires that the preceding covariance vary inversely with the spot rate i_1. But by its nature, heterogeneity induces time variation into this covariance: recall the discussion about time variation in both θ_R and δ_R.[56]

[55] The equation, which was derived for the case of $t = 1$ and $t = 2$, is easily generalized.

[56] The covariance in question is between marginal utility of aggregate consumption growth at date 2 and the date 1 spot interest rate. Suppose that this covariance is positive. In this case, the forward rate will

To illustrate the impact of heterogeneity on the term structure, consider the binomial example developed in sections 3 and 4, there being no term structure in the one-period example used in the previous three sections.

In sections 3 and 4, I discussed why heterogeneity can cause the short-term interest rate to be stochastic. In turn, this implies that the entire term-structure is stochastic. In the homogeneous beliefs case, the interest rate stayed constant at 1.868% over time, which I rounded to 1.87%. By applying the argument developed in section 4 to find the entire term structure of interest rates, we can determine that in the homogeneous case described there, the yield curve is flat at a rate of 1.868%. However, for the example discussed in section 4, a little computation shows that heterogeneity causes the date 0 yield curve to slope downward, with successive values of $1.868\%, 1.863\%, 1.858\%, 1.852\%, 1.847\%$.

What about the forward rates? Equation (25) implies that on the date 0 market, the date 2 forward rate will be 1.857%. The expected spot rate under the representative trader's probabilities turns out to be 1.868%, implying a negative risk premium of -0.011%.[57] By repeating the calculation after an up-move at date 0, I find that the forward rate has increased to 1.200%, the corresponding expected spot interest rate will have increased to 1.201%, and the associated risk premium will have changed to -0.010%. This small change in the risk premium implies that in the example, the expectations hypothesis does not hold.

---

tend to be greater than the expected future spot rate. What gives rise to a positive covariance in (26)? Consider a state where date 2 consumption is high relative to its mean, and so marginal utility is low. A positive covariance implies that the high consumption growth was likely to have been preceded by a low spot interest rate at the previous date (1). What keeps the interest rate low? The representative trader's pessimistic expectations about future consumption growth. When the covariance is negative, the reverse holds: high consumption growth at date 2 is likely to have been preceded by a high spot rate at date 1.

[57] The risk premium is negative because traders hold I.I.D. beliefs, and spot interest rates are stochastic. Given that risk-free bonds are denominated in real terms, traders demand a premium to accept a sequence of risky short-term returns instead of a certain long-term return. This occurs in the model because exponentiation amplifies small differences in exponents, and the largest discount factor (lower interest rate) takes on the greatest relative importance, dominating for higher values of t. Recall that bond prices, discount factors, are obtained as wealth- weighed convex combinations of homogenous-based bond prices.

12 Risk-neutral Densities and Option Pricing

In this section, I consider the impact of heterogeneity on the structure of European option prices. The discussion in this section reinforces the argument I presented in section 5, where I provided an example to demonstrate that heterogeneous beliefs can lead equilibrium option prices to deviate from their corresponding Black-Scholes values. In this section, I augment the argument and relate it to sentiment. The point is as follows: take an economy that features Black-Scholes equilibrium pricing when sentiment is zero. Then the introduction of non-zero sentiment leads the equilibrium prices of some options to deviate from their corresponding Black-Scholes values. A secondary issue concerns the value for volatility to be used in the Black-Scholes formula when traders disagree about its true value.

In this section I present three equivalent option pricing expressions. The first expression is developed in theorem 6. This expression is based on the standard risk-neutral density approach, and involves the arguments used to demonstrate how discrete time option pricing formulas converge to the Black-Scholes formula in the limit. See Cox, Ross, and Rubinstein (1979), and Madan, Milne and Shefrin (1989). I develop the other two option pricing expressions to bring out some important features about the variables that determine the prices of options.

In theorem 7, I present a second option pricing expression that demonstrates how traders' beliefs, operating through the beliefs of the representative trader, affect option prices. The risk-neutral based option pricing expression in theorem 6 obscures the relationship between traders' beliefs and the prices of options. And the traditional risk-neutral approach to option pricing appears to have led researchers to the view that option prices are independent of traders' beliefs, since that is the case in partial equilibrium option models. However, as I discussed in section 4, traders' beliefs impact option prices.

The third option pricing expression reflects what I call a "snapshot in time" approach. The "snapshot in time" expression, also described in theorem 7, depends only on variables associated with the expiration date. In particular it relies on the long-term interest rate and the risk-neutral density at the expiration date. This contrasts with the first expression, in theorem 6, that relies on the co-evolution of the short-term interest rate process and the risk-neutral process over the life of the option. The "snapshot in time" approach is useful for pointing out that the differences between continuous time option pricing models and discrete time option pricing models are less important than the character of the risk-neutral process, a point to which I return at the end of the section. This approach serves to provide a link between the two modeling techniques.

Theorem 6 below describes the first option pricing formula, expressed in terms of the risk-neutral process and the process for short-term interest rates.

Theorem 6 *Given (8), the general expression for the price of a European call option on a security Z, featuring exercise price K and expiration date t, is determined as follows.*

(1) Let $S(x_{t-1})$ be the set of successor nodes x_t to x_{t-1}. The risk-neutral density $\eta(x_t)$ associated with event $\{x_t\}$, conditional on x_{t-1}, is defined by:

$$\eta(x_t) = \frac{v(x_t)}{\sum_{y_t \in S(x_{t-1})} v(y_t)} \tag{27}$$

(2) Let A_E denote the event $\{q_z(x_t) \geq K\}$, in which the call option is exercised, and $P_\eta\{A_E\}$ be its probability under the risk-neutral density P_η. The product of the single period interest rates defines the cumulative return $i_c^t(x_t) = i_1(x_0)i_1(x_1)\cdots i_1(x_{t-1})$ to holding the short-term risk-free security, with reinvestment, from date 0 to date t. Then the x_0-price of the call option is given by:

$$q_c(x_0) = E_\eta\{(q_z(x_t) - K)/i_c^t(x_t)|A_E, x_0\}P_\eta\{A_E|x_0\} \tag{28}$$

Risk-neutral density pricing equations, such as (28) tend to obscure how the properties of the representative trader's beliefs affect asset prices. As I mentioned above, I present two alternative option pricing expressions.

Theorem 7 *(1) Given (8), a second expression for the price of a European call option on a security Z, featuring exercise price K and expiration date t, is determined as follows. Let A_E denote the event $\{q_z(x_t) \geq K\}$, in which the call option is exercised, and $P_R\{A_E\}$ be its probability under the representative trader's probability distribution P_R. Then the x_0-price of the call option is given by:*

$$q_c(x_0) = \delta_{R,t}^t E_R\{(q_z(x_t) - K)g(x_t)^{-\theta_R(x_t)}|A_E\}P_R\{A_E\} \tag{29}$$

(2) Define the t-step probability distribution $\phi(x_t)$ over date t events x_t, conditional on x_0 as follows:

$$\phi(x_t|x_0) = \frac{v(x_t)}{\sum_{y_t} v(y_t)} \tag{30}$$

Then the third expression for q_c is:

$$q_c(x_0) = E_\phi\{(q_z(x_t) - K)|A_E, x_0\}P_\phi\{A_E|x_0\}/i_t^t(x_0) \tag{31}$$

Expressions (29) and (31) describe the direct impact of the representative trader's beliefs on call option prices. (29) prices the option using the state price representation (8).[58] (31) indicates the connection between the term structure and option prices, in that the $t-$period bond is used to price the option. In the remainder of this section, I emphasize the insights offered by theorem 7 for the pricing of options.

Disagreement among traders over the true probabilities governing the evolution of the system typically leads the representative trader's probability beliefs to be multi-modal and fat-tailed. To illustrate the implication of these features for option prices, I continue with the example discussed in sections 8, 9, and 10. When both traders hold objectively correct beliefs, the interest rate will be 5.35%,[59] and the representative trader's density function will be (approximately) lognormal. Consider a security whose date 0 price is 1, and whose return has the same (lognormal) distribution as the market portfolio. In this case, the Black-Scholes formula can be used to compute the equilibrium price of an option defined on this security. For example, a call option with an exercise price of $K = 1.05$, expiring at date 1, will have an equilibrium price of 0.0166.[60]

When traders hold heterogeneous beliefs, as described in section 8, the interest rate falls from 5.35% to 4.95%. The equilibrium option price, computed directly using the one-period interest rate and pricing kernel (from equation (31)) is 0.0314. However, the Black-Scholes price is 0.0148, a lower value. I note that the equilibrium price of 0.0314 is higher than the Black-Scholes price of 0.0148 because the bulls drive up the price of the call option. That is, traders' beliefs affect option prices, that being the focus of option pricing expression (29).

The difference between the equilibrium option price and Black-Scholes price is caused by nonzero sentiment. Nonzero sentiment affects more than the interest rate. Nonzero sentiment also distorts the risk-neutral density function, causing it to depart from lognormality. See figure 9 which contrasts the risk-neutral density when sentiment is zero

[58] Equation (28) makes use of the definition of conditional probability, $Prob\{x_t|\eta, A_E\} = Prob\{x_t|\eta\}/Prob\{A_E|\eta\}$, where $P_\eta(A_E) = Prob\{A_E|\eta\}$.

[59] The continuously compounded rate is 5.21%.

[60] The Black- Scholes price is actually 0.0167, the difference in the last decimal point stemming from the lognormal approximation used in the discrete time example.

(the objective case) with the bimodal, fat-tailed density associated with nonzero sentiment in the example.

The lognormal density is single peaked, not multi-peaked. And coming back to section 5, let me say that it is the smile pattern in the pricing kernel induced by nonzero sentiment, that generates a smile pattern in the implied volatility graph. Figure 10 displays the $phi-$function associated with the expiration date, (and its underlying lognormal generators), for the continuous time example in section 5. Note that it is not all that different from the one period discrete time example portrayed in figure 9.[61] My point is that in a heterogeneous environment, the appearance of "volatility smile" effects and "crashophobia" humps (Rubinstein, 1994) can stem from multi-modality and fat tails that are absent in a lognormal environment.[62] The smile effect in the left tail in the risk-neutral density functions in my examples reflects the beliefs of the pessimists; the crashophobia hump reflects the mode of the pessimists' density functions. Moreover, multi-modality is counterintuitive for most. In this connection, Jackwerth and Rubinstein (1996) use a smoothing procedure that produces a unimodal function when they estimate the risk-neutral density from actual option prices.[63]

One final issue: all the lognormal examples in the paper assume that traders agree about the value of volatility. I have made this assumption in order that the Black-Scholes

[61]When P_R is I.I.D., implying a constant interest rate over time, phi and η coincide. Hence, for the continuous time example described in section 5, phi can be constructed from the constituent risk-neutral processes. By definition, phi and eta also coincide in the single period example. In the appendix I establish that the risk-neutral density associated with an event x_t has the same form as the associated state price, but features the discounted value of $g(x_t^{\theta_R}$, not the undiscounted value. Hence, the $phi-$function describes the density function associated with the stochastic process featuring the ratio of the cumulative future value i_c^t and eta, this being the key variable used to price options in (28).

[62]Although my examples pertain to options on the market portfolio, the results apply to any underlying asset. Smile effects stem from the extent of disagreement on payoff relevant states for the underlying asset, the joint distribution of trader errors, risk tolerance parameters, wealth levels, and discount factors. By no means is it true that the results for options on the market portfolio need carry over to options on individual securities.

[63]There are competing explanations for what gives rise to smile effects in option pricing. The Black-Scholes option pricing framework has been extended to feature jumps, stochastic interest rates, and stochastic volatility. As I argued in section 5, when sentiment is nonzero, heterogeneity gives rise to stochastic interest rates, and can give rise to stochastic volatility and jumps. However, existing models make no reference to heterogeneity. Notably, Bakshi, Cao, and Chen (1997) report that existing models do a poor job of explaining the option smile. For a discussion of the role that heterogeneity plays in option smile effects, see Shefrin (1999).

price be well-defined. Relaxing the assumption in no way prevents equilibrium option prices from being computed. However, relaxing the assumption does raise the question of which volatility to use in the computation of the Black-Scholes price. In the standard approach, it is the objective volatility that enters as an argument of the Black-Scholes formula. On this point, I remind readers that it is actually the volatility of the risk-neutral density that enters the Black-Scholes formula. However, in the Black-Scholes framework, the risk-neutral volatility and objective volatility are equal. In my framework such equality need not hold.[64]

13 The Market Portfolio

How sensitive is the objective return distribution of the market portfolio to sentiment? The answer to this question depends on the representative trader's risk tolerance, as the following theorem demonstrates.

Theorem 8 *The x_0-price q_ω of the market portfolio has the form:*

$$q(Z_\omega) = \omega(x_0) E_R\{\sum_{t=1}^{T} \delta_{R,t}^t g(x_t)^{\theta_R(x_t)-1}\} \tag{32}$$

Let $r_\omega(x_1)$ denote the return to holding the market portfolio from x_0 to the beginning of x_1. Then:

$$r_\omega(x_1) = (g(x_1)/\delta_{R,1}) \frac{\sum_1^T E_R\{\delta_{R,1}^t g(x_t)^{1-\theta_R(x_t)}|x_1\}}{\sum_1^T E_R\{\delta_{R,0}^t g(x_t)^{1-\theta_R(x_t)}|x_0\}} \tag{33}$$

In (33) the base from which growth is measured in the numerator is $\omega(x_1)$, whereas in the denominator the base is $\omega(x_0)$.

The probabilities that underlie the return distribution for the market portfolio are given by Π. The support of the distribution is given by (33). Theorem 8 establishes how beliefs affect the support. The return to the market portfolio is a product of three terms, the growth rate in aggregate consumption, the inverse of δ_R, and the ratio of two expectations.

To interpret expression (33), consider the case of logarithmic utility, meaning the case when $\theta_R = 1$. Here, the expectation ratio in (33) is unity, so the return to the market portfolio is $g(x_1)/\delta_R$. This implies that the return on the market portfolio is the

[64]Note that in the numerical examples I presented in this paper, all risk-neutral moments are wealth-weighted convex combinations of the corresponding values for the individual trader density functions.

consumption growth rate, scaled by the inverse discount factor. Scaling is necessary with discounting in order to induce saving. Take the logarithmic situation as the base case, and consider how r_ω changes relative to the base case as we increase the value of θ_R.

When $\theta_R > 1$, the expectation ratio in (33) is not unity. Notice that the numerator of the expectations ratio is conditional on the x_1 while the denominator is the same expectation conditional on x_0. Because of the different bases from which growth is measured in numerator and denominator, a positive trend in expected growth rates leads the expectation ratio in (33) to lie above unity. Hence theorem 8 implies that a shift in optimism about consumption growth causes the return r_ω to be higher than its value under logarithmic utility. In other words, theorem 8 demonstrates how the value of θ_R affects the sensitivity of the return distribution of the market portfolio to trader beliefs.

Under logarithmic utility, the support of the return distribution is independent of traders' beliefs. This can be seen in the example of section 3. Theorem 8 makes clear that the logarithmic utility case is special. The lower the representative trader's risk tolerance, the greater the influence of expectations on the value of r_ω.

There is something else about the impact of lower risk tolerance. It strengthens the correlation between the return on the market portfolio and the change in the yield curve. This can be seen by comparing (23) and (33).

14 Specific Behavioral Assumptions

The results in the paper pertain to heterogeneity, not to specific behavioral features. Yet, readers of this paper have consistently asked me to relate the results to the literature that focuses on behavioral phenomena; hence, I have added this section in response to the request. The literature in behavioral finance is principally concerned with two types of phenomena: (1) judgmental errors in probability beliefs; and (2) prospect theoretic preferences (gains, losses, reference points, and an S-shaped utility function). In this section, I discuss both phenomena, beginning with judgmental errors. Please note that I intend for the discussion in this section to be qualitative, rather than precise.

In a behavioral pricing framework, there are two processes to model, the fundamental process based on aggregate consumption growth g, and the process governing the evolution of sentiment λ.[65] The variables that underlie the sentiment process are a mix:

[65] In a continuous time framework, it is typical to assume that the growth rate in aggregate consumption

individual traders' beliefs, risk-tolerance parameters, wealth levels, and the fundamental process itself. Consider each, beginning with beliefs.

Shefrin and Statman (1994) propose two classes of quasi-Bayesian behaviorally-motivated learning structures that underlie beliefs $P_h(x_t)$.[66] One group of traders under-weights base rates in applying Bayes' rule to update conditional probabilities. This error induces traders to predict the continuation of recent trends. The second group succumbs to the "law of small numbers," and tends to predict the reversal of recent trends. Notably, both types of quasi-Bayesian traders are implicitly overconfident. Their priors are excessively tight, relative to true Bayesians.[67] If the objective process Π is Markovian and Ergodic, true Bayesians will eventually learn the true probabilities. However, quasi-Bayesians will not, since their beliefs do not typically converge. There is a view in traditional finance that traders' errors are temporary, and will disappear with learning. In contrast, the literature in behavioral decision making contains many studies showing that people learn very slowly, and that errors persist in the face of experience.[68]

Sentiment λ is the aggregate reflection of traders' errors. The degree to which an individual trader's error process affects sentiment depends on the size of the trader's trades. λ depends on risk tolerance and wealth because these variables provide the weights used to aggregate trader errors. Traders who are wealthier and more tolerant of risk take larger positions than traders who are less wealthy and less tolerant of risk.

Sentiment is time varying. The magnitude of λ varies as wealth shifts between traders who have taken the opposite sides of trades, a point I emphasized in section 4. This is because the weight attached to a trader's beliefs is an increasing function of his trading success. Along a high consumption growth sequence, base rate underweighters will become unduly optimistic, and sentiment will move in the positive direction. Along a low consumption growth sequence, the reverse will occur. In a volatile segment, with frequent alternation between high and low consumption growth, weight will shift to the traders who believe in the law of small numbers. These traders overestimate the degree of volatility. During those segments when consumption growth is volatile, the kernel will accord their

growth follows a square root (Cox-Ingersoll-Ross) process. How about the process governing λ?

[66] Both stem from the "representativeness" heuristic (Tversky and Kahneman, 1974).

[67] However, neither type is uniformly optimistic nor uniformly pessimistic. Rather optimism and pessimism stem from the interaction of the heuristics they employ and the recent history they observe.

[68] My examples in sections 3 and 4 feature zero learning: traders never adjust their estimates of the underlying branch probabilities. This is an extreme case. Traders do learn, albeit inefficiently.

beliefs greater weight, and returns will amplify the volatility in consumption growth. What determines the relative occurrence of the different types of segments? The objective process $\{\Pi, g\}$.

There is a widespread view that CRRA-based consumption growth models are incapable of explaining security returns. One reason for this view is the equity premium puzzle (Mehra and Prescott, 1985): a consumption based model seems incapable of explaining the historical equity premium in the U.S. market. This might be the case. However, it is important to recognize that the traditional arguments based on CRRA-based consumption models ignore traders' errors, and hence ignore sentiment. Instead, the burden of explaining the equity premium falls to complex dynamics in the underlying process, such as regime shifts (Whitelaw, 2000). However, regime shifts occur naturally when beliefs are heterogeneous and the beliefs of the representative trader are a formed as a wealth weighted combination of the beliefs of the individual traders.

In theory, the relative contribution of sentiment on the pricing kernel can be much larger than that of consumption growth. Fundamentals can get short shrift, at least along particular segments of the realized path. Think about theorems 4 and 5, which characterize the term structure in terms of aggregate consumption growth. These theorems suggest that information releases about consumption growth should be one of the most important pieces of news bond traders receive. Nevertheless, Balduzzi, Elton, and Green (1997) find that consumption growth is one of the least important variables influencing the Treasury market. Why? I suggest that the answer concerns the contribution of sentiment relative to fundamentals in determining returns in the short-run.

The process governing the evolution of sentiment $\lambda = \ln(P_R/\Pi)$ is based on the beliefs of a representative trader. I remind readers that in the presence of heterogeneity, the beliefs of the representative trader need not have the same structure as those of individual traders, and typically do not. The representative trader aggregates individual beliefs, but with stochastic, time-varying weights. In a prolonged period of favorable fundamentals, the representative trader will move in the direction of irrational exuberance. In a prolonged period featuring unfavorable fundamentals, the reverse will be true. In a period of rapidly fluctuating fundamentals, the representative trader will display frequent changes of opinion. Market returns will tend to become more positively autocorrelated along long runs, and move to being negatively autocorrelated during periods that feature short runs. In other words, the process governing λ will have a stochastic autocorrelation structure that depends on the history.

The discussion in the previous paragraph leads me to reiterate a methodological warning to theorists whose models begin by assuming a representative trader. It is dangerous to treat the representative trader as an individual trader. This applies to both models of rational behavior and non-rational behavior. In the presence of heterogeneity, the representative trader will typically be neither rational, nor be subject to the same type of error structure as a single individual trader would. Rather the representative trader is a dynamically evolving amalgam of all the individual traders.

There are additional behavioral elements concerning preferences, most notably loss aversion and reference point-based mental accounting. These features can be accommodated within the present framework, with minor modifications. Shefrin and Statman (1989) explore the implications for the pricing kernel that stem from the introduction of prospect theory preferences (Kahneman and Tversky, 1979). A prospect theory utility function is S-shaped, and is defined over consumption changes (i.e., gains and losses) relative to a reference point. The S-shape depicts risk aversion in the domain of gains, and risk seeking in the domain of losses.[69]

Shefrin and Statman (1989) demonstrate that the introduction of prospect theory traders tends to flatten the graph of the pricing kernel. Prospect theory traders tend to shun claims that pay off in loss states, typically those for which $m(x_t)$ is high, in exchange for claims that pay off in gains states, typically those for which $m(x_t)$ is low.[70] As the relative proportion, and wealth, of prospect theory traders increases, the graph of the pricing kernel begins to flatten in the loss states segment. If it flattens completely, turning horizontal, then risk averse agents will choose to hold portfolio insurance in equilibrium. They will do so because the conditional risk premium vanishes, where the conditioning is for states in which prospect theory traders register losses.

The pricing kernel equation (15) includes time varying stochastic terms for both risk tolerance θ_R and time preference δ_R. Notably, during periods where wealth shifts from

[69] See also Shefrin-Statman (2000).

[70] Prospect theory traders act as if they were CRRA-traders who attach low or zero probability to loss states. Why? Because nonconvexity in the indifference map for the domain of losses leads to boundary choices for prospect theory traders, just as zero probabilities do for CRRA-traders. In addition, Tversky and Kahneman (1992) postulate a power function for u_h, at least over the domain of gains and losses. One implication of using a probability transformation to mimic the effect of prospect theory traders is that it leaves the structure of the pricing kernel intact.

those who are less risk-tolerant to those who are more risk-tolerant, the representative trader becomes more risk tolerant. This phenomenon tends to make the representative trader more risk tolerant during up markets, a point made by Benninga and Mayshar (1993, 1997). A similar remark applies to time discounting. Recently, Barberis, Huang, and Santos (1999) have developed a prospect theory motivated model with the same time varying, stochastic risk tolerance properties described above. Specifically, they explain the equity premium puzzle by establishing that prior gains lead to an increase in the market's tolerance for risk, whereas prior losses lead to the reverse.

15 Conclusion

I conclude by summarizing the main results about heterogeneity. The central issues in this paper concern a formal definition for market sentiment, and the manner in which market sentiment distorts the pricing kernel and the returns to major asset classes: bonds, stocks, and options. I establish two main results. First, sentiment manifests itself through the log-likelihood ratio λ. Second, the (or more correctly, a) log-pricing kernel is the sum of two stochastic processes, a sentiment process λ and a fundamental process based on aggregate consumption growth.[71] Markets are efficient if and only if sentiment $\lambda = 0$ uniformly. When sentiment is nonzero, heterogeneity typically induces smile effects into option pricing, frown effects into mean-variance efficient portfolios, distortions that disrupt the expectations hypothesis of the term structure, and alterations to the return on the market portfolio, depending on the risk tolerance profile of the individual traders.

[71] As mentioned earlier, δ_R and θ_R are also stochastic, but the focus of attention is on λ.

References

Abel, A., 1988. "Stock Prices Under Time Varying Dividend Risk, An Exact Solution in an Infinite-Horizon General Equilibrium Model," *Journal of Monetary Economics* 22, 375-393.

Balduzzi, P., E. Elton, and T. Green, 1997. "Economic News and the Yield Curve: Evidence from the U.S. Treasury Market," Boston College/New York University working paper.

Bakshi, G., C. Cao, and Z. Chen, 1997. "Empirical Performance of Alternative Option Pricing Models, " *Journal of Finance* 52, 2003-2049.

Barberis, N., A. Shleifer, and R. Vishny, (1998) "A Model of Investor Sentiment," *Journal of Financial Economics*, Vol. 49, No. 3, 307-344.

Barberis, Nicholas, Ming Huang, and Tano Santos, 1999. "Prospect Theory and Asset Prices," Working paper, University of Chicago.

Bates, D., 1991. "The Crash of 87: Was it Expected? The Evidence from Options Markets," *Journal of Finance* 46, 1009-1044.

Bates, D., 1996. "Testing Option Pricing Models", in G.S. Maddala and C.R. Rao, *Statistical Methods in Finance/ Handbook of Statisitics*, Amsterdam: Elsevier, pp. 567-611.

Basak, S., 2000. "A Model of Dynamic Equilibrium Asset Pricing with Heterogeneous Beliefs and Extraneous Risk," *Journal of Economic Dynamics and Control*, 24, 63-95.

Beja, Avraham, 1978. "State Preference and the Riskless Interest Rate: A Markov Model of Capital Markets," *Review of Economic Studies*, 46, 435-446.

Benninga, S. and J. Mayshar, 1993. "Dynamic Wealth Redistribution, Trade, and Asset Pricing," Working Paper 8-93, Wharton School.

Benninga, S. and J. Mayshar, 1997. "Heterogeneity and Option Pricing," Working Paper, Wharton School.

Benninga, S. and A. Protopapadakis, 1983. "Real and Nominal Interest Rates under Uncertainty: The Fisher Theorem and the Term Structure," *Journal of Political Economy.* Vol 91, No.5, 856-867.

Campbell, J. and R. Shiller, 1984. "A Simple Account of the Behavior of Long Term Interest Rates," *American Economic Review*, 74, p. 44-48.

Campbell, J. and R. Shiller, 1991. "Yield Spreads and Interest Rate Movements: A Birds Eye View," *Review of Economic Studies*, 58, 495-514.

Campbell, J., 1995. "Some Lessons from the Yield Curve," *Journal of Economics Perspec-*

tives, 129-152.

Campbell, J., A. Lo, and A.C. MacKinlay, 1997. *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Carr, P. and D. Madan, "Optimal Positioning in Derivative Securities," Morgan Stanley/University of Maryland working paper.

Cox, J., J.E. Ingersoll, and S. Ross, 1985. "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385-407.

De Bondt, W. F. M. 1993. "Betting on Trends: Intuitive Forecasts of Financial Risk and Return," *International Journal of Forecasting*, Vol 9, 355-371.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1998. "A Theory of Overconfidence, Self-Attribution, and Security Market Under- and Over-reactions," *Journal of Finance*, Vol. 53, 1839-1886.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1999. "Investor Overconfidence, Covariance Risk, and Predictors of Securities Returns," working paper Northwestern University.

David, A. and P. Veronesi, 1999. "Option Prices with Uncertain Fundamentals," working paper, Board of Governors of the Federal Reserve System.

Derman, E. and I. Kani, "Riding on a Smile," *Risk* 7, 32-39.

Detemple, J. and S. Murthy, 1994. "Intertemporal Asset Pricing with Heterogeneous Beliefs," *Journal of Economic Theory* 62, 294-320.

Detemple, J. and S. Murthy, 1997. "Equilibrium Asset Prices and No-Arbitrage with Portfolio Constraints," McGill University/Rutgers University working paper.

Diz, F. and T.J. Finucane, 1993. "Do the Options Markets Really Overreact?," *Journal of Futures Markets*, 13, 298-312.

Emmanuel, D.C. and J.D. MacBeth, 1982. "Further Tests on the Constant Elasticity of Variance Option Pricing Model." *Journal of Financial and Quantitative Analysis*, 17, 533-554.

Glassman, J. and K. Hassett, 1999. *Dow 36,000: The New Strategy for Profiting from the Coming Rise in the Stock Market*, New York: Times Books.

González de la Mota, A., 2000. "The Relevance of the Market Price of Risk and Multi-Scale Stochastic Volatility for the Dynamics of Smile Curves: Insights from Endogenous Uncertainty and Heterogeneous Beliefs," Working paper Stanford University.

González de la Mota, A., 2000. “Essays on Asset Pricing and Risk- Management under Endogenous Uncertainty: the Infinite Dimensional Case,” Working paper Stanford University

Gorman, W., 1953. “Community Preference Fields,” *Econometrica*, Vol. 21

Grether, D., 1980. “Bayes Rule as a Descriptive Model: The Representativeness Heuristic,” *Quarterly Journal of Economics*, 95, 537-57.

Harris, M. and A. Raviv, 1991. “Differences of Opinion Make a Horserace,” *Review of Financial Studies*, v6(3), 473-506.

Heineke, J. and H. Shefrin, 1988. “Exact Aggregation and the Finite Basis Property,” *International Economic Review*, Vol. 29, No. 3, 525-538.

Jackwerth, J.C. and M. Rubinstein, 1996. Recovering Probability Distributions from Contemporaneous Security Prices,” *Journal of Finance*, Vol. 51, No. 5, pp. 1611-1631.

Karolyi, G.A., 1993. “A Bayesian Approach to Modelling Stock Return Volatility for Option Evaluation,” *Journal of Financial and Quantitative Analysis*, 28, 579-594.

Kurz, M. , 1997. *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs.* Studies in Economic Theory No. 6, Springer-Verlag: Berlin and New York.

Leland, H., 1999. “Beyond Mean-Variance: Performance Measurement in an Nonsymmetrical World,” *Financial Analysts Journal*, January/February, 27-36.

Lucas, R., 1978. “Asset Pricing in an Exchange Economy,” *Econometrica*, 46, 1429-1445.

Madan, D., F. Milne, and H. Shefrin, 1989. “The Multinomial Option Pricing Model and Its Brownian and Poisson Limits,” *Review of Financial Studies* 2, 251-265.

MacBeth, J.D. and Merville, 1980. “Tests of the Black-Scholes and Cox Call Option Valuation Models,” *Journal of Finance* 35, 285-301.

Mayshar, J., 1983. “On Divergence of Opinion and Imperfections in Capital Markets,” *American Economic Review*, Vol. 73, pp. 114-128.

Mehra, R. and E.C. Prescott, 1985.“The Equity Premium Puzzle,” *Journal of Monetary Economics*, Vol. 40, No. 2, 145-161.

Milne, F. and S. Turnbull, 1996. “Theoretical Methods for Security Pricing,” Queens University working paper.

Naik, V. and M.H. Lee, 1990. “General Equilibrium Pricing of Options on the Market Portfolio with Discontinuous Returns,” *Review of Financial Studies* 3, 493-522.

Odean, Terrance, 1998. “Volume, Volatility, Price, and Profit When All Traders Are Above

Average," *Journal of Finance*, Vol. 53, 1887-1934.

Rubinstein, M., 1973. "The Fundamental Theorem of Parameter-Preference Security Valuation," *Journal of Financial and Quantitative Analysis*, 8, 61-69.

Rubinstein, M., 1976. "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, 7, 407-425.

Rubinstein, M., 1985. "Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978," *Journal of Finance*, 40, 455-480.

Rubinstein, M., 1994. "Implied Binomial Trees," *Journal of Finance*, 49, 771-818.

Shefrin, H., 1984. "Inferior Forecasters, Cycles, and the Efficient-Markets Hypothesis: A Comment". *Journal of Political Economy*, Vol. 92, pp. 156-161.

Shefrin, H., 1999. "Irrational Exuberance and Option Smiles." *Financial Analysts Journal*, November/December, 91-103.

Shefrin, H. And M. Statman, 1989. "Introducing Prospect Theory into General Equilibrium: Implications for CAPM and Portfolio Insurance." Working paper, Santa Clara University.

Shefrin, H. And M. Statman, 1994. "Behavioral Capital Asset Pricing Theory," *Journal of Financial and Quantitative Analysis* 29, 323-349.

Shefrin, H. And M. Statman, 2000. "Behavioral Portfolio Theory," *Journal of Financial and Quantitative Analysis* 35, 127-151.

Shiller, R., 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends," reprinted as Ch. 4 in *Advances in Behavioral Finance*, R. Thaler, ed., Russell Sage Foundation, New York, NY.

Shiller, R., 1990. "The Term Structure of Interest Rates," in Friedman, B. and F. Hahn, eds., *Handbook of Monetary Economics.* Volume 1, Amsterdam: North Holland, 627-722.

Shiller, R., 2000. *Irrational Exuberance*, Princeton: Princeton University Press.

Siegal, J. and R. Thaler, 1997. "The Equity Premium Puzzle," *Journal of Economics Perspectives*, 11,1: 191-200.

Stein, J., 1989. "Overreactions in the Options Market," *Journal of Finance* 44, 1011-1023.

Tversky, A. and D. Kahneman, 1971. "Belief in the Law of Small Numbers," *Psychological Bulletin*, 105-110.

Tversky, A. and D. Kahneman, 1974. "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185, pp. 1124-1131.

Whitelaw, R., 2000. "Stock Market Risk and Return: An Equilibrium Approach," *Review of Financial Studies*, Vol. 13, No. 3, 521-547.

Appendix

This appendix contains proofs of theorems 1,3,6, and 8. Theorem 2 is essentially proved in the body of the paper. Theorem 4 follows directly from equation (6). Theorem 5 is proved in the body of the paper. The proof of theorem 7 is similar to the proof of theorem 6, and relies on equation (6).

Proof of theorem 1.

Define

$$\gamma_h(x_t) = \frac{c_h(x_0)(D_h(x_t))^{1/\theta_h}}{\sum_{j=1}^{H} c_j(x_0)}$$

$$\gamma(x_t) = \sum_{h=1}^{H} \gamma_h(x_t)$$

Define trader h's x_0-consumption share

$$\xi_h = \frac{c_h(x_0)}{\sum_{j=1}^{H} c_j(x_0)}$$

Now use (5) to compute the equilibrium value of $g(x_t)$. Obtain

$$g(x_t) = \frac{\sum_h c_h(x_t)}{\sum_h c_h(x_0)}$$

$$= \sum_h \xi_h (D_h(x_t)/v(x_t))^{1/\theta_h}$$

$$= \sum_h \gamma_h(x_t) v(x_t)^{-1/\theta_h}$$

where $\gamma_h(x_t)$ is defined in the statement of Theorem 1. Notice that $\gamma_h(x_t)$ is a consumption weighted discounted probability moment.

Invert equation (6) to obtain an expression for $g(x_t)$, which can be equated to the preceding expression. This yields

$$\sum_h \gamma_h(x_t) v(x_t)^{-1/\theta_h}$$

$$= (\delta_R^t P_R(x_t)/v(x_t))^{1/\theta_R(x_t)}$$

where $P_R(x_t)$ is endogenously determined.

Theorem 1 specifies two alternative terms for θ_R, (9) and (13). The first is a function of Π, and is derived from Benninga-Mayshar (1993). See below. Note that (5) and $\sum_h c_h(v) = \sum_h \omega_h$ imply that $\sum_h \alpha_h(x_t) = 1$ for all x_t.

The second is defined independently of Π. Note that the term $\gamma(x_t)$ is the sum of the weights used to define θ_R in (13). This term provides the structure of $\theta_R(x_t)$ (in equation (13)), or more properly $1/\theta_R(x_t)$.

Define the representative trader's discounted probability moment by

$$(\delta_{R,t}^t P_R(x_t))^{1/\theta_R(x_t)} = \gamma(x_t)$$

Next substitute into (6) and take the logarithm of both sides to obtain

$$1/\theta_R(x_t) = \frac{\ln(\gamma(x_t)) - \ln(g(x_t))}{\ln(v(x_t))}$$

Having θ_R in hand, either (9) or (13), enables us to find $\delta_R^t P_R(x_t)$ by taking $\gamma(x_t)$ to the power $\theta_R(x_t)$. Obtain $\delta_{R,t}^t$ as the sum, for fixed t,

$$\sum_{x_t} \gamma(x_t)^{\theta_R(x_t)}$$

In view of the normalization, obtain $P_R(x_t)$ as the ratio

$$P_R(x_t) = \frac{\gamma(x_t)^{\theta_R(x_t)}}{\delta_{R,t}^t}$$

For sake of completeness, I sketch the proof provided by Benninga and Mayshar (1993) characterizing θ_R. Based on (5), the equilibrium condition $\sum_h (c_h(v) - \omega_h) = 0$, and the kernel variable $V = v/\Pi$, Benninga-Mayshar define the implicit function

$$F(C, V) = \sum_h (c_h(x_0)/C)[\delta_h^t/V]^{1/\theta_h} = 1$$

They note that by the principle of expected utility maximization, the representative trader's marginal utility at C will be proportional to V. In turn, this implies that θ_R, the Arrow-Pratt coefficient of relative risk aversion can be defined locally by $-CV'(C)/V(C)$. By computing $\partial F/\partial C$ and $\partial F/\partial V$, they observe that

$$V'(C) = \frac{-\partial F/\partial C}{\partial F/\partial V} = \frac{-(V'(C)/C)}{\sum_h \alpha_h(x_t)/\theta_h}$$

which, taken together with the local Arrow-Pratt measure, completes their proof.

Proof of theorem 3.

To prove Theorem 3, compute the first-order-condition associated with maximizing expected quadratic utility.

$$r_{MV}(x_1) = \nu^{-1}(1 - (\xi v(x_1)/\Pi(x_1))) \tag{34}$$

where ξ is the Lagrange multiplier for the optimization and has the form:

$$\xi = \frac{\sum_{y_1} v(y_1) - \nu}{\sum_{y_1} v(y_1)^2/\Pi(y_1)} \tag{35}$$

Next observe that

$$M(x_1) = v(x_1)/\Pi(x_1) = \delta_{R,1}\Lambda(x_1)g(x_1)^{-\theta_R}$$

$$v(x_1)^2/\Pi(x_1) = \Pi(x_1)M(x_1)^2$$

$$\sum_{y_1} v(y_1) = E_\Pi\{M(y_1)\} = \delta_{R,1}E_\Pi\{\Lambda(x_1)g(x_1)^{-\theta_R(x_1)}|x_0\}$$

$$\sum_{y_1} v(y_1)^2/\Pi(y_1) = E_\Pi\{M(y_1)^2\}$$

Substitution into (35) completes the proof.

Proof of theorem 6.

From the perspective of x_{t-1}, $\eta(x_t)$ is the future value of a contingent x_t real dollar payoff. Given x_{t-1}, the future value of a contract which delivers a certain dollar at date t must be one dollar. This is why $\sum_{y_t \in S(x_{t-1})} \eta(y_t) = 1$. In other words, the future value of y_t-claims are nonnegative and sum to unity. Hence they constitute a probability distribution. Since they deal with the transition from x_{t-1}, $\{\eta(y_t)\}$ are one-step branch probabilities of a stochastic process.

Under the stochastic process, the probability attached to the occurrence of x_t is obtained by multiplying the one-step branch probabilities leading to x_t. To interpret this product, consider the denominator of (27). This term can be matched with the numerator of the x_{t-1} one-step branch probability to form $v(x_{t-1})/\sum_{y_t \in S(x_{t-1})} v(y_t)$. The latter term is simply one plus the single period risk free interest rate $i_1(x_{t-1})$ that applies on the x_{t-1}-market. Therefore the probability of the branch leading to x_t is the product of the single period stochastic interest rates and the present value of an x_t-claim: $i_1(x_0)i_1(x_1)\cdots i_1(x_{t-1})v(x_t)$. The product of the single period interest rates defines the cumulative return $i_c^t(x_t)$ to holding the short-term risk-free security, with reinvestment, from date 0 to date t.

A call option pays $q_z(x_t)-K$ at date t, if $x_t \in A_E$, the set of date-event pairs where the option expires in-the-money. The present value of the claims that make up the option payoff is computed using state prices v. But the present value of an x_t-contingent dollar is its future value discounted back by the product of the one-period risk-free rates. The discounted contingent future dollar is simply the ratio of a risk-neutral probability $\eta(x_t)$ to a compounded interest rate $i_c(x_t)$. Finally, the risk-neutral probability $\eta(x_t)$ is unconditional. To convert to a distribution conditional on exercise, divide $\eta(x_t)$ by $P_\eta\{A_E|x_0\}$. Using the conditional expectation in place of the unconditional expectation leads to the appearance of $P_\eta\{A_E|x_0\}$ in (28).

Proof of theorem 8.

The proof of this theorem is computational. The one-period return to the market portfolio is the sum of the date 1 dividend and date 1 price, divided by the date 0 price, i.e. $(\omega(x_1) + q_1(Z_\omega))/q_0(Z_\omega)$. Use (8) to compute the present values two future aggregate consumption stream: the value of the unconditional process under v, and the value of the process, conditional on x_1. The present value of these two streams appear respectively, in the denominator and numerator of (33), with the numerator value divided by $v(x_1)$. This completes the proof.