Asset Prices, Heterogeneous Expectations, and Limited Short Sales

Cheolbeom Park

July 23, 2003

Abstract: This paper extends the Harrison-Kreps model by allowing limited short sales. The main results of this paper are: (1) investors pursue short-term gains when perceiving heterogeneous expectations; (2) important properties of the equilibrium price in the Harrison-Kreps model still hold even when limited short sales are allowed; (3) an increase in the dispersion of expectations about future dividends raises the risky asset price; and (4) an increase in short-sale costs also raises the risky asset price.

JEL classification: G10, G12
Keywords: Heterogeneous expectations, Dispersion in expectations, Limited short sales, Short-sale cost
Introduction

This paper is an extension of Harrison and Kreps’ (1978) model focusing on the relationship between the price of a risky asset and heterogeneous expectations. In their seminal paper, Harrison and Kreps (1978) show that the equilibrium asset price generally exceeds the most optimistic investor’s present value of future dividend stream, assuming heterogeneous expectations with no Bayesian learning, no short sales, and infinite wealth. Although the model did not attract economists’ attention immediately, their insights have shed light on certain puzzles in the financial markets that have been observed over the last few decades. Morris (1996) generalizes the Harrison-Kreps model by allowing the Bayesian learning process, and applies the model to explain the puzzle of initial public offerings. Scheinkman and Xiong (2003) extend the Harrison-Kreps model in a continuous time framework, and derive a relation between price, price volatility, and trading volume which is consistent with empirical regularities in the stock market.

This paper generalizes the Harrison-Kreps model by allowing limited short sales, and shows that the risky asset price increases with the dispersion of expectations on future dividends. This paper also predicts that the risky asset price increases with short-sale costs. These predictions are consistent with recent empirical findings in Chen et al. (2002), Diether et al. (2002), and Jones and Lamont (2002).

The explanation of the predictions in this paper is simple. If limited short sales make the equilibrium risky asset price reflect a more optimistic evaluation than the median expectation across investors, then an increase in the dispersion of expectations while keeping the median expectation constant will result in a higher price because the marginal investor’s expectation \(^1\) becomes more optimistic due to the increase in the
dispersion. In other words, the increase in the dispersion of expectations makes demand equal to supply at a more optimistic willingness to pay for the risky asset. In addition, an increase in short-sale cost results in a lower fraction of risky-asset holders to all investors because of the reduction in the short sales, which is the supply created by investors. As a result, the equilibrium will be regained at a higher willingness to pay, which results in a higher equilibrium price. These results are consistent with recent empirical findings in the literature. Hence, this paper can be interpreted as a formalized version of Miller (1977).

While demonstrating the above, this paper also clarifies that all investors pursue short-term gains when perceiving heterogeneous expectations, and that their willingness to pay for the risky asset includes not only the present value of future dividends but also the component of the speculative premium, which is the expected present value of the gap between the future equilibrium price and willingness to pay. This is the consequence of the perception of differences in expectations, and the existence of an alternative risk-free asset. In addition, the speculative premium could be a reason for the failure of the present value model in the stock market, which is widely reported in empirical studies.

This paper is structured as follows. Section I presents a model to show how investors determine their willingness to pay for a stock when perceiving heterogeneous expectations. Section II shows that important properties of the equilibrium price in the Harrison-Kreps model still hold even when limited short sales are allowed. Section III provides an application of the model to examine relations between the risky asset price, the dispersion in expectations, and short-sale costs. Section IV contains concluding remarks.
Consider a financial market with a risky asset and a risk-free asset. If an investor holds a share of the risky asset between time $t$ and $t+1$, she receives dividends, denoted by $D_{t+1}$, immediately prior to time $t+1$ and can resell at time $t+1$. The dividend process is perceived by investors as a stationary Markov process with state space $\{s_1, s_2, \ldots, s_N\}$. The dividends are characterized by the possible states, $D(s_i) \geq 0$, where $i = 1, 2, \ldots, N$. That is, in every period, dividends will be one element from the set $\{D(s_1), \ldots, D(s_N)\}$, which is time independent. Without loss of generality, I assume that $D(s_1) > \cdots > D(s_N)$.

If an investor holds a risk-free asset, she receives a constant return $r$. For simplicity, investors are assumed to be risk-neutral and to have an identical discount rate ($\gamma$) equal to $\frac{1}{1+r}$. Investors can go short by a finite amount (e.g. a fraction of their wealth) when they become pessimistic about the risky asset. There are an infinite number of investors indexed by an $N$-dimensional vector whose elements are randomly drawn real numbers between 0 and 1. Once numbers are drawn for the index vector, the investor will keep the identity index forever. More formally, the relation between an individual investor’s identity and priors (i.e. transition matrices) can be written as follows:

**Definition 1.** An investor’s identity “$a$” is defined as an $N \times 1$ vector $a = [a_1, a_2, \ldots, a_N]'$ and $a_i \in [0, 1]$ for $i = 1, 2, \ldots, N$.

**Definition 2.** Define $A$ as an $N \times N$ matrix corresponding to “$a$” such as:
\[ A_{ij} = a_i \quad \text{if } i = j \]
\[ A_{ij} = 0 \quad \text{otherwise.} \]

I assume that an individual investor’s transition matrix \((M^a)\) can be written as \(M^a = AM_0 + (I - A)M_1\), where each row in \(M_0\) and \(M_1\) is a point in the \(N\)-dimensional simplex. \(M_0\) (\(M_1\)) can be interpreted as a matrix containing the most pessimistic (optimistic) probability assessments about the dividends in each state. Then, each row in \(M_1\) first-order stochastically dominates the corresponding row in \(M_0\) given \(\{D(s_1), \ldots, D(s_N)\}\), and each element of \((M_1 - M_0)D\) is positive where \(D = [D(s_1), D(s_2), \ldots, D(s_N)]^\top \neq 0\). This assumption regarding \(M_0\) and \(M_1\) ensures that an investor’s expectations of the next period dividends will decrease with \(a_i\) in each state without loss of generality. Once an investor’s identity and prior (i.e. transition matrix) are determined at the beginning, then they will remain fixed for the remaining periods.\(^3\)

Under this set-up, an investor decides whether to buy the risky asset or the risk-free asset in her portfolio and how long to hold the asset to maximize wealth.

**Definition 3.** \(P_t^a(s_i)\) is defined as an investor’s willingness to pay for one unit of the risky asset at time \(t\) in state \(s_i\) where the optimal investment decision rule is:

Buy the risky asset if \(P_t(s_i) < P_t^a(s_i)\), and buy the risk-free asset if \(P_t(s_i) > P_t^a(s_i)\),

where the superscript “\(a\)” denotes an investor, and \(P_t(s_i)\) is the price of the risky asset at time \(t\) in state \(s_i\).
Definition 4. \( A^a_t(s_i) \equiv \{ E^a_t[D_{t+1} + \gamma P_{t+1} \mid s_i], E^a_t[D_{t+1} + \gamma^2 D_{t+2} + \gamma^3 P_{t+2} \mid s_i] \ldots \} \),

\[
E^a_t[\sum_{i=1}^{\infty} \gamma^i D_{t+i} \mid s_i] \text{, where } E^a_t[\cdot \mid s_i] \text{ is the conditional expectation operator for investor}
\]

“a” at time \( t \) in state \( s_i \).

\( A^a_t(s_i) \) is defined as an investor’s set of all strategies for buying one unit of the risky asset and holding it for \( k \) (\( k = 1, 2, \ldots \)) periods at time \( t \) in state \( s_i \). The \( k \)-th element in \( A^a_t(s_i) \) represents the expected present worth of the risky asset when buying at time \( t \) and holding it for \( k \) periods. An investor decides which element in \( A^a_t(s_i) \) should be used as \( P^a_t(s_i) \) in order to determine whether to buy the risky asset or the risk-free asset and how long the selected asset should be held based on all available information and her prior. An investor’s optimization problem is, therefore, a choice of elements in \( A^a_t(s_i) \) to compare to \( P^a_t(s_i) \), which represents the expected present value of the wealth when \( P^a_t(s_i) \) is invested in the risk-free asset regardless of the holding period. In other words, the investor decides which strategy in \( A^a_t(s_i) \) should measure the willingness to pay for the risky asset (\( P^a_t(s_i) \)). Note that strategies in \( A^a_t(s_i) \) do not require any actual commitment after time \( t \) since there is no holding period obligation for the risky asset or the risk-free asset. In other words, investors can change their investment strategy without cost as their expectations change.

Under the assumption of homogeneous expectations, all elements in \( A^a_t(s_i) \) are equal. As a result, the optimization problem is trivial, and the willingness to pay for the
risky asset can be written as the present value of all future dividends. In equilibrium, the present value of all future dividends equals the price of the risky asset. Under the assumption of heterogeneous expectations, however, an element in $A^a_t(s)$ does not necessarily equal other elements since an investor’s valuation can differ from the market equilibrium price. In the following proposition, I show that only the first element in $A^a_t(s)$ will be used as the willingness to pay for a stock when an investor maximizes her expected wealth.

**Proposition 1.** When an investor maximizes her expected wealth,

(1) \[ P^a_t(s) = E^a_t[\rho D_{t+1} + \rho P_{t+1} | s] \].

*Proof.* The proof is provided in the Appendix. ■

One might conjecture that the maximum in $A^a_t(s)$ should be the willingness to pay for the risky asset because it shows the maximum expected present worth that an investor can obtain by purchasing the risky asset at time $t$. However, this conjecture neglects the opportunity cost of not buying the risk-free asset. Suppose that an investor is expecting that the price of the risky asset will rise sharply at time $t+k$. Although the risky asset should be attractive to this investor at time $t$, the investor should not buy the stock at time $t$ if she expects the rate of return from the risky asset between time $t$ and $t+1$ to be lower than that from the risk-free asset. The investor should postpone the purchase of the risky asset until it can yield a relatively higher one-period rate of return than the risk-free asset.

In the proof, I show that one can always find an alternative way of increasing the expected discounted wealth if the holding period is longer than one period. Intuitively,
the trading strategy using Equation (1) as the willingness to pay for the risky asset is optimal because it does not sacrifice any option of buying the risk-free asset while other strategies must forgo some of these options. Although Harrison and Kreps (1978) and Scheinkman and Xiong (2003) show that the equilibrium price has the property shown in Proposition 1 under the assumptions of heterogeneous expectations and no short sales, Proposition 1 states that all investors are targeting only short-term gains when they perceive heterogeneous expectations among themselves. This investment strategy is the consequence of the perception of differences in expectations, and the existence of an alternative risk-free asset.

Proposition 1 can also justify the short-horizon assumption assured in the model of De Long et al. (1990) with two types of investors (noise traders and rational investors) and overlapping generations. Although De Long et al. (1990) justify the assumption by frequent evaluation of financial managers, Proposition 1 shows that this is unnecessary because even long-lived investors care about short-term performance if they are aware of heterogeneous expectations.

With Proposition 1, one may ask how stock prices should respond to an announcement that dividends at $t+k$ ($k>1$) will be raised. The answer is given in the following corollary, which can be obtained by the use of Proposition 1.

**Corollary 1.** Suppose that \[ \lim_{j \to \infty} \gamma^j E_t^a [P_{t+j} - P_{t+j}^a | s_i] = 0 \]. Then,

\[
(2) P_t^a (s_i) = E_t^a [\mathcal{P}_{t+1} + \mathcal{P}_{t+1} | s_i] = E_t^a \left[ \sum_{j=1}^{\infty} \gamma^j D_{t+j} | s_i \right] + E_t^a \left[ \sum_{j=1}^{\infty} \gamma^j (P_{t+j} - P_{t+j}^a) | s_i \right]
\]

for all $t$ and $s_i$. 

Proof. \( P^a_t(s_i) = E_i^a[\gamma D_{t+1} + \gamma P_{t+1} | s_i] = E_i^a[\gamma D_{t+1} + \gamma P_{t+1} - \gamma P_{t+1} | s_i] \)

\[ = E_i^a[\gamma D_{t+1} + \gamma P_{t+1} - \gamma P_{t+1} + \gamma (P_{t+1} - P_{t+1}) | s_i] \]

\[ = E_i^a[\gamma D_{t+1} + \gamma D_{t+1} + \gamma (P_{t+1} - P_{t+1}) + \gamma^2 P_{t+1} | s_i] \]

\[ = ... = E_i^a[\sum_{j=1}^{\infty} \gamma^j D_{t+j} | s_i] + E_i^a[\sum_{j=1}^{\infty} \gamma^j (P_{t+j} - P_{t+j}) | s_i]. \]

The fourth equality holds due to the of law iterative projections. ■

Corollary 1 shows that investors actually consider the dividend and resale option in the distant future. In other words, Corollary 1 states that an investor’s willingness to pay incorporates both the present value of future dividends and adjustments to market valuations. Especially, for example, investors whose future evaluation is pessimistic increase their current willingness to pay because of expected gains from a market price in the future. In contrast, relatively optimistic investors reduce their current willingness to pay because of expected losses from a market price in the future. 4

Corollary 1 might be interpreted as a formalization of Keynes’ “beauty contest,” where investors are interested in the opinions of others as well as their own estimates. Harrison and Kreps (1978) and Morris (1996) call the second term in Equation (2) the speculative premium, while Scheinkman and Xiong (2003) call it a speculative bubble. The speculative premium is the expected resale gain due to the awareness of heterogeneity in expectations, and is contained in individual investors’ willingness to pay for the risky asset. Like the rational bubbles in Blanchard and Watson (1982), it has no relationship with fundamentals. However, unlike the rational bubbles, it can be non-explosive (shown in Harrison and Kreps (1978)) and can affect expected returns. Finally,
the speculative premium could be another source of stock price volatility, as Scheinkman and Xiong (2003) point out.

II. Market Equilibrium

This section discusses the characteristics of the price of the risky asset when investors have heterogeneous expectations and when limited short sales are allowed. In practice, short sales require greater costs than regular sales or purchases. For example, most short sellers do not receive the full sale proceeds immediately, and margin funds must be maintained with the broker through whom a short seller makes the short sale. Thus, it is assumed that investors can go short with only a fraction of their wealth.

In addition to the assumption of limited short sales, every investor is assumed to have exactly the same level of financial wealth so that each investor’s perception of dividends (i.e. the transition matrix) is independent of her financial wealth level. The level of financial wealth is exogenously given and independent of states, and investors can go long (short) up to the level (a fraction) of their own financial wealth. Therefore, the demand schedule will be:

\[
Q_t^a = \begin{cases} 
  w_t^a & \text{if } P_t^a(s_i) \geq P_t(s_i) \\
  -\frac{w_t^a}{c} & \text{if } P_t^a(s_i) < P_t(s_i)
\end{cases}
\]

where \(w_t^a\) denotes investor \(a\)’s financial wealth in terms of the risky asset and \(c \geq 1\) represents the cost of short sales. If \(c\) is equal to 1, then the assumption implies that short sales incur no cost. As \(c\) increases, short sales incur more cost. If \(c\) is infinite, no short sales are allowed. Finally, the supply of the risky asset originally issued by firms is also assumed to be an exogenously given constant.
Since investors are interested in information about the next period, the date-state pair for investors’ plan is finite. In addition, investors’ preferences, demand schedules, and budget constraints satisfy conditions in Radner (1972). This implies that the market under the above assumptions must have an equilibrium.

**Proposition 2.** Under the assumptions made in sections I and II, there exists an equilibrium.

*Proof.* See Radner (1972). □

Note that all investors agree on what the equilibrium price is in each state, but they have different subjective probability assessments associated with the states. Each investor knows her optimal investment decision in each state, according to Proposition 1, and investors’ interactions will have the equilibrium price. Another interesting result will emerge in this market.

**Proposition 3.** The fraction of the risky asset holders to all investors is constant and independent of time and state.

*Proof.* The equilibrium condition in state \(i\) at \(t\) can be written as:

\[
W_t \int_{\{p_t^i(s_i) \geq p_t(s_i)\}} \varphi_t(a_i) da_i - \frac{W_t}{c} \int_{\{p_t^i(s_i) < p_t(s_i)\}} \varphi_t(a_i) da_i = \overline{Q}_t,
\]

where \(W_t\) is the aggregate financial wealth in terms of the risky asset, \(\varphi_t(a_i)\) is the probability density function of the willingness to pay across investors in state \(i\) at period \(t\), and \(\overline{Q}_t\) is the number of shares originally issued by firms. Define \(\alpha_t = \int_{p_t^i(s_i) \geq p_t(s_i)} \varphi_t(a_i) da_i\).
Since both $\overline{Q}_i$ and $W_i$ are constant numbers, $\alpha_i = \frac{1}{c+1} + \frac{c\overline{Q}_i}{(c+1)W_i}$ is also constant over time and across states. Therefore, $\alpha_i = \alpha$ for all $t$ and $i$. ■

The fact that the fraction of risky asset holders among all potential investors is always constant has two interesting implications. First, movements of the risky asset price can be explained by the changes of the distribution of the willingness to pay across investors, $W_i$, and $\overline{Q}_i$. For example, given $W_i$ and $\overline{Q}_i$, if investors’ willingness to pay is normally distributed, movements of the risky asset price can be explained by the movements in the first and second moments in the distribution. Second, the information of the fraction of risky asset holders to all investors makes it easy to find the equilibrium price without relying on the inductive argument when investors’ expectations are weighted averages between two extreme views. This will be discussed after considering the following proposition concerning the uniqueness of the stationary price of the risky asset.

**Proposition 4.** If $a_i$ is identically distributed in each state, and if each row in $M_0$ and $M_1$ first-order stochastically dominates the next row from the top row of each matrix with $D = [D(s_1),\ldots,D(s_N)]'$ given, then there exists a unique stationary equilibrium price that is time independent but state dependent.

*Proof.* The proof is provided in the Appendix. ■
The condition in Proposition 4 (each row in $M_0$ and $M_1$ first-order stochastically dominates the next row from the top row of each matrix with $D = [D(s_1), \ldots, D(s_N)]'$ given) means that investors become more optimistic when more dividends are paid out or when the economy is in a better state. Then, Proposition 4 states the uniqueness of the stationary price even when limited short sales are allowed. In addition, the proof of Proposition 4 shows an easy way of finding the prior (the transition matrix) of the marginal investor whose willingness to pay is equal to the equilibrium price in every state. When all investors’ expectations can be expressed as weighted averages between the most pessimistic and the most optimistic expectations in every state, the equilibrium price is also a weighted average of two extreme views. If $a_i$ is identically distributed in each state, then the weight for the equilibrium price, which is the marginal investor’s identification number in each state, is determined by the fraction of the risky asset holders as follows:

$$A^*_j = a^*_i = a^*_j = F_i^{-1}(\alpha) \text{ if } i = j, \text{ and } A^*_j = 0 \text{ if } i \neq j,$$

where $A^*$ is the $N \times N$ weight matrix for the marginal investor and $F_i(a_i)$ is the distribution function of $a_i$. Hence, the transition matrix of the marginal investor can be written as $M^* = A^* M_0 + (1 - A^*) M_1$.

If one knows the transition matrix of the marginal investor, one can easily find the equilibrium price without relying on the inductive argument in previous studies. Since the marginal investor decides the willingness to pay as in Proposition 1, and since the marginal investor’s willingness to pay is equal to the equilibrium price, the equilibrium price will be:
Since Equation (5) can be written as $P^* = (I - \gamma M^*)^{-1} \gamma M^* D = \gamma M^* D + \gamma^2 M^* D + \cdots$, the equilibrium price is the marginal investor’s present value of future dividends. Only the marginal investor considers that the equilibrium price has no speculative premium. From the viewpoints of all other investors, the equilibrium price does contain a speculative premium.6

III. Dispersion in Expectations and the Equilibrium Price

This section investigates the application of the model presented in the previous sections to explain recent empirical findings in the literature. Chen et al. (2002) and Diether et al. (2002) report that high dispersion in expectations results in low subsequent stock returns, and interpret this as a piece of evidence that stock prices are overpriced due to binding short-sale constraints. This empirical finding can be explained by the original Harrison-Kreps model. Since no short-sale assumption makes the equilibrium price be the maximum willingness to pay across investors, the increase in the dispersion of expectations will raise the stock price in the Harrison-Kreps model. However, previous studies with no short-sale assumption have difficulty in explaining another empirical finding in Jones and Lamont (2002) that high short-sale cost also predicts low stock returns in the future. This section shows that the model presented in this paper is able to show that the price of the risky asset can be raised not only by high dispersion in expectations but also by high short-sale cost.

A simple comparative static exercise in the following corollary can provide an explanation for how variations in short-sale cost influence the risky asset price.
Corollary 2. As short-sale cost increases, the fraction of the risky asset holders to all potential investors will decline.

Proof. Since \( \alpha = \left( \frac{1}{c+1} + \frac{cQ_t}{(c+1)W_t} \right) \) from Proposition 3, and since \( W_t > Q_t \),

\[
\frac{d\alpha}{dc} = \left( -\frac{1}{(c+1)^2} + \frac{Q_t}{(c+1)^2 W_t} \right) < 0 .
\]

Since \( F_i(a^*) = \alpha = \frac{1}{c+1} + \frac{cQ_t}{(c+1)W_t} \) from Proposition 3 and Equation (4), the increase in short-sale cost results in a decrease in \( \alpha \) as well as \( a^* \). Since willingness to pay decreases with \( a_i \) in state \( i \), the new transition matrix of the marginal investor becomes more optimistic and the new equilibrium price will rise. As a result, short-sale cost can predict low stock returns in the future.

Finally, this section also examines under what conditions high dispersion predicts low stock returns when limited short sales are allowed. This paper interprets the increase in the dispersion in expectations as the situation when \( M_1 \) becomes more optimistic and \( M_0 \) becomes more pessimistic in all states, keeping the median expectation constant. For simplicity, I assume that there are only two states where \( D(s_1) > D(s_2) \), and that every entry in both \( M_0 \) and \( M_1 \) is positive. Then, the dispersion variable, \( x \), modifies \( M_0 \) and \( M_1 \) as follows:

\[
M_0 = \begin{bmatrix} a_{11} - x & 1 - a_{11} + x \\ a_{21} - x & 1 - a_{21} + x \end{bmatrix} \quad \text{and} \quad M_1 = \begin{bmatrix} b_{11} + x & 1 - b_{11} - x \\ b_{21} + x & 1 - b_{21} - x \end{bmatrix}.
\]
**Proposition 5.** If $a_i$ is identically uniformly distributed on the [0, 1] interval in each state, and if the first row in $M_0$ and $M_1$ first-order stochastically dominates the second row with $D = [D(s_1), D(s_2)]'$ given, then

$$\frac{\partial P(s_i)}{\partial x} = \frac{\gamma(D(s_1) - D(s_2))(1 - 2a^*)}{(1 - \gamma)(1 - \gamma(c_{11} - c_{21}))} \text{ for } i = 1, 2,$$

where $c_{11} = a^*a_{11} + (1 - a^*)b_{11}$ and $c_{21} = a^*a_{21} + (1 - a^*)b_{21}$.

**Proof.** The proof is provided in the Appendix. 

From Equation (4), $a^*$ (the identity number of the marginal investor) has a positive relation with $\alpha$ (the fraction of the risky asset holders to all investors). Hence, if high short-sale cost makes $\alpha$ small enough so that $a^* < 0.5$, then the increase in the dispersion of expectations raises the price of the risky asset regardless of whether the economy is in a good state or a bad state. Like Proposition 5, if $a_i$ is uniformly distributed between 0 and 1, then $a^* = \alpha$. In this case, the increase in the dispersion of expectations raises the price of the risky asset when the fraction of the risky asset holders to all investors is less than 50%. Intuitively, this means that investors perceive that the effect from greater resale gains will be more likely to dominate the effect from more short sales as expectations become more diverse.

Interestingly, if there is no short-sale cost ($c = 1$), then $\alpha > 0.5$, and the increase in the dispersion could lower the risky asset price when $a_i$ is uniformly distributed between 0 and 1. Therefore, the empirical finding in recent studies that high dispersion in expectations predicts lower stock returns in the future could be interpreted by the model.
as a piece of evidence that the stock price is overpriced due to high short-sale cost in practice.

IV. Conclusions

This paper has explicitly shown that important characteristics in the Harrison-Kreps model still hold even when limited short sales are allowed. The paper has also clarified that individual investors’ willingness to pay for the risky asset contains the speculative premium component, which has no relation to the fundamentals of the risky asset and is the present value of future resale gains. Finally, the predictions of the model are consistent with recent empirical findings that both high dispersion in expectations and high short-sale costs forecast low stock returns in the future.

The number of theoretical models that analyze the consequences of interactions among investors with heterogeneous expectations is increasing. This may be because there are more and more pieces of evidence that cannot be explained by models with homogeneous expectations. The predictive power of the dispersion in analysts’ forecasts (Diether et al. (2002)), the predictive power of the breadth of ownership (Chen et al. (2002)), and trading volume and heterogeneous expectations (Kandel and Pearson (1995)) are just a few examples of such pieces of evidence. The Harrison-Kreps model is believed to be an important cornerstone to an understanding of those phenomena, and this paper has provided a limited extension of the model. However, there are still many aspects of the model which should be generalized further. For example, interactions among risk-averse investors with heterogeneous expectations, and a full consideration of wealth accumulation in the model would be interesting topics for future research.
Appendix

Proof of Proposition 1. Define $\hat{P}_t^a(s_t) = E_t^a[\sum_{i=1}^k \gamma^i D_{t+i} + \gamma^k P_{t+k} | s_t]$ for $k > 1$. Also, define $\bar{P}_t^a(s_t) = E_t^a[\gamma D_{t+1} + \gamma P_{t+1} | s_t]$. Suppose that an investor has found an optimal investment strategy after time $t+k$.

i) Suppose that $P_t(s_t) < \bar{P}_t^a(s_t) < \hat{P}_t^a(s_t)$. Since both $\bar{P}_t^a(s_t)$ and $\hat{P}_t^a(s_t)$ are greater than $P_t(s_t)$, the investor buys the risky asset when either $\bar{P}_t^a(s_t)$ or $\hat{P}_t^a(s_t)$ is used as the willingness to pay at time $t$. However, $\hat{P}_t^a(s_t)$ requires that the investor hold the risky asset for the first $k$ periods, which is not optimal unless $E_t^a[P_{t+1} | s_t] < E_t^a[\gamma D_{t+2} + \gamma P_{t+2} | s_t]$, $E_t^a[P_{t+2} | s_t] < E_t^a[\gamma D_{t+3} + \gamma P_{t+3} | s_t]$, ..., and $E_t^a[P_{t+k-1} | s_t] < E_t^a[\gamma D_{t+k} + \gamma P_{t+k} | s_t]$ hold. If any $E_t^a[P_{t+h-1} | s_t] > E_t^a[\gamma D_{t+h} + \gamma P_{t+h} | s_t]$ for $1 < h < k$, then the investor expects greater returns by switching the risky asset to the risk-free asset at time $t+h-1$ than by buying the risky asset at time $t$ and holding it for the remaining periods consecutively. Even when $E_t^a[P_{t+1} | s_t] < E_t^a[\gamma D_{t+2} + \gamma P_{t+2} | s_t]$, $E_t^a[P_{t+2} | s_t] < E_t^a[\gamma D_{t+3} + \gamma P_{t+3} | s_t]$, ..., and $E_t^a[P_{t+k-1} | s_t] < E_t^a[\gamma D_{t+k} + \gamma P_{t+k} | s_t]$, using $\bar{P}_t^a(s_t)$ as the willingness to pay will not terminate the option to obtain the return from $\hat{P}_t^a(s_t)$. Actually, using $\bar{P}_t^a(s_t)$ for $0 \leq i \leq k$ consecutively will result in the same level of wealth as $\hat{P}_t^a(s_t)$. Therefore, although both $\bar{P}_t^a(s_t)$ and $\hat{P}_t^a(s_t)$ give the same signal, $\hat{P}_t^a(s_t)$ is not the optimal strategy for the first $k$ periods, except in the above single case. Furthermore, even when $\hat{P}_t^a(s_t)$ is the
optimal choice, the return from the use of $\hat{P}^a_t(s_i)$ as the willingness to pay for the risky asset is not given up by the use of $\bar{P}^a_t(s_i)$. The same result can be exactly replicated by using $\bar{P}^a_t(s_i)$ consecutively.

ii) Suppose that $P_t(s_i) < \hat{P}^a_t(s_i) < \bar{P}^a_t(s_i)$. Although both $\bar{P}^a_t(s_i)$ and $\hat{P}^a_t(s_i)$ are greater than $P_t(s_i)$, buying the risky asset, holding it for one period, and then switching the risky asset to the risk-free asset is expected to yield higher returns than buying the risky asset and holding it for $k$ periods. The reason for this is that

$$\hat{P}^a_t(s_i) = E_t^a \left[ \sum_{i=1}^{k} \gamma^i D_{t+i} + \gamma^k P_{t+k} \mid s_i \right] < \bar{P}^a_t = E_t^a [\gamma D_{t+1} + \gamma P_{t+1} \mid s_i]$$
 implies that

$$E_t^a \left[ \sum_{i=2}^{k} \gamma^{i-1} D_{t+i} + \gamma^{k-1} P_{t+k} \mid s_i \right] < E_t^a [P_{t+1} \mid s_i]$$,

which states that reselling the risky asset and buying the risk-free asset at time $t+1$ is expected to result in greater returns than continuing to hold the risky asset for the remaining $k-1$ periods.

iii) Suppose that $\hat{P}^a_t(s_i) < \bar{P}^a_t(s_i) < P_t(s_i)$. Although the investor will buy the risk-free asset following the signal from either $\bar{P}^a_t(s_i)$ or $\hat{P}^a_t(s_i)$, buying the risk-free asset and holding it for $k$ periods is not the optimal strategy for the first $k$ periods unless

$$E_t^a [P_{t+1} \mid s_i] > E_t^a [\gamma D_{t+2} + \gamma P_{t+2} \mid s_i]$$,

$$E_t^a [P_{t+2} \mid s_i] > E_t^a [\gamma D_{t+3} + \gamma P_{t+3} \mid s_i]$$,

and

$$E_t^a [P_{t+k-1} \mid s_i] > E_t^a [\gamma D_{t+k} + \gamma P_{t+k} \mid s_i]$$.

Also, even when

$$E_t^a [P_{t+1} \mid s_i] > E_t^a [\gamma D_{t+2} + \gamma P_{t+2} \mid s_i]$$,

$$E_t^a [P_{t+2} \mid s_i] > E_t^a [\gamma D_{t+3} + \gamma P_{t+3} \mid s_i]$$,

and

$$E_t^a [P_{t+k-1} \mid s_i] > E_t^a [\gamma D_{t+k} + \gamma P_{t+k} \mid s_i]$$, the expected wealth from holding the risk-free asset for $k$ periods can be achieved by using $\bar{P}^a_t(s_i)$ consecutively during the periods between $t$ and $t+k$. 
iv) Suppose that \( \overline{P}_t^a(s_i) < \hat{P}_t^a(s_i) < P_t(s_i) \). The investor will buy the risk-free asset according to the signal from either \( \overline{P}_t^a(s_i) \) or \( \hat{P}_t^a(s_i) \). However, holding the risk-free asset for \( k \) periods is not optimal because switching the risk-free asset to the risky asset at time \( t+1 \) is more profitable. Since

\[
\hat{P}_t^a(s_i) = E_t^a \left[ \sum_{i=1}^{k} \gamma^i D_{t+i} + \gamma^k P_{t+k} \mid s_i \right] > \overline{P}_t^a = E_t^a \left[ \gamma D_{t+1} + \gamma P_{t+1} \mid s_i \right],
\]

which implies

\[
E_t^a \left( \sum_{i=2}^{k} \gamma^{i-1} D_{t+i} + \gamma^{k-1} P_{t+k} \mid s_i \right) > E_t^a \left[ P_{t+1} \mid s_i \right],
\]

buying the risky asset at time \( t+1 \) is expected to be more profitable than buying or holding the risk-free asset.

v) Suppose that \( \hat{P}_t^a(s_i) < P_t(s_i) < \overline{P}_t^a(s_i) \). Obviously, buying the risky asset and holding it for only one period at time \( t \) is more profitable than buying the risk-free asset at time \( t \), which is expected to be more profitable than buying the risky asset and holding it for \( k \) periods at time \( t \). Thus, the investor should buy the risky asset and hold it for one period.

vi) Suppose that \( \overline{P}_t^a(s_i) < P_t(s_i) < \hat{P}_t^a(s_i) \). Then, buying the risk-free asset at time \( t \), reselling the risk-free asset, and buying the risky asset financed by the resale at time \( t+1 \) will be more profitable than buying the risky asset and holding it for \( k \) periods. The reason for this is shown by the following:

\[
\hat{P}_t^a(s_i) = E_t^a \left[ \sum_{i=1}^{k} \gamma^i D_{t+i} + \gamma^k P_{t+k} \mid s_i \right] \\
= E_t^a \left[ \gamma D_{t+1} + \gamma P_{t+1} + \sum_{i=2}^{k} \gamma^{i-1} D_{t+i} + \gamma^{k-1} P_{t+k} \mid s_i \right] \\
< E_t^a \left[ P_t - \gamma P_{t+1} + \sum_{i=2}^{k} \gamma^{i-1} D_{t+i} + \gamma^{k-1} P_{t+k} \mid s_i \right]
\]
\[
E_t^a[P_{i+1} \left( \frac{(1+r)P_t}{P_{i+1}} - 1 \right) + \sum_{i=2}^{k} \gamma^i D_{t+i} + \gamma^k P_{t+k} | s_i]
\]

\[
< E_t^a[\gamma \left( \sum_{i=2}^{k} \gamma^{-i} D_{t+i} + \gamma^{-k} P_{t+k} \right) \left( \frac{(1+r)P_t}{P_{t+1}} - 1 \right) + \sum_{i=2}^{k} \gamma^i D_{t+i} + \gamma^k P_{t+k} | s_i]
\]

\[
= E_t^a[\gamma \left( \sum_{i=2}^{k} \gamma^{-i} D_{t+i} + \gamma^{-k} P_{t+k} \right) | s_i].
\]

The last inequality holds because \(E_t^a[\sum_{i=2}^{k} \gamma^{-i} D_{t+i} + \gamma^{-k} P_{t+k} | s_i] > E_t^a[P_{t+1} | s_i]\).

Furthermore, the last expression means that the expected return from buying the risk-free asset first at time \(t\), reselling it, and buying the risky asset financed by the resale of the risk-free asset at time \(t+1\) is greater than buying the risky asset and holding it for \(k\) periods. Therefore, the investor should not buy the risky asset at time \(t\). The investor must buy the risk-free asset at time \(t\), resell the risk-free asset, and buy the risky asset immediately before the risky asset price is expected to rise.

So far, this proof has shown that one can always find a more profitable alternative strategy or an equally profitable alternative strategy when \(\hat{P}_t^a(s_i)\) is used as the willingness to pay for the risky asset rather than \(\bar{P}_t^a(s_i)\). Therefore,

\[
E_t^a[\sum_{i=1}^{k} \gamma^i D_{t+i} + \gamma^k P_{t+k} | s_i]
\]

when \(k > 1\) cannot be the willingness to pay for the risky asset when the investor maximizes her expected wealth.

Proof of Proposition 4. Define the terminal period \((T)\) as the period when no further future resales are allowed. Suppose that investors are at the terminal period \(T\). Since each
row in $M_1$ first-order stochastically dominates the corresponding row in $M_0$, investors’ willingness to pay at $T$ decreases uniformly with $a_i$ in each state. Define $a_i^*$ as the $i$-th element in an investor’s identity “a” when the investor’s willingness to pay in state $i$ is equal to the equilibrium price. Note that $a_1^* = a_2^* = \cdots = a_N^*$ because $a_i$ is identically distributed in each state and because investors’ willingness to pay at $T$ decreases uniformly with $a_i$. Define $A^*$ as an $N \times N$ matrix such as $A_{ij}^* = F_i^{-1}(\alpha)$ if $i = j$, and $A_{ij}^* = 0$ if $i \neq j$, where $F_i(a_i)$ is the distribution function of $a_i$. Therefore, the equilibrium price at $T$ will be $P_T^* = [P_T^*(s_1), P_T^*(s_2), \ldots, P_T^*(s_N)] = \gamma M^* D$, where $D = [D(s_1), D(s_2), \ldots, D(s_N)]^\top$ and $M^* = A^* M_0 + (I - A^*) M_1$. Since each row in $M_0$ and $M_1$ first-order stochastically dominates rows below it, the first row of $M^*$ first-order stochastically dominates the second row, the second row first-order stochastically dominates the third row, and so on. As a result, $P_T^*(s_1) > P_T^*(s_2) > \cdots > P_T^*(s_N)$.

At $T-1$, the willingness to pay of any investor $a$ can be written as $P_{T-1}^a = \gamma M^* (D + P_T)$. Since $D(s_1) > D(s_2) > \cdots > D(s_N)$, and since $P_T^*(s_1) > P_T^*(s_2) > \cdots > P_T^*(s_N)$, the order of willingness to pay across investors in any state at $T-1$ will be the same as that in the corresponding state at $T$. Therefore, investors’ willingness to pay at $T-1$ will also decrease monotonically with $a_i$. In addition, $A^*$ and $M^*$ at $T-1$ will be the same as those at $T$. As a result, the equilibrium price at $T-1$ can be written as $P_{T-1}^* = \gamma M^* D + \gamma^2 (M^*)^2 D$.

This logic can be applied to any arbitrary period from $T$. Willingness to pay will decrease with $a_i$, and $A^*$ and $M^*$ will be the same every period. Hence, the equilibrium price for
any arbitrary period from $T$ can be written as $P^*_t = \sum_{i=1}^{T+t}[(\gamma M^*)^i D]$. The negative relation between willingness to pay and identity will also be true even when $T \to \infty$. Since $M^*$ is unique, the equilibrium price when $T \to \infty$ will be stationary and can be written as

$$P^*_t = \sum_{i=1}^{\infty} (\gamma M^*)^i D = (I - \gamma M^*)^{-1} \gamma M^* D.$$  

**Proof of Proposition 5.** Define $M^*$ and $\overline{M}^*$ as the transition matrices of the marginal investor when $x = 0$ and $x > 0$, respectively. Also, define $P^*(s_i)$ and $\overline{P}^*(s_i)$ as the equilibrium price of the risky asset in state $i=1, 2$ when $x = 0$ and $x > 0$, respectively. Since the eigenvalues of $M^*$ are 1 and $c_{11} - c_{21}$, and since the corresponding eigenvectors are $[0.5 \ 0.5]'$ and $\left[\frac{c_{11} - 1}{c_{21}} \ 1\right]'$, respectively,

$$(M^*)^k D = m_1 \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + m_2 \begin{bmatrix} (c_{11} - 1)/c_{21} \\ 1 \end{bmatrix} (c_{11} - c_{21})^k,$$

where $m_1 = \frac{2c_{21}}{c_{21} - c_{11} + 1} \left(D(s_1) + \frac{1 - c_{11}}{c_{21}} D(s_2)\right)$ and $m_2 = \frac{c_{21}}{c_{21} - c_{11} + 1} (D(s_2) - D(s_1)).$

Also, since the eigenvalues of $\overline{M}^*$ are 1 and $c_{11} - c_{21}$, and since the corresponding eigenvectors are $[0.5 \ 0.5]'$ and $\left[\frac{c_{11} + \delta - 1}{c_{21} + \delta} \ 1\right]'$, respectively,

$$(\overline{M}^*)^k D = \overline{m}_1 \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + \overline{m}_2 \begin{bmatrix} (c_{11} + \delta - 1)/(c_{21} + \delta) \\ 1 \end{bmatrix} (c_{11} - c_{21})^k,$$
\[ \delta = (1 - 2a^*)x \], \quad \bar{m}_i = \frac{2(c_{21} + \delta)}{c_{21} - c_{11} + 1} \left( D(s_i) + \frac{1 - c_{11} - \delta}{c_{21} + \delta} D(s_2) \right) \quad \text{and} \quad \bar{m}_2 = \frac{c_{21} + \delta}{c_{21} - c_{11} + 1} \left( D(s_2) - D(s_1) \right).

Hence,
\[
((\bar{M}^*)^k - (M^*)^k)D = \left( \frac{(D(s_1) - D(s_2))\delta}{c_{21} - c_{11} + 1} + \frac{(D(s_2) - D(s_1))\delta}{c_{21} - c_{11} + 1} (c_{11} - c_{21})^k \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
\]

Since
\[
\begin{bmatrix} P^*(s_1) \\ P^*(s_2) \end{bmatrix} = \gamma M^*D + (\gamma M^*)^2 D + \cdots, \quad \text{and since} \quad \begin{bmatrix} \bar{P}^*(s_1) \\ \bar{P}^*(s_2) \end{bmatrix} = \gamma \bar{M}^*D + (\gamma \bar{M}^*)^2 D + \cdots
\]

\[
\begin{bmatrix} \bar{P}^*(s_1) \\ \bar{P}^*(s_2) \end{bmatrix} - \begin{bmatrix} P^*(s_1) \\ P^*(s_2) \end{bmatrix} = \frac{\gamma(D(s_1) - D(s_2))(1 - 2a^*)x}{(1 - \gamma)(1 - \gamma(c_{11} - c_{21}))} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
\]

Therefore,
\[
\frac{\partial P(s_i)}{\partial x} = \frac{\gamma(D(s_1) - D(s_2))(1 - 2a^*)}{(1 - \gamma)(1 - \gamma(c_{11} - c_{21}))} \quad \text{for } i = 1, 2.
\]
References


Scheinkman, J. and Xiong, W. (2003). Overconfidence, Short-Sale Constraints, and
Notes

1 The definition of the marginal investor is the investor whose willingness to pay is equal to the equilibrium price in all states.
2 The limited short-sale assumption is necessary since investors are assumed to be risk-neutral.
3 Although investors’ learning is not allowed, the results in this paper will hold when the Bayesian learning process is allowed, as shown in Morris (1996).
4 Even with adjustments to market valuations, investors’ valuations will be different due to different prior transition matrices.
5 Expectations can affect investors’ wealth accumulation and asset prices depend on the distribution of wealth in practice. However, consideration of these connections will make the problem intractable. As a result, previous studies, such as Harrison and Kreps (1978), Morris (1996), and Scheinkman and Xiong (2003), assume an infinite amount of wealth to avoid any effect of heterogeneous expectations on wealth accumulation.
6 Although the marginal investor always exists conceptually, no real investor could be the marginal investor. This can especially happen when investors’ identity numbers come from the distribution of a discrete random variable.