

NBER WORKING PAPER SERIES

THE DISPOSITION EFFECT AND MOMENTUM

Mark Grinblatt  
Bing Han

Working Paper 8734  
<http://www.nber.org/papers/w8734>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
January 2002

The authors thank Andrew Lo and Jiang Wang for providing the MiniCRSP database used in the paper's empirical work, the UCLA Academic Senate for financial support, David Hirshleifer, Francis Longstaff, Monika Piazzesi, Shlomo Benartzi, Mark Garmaise, and Jun Liu for invaluable discussions, and seminar participants at Emory for comments on earlier drafts. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2002 by Mark Grinblatt and Bing Han. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Disposition Effect and Momentum  
Mark Grinblatt and Bing Han  
NBER Working Paper No. 8734  
January 2002  
JEL No. G0, G1

### ABSTRACT

Prior experimental and empirical research documents that many investors have a lower propensity to sell those stocks on which they have a capital loss. This behavioral phenomenon, known as “the disposition effect,” has implications for equilibrium prices. We investigate the temporal pattern of stock prices in an equilibrium that aggregates the demand functions of both rational and disposition investors. The disposition effect creates a spread between a stock’s fundamental value -- the stock price that would exist in the absence of a disposition effect -- and its market price. Even when a stock’s fundamental value follows a random walk, and thus is unpredictable, its equilibrium price will tend to underreact to information. Spread convergence, arising from the random evolution of fundamental values, generates predictable equilibrium prices. This convergence implies that stocks with large past price runups and stocks on which most investors experienced capital gains have higher expected returns than those that have experienced large declines and capital losses. The profitability of a momentum strategy, which makes use of this spread, depends on the path of past stock prices. Crosssectional empirical tests of the model find that stocks with large aggregate unrealized capital gains tend to have higher expected returns than stocks with large aggregate unrealized capital losses and that this capital gains “overhang” appears to be the key variable that generates the profitability of a momentum strategy. When this capital gains variable is used as a regressor along with past returns and volume to predict future returns, the momentum effect disappears.

Mark Grinblatt  
UCLA  
Anderson Graduate School of Management  
110 Westwood Plaza  
Los Angeles, CA 90095-1481  
and NBER

Bing Han  
UCLA  
Anderson Graduate School of Management  
110 Westwood Plaza  
Los Angeles, CA 90095-1481  
Tel: 310-825-8160  
Email: [bhan@anderson.ucla.edu](mailto:bhan@anderson.ucla.edu)

Momentum, which is the persistence in the returns of stocks over horizons between three months and one year, remains one of the most puzzling anomalies in finance. Jegadeesh and Titman (1993), for example, found that past winning stocks, as measured by returns over the prior six months tended to subsequently outperform past losing stocks by about twelve percent per year. Various explanations for momentum have been advanced, but few have stood up to rigorous empirical tests or calibration. Moreover, most of these explanations are more rooted in attempts to explain momentum than in developing a general theory of how demand for stocks affects equilibrium pricing.

While momentum-motivated theories have been categorized as belonging to the field of behavioral finance, there is a seemingly unrelated branch of the literature in behavioral finance that documents stylized facts about investor behavior. It is motivated by experiments on individual investors and observations of their behavior in real financial markets. It rarely attempts to link its stylized facts about investor behavior to equilibrium prices. Perhaps the most well-documented behavioral regularity in this literature is what Shefrin and Statman (1985) termed “the disposition effect.” This is the tendency of investors to hold onto their losing stocks to a greater extent than they hold onto their winners.<sup>1</sup>

This paper analyzes how aggregate demand and equilibrium prices evolve over time when they are affected by the existence of a fixed proportion of investors who exhibit the disposition effect. It is possible to analytically prove that if some investors are subject to the disposition effect, then stocks with aggregate unrealized capital gains tend to outperform stocks with aggregate unrealized capital losses. This paper shows, both theoretically and empirically, that the disposition effect may account for the tendency of past winning stocks to subsequently outperform past losing stocks.

The intuition for our model is rather simple. Assume for the moment that some investors, the *disposition* investors, perturb otherwise rational demand functions for a stock they own because they have experienced unrealized capital gains or losses in the stock. Such investors would tend to have higher demand for losing stocks than for winning stocks, other things equal. If demand for that same stock by other investors is not perfectly elastic, then such demand perturbation, induced by a disposition effect, tends to generate price underreaction to public information. Stocks that have been privy to good news in the past would have excess

---

<sup>1</sup>The disposition effect is sometimes linked to, but is really distinct from *loss aversion*. Loss aversion occurs when two conditions are met: (i) the decline in utility for a loss (measured relative to a reference point) exceeds the increase in utility for an equal sized gain (relative to the same reference point); (ii) a gamble that is always a loss relative to the reference point generates higher utility than a certain loss with the same mean while the reverse preference applies to gambles that are always gains. In a market with a single risky asset, one can see how an investor might be more likely to substitute a risk-free asset for a risky asset in a gain situation (where utility is locally concave) than a loss situation (where utility is locally convex). However, in a multiasset multiperiod framework, it is necessary to argue that reference prices change to induce substitution of one risky asset with a paper gain or loss for another. This complex setting is atypical in the modeling of loss aversion. Moreover, once we allow for such reference price changes, there is little need for the remaining utility assumptions of loss aversion.

selling pressure relative to stocks that have been privy to adverse information. This produces a spread between the fundamental value of the stock – its equilibrium price in the absence of any disposition effect among investors – and the market price of the stock. Aggregate investor demand equals supply in the model when the stock price is a weighted average of its fundamental value and a reference price (related to the basis at which disposition investors acquired the stock).

The demand perturbation and associated spread does not, per se, generate momentum in stock returns. For momentum to arise, there must be some mechanism that forces market prices to revert to fundamental values. As one example of the importance of this reversion, note that if all investors are identical and subject to the disposition effect, market prices deviate from fundamental values. However, in the absence of trading, it would be reasonable to conclude that the reference price (such as the cost basis) of disposition investors would not change. In this case, the associated demand perturbation generates a market price that is a constant weighted average of the stock's fundamental value and a fixed reference price. Because the weights are constant, changes in market prices are just dampened versions of changes in fundamental values. Hence, if fundamental values follow a random walk, so do market prices, despite their underreaction to information about fundamentals.

One could imagine various forces that drive market prices to their fundamental values, thus generating momentum. A press release about an acquisition, an earnings announcement, and a reduction in uncertainty all are events that may make some investors rely less on their behavioral tendency to sell winners and hold onto losers. However, it is surprising to discover that a random process for a stock's fundamental value, in and of itself, makes a stock's reference price, and hence its market price, catch up with its associated fundamental value. In other words, trading, which arises in the model only because some investors exhibit the disposition effect, produces mean reversion in the *spread* between a stock's fundamental value and its market price. By definition, the stochastic process for a stock's market price is the stochastic processes for its fundamental value less that for its spread; hence, any mean reversion in the stochastic process for the spread implies that stocks with positive spreads have risk-adjusted mean returns that exceed the risk-free return, while those with negative spreads have risk-adjusted mean returns that are below the risk-free return. The mean return of a stock that is a big winner, which tends to have a positive spread between its fundamental value and market price, thus tends to exceed that of a loser, which tends to have a negative spread.

The difference between the market price and the reference price is proportional to the spread between the fundamental value and the market price. Thus, it, too, is of the same sign as the expected return. If the reference price is related to some aggregate cost basis in the market, as the motivation for this paper suggests, this difference is a gain measured relative to a reference price that is the cost basis.

The empirical implications of the model, outlined above, are verified with cross-sectional

“Fama-MacBeth” regressions that make use of a “capital gains overhang” regressor to proxy for aggregate unrealized capital gains or losses. In all of our regression specifications, the gain variable predicts future returns, even after controlling for the effect of past returns, but the reverse is rarely true. Indeed, in most of our regression specifications, there is no momentum effect once the disposition effect is controlled for with the gain regressor.

Section I of the paper presents results derived from a model with two types of investors: One type has no disposition effect and is fully rational; the other is identical to the first type except that the disposition effect perturbs his demand function for stocks. These include results about the temporal pattern of equilibrium prices conditional on both aggregate capital gains and past returns. It also includes results about the determinants of trades and volume. Section II presents empirical data and provides numerous tests illustrating that our findings are not due to omitted variables that others have used in the literature to analyze momentum. Our main finding here is that the capital gains overhang is a critical variable in any study of the relation between past returns and future returns, as the theory would predict. Section III discusses the relation of our work to prior literature, both theoretical and empirical. Section IV concludes the paper.

## I. The Model

In an attempt to make the model as simple and analytically tractable as possible, we focus on how the partial demand function of a single risky stock among possibly many assets has its equilibrium price path affected by the disposition effect

- The risky stock that we focus on is in fixed supply normalized to one unit. Public news about the fundamental value of the stock arrives at discrete dates  $t = 0, 1, 2, \dots$  just prior to trading on those dates
- There are two types of investors:
  - The fraction  $1 - \mu$  are type- $r$  “rational” investors
  - The fraction  $\mu$  are type- $d$  “disposition” investors
- Date  $t$  demand functions per unit of each investor-type’s mass in the market are respectively given by

$$D_t^r = 1 + b_t(F_t - P_t) \tag{1}$$

$$D_t^d = 1 + b_t[(F_t - P_t) + \lambda(R_t - P_t)] \tag{2}$$

where

- $b_t$  is a positive parameter that represents the slope of the rational component of the demand functions for the stock.<sup>2</sup> This parameter is actually a function that can be contingent on just about anything in the date  $t$  information set of investors, including  $P_t$ ,  $F_t$ ,  $R_t$ ,  $\lambda$ ,  $\mu$ , historical values of these variables, or corresponding information for other assets.<sup>3</sup>
- $\lambda$  is a positive constant parameter measuring the relative importance of the disposition component of demand for type- $d$  investors
- $P_t$  is the price of the stock
- $R_t$  is a reference price representing a weighted average of the past prices at which type- $d$  investors executed trades, and
- $F_t$  is the stock's rational equilibrium price, or fundamental value<sup>4</sup>

For  $\lambda > 0$ , one of the two investor-types is relatively more averse to realizing losses. When this investor-type has a paper capital loss, he holds more shares than his rational counterpart. It would be wrong to model him as focused only on stocks that have declined in value since their purchase. Disposition investors are not characterized by a propensity to accumulate more stock than rational behavior would suggest. Rather, the experimental and empirical evidence only indicates that their propensity to sell stocks they have lost money on is less than their propensity to sell stocks they have made money on. Hence, the other side of the coin of loss aversion is a relatively greater propensity to sell stocks experiencing a gain. As our demand function suggests, when good news about the fundamental of the stock arrives after the loss averse investor has bought it, he experiences a paper gain, and sells more of the stock than he would if he were fully rational and the stock was trading at the same price.

The experimental and empirical evidence focuses on active buying and selling behavior. Modeling equilibrium requires that buying and selling behavior be endogenously derived from demand functions. For an investor to have a greater propensity to sell a stock with a paper gain, the excess demand of that investor must be lower for such a stock than it

---

<sup>2</sup>To isolate the impact of the disposition effect, and avoid possible confounding effects arising, for example, from differences in risk aversion across investor-types, we assume that all investor-types have the same slope to the rational component of their demand functions,  $b_t$ .

<sup>3</sup>We can think of  $b_t$  as being whatever solves for the optimal demand function given a utility function.

<sup>4</sup>For our results, it is not necessary to be precise about how one arrives at this fundamental value. We would merely like the fundamental value to converge to the equilibrium price as the number of disposition agents converges to zero. Two specific alternatives come to mind. In the first, the fundamental value at date  $t$  is the market price that would prevail at date  $t$  if all agents were fully rational but behaved under the assumption that, in the future there would be disposition agents and equilibria as specified in the model. In the second, the date  $t$  fundamental value is the price that would prevail if all agents were fully rational and assumes that, in the future, all agents continue to be fully rational. In the appendix, we employ the former alternative to analytically compute fundamental values and market prices for a multiperiod exponential model. This interpretation of  $F$  makes the fundamental value a function of  $\lambda$ .

would be in the absence of disposition behavior (as described above). For an investor to have a lower propensity to sell a stock with a paper loss, the excess demand of that investor must be higher for that stock. One can quibble about what the experimental and empirical evidence on buying and selling behavior has on the functional form of the demand function. We have no interest in debating this issue, as there is little evidence about what the true demand functions of such investors look like.<sup>5</sup> In the interest of parsimony and tractability, equation (2) simply assumes a functional form for the deviation from rational behavior where the deviation parameter  $\lambda$  is constant.<sup>6</sup>

Note that if  $\lambda$  is zero, (or, alternatively, if  $\mu = 0$ ), so that all investors are “rational,” aggregate demand, a weighted average of these two demand functions is

$$(1 - \mu)D_t^r + \mu D_t^d = 1 + b_t(F_t - P_t)$$

and thus, with  $b_t \neq 0$ , supply equals demand when  $F_t = P_t$ . In this sense, the fundamental value  $F$  is simply the price that would prevail in the absence of a disposition effect. The fundamental value is thus the present value of the free cash flow stream of the stock conditional on all information currently available, adjusted for the risk premium.<sup>7</sup> One of the main goals of this paper is to assess how the disposition effect alters the stochastic process for equilibrium prices. As a benchmark, and consistent with an extensive literature in finance, we assume that the fundamental value follows a random walk:

$$F_{t+1} = F_t + \epsilon_{t+1} \tag{3}$$

where the  $\epsilon$ s are i.i.d. and mean zero.<sup>8</sup>

---

<sup>5</sup>Grinblatt and Keloharju (2001), however, estimate that large positive (negative) price deviations from the cost basis generate more (fewer) sales.

<sup>6</sup>In the interest of lucidity, we see no point to artificially complicating the model with utility functions. The solution to rational investor demand may affect the fundamental value; beyond this, however, it is not present in the equation that determines the equilibrium price.

<sup>7</sup>The precise structure of dividends, free cash flows, or information is not important for our purpose. For example, we can assume the stock pays a dividend each period or only one liquidating dividend at some far away future date.

<sup>8</sup>As we demonstrate in a subsequent footnote, the absence of a random walk for fundamental values does not alter any of our results if we interpret all expectations in the paper as risk-neutral expectations. Alternatively, if a random walk does not apply to fundamental values, our results are simply measuring the incremental stochastic processes arising from disposition agents. That is, we are providing closed form solutions for how the stochastic process for equilibrium prices deviates from the stochastic process that would prevail in the absence of disposition agents. However, our desire to ignore the confounding effects of changing risk premia arising from in an equilibrium generated by an intertemporal representative agent model has precedent: We are unaware of *any* nontrivial models of this class where the equilibrium stochastic process for prices has been derived analytically. A trivial model that achieves this rare feat is provided in the appendix.

## A. Equilibrium

When  $\lambda = 0$ , both investor types are rational and hold  $1 - \mu$  and  $\mu$  shares, respectively. Thus, there is no trading. However, when  $\lambda$  is nonzero, the demand of the type- $d$  investors is also affected by the unrealized capital gain as measured relative to their reference price. This provides the key motive for trade between the two types of investors. Thus, at each date  $t$ , aggregate demand equals aggregate supply when

$$1 + b_t(F_t - P_t) + \mu\lambda b_t(R_t - P_t) = 1 \quad (4)$$

and date  $t$  volume, the number of shares changing hands between the two investor-types, is

$$V_t = \mu |D_t^d - D_{t-1}^d|$$

computed at the equilibrium prices for dates  $t$  and  $t - 1$ . Volume and turnover ratio are identical here since there is one share outstanding.

The market clearing condition, equation (4), is equivalent to

$$P_t = wF_t + (1 - w)R_t, \quad \text{where } w = \frac{1}{1 + \mu\lambda} \quad (5)$$

demonstrating that the equilibrium market price is a weighted average of the fundamental value and the reference price. Since  $0 < w < 1$ , the market price underreacts to public information about the fundamental value, holding the reference price constant. (Market prices also underreact for reference prices that follow equilibrium paths, as we will see shortly.) The degree of underreaction, measured by  $w$ , depends on the proportion of disposition investors,  $\mu$ , and the relative intensity of the demand perturbation induced by the disposition effect,  $\lambda$ . The fewer the number of disposition investors, and the smaller the degree to which each perturbs demand, the closer the market price will be to its fundamental value.

To illustrate how changes in the fundamental value of the stock affect equilibrium prices, consider a case where type- $d$  investors' entire holding of shares was purchased at a fundamental, market, and reference price of \$100 per share last period. Because the three prices were equal, type- $d$  holdings equal  $\mu$  shares. Suddenly, bad news arrives and the fundamental value drops to \$80. The equilibrium price must end up between the fundamental value of \$80 and the reference price of \$100. At a price at or below \$80, the stock is too attractive as the demand functions of both types of investors suggest that they would want to hold more than they currently own. Equilibrium prices cannot be those at which aggregate demand for the stock exceeds one share. Similarly, the equilibrium price cannot be at or above \$100 either, since aggregate demand would be below one share and both investor-types would like to sell stock at such prices.

If the fundamental value of the stock is \$80, type- $r$  investors want to sell some of the stock they own at prices above \$80. At some price between \$80 and \$100, these type- $r$

investors exactly accommodate the extra buying pressure from the type- $d$  investors. The market will be in equilibrium at this price, a weighted average of the fundamental value and the reference price.<sup>9</sup> Since the equilibrium price must lie between \$80 and \$100 per share, the downward price response to the news is sluggish, settling at a weighted average of the past price at which shares were purchased by disposition investors and the fundamental value. (A similar argument applies for good news that generates a fundamental value above \$100.)

In the illustration above, we had some trades at \$100 per share (by assumption) at what we will refer to as date 0 and then another set of trades occurring at some price between \$80 and \$100 at date 1. Knowing the parameters of the model would have allowed us to solve for the date 1 equilibrium price as the single unknown of a linear equation. Assume for the moment that this price turned out to be \$85. To solve for the date 2 equilibrium price as a function of the date 2 fundamental value, it is critical that we know the date 2 reference price. It seems reasonable to think that this reference price is going to be some weighted average of \$100 and \$85. If we know this reference price, solving for the equilibrium price as a function of the date 2 fundamental value is again trivial.

## B. Reference Price Dynamics

The reference price is established when an investor first enters into a position and it is updated as the investor trades. Our specification for its dynamics here, which is consistent with the existing experimental and empirical evidence, is that the type- $d$  investors' reference price satisfies the difference equation

$$R_{t+1} = \nu_t P_t + (1 - \nu_t) R_t \quad (6)$$

where  $\nu_t$ , a function of the type- $d$  investors' date  $t$  information, lies between 0 and 1.

This specification implies that the reference price can be any weighted average of past prices<sup>10</sup>

$$R_{t+1} = \sum_{\tau \geq 0} \omega_{t-\tau} P_{t-\tau}, \quad \sum_{\tau} \omega_{\tau} = 1, \quad \omega_{\tau} > 0, \quad (7)$$

For economic content, such as ties to a cost basis, it is useful to think of  $\nu_t$  as linked to the trading volume of type- $d$  investors. However, our theoretical results are sufficiently general as to not require even this restriction. For the theoretical findings in this section, the interval restriction for  $\nu_t$  should hold in any reasonable specification of the reference price dynamics. For example, the reference price dynamics specified in equation (6) is not only consistent with the type- $d$  investors' cost basis as the reference price; it is also consistent with the

<sup>9</sup>Clearly, although we do not make use of utility functions, the fact that demand functions are not perfectly elastic reflects some risk aversion, capital constraint, or other force restraining unlimited trade by investors.

<sup>10</sup>Iteratively expanding equation (6) implies that the weight on date  $t - \tau$ 's price,  $\omega_{t-\tau} = (1 - \nu_t)(1 - \nu_{t-1}) \cdots (1 - \nu_{t-\tau+1})\nu_{t-\tau}$ . On the other hand, equation (7) is consistent with  $\nu_{t-\tau} = \frac{\omega_{t-\tau}}{1 - \omega_t - \omega_{t-1} - \cdots - \omega_{t-\tau+1}}$ .

reference price being a weighted average of past prices with weights proportional to trade size, to some function of time proximity of trades in the past, or to any combination of the two.

Empirical analyses, presented later in the paper, requires a more precise specification of  $\nu_t$ . As one example, we might assume that each share is as likely to trade as any other and that the probability of a trade in a share on any one day is independent of when it has traded in the past. In this case, a proxy for the aggregate cost basis of the outstanding shares is obtained by setting  $\nu_t$  as date  $t$  turnover,  $V_t$ . The associated expression for  $\omega_{t-\tau}$ , the weight on date  $t - \tau$ 's price, is the product of the probabilities that a share did not trade between dates  $t - \tau + 1$  and  $t$ ,  $(1 - V_{t-\tau+1}) \cdots (1 - V_{t-1})$  times the probability that it traded on date  $t - \tau$ ,  $V_{t-\tau}$ . The resulting reference price is the expected cost basis of an outstanding share under the assumptions given above.

Earlier, we argued that market prices,  $P_t$ , respond sluggishly to changes in the fundamental value, *ceteris paribus*. Because  $\nu_t$  lies between 0 and 1, the reference price also responds sluggishly to changes in the fundamental value. We can see this by substituting equation (5) into (6), obtaining

$$R_{t+1} = w\nu_t F_t + (1 - w\nu_t)R_t \quad (8)$$

This equation also points out that the reference price is always reverting to the fundamental value.

### C. Can the Equilibrium Degenerate?

The previous subsections argued that we can solve for an equilibrium in closed form by assuming fully rational behavior on the part of one class of investors. Our closed form solutions arise from modeling the type- $d$  investor behavior as a specific perturbation of this fully rational behavior. Modeling disposition behavior in this form allows us to calculate closed form equilibrium prices as functions of fully rational prices, the fundamental values. This appears to be almost too good to be true as the fundamental values cannot be solved for directly, except in very special cases.<sup>11</sup> Moreover, fully rational behavior recognizes that disposition investors exists. Rest assured that with virtually any reasonable intertemporal utility function, this does not generate an infinite loop that fails to generate an equilibrium. The type- $r$  investors, unless they are risk neutral, do not fully undo the impact of the type- $d$  investors perturbation. As a consequence, the equilibrium does not collapse. To illustrate this point – that the type- $r$  investors can solve for optimal demand recognizing the impact of type- $r$  investors who partly mimic them, and that a Walrasian auctioneer can solve for prices that arise from the specified demand structure – we derive closed form solutions for demand and prices in one very special case: In this case, we solve for a date 0

---

<sup>11</sup>As noted earlier, closed form solutions for equilibria in intertemporal models have been found only in trivial cases.

equilibrium involving type- $r$  agents who maximize the expected negative exponential utility of date 2 wealth, and trade with type- $d$  agents at dates 0 and 1. We also assume that the fundamental value at date 2 is normally distributed, and that the information about the date 2 value that arrives at date 1 also is normally distributed. The closed form solutions for the date 0 and date 1 demand functions of type- $r$  investors, along with the date 0 and date 1 equilibrium price and fundamental value functions, are found in the appendix. This should be sufficiently convincing that there is nothing aberrational going on within our model, despite an underlying complexity that has been deliberately masked.

#### D. Expected Price Changes

Although the fundamental value, by assumption, follows a random walk, changes in equilibrium prices are predictable in this model. To see this, note that by equation (5), the price change can be expressed as:

$$P_{t+1} - P_t = w(F_{t+1} - F_t) + (1 - w)(R_{t+1} - R_t) \quad (9)$$

and thus the expected price change is

$$\mathbf{E}_t[P_{t+1} - P_t] = (1 - w)(R_{t+1} - R_t) \quad (10)$$

Subtracting equation (10) from (9) and substituting in (3) implies

$$P_{t+1} - P_t - \mathbf{E}_t[P_{t+1} - P_t] = w\epsilon_{t+1}$$

proving that price changes underreact to news about fundamentals. Although this may contradict the suggestion in Shiller (1981) that prices are too volatile relative to fundamentals, the lower price volatility feature is a necessary feature of any model of price underreaction.

Equation (10) implies that changes in equilibrium prices are predictable. Reference prices for the date  $t + 1$  equilibrium are known at date  $t$ . Moreover, from equation (8),

$$R_{t+1} - R_t = w\nu_t(F_t - R_t) \quad (11)$$

which, substituted into equation (10), implies

$$\mathbf{E}_t[P_{t+1} - P_t] = w(1 - w)\nu_t(F_t - R_t) \quad (12)$$

Alternatively, this equation is equivalent to

$$\mathbf{E}_t[P_{t+1} - P_t] = w\nu_t(F_t - P_t) \quad (13)$$

and to

$$\mathbf{E}_t[P_{t+1} - P_t] = (1 - w)\nu_t(P_t - R_t) \quad (14)$$

by the equilibrium pricing condition, equation (5). These findings are summarized below:

**Proposition 1** *If the date  $t$  fundamental value exceeds date  $t$ 's market price (or reference price), or if the date  $t$  market price exceeds date  $t$ 's reference price, then the stock price is expected to increase next period. Similarly, if the date  $t$  fundamental value is exceeded by date  $t$ 's market price (or reference price), or if the date  $t$  market price is exceeded by date  $t$ 's reference price, then the stock price is expected to decrease next period. The expected price change is proportional to the change in the reference price. Moreover, the expected return is increasing in the difference between the fundamental value and the market (or the reference) price, or in the difference between the market and reference prices, and in the weight placed on the current market price in updating the reference price.*

Hence the date  $t$  unrealized gain  $g_t$ , defined as the difference between the market and reference prices,

$$g_t = P_t - R_t \quad (15)$$

or, alternatively, the date  $t$  spread between the fundamental value and the equilibrium price

$$s_t = F_t - P_t = \frac{1-w}{w}g_t \quad (16)$$

or the date  $t$  gap between the fundamental value and the reference price determine (along with the weights  $w$  and  $\nu_t$ ) the expected price change from  $t$  to  $t+1$ . All of these are driven by the innovation in the fundamental value, which is the model's only source of uncertainty.

The sign of the date  $t$  spread between the fundamental value and the equilibrium price is the same as the sign of the expected future price change,  $E_t[P_{t+1} - P_t]$ , because the former spread is mean reverting and by definition,

$$P_{t+1} - P_t = (F_t - P_t) - (F_{t+1} - P_{t+1}) + (F_{t+1} - F_t)$$

which has an expectation of

$$E_t[P_{t+1} - P_t] = (F_t - P_t) - E_t[F_{t+1} - P_{t+1}] \quad (17)$$

since the fundamental value follows a random walk. However, with mean reversion in the spread, the expectation on the right side of equation (17) is smaller in absolute terms than the term in parentheses it is subtracted from. Hence, if the spread,  $F_t - P_t$ , is positive, the difference between the spread and the expected spread is positive; if the spread is negative, the difference is negative. To understand why this mean reversion exists, we prove the following proposition.

**Proposition 2** *The gain  $g_t$ , which is the difference between the market and reference prices, mean reverts towards zero. The rate of mean reversion is greater the larger is the absolute magnitude of  $\nu_t$ , and the smaller is  $\lambda$  and  $\mu$ . The same result applies to the spread,  $F_t - P_t$ , and to the gap,  $F_t - R_t$ .*

*Proof:* By equations (3), (11), and (15),

$$g_{t+1} - g_t = w\epsilon_{t+1} - w\nu_t g_t \tag{18}$$

Hence,  $E_t[g_{t+1} - g_t] = -w\nu_t g_t$ . The mean reversion speed is not constant, but is increasing in  $\nu_t$  and  $w$ . The spread and gap are constant proportions of the gain so this result also applies to the spread or gap. ■

Proposition 2 states that the reference price is always trying to catch up to the latest fundamental value (and market price). As trading occurs, the reference price gets updated with some weighting of the latest market price, which, in turn, is updated by the latest fundamental value. This implies that the gain, the spread, and the gap tend to narrow over time. As the next result indicates, these will widen or change sign only as a consequence of extraordinary innovations in the fundamental value.

**Proposition 3** *If the date  $t$  gain,  $g_t$ , is positive (negative), it will continue to be positive (negative) unless a sufficiently large negative (positive) shock to the fundamental arrives. Assume that  $\nu_t$  is positive. Then, the absolute magnitude of the gain will increase next period only if a significant shock of the same sign as the gain arrives. Otherwise, the absolute magnitude of the gain will decrease next period. The same results apply to the spread and the gap.*

*Proof:* All statements follow from equation (18) which states that  $g_{t+1} = w\epsilon_{t+1} + (1 - w\nu_t)g_t$ . Assume  $g_t > 0$ . Since  $0 < \nu_t < 1$  and  $0 < w < 1$ ,  $g_{t+1} > 0$  unless  $\epsilon_{t+1}$  is sufficiently negative. Equation (18) also implies  $g_{t+1} - g_t < 0$  if  $\epsilon_{t+1} < 0$  or if  $\epsilon_{t+1}$  is positive but not too large. An analogous result applies to the negative gain. Since the spread  $s_t$  and gap are constant multiples of the gain  $g_t$ , the same conclusions necessarily apply to them. ■

The results above, applied to the unrealized gain  $g_t$ , describe expected future price changes that are conditional on the relation between the market price and the reference price. This suggests there is an interesting line of empirical work that explores the relationship between aggregate capital gains and the cross-section of expected returns. It also may explain why momentum strategies are profitable, as stocks with large positive differences between the market price and reference price tend to be winning stocks and vice versa. However, this difference is path dependent. There are historical price paths for a stock, when combined with reasonable specifications for the reference price updating parameter,  $\nu_t$ , that generate negative gains, and hence negative expected returns, even when past prices have increased. Similar, anomalous paths exist for losing stocks. Thus, while momentum in stock returns may be an artifact of the disposition effect because past returns are correlated with variables like aggregate capital gains, our model implies that for a given past return, some types of paths will generate higher expected returns than others. Lacking a specific functional form for  $\nu_t$ , it is difficult to quantify which paths have higher expected future

returns than others. However, it is fair to say that past returns are merely noisy proxies for behavioral variables, like capital gains, and are likely to be poorer predictors of expected returns than capital gains proxies if our model is an accurate portrayal of how demand for stock is generated.

There is one class of past return paths for which the gain and spread are necessarily of the same sign as the past return. Stocks that have reached a new high (or low) relative to a reasonably lengthy historical period are those for which past returns, irrespective of the past return horizon in the historical period are all positive (or negative). Such “consistent winning” (or “consistent losing”) stocks necessarily have investors who acquired the stock at a basis below the current price – thus experiencing a capital gain (or, in the case of the consistent losers, a capital loss). Given the reference price updating rule, and Proposition 1, the following result must hold:

**Proposition 4** *When the market price is at a new high (new low), the gain is positive (negative) and over the next period, the expected change in the market price is positive (negative).*

Although there are occasional historical paths for winning stocks that generate negative spreads and negative expected future returns, intuition tells us that it is unlikely, particularly if the past return is large, that these paths could dominate the abundant number of paths for which the spread is positive (and thus, for example, most investors experience capital gains). This intuition is indeed correct.

Before demonstrating this formally, it is useful to first express the reference price explicitly as a weighted average of the fundamental values at previous dates. This is done by applying equation (11) iteratively.

$$R_t = w\nu_{t-1}F_{t-1} + (1 - w\nu_{t-1})w\nu_{t-2}F_{t-2} + \cdots + (1 - w\nu_{t-1}) \cdots (1 - w\nu_{t-n-1})w\nu_{t-n}F_{t-n} + (1 - w\nu_{t-1}) \cdots (1 - w\nu_{t-n-1})(1 - w\nu_{t-n})R_{t-n} \quad (19)$$

This is a complex expression, with coefficients on the  $F$ s and  $R_{t-n}$  that may be path dependent. But notice that these nonnegative coefficients sum to one (which is not surprising given that the reference price is a weighted average of the current price and the prior period’s reference price and the current price is a weighted average of the current fundamental value and the current reference price). More recent prices have more influence on the current reference price. As we push further back into history (larger  $n$ ), we can see that what happens at a remote historic date matters little for the current reference price. In this sense, the market is slowly forgetful. The larger the  $\nu$ s or  $w$ , the faster the market forgets. For this reason, any insights from this model do not depend on initial conditions, provided that the security under study is sufficiently long-lived.

The low weight placed on the distant past is not the only justification for our assertion that initial conditions are unimportant. We can also justify initial conditions with no spread

as a good approximation because the fundamental value and the market price tend to revisit each other as time evolves. As the following proposition proves, in continuous time, this occurs with probability one.

**Proposition 5** *Assume that  $F$  follows a diffusion process in continuous time and trading occurs continuously. Given any date  $t$  spread,  $s_t$ , with probability 1, there is a date in the future when the market price equals the fundamental value (a spread and gain of zero).<sup>12</sup>*

*Proof:* Suppose the gain  $g_t$  is negative. Let  $F^*$  be the all time high of the fundamental value up until date  $t$ . A basic property of Brownian motion is with probability 1, it will eventually hit any number. Hence with probability 1, there exists a future date  $\tau$  when  $F_\tau = 2F^*$  for the first time. That means  $F$  has a new high at  $\tau$ , and hence by Proposition 4,  $g_\tau > 0$ . Since the reference price is a weighted average of historical  $F$ 's and  $F$  has a continuous sample path,  $g_t$  will also have continuous sample path. Hence, by the intermediate value theorem, sometime between  $t$  and  $\tau$  there will be a date when the gain is zero. The proof is the same when the gain is positive. ■

We now derive a closed form solution that quantifies momentum in stock returns. Using equation (19), the gain  $g_t = P_t - R_t$  can be written as:

$$g_t = w\epsilon_t + w(1 - w\nu_{t-1})\epsilon_{t-1} + \dots + w(1 - w\nu_{t-1}) \dots (1 - w\nu_{t-n+1})\epsilon_{t-n+1} + w(1 - w\nu_{t-1}) \dots (1 - w\nu_{t-n})(F_{t-n} - R_{t-n}) \quad (20)$$

This relation turns out to be useful in analyzing how past changes in the fundamental value affect future expected returns. The next proposition actually provides a closed form solution for this conditional expectation.

**Proposition 6** *Given a historical horizon of  $n$  periods*

$$\begin{aligned} & \mathbb{E}[P_{t+1} - P_t | F_t - F_{t-n} = x] \\ &= w(1-w) \frac{x}{n} \mathbb{E}[\nu_t(1 + (1 - w\nu_{t-1}) + \dots + (1 - w\nu_{t-1}) \dots (1 - w\nu_{t-n+1}))] \\ & \quad + w(1-w) \mathbb{E}[\nu_t(1 - w\nu_{t-1}) \dots (1 - w\nu_{t-n})(F_{t-n} - R_{t-n})] \end{aligned} \quad (21)$$

*This conditional expectation is increasing in  $x$ , ceteris paribus, and for positive (negative)  $x$ , is positive (negative) if either the absolute magnitude of  $x$  or  $n$  is sufficiently large, or is positive (negative) for any positive (negative)  $x$  provided that  $F_{t-n} - R_{t-n} \geq 0$  ( $\leq 0$ ) or negative (positive) but of sufficiently small absolute magnitude.*

*Proof:* By the law of iterated expectations and equation (12),

$$\mathbb{E}[P_{t+1} - P_t | F_t - F_{t-n} = x] = (1-w) \mathbb{E}[\nu_t g_t | F_t - F_{t-n} = x]$$

---

<sup>12</sup>An analogous result holds as an approximation in discrete time with high frequency trading.

Since the innovation in  $F$  is i.i.d.,

$$\mathbf{E}[F_t - F_{t-i} | F_t - F_{t-n} = x] = \frac{i}{n} x$$

Using this equation, the law of iterated expectations, and equation (20), we obtain

$$\begin{aligned} & \mathbf{E}[\nu_t g_t | F_t - F_{t-n} = x] \\ &= \mathbf{E}(\mathbf{E}[\nu_t g_t | F_t - F_{t-n} = x, \nu_t, \nu_{t-1}, \dots, \nu_{t-n}]) \\ &= \mathbf{E}(\nu_t \mathbf{E}[g_t | F_t - F_{t-n}, \nu_t, \nu_{t-1}, \dots, \nu_{t-n}]) \\ &= w \frac{x}{n} \mathbf{E}[\nu_t (1 + (1 - w\nu_{t-1}) + \dots + (1 - w\nu_{t-1}) \cdots (1 - w\nu_{t-n+1}))] \\ &\quad + w \mathbf{E}[\nu_t (1 - w\nu_{t-1}) \cdots (1 - w\nu_{t-n}) (F_{t-n} - R_{t-n})] \end{aligned}$$

The sign conclusions are obvious given that the  $w\nu$  terms are between 0 and 1 and because this implies that the maximum absolute value of the final expectation term is  $|\mathbf{E}[F_{t-n} - R_{t-n}]|$ , which is bounded. ■

The expectations in Proposition 6 are those of an econometrician who has a prior distribution for the initial conditions at date 0. However, we have been overly generous in our bound on the absolute value of the last term in the proof. For virtually any reasonable set of date 0 values for the fundamental value and the reference price, the final term within the expectation is likely to be negligibly small. Hence, from the perspective of an investor at any date between dates 0 and  $t - n$ , who is aware of the initial conditions, and the path taken by the fundamental value, the expected price change between dates  $t$  and  $t + 1$  will have the properties described in Proposition 6. By date  $t$ , however, and depending on the initial conditions, there are occasional, albeit rare paths, for which the gain is negative despite an increase in the fundamental value between dates  $t - n$  and  $t$ . For this reason, past changes in the fundamental value, and to a greater extent past changes in prices, should be noisier predictors of future price changes than proxies for the unrealized gains.<sup>13</sup> We explore this issue in the empirical section of the paper.

### *E. A Back of the Envelope Calculation of the Expected Price Change*

To assess whether our model generates expected price changes that bear any resemblance to those observed by empiricists, we now undertake a back of the envelope calculation of  $\mathbf{E}[P_{t+1} - P_t | F_t - F_{t-n} = x]$ .

---

<sup>13</sup>If the econometrician can view the distribution of price changes between two consecutive dates within an interval over which a price change has occurred as being symmetric with respect to any pair of consecutive dates, the above proposition goes through in exactly the same form with prices replacing fundamental values. Indeed, as long as expected price changes between two consecutive dates are of the same sign as the price change over the surrounding interval, positive past price changes generate positive expected returns in the period immediately following the interval and vice versa.

Assume for simplicity that for all  $\tau$ ,  $\nu_\tau = \nu$ . By equations (12) and (20),

$$\begin{aligned} \mathbb{E}[P_{t+1} - P_t | F_t - F_{t-n} = x] &= (1-w)\nu \mathbb{E}[g_t | F_t - F_{t-n} = x] \\ &= \frac{x}{n}(1-w)w\nu(1 + (1-w\nu) + (1-w\nu)^2 + \dots) = \frac{1}{n}(1-w)x \end{aligned}$$

Suppose that a trading period corresponds to a month, the fundamental value has increased by 50% over the last twelve months, and  $w = .75$ , as would be the case if 1/3 of a stock's ownership was by type- $d$  investors, each of them had a gain-related disposition effect that influenced their demand function half as much as the spread between the fundamental value and the market price. Then, next month the price is expected to increase by  $\frac{50}{48}\%$  or slightly over 1%.

Of course, despite our best attempts at arguing that this is similar to the size of the Jegadeesh and Titman (1993) momentum effect, we have no way to assess if the 0.75 value for  $w$  is truly reasonable. However, the model does provide other predictions, such as those about volume in the next subsection.

## F. Determinants of Equilibrium Trades and Trading Volume

To obtain an expression for trading volume in the model, substitute the equilibrium price  $P_t$ , as given in equation (5), into the demand equations (1) and (2), and multiply the respective demands by  $1 - \mu$  and  $\mu$ , respectively. This gives the equilibrium aggregate shareholding of each investor-type as a function of the unrealized gain  $g_t = P_t - R_t$ . Specifically,

$$(1 - \mu)D_t^r = 1 - \mu + c_t g_t; \quad \mu D_t^d = \mu - c_t g_t$$

where  $c_t = b_t \lambda \mu (1 - \mu)$  is a positive parameter. It follows that the change in the aggregate equilibrium shareholdings of each investor-type is proportional to the change in the unrealized gain. For type- $r$  investors:

$$(1 - \mu)(D_{t+1}^r - D_t^r) = c_t(g_{t+1} - g_t) + (c_{t+1} - c_t)g_{t+1}$$

while for type- $d$  investors:

$$\mu(D_{t+1}^d - D_t^d) = -c_t(g_{t+1} - g_t) - (c_{t+1} - c_t)g_{t+1} \tag{22}$$

implying the following result:

**Proposition 7** *Assume that  $c_{t+1} - c_t$  is sufficiently small. Then type- $d$  investors sell stock to type- $r$  investors when their unrealized gain increases and buy stock from type- $r$  investors when their gain decreases.*

To elaborate on this point, assume  $c_t = c$  and substitute equation (18) into the right hand side of (22) to obtain

$$\mu(D_{t+1}^d - D_t^d) = -cw(\epsilon_{t+1} - \nu_t g_t)$$

Hence, whether type- $d$  investors buy or sell depends on the sign and the magnitude of the innovation in the fundamental value,  $\epsilon_{t+1}$ , in relation to the hurdle  $\nu_t g_t$ .

- Case 1:  $g_t > 0$ . Then, at  $t+1$ , type- $d$  investors sell (these “winners”) only on sufficiently good news. The bigger the gain  $g_t$ , the better the news must be to induce a sale.
- Case 2:  $g_t < 0$ . Then, at  $t+1$ , type- $d$  investors buy (these “losers”) only on sufficiently bad news. The more negative  $g_t$  is, the worse the news must be to induce additional purchase.

Trading volume, obtained by taking the absolute value of the prior equation is thus

$$V_{t+1} = cw|\epsilon_{t+1} - \nu_t g_t| \tag{23}$$

By substituting equation (9) into (23), we can also express the volume in terms of price changes as opposed to changes in the fundamental value.

$$V_{t+1} = \frac{c}{w} |(P_{t+1} - P_t) - \nu_t g_t| = \frac{c}{w} \left| (P_{t+1} - P_t) - \frac{\mathbf{E}_t[P_{t+1} - P_t]}{1 - w} \right|$$

This implies the following:

**Proposition 8** *Assume that  $c_{t+1} - c_t$  is sufficiently small. Then the volume associated with an increase in the market price is going to be larger when the reference price exceeds the market price than when the market price exceeds the reference price. Also, the volume associated with an decrease in the market price is going to be larger when the market price exceeds the reference price than when the reference price exceeds the market price.*

Stocks that have generally been experiencing price increases are those for which the market tends to exceed the reference price and vice versa. Proposition 8 suggests that major reversals of fortune are more likely to beget larger volume than repetitions of past trends. That is, large price decreases (increases) will be associated with the greatest volume impact for stocks that have experienced major and consistent increases (decreases) in value.

## G. Numerical Findings and the Length of the Past Return Horizon for Momentum

Two types of numerical simulations generated several interesting findings. These simulations assume  $b_t$  is constant over time and that the reference price updating weight is proportional to turnover.

In a binomial simulation of the model where each period  $F$  either goes up (“+” move) or down (“-” move) by a fixed amount:

- Fix  $n > m$ . Among paths with  $n$  “+” moves and  $m$  “-” moves, the path with the highest expected positive price change has all the “-” moves at the beginning, while  $E_t[P_{t+1} - P_t]$  is small or even negative along paths for which all the “-” moves occur at the end. One should avoid buying winners from the more distant past that recently have begun to decline in value in a fairly persistent manner.

When the simulations are based on i.i.d. normal innovations in  $F$ :

- Stocks with high current volume and low past volume tend to have larger momentum. If  $F_t - F_{t-n} > 0$ , then  $E_t[P_{t+1} - P_t]$  is strongly positively correlated with  $V_t$ , and slightly negatively correlated with average past volume over  $[t - n, t]$ . If  $F_t - F_{t-n} < 0$ , then  $E_t[P_{t+1} - P_t]$  is strongly negatively correlated with  $V_t$  and slightly positively correlated with average past volume over  $[t - n, t]$ .
- Volume and absolute price change are strongly positively correlated.
- The difference in 1-month returns of the top decile and bottom decile of past performing stocks, plotted as a function of past return horizon, has a humped shape. See Figure 1.

This last finding deserves special mention as one of the great curiosities of momentum is that it only seems to function at intermediate horizons. Our model does not generate reversals at short or long horizons, but it does generate less profit from momentum when the momentum portfolio is formed using past returns over short or long horizons. The model suggests that the most profitable horizons generated from the model are those that use intermediate horizon past returns for portfolio formation. The numerical finding is difficult to prove analytically; however, it seems rather intuitive. Over very short horizons, it is rather difficult for the gain to deviate from zero by a large amount, as fundamentals have not had much time to move, even among the best and worst performing stocks. While the volatility of the change in a fundamental value that follows a random walk is proportional to the square root of the past horizon’s length, one must also consider how horizon affects the stochastic process for the reference price. The reference price reverts to the fundamental value. Over short horizons, such reversion cannot have much of an effect. However, over a long horizon, reversion to the fundamental value is likely to have a tremendous effect. Indeed, as we learned earlier, the gain will be zero with probability one. Reversion in the gain to zero only enhances the frequency with which this occurs.

A good analogy is a race between two thoroughbred horses with equal expected speed: the fundamental value horse and the reference price horse. When we sort on past winners, we are saying that the fundamental value horse is in the lead. However, both shortly after

the start of the race, and towards the end of the race, he cannot have a very big lead. Near the start, (and even assuming that the initial acceleration from the starting gate took place instantly), the top decile fundamental value horse has not had enough time to get far ahead, despite the average speed being greatest for this leg of the race. Moreover, throughout the race, whenever the fundamental value horse gets far ahead, the reference price horse speeds up. However, the fundamental value horse's speed is not persistent. Hence, conditional on him being far ahead in the first part of the race, he is likely to slow down. This means that the point at which the gap between the top decile fundamental value horse and the reference price horse is likely to be greatest is somewhere in the middle of the race.

## II. Empirical Tests

Our empirical work utilizes weekly returns, turnover (weekly trading volume divided by the number of outstanding shares), and market capitalization data from the MiniCRSP database. The dataset includes all ordinary common shares traded on the NYSE and AMEX exchanges. NASDAQ firms are excluded because of multiple counting of dealer trades. The sample period, from July 1962 to December 1996, consists of 1799 weeks.

### A. Regression Description

We analyze the average slope coefficients of weekly cross-sectional regressions and their time series  $t$ -statistics, as in Fama and MacBeth (1973). The week  $t$  return of stock  $j$ ,  $r_t^j = \frac{P_t^j - P_{t-1}^j}{P_{t-1}^j}$ , is the dependent variable. Denote  $r_{t-t_2:t-t_1}^j$  as stock  $j$ 's cumulative return from weeks  $t-t_2$  to  $t-t_1$ . The prior cumulative returns over short, intermediate, and long horizons are used as control regressors for the return effects described in Jegadeesh (1990), Jegadeesh and Titman (1993), and DeBondt and Thaler (1995). Regressor  $s_{t-1}^j$ , the logarithm of firm  $j$ 's market capitalization at the end of week  $t-1$ , controls for the return premium effect of firm size. We also control for the possible effects of volume, including those described in Lee and Swaminathan (2000) and Gervais, Kaniel, and Minelgrin (2001), by including  $\bar{V}_{t-52:t-1}^j$ , stock  $j$ 's average weekly turnover over the 52 weeks prior to week  $t$  as a regressor (and in later regressions, three interaction terms, computed as the product of the former volume variable and returns over the three past return horizons). We then study the coefficient on  $g_{t-1}^j$ , a capital gains related proxy. Formally, we analyze the regression,

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g \quad (24)$$

and variants of it, where, for brevity, we have dropped  $j$  superscripts and  $t$  subscripts.

Recall that the theoretical model states that

$$E_{t-1} \left[ \frac{P_t - P_{t-1}}{P_{t-1}} \right] = (1 - w) \nu_{t-1} \frac{P_{t-1} - R_{t-1}}{P_{t-1}}$$

This equation suggests that a measurable variable that predicts expected returns is the percentage difference between the market price and the reference price at the beginning of week  $t$ . Our proxy for this variable, the capital gains overhang, is

$$g_{t-1} = \frac{P_{t-2} - R_{t-1}}{P_{t-2}}$$

Theory says that this key regressor should employ  $P_{t-1}$  instead of  $P_{t-2}$ . We lag the market price by one week to avoid confounding market microstructure effects, such as bid-ask bounce.<sup>14</sup>

## B. Specifying a Reference Price

Theory allows the reference price to be any weighted average of historical market prices. In empirical work, we have to specify how the weights used to update the reference prices are determined. Because our theory was motivated by the disposition effect, we believe that reference prices should represent the best estimate of a stock's cost basis to disposition investors. Since we cannot identify who these disposition investors are, our proxy for the cost basis of a stock is an estimate of the aggregate cost basis for all outstanding shares. The date  $t$  aggregate basis thus requires us to use price and volume data to estimate the fraction of shares purchased at date  $t - n < t$ , at price  $P_{t-n}$ , that are still held by their original purchasers at date  $t$ . Summing the products of these fractions and the prices at the relevant prior dates generates the aggregate cost basis for the market.

Estimating these fractions requires us to model trading behavior. We model the fraction of shares purchased at week  $t - n < t$  and held by the week  $t - n$  purchaser through week  $t$  as given by

$$V_{t-n} \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}]$$

If we truncate the reference price estimation process at the price five years prior to week  $t$ , the date  $t$  reference price

$$R_t = \frac{1}{k} \sum_{n=1}^{260} \left( V_{t-n} \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}] \right) P_{t-n} \quad (25)$$

where the scaling constant that makes the fractions sum to one,

$$k = \sum_{n=1}^{260} V_{t-n} \left( \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}] \right)$$

---

<sup>14</sup>Our findings are strengthened without the lag in the price, but possibly for spurious reasons. For this reason, we do not report these regressions in a table. Also, we obtain essentially the same results when we multiply the gain variable by turnover, as specified above. We opt for the more parsimonious representation, which omits this factor, because there may be a cross-sectional relation between a firm's typical  $\nu_t$  and  $w$ , which we cannot estimate.

This model is equivalent to assuming that all shares are symmetric. That is, irrespective of its trading history, each outstanding share is equally likely to be sold at any date. It can be shown that with a constant weekly turnover of  $V$ , the average holding period is  $\frac{1}{V}$  weeks or approximately  $\frac{2}{V}$  years. Assuming a constant weekly turnover of 1%, which is approximately the mean for entire sample period, this implies an average holding period of 2 years.

As noted earlier in the paper, the logic behind the expression for the reference price in equation (25) is straightforward if we assume  $k = 1$ , as is the case when the sum is infinite rather than over 260 weeks. The turnover ratios, the  $V$ s, are then probabilities and each of the bracketed factors inside the product symbol represents the probability that a share did not trade at date  $t - n + \tau$ ; the term in front of the product symbol,  $V_{t-n}$ , represents the probability that the share traded at date  $t - n$ ; the term in large parentheses in equation (25) is the probability that the share's basis is the price at date  $t - n$ ; and the sum is the expected cost basis.

Observe that more recent trading prices have more weight on the reference price, other things equal. This is because the survival probability for a historical price declines geometrically with the passage of time. Indeed, distant prices negligibly influence the reference price. Recognizing that distant market prices have little influence on the regressor, we truncate the estimation at five years and effectively rescale the weights to sum to one by having a  $k < 1$ . This allows us to estimate the reference price in a consistent manner across the sample period. The 5-year cutoff, while arbitrary, allows us to analyze a reasonable portion of our sample period: July 1967 on. Stocks that lack at least five years of historical return and turnover data at a particular week are excluded from the cross-sectional regression for that week.<sup>15</sup>

### *C. Summary Statistics*

Figure 2 plots the weekly time series of the 10th, 50th, and 90th percentile of the capital gains regressor. It indicates that there is wide cross-sectional dispersion in this regressor and a fair amount of time series variation as well. For most firms, the time series of this variable exhibits significant comovements with the past returns of the S&P 500. For the 10th, 50th, and 90th percentile of the regressor, plotted in Figure 1, the correlations between the weekly time series of the regressor and the past one-year percentage change in the S&P 500 index are respectively 0.50, 0.60, and 0.62.

Table 1 Panel A reports summary statistics on each of the variables used in the regression described above. These include time series means and standard deviations of the cross-sectional averages of the dependent and independent variables, along with time series means of their 10th, 50th and 90th percentiles.

---

<sup>15</sup>We verified that our regression results remain about the same when return and turnover data over three or seven prior years are used to calculate the reference price.

We obtain further insight into what determines the critical capital gains regressor by regressing it (cross-sectionally) on stock  $j$ 's cumulative return and average weekly turnover for three past periods: very short term (defined as the last four weeks), intermediate horizon (between one month and one year ago) and long horizon (between one and three years ago). Size is also included as a control regressor. Panel B of Table I reports that, on average, about 59% of the cross-sectional variation in the capital gains variable can be explained by differences in past returns, past turnover, and firm size. Earlier, we explained that the reference price is always trying to catch up to a fundamental value that deviates from the reference price for large return realizations. Consistent with this, Panel B shows that our capital gains variable, in both cases, is positively related to past returns and negatively related to past turnover. Also, consistent with the thoroughbred horse analogy explaining why intermediate horizons are most important, we find that the effect of intermediate horizon turnover on the capital gains variable is much stronger than the effect of turnover from the other two horizons. Controlling for past returns, a low volume winner has a larger capital gain, while a high volume loser has a larger capital loss. Finally, the size coefficient in this regression is significantly positive, perhaps reflecting that large firms have grown in the past at horizons not captured by our past return variables and thus tend to have experienced larger capital gains.

#### *D. Expected Returns, Past Returns, and the Capital Gains Overhang*

Table 2 presents the average coefficients and time-series  $t$ -statistics for the regression described by equation (24) and variations of it that omit certain regressors. Each panel reports average coefficients and test statistics for all months in the sample, for January only, for February-November only, and for December only. All panels include the firm size regressor. Panel A employs only the three past return regressors. Panel B adds volume as a fourth regressor. Panel C adds the capital gains overhang as a fifth regressor.

Panels A and B contain no surprises. As can be seen, when the capital gains overhang variable is excluded from the regression, there is a reversal of returns at both the very short and long horizons, but continuations in returns over the intermediate horizon. Consistent with prior research, the long horizon reversal appears to be due to January. Panel B indicates that there is a volume effect, albeit one that is hard to interpret, but it does not seem to alter the conclusion about the horizons for profitable momentum and contrarian strategies.

Panel C is rather astounding, however. When the capital gains overhang regressor is included in the regression, there is no longer an intermediate horizon momentum effect. The coefficient,  $a_2$ , is insignificant, both overall and from February through November. However, except for January, there is a remarkably strong cross-sectional relation between the capital gains overhang variable and future returns, with a sign predicted by the model.

## E. Explaining Seasonalities

The seasonalities observed in Table II are consistent with what other researchers have found.<sup>16</sup> They are fairly easy to explain within the context of our theoretical model if we accept that there is an additional perturbation in demand arising from tax loss selling.

Grinblatt and Keloharju (2001), for example, found that there was no disposition effect in December, and attributed this to the marginal impact of tax loss selling. If we generalize the demand function of the disposition investor,

$$D_t^d = 1 + b_t[(F_t - P_t) + \lambda_t(R_t - P_t)] \quad (26)$$

and assume that  $\lambda_t$  drifts downward in December for certain, as might be expected because of tax loss selling (possibly, but not necessarily, becoming negative) and reverts to its normal positive value sometime in early January, we would find that the equilibrium effects of this seasonal demand perturbation would be consistent with our empirical findings. The downward drift in  $\lambda$  in December implies that market prices move closer to fundamental values. For stocks with capital losses, implying that the fundamental value is below the market price, convergence towards the fundamental value from the decline in  $\lambda$  represents an added force that makes the market price decline even further than it would were  $\lambda$  to remain constant. Similarly, the increase in  $\lambda$  in early January would make the prices of these same stocks with capital losses deviate again from their fair values, leading to a January reversal.

To understand this more formally, note that with the generalized disposition demand, equation (26), the expected price change, formerly in equation (14), generalizes to

$$E_t[P_{t+1} - P_t] = \left( (1 - w_t)\nu_t + \frac{(w_{t+1} - w_t)(1 - \nu_t w_t)}{w_t} \right) (P_t - R_t)$$

Hence, if we know that  $\lambda_{t+1}$  is going to be lower than  $\lambda_t$ , which makes  $w_{t+1} - w_t$  positive, the expected return between dates  $t$  and  $t + 1$  is going to be larger. The evidence in Grinblatt and Keloharju (2001) suggests that over the course of December,  $\lambda$  declines to zero but is positive during the rest of the year. Viewed from the end of November, this would be like knowing that  $w_{t+1} = 1$  and larger than  $w_t$ , thus generating a larger coefficient on the gain regressor in December than would be observed in months with  $w_{t+1} = w_t$ . Viewed from the end of December,  $w_t = 1$  and larger than  $w_{t+1}$ . This makes the expected price change during January negatively related to the gain regressor.

---

<sup>16</sup>For example, momentum strategies that form portfolios from past returns over intermediate horizons appear to be most effective in December, and there is a strong January reversal in the direction of expected returns when using past returns over any horizon. See, for example, Jegadeesh and Titman (1993), Grundy and Martin (2001) and Grinblatt and Moskowitz (2001)

## *F. Robustness Across Subperiods and Gain Definitions*

To most observers, the first and second half of our sample period present different portraits of the stock market. From July 1967 to March 1982, average returns were low, liquidity was low, and trading costs including commissions were high. The second half of our sample period, April 1982 to December 1996 corresponds to a sea change in the stock market. Beginning in August 1982, average returns and trading volume appeared to explode and trading costs rapidly declined. These subperiods also demarcate an important turning point in the strength of the firm size effect. In the second half of our sample period, size was far less important as a determinant of return premia. Despite these differences, if our theory is part of the core foundation of equilibrium pricing, there should be little difference in the coefficient on our capital gains regressor. Panels D and E of Table II confirm this hypothesis. There is only about a one standard error difference between the average coefficients on the capital gains regressor in the two subperiods. In both subperiods, the average coefficient is highly significant.

Although we do not report this formally in a table, the signs and significance of the capital gains overhang regressor are not drastically altered by restricting the sample to various size quintiles, either. Alternative definitions of the capital gain percentage, such as

$$g_{t-1} = \frac{P_{t-2} - R_{t-1}}{R_{t-1}} \text{ or } \frac{P_{t-2} - R_{t-2}}{P_{t-2}}$$

are also significantly related to the future return and knock out past returns over intermediate horizons as a significant predictor of future returns.

## *G. Alternative Explanations*

Could the strength of the capital gains variable as a predictor of returns be due to some alternative explanation? Table III investigates this issue with respect to two alternatives. First, Panels A and B examine whether there is some sort of interaction between a firm's average historical turnover and future returns. For example, the results in Table II Panel C may have arisen because cross-sectional differences in liquidity imply that the reference prices of more liquid stocks place greater weight on more recent prices than the reference prices of less liquid stocks. By formulating a reference price using the average turnover over the past year in place of each week's actual turnover, we assess whether it is only the cross-sectional difference in liquidity that is responsible for the predictive power of our original gain variable, or whether the information about a stock's capital gain inherent in the time series of its historical weekly turnover also contributes to the predictive power of our findings in Table II.

In Panels A and B of Table III, we compute an alternative week  $t$  reference price using  $\bar{V}_t^j$ , firm  $j$ 's average weekly turnover from weeks  $t - 52$  to  $t - 1$  for all of the 260  $V$ s in

equation (25). Panel A replicates Panel C of Table II, except that in place of the gain variable, we compute an alternative gain variable using the alternative reference price. As Panel A indicates, using a firm’s average turnover for the reference price computation instead of the actual weekly turnover generates a significant coefficient on the gain variable. The results are similar to those of Table II Panel C, in that past returns have no predictive power. Moreover, the coefficients and  $t$ -statistics on the alternative gain variable are similar to those in Table II Panel C.

Table III Panel B runs a horse race between the two gain variables. It is identical to Table III Panel A, except that the Table II proxy for firm  $j$ ’s capital gain is added as a regressor. The inclusion of this variable eliminates the significance of the alternative gain variable, and its coefficient is about the same size as that in Table II Panel C. While our original gain variable is based on an imperfect model of the actual capital gains overhang in the market, it is probably a more precise estimate of aggregate capital gains than the alternative capital gains proxy constructed from average historical turnover. The fact that it “knocks out” the alternative as a predictor of future returns is consistent with more precise estimates of the aggregate capital gain being better predictors of future returns.

A second concern about the significance of our capital gains proxy in Table II is that it represents some complicated interaction between volume and past returns. For example, Lee and Swaminathan (2000) suggested that high volume losers should have lower returns than average volume losers and empirically documented that this was indeed the case. Hence, it is possible that our findings in Table II arise from the capital gains overhang variable being correlated with some interaction between intermediate horizon past returns and volume. Panels C and D of Table III test this hypothesis by adding three turnover and past return interaction terms.

Table 3 Panel C analyzes the impact of these regressors in the absence of a capital gains regressor. Even though two of the three turnover-return coefficients are significant, the inclusion of these regressors does not subsume the intermediate horizon momentum effect. Rather, the volume-return interaction seems to work in part by moderating the strong one-month return reversal.

Once the capital gains variable is added to the regression, as in Table III Panel D, the intermediate horizon past return becomes insignificant, while the capital gains coefficient is highly significant. Comparing Table II Panel C with Table III Panel D, the average regression coefficient for the capital gains variable and its  $t$ -statistic are almost unchanged in the presence of the three turnover and past return interaction terms.

### III. Relation to Prior Research

Jegadeesh and Titman (1993) popularized the notion that strategies of buying stocks with high returns over the prior three to twelve months and selling stocks with poor returns over the same past horizon dominates a buy and hold strategy,<sup>17</sup> Ever since then, researchers have attempted to come up with explanations for the phenomenon.

Conrad and Kaul (1998) argue that the profitability of momentum strategies could be due to cross-sectional variation in the unconditional expected returns rather than any predictable time-series variation in stock returns. Yet Jegadeesh and Titman (2000) find that the cumulative return in months 13 to 60 after the formation of momentum portfolio is negative, which is inconsistent with the Conrad and Kaul hypothesis. Moskowitz and Grinblatt (1999) show the component of momentum profits due to cross-sectional variation in unconditional expected returns is small. Grundy and Martin's (2001) evidence also appears to contradict this hypothesis. They find that the risk-adjusted profitability of a total return momentum strategy is more than 1.3% per month and remarkably large and stable across subperiods, even after subtracting each stock's mean return from its return during the investment period.

Moskowitz and Grinblatt (1999) find that a significant component of momentum can be explained by industry effects. However, this does not mean that individual stock momentum does not exist. Moskowitz and Grinblatt (1999), Grundy and Martin (2001), and Chordia and Shivakumar (2000) show that a component of individual stock momentum is distinct from industry momentum. The latter paper also argues that momentum profits are driven by time varying conditional expected returns that are related to the business cycle.

Another strand of the literature uses behavioral models to explain momentum profits.<sup>18</sup> These models can be divided into two camps, depending on whether investor behavior generates overreaction or underreaction. In the positive feedback trader model of DeLong et al (1990b), prices initially overreact to news about fundamentals, and continue to overreact for a period of time. Daniel, Hirshleifer and Subrahmanyam (1998) present a model where investors are overconfident. This implies overreaction to private information and underreaction to public information arrival. The investors also suffer from a self-attribution bias. Their

---

<sup>17</sup>This finding appears to be fairly universal and robust to methodological tweaking. Rouwenhorst (1997), for example, finds that momentum strategies work in twelve European markets. Chui, Titman, and Wei (2000) document that with the exception of Japan and Korea, momentum profits can be earned in Asian markets. Jegadeesh and Titman (2000) document that momentum profits persisted throughout the 1990s. In contrast, other well known anomalies such as small firm effect and book-to-market effect disappeared after being well-publicized. Jegadeesh and Titman (1993) and Fama and French (1996) find that risk adjustment tends to accentuate momentum profits. Chan, Jegadeesh and Lakonishok (1996) show that intermediate horizon return continuation can be partially explained by underreaction to earnings news but that price momentum is not subsumed by earnings momentum. Lee and Swaminathan (2000) show that past trading volume predicts both the magnitude and the persistence of future price momentum.

<sup>18</sup>Hirshleifer (2001) gives a comprehensive account of psychological biases and empirical evidence on the importance of investor psychology for security prices.

behavior generates delayed overreaction to the information which is eventually reversed.

Barberis, Shleifer and Vishny (1998) argue that the representative heuristic<sup>19</sup> may lead investors to extrapolate current earnings growth well into the future. At the same time, investors' conservatism bias leads to underreaction to new public information. In Hong and Stein (2000), agents can use only part of the information about the economy because of communication frictions. In their model, private information diffuses slowly through the population of investors, which causes underreaction in the short run. Momentum traders can profit by trend-chasing, but cause overreaction at long horizons in doing so.

Our explanation of the profitability of momentum strategies is distinct from explanations in prior research. Our investors have no cognitive biases such as those based on overconfidence, self attribution, conservatism, or representativeness. There is no mistaken belief about the fundamental value. There is no time variation in risk, risk aversion, or investor sentiment driving our results. There are no hidden factors such as those based on industry. Information is symmetric. In asymmetric information models, trading volume reflects investors' disagreements about a stock's intrinsic value, and often requires the existence of noise traders to generate trading volume. There is no information asymmetry in our model, and no noise traders, but there is volume. Trading occurs because some investors are subject to the disposition effect. Past trading volume affects the equilibrium price through its influence on the reference price, while most extant models have a representative agent and volume plays no role.

Most importantly, our model is based on well-documented investor behavior and principles of psychology. Disposition behavior, while inconsistent with the standard neoclassical framework, has been justified as a consequence of theories of behavior including prospect theory,<sup>20</sup> regret theory,<sup>21</sup> and cognitive dissonance theory.<sup>22</sup> Camerer and Weber (1998) and Heilmann, Lager and Oehler (2000) found evidence for disposition behavior in experimental markets. Evidence of disposition behavior among actual investors is found in Odean (1998), Heath, Huddart and Lang (1998), Grinblatt and Keloharju (2001), and Locke and Mann (1999). Odean (1998) analyzes accounts at a large brokerage house and found that there was a greater tendency to sell stocks with paper capital gains than those with paper losses. Grinblatt and Keloharju (2001) find a similar effect among all types of investors in Finland, even after controlling for a variety of variables that may determine trading. They also observe that the disposition behavior interacts with past returns in a multiplicative fashion and has a pronounced seasonality: it disappears in December. Using data from a major Israeli brokerage house during 1994, Shapira and Venezia show that both professional and independent investors exhibit the disposition effect, although the effect is stronger for independent investors. Heath, Huddart and Lang (1998) uncover disposition behavior relative to

---

<sup>19</sup>See Tversky and Kahneman (1974).

<sup>20</sup>See Shefrin and Statman's (1985) interpretation of Kahneman and Tversky (1979).

<sup>21</sup>See Shefrin and Statman (1985)

<sup>22</sup>See Shefrin and Statman (1985), Camerer and Weber (1998).

a reference price of a prior high for the stock price by studying the option exercise behavior of over 50,000 employees at seven corporations. Locke and Mann (1999) present evidence for the existence of a disposition effect within a sample of professional futures traders. In their study, traders held losing trades longer than winning trades and average position sizes for losing trades were larger than for winners. Ferris, Haugen and Makhija (1988) argue that a disposition effect has to exist by studying the relationship between volume at a given point in time with historical volume at differential prices, controlling for seasonal effects to isolate tax motivated trading. The disposition effect also influences agents in the IPO and housing markets.<sup>23</sup>

There are a set of papers that are linked to modeling how loss aversion affects equilibrium prices and portfolio holdings.<sup>24</sup> Barberis, Huang and Santos (2001) model investor preferences to reflect a combination of loss aversion and the “house money” effect of Thaler and Johnson (1990). Their goal is to show how changing risk aversion explains the high mean, high volatility, and significant predictability of stock returns. They have a representative agent and no trading, but find that the variation in risk aversion of their representative agent allows returns to be much more volatile than the underlying dividends. Moreover, asset return predictability found in their model is consistent with the profitability of contrarian strategies. Our model, by contrast, is consistent with contrarian strategies being money losers and our investors limited willingness to take positions is consistent with risk aversion. Trading arises only because of the disposition effect and plays an important role in generating momentum. Although both the house money effect and the disposition effect can be rooted in prospect theory, the fact that they lead to such opposite results suggests that they are truly distinct phenomena. Barberis and Huang (2001) extend this paper to multiple assets in order to address cross-sectional expected return patterns, such as the value premium. Ang, Berkaert, and Liu (2001) study portfolio choice under the disappointment aversion preference, where outcomes below the certainty equivalent are weighted more heavily than above the certainty equivalent. They study optimal “non-participation” in the stock market and cross-sectional variation in portfolio holdings, and they contrast their preference structure with loss aversion, which, for many parameter values leads to troublesome portfolio predictions. Both Gomes (2000) and Berkelaar and Kouwenberg (2000) study optimal portfolio choice under loss aversion. Both papers find that demand under loss aversion shares some common features with the disposition effect, pointing to the possibility that loss aversion can be consistent with the disposition effect.

---

<sup>23</sup>See Case and Shiller (1988), Genesove and Mayer (2001) and Kaustia (2001).

<sup>24</sup>In empirical work, Coval and Shumway (2001) document Chicago Board of Trade proprietary futures traders are highly loss averse, as they assume significantly more afternoon risk following morning losses than following morning gains.

## IV. Conclusion

Our paper has developed a model of equilibrium asset prices based on the disposition effect. By restricting loss realization aversion to be a geometric deviation from fully rational behavior, we are able to generate closed form solutions and a set of powerful propositions about conditional expected future returns without solving the optimal dynamic portfolio problem of rational agents. The solution to the portfolio problem drops out of the equations we are interested in, which focus on deviations from the rational norm. We then test the model and surprisingly show that the Jegadeesh and Titman (1993) momentum effect largely disappears. This suggests that it is the correlation between past returns and variables related to the disposition effect that may be driving momentum in stock returns.

In the model presented here, the critical variable determining the sign and magnitude of a stock's expected return is the difference between the stock's market price and its reference price. Despite having fully rational investors in the model, they cannot eliminate the impact of the gain on equilibrium prices. Although, on average, the gain shrinks, the payoff to more rational investors is uncertain. Hence, rational investors cannot ascertain when reference prices, and hence market prices, will converge to fundamental values.

DeLong et al (1990b) show that when there are positive feedback traders in the economy, rational arbitrageurs who anticipate their impact on demand can front-run the other investors and destabilize prices. Specifically, when the rational investor receives good news today, he buys more shares to drive up the price. This, in turn, attracts the positive feedback investors who buy tomorrow so that the rational investor can exit with a profit. In our model, there is no way to anticipate the disposition demand in advance, as it is determined by the future realization of the fundamental value, which follows a random walk and hence is unpredictable. Any degree of risk aversion on the part of rational agents thus prevents the model from collapsing.

Our model falls in the class of “underreaction models.” However, it also points out some interesting implications of underreaction and suggests that our field may have to better clarify what we mean by the term. For example, Barberis *et al.* (1998) and Shleifer (2000) define underreaction as occurring when the average return on the stock following good news is higher than the average return following bad news. Our model clearly has underreaction, but the fit with this definition is imperfect because path dependency generates cases where this definition does not hold.<sup>25</sup> Our model points to the difficulty of measuring underreaction

---

<sup>25</sup>The future expected return in our model is of the same sign as the current gain. If the gain is negative now, it will continue to be negative even after good news is announced, assuming that the news is not good enough. Hence the expected return is negative following good news following paths with capital losses. On the other hand, if the current gain is positive, it may still be positive and hence the expected return is positive, even if the news is bad (but not too bad). Hence, because the path associated with the good news had a capital loss and the path associated with the bad news had a capital gain, the expected return after good news was lower than the expected return following bad news. While this is generally not the case, it

(or overreaction) in terms of subsequent price changes without being able to measure the degree of underreaction (or overreaction) existing in the market at all times used for these computations. Similarly, special cases of our model have underreaction at all times, but no tendency for prices to converge to fundamental values. In these cases, there is not even the positive autocorrelation that is typically associated with underreaction.

Our assumptions are quite general, allowing the model's analysis of a partial equilibrium for a single asset to be consistent with more comprehensive modeling of a multi-asset equilibrium. The process by which the market arrives at a fair value in an intertemporal multi-asset economy can be quite complicated, but that is not our concern. We simply want to understand as clearly and analytically as possible how a perturbation of investors' demand for a single stock, due to the disposition effect, generates deviations from the fully rational equilibrium. Our "partial" equilibrium approach should not generate conclusions that differ from those obtained by postulating utility functions and solving for demand functions over multiple assets that optimize utility unless momentum strategies are true arbitrages that lack risk. However, prior empirical evidence indicates that momentum strategies are quite risky. Moreover, we have found that the capital gains overhang of individual stocks is strongly correlated with the cumulative past return of a broad market index like S&P 500. Thus the disposition components of demand across stocks are also likely to be positively correlated, making the disposition effect a systematic risk to potential arbitrageurs. In short, we believe that our approach, despite lacking a closed form solution for demand functions, does not generate aberrational conclusions.

Similarly, using what effectively are two representative agents is an oversimplification. However, such a simplification is reasonable if aggregate demand generates effective aggregate reference prices that are weighted averages of current prices and past aggregate reference prices. While this aggregation cannot be done analytically, we have been quite general in allowing weights for reference prices to time vary and be path dependent. With this level of generality in reference price construction, we would be surprised if the two representative agent paradigm used here does not hold up to closer scrutiny.

The generality of the reference price updating rule in the paper's theoretical section is both a strength and a weakness. Any attempt at defining an aggregate reference price for empirical work requires a concrete updating weight, with little guidance from our theory. Real-world equilibria, with multiple investors, requires aggregation of investors' demand and reference prices. Our specification of the aggregate reference price for empirical work is not the only solution to this problem. While we have analyzed modest variations in the reference price updating rule and found nothing to refute our conclusions, the exploration of appropriate and alternative reference price rules is certainly an interesting avenue for future research.

---

points to the need for more precision in any definition of underreaction.

**Table I**  
**Summary Statistics**

This table presents summary statistics of weekly data on NYSE and AMEX securities from July 1967 to December 1996, obtained from mini-CRSP. Panel A provides time series averages of the cross-sectional mean, median, standard deviation, and 10th, 50th, and 90th percentiles of each of the variables used in the regression

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g$$

where  $r$  is the week  $t$  return,  $r_{-t_1:-t_2}$  is the cumulative return from week  $t - t_1$  through  $t - t_2$ ;  $\bar{V}$  is the average weekly turnover ratio over the prior 52 weeks, the ratio of the week's share volume to the number of outstanding shares;  $s$  is  $\log(\text{market capitalization})$  measured at the beginning of week  $t$ ;  $g$  is the capital gains regressor, computed as one less the ratio of the end of week  $t - 1$  reference price to the end of week  $t - 2$  price, where the week  $t - 1$  reference price is the average cost basis calculated from the formula

$$R_t = \frac{1}{k} \sum_{n=1}^{260} \left( V_{t-n} \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}] \right) P_{t-n}$$

with  $k$  a constant that makes the weights on past prices sum to one. Panel B presents more detailed data on the association between the capital gains regressor and other variables. It contains the time-series average of the coefficients and their associated time series  $t$ -statistics for 1539 weekly Fama-MacBeth type cross-sectional regressions of the form

$$g = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 V_{-4:-1} + a_5 V_{-52:-5} + a_6 V_{-156:-53} + a_7 s$$

where  $V_{-t_1:-t_2}$  is the average weekly turnover from  $t - t_1$  through  $t - t_2$ .  $R_{\text{adj}}^2$  is the average of the weekly cross-sectional regression  $R^2$ s adjusted for degrees of freedom.

Panel A: Time series average of summary statistics of the regressors in the regression

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g$$

	$r_{-4:-1}$	$r_{-52:-5}$	$r_{-156:-53}$	$\bar{V}$	$s$	$g$
Mean	0.0119	0.1493	0.3487	0.0092	18.7207	0.0560
Median	0.0045	0.0940	0.2098	0.0072	18.7251	0.1062
Std	0.1073	0.4192	0.7585	0.0079	1.9441	0.2508
10 percentile	-0.0959	-0.2538	-0.3227	0.0025	16.1399	-0.2810
90 percentile	0.1223	0.5816	1.1097	0.0181	21.2322	0.3122

Panel B: Average coefficients and  $t$ -statistics (in parentheses) for the regression

$$g = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 V_{-4:-1} + a_5 V_{-52:-5} + a_6 V_{-156:-53} + a_7 s$$

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$R_{\text{adj}}^2$
0.5527	0.4907	0.1771	-0.9159	-6.4051	-2.7843	0.0504	0.5879
(73.0290)	(51.7965)	(37.5209)	(-7.6351)	(-45.0322)	(-27.8215)	(55.9642)	

**Table II**  
**Cross-sectional Regression Estimates**

This table presents the results of Fama-MacBeth (1973) cross-sectional regressions run each week on NYSE and Amex securities from July 1967 to December 1996. The weekly cross-sectional regressions include all stocks that have at least five years of historical trading data on mini-CRSP. The cross section of stock returns in week  $t$ , denoted  $r$ , are regressed on a constant and some or all of the following variables:  $r_{-t_1:-t_2}$  = the cumulative return from week  $t-t_1$  through  $t-t_2$ , computed over three past return horizons;  $\bar{V}$  = the average weekly turnover ratio over the prior 52 weeks, with turnover being the ratio of the week's share volume to the number of outstanding shares;  $s = \log(\text{market capitalization})$  measured at the beginning of week  $t$ ; and  $g$  = the capital gains regressor, computed as one less the ratio of the end of week  $t-1$  reference price to the end of week  $t-2$  price, where the week  $t-1$  reference price is the average cost basis calculated from the formula

$$R_t = \frac{1}{k} \sum_{n=1}^{260} \left( V_{t-n} \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}] \right) P_{t-n}$$

with  $k$  a constant that makes the weights on past prices sum to one.  $R_{\text{adj}}^2$  is the average of the weekly cross-sectional regression  $R^2$ s adjusted for degrees of freedom. There are a total of 1539 weekly regressions. The parameter estimates and  $t$ -statistics (in parentheses) are obtained from the time series of the corresponding cross-sectional regression coefficients. We report the results of regressions over all months, for January only, February through November only, and December only. Panel A omits the capital gains and turnover variables. Panel B omits the capital gains variable. Panel C contains the full set of regressors. Panels D and E report results for the full set of regressors over the first and second half of the sample period.

Panel A

$$r_t = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 s$$

Period	$a_1$	$a_2$	$a_3$	$a_4$
All	-0.0482 (-35.6415)	0.0012 (2.9527)	-0.0005 (-3.0054)	-0.0004 (-4.2733)
Jan	-0.0700 (-9.6647)	-0.0087 (-4.5972)	-0.0068 (-6.6744)	-0.0040 (-10.9146)
Feb-Nov	-0.0459 (-34.0613)	0.0018 (4.3344)	-0.0001 (-0.6243)	-0.0001 (-1.4488)
Dec	-0.0491 (-9.9440)	0.0051 (3.8921)	0.0015 (2.8930)	0.0008 (3.0164)

Panel B

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
All	-0.0488 (-37.2470)	0.0014 (3.5703)	-0.0005 (-2.6700)	-0.0540 (-2.5732)	-0.0004 (-4.4200)
Jan	-0.0706 (-9.7366)	-0.0086 (-4.5561)	-0.0069 (-6.5561)	0.0681 (0.9793)	-0.0042 (-11.2309)
Feb-Nov	-0.0465 (-36.0594)	0.0021 (5.1324)	-0.0000 (-0.1979)	-0.0729 (-3.1591)	-0.0001 (-1.5202)
Dec	-0.0489 (-10.2429)	0.0049 (3.7745)	0.0015 (2.8046)	0.0088 (0.1214)	0.0009 (3.1917)

Panel C

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
All	-0.0425 (-35.9364)	-0.0002 (-0.6794)	-0.0007 (-5.0871)	-0.0188 (-0.9364)	-0.0004 (-5.2885)	0.0040 (7.7885)
Jan	-0.0520 (-10.9905)	-0.0001 (-0.0477)	-0.0025 (-3.8964)	-0.0620 (-0.9768)	-0.0026 (-8.4381)	-0.0117 (-4.9519)
Feb-Nov	-0.0407 (-32.6251)	-0.0000 (-0.0768)	-0.0006 (-3.6950)	-0.0291 (-1.3143)	-0.0002 (-2.8816)	0.0050 (9.4191)
Dec	-0.0498 (-10.8151)	-0.0022 (-1.8953)	-0.0005 (-1.3410)	0.1238 (1.7980)	0.0001 (0.2702)	0.0104 (6.2673)

Panel D: July 1967 to March 1982

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
All	-0.0552 (-31.6943)	-0.0005 (-0.9578)	-0.0013 (-5.7743)	-0.0143 (-0.4054)	-0.0007 (-5.2407)	0.0046 (6.1793)
Jan	-0.0631 (-7.9314)	-0.0005 (-0.2847)	-0.0045 (-4.5862)	-0.1711 (-1.7864)	-0.0038 (-8.4704)	-0.0123 (-4.0505)
Feb-Nov	-0.0532 (-29.2124)	-0.0004 (-0.6562)	-0.0011 (-4.2579)	-0.0231 (-0.5866)	-0.0004 (-3.0217)	0.0058 (7.5394)
Dec	-0.0666 (-10.9771)	-0.0016 (-0.8759)	-0.0007 (-1.2674)	0.2267 (1.9665)	0.0001 (0.2758)	0.0102 (4.2340)

Panel E: April 1982 to December 1996

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
All	-0.0297 (-20.3628)	0.0000 (0.1063)	-0.0001 (-0.6985)	-0.0233 (-1.2045)	-0.0002 (-1.8569)	0.0035 (4.8216)
Jan	-0.0401 (-8.9945)	0.0004 (0.2767)	-0.0004 (-0.4923)	0.0540 (0.6699)	-0.0013 (-3.6897)	-0.0110 (-3.0077)
Feb-Nov	-0.0284 (-18.1436)	0.0003 (0.6909)	-0.0001 (-0.4204)	-0.0350 (-1.7047)	-0.0001 (-0.8256)	0.0042 (5.7574)
Dec	-0.0325 (-5.1506)	-0.0028 (-1.9620)	-0.0003 (-0.5609)	0.0177 (0.2447)	0.0000 (0.0839)	0.0106 (4.6193)

**Table III**  
**Alternative Explanations**

This table investigates alternative explanations for the significance of the coefficient on the capital gains regressor. For Panels A and B,  $\bar{g}$  is calculated from a reference price using  $\bar{V}_t^j$ , firm  $j$ 's average weekly turnover from weeks  $t - 52$  to  $t - 1$  in the formula for the gain variable used in week  $t$ 's cross-sectional regression. Panel A replicates Panel C of Table II, replacing our original capital gains variable by  $\bar{g}$ . In Panel B, the relative significance of the two gain variables are compared by including both as regressors. Panels C and D investigate whether significance was generated by the capital gains variable being correlated with some interaction between past returns and volume over several horizons. Panels C and D add three turnover and past return interaction terms without and with our original capital gains variable, respectively. The parameter estimates and t-statistics (in parentheses) are obtained from the time series of the corresponding cross-sectional regression coefficients. There are a total of 1539 weekly regressions.

Panel A

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 \bar{g}$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
All	-0.0419 (-35.3749)	-0.0003 (-0.9434)	-0.0008 (-5.6612)	-0.0160 (-0.8074)	-0.0003 (-4.3955)	0.0043 (8.0694)
Jan	-0.0511 (-10.8551)	-0.0004 (-0.3277)	-0.0026 (-4.0810)	-0.0553 (-0.8509)	-0.0030 (-9.4107)	-0.0097 (-3.9209)
Feb-Nov	-0.0403 (-32.0373)	-0.0001 (-0.3395)	-0.0007 (-4.2724)	-0.0266 (-1.2182)	-0.0002 (-1.8236)	0.0051 (9.2848)
Dec	-0.0488 (-10.7502)	-0.0019 (-1.7329)	-0.0005 (-1.3532)	0.1250 (1.8159)	0.0003 (1.2605)	0.0103 (6.0802)

Panel B

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 s + a_6 g + a_7 \bar{g}$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$
All	-0.0424 (-34.7482)	-0.0003 (-1.0308)	-0.0008 (-5.1146)	-0.0133 (-0.6459)	-0.0003 (-3.8746)	0.0028 (2.4381)	0.0014 (1.2590)
Jan	-0.0524 (-10.9145)	-0.0010 (-0.7809)	-0.0029 (-4.2183)	-0.0193 (-0.2889)	-0.0022 (-7.7160)	-0.0238 (-4.0727)	0.0154 (2.5766)
Feb-Nov	-0.0405 (-31.3421)	-0.0001 (-0.2966)	-0.0006 (-3.6772)	-0.0260 (-1.1447)	-0.0001 (-1.8369)	0.0042 (3.6161)	0.0007 (0.5650)
Dec	-0.0513 (-10.7503)	-0.0020 (-1.6643)	-0.0004 (-0.9656)	0.1160 (1.5729)	0.0002 (0.9717)	0.0152 (4.5511)	-0.0048 (-1.4017)

Panel C

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 \bar{V} * r_{-4:-1} + a_6 \bar{V} * r_{-52:-5} + a_7 \bar{V} * r_{-156:-53} + a_8 s$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$
All	-0.0601 (-39.0925)	0.0013 (2.7829)	-0.0007 (-2.9421)	-0.0805 (-3.1500)	1.2308 (13.2851)	0.0169 (0.8890)	0.0178 (1.8128)	-0.0004 (-4.4309)
Jan	-0.0863 (-10.3072)	-0.0115 (-4.7567)	-0.0086 (-6.3218)	-0.0261 (-0.3156)	1.6915 (4.8222)	0.2757 (2.8754)	0.1494 (3.0648)	-0.0041 (-11.3921)
Feb-Nov	-0.0569 (-36.9647)	0.0020 (4.3620)	-0.0002 (-0.8066)	-0.1055 (-3.7325)	1.1440 (11.3819)	0.0074 (0.3675)	0.0132 (1.2977)	-0.0001 (-1.5696)
Dec	-0.0647 (-12.7619)	0.0067 (4.6382)	0.0023 (3.5868)	0.1080 (1.2784)	1.6171 (5.1047)	-0.1455 (-3.2586)	-0.0672 (-2.1602)	0.0008 (3.1884)

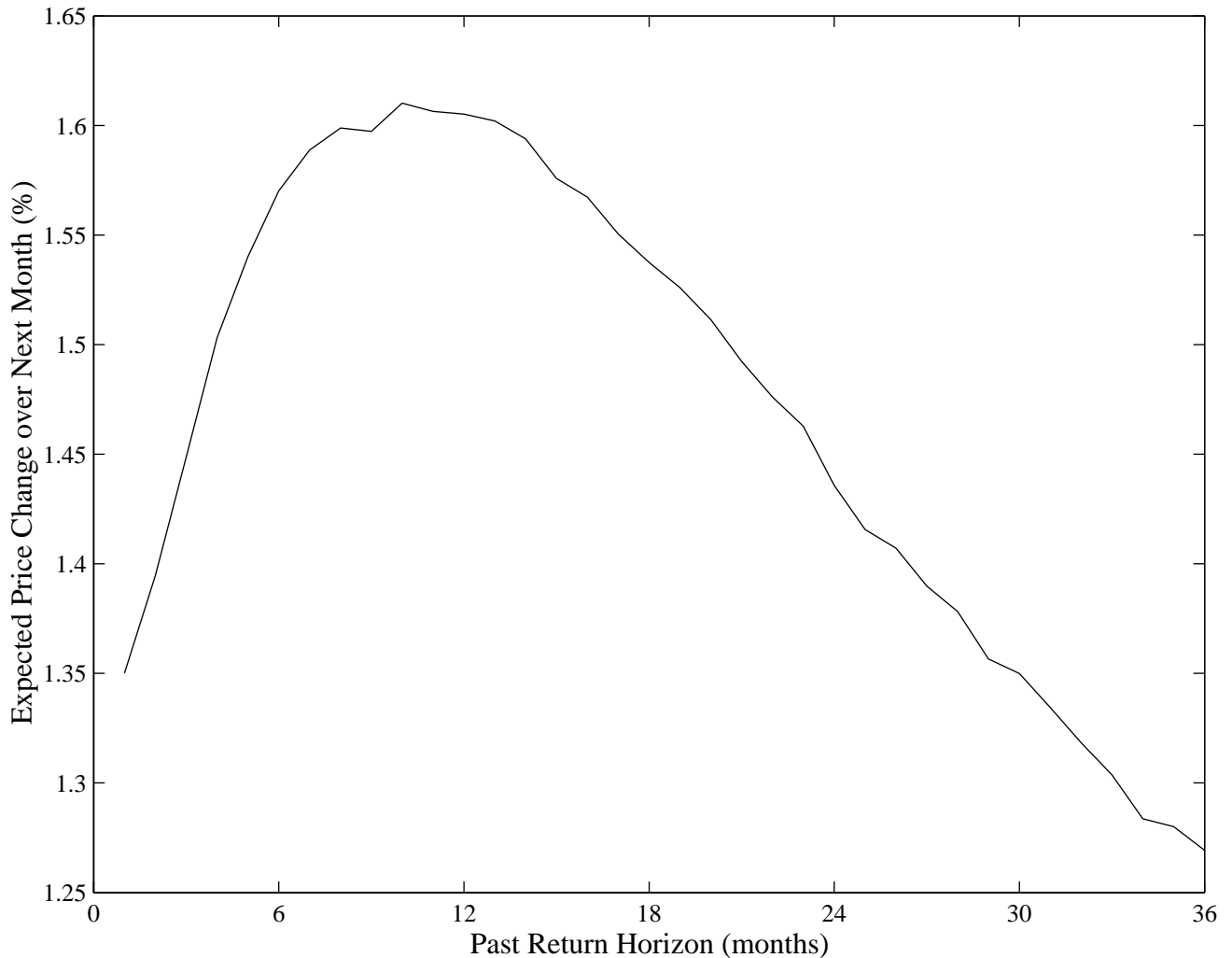
Panel D

$$r = a_0 + a_1 r_{-4:-1} + a_2 r_{-52:-5} + a_3 r_{-156:-53} + a_4 \bar{V} + a_5 \bar{V} * r_{-4:-1} + a_6 \bar{V} * r_{-52:-5} + a_7 \bar{V} * r_{-156:-53} + a_8 s + a_9 g$$

Period	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$
All	-0.0505 (-36.9999)	-0.0004 (-1.0135)	-0.0010 (-5.2744)	-0.0448 (-1.8042)	0.8291 (9.8004)	0.0172 (0.9803)	0.0203 (2.0141)	-0.0004 (-5.4115)	0.0041 (7.7062)
Jan	-0.0631 (-10.9625)	0.0004 (0.2253)	-0.0020 (-2.1356)	-0.0606 (-0.7509)	1.2519 (4.2349)	0.0256 (0.3194)	-0.0324 (-0.7282)	-0.0026 (-8.6123)	-0.0127 (-5.2671)
Feb-Nov	-0.0481 (-33.5092)	-0.0004 (-0.8019)	-0.0010 (-4.7637)	-0.0668 (-2.4386)	0.7449 (8.0398)	0.0230 (1.2122)	0.0295 (2.7375)	-0.0002 (-3.0289)	0.0052 (9.5347)
Dec	-0.0619 (-12.0063)	-0.0017 (-1.3278)	-0.0004 (-0.8323)	0.1836 (2.2080)	1.2275 (4.2620)	-0.0471 (-1.0533)	-0.0171 (-0.5460)	0.0001 (0.3912)	0.0102 (5.9292)

**Figure 1**  
**Momentum for Different Past Return Horizons**

This figure plots the expected price change over next month in a simulated economy for momentum strategies that sort stocks on past returns over different horizons. First, 100,000 paths of fundamental values  $F$  over 60 months are simulated according to a random walk model, assuming initial  $F = R = 1$ . The reference price, the market price and agents' equilibrium holdings are calculated along each path, assuming  $R_{t+1} = V_t P_t + (1 - V_t) R_t$ , where  $V_t$  is turnover ratio. At month 60, a momentum portfolio is formed that buys the top decile (winners) and shorts the lowest decile (losers) sorted according to the change in the fundamental value over past  $n$  months, for  $n = 1, 2, \dots, 36$ . The expected price change over the next month of this winner minus loser portfolio is plotted against the past return horizon used to identify winners and losers. The graph is generated with following model parameters: annual volatility of  $F$  is  $\sigma = 30\%$ ;  $b$  constant and consistent with type- $r$  investors' absolute risk aversion coefficient of  $\gamma = 2$  (a myopic exponential utility function);  $\mu = 1/3$ , implying that  $1/3$  of investors are subject to the disposition effect and  $\lambda = 1$ .

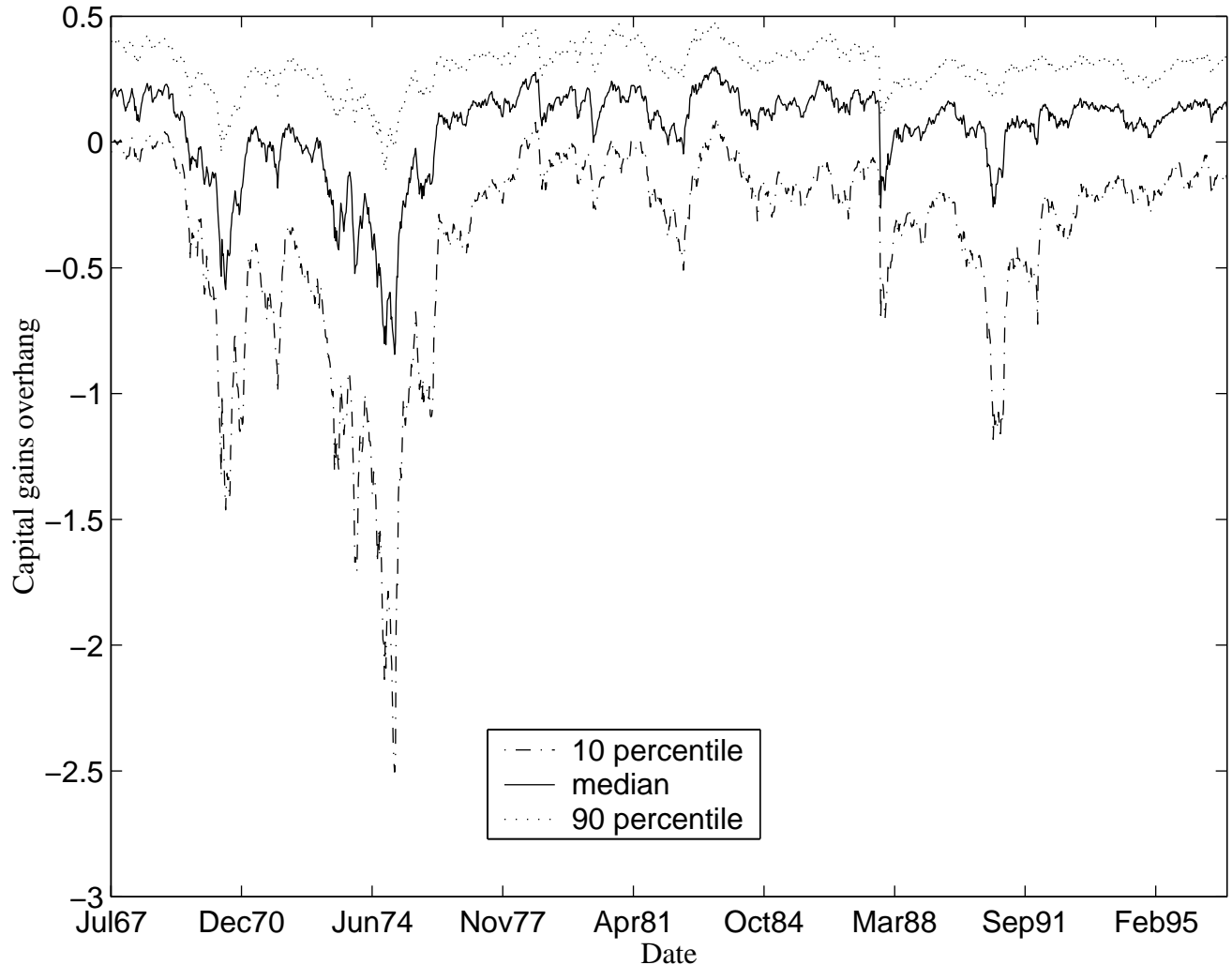


**Figure 2**  
**Time Series of Cross-Sectional Percentiles of the Capital Gains Regressor**

This figure plots the time series of the empirical 10th, 50th and 90th percentiles of the cross-sectional distribution of the capital gains regressor. The sample period is from July 1967 to December 1996, for a total of 1539 weeks. Each week, we include all stocks (with sharecode 10 or 11) listed on NYSE and AMEX which have at least five years of historical trading data from mini-CRSP. The previous five years of return and turnover data are used to calculate the capital gains variable as one less the ratio of the end of week  $t - 1$  reference price to the end of week  $t - 2$  price, where the week  $t - 1$  reference price is the average cost basis obtained from the formula

$$R_t = \frac{1}{k} \sum_{n=1}^{260} \left( V_{t-n} \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}] \right) P_{t-n}$$

with  $k$  a constant that makes the weights on past prices sum to one.



## Appendix

### Equilibrium and Demand Functions in a 3-Date Exponential Utility Model

Within the context of the model described in Section I, we consider the date 0 valuation of two securities: a risk-free asset with a return of 0 in infinitely elastic supply and a risky stock that pays a liquidating dividend  $\tilde{F}_2$  at date 2. Trading at date 1 occurs after receiving a normally distributed signal about  $F_2$  that resolves half of the uncertainty about the final payoff. That is,

$$E_1[F_2] = E_0[F_2] + \epsilon_1, \quad \epsilon_1 \sim N(0, \sigma^2)$$

and  $\text{Var}_1(F_2) = \sigma^2$ . It follows that

$$\text{Var}_0(F_2) = E_0[\text{Var}_1(F_2)] + \text{Var}_0(E_1[F_2]) = 2\sigma^2$$

There are two types of price-taking investors: type- $r$ , whose demand function has weight  $1 - \mu$ , has CARA utility over terminal wealth without intermediate consumption. This type chooses date 0 risky asset shareholding  $\alpha_0$  and date 1 shareholding  $\alpha_1$  in the stock to maximize expected utility of final wealth:

$$\text{Max}_{\alpha_0, \alpha_1} E[-e^{-\gamma W_2}]$$

where  $\gamma$  is his absolute risk aversion coefficient. Denote  $D_t^r(P_t)$  as type  $r$ 's optimal demand at date  $t$  given price  $P_t$ , and define  $F_t$  as the price at which he would optimally hold one share of the stock, assuming knowledge of how type- $d$  investors influence future equilibrium prices. The latter investor-type, with weight  $\mu$  in the economy has date  $t$  demand given by

$$D_t^r(P_t) + \lambda \frac{D_t^r(P_t) - D_t^r(F_t)}{F_t - P_t} (R_t - P_t)$$

with variables and their dynamics defined in the body of the paper.

We calculate the type- $r$  investors optimal demand function (in shares) for the stock, given the conjecture that the linear equilibrium price function

$$P_t = wF_t + (1 - w)R_t, \text{ where } w = \frac{1}{1 + \mu\lambda}$$

applies. By the argument in the body of the text, the market will clear at the above conjectured price as long as type  $r$ 's date 0 demand  $D_0^r(P_0)$  exists (and is finite).

We now explicitly calculate  $D_0^r(P)$  by backwards induction, and show that  $b_0 = \frac{D_0^r(P_0) - 1}{F_0 - P_0}$  is positive. At date 2, the risky asset's value is  $F_2$  as there is only a liquidating payout. Hence at date 1, the rational agent's optimal demand at any price  $P_1$  will be the same as

that in a fully rational economy. With CARA utility and a normally distributed date 2 payoff, the date 1 demand of type- $r$  investors must satisfy:

$$\alpha_1(P_1) = \frac{\mathbf{E}_1[F_2] - P_1}{\gamma\sigma^2} = \frac{\theta + \epsilon_1}{\gamma\sigma^2}$$

which generates a fully rational equilibrium price satisfying

$$F_1 = \theta + \epsilon_1 - \gamma\sigma^2 \quad (27)$$

We obtain the date 1 indirect utility function  $J_1(W_1, P_1)$  by evaluating the expected utility at the optimal demand above:

$$J_1(W_1, P_1) = -e^{-\gamma\left(W_1 + \frac{\mathbf{E}_1[F_2] - P_1}{2\gamma\sigma^2}\right)^2}$$

We now turn to the optimal demand at date 0. At any price  $P_0$  at date 0, the type- $r$  investors choose  $\alpha_0 = \alpha_0^*(P_0)$  to

$$\text{Maximize } \mathbf{E}_0[J_1(W_1, P_1^*)]$$

where  $P_1^* = wF_1 + (1-w)R_1$  is the equilibrium price at date 1, and

$$W_1 = W_0 + \alpha_0(P_1^* - P_0) = W_0 + \alpha_0(wF_1 + (1-w)R_1 - P_0)$$

which is equivalent to

$$\text{Max}_{\alpha_0} - \mathbf{E}_0 \left[ e^{-\gamma\alpha_0(wF_1 + (1-w)R_1 - P_0) - \frac{(\theta + \epsilon_1 - wF_1 - (1-w)R_1)^2}{2\sigma_1^2}} \right] \quad (28)$$

Upon substituting (27) and collecting terms, equation (28) can be rewritten as

$$\begin{aligned} \text{Max}_{\alpha_0} - \mathbf{E}_0 \left[ e^{-(m_0 + m_1\epsilon_1 + m_2\epsilon_1^2)} \right] \\ \text{or equivalently, } \text{Max}_{\alpha_0} - e^{-(m_0 - \frac{m_1^2}{4m_2})} \mathbf{E}_0 \left[ e^{-m_2\left(\epsilon_1 + \frac{m_1}{2m_2}\right)^2} \right] \end{aligned} \quad (29)$$

where  $m_0, m_1$  and  $m_2$  are deterministic functions of model parameters, price  $P_0$ , reference price  $R_1 = \nu P_0 + (1-\nu)R_0$  (which is known at date 0 given  $R_0$  and exogenous  $\nu$ ), and shares invested in the stock  $\alpha_0$ , as follows:

$$\begin{aligned} m_0(\alpha_0) &= \gamma h_1 \alpha_0 + \frac{h_2^2}{2\sigma^2} \\ m_1(\alpha_0) &= \frac{h_2(1-w)}{\sigma^2} + \gamma w \alpha_0 \\ m_2 &= \frac{(1-w)^2}{2\sigma^2} \end{aligned}$$

where  $h_1$  and  $h_2$  are functions of  $P_0$

$$\begin{aligned} h_1 &= w(\theta - \gamma\sigma^2) + (1 - w)R_1 - P_0 \\ h_2 &= (1 - w)(\theta - R_1) + w\gamma\sigma^2 \end{aligned}$$

Denote  $y = \epsilon_1 + \frac{m_1}{2m_2}$ , then  $y \sim N(\frac{m_1}{2m_2}, \sigma^2)$ . Note that the probability density function of the normal random variable  $y$  is  $e^{-\frac{1}{2\sigma^2}(y - \frac{m_1}{2m_2})^2}$ . This is of the same functional form as  $e^{-m_2y^2}$ , and greatly simplifies the computation of the expectation  $\mathbb{E}[e^{-m_2y^2}]$ . Note that the expression to be maximized here involves only the expectation of an exponential of a normal random variable  $\epsilon_1$  and its square.

We now make use of moment generating functions of normal and  $\chi^2$  random variables to derive the solution. By completing squares and using the fact that the integral of a probability density function is 1, we get

$$\mathbb{E}[e^{-m_2y^2}] = \frac{1}{\sqrt{1 + 2m_2\sigma^2}} e^{-\frac{m_1^2}{4m_2(1 + 2m_2\sigma^2)}}$$

Substituting this into (29), type  $r$  investors solve the following problem:

$$\begin{aligned} & \text{Max}_{\alpha_0} - \frac{1}{\sqrt{1 + 2m_2\sigma^2}} e^{-(m_0 - \frac{m_1^2}{4m_2})} e^{-\frac{m_1^2}{4m_2(1 + 2m_2\sigma^2)}} \\ \text{or equivalently,} & \quad \text{Max}_{\alpha_0} (m_0 - \frac{m_1^2}{4m_2}) + \frac{m_1^2}{4m_2(1 + 2m_2\sigma^2)} \\ \text{or equivalently,} & \quad \text{Max}_{\alpha_0} \gamma b_1 \alpha_0 + \frac{h_2^2}{2\sigma_1^2} - \frac{m_1^2\sigma^2}{2(1 + 2m_2\sigma^2)} \end{aligned}$$

Since  $m_0$  and  $m_1$  are linear in  $\alpha_0$ , the function being optimized above is quadratic in  $\alpha_0$ . The coefficient for the  $\alpha_0^2$  term is negative, and hence a maximum exists. The first order condition implies that the optimal choice  $\alpha_0^* = \alpha_0^*(P_0)$  satisfies

$$(\gamma w^2 \sigma^2) \alpha_0^*(P_0) = h_1(P_0) + (1 - w)(1 - w - h_2(P_0)w)$$

For a given price  $P_0$ , the type- $r$  investor's optimal demand at date 0 is

$$\alpha_0^*(P_0) = \frac{1}{\gamma w^2 \sigma^2} (h_1(P_0) + (1 - w)(1 - w - h_2(P_0)w)) \quad (30)$$

By definition,  $F_0$  satisfies

$$\begin{aligned} \gamma w^2 \sigma^2 &= h_1(F_0) + (1 - w)(1 - w - h_2(F_0)w) \\ \text{or} \quad 1 &= \frac{1}{\gamma w^2 \sigma^2} (h_1(F_0) + (1 - w)(1 - w - h_2(F_0)w)) \end{aligned} \quad (31)$$

Subtracting (31) from (30),

$$\alpha_0^*(P_0) - 1 = \frac{1}{\gamma w^2 \sigma^2} ((h_1(P_0) - h_1(F_0)) + (1 - w)w(h_2(F_0) - h_2(P_0))) \quad (32)$$

Upon substituting  $h_1, h_2$ , and using  $R_1 = \nu P_0 + (1 - \nu)R_0$ ,

$$\begin{aligned} h_1(P_0) - h_1(F_0) &= (1 - \nu(1 - w))(F_0 - P_0) \\ h_2(F_0) - h_2(P_0) &= (1 - w)\nu(P_0 - F_0) \end{aligned}$$

Plugging these back into (32),

$$\alpha_0^*(P_0) - 1 = \frac{1}{\gamma w^2 \sigma^2} ((1 - (1 - w)\nu) - (1 - w)^2 w \nu) (F_0 - P_0) \quad (33)$$

implying

$$b_0 = \frac{1}{\gamma w^2 \sigma^2} (1 - (1 - w)\nu - (1 - w)^2 w \nu) > 0$$

Also note that when  $P_0 = F_0$ , both sides of (33) are 0, since by definition,  $\alpha_0^*(F_0) - 1 = 0$ .

## REFERENCES

- [1] Ang, Andrew, Geert Bekaert, and Jun Liu, 2001, "Why stocks may disappoint," Working Paper, Columbia University and UCLA.
- [2] Barberis, Nicholas, Andrei Shleifer and Robert Vishny, 1998, "A model of investor sentiment," *Journal of Financial Economics* 49, 307-343
- [3] Barberis, N., Ming Huang and Tano Santos, 2001, "Prospect theory and asset prices," *Quarterly Journal of Economics*, February, 2001
- [4] Barberis, Nicholas, Ming Huang, 2001, "Mental accounting, loss aversion, and individual stock returns," *Journal of Finance*, 56, 1247-1292
- [5] Berkelaar, Arjan, and Roy Kouwenberg, 2000, "Optimal portfolio choice under loss aversion," Econometric Institute Report 2000-08, Erasmus University Rotterdam, The Netherlands
- [6] Camerer, Colin F., and Martin Weber, 1998, "The disposition effect in securities trading: an experimental analysis," *J. Economic Behavior And Organization*, Vol. 33 (2) pp. 167-184.
- [7] Case, Karl, and Robert Shiller, 1988, "The behavior of home buyers in boom and post-boom markets," *New England Economic Review*, November/December, 29-46.
- [8] Chan, Louis K., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, "Momentum strategies," *Journal of Finance* 51, 1681-1713.
- [9] Conrad, Jennifer S., and Gautam Kaul, 1998, "An anatomy of trading strategies," *Review of Financial Studies*, 11, 489-519.
- [10] Coval, Joshua, and Tyler Shumway, 2001, "Do behavioral biases affect prices," Working Paper, University of Michigan, May 2001.
- [11] Daniel, Kent, David Hirshleifer and Avanidhar Subrahmanyam, 1998, "Investor psychology and security market under- and overreactions," *Journal of Finance* 53, 1839-1886
- [12] DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert Waldmann, 1990a, "Noise trader risk in financial markets," *Journal of Political Economy* 98, 703-738
- [13] DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert Waldmann, 1990b, "Positive feedback investment strategies and destabilizing rational speculation," *Journal of Finance* 45, 379-395

- [14] Fama, Eugene F., and James MacBeth, 1973, "Risk, return, and equilibrium: empirical tests," *Journal of Political Economy*, 81, 607-636.
- [15] Fama, Eugene F., and Kenneth R. French, 1996, "Multifactor explanations of asset pricing anomalies," *Journal of Finance* 51, 55-84
- [16] Ferris, Stephen P., Robert A. Haugen and Anil K. Makhija, 1988, "Predicting contemporary volume with historic volume at differential price levels: evidence supporting the disposition effect," *Journal of Finance* 43, 677-697
- [17] Genesove, David, and Christopher Mayer, "Loss aversion and seller behavior: evidence from the housing market," NBER Working Paper No. W8143, March 2001
- [18] Gomes, Francisco, 2000, "Loss aversion and the demand for risky assets," job market paper, Harvard University, Jan 2000
- [19] Grinblatt, Mark, and Tobias Moskowitz, 2001 "What we really know about the cross-sectional relation between past and expected returns," Working Paper, UCLA and University of Chicago, August.
- [20] Grinblatt, Mark, and Matti Keloharju, 2001, "What makes investor trade?," *Journal of Finance*, 56, No. 2, 589-616
- [21] Grundy, Bruce D., and Spencer J. Martin, 2001, "Understanding the nature of risks and source of rewards to momentum investing," *Review of Financial Studies*, Vol. 14, No. 1, pp29 – 78.
- [22] Heath, Chip, Steven Huddart, and Mark Lang, 1998, "Psychological factors and stock option exercise," *Quarterly Journal of Economics* 114, No.2, 601-628
- [23] Heilmann, Klaus, Volker Laeger, and Andreas Oehler, 2000, "Evidence about the investors aversion to realize losses: A contribution to behavioral finance through the use of experimental call markets," working paper, University of Bamberg.
- [24] Hirshleifer, David, 2001, "Investor psychology and asset Pricing," *Journal of Finance*, 56, 1533-1597.
- [25] Hong, Harrison, and Jeremy Stein, 2000, "A unified theory of underreaction, momentum trading and overreaction in asset markets," *Journal of Finance* 54, 2143 – 2184.
- [26] Jegadeesh, Narasimhan, 1990, "Evidence of predictable behavior of security returns," *Journal of Finance*, 45, 881-898.
- [27] Jegadeesh, Narasimhan, and Sheridan Titman, 1993, "Returns to buying winners and selling losers: implications for stock market efficiency," *Journal of Finance* 48, 65 – 91

- [28] Jegadeesh, Narasimhan, and Sheridan Titman, 2000, "Profitability of momentum strategies: An evaluation of alternative explanations," *Journal of Finance*
- [29] Daniel Kahneman, and Amos Tversky, 1979, "Prospect theory: An analysis of decision under risk," *Econometrica*, 47, 263-291.
- [30] Kaustia, Markku, "Can loss aversion move the market? Evidence from IPO trading volume," Working Paper, Helsinki School of Economics and Business Administration
- [31] Lee, Charles M.C. and Bhaskaran Swaminathan, 2000, "Price momentum and trading volume," *Journal of Finance* 55, No.5, 2017-2069.
- [32] Lo, Andrew, and Jiang Wang, 2000a, "Trading volume: definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 13, 257-300.
- [33] Locke, Peter R., and Steven C. Mann, 1999, "Do Professional Traders Exhibit Loss Realization Aversion," manuscript, Division of Economic Analysis, Commodity Futures Trading Commission and Neeley School of Business
- [34] Moskowitz, Tobias, and Mark Grinblatt 1999, "Do industries explain momentum?," *Journal of Finance*, 54, No. 4, 1249-1290.
- [35] Odean, Terrance, 1998, "Are Investors Reluctant to Realize Their Losses?," *Journal of Finance* 53, 1775-1798
- [36] Rouwenhorst, K. Geert, 1998, "International momentum strategies," *Journal of Finance*, 53, 267-284
- [37] Shapira, Zur, and Itzhak Venezia, 2001, "Patterns of behavior of professionally managed and independent investors," *Journal of Banking and Finance*, 25 (2001), 1573-1587
- [38] Shiller, Robert, 1981, "Do Stock Prices Move too Much to be Justified by Subsequent Changes in Dividends," *American Economic Review*, 71, no.3, 421-436.
- [39] Shleifer, Andrei, 2000, "Inefficient markets: an introduction to behavioral finance," *Clarendon Lectures in Economics*, Oxford University Press, 2000.
- [40] Shefrin, Hersh, and Meir Statman, 1985, "The disposition to sell winners too early and ride losers too long: Theory and evidence," *Journal of Finance* 40 777-90
- [41] Thaler, Richard H., and Eric J. Johnson, "Gambling with house money and trying to break even: the effects of prior outcomes on risky choice," *Management Science*, 36 (1990), 643-660.