THE JOURNAL OF FINANCE • VOL. LVII, NO. 2 • APRIL 2002

Rational Momentum Effects

TIMOTHY C. JOHNSON*

ABSTRACT

Momentum effects in stock returns need not imply investor irrationality, heterogeneous information, or market frictions. A simple, single-firm model with a standard pricing kernel can produce such effects when expected dividend growth rates vary over time. An enhanced model, under which persistent growth rate shocks occur episodically, can match many of the features documented by the empirical research. The same basic mechanism could potentially account for underreaction anomalies in general.

THERE WOULD APPEAR TO BE FEW more flagrant affronts to the idea of rational, efficient markets than the existence of large excess returns to simple momentum strategies in the stock market. So naturally do these profits suggest systematic underreaction by the market, and so unpromising seems the attempt to associate the rewards with risk factors, that asset pricing theorists have mostly seen the task as simply one of deciding which sort of investor irrationality is at work.[1]

This article suggests that the case for rational momentum effects is not hopeless, however. In fact, a simple, standard model of firm cash-flows discounted by an ordinary pricing kernel can deliver a strong positive correlation between past realized returns and current expected returns. The framework is simplified and ignores many features crucial for valuing real firms. The point is just to call attention to a direct, plausible, and rational mechanism that may contribute to this puzzling phenomenon.

The key to the model is stochastic expected growth rates. By their nature, such growth rates affect returns in a highly nonlinear way, and the dynamics they imply differ qualitatively from those of familiar linear factor models.

Specifically, the curvature with respect to growth rates of equity prices is extreme: Their log is convex. This property means that growth rate risk rises with growth rates. Assuming that exposure to this risk carries a positive price, expected returns then rise with growth rates. Other things equal, firms that have recently had large positive price moves are more likely to have had positive growth rate shocks than other firms, with negative growth shocks more likely among poor performers. Hence, a momentum sort will

* London Business School. I am indebted to the referee for several helpful comments.

[1] The original momentum findings are in Jegadeesh (1990) and Lehmann (1990). Behavioral explanations appear in Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hong and Stein (1999).

tend to sort firms by recent growth rate changes. In the absence of information about starting growth rates, sorting by changes will thus also tend to sort according to growth rate levels, and hence by end-of-period expected return.

When it comes to mimicking actual empirical results, the basic model runs into some problems. Most noticeably, to achieve large effects, growth rate shocks must decay quite slowly. But this persistence implies risk premia—and the associated risks—will also be persistent. By contrast, excess returns to portfolios formed according to momentum vanish for holding periods beyond one year. Moreover, volatility differences between high and low momentum portfolios are not large in postformation periods, suggesting that risk changes too are transitory.

I address these and other shortcomings of the original model with a natural extension allowing shocks to growth rates to be episodic. More precisely, I envision a two-regime process in which persistent growth shocks occur only in the more infrequent, short-lived state. This introduces a characteristic time scale beyond which effects will be undetectable. The switching model can also explain the curious fact that neither short nor long portfolio formation periods capture changes in subsequent expected returns.

While the enhanced model sacrifices the tractability of the original (and no closed-form results are available), its premise is not artificial. The intuition is simply that persistent growth rate shocks represent major changes in business conditions, like those associated with fundamental technological innovation. Such innovations do tend to be rare and episodic. Moreover, technological shocks are likely to be common within sectors, which might account for the industry component of momentum profits reported in Moskowitz and Grinblatt (1999).

I do not, however, take the analysis to the multifirm level. Nor are general equilibrium effects considered. Furthermore, no strong claim is made as to the robustness of the results. The aim is merely to show that momentum effects are not intrinsically at odds with rational behavior.[2]

The paper contributes to the effort to understand the cross section of expected returns in terms of the time-varying risk characteristics of individual firms. The role of changing capital structure (leverage effects) in altering expected returns on equity was recognized as early as Merton (1974). However, this line did not prove particularly fruitful in accounting for asset pricing anomalies. Recent work by Berk, Green, and Naik (1999) demonstrates that a rich variety of return patterns, including momentum effects, can result from the variation of exposures over the life-cycle of a firm's endogenously chosen projects. I complement this line of research by pointing out a more direct channel from cash flows to momentum in returns. The model is

[2] Conrad and Kaul (1998) also suggest a nonbehavioral mechanism, namely, that momentum sorts simply select stocks according to their unconditional expected returns. This does not seem to work empirically, however, mainly because stocks selected this way do not realize persistently different returns. See Grundy and Martin (2001) and Jegadeesh and Titman (2001).

sufficiently well developed that it offers precise quantitative predictions that can be compared in detail to the empirical evidence.

The outline of the paper is as follows. Section I presents the basic setup. Theoretical results on the existence of momentum effects are established, and numerical examples presented. Section II develops the regime-switching extension and illustrates its consequences. Simulations are used to demonstrate the ability of the model to produce realistic effects. The final section summarizes the project and highlights some areas for empirical investigation.

I. The Model

The setting used throughout the paper is a standard, continuous-time economy, with full rationality and complete information. The assumptions are as follows:

- The economy is characterized by a state-price density process Λ_t, which evolves as a geometric Brownian motion

$$\frac{d\Lambda_t}{\Lambda_t} = -r\,dt + \sigma_\Lambda\,dW_t^{(\Lambda)}, \tag{1}$$

  where r and σ_Λ are fixed constants. This assumption is tantamount to stipulating that assets are priced by a representative agent for whom Λ_t is the marginal utility process.
- The equity is an unlevered claim to a perpetual, nonnegative cash-flow process D_t with a random, stationary growth rate.

$$\frac{dD_t}{D_t} = \mu_t\,dt + \sigma_D\,dW_t^{(D)} \tag{2}$$

$$d\mu_t = \kappa(\bar{\mu} - \mu_t)dt + s\,dW_t^{(\mu)} \tag{3}$$

  Here σ_D, κ, $\bar{\mu}$, and s are constant, as are the three correlations between the Brownian motions, denoted $\rho_{\Lambda D}$, $\rho_{\Lambda\mu}$, and $\rho_{D\mu}$. Note that positive covariation with Λ is desirable in a security for offsetting fluctuations in marginal utility. The market price of D and μ risk are then $-\rho_{\Lambda D}\sigma_\Lambda$ and $-\rho_{\Lambda\mu}\sigma_\Lambda$, respectively.

Pricing the equity claim is straightforward under these assumptions. The solution was recently derived by Brennan and Xia (2001), who give the results summarized in the following proposition.

Proposition 1: *(Brennan and Xia, (2001)) Let $P = P(D,\mu)$ be the price of a claim to the dividend stream D. Then,* *(a) A necessary and sufficient condition for P to be finite is*

$$\zeta_1 \equiv \bar{\mu} - r + (\sigma_D \rho_{D\mu} + \sigma_\Lambda \rho_{\Lambda\mu}) s/\kappa + \sigma_\Lambda \sigma_D \rho_{\Lambda D} + s^2/(2\kappa^2) < 0.$$

(b) If $\zeta_1 < 0$, then $P(D,\mu) = D_t \cdot U(\mu)$ and

$$U(\mu) = e^{((\mu/\kappa)-z_0)} \int_0^\infty e^{\{\zeta_1 y + (z_2 - (\mu/\kappa)) e^{-\kappa y} + \zeta_3 e^{-2\kappa y}\}} dy$$

where

$$z_0 = \frac{\mu^\star}{\kappa} - 3\zeta_3$$

$$z_2 = \frac{\mu^\star}{\kappa} - 4\zeta_3$$

$$\zeta_3 = -\frac{s^2}{4\kappa^3}$$

$$\mu^\star = \bar{\mu} + (\sigma_D \rho_{D\mu} + \sigma_\Lambda \rho_{\Lambda\mu}) s/\kappa.$$

To study momentum, the two key processes are the cumulative excess returns accruing to the holder of a unit investment in the stock (from some specified starting date), and the instantaneous expected excess returns, which is just the drift of the first quantity. I label these processes CER_t and EER_t. The latter is given by the risk premia associated with the exposures of the equity, which is given by Itô's lemma. Hence,

$$\mathrm{EER}_t = -\rho_{\Lambda D} \sigma_\Lambda \sigma_D - \rho_{\Lambda\mu} \sigma_\Lambda \frac{U'}{U} s.$$

Then, the dynamics of the cumulative excess return process are

$$\begin{aligned} d\mathrm{CER}_t &\equiv \frac{dP_t}{P_t} - r\,dt + \frac{D}{P}\,dt \\ &= \mathrm{EER}_t\,dt + \sigma_D\,dW_t^{(D)} + \frac{U'}{U} s\,dW_t^{(\mu)}. \end{aligned}$$

Clearly the process $U'/U = U'(\mu_t)/U(\mu_t)$, the derivative of the log price–dividend ratio, is central to the evolution of the system. While its explicit form is not very revealing, the characteristic behavior of the system may be seen with the help of the following lemma.

LEMMA 1: *Let $U(x)$ be as defined in (b) of Proposition 1 and assume the condition in (a) is satisfied.*[3] *Then, for all x, $U'(x)/U(x)$ is a positive, increasing function.*

Note: *All proofs are given in the appendix.*

The lemma establishes the property mentioned in the introduction: that growth rate risk ($1/P \cdot \partial P/\partial\mu \propto U'/U$) rises with growth rates, regardless of the values of the parameters chosen. Mathematically, this means that the sensitivity of the pricing function to this state variable is stronger than exponential. Economically, such extreme sensitivity can lead to purely rational price paths that display bubblelike characteristics. For that reason, this class of models deserves careful scrutiny by those inclined to interpret such behavior as evidence of expectational cascades, irrationality, or chaos. Nothing like that needs to be involved.

As described above, if growth rate risk has a positive price, then higher growth rates must entail higher expected returns. And momentum effects then follow because positive (negative) cumulative returns typically imply ex post that recent growth rate shocks have been positive (negative). To verify the intuition of this simple conditioning argument, fix a time horizon, ℓ, and consider how total excess returns from t (today) to ℓ will covary with the expected excess return after ℓ.

PROPOSITION 2: *Let $\mathcal{F}_t$ be the time t information set. Then, assuming $\rho_{D\mu} \geq 0$ and $\rho_{\Lambda\mu} < 0$,*

$$\mathrm{E}[(\mathrm{CER}_{t+\ell} - \mathrm{E}[\mathrm{CER}_{t+\ell}|\mathcal{F}_t]) \cdot (\mathrm{EER}_{t+\ell} - \mathrm{E}[\mathrm{EER}_{t+\ell}|\mathcal{F}_t])|\mathcal{F}_t] > 0.$$

The conclusion of the proposition just tells us that, given large returns to ℓ, we would indeed expect to see larger subsequent returns. The two correlation restrictions are sufficient but far from necessary, as can be seen from the proof. The requirement $\rho_{D\mu} \geq 0$ ensures that growth rate increases are unambiguously "good news" and will, in general, coincide with increasing stock prices. But a negative correlation does not rule this out if the sensitivity to growth rates outweighs the sensitivity to dividends, which occurs for many natural parameter choices. On the other hand, the requirement $\rho_{\Lambda\mu} < 0$, which ensures that growth rate risk has a positive price, is harder to relax.[4] Still, it can be the case that a countercyclical firm (whose growth rate tends to increase in recessions, say) still exhibits momentum if *also* $\rho_{D\mu} < 0$.

The exact covariance function whose sign the proposition gives is a function of the starting growth rate, μ_t, and is not available in closed form. However, it may be found by integrating forward the expected instantaneous

[3] The latter will be assumed implicitly in the remainder of the section.

[4] Positive covariance of consumption with growth rates (hence negative covariance of μ with marginal utility) holding dividends fixed is a standard result for an endowment economy with an elastically supplied riskless storage technology.

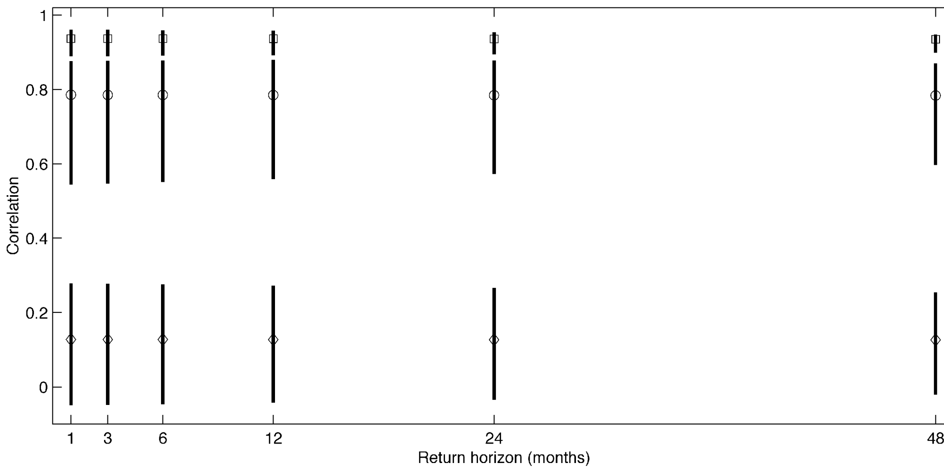

Figure 1. Correlations: past cumulative return and current expected return. The figure shows the correlation of realized returns over different horizons with the instantaneous expected excess return at the end of that period. Three cases are shown. The top case (squares) has $\rho_{\Lambda\mu} = -0.5$, $\rho_{\Lambda D} = -0.5$, $\rho_{D\mu} = +0.5$, and $\sigma_D = 0.2$. The middle case (circles) has $\rho_{\Lambda\mu} = -0.5$, $\rho_{\Lambda D} = -0.5$, $\rho_{D\mu} = -0.5$, and $\sigma_D = 0.1$. The bottom case (diamonds) has $\rho_{\Lambda\mu} = +0.5$, $\rho_{\Lambda D} = -0.5$, $\rho_{D\mu} = -0.5$, and $\sigma_D = 0.4$. All cases use $r = 0.05$, $\sigma_\Lambda = 0.40$, $s = 0.03$, $\kappa = 0.09$. The vertical bars depict the spread implied by letting the time-zero growth rate range over ± 3 times its stationary standard deviation.

covariances from t to $t + \ell$. These expectations can be readily calculated from the Kolmogorov forward equations, since their time t values are known for all values of μ_t.

Slightly better, we may standardize these covariances and turn them into the correlation function,

$$\Gamma(\ell;\mu_t) = \frac{\displaystyle\int_t^{t+\ell} \operatorname{cov}_t(\mathrm{CER}_{u+\ell}, \mathrm{EER}_{u+\ell})\,du}{\left(\displaystyle\int_t^{t+\ell} \operatorname{var}_t(\mathrm{CER}_{u+\ell})\,du\right)^{1/2} \cdot \left(\displaystyle\int_t^{t+\ell} \operatorname{var}_t(\mathrm{EER}_{u+\ell})\,du\right)^{1/2}},$$

and compute all the moments in the same manner.

Figure 1 plots $\Gamma(\ell)$ for time horizons out to four years for some different parameter configurations. (The vertical bars in the figure delimit the correlations for values of the starting growth rate of ± 3 times its stationary standard deviation.) The highest values correspond to a "normal" firm, which satisfies the conditions of the proposition. Here the correlation is nearly one for all time horizons. The values below these correspond to the same parameter configuration, but with $\rho_{D\mu} = -0.5$, instead of $+0.5$. These correlations too are large and positive, illustrating the secondary importance of that parameter. Finally, the bottom values correspond to a "countercyclical" firm, as described above. These values again are positive, though small, except for very low values of μ_0. Apparently momentum effects, as measured by this statistic, are a robust occurrence. In fact, it is not easy to generate antimomentum configurations that are at all realistic and also large, without violating the regularity condition (a) in Proposition 1.

For comparison with empirical work, and to gauge the magnitude of the theoretical effect, one would like to know the exact relationship between a given observed return and the subsequent expected return, as in

$$\Theta(\iota,\ell) \equiv \mathrm{E}(\mathrm{EER}_\ell | \mathrm{CER}_\ell = \iota)$$

(where the initial time is now being taken to be zero). To be clear, the conditioning information here is *only* the realized excess return, and not the subsequent path of the growth rate.[5] Were μ_ℓ known, EER_ℓ would be determined. While the model envisions μ being observable, it is not readily available to econometricians. So, in the empirical literature, momentum is nearly always analyzed by tabulating subsequent returns for portfolios formed on the basis of cumulative returns from 0 to ℓ. More generally, those subsequent returns can be measured at varying horizons τ. Then, in terms of the model, one would want to compare these to the theoretical function of three parameters:

$$\Phi(\iota,\ell,\tau) \equiv \mathrm{E}\left(\int_\ell^{\ell+\tau} \mathrm{EER}_u \, du | \mathrm{CER}_\ell = \iota\right) = \mathrm{E}(\mathrm{CER}_{\ell+\tau} - \mathrm{CER}_\ell | \mathrm{CER}_\ell = \iota).$$

The functions Θ and Φ, though not available analytically, can be computed by Monte Carlo techniques. In Table I, I present these for some chosen parameter configurations.

First, the table shows the expected excess return following a one-year period, conditioned on 10 possible return intervals. The intervals were chosen to match typical intervals used by studies in forming portfolios. Specifically, I used the average of the performance decile breakpoints for NYSE-listed stocks from 1977 to 1992.[6] For example, the range labeled I1 would, on average, have put the stock in the bottom 10 percent of all firms. (The exact breakpoints are given in the table caption.) The second panel shows the parameters used in the different cases, along with some statistics describing the stock price process they imply.[7]

[5] Formally, the right side is the derivative of the regular conditional measure $\mathrm{E}(\mathrm{EER}_\ell | \mathrm{CER}_\ell < \iota)$. If the initial growth rate is also taken as a parameter, $\Theta(i,\mu_0;\ell)$ may be defined likewise.

[6] This is the period used in Chan, Jegadeesh, and Lakonishok (1996). They report average six-month returns by decile. I fit a normal distribution to these, scaled that by the square root of my formation period (one year in the table), and calculated decile breaks from that.

[7] There is no claim to generality here. The cases were chosen to show the potential of the model.

Table I

Theoretical Momentum Effects

The first panel shows the instantaneous expected excess return (continuously compounded annualized percentage) under four different sets of parameters, subsequent to a year in which the cumulative return has fallen into 1 of the 10 intervals labeled I1 to I10. The return intervals are defined by the breakpoints (−19.53, −7.69, 0.58, 7.70, 14.30, 20.90, 28.03, 36.31, 47.85). I1 corresponds to returns below −19.53 percent, I2 to returns between −19.53 percent and −7.69 percent, and so on up to returns over 47.85 percent in I10. The values are calculated by Monte Carlo simulation of the model of Section I. The second panel lists the parameter settings for the cases. All cases put $r = 0.05$, $\sigma_\Lambda = 0.40$, $\bar{\mu} = \mu_0 = 0.00$. The last three columns show the initial risk premium, volatility, and price–dividend ratio for the stock that are implied by the settings. The risk premium and volatility are annualized percentages.

Panel A: Expected Return as a Function of Realized Return

Case	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
A	13.83	13.86	13.86	13.86	13.87	13.87	13.87	13.88	13.89	13.90
B	10.79	11.30	11.43	11.52	11.60	11.68	11.77	11.83	11.93	12.26
C	9.23	9.98	10.28	10.49	10.72	10.93	11.12	11.35	11.63	12.39
D	9.87	11.06	11.64	12.05	12.37	12.77	13.12	13.48	14.01	15.46
E	9.54	10.83	11.38	11.74	12.16	12.55	12.85	13.27	13.79	15.38
F	8.60	9.78	10.48	11.00	11.56	12.09	12.63	13.21	14.02	16.14

Panel B: Parameter Settings

	σ_D	s	κ	$\rho_{D\mu}$	$\rho_{\Lambda D}$	$\rho_{\Lambda\mu}$		EER_0	VOL_0	U_0
A	0.10	0.60	1.00	0.00	−0.50	−0.50		13.9	60.1	86.2
B	0.10	0.06	0.10	0.00	−0.50	−0.50		11.7	49.8	28.8
C	0.10	0.04	0.06	0.00	−0.60	−0.60		10.8	36.4	14.0
D	0.08	0.03	0.04	0.00	−0.70	−0.70		12.6	37.8	14.5
E	0.10	0.03	0.04	0.20	−0.20	−0.80		12.3	39.3	15.4
F	0.03	0.035	0.04	0.10	−0.40	−0.95		11.8	30.2	11.1

For these plausible cases, a strong and monotonic relationship between past return and future expected return is shown clearly in the table, with the magnitudes (given in annualized percentage) being economically large. The empirical effect, as measured by the average difference between post-formation returns of top and bottom intervals, is larger still: typically around 8–12 percent per year for six-month holding periods (Jegadeesh and Titman (1993), Rouwenhorst (1998)) in the postwar period. Matching this, while obeying the integrability condition of Proposition 1, appears to be unachievable. But when μ shocks are sufficiently persistent and growth rates are highly correlated with marginal utility, over half of this can be accounted for.

Some care is required in making comparisons between the table, which shows the expected return for a single firm conditioned on its own performance, and the results of portfolio studies. For one thing, the conditioning information embodied in a performance sort is about *relative* returns. Being in the 10th decile literally means doing less well than 90 percent of other stocks, not having return below (say) −15 percent.

Capturing that relative condition would entail modeling the full covariance structure of returns. Whether or not this would lead to smaller predicted momentum effects, however, is ambiguous. Intuitively, in up markets, using relative performance breakpoints will lead to higher beta stocks in the top decile than using absolute performance levels, and lower beta stocks in the bottom decile, whereas in down markets, the reverse would be true. (This argument is formalized in Grundy and Martin (2001).) Assuming market risk is correlated (negatively) with the state–price density, this implies that using fixed breakpoints understates the magnitude of predicted momentum effects in up markets and overstates it in down markets. To the extent that the U.S. postwar market can be regarded as having experienced positive excess returns, then, my reported effects are conservative.

A second caution in interpreting the results is that the single-firm numbers in the table do not reflect the information about the firm parameters (σ_D, s, and κ) that a performance sort captures. For example, the extreme decile portfolios are more likely to be composed of firms whose unconditional volatility is higher, meaning bigger σ_D and s and smaller κ. Likewise, since time zero volatility is an increasing function of growth rate, higher μ_0 and $\bar{\mu}$ may be more likely for the extreme performers. Here the effect is ambiguous for the poor performers, though, because higher initial μ also implies higher drift for the stock price.

Fully analyzing the effects of this additional parameter information would require specification of a prior distribution over the possible configurations (as well as the covariance structure again). Such an effort is beyond the scope of this work. But the general effect is likely to work against the monotonicity exhibited in Table I. The low performers would be unconditionally more risky, leading to something of a U-shape in expected returns.

One calculation that can be readily performed to illustrate the point is to integrate out the dependence on μ_0. Since the growth rate follows a stationary Ornstein–Uhlenbeck process, it has a steady-state distribution (normal with mean $\bar{\mu}$ and variance $s^2/2\kappa$), which is a natural candidate for the unconditional distribution of μ_0. Table II shows the effect on the cases in Table I of integrating over this distribution. Now, for all the configurations, low realized returns imply higher expected returns than in the previous table, because a high initial μ is more likely. The volatility effect outweighs the drift effect. Overall, comparing I1 to I10, there is still a strong momentum effect. However, the empirical literature finds a monotonic relation here.[8] This presents something of a problem for the model, and suggests that the picture—at least as far as the worst performers is concerned—remains incomplete.

The model also has difficulty matching another feature of empirical studies: the dependence of the strength of the effect on both the formation period

[8] Moreover, for individual stocks, the majority of momentum profits come from the underperformance of the losers, which also contrasts with the model's prediction. This is not the case, however, for industry or country momentum portfolios (Moskowitz and Grinblatt (1999), Asness, Liew, and Stevens (1997)).

Table II

Theoretical Momentum Effects (Continued)

The table repeats the calculation of Table I taking the beginning-of-period growth rate μ_0 to be distributed according to its stationary distribution (instead of being set to its unconditional mean).

Case	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
A	13.83	13.85	13.86	13.86	13.87	13.87	13.87	13.88	13.88	13.89
B	10.72	10.77	10.88	11.05	11.16	11.25	11.35	11.48	11.64	12.22
C	10.48	9.63	9.62	9.68	9.82	10.15	10.51	10.91	11.72	13.35
D	13.59	10.83	10.23	10.21	10.39	10.72	11.88	12.88	14.15	18.02
E	12.90	10.34	10.00	10.14	10.24	10.69	11.54	12.68	13.81	18.14
F	15.09	10.94	9.77	9.88	10.35	10.69	12.15	13.76	16.05	22.71

(over which momentum is measured to select portfolios) and on the subsequent holding period. The typical patterns here are that (a) there are no extra excess returns to holding momentum portfolios much beyond a year; and (b) there are no excess returns at all when portfolios are formed on the basis of performance over periods longer than a year or shorter than a few months.

This is too much complexity to reproduce in the current set up. As indicated by Figure 1, the strength of the correlation between CER and EER does not vary much with the formation period. Neither do expected excess returns differentials decline much with holding period. They *do* decline, since μ is a mean-reverting process (and hence EER is). However, as already remarked, the decay rate κ—which is in units of inverse years—must be quite small to produce large expected return differences.

The situation is summarized graphically by Figure 2. Here, I measure the strength of the momentum effect by the differences in expected excess return between the highest and lowest performance brackets using different lengths of formation and holding periods. The top panel (which uses parameter case A from Table I) is qualitatively similar to the empirical findings. The strength of the effect is indeed maximized by using formation period of about a year. And the anticipated excess returns do decay quickly with holding period. This is achievable because the parameter κ has been set to unity here, so one year is the characteristic decay period of growth rate shocks.

Unfortunately, the effect is minuscule (the vertical axis is in annualized percentage points). Shocks that decay this quickly do not have big consequences in terms of discounted dividend streams. The bottom panel shows behavior typical of smaller values of κ. Here the characteristic decay length is 10 years. Now the decay with respect to holding period is very slow, and the effect only grows stronger as longer term returns are used to define momentum. The model simply lacks the flexibility to capture the enigmatic pattern observed in real stocks while also producing strong effects.

The lack of a rapid mean reversion of expected returns points to another weakness of the model as well. It predicts that volatility differentials across performance levels should be persistent. The model implies that poor performers have low expected returns because their future risk is low. And the

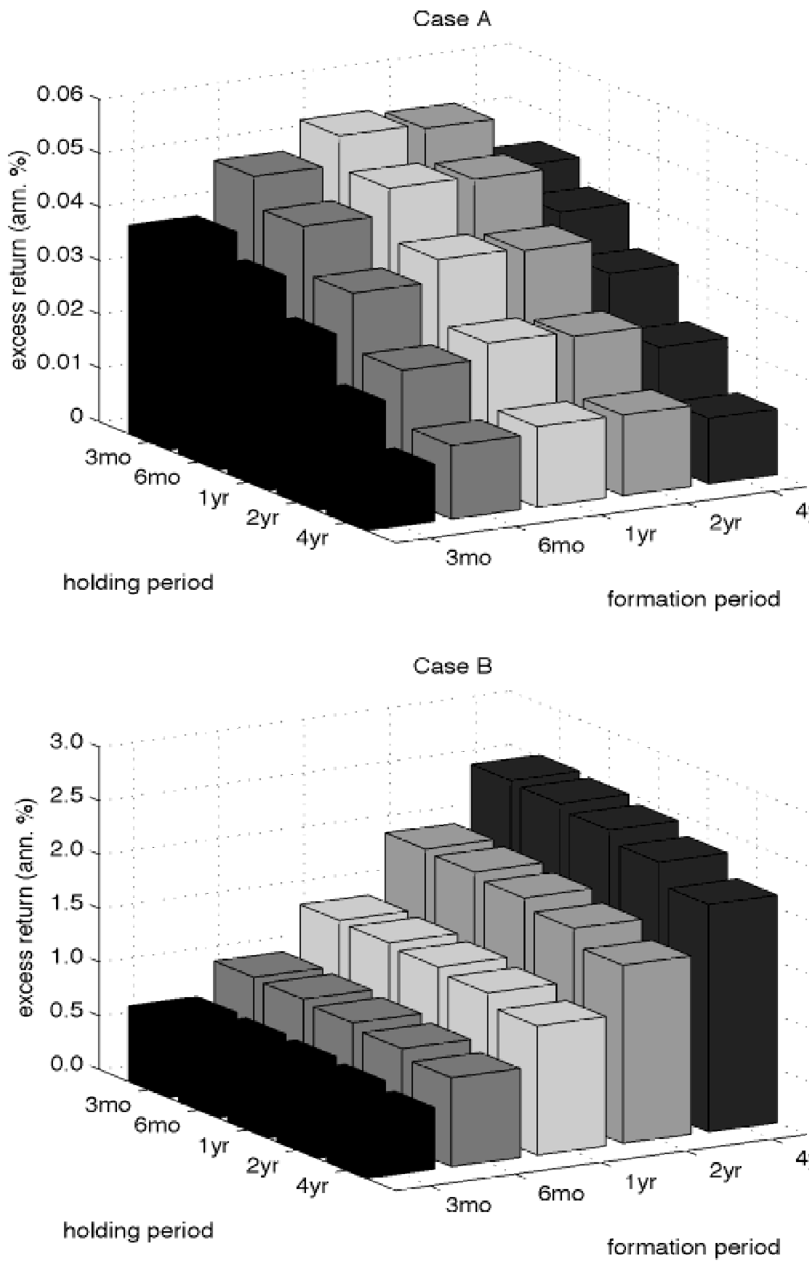

Figure 2. Momentum effect as a function of holding and formation periods. The graphs show the difference between expected excess returns conditional on high realized return and that conditional on low realized return. The return differences are plotted as a function of the expected holding period and as a function of the formation period over which realized returns are measured. High (low) realized returns are defined as returns that would be in the top (bottom) decile if returns were normally distributed with annual mean and standard deviation matching the unconditional distribution of NYSE stocks from 1977 to 1992. The two cases shown correspond to the parameter settings A and B shown in Table I.

future risk of good performers should be high. Empirical studies fail to find such differences, in either systematic or unsystematic risk, during the post-formation period. This suggests that risk changes also decay quickly.

Finally, one may mention two further counterfactual implications of the model when taken beyond the single-firm setting. First, it suggests that high expected returns should be associated with *high* price–dividend ratios, since, under the model, both things increase with μ. The model thus makes one asset pricing anomaly worse even as it addresses another. Second, since aggregate growth rates would also be stochastic and command a positive risk price, the model would seem to imply an aggregate momentum effect.[9]

There is no point in being too harsh on the model, however. Its virtue is its simplicity. Clearly, real firms are not continuous, nonnegative dividend streams. The remarkable thing is that one can generate such an apparently irrational phenomenon from such an uncomplicated depiction. Still, the next section asks whether the basic mechanism of the model, generalized somewhat, can indeed address some of the shortcomings outlined here.

II. A Generalization

There are a number of obvious ways to make the model of the last section more realistic. This section implements one that retains the basic mathematical structure and preserves the original intuition, while adding significant flexibility to the growth rate dynamics. Specifically, the nature of the innovation process itself is permitted to change intermittently. The idea is to introduce a characteristic time scale—the length of time between such structural changes—that can allow the model to match the apparent short duration of momentum-induced changes in excess returns and risks that real stocks undergo. As a side benefit, the generalization brings the model closer to the data on some other dimensions as well.

Formally, this is accomplished by augmenting the system with a two-state regime indicator variable, S, upon which the dynamics of the growth rate process may depend. Intuitively, I think of one of the regimes ($S = 1$, say) as standing for periods of fundamental technological change in which growth rate innovations are more or less permanent. The other regime ($S = 0$) would correspond to the more normal state of affairs in which there may still be growth rate shocks, lasting for a quarter, a year, or even a business cycle, but not changing the long-term fundamentals.[10]

[9] This criticism applies generally to momentum models that focus on firm level autocorrelations, as pointed out by the referee.

[10] The notion of small but persistent shocks to growth rates was suggested by Barsky and DeLong (1993) as an explanation of the apparent “excess” volatility of the stock market. Donaldson and Kamstra (1996) find evidence for such dynamics in dividend growth in the 1920s sufficient to account for both the rise and fall of stock prices in that decade ex ante. Recently, Bansal and Yaron (2000) showed that the same mechanism could potentially explain the equity premium puzzle. Johnson (2001) introduces the idea of time-varying persistence to account for predictable patterns of volatility dynamics.

In continuous time, the process S is characterized by two switching intensities, denoted λ_0 and λ_1, whose units are inverse years. So if, for instance, $\lambda_0 = 1/10$, then the expected duration of $S = 0$ episodes is 10 years. (Also, over a small time interval Δt, the probability of a switch from $S = 0$ to $S = 1$ is $\lambda_0 \Delta t$.) The ratio $\lambda_0/(\lambda_0 + \lambda_1) \equiv \bar{S}$ represents the fraction of time spent in the $S = 1$ state and is also the unconditional expected value of S. The intuition in the preceding paragraph then suggests that $\bar{S}$ is small (transient shocks are more likely) and λ_1 is large (persistent shock episodes do not last long). For simplicity, S will be taken to be independent of the other stochastic processes in the economy.

To model the changing degree of persistence between regimes, the growth-rate process will be decomposed into two component processes representing the cumulative long-term and short-term shocks. I call these x_t and y_t, respectively, and define them as follows:

$$dx_t = \kappa_1(\bar{x} - x_t)\,dt + s_1 S_t\, dW_t^{(\mu)} \tag{4}$$

$$dy_t = \kappa_0(\bar{y} - y_t)\,dt + s_0(1 - S_t)\, dW_t^{(\mu)} \tag{5}$$

$$\mu_t = x_t + y_t, \tag{6}$$

with $\kappa_1 \leq \kappa_0$. This formulation captures the idea of a given "shock" $dW_t^{(\mu)}$ possessing an intrinsic trait, coded by S_t, corresponding to the length of time it takes for its effect on μ to decay. Another helpful way to write μ is in integral form:

$$\mu_t = x_0 e^{-\kappa_1 t} + y_0 e^{-\kappa_0 t} + s_1 \int_0^t e^{-\kappa_1(t-u)} S_u\, dW_u^{(\mu)}$$

$$+ s_0 \int_0^t e^{-\kappa_0(t-u)} (1 - S_u)\, dW_u^{(\mu)}.$$

(Here for brevity I am taking the long-run values to be $\bar{x} = \bar{y} = 0$.) This shows explicitly how the effect over time of a shock experienced at time t_0 declines with t as $\exp(-\kappa_i(t - t_0))$, with κ_i fixed forever by S_{t_0}.

A more parsimonious (though unsuccessful) model for μ is also nested in this one. If $\kappa_1 = \kappa_0 = \kappa$, then we have

$$d\mu_t = \kappa(\bar{\mu} - \mu_t)\,dt + s(S_t)\,dW_t^{(\mu)},$$

where now s switches between two values according to S_t. This is just a simplified way of introducing stochastic volatility to the μ process. The corresponding case where $s_1 = s_0$ in (4) and (5) also turns out to be inadequate.

Although the model now has four stochastic state variables, it remains fairly tractable. The dynamics are summarized in the following proposition.

PROPOSITION 3: *With the processes* Λ *and* D *defined by equations (1) and (2), and with the growth rate process given by (4)–(6), the stock price is*

$$P(D,x,y,S) = D_t \cdot (u^{(0)}(x_t, y_t) \cdot (1 - S_t) + u^{(1)}(x_t, y_t) \cdot S_t), \tag{7}$$

where $u^{(0)}(\;)$ *and* $u^{(1)}(\;)$ *satisfy the coupled partial differential equations*

$$\frac{s_0^2}{2} u_{yy}^{(0)} + [\kappa_0(\bar{y} - y) + s_0(\rho_{\Lambda\mu}\sigma_\Lambda + \rho_{D\mu}\sigma_D)]u_y^{(0)}$$
$$+ [\kappa_1(\bar{x} - x)]u_x^{(0)} + [(x + y) - r + \rho_{\Lambda D}\sigma_\Lambda\sigma_D]u^{(0)} + \lambda_0(u^{(1)} - u^{(0)}) - 1 = 0$$

$$\frac{s_1^2}{2} u_{xx}^{(1)} + [\kappa_1(\bar{x} - x) + s_0(\rho_{\Lambda\mu}\sigma_\Lambda + \rho_{D\mu}\sigma_D)]u_x^{(1)}$$
$$+ [\kappa_0(\bar{y} - y)]u_y^{(1)} + [(x + y) - r + \rho_{\Lambda D}\sigma_\Lambda\sigma_D]u^{(1)} + \lambda_1(u^{(0)} - u^{(1)}) - 1 = 0.$$

The instantaneous expected excess return and volatility of P *are given by*

$$\mathrm{EER}_t = -\sigma_\Lambda\{\rho_{\Lambda D}\sigma_D + \rho_{\Lambda\mu}\Psi_t\}$$

$$\mathrm{VOL}_t = (\sigma_D^2 + 2\rho_{D\mu}\sigma_D\Psi_t + \Psi_t^2)^{1/2},$$

where

$$\Psi_t \equiv (1 - S_t)\cdot s_1 \cdot \frac{u_y^{(0)}}{u^{(0)}} + S_t \cdot s_0 \cdot \frac{u_x^{(1)}}{u^{(1)}}.$$

The covariance between expected excess returns and cumulative excess returns is

$$-\rho_{\Lambda\mu}\sigma_\Lambda\Upsilon_t\cdot(\rho_{D\mu}\sigma_D + \Psi_t)$$

with

$$\Upsilon_t \equiv (1 - S_t)\cdot s_1 \cdot \frac{\partial}{\partial y}\left(\frac{u_y^{(0)}}{u^{(0)}}\right) + S_t \cdot s_0 \cdot \frac{\partial}{\partial x}\left(\frac{u_x^{(1)}}{u^{(1)}}\right).$$

Under this model, the stock price process follows a jump diffusion. As the proof notes, having a continuous process for the pricing kernel is tantamount to having jump risk be unpriced. Thus, expected excess returns do not depend on the jump parameters. Instead, the EER process is exactly analogous to that of the model derived in the last section. Again, the key ingredient is the log derivative of the price–dividend ratio. Now, though, there are two such ratios: $u^{(0)}(x, y)$ and $u^{(1)}(x, y)$. Expected returns and volatility simply toggle between the processes implied by each of these as S switches. In particular, if momentum effects result mainly from the $S = 1$ regime, these will only last on average $1/\lambda_1$ years. If this duration is, say, 0.5 to 1, that might account for the empirically observed dissipation of the effects for longer holding periods. Moreover, it might also explain why the effects are strongest for formation periods of about this length: Large returns over a longer period might no longer imply that $S = 1$ at the end of the period; large returns over a much shorter period would simply be too noisy.

Against this potential, the model now threatens to dilute the strength of predicted effects, whereas the original model already fell short of the return differentials seen in the data. The total effect mixes that of the two possible states of S at the end of the formation period. If unconditionally the $S = 1$ state is unlikely, then so are any expected return differentials. There is hope, however, for two reasons. First, stronger effects can be induced in the $S = 1$ state of this model than could be in the original one.[11] Second, the $S = 1$ state implies higher stock volatility, making it relatively more likely than the $S = 0$ state among extreme performers.

To investigate the net result, I explore some numerical examples. The coupled differential equations defining $u^{(0)}$ and $u^{(1)}$ resist analytical solution, but may be solved with standard techniques. As a first case, I set $\lambda_0 = 1/36$ and $\lambda_1 = 1$ so that $\bar{S} = 2.7$ percent; hence persistent shocks occur on average every 36 years and last around 1 year. The persistent shocks are taken to have a decay constant of $\kappa_1 = 0.05 = 1/20$ years, which implies a half-life of about 14 years for these shocks. (The transient shocks have $\kappa_0 = 1.0$. For the other parameter choices see Table III.)

Figure 3 shows the resulting momentum effects for different holding periods and formation periods. Now the model is able to achieve quite rapid decay of expected excess returns with holding period, closely matching the rate reported by Rouwenhorst (1998) and Jegadeesh and Titman (2001). Even better, the model is able to achieve the empirically observed peak of maximum effect at a six-month formation window. Most strikingly of all, the size of the effect can exceed that of the earlier model in which all shocks are persistent. The case in the figure even approaches the magnitude found in the postwar data. This despite the fact that the firm being modeled here only experiences persistent rate shocks once every 40 years.

Looked at another way, this last finding is even more unusual. If *all* stocks had these parameter values, it would seem to suggest that the entire momentum effect at any one time could be due to the dynamics of a mere three percent of them. This could at once address several of the difficulties with the basic model mentioned in the previous section.

[11] The proposition does not provide explicit integrability conditions under which the price is guaranteed to be finite. Heuristically, the infrequency of the persistent state allows larger and more persistent shocks, while keeping the long-run growth rate below the discount rate.

Table III

Theoretical Momentum Effects

The first panel shows the instantaneous expected excess return (continuously compounded annualized percentage), using the model of Section II, for two sets of parameters, conditional upon the previous 12-month performance. That performance is partitioned into 10 intervals of cumulative return, whose intervals are defined by the breakpoints (−19.53, −7.69, 0.58, 7.70, 14.30, 20.90, 28.03, 36.31, 47.85). I1 corresponds to returns below −19.53 percent, I2 to returns between −19.53 percent and −7.69 percent, and so on up to returns over 47.85 percent in I10. The second panel lists the parameter settings for the cases. All cases put $r = 0.05$, $\bar{x} = \bar{y} = 0.00$. The initial values of the growth rate components are distributed according to their steady-state density.

Panel A: Expected Return as a Function of Realized Return

Case	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
G	9.63	5.94	5.78	5.70	5.71	5.75	5.80	5.95	6.68	17.53
K	4.26	5.41	5.47	5.90	5.91	6.04	6.71	7.85	8.01	11.12

Panel B: Parameter Settings

	σ_Λ	σ_D	$\rho_{D\mu}$	$\rho_{\Lambda D}$		λ	s	κ	$\rho_{\Lambda\mu}$
G	0.40	0.05	0.00	−0.70	$S = 0$:	0.028	0.16	1.00	−0.70
					$S = 1$:	1.00	0.08	0.05	−0.70
K	0.90	0.03	0.00	−0.30	$S = 0$:	0.20	0.45	1.00	−0.90
					$S = 1$:	1.00	0.033	0.06	0.00

First, it could help to explain the difficulty of picking up the incremental risk (measured by market beta, for example) in portfolios of winners, or its lack in losers. Second, it could mitigate the unconditional positive relation between price–dividend ratios and expected returns. Momentum sorts now work because they *both* condition on the level of μ *and* on the likelihood of being in the persistent-shock state. In contrast, a D/P sort would not achieve the second: There would be no reason to expect a concentration of $S = 1$ firms in extreme D/P deciles. Finally, aggregating across firms, the percentage in that state would not fluctuate much. So aggregate growth rates would not exhibit the regime-switching behavior that drives this model. Instead, aggregate growth shocks would be almost entirely transient, leading to little or no autocorrelation in the market as a whole.

One respect in which the model has not been improved is in matching the smooth, monotonic increase of expected return with past return. In fact, the U-shape seen in Table II is even more exaggerated. This is because, conditional on middling performance, the unvolatile, transient state is much more likely. In this state, risks are lower, so expected returns are depressed.

One way to mitigate this problem is to raise the amplitude of the transient shocks. This might be plausible: Low-frequency fluctuations in growth rates might well be of smaller magnitude than high frequency ones. Due to their rapid decay, these would still have little impact on expected returns, but

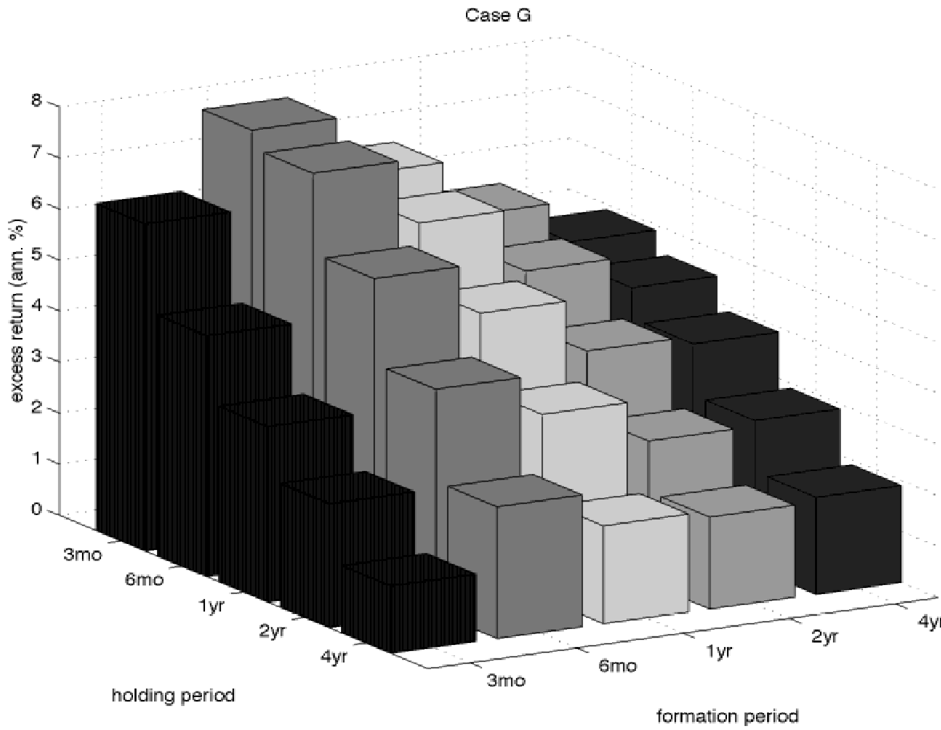

Figure 3. Momentum effect as a function of holding and formation periods. The figure shows the difference between expected excess returns conditional on high realized return and that conditional on low realized return. The return differences are plotted as a function of the expected holding period and as a function of the formation period over which realized returns are measured. High (low) realized returns are defined as returns that would be in the top (bottom) decile if returns were normally distributed with annual mean and standard deviation matching the unconditional distribution of NYSE stocks from 1977 to 1992. The figure employs the parameter settings of Case G (see Table III).

would command a higher premium. Table III shows the momentum effect by performance interval for two cases which use this assumption.

The first case (labeled G), is the one shown in Figure 3. Here, it is still the case that expected returns initially *decline* with performance. All of the momentum effect is in the highest two or three brackets. The second case is able to achieve a smoother dependence. Here, however, another degree of flexibility has been added: The transient shocks have been assumed to have a lower price of risk.[12] This does not seem unduly demanding. In fact, the idea that marginal utility would be more affected by long-run shocks than short-run ones is appealing. In terms of a consumption-based model, this would just indicate a nonmyopic policy. The effect here is to make the Sharpe

[12] That is, the two component processes x and y are allowed to have differing correlations with the pricing kernel: $|\rho_{\Lambda y}| < |\rho_{\Lambda x}|$. The modifications to Proposition 3 are straightforward.

ratio under $S = 1$ much higher than under $S = 0$. This causes the latter cases to be relatively more likely the worse the observed performance.

This section has pursued the insight of the Section I that growth-rate risks might rise with growth rates. Embedding the basic model in a more flexible and realistic one, the theory can capture the peculiar dependency of momentum profits on the length of formation and holding periods. Using plausible parameter values, the magnitude of the effect can attain roughly that in the data. The key addition is the possibility of time-varying persistence in growth rate innovations. The model implies that the effects might be entirely attributable to very infrequent, but highly persistent, shocks.

III. Conclusion

This paper advances the hypothesis that stochastic growth rates may account for some or all of the momentum anomaly. The argument works because stock prices depend on growth rates in a highly sensitive, nonlinear way. Other things equal, recent performance is correlated with levels of expected growth rate, which is monotonically related to risk. This relationship was demonstrated analytically by means of a simple partial-equilibrium model that has previously appeared in the literature. A more sophisticated version incorporating the notion of episodic, highly persistent growth rate shocks was able to achieve agreement with observation along a number of challenging dimensions.

The results raise the possibility that the same basic mechanism could play a role in *all* the anomalies that fall under the general heading of underreaction. As Mitchell and Stafford (2000) have argued, the mispricing evident in many long-horizon event studies seems to be due to common exposure of event firms to the same source of benchmark error. The model here suggests an economic rationale: Conditioning on a large stock return (the event) is like conditioning on a persistent shock to dividend growth, which *should* alter expected returns in the same direction.

Of course, investors could also systematically underreact to news. The point is not to insist that markets *are* rational, but only to point out one alternative that does not rely on the opposite assumption. Whether or not the story works empirically hinges on the answer to two questions.

First, do momentum portfolios, in fact, differ in their expected dividend (or earnings or cash-flow) growth? (For some extremely suggestive evidence based on country momentum portfolios, see Figure 4.) The model implies that past performance is essentially acting as an instrument for persistent changes in this expectation, which is readily testable.

Second, is growth rate risk priced? Perhaps the most fundamental objection to risk-based explanations of momentum (or any other cross-sectional anomaly) is that the risk part of the story seems absent in the data. Momentum strategies do not appear especially dangerous. This paper has skirted that issue by not identifying the state–price density covariance with which is the relevant measure of dangerousness. For some, the explanation will

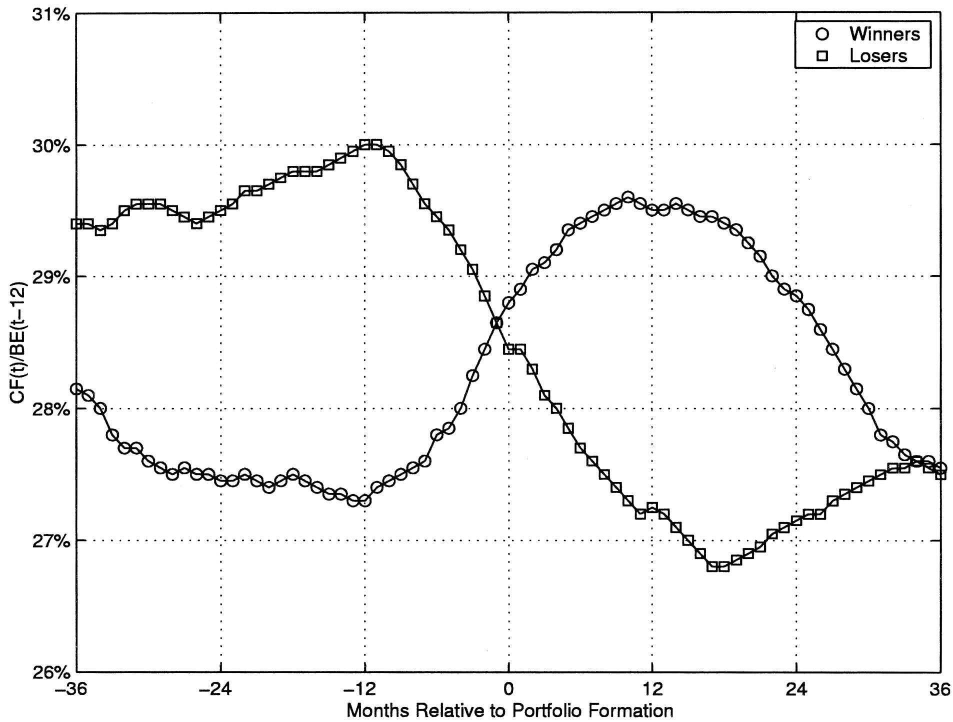

Figure 4. Cash-flows of country momentum portfolios. The figure plots the average cash-flow over book-value for portfolios formed from value-weighted country indices from 1970 through 1994. The winner and loser portfolios are, respectively, the top and bottom third of 18 countries sorted by prior 12-month return, with 1 month skipped before formation. The data are from Asness, Liew, and Stevens (1997, exhibit 7).

remain unconvincing until plausible candidates are found. (Recently Chordia and Shivakumar (2001) have uncovered evidence of systematic variation in momentum profits with certain business cycle variables.) The empirical task here is to establish (a) whether there is a systematic and persistent component to growth rate shocks at all; and (b) whether exposure to that component is associated with positive expected returns, independent of momentum.

Whatever one's view of the existing behavioral explanations, the theory presented here has the benefit of offering clearly defined quantitative predictions about risk and return in terms of familiar and (to some extent) observable characteristics of individual firms or portfolios. Exploring these is the subject of ongoing research.

Appendix: Proofs

LEMMA 1: *Let $U(x)$ be as defined in (b) of Proposition 1 and assume the condition in (a) is satisfied. Then, for all x, $U'(x)/U(x)$ is a positive, increasing function.*

Proof: First, the integrand in the definition of $U(\quad)$ is positive, so $U > 0$ for all x. Next, by assumption, $\zeta_1 < 0$. So regardless of the signs of the other terms in the exponential, that integrand is bounded by $\exp(\zeta_1 y)$ (where y is the integration variable). Hence, differentiation with respect to x may be taken inside the integral. Call the integrand $h(y)$. Then

$$\frac{d}{dx}\int h(y)\,dy = -\frac{1}{\kappa}\,e^{-\kappa y}h(y)\,dy$$

and $\int e^{-\kappa y}h(y)\,dy < \int h(y)\,dy$. So

$$\begin{aligned} U'(x) &= \frac{1}{\kappa}\left[U(x) - e^{((x/\kappa)-z_0)}\int e^{-\kappa y}h(y)dy\right] \\ &= \frac{1}{\kappa}\,e^{((x/\kappa)-z_0)}\left[\int h(y)dy - \int e^{-\kappa y}h(y)dy\right] > 0. \end{aligned}$$

To see that $U'(x)/U(x)$ is increasing, write

$$\begin{aligned} \left(\frac{U'}{U}\right)' &= -\frac{1}{\kappa}\left(\frac{\displaystyle\int e^{-\kappa y}h(y)dy}{\displaystyle\int h(y)dy}\right)' \\ &= \frac{1}{\kappa^2}\left[\frac{\displaystyle\int e^{-2\kappa y}h(y)dy}{\displaystyle\int h(y)dy} - \left(\frac{\displaystyle\int e^{-\kappa y}h(y)dy}{\displaystyle\int h(y)dy}\right)^2\right] \\ &= \frac{1}{\kappa^2}\,\frac{1}{\left(\displaystyle\int h(y)dy\right)^2} \\ &\quad\times\left[\left(\int e^{-2\kappa y}h(y)dy\right)\left(\int h(y)dy\right) - \left(\int e^{-\kappa y}h(y)dy\right)^2\right]. \end{aligned}$$

The third term in the last expression is positive by an application of the Cauchy–Schwartz inequality. Q.E.D.

PROPOSITION 2: *Let $\mathcal{F}_t$ be the time t information set. Then, assuming $\rho_{D\mu} \geq 0$ and $\rho_{\Lambda\mu} < 0$,*

$$\mathrm{E}[(\mathrm{CER}_{t+\ell} - \mathrm{E}[\mathrm{CER}_{t+\ell}|\mathcal{F}_t])\cdot(\mathrm{EER}_{t+\ell} - \mathrm{E}[\mathrm{EER}_{t+\ell}|\mathcal{F}_t])|\mathcal{F}_t] > 0.$$

Proof: The covariance to $t + \ell$ is the integrated expected instantaneous cross-variation of the two processes. From Itô's lemma, the diffusion term of the EER process is

$$-\rho_{\Lambda\mu}\sigma_{\Lambda}s\left(\frac{U'}{U}\right)' dW_t^{(\mu)}.$$

So the instantaneous covariance is

$$-\rho_{\Lambda\mu}\sigma_{\Lambda}s\left(\frac{U'}{U}\right)' \cdot \left(\sigma_D\rho_{D\mu} + \frac{U'}{U}s\right).$$

The terms involving $U(\ )$ are positive by the lemma. So are σ_{Λ}, σ_D, and s. The assumption about the correlations then ensures that the cross variation is *always* positive. Hence, its integrated expected value from t to $t + \ell$ is. Q.E.D.

PROPOSITION 3: *With the processes Λ and D defined by equations (1) and (2), and with the growth rate process given by (4)–(6), the stock price is*

$$P(D,x,y,S) = D_t \cdot (u^{(0)}(x_t,y_t)\cdot(1 - S_t) + u^{(1)}(x_t,y_t)\cdot S_t), \tag{A1}$$

where $u^{(0)}(\ )$ and $u^{(1)}(\ )$ satisfy the coupled partial differential equations

$$\frac{s_0^2}{2}u_{yy}^{(0)} + [\kappa_0(\bar{y} - y) + s_0(\rho_{\Lambda\mu}\sigma_{\Lambda} + \rho_{D\mu}\sigma_D)]u_y^{(0)}$$
$$+ [\kappa_1(\bar{x} - x)]u_x^{(0)} + [(x + y) - r + \rho_{\Lambda D}\sigma_{\Lambda}\sigma_D]u^{(0)} + \lambda_0(u^{(1)} - u^{(0)}) - 1 = 0$$

$$\frac{s_1^2}{2}u_{xx}^{(1)} + [\kappa_1(\bar{x} - x) + s_0(\rho_{\Lambda\mu}\sigma_{\Lambda} + \rho_{D\mu}\sigma_D)]u_x^{(1)}$$
$$+ [\kappa_0(\bar{y} - y)]u_y^{(1)} + [(x + y) - r + \rho_{\Lambda D}\sigma_{\Lambda}\sigma_D]u^{(1)} + \lambda_1(u^{(0)} - u^{(1)}) - 1 = 0.$$

The instantaneous expected excess return and volatility of P are given by

$$\mathrm{EER}_t = -\sigma_{\Lambda}\{\rho_{\Lambda D}\sigma_D + \rho_{\Lambda\mu}\Psi_t\}$$

$$\mathrm{VOL}_t = (\sigma_D^2 + 2\rho_{D\mu}\sigma_D\Psi_t + \Psi_t^2)^{1/2},$$

where

$$\Psi_t \equiv (1 - S_t)\cdot s_1 \cdot \frac{u_y^{(0)}}{u^{(0)}} + S_t \cdot s_0 \cdot \frac{u_x^{(1)}}{u^{(1)}}.$$

The covariance between expected excess returns and cumulative excess returns is

$$-\rho_{\Lambda\mu}\sigma_{\Lambda}\Upsilon_t\cdot(\rho_{D\mu}\sigma_D + \Psi_t)$$

with

$$\Upsilon_t \equiv (1 - S_t)\cdot s_1\cdot\frac{\partial}{\partial y}\left(\frac{u_y^{(0)}}{u^{(0)}}\right) + S_t\cdot s_0\cdot\frac{\partial}{\partial x}\left(\frac{u_x^{(1)}}{u^{(1)}}\right).$$

Proof: The proof is an application of the generalized Itô formula for jump processes (cf. Gihman and Skorohod (1972), II.2.6) to the product $\Lambda_t P_t$. Using this and the specification of equation (1), the expected instantaneous change in this product is

$$\begin{aligned}&-r\cdot P_t + \mathcal{D}P_t\cdot\Lambda_t + \langle\Lambda, P_t\rangle\\ &\quad+\{\lambda_0(P(S=1) - P(S=0))(1-S) + \lambda_1(P(S=0) - P(S=1))S\}\cdot\Lambda_t,\end{aligned} \tag{A2}$$

where $\mathcal{D}P_t$ is the usual Itô drift

$$\begin{aligned}&\frac{\sigma_D^2}{2}D^2\frac{\partial^2 P}{\partial D^2} + \frac{s_0^2}{2}(1-S)\frac{\partial^2 P}{\partial y^2} + \frac{s_1^2}{2}(S)\frac{\partial^2 P}{\partial x^2}\\ &\quad+\rho_{Dy}\sigma_D D s_0(1-S)\frac{\partial^2 P}{\partial D\partial y} + \rho_{Dx}\sigma_D D s_1(S)\frac{\partial^2 P}{\partial D\partial x} + \rho_{xy}s_0 s_1(S)(1-S)\frac{\partial^2 P}{\partial x\partial y}\\ &\quad+(\mu D)\frac{\partial P}{\partial D} + (\kappa_0(\bar{y} - y))\frac{\partial P}{\partial y} + (\kappa_1(\bar{x} - x))\frac{\partial P}{\partial x}\end{aligned}$$

and $\langle\Lambda_t, P_t\rangle$ the instantaneous covariance

$$\sigma_{\Lambda}\left\{(\rho_{\Lambda D}\sigma_D D)\frac{\partial P}{\partial D} + (\rho_{\Lambda y}s_0(1-S))\frac{\partial P}{\partial y} + (\rho_{\Lambda x}s_1 S)\frac{\partial P}{\partial x}\right\}.$$

The final term in (A2) is the contribution to the expected change from the possibility of a jump in P, which is multiplied by the jump intensity relevant to the current state. (Note that the continuous process Λ can have no covariation with the jump component of P.)

Next, by definition, the pricing kernel Λ (whose existence is assumed) determines P by the equation

$$\Lambda_t P_t = \mathrm{E}_t\left[\int_t^{\infty}\Lambda_u D_u\,du\right].$$

Since the process $\mathrm{E_t}\{\int_0^\infty \Lambda_u D_u \, du\}$ is a martingale, the expected change in $\Lambda_t P_t$ must also be given by $-\Lambda_t D_t$. Equate this to (A2), and use $S(1-S)=0$ to simplify. Also noting that we have defined the innovation to both x and y to be the same process dW^μ, the six correlations collapse to three (labeled in the obvious manner). Plugging in a solution of the form (A1) and dividing by D yields a partial differential equation that the price must satisfy. This one equation must be satisfied whether $S=1$ or $S=0$. The two equations given in the proposition correspond to these two cases.

The derivation of the moments of the return process in terms of the solution is then a straightforward application of Itô's lemma. Q.E.D.

REFERENCES

Asness, Clifford S., John M. Liew, and Ross L. Stevens, 1997, Parallels between the cross-sectional predictability of stock and country returns, *Journal of Portfolio Management* Spring, 79–87.

Bansal, Ravi, and Amir Yaron, 2000, Risks for the long-run: A potential explanation for asset pricing puzzles, Working paper, Duke University.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–344.

Barsky, Robert B., and J. Bradford DeLong, 1993, Why does the stock market fluctuate? *Quarterly Journal of Economics* 108, 291–311.

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1153–1607.

Brennan, Michael J., and Yihong Xia, 2001, Stock price volatility and the equity premium, *Journal of Monetary Economics* 47, 249–283.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *Journal of Finance* 51, 1681–1714.

Chordia, Tarun, and Lakshmanan Shivakumar, 2001, Momentum, business cycle and time-varying expected returns, *Journal of Finance* 57, this issue, 985–1019.

Conrad, Jennifer, and Gautam Kaul, 1998, An anatomy of trading strategies, *Review of Financial Studies* 11, 489–519.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and over-reactions, *Journal of Finance* 53, 1839–1886.

Donaldson, R. Glen, and Mark Kamstra, 1996, A new dividend forecasting procedure that rejects bubbles in asset prices: The case of 1929's stock crash, *Review of Financial Studies* 9, 333–383.

Gihman, I. I., and A. V. Skorohod, 1972, *Stochastic Differential Equations* (Springer-Verlag: Berlin).

Grundy, Bruce D., and Spencer J. Martin, 2001, Understanding the nature of risks and the sources of rewards to momentum investing, *Review of Financial Studies* 14, 29–78.

Hong, Harrison, and Jeremy Stein, 1999, A unified theory of underreaction, momentum trading and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 56, 699–720.

Johnson, Timothy C., 2001, Return dynamics when persistence is unobservable, *Mathematical Finance* 11, 415–445.

Lehmann, Bruce, 1990, Fads, martingales, and market efficiency, *Quarterly Journal of Economics* 105, 1–28.

Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–469.

Mitchell, Mark, and Erik Stafford, 2000, Managerial performance and long-term stock price performance, *Journal of Business* 73, 287–328.

Moskowitz, Tobias, and Mark Grinblatt, 1999, Does industry explain momentum? *Journal of Finance* 54, 1249–1290.

Rouwenhorst, K. Geert, 1998, International momentum strategies, *Journal of Finance* 53, 267–284.