Stock Market Predictability: Is it There? 
A Critical Review

David Rey
University of Basel/University of St. Gallen

December 2003

Author:
David Rey
University of Basel
WWZ/Department of Finance
Holbeinstrasse 12
4051 Basel
david.rey@unibas.ch
Phone: +41 (0)61 267 3307

WWZ/Department of Finance, Working Paper No. 12/03
Stock Market Predictability: Is it There?
A Critical Review

David Rey*

University of Basel (WWZ)
and
University of St. Gallen (HSG)

December 15, 2003

Abstract

This paper aims to survey selected recent papers presenting new evidence on an age-old question in financial economics: “Are stock market returns predictable?” The hypothesis that equity returns are predictable (specifically at long horizons) has been called a “new fact in finance” by Cochrane (1999). Thus, stock market predictability is now taken almost as a feature of the data. However, there is less consensus on what drives this predictability. It may reflect time-varying risk premiums, it may reflect irrational behavior on the part of market participants, or it may simply not present in the data – a statistical fluke due to poor statistical inference. It is this last possibility that seems to gain increasing credibility considering the long list of authors criticizing the statistical methodologies in the predictability literature. The results of these studies typically show that findings against the constant expected excess return hypothesis based on standard statistical inference can appear much more significant than they really are. Of course, the strength and usefulness of these critical results are still hotly debated – consistently evaluating their significance is thus no easy task. Still, they raise questions about the implications of these findings for the way academics and practitioners use financial theory. In particular, the implications for the asset management industry (and dynamic asset allocation strategies) may be serious.

*Department of Finance, WWZ University of Basel, Holbeinstrasse 12, 4051 Basel, Switzerland; Phone: +41-61-267 33 07, Mail: david.rey@unibas.ch. Preliminary version. Financial support from the National Center of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK) is gratefully acknowledged. The NCCR FINRISK is a research program supported by the Swiss National Science Foundation. A short version appears in Financial Markets and Portfolio Management 17/3, 2003 (Current Research Topics).
This paper aims to survey selected recent papers presenting new evidence on an age-old question in financial economics: "Are stock market returns predictable?". The hypothesis that equity returns are predictable (specifically at long horizons) has been called a "new fact in finance" by Cochrane (1999a). Thus, stock market predictability is now taken almost as a feature of the data. However, there is less consensus on what drives this predictability. It may reflect time-varying risk premiums, it may reflect irrational behavior on the part of market participants, or it may simply not present in the data – a statistical fluke due to poor statistical inference. It is this last possibility that seems to gain increasing credibility considering the long list of authors criticizing the statistical methodologies in the predictability literature. The results of these studies typically show that findings against the constant expected excess return hypothesis based on standard statistical inference can appear much more significant than they really are. Of course, the strength and usefulness of these critical results are still hotly debated – consistently evaluating their significance is thus no easy task. Still, they raise questions about the implications of these findings for the way academics and practitioners use financial theory. In particular, the implications for the asset management industry (and dynamic asset allocation strategies) may be serious.

The paper is organized as follows. The first section aims to provide a general and conceptional overview of the stock market predictability literature. Section 2 reviews four selected current papers that call the evidence of return predictability into question. In that they point out a number of pitfalls in the calculations of the usual test statistics and interpretations, they may constitute important references for future research in the area. The final section concludes and develops possible implications for the asset management industry and dynamic asset allocation strategies.

1 Introduction and Overview

Perhaps the most elusive goal in finance is the ability to predict stock market returns. Not surprisingly, thus, the literature on stock market predictability has evolved considerably over the last, say, thirty years. Originally, tests of the predictability of stock market returns were motivated by market efficiency. It was commonly assumed that predictability would be inconsistent with the constant expected returns, efficient markets paradigm. Indeed, it was long thought that stock returns are not predictable, at least not in an economically significant manner (Fama, 1970). The possibility that they could be predicted, be it only partially, by other observable variables seemed to have been unthinkable until the early 80’s.

Stock Market Predictability: A New Fact in Finance... Little by little, however, a series of articles have documented a small degree of predictability in stock returns based on prior information, specifically at long horizons. This evidence is statistically rather weak when only past returns are used to pre-
dict future returns (this is the univariate mean-reversion literature, starting with Fama and French, 1988b, and Poterba and Summers, 1988), but it seems considerably stronger when other variables are brought into the analysis. Examples of such other predictive variables include (recent changes in) short-term interest rates (e.g., Fama and Schwert, 1977; Campbell, 1991), yield spreads between long-term and short-term interest rates and between low- and high-quality bond yields (e.g., Keim and Stambaugh, 1986; Campbell, 1987; Fama and French, 1989), stock market volatility (e.g., French, Schwert, and Stambaugh, 1987; Goyal and Santa-Clara, 2003), Eurodollar-U.S. Treasury (TED) spread (e.g., Ferson and Harvey, 1993), book-to-market ratios (e.g., Kothari and Shanken, 1997; Pontiff and Schall, 1998), dividend-payout and price-earnings ratios (e.g., Lamont, 1998; Campbell and Shiller, 1988a), and more complex measures based on analysts’ forecasts (e.g., Lee, Myers, and Swaminathan, 1999). Recently, Baker and Wurgler (2000) have shown that the share of equity in new finance is a negative predictor of future equity returns. Lettau and Ludvigson (2001) find evidence for predictability using a consumption-wealth ratio, the level of consumption relative to income and wealth. The most influential papers, however, are Fama and French (1988a, 1989) and Campbell and Shiller (1988b). They argue that dividend yields, $D_t/P_{t-1}$, and particularly dividend-price ratios, $D_t/P_t$, on aggregate stock portfolios should and do predict expected (long-horizon) returns with some success.

Given this mounting empirical evidence of predictability, it is fair to say that the pendulum has swung in the opposite direction of the constant expected returns paradigm: time-varying expected returns as well as stock market predictability are now taken almost as a feature of the data – indeed, the hypothesis that equity returns (more precisely, excess returns over short-term interest rates) are predictable (specifically at long horizons) has been called a “new fact in finance” by Cochrane (1999a).

...and its Economic Interpretation At the same time, there is less consensus on what drives this predictability. Bekaert (2001) differentiates between three possibilities: it may reflect time-varying risk premiums (which he calls the “risk view”), it may reflect irrational behavior on the part of market participants (the “behavioral view”) or it may simply not present in the data – a statistical fluke due to poor statistical inference (the “statistical view”).

Of course, return predictability can be interpreted only in conjunction with an intertemporal equilibrium model of the economy. To formally test whether predictability is consistent with market efficiency, a model is needed that tells how assets are priced. Inevitably, thus, all theoretical attempts at interpretation of predictability will be model-dependent, and hence inconclusive. Nevertheless, while much work has focused on the second possibility, recent advances in (rational) asset pricing theory seems to have persuaded a majority of researchers that a certain degree of time-varying expected excess returns is necessary to reward investors for bearing certain dynamic risks associated with the business cycle. Loosely, it is claimed that the equity premium rises during an economic slow-
down and falls during periods of economic growth, so that expected returns and business conditions move in opposite directions (e.g., Fama and French, 1989; Chen, 1991; Fama, 1991; Ferson and Harvey, 1991). Consequently, stock market predictability on its own would not imply stock market inefficiency (and irrational behavior).

On the other hand, the risk view has sometimes been discredited because the empirical evidence suggests substantial time variation in risk premiums that cannot (yet) be delivered by standard (still extremely stylized) models of risk. Moreover, other aspects of the empirical research on predicting returns remain controversial. Specifically, it seems that the statistical view gains increasing credibility considering the long list of authors criticizing the statistical methodologies in the predictability literature. Recall that almost all of the empirical research in this field has relied on regression models in which the realized equity premium is regressed on one (or more) of the predictive variables indicated above, and significant t- or F-statistics and high $R^2$s are interpreted in favor of stock market predictability.  

The Typical Predictive Regression Specification  In particular, the typical predictive regression specification is often as simple as

$$e_{t+1} = \alpha + \beta x_t + \xi_{t+1},$$  

(1)

where $e_{t+1} = r_{t+1} - r_{f,t+1}$ is the log excess return, $r_{t+1} \equiv \ln (1 + R_{t+1})$ the log (nominal or real) equity return (including dividends), $r_{f,t+1} \equiv \ln (1 + R_{f,t+1})$ the log (nominal or real) short-term risk-free rate, and $x_t$ the predictive variable, known at the beginning of the return period. The time $t$ conditional equity premium is thus given by $E_t(e_{t+1}) = \alpha + \beta x_t$. It is easy to show that $\beta$ must be zero if expected excess returns are constant (the “random walk” null hypothesis). The alternative hypothesis of predictability is thus $\beta \neq 0$.

The predictive variable is often assumed to follow a (stationary) first-order autoregressive (AR(1)) process,

$$x_{t+1} = \gamma + \delta x_t + \eta_{t+1},$$  

(2)

When the AR coefficient $\delta$ is close to (but strictly less than) one, the process for the predictive variable is highly persistent.

Although this regression specification is rather simple at first sight, the econometric problems that are involved in testing for predictability are manifold. Standard statistical inference relies on first-order asymptotic distribution

\footnote{Instead of realized equity premiums (i.e., realized excess returns over short-term interest rates), some studies use nominal or real returns. In empirical studies, however, the impact on the results is quantitatively minuscule: relative to equity returns, short-term interest rates are generally much less volatile.}

\footnote{Using continuously compounded returns instead of simple returns is more common, particularly in the long-horizon regression specification described below. However, the change from log returns to levels of returns might have a non-negligible effect on the results (Cochrane, 2001, Ch. 20, p. 412).}
theory, which implies that $t$-statistics are approximately standard normal in large samples. However, there is some question as to whether “asymptotic” should be measured in terms of years, decades, or even centuries, especially for the long-horizon regressions discussed below. Hence, an important question is whether the large sample theory provides an accurate approximation to the actual finite sample distribution of the $t$-statistics. Unfortunately, this may not be the case: first-order asymptotics are often a poor approximation in finite samples. Indeed, findings against the constant expected excess return hypothesis based on standard statistical inference can appear more significant than they really are.

**Persistent and Predetermined Predictive Variables**
To begin with, most popular and economically sensible candidates for predictive variables are (i) highly persistent (i.e., the autoregressive root $\delta$ in the univariate representation is close to one, $\delta \approx 1$) and (ii) are not really exogenous but lagged endogenous (i.e., predetermined), such that $\xi_t$ is correlated with $x_t$, leading to violation of one of the standard assumptions of ordinary least squares (OLS), namely the independence at all leads and lags. Together with the fact that the correlation between return and predictive variable innovations is often strongly negative — mostly so in the case where the predictive variable is a financial ratio! — (i) and (ii) result in high simultaneity in the above regression system. Accordingly, the autoregressive root estimate $\hat{\delta}$ is biased downward, negatively skewed, and more variable than suggested by OLS in finite samples. Taking into account the negative correlation between the innovations $\xi_t$ and $\eta_t$, these properties imply that in finite samples the predictive slope estimate $\hat{\beta}$ is biased upward, positively skewed, and more variable than suggested by OLS, leading to sizeable overrejection of the null hypothesis using conventional critical values. Notationally, as shown in Campbell, Lo, and MacKinlay (1997, Ch. 7), Stambaugh (1999), and Lewellen (2003), these biases can be summarized as

$$E \left( \hat{\beta} - \beta \right) = - \left( \frac{1 + 3\delta}{T} \right) \frac{\sigma_{\eta \xi}}{\sigma_{\xi}^2} = E \left( \hat{\delta} - \delta \right) \frac{\sigma_{\eta \xi}}{\sigma_{\xi}^2},$$

(3)

where $T$ is the sample size, $\sigma_{\eta \xi} \equiv Cov(\xi_t, \eta_t)$ the covariance between the innovations $\xi_t$ and $\eta_t$, and $\sigma_{\xi}^2 \equiv Var(\xi_t)$ the variance of $\xi_t$, $\sigma_{\xi}^2 = (1 - \delta^2) \sigma_x^2$, $\sigma_x^2 \equiv Var(x_t)$.

**The Special Status of the Dividend-Price Ratio, Dividend Growth, and the Errors-in-Variables Problem**
The hypothesis that dividend yields (and dividend-price ratios) predict returns has a long tradition among practitioners and academics (e.g., Ball, 1978; Rozeff, 1984). The intuition of the “efficient markets” version of the hypothesis is that stock prices are low relative to dividends when discount rates and expected returns are high, and vice versa, so that dividend yields vary with expected returns. Interestingly, thus, in contrast to the bulk of the other predictive variables indicated above, the risk-based business-cycle argumentation is not explicitly needed to explain the predictive
power of the dividend-price ratio. It can be shown that, without relying on any asset pricing model, dividend-price ratios can only move at all if they predict expected future returns, if they forecast expected future dividend growth, or if there is a bubble – if the inverse of the dividend-price ratio is non-stationary and is expected to grow explosively. When prices are high relative to dividends (or earnings, cash flow, book value or some other divisor), one of three things must be true. First, investors expect dividends to rise in the future. Second, investors expect returns to be low in the future. Third, investors expect prices to rise forever, giving an adequate return even if there are no dividends. This statement is not a theory, it is an identity. If the dividend-price ratio is low, either dividends must rise, prices must decline, or the inverse of the dividend-price ratio must grow explosively. The open question is, which option holds for the stock market? Are prices high now because investors expect high dividend growth in the future, because they expect low returns in the future, or because they expect prices to go on rising forever?

To start with, consider the well-known discrete-time perfect-certainty model in which \( D \), the dividend per share, grows at the constant rate \( G \), and the market interest rate that relates the stream of future dividends to the stock price \( P \) is the constant \( R \). In this model, the stock price at time \( t \) is given by

\[
P_t = \sum_{j=1}^{\infty} D_t (1 + G)^j / (1 + R)^j = \frac{D_{t+1}}{R - G},
\]

where \( G < R \). Hence, the dividend yield is the interest rate less the dividend growth rate,

\[
\frac{D_{t+1}}{P_t} = R - G.
\]

In this deterministic dividend discount model, the interest rate \( R \) is the discount rate for dividends and the expected return on the stock. Although the model is not directly applicable to the case in which dividend growth and discount rates vary through time, the model already suggests that the dividend yield captures expectations about dividend growth as well as expected returns. The transition from certainty to a model that (i) accommodates uncertain future dividends and discount rates and (ii) shows the correspondence between discount rates and time-varying expected returns is more involved, however.

Campbell and Shiller (1987, 1988a,b) – accessible textbook treatments are Campbell, Lo, and MacKinlay (1997, Ch. 7) and Cochrane (2001, Ch. 20) – develop an approximate present value relation with time-varying expected returns. Their approach is to use a loglinear approximation that has the advantage that it is tractable under the empirically plausible assumption that dividends and returns follow loglinear driving processes. The loglinear approximation starts with the following identity

\[
1 = (1 + R_{t+1})^{-1} (1 + R_{t+1}) = (1 + R_{t+1})^{-1} \left( \frac{P_{t+1} + D_{t+1}}{P_t} \right).
\]
Multiplying both sides by \( P_t/D_t \) and massaging the result, the inverse of the dividend-price ratio can be expressed as

\[
\frac{P_t}{D_t} = (1 + R_{t+1})^{-1} \frac{D_{t+1}}{D_t} \left( 1 + \frac{P_{t+1}}{D_{t+1}} \right).
\]

Taking logs, and with lowercase letters denoting logs of uppercase letters, the log dividend-price ratio is given by

\[
\ln \left( \frac{D_t}{P_t} \right) \equiv d_t - p_t = r_{t+1} - \Delta d_{t+1} - \ln \left( 1 + e^{p_{t+1}-d_{t+1}} \right),
\]

with \( r_t \) the continuously compounded (real) stock return over period \( t \), and \( \Delta d_{t+1} = d_{t+1} - d_t \) (real) dividend growth. Taking a (first-order) Taylor approximation of the last term about a point \( P/D = e^{p-d} \),

\[
\ln \left( 1 + e^{p_{t+1}-d_{t+1}} \right) \approx \ln \left( 1 + \frac{P}{D} \right) + \frac{P/D}{1+P/D} \left( \left( p_{t+1} - d_{t+1} \right) - \left( p - d \right) \right),
\]

and defining

\[
k \equiv \ln \left( 1 + \frac{P}{D} \right) = \frac{P/D}{1+P/D} (p-d) \quad \text{and} \quad \rho \equiv \frac{P/D}{1+P/D}
\]

as parameters of linearization, the log dividend-price ratio can be approximated by

\[
d_t - p_t \approx r_{t+1} - \Delta d_{t+1} - k + \rho (d_{t+1} - p_{t+1}). \tag{6}
\]

Solving forward and assuming that the standard transversality condition holds, the approximate present value relation is obtained as

\[
d_t - p_t \approx -c + \sum_{j=1}^{\infty} \rho^{j-1} (r_{t+j} - \Delta d_{t+j}),
\]

with \( c \equiv k/(1 - \rho) \).

Since this approximate identity for the dividend-price ratio holds ex-post, we can take conditional expectations and relate the dividend-price ratio to ex-ante dividend growth and return forecasts as

\[
d_t - p_t \approx -c + E_t \sum_{j=1}^{\infty} \rho^{j-1} (r_{t+j} - \Delta d_{t+j}). \tag{7}
\]

This is a dynamic generalization of the Gordon (1962) formula for a stock price with constant required returns and dividend growth. Like the original Gordon growth model, the dynamic Gordon growth model says that stock prices are high when dividends are expected to grow rapidly or when dividends are discounted at a low rate; but the effect on the stock price of a high dividend growth rate (or a low discount rate) now depends on how long the dividend growth rate
is expected to be high (or how long the discount rate is expected to be low), whereas in the original model these rates are assumed to be constant at their initial levels forever.

The loglinear relation between prices, dividends, and returns provides an accounting framework. The variance of the dividend-price ratio can thus be decomposed as

$$\text{Var} \left( d_t - p_t \right) \approx \text{Cov} \left( d_t - p_t, \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \right) - \text{Cov} \left( d_t - p_t, \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} \right).$$

In words, in a rational no-bubble model, dividend-price ratios can only vary (i.e., $\text{Var} \left( d_t - p_t \right) > 0$) if they predict changing dividend growth or if they forecast changing returns. If returns are unpredictable (the “random walk” null hypothesis) and dividend growth is not forecastable either, the dividend-price ratio would have to be a constant. Thus, the fact that the dividend-price ratio varies at all means that either dividend growth or returns must be forecastable—loosely, that the world is not independently and identically distributed (i.i.d.).

The dividend-price ratio has a special status since it must predict something. To believe that the dividend-price ratio is stationary and varies, but does not predict expected future returns, one has to believe that it forecasts dividend growth, or both. But which is it?

At a simple level, matching the corresponding return regressions, Cochrane (2001, Ch. 20, Table 20.1) includes regressions of (long-horizon) dividend growth on dividend-price ratios. The coefficients in the dividend growth case are much smaller and statistically insignificant, and the $R^2$s are tiny. Worse, the signs are wrong. To the extent that a low dividend-price ratio forecasts any change in dividends, it seems to forecast a small decline in dividends! So, confirming his earlier findings (Cochrane, 1992, 1997) and those of Campbell (1991), for example, he concludes that historically virtually all variation in dividend-price ratios has reflected time-varying expected returns. Similarly, Campbell and Shiller (1998) show that after a deviation from its average value, the dividend-price ratio (and the price-earnings ratio) does a poor job as a forecaster of future dividend growth to the date when the ratio is borne back to its average value again. It is instead the stock price that moves to restore the ratio to its mean value, an offsetting adjustment to dividends (and earnings) seems to be absent. The standard conclusion in the literature seems thus to be threefold: (i) the bulk of the variance of dividend-price ratios must be accounted for by changing forecasts of discount rates, (ii) expected dividend growth is approximately constant, and (iii) high prices reflect low expected returns!

However, recent work by Ang and Bekaert (2003), Engstrom (2003), Goyal and Welch (2003), and Lettau and Ludvigson (2003) suggest that this view is probably somewhat premature, at least incomplete.

First, while they would not deny that the accounting identity in equation (7) demonstrates that the dividend-price ratio must either forecast expected future returns or dividend growth in the long run, Goyal and Welch (2003) argue
that for finite-period-ahead predictions, equation (6) is more informative. The dividend-price ratio must thus predict either the next-period stock return, the next-period dividend growth rate, or the next-period dividend-price ratio. They show that although in the early part of their sample, in fact, the dividend-price ratio used to be a good predictor of future dividend growth, in recent years the ratio’s predictive ability has shifted towards an ability to predict its own future value (higher autoregressive root) rather than one-year-ahead equity returns or dividend growth rates. Empirically, it is only over horizons longer than about 5 to 10 years where the long-run accounting identity begins to dominate the self-predictive properties of the dividend-price ratio. Over shorter horizons, dividend-price ratios primarily forecast themselves, possibly explaining why the predictability of stock returns with dividend-price ratios has been weak for most of their sample period.

Second, since discount rates consist of a risk-free and risk premium component, \( r_t = r_{f,t} + e_t \), equation (7) might be rewritten

\[
d_t - p_t \approx -c + E_t \sum_{j=1}^{\infty} \rho^{j-1} (e_{t+j} - r_{f,t+j} - \Delta d_{t+j}),
\]

similar to Campbell (1991). Thus, the variation of dividend-price ratios should not only be attributed to the variation of expected future excess returns and dividend growth, but also of expected future interest rates. Ang and Bekaert (2003) is one of the few studies that examine the fraction of the variation in dividend-price ratios that reflects the predictability of future interest rates as well. Indeed, while somewhat in contrast to Campbell (1991), they find that the dividend-price ratio significantly predict interest rates!

In any case, because it might also reflect expected future dividend growth and future interest rates, the dividend-price ratio (as well as other predictive variables) is a noisy proxy for expected excess returns. Variation in the dividend-price ratio due to changes in expected dividend growth (and, additionally, due to changes in the dividend policies of firms!) can thus cloud the information about time-varying expected excess returns. More generally, any variation in the dividend-price ratio that is unrelated to variation in future expected excess returns is noise that tends to cause estimates of equation (1) (and equation (16) below) to understate the true variation of expected excess returns. The unanticipated component of dividend growth acts as a measurement error that biases the predictive regression’s explanatory power downward. This “errors-in-variables” problem is thus an additional source of bias in predictive regressions. However, in contrast to the bias arising from the persistent and predetermined character of the predictive variable (and the negative correlation between return and predictive variable innovations), estimates of the slope coefficient, \( \hat{\beta} \), are now biased downward, obscuring the predetermined variable’s predictive power.\(^3\)

Kothari and Shanken (1992) provide evidence that the errors-in-variables

\(^3\) A more detailed discussion is for example given in Fama and French (1988a), Kothari and Shanken (1992), Goetzmann and Jorion (1995), and, more recently (and challenging), Lettau and Ludvigson (2003).
problem is a potentially major one. They examine the extent to which aggregate stock return variation is explained by variables chosen to reflect revisions in expectations of future dividends. In effect, they decompose realized dividend growth into expected and unexpected components using information in aggregate investment, dividend yield, and future returns. The inclusion of these forward-looking proxies for the market’s expectations of future dividends to equation (1) (or equation (16) below) explains a substantial fraction of return variation. Both the regression $R^2$ and the coefficient on the dividend-price ratio increases, indicating a reduction of the downward bias. Fama (1990) also finds a substantial increase in $R^2$ when using measures of future industrial production as additional regressors.

Of course, this approach uses future variables and does thus not directly address the issue of the (out-of-sample) predictive ability of the dividend-price ratio in (univariate) regressions conditional only on past and present information. Nevertheless, if analysts have some ability to forecast expected dividend growth, these forecasts should be included in the predictive regression, in which case the dividend-price ratio might be a more useful predictor of future stock returns.

Still another interesting way to describe the conditional relationship between the equity premium and the dividend-price ratio is pointed out in Engstrom (2003). He demonstrates that within a very broad class of economic models of risk, predictive regressions may be misspecified and have almost no power against the specific form of predictability suggested by reasonable treatments of risk. Simple predictive regressions can therefore produce estimates of the conditional risk premium which may be very different from the true values.

In particular, he shows that a very general model of risk implies an intrinsically time-varying relationship between the dividend-price ratio and the conditional equity premium.

The starting point of Engstrom’s (2003) argument is the by now familiar and general conditional asset pricing relation,

$$E_t (m_{t+1} (1 + R_{t+1})) = 1, \quad (9)$$

where $m_{t+1}$ is the stochastic discount factor (SDF, or pricing kernel). Loosely, this general expression subsumes all economic models of returns in which the law of one price is respected, including the Consumption CAPM (CCAPM) and many recent extensions (Cochrane, 2001, provides a very comprehensive discussion).

By applying the covariance decomposition and noting that $E_t (m_{t+1}) = 1/(1 + R_{f,t+1})$, the relationship can be rewritten as

$$E_t (m_{t+1} (1 + R_{t+1})) = E_t (m_{t+1}) E_t (1 + R_{t+1}) + Cov_t (m_{t+1}, R_{t+1})$$

$$= \frac{1}{1 + R_{f,t+1}} E_t (1 + R_{t+1}) + Cov_t (m_{t+1}, R_{t+1})$$

$$= 1,$$
with \( \text{Cov}_t(.) \) the conditional covariance operator. Hence,

\[
E_t (R_{t+1}^e) \equiv \frac{E_t (1 + R_{t+1})}{1 + R_{f,t+1}} - 1 = -\text{Cov}_t (m_{t+1}, R_{t+1}),
\]

(10)

where \( E_t (R_{t+1}^e) \) can be understood as a kind of expected excess return.

In this framework, thus, excess equity returns have a (positive) non-zero conditional expectation only to the extent that they covary (negatively) with the stochastic discount factor. The economic reasoning behind such a model is easily illustrated in the context of the CCAPM, where the stochastic discount factor is given by the ratio of marginal utilities of consumption (of a representative agent) as

\[
m_{t+1} = \tau \frac{U'(C_{t+1})}{U'(C_t)},
\]

(11)

with \( C_t \) the aggregate real per capita consumption, \( U(.) \) the period utility function, and \( \tau \) a (time) discount factor, \( 0 < \tau < 1 \). Assuming utility is increasing and concave in consumption, marginal utility and the SDF will be relatively high in times of relatively low consumption. If stocks are expected to have high returns in states of the world where marginal utility is low, the SDF and returns will be negatively correlated, stocks are viewed more risky by investors, and the risk premium will be higher. An asset whose covariance with the SDF is large and negative tends to have low returns when the SDF is high, i.e., when marginal utility is high. In equilibrium such an asset must have a high excess return to compensate for its tendency to do poorly in states of the world where wealth is particularly valuable to investors.

By inserting the definition of equity returns, equation (10) can be further manipulated to elucidate the role of the dividend-price ratio in determining the equity premium,

\[
E_t (R_{t+1}^e) = -\text{Cov}_t \left( m_{t+1}, \frac{P_{t+1} + D_{t+1}}{P_t} \right) = -\text{Cov}_t \left( m_{t+1}, \frac{P_{t+1}}{P_t} \right) - \text{Cov}_t \left( m_{t+1}, \frac{D_{t+1}}{D_t} \right) \left( \frac{D_t}{P_t} \right). \quad (12)
\]

The second line of equation (12) shows another way why the dividend-price ratio can be viewed as playing a part in measuring (and predicting) the equity premium. In particular, this representation decomposes the risk premium into two components. The first, represented by the first term on the right-hand side of equation (12), captures risk due to capital appreciation of the stock market. Only to the extent that capital appreciation covaries negatively with the pricing kernel will this “intercept” term be positive. The second component of the risk premium arises from equity risk due to fundamental (dividend) growth. It is the covariance of the pricing kernel with dividend growth scaled by the dividend-price ratio. Thus, equation (12) illustrates that expected excess returns are indeed affine in the dividend-price ratio – conditional on the component premiums represented by covariances with the stochastic discount factor.
Equation (12) also illuminates the economic content of the hypothesis tested in standard regression analysis. In particular, the null hypothesis of no predictability implies that the slope coefficient of the dividend-price ratio is zero. In the general context of the stochastic discount factor framework, this hypothesis is equivalent to the dubious notion that investors are consistently risk neutral with respect to short-term equity cash flows. To the extent that cash flows are risky, this null hypothesis is false. More importantly, to the extent that cash flow risk varies through time, these regressions may be misspecified.

Under the additional assumption ruling out rational bubbles (i.e., as above, that the standard transversality condition holds),

$$ P_t = E_t \left( m_{t+1} (P_{t+1} + D_{t+1}) \right) $$

can be regarded as a stochastic difference equation that has the fundamental and complete solution

$$ P_t = E_t \left( \sum_{i=1}^{\infty} \left( \prod_{j=1}^{i} m_{t+j} \right) D_{t+i} \right) = D_t E_t \left( \sum_{i=1}^{\infty} \prod_{j=1}^{i} m_{t+j} G_{t+j} \right), \quad (13) $$

where $G_{t+j} \equiv D_{t+j}/D_{t+j-1}$ is the gross rate of dividend growth. Engstrom (2003) shows that equation (12) can be rewritten as a function of future dividend growth and stochastic discount factors,

$$ E_t \left( R_{t+1}^e \right) = \left( -\text{Cov}_t \left( m_{t+1}, G_{t+1} \left( 1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} m_{t+j+1} G_{t+j+1} \right) \right) \right) \left( \frac{D_t}{P_t} \right) \equiv \beta_t \left( \frac{D_t}{P_t} \right). \quad (14) $$

Written in this form, expected excess returns are revealed to be linear (not affine) in the dividend-price ratio, conditional on a potentially time-varying $\beta_t$. The intuition of this conditional compound premium is that it captures not only short-term fundamental risk, as in the second term of equation (12), but also that of an indefinitely long chain of future stochastic discount factors and dividend growth.

The representations in equations (12) and (14) immediately suggest that the dividend-price ratio regression coefficient may in fact vary through time, and the common components of the dividend-price ratio in equation (13) and $\beta_t$ hint that they may coevolve as well. Of course, the specific nature of this relationship depends on the structural economic environment, but it may well be the case that shocks to the conditional mean of the pricing kernel induce a negative correlation between the dividend-price ratio and the component premiums in equation (12). As a result, if $\beta_t$ is time varying and negatively correlated with the component premiums, then unconditional predictive regressions (where the component premiums are simply modeled as constant) suffer in two important ways. First, OLS regression statistics have almost no power to reject the null hypothesis of no predictability under the alternative hypothesis. Second, the
OLS slope coefficient estimate, if interpreted as the unconditional mean of $\beta_t$, provides a severe underestimate as a result of omitted variable bias (which therefore cannot be expected to improve even in asymptotically large samples). After all, if the variance of $\beta_t$ is large, past unconditional estimates of $\beta_t$ are likely to be quite different from the current prevailing value, which would correctly predict future excess returns. This suggests that the out-of-sample performance of unconditional models is likely to be low!

Since equation (14) suggests that the coefficient on the dividend-price ratio represents a conditional covariance between the stochastic discount factor and future pricing kernels and dividend growth, a quick and easy first check for state dependence of this quantity is to model $\beta_t$ as a non-stochastic, affine function of a state vector such as

$$
\begin{align*}
e_{t+1} &= \alpha + \beta_1 x_t + \xi_{t+1} \\
&= \alpha + (\beta_0 + \beta_1 x_{1,t}^s + ... + \beta_n x_{n,t}^s) x_t + \xi_{t+1},
\end{align*}
$$

(15)

where $x_t$ represents the dividend-price ratio and $x_{i,t}^s$, $i = 1, ..., n$, state variables that are expected to drive conditional expectations in the economy, for example, as in Engstrom (2003), the dividend growth rate, inflation, and the risk-free rate. Engstrom (2003) shows that equation (15) exhibit quite a bit more return predictability than their static coefficient counterparts in equation (1); the null hypothesis $\beta_0 = \beta_1 = ... = \beta_n = 0$ is strongly rejected.

Altogether, thus, the unobserved conditional equity premium can be decomposed into two factors, only one of which – the dividend-price ratio – is observed, giving rise to the potential use of the dividend-price ratio as a tool for measuring the conditional equity premium, indeed. However, any empirical investigation which uses the dividend-price ratio in predicting the conditional equity premium implicitly makes assumptions about the unobserved factor, $\beta_t$. If $\beta_t$ is constant (i.e., $\beta_t = \beta$), then an unconditional regression of excess returns onto the dividend-price ratio is appropriate for this purpose and the estimated slope coefficient can be used to describe, ex-post and in conjunction with the current dividend-price ratio, the prevailing conditional equity premium at each point in the sample. However, if $\beta_t$ evolves according to a more general process, then an unconditional regression specification may be misspecified. Using popular asset pricing models, Engstrom (2003) examines the implied behavior of $\beta_t$ and the implications for small and large sample properties of the slope coefficient in a regression of excess stock returns on the lagged dividend-price ratio. He generally finds that $\beta_t$ has large and significant variation and a strong negative correlation with the dividend-price ratio. In particular, he shows that under alternatives in which $\beta_t$ exhibits variation suggested by a reasonable structural model, unconditional regressions have almost no power to reject the null hypothesis of no predictability. Moreover, estimates of the conditional risk premium suggested by unconditional regressions can be very different from the true values. Thus, the dynamics between $\beta_t$ and the dividend-price ratio render unconditional regressions nearly useless, linear (or affine) regression specifications extract almost none of the information in the dividend-price ratio about the equity premium!
Nevertheless, in practice, the most common approach used to document return predictability is to fit a linear regression model to the data – possibly because it is simple to implement and the resulting estimates of the slope coefficient(s) and regression $R^2$ have an appealing interpretation.

**Long-Horizon Regressions** However, instead of regressing realized short-period excess returns on one (or a set) of predictive variables as in equation (1), much of the empirical research has focused on long-horizon regressions of the form

$$e_{t+1→t+K} = \alpha(K) + \beta(K) x_t + \xi_{t+1→t+K},$$

(16)

where $e_{t+1→t+K} = e_{t+1} + e_{t+2} + ... + e_{t+K}$ is the cumulative log excess return from period $t+1$ to $t+K$.

There are all sorts of (partly overlapping) motivations for such long-horizon regressions. It is argued that (i) since asset prices are influenced by expectations of returns into the distant future (see equation (7) for an approximation), long-horizon regressions, best with as many independent, non-overlapping observations as possible, are necessary to understand price behavior; it may be only at these lower frequencies that the impact of economic factors such as the business cycle is detectable (e.g., Campbell, Lo, and MacKinlay, 1997, Ch. 7), (ii) using long-horizon returns rather than one-period returns increases statistical power and improves statistical efficiency (e.g., Hodrick, 1992; particularly, Campbell, 2001), and (iii), intuitively, long-run regressions produce more accurate results by strengthening the signal coming from the data while eliminating the noise (e.g., Torous and Valkanov, 2002; Valkanov, 2003).

Furthermore, these long-horizon studies typically have a number of characteristics in common. First, they use effective (i.e., independent, non-overlapping) sample sizes that are very small. Second, as a consequence, they assess the ability to predict returns based on a regression specification that, more often than not, uses overlapping returns. Finally, in the majority of cases, the strongly increasing in-sample $t$-statistics and sample $R^2$s from the long-horizon regression specification are treated as a measure of the economic significance of the predictable component of returns. Fama and French (1989), for example, attribute the dramatic increase in explanatory power at long horizons to low-frequency oscillations in expected returns. They contend that these low-frequency oscillations reflect the rational response of investors to slowly changing business conditions.

Still, while the tendency to treat sample $R^2$s as a measure of the economic significance of predictability is intuitively appealing (Cochrane, 2001, Ch. 20), it is nonetheless important to understand how the use of small sample sizes and overlapping returns affect the distributional properties of the $t$-statistics and hence of the sample $R^2$s. There are good reasons to be wary of such inferences when they are based on such long-horizon regressions. For example,

---

4A technical analysis (based on persistent first-order autoregressive processes) is provided for example in Fama and French (1988a), Campbell (1991, 2001), and Campbell, Lo, and MacKinlay, 1997, Ch. 7).
Foster, Smith, and Whaley (1997) and particularly Kirby (1997) show that the distribution of sample $R^2$s is severely shifted to the right in long-horizon regressions, considerably weakening previous conclusions about predictability. Indeed, the focus on long-horizon regressions is problematic from a number of different perspectives. Generally, it is not obvious whether tests of predictability in long-horizon regressions have greater power than short-horizon regression tests, or more serious size distortions, or both.

**Overlapping Data**  Aside from the extremely small sample sizes (e.g., from 1926 to 1985, a standard time period in the (early) literature, there are only 12 non-overlapping and therefore independent five-year return periods!), the main concern in long-horizon regressions follows from the use of overlapping data. This causes the error term to be strongly serially correlated. As a result, OLS standard errors understate the variance of the least-squares estimator of the predictive slope coefficient, $\hat{\beta}$. The evidence for long-horizon predictability thus critically depends on the choice of standard errors for making statistical inference. Yet, the estimators of the (heteroskedasticity and) autocorrelation consistent standard errors for $\hat{\beta}$ are known to be poorly behaved in small samples, and the presence of heteroskedasticity (of unknown form) is likely to further complicate the analysis (e.g., Richardson and Stock, 1989).

**Simulation Methods**  A few early attempts to resolve the econometric problems in testing for long-horizon predictability are based on simulation methods (bootstrap and its modifications). Hodrick (1992) examines the statistical properties of three alternative methods for conducting inference in long-horizon regressions. His VAR simulation-based methods indicate that substantial biases can arise in test statistics in long-horizon regressions. But he still concludes that dividend yields have some predictive power for stock returns in the U.S. stock market. Nelson and Kim (1993) analyze small-sample biases in simulations of a VAR system for returns and dividend yields, under the null hypothesis of no predictability. Using randomized (bootstrapped without replacement) null distributions for the test statistics, they find some, but little, predictability of postwar U.S. stock returns, too. On the other hand, also employing a bootstrapping and randomization approach, Goetzmann and Jorion (1993) find that the Fama and French (1988a) coefficient estimates are upward biased and fail to find any strong evidence for predictability of U.S. stock returns by dividend yields. Altogether, thus, these studies conclude that the evidence for predictability is not nearly as overwhelming at long horizons as previous studies had suggested, but the distortions in the distributions of the $t$-statistics are not always enough to completely overturn the evidence of stock market predictability.

**Near Unit Roots,...**  However, it is difficult to reconcile or compare the conclusions from these studies, because, based on simulations or bootstrap methods, they fail to yield general results. Their conclusions are ultimately a function of how the artificial data are being generated. Moreover, there is no way of know-
ing what features of the data would influence the statistics of interest and in what fashion. More importantly, it is even argued that the bootstrap results of Hodrick (1992), Nelson and Kim (1993), and Goetzmann and Jorion (1993), as well as closely related approaches such as Wolf’s (2000) subsampling method, are theoretically not justified. Recent theoretical results in the econometric literature indicate that these methods fail to provide an asymptotically valid method of inference in regression models in which the predictive variable has a near unit root, i.e., $\delta \approx 1$. Indeed, statistical inference in predictive regressions depends critically on the predictive variable’s stochastic properties one is willing to consider, notably its order of integration ($\delta < 1$ for stationarity versus $\delta = 1$ for a non-stationary unit root or “random walk”)! Incorporating information about the predictive variable’s order of integration can result in large efficiency gains and therefore have a significant effect on inferences drawn in predictive regressions (Campbell and Yogo, 2002; Torous, Valkanov, and Yan, 2002; Lewellen, 2003).

...Stationarity versus Non-Stationarity,... Of course, one might argue that it makes little sense to predict returns with non-stationary variables. Economically, predictive variables should be stationary unless there is an explosive bubble in stock prices. Suppose, for example, that the predictive variable equals the log dividend-price ratio. Then the dividend-price ratio is stationary if log dividends and log prices are cointegrated, implying that, in the long run, dividends and prices grow at the same rate. Certainly, this assumption is intuitively appealing (e.g., Lamont, 1998; Lewellen, 2003). On the other hand, dividend-price ratios might exhibit other forms of non-stationarity that do not imply explosive behavior. For example, Jagannathan, McGratten, and Scherbina (2000) and in particular Fama and French (2002) suggest that the equity premium dropped sharply over the last fifty years. If this drop is permanent, not caused by transitory sentiment or the business cycle, it should lead to a permanent drop in the dividend-price ratio. But if the recent drop in the dividend-price ratio is really caused by a permanent drop in the risk premium (and, for instance, not because of a change in payout policies), then at a basic level, it is already acknowledged that the dividend-price ratio tracks changes in expected returns. Thus, this story does not say that one might falsely reject the null hypothesis, because it inherently assumes that the null hypothesis is false (Lewellen, 2003). Altogether, thus, studies such as Stambaugh (1999), Campbell and Yogo (2002), and Lewellen (2003) explicitly exclude the possibility that predictive variables follow explosive non-stationary processes. However, Torous, Valkanov, and Yan (2002) argue that the a priori assumption of stationarity is not completely satisfactory. For example, based on an argument of Roll (2000), they show that rational expectations can imply non-stationarity in predictive variables which are functions of asset prices (such as the dividend-price and the book-to-market ratio). In particular, since rational expectations must not be expected to change, an expectation about a future quantity must follow a “random walk” if innovations in expectations are independent and identically
distributed. Since asset prices are functions of expectations about future quantities, asset prices and predictive variables which are functions of asset prices will also follow a “random walk”.

...and the Local to Unity Framework  Whatever, in empirical research, confidence intervals for the largest autoregressive root estimate $\delta$ often confirm uncertainty surrounding the predictive variable’s order of integration. In particular, unit root tests, such as the augmented Dickey and Fuller (1979) (ADF) $t$-test, are often not able to reject the null hypothesis of a unit root and are very sensitive to model misspecifications (Schwert, 1987, 1989a). Thus, with currently available econometric tools, the degree of persistence is hard to identify with sufficient precision. As a result, one can take different stands on the stationarity of the predictive variables, without being disproved by the data. Worse, Elliott and Stock (1994) and Cavanagh, Elliott, and Stock (1995) demonstrate that standard inference is not, in general, even asymptotically valid when the order of integration of the predictive variable is unknown, and that the use of a nearly integrated predictive variable introduces potentially substantial size distortions in conventional hypothesis testing, both in small samples and asymptotically. However, there are by now several asymptotically valid procedures of inference in simple regression models when the order of integration of the predictive variable is unknown, which provide a more accurate approximation to the finite sample distribution, too. Based on the theory of nearly integrated processes of, for example, Stock (1991), Elliott and Stock (1994) and Cavanagh, Elliott, and Stock (1995) derive an alternative asymptotic distribution theory in which the predictive variable is modeled as having a so-called local to unit root, i.e., $\delta = 1 + c/T$, where the nuisance parameter $c$ measures deviations from the unit root in a decreasing (at rate $T$, the sample size) neighborhood of one. The unit root case corresponds to $c = 0$. In other words, the predictive variable is assumed to follow an autoregression with a root near to unity in the sense that for a given sample size, however arbitrarily large, one is unable to distinguish the assumed stationary specification from the unit root alternative.

Some recent research has begun to empirically investigate stock return predictability within this local to unity framework. For instance, Torous, Valkanov, and Yan (2002) develop a test procedure, extending the work of Elliott and Stock (1994) and Cavanagh, Elliott, and Stock (1995), and find some evidence for predictability at short horizons but not at long horizons, rather contrary to conventional wisdom. Valkanov (2003) also shows that when appropriate testing procedures are employed, the evidence of return predictability is not as strong as previously maintained. Using a more conservative test procedure, Lanne (2002) finds no evidence that stock returns can be predicted by any highly persistent variable. While primarily focused on tests for structural breaks, Viceira (1997) concludes that under conventional asymptotics, there is some evidence of predictability in one-month returns from the dividend yield, but there is none under local to unity asymptotics. Thus, the reported evidence of predictability may simply follow from a neglected near unit root problem, once the uncertainty
of the predictive variable’s order of integration is properly incorporated in test statistics, the evidence for return predictability weakens considerably.

**Data Snooping, Out-of-Sample Evidence, Survivorship, and Structural Breaks** Finally, Lo and MacKinlay (1990), Foster, Smith, and Whaley (1997), and Ferson, Sarkissian, and Simin (2003), among others, show that repeated visits of the same dataset lead to a problem that statisticians refer to as model overfitting, i.e., the tendency to discover spurious relationships when applying tests that are inspired by evidence from previous work. Lo and MacKinlay (1990) call this “data snooping”. These studies point out that a specification search over even a small number of predictive variables can seriously bias standard procedures for drawing inferences – where the implicit assumption is that only one test is made with a particular dataset – and may produce spurious predictability that is of the same magnitude as that typically reported. Indeed, because the identity of the predictive variables is unknown (possibly with the exception of the dividend-price ratio, as discussed above), data snooping must be a serious concern as well. Moreover, as is shown by Ferson, Sarkissian, and Simin (2003), highly persistent series are more likely to be found significant in the search for predictive variables, implying a kind of spurious regression bias, even outside the classic setting of Yule (1926) and Granger and Newbold (1974). Consequently, if past work uses similar data, the possibility of decades of data snooping could cloud any inference regarding stock market predictability further.

Of course, the unfortunate consequences of data snooping might be mitigated or even eliminated by out-of-sample tests. First, some gain comfort in knowing that the evidence for predictability extends to other markets internationally (e.g., Solnik, 1993; Bossaerts and Hillion, 1999; Neely and Weller, 1999; Ang and Bekaert, 2003). However, whether or not this is effective depends on the extent to which the new data is truly out of sample. If returns on the other markets are correlated with those that have been previously mined, one might derive false confidence from tests of this nature (Foster, Smith, and Whaley, 1997; Kirby, 1998). Still, the predictability literature has focused almost exclusively on in-sample evidence (and the U.S. stock market). Worse, the rare exceptions such as Bossaerts and Hillion (1999), Neely and Weller (1999), and particularly Goyal and Welch (2003) generally conclude that even the best prediction models have no out-of-sample forecasting power and fail to generate robust results that outperform simple unconditional benchmark models!

A second way to deal with data snooping is by examining longer time series. Goetzmann and Jorion (1995) use two new time series beginning in 1871, a monthly series for the U.S. and an annual series for the U.K. stock market, and perform dividend yield tests on those. Over the entire time period, dividend yields display only marginal ability to predict stock market returns in either country. However, in the “modern” era (1926-1992), the U.K. market displays strong evidence of predictability. Yet, a closer look at the data indicates that
most of the effect is due to the extraordinary crash and rebound in the early 70’s: 1975 was such an extreme positive outlier that, amplified by the strongly negative years in 1973 and 1974, it must have pulled up the regression line. Of course, this does not necessarily mean that this effect is economically meaningless, however. If the dividend yield can forecast a market rebound in times of extreme economic distress and political upheaval, then it clearly has value to risk-averse investors. But detecting outliers is not the main argument in Goetzmann and Jorion (1995). They rather argue that tests over long time periods may be affected by survivorship, akin to the “peso problem”. Had the market closed in the early 70’s, and remained shut for an appreciable length of time (as for example in Portugal), one might not be using it to study dividend yield regressions. The question of whether dividend yields predict stock returns in the U.K. thus ultimately rests on the chances of market closure at that critical juncture. As a result, the predictive power of dividend yields is strongly conditioned on survival. Indeed, Goetzmann and Jorion’s (1995) simulations of market disappearance show that regression statistics based on a sample drawn solely from surviving markets can be seriously biased toward finding return predictability. In addition, the biases seem to grow with the length of the horizon; differences in p-values are systematically higher for 5-year β’s than for 1-year β’s. Overall, their results show that using long time series might not be the suggested remedy for data snooping, but instead can seriously bias inference when survivorship is an issue.

Similarly, since the inference about return predictability has often been shown to depend on the sample period used to estimate the test statistics, it is not only the issue of survivorship that inevitably arises over long horizons, but also the risk of structural breaks. Viceira (1997) develops asymptotic results on the distribution of recursive, reverse-recursive and sequential least squares estimators under general conditions of heteroskedasticity and autocorrelation, and shows that structural breaks possibly invalidate any inference based on full sample estimators.

The dilemma is obvious. On the one hand, the use of long-horizon regressions and, in particular, the use of long time series are clearly recommendable. On the other hand, however, survivorship and structural breaks are likely to overturn the advantages of using those historical time series.

Rational Asset Pricing and Restrictions on Predictability It has already been argued that return predictability can be interpreted only in conjunction with an intertemporal equilibrium model of the economy. However,

---

5 Interestingly, Goyal and Welch (2003) find similar results for the U.S. stock market.

6 Generally, peso problems can be interpreted as a failure of the paradigm of rational expectations econometrics, which requires that the ex-post distribution of endogenous variables be a good approximation to the ex-ante distribution that agents think may happen. The failure may not be that of the economic agent, but that of the econometrician, who only analyzes series with continuous histories. Unusual events with a low probability of occurrence but severe effects on prices, such as wars or nationalizations, are not likely to be well represented in samples and may be totally omitted from survived series.
the choice of a particular model for the time-varying behavior of expected returns is, by nature, somewhat arbitrary. At best, an equilibrium model should specify both the stochastic process for and the underlying economic determinants of expected returns. But existing asset pricing theories do hardly ever specify any particular a priori restrictions on the variation through time in expected returns. For example, one reasonable restriction is that equilibrium in an efficient market never implies predictable price declines (negative expected nominal or even real returns). Yet, even the models by Merton (1973), Lucas (1978), Breeden (1979), and Cox, Ingersoll, and Ross (1985) do not rule out negative expected returns. At least, there are good reasons to think that expected stock returns may be persistent. In the Gordon growth model, for example, both investment and dividends rise when earnings increase. In practice, investment is generally a sequential process that involves various stages of planning and execution over time. Once the initial costs are sunk, the firm often has an economic incentive to follow through with additional investment, even in the face of lowered expectations about future benefits. As a result, changes in investment should have a significant predictable component. Asset pricing models like the consumption model of Lucas (1978) describe expected stock returns as functions of expected economic growth rates. Merton (1973) and Cox, Ingersoll, and Ross (1985) propose real interest rates as candidate state variables, driving expected returns in intertemporal models. Such variables are likely to be highly persistent.

Consequently, already the early empirical studies for stock return dynamics by Conrad and Kaul (1988), Fama and French (1988b), Lo and MacKinlay (1988), or Huberman and Kandel (1990) involve persistent, autoregressive expected returns. The standard regression specification in equations (1), (2), and (16) with $\delta \approx 1$ is only a natural consequence in this regard, and, as shown for example in Campbell (1991), Campbell, Lo, and MacKinlay (1997, Ch. 7), and Cochrane (2001, Ch. 20), sufficient (and, in this setting, necessary) to explain the “excess volatility” puzzle of Shiller (1981) and LeRoy and Porter (1981), reviewed for example in Cochrane (1991) and Schwert (1991). Persistence in the expected return process increases the variability of realized returns, for small but persistent changes in expected returns have large effects on prices and thus on realized returns.\footnote{In general, if expected returns are represented by such a process, realized returns can be characterized by an ARMA(1,1) process. Accordingly, autocorrelated expected returns and the opposite response of prices to expected return shocks (the “discount-rate effect”, Fama and French, 1988a) can combine to produce mean-reverting components of stock prices. Fama and French (1988b) and Poterba and Summers (1988) show that mean-reverting price components tend to induce negative autocorrelation in long-horizon returns. But a mean-reverting, positively autocorrelated expected return does not necessarily imply negative autocorrelated returns or a mean-reverting component of prices! If shocks to expected returns and expected dividends are positively correlated, the opposite response of prices to expected return shocks can disappear. In this case, the positive autocorrelation of expected returns will imply positively autocorrelated returns, and time-varying expected returns will not generate mean-reverting price components (the MA part stems from “news about future dividends” (Campbell, 1991); if the roots cancel, realized returns are white noise). However, as Cochrane (2001, Ch. 20) puts it, any positive correlation between dividend growth and expected return...}
Moreover, since almost all of the empirical work on predicting returns takes the form of regression analysis, it would be easier to interpret the evidence on predictability if one could show how asset pricing theory restricts the parameters of linear forecasting models. Hansen and Singleton (1983) and particularly Kirby (1998) show that asset pricing models imply that the parameters in predictive regressions must take on certain values and thus restrict the regression-based criteria commonly used to measure return predictability. These restrictions then provide a way to directly assess whether the predictability uncovered using simple regression analysis is consistent with rational asset pricing.

Hansen and Singleton (1983) conduct the first comprehensive study of rational pricing and return predictability. They study the time-series behavior of asset returns and aggregate consumption. If agents effect their consumption plans by trading shares of ownership of firms in a competitive stock market, an implication of this trading is that the serial correlation properties of stock returns are intimately related to the stochastic properties of consumption and the degree of risk aversion of investors.

First, in general equilibrium models of stock price behavior with risk-neutral agents (i.e., linear utility), share prices will be set so that the expected return on each asset is constant. Thus, asset returns will be serially uncorrelated and, in particular, past values of consumption will be uncorrelated with current-period asset returns. Second, if agents are risk averse, then the temporal covariance structure of consumption and asset returns will be non-trivial, except under very strong restrictions on the underlying production technology (Rubinstein, 1976).

Based on a fairly standard framework (production-exchange single-good economy of identical consumers, time-additive von Neumann-Morgenstern utility, lognormal joint distribution of consumption and returns, constant relative risk aversion (CRRA) preferences), Hansen and Singleton (1983) show that these restrictions imply that the predictable components of log asset returns are proportional to the predictable component of the change in log consumption, with the proportionality factor being minus the coefficient of relative risk aversion.

In a single-good economy of identical consumers, whose utility functions are of the CRRA type

\[ U(C_t) = \frac{1}{1-\zeta}C_t^{1-\zeta}, \]

where \( \zeta \) denotes the coefficient of relative risk aversion, \( \zeta \geq 0 \), and where the representative consumer is assumed to choose a stochastic consumption plan so

shocks is difficult to reconcile with the business cycle, consumption smoothing explanation of a time-varying risk premium. If anything, since expected returns are assumed to rise in “bad times” when risk or risk aversion increases, one should see a positive shock to expected returns associated with a negative shock to current or future dividend growth! Of course, changes through time in the autocovariance of expected returns, or in the relation between shocks to expected returns and expected dividends, can change the time-series properties of returns and obscure tests of forecast power based on autocovariance (e.g., Campbell, 1991, 2001; Campbell, Lo, and MacKinlay, 1997, Ch. 7; Cochrane, 2001, Ch. 20).
as to maximize the expected value of his time-additive utility function,

$$E_0 \sum_{t=0}^{\infty} \tau^t U(C_t),$$

the first-order necessary condition is given by

$$E_t \left( \frac{U'(C_{t+1})}{U'(C_t)} (1 + R_{t+1}) \right) = E_t \left( \tau \left( \frac{C_{t+1}}{C_t} \right)^{-\zeta} (1 + R_{t+1}) \right) = 1.$$ 

The stochastic discount factor is thus given by

$$m_{t+1} = \tau \left( \frac{C_{t+1}}{C_t} \right)^{-\zeta}.$$ 

By defining the information set \( \{ \Psi_{t-s} : s \geq 0 \} \), denoted by \( \psi_t, \Psi_t \equiv (C_t, R_t) \), and assuming that the joint distribution of consumption and returns is conditionally lognormal with constant variance,

$$\ln \left( \left( \frac{C_{t+1}}{C_t} \right)^{-\zeta} (1 + R_{t+1}) \right) \equiv \ln (CR_{t+1}) \sim N(\mu_t, \sigma^2),$$

where \( \mu_t \equiv E(\ln (CR_{t+1}) | \psi_t) \),

$$E_t (CR_{t+1}) \equiv E(CR_{t+1} | \psi_t) = e^{\left( \mu_t + \frac{1}{2} \sigma^2 \right)} = 1/\tau,$$

and therefore

$$\mu_t = -\ln (\tau) - \frac{1}{2} \sigma^2.$$ 

Thus,

$$\ln (CR_{t+1}) - \mu_t = -\zeta \Delta c_{t+1} + r_{t+1} + \ln (\tau) + \frac{1}{2} \sigma^2,$$

with \( \Delta c_{t+1} \equiv c_{t+1} - c_t \) denoting log consumption growth.

Since \( E_t (\ln (CR_{t+1}) - \mu_t) = 0 \) by definition,

$$E_t (r_{t+1}) + \frac{1}{2} \sigma^2 = \zeta E_t (\Delta c_{t+1}) - \ln (\tau)^8$$

Equation (17) summarizes the relationships among serial correlation of consumption, the level of risk aversion, and serial correlation of asset returns implied by the first-order condition. Risk neutrality, for example, corresponds to

\[\text{Campbell (2000) and Campbell and Viceira (2002, Ch. 2) derive logarithmic versions of the stochastic discount factor equations in lognormal intertemporal equilibrium models. By assuming that the pricing kernel and asset returns are conditionally lognormal, it follows that, in general, the expected excess return on a risky asset is given by } E_t (r_{t+1}) - r_{f,t+1} + 1/2\sigma_t^2 = -\zeta \text{Cov}_t (\ln (m_{t+1}), r_{t+1}). \text{ In the power utility case of constant relative risk aversion, it is easy to show that } E_t (r_{t+1}) - r_{f,t+1} + 1/2\sigma_t^2 = \zeta \text{Cov}_t (\Delta c_{t+1}, r_{t+1}). \text{ In equilibrium, the expected excess return on a risky asset must equal risk aversion } \zeta \text{ times the conditional covariance of the asset return with log consumption growth.}\]
the case of $\zeta = 0$, which implies that $r_{t+1}$ is equal to a constant and serially uncorrelated.

After all, to translate these observations into statements about predictability of asset returns, Hansen and Singleton (1983) derive the following expression for the coefficient of determination, $R^2$, from the projection of $r_{t+1}$ onto $\psi_t$ implied by equation (17),

$$R^2 = \frac{\zeta^2 \text{Var}(E_t(\Delta c_{t+1}))}{\text{Var}(E_t(r_{t+1})) + \zeta^2 \text{Var}(E_t(\Delta c_{t+1}))}.$$  (18)

From equation (18) it follows immediately that a necessary condition for asset returns to have predictable components is that agents be risk averse ($\zeta \neq 0$). Risk aversion, however, is not a sufficient condition for predictability. For the special case where the projection $E_t(\Delta c_{t+1})$ is constant, the $R^2$ is equal to zero or, equivalently, the projection of $r_{t+1}$ onto $\psi_t$ is constant. When there are non-trivial predictable components in $\Delta c_{t+1}$ and $\zeta \neq 0$, then (real) asset returns will also have predictable components. Under their model, thus, the predictable component of asset returns is proportional to the predictable component of consumption growth.

Kirby (1998) develops a more general approach for studying predictability within a regression setting and generalizes the framework of Hansen and Singleton (1983), which is largely based on the strong assumptions of constant relative risk aversion and joint lognormality of asset returns and aggregate consumption growth. He derives restrictions that apply quite generally and can therefore aid in understanding the role that predictability plays in the econometric evaluation of asset pricing models. In particular, they make it possible to construct the linear specification for predicting asset returns that is implied by the asset pricing model under consideration. The forecasts generated in this manner provide a common frame of reference for studying the ability of different asset pricing models to explain the predictable variation in returns. It might be the case, for example, that the variance of the forecasts associated with one model is too small, while that for another model is too large.

The general conditional asset pricing relation

$$E_t(m_{t+1} (1 + R_{t+1})) = 1$$

establishes a necessary condition for the market to be intertemporally efficient. It implies that, in an efficient market, the stochastic process defined by the

---

9 Ferson and Harvey (1991) also look at the issue of return predictability and rational asset pricing in a regression setting. They use a multi-beta CAPM to decompose the variance of the fitted values from a regression of returns on a set of instrumental variables into explained and unexplained components. If the ratio of explained to unexplained variance is large, they argue, then this is evidence that predictability is driven by rational variation of expected returns. They find that this ratio is reasonably close to one and consequently conclude that the ability to predict returns is, for the most part, a product of rational market forces. However, their methodology is subject to errors-in-variables biases and does not impose the full set of restrictions implied by their pricing model. Their cross-sectional estimation approach leaves the intercept and slope coefficients unrestricted. As a result, their variance-ratio tests may have low power to detect violations of the restrictions implied by the pricing model.
product of the total return on an asset and the SDF must be unpredictable by information used in the conditional expectation. But this does not require that the returns themselves be unpredictable. In general, the only circumstance in which it rules out the ability to predict returns is when the pricing kernel is equal to a constant – when risk-neutral pricing prevails.

Although restrictions could be developed in terms of $R_{t+1}$, the analysis is more tractable if excess returns, defined as $R_{t+1}^e = R_{t+1} - R_{f,t+1}$ (slightly different than in equation (10), once again!), are used, and if the normalization $\tilde{m}_{t+1} = m_{t+1}/E(m_{t+1})$ is imposed. This yields

$$E_t(\tilde{m}_{t+1} R_{t+1}^e) = 0$$

with $E(\tilde{m}_{t+1}) = 1$. Taking any predictive variable, $x_t$, multiplying both sides of equation (19) by $x_t$,$$E_t(\tilde{m}_{t+1} R_{t+1}^e x_t) = 0,$$or

$$E(\tilde{m}_{t+1} R_{t+1}^e x_t) = 0$$

by applying the law of iterated expectations. To see that equation (20) places tight restrictions on the ability to predict excess returns, note that the unconditional analogue of equation (19) implies the pricing relation

$$E(R_{t+1}^e) = -Cov(\tilde{m}_{t+1}, R_{t+1}^e).$$

We can thus use the familiar covariance decomposition to write equation (20) as

$$Cov(R_{t+1}^e, x_t) = -E(R_{t+1}^e) E(x_t) - Cov(\tilde{m}_{t+1}, R_{t+1}^e x_t) = -Cov(\tilde{m}_{t+1}, R_{t+1}^e (x_t - E(x_t))).$$

This equation provides a general characterization of the restrictions on predictability that must be satisfied in an efficient market. It indicates that the ability to predict $R_{t+1}^e$, as measured by its unconditional covariance with $x_t$, has a well-defined theoretical value.\(^{10}\)

By now, the restrictions shown in equation (21) are not directly applicable to most of the research on predicting asset returns, namely, to fit a linear regression model to the data. Nevertheless, the analysis easily generalizes so that it applies to regression-based measures of return predictability, as in a multivariate generalization of equation (1),

$$R_{t+1}^e = z_t^t / \beta + \xi_{t+1},$$

where $z_t = (x_t^t)'. From least-squares theory, the vector of regression coefficients in equation (22) is given by

$$\beta = (z_t z_t')^{-1} E(R_{t+1}^e z_t).$$

\(^{10}\)The covariance between $R_{t+1}^e$ and $x_t$ is the expected excess payoff an a dynamic trading strategy that exploits the information conveyed by the realization of $(x_t - E(x_t))$ (Hansen and Richard, 1987).
and from equation (20), the pricing relation

\[ E(R_{t+1}^e z_t) = -\text{Cov}(\tilde{m}_{t+1}, R_{t+1}^e z_t) \]

gains. Thus,

\[ \beta = -E(z_t z_t')^{-1} \text{Cov}(\tilde{m}_{t+1}, R_{t+1}^e z_t). \quad (23) \]

If predictability is rational, then the regression coefficients must satisfy the restrictions shown in equation (23).

A similar approach can be used to derive restrictions on the value of the regression \( R^2 \). If \( \beta \) is partitioned as \((\beta_0 \beta_x)^t\), where \( \beta_0 \) is the intercept and \( \beta_x \) the vector of slope coefficients, then the population \( R^2 \) for the model can be written as

\[ R^2 = \left( \frac{\beta_x' \Sigma_{xx} \beta_x}{\text{Var} (R_{t+1}^e)} \right) = \left( \frac{\sigma_{m,R^e x} \Sigma_{xx}^{-1} \sigma_{m,R^e x}}{\text{Var} (R_{t+1}^e)} \right), \quad (24) \]

where \( \sigma_{m,R^e x} = \text{Cov}(\tilde{m}_{t+1}, R_{t+1}^e (x_t - E(x_t))) \) and \( \Sigma_{xx} \) denotes the variance-covariance matrix of \( x_t \). This is the only value of the population \( R^2 \) that is consistent with rational pricing!

These restrictions make it easy to assess whether the evidence of predictability uncovered using regression analysis is consistent with any given specification of the stochastic discount factor. If it is, then the view that predictability is driven by rational variation in expected returns gains credibility. If not, alternative mechanisms, like market inefficiencies or collective data mining, may explain the observed ability to predict returns.

Kirby’s (1998) empirical analysis looks at whether the observed predictability is consistent with the implications of the following well-known asset pricing models: the utility-based models of Lucas (1978), Abel (1990), and Epstein and Zin (1991), simple conditional versions of the Sharpe (1964) and Lintner (1965) CAPM and the Fama and French (1993) three-factor model.

The empirical tests reveal that excess returns are generally too predictable to be compatible with those asset pricing models; none of the pricing models can generate the degree of predictability that is observed in the data. There is nothing in the results to suggest that the power utility model can explain the observed return predictability. Like the power utility specification, the habit persistence model also fails to generate the return predictability that shows up in the unrestricted regressions. In the same way, it seems that breaking the link between risk aversion and intertemporal elasticity of substitution does little to improve the poor fit of the consumption-based models. Although these findings for the consumption-based models may seem unremarkable in light of the statistical rejections of these models reported in the asset pricing literature, they do serve to illustrate that predictability almost certainly plays a key role in these rejections. The low values of the regression \( R^2 \) generated by the consumption-based specifications point to a lack of predictability as the driving force behind these rejections.
Ferson and Harvey (1991) and Ferson and Korajczyk (1995) argue that conditional beta pricing models explain a large fraction of the predictable variation in portfolio returns. The evidence in Kirby (1998) is generally consistent with their results. But it may be premature to conclude that these models truly explain return predictability. In conditional beta models, return predictability is driven by predictable variation in both the betas and the factors. So, if one or more of the factors is the return on a portfolio, we know before we conduct the empirical tests that the model will almost certainly imply some level of predictability. This is an unavoidable consequence of the contemporaneous correlation in returns. The real question, however, is not whether conditional beta pricing models capture a substantial fraction of the predictable variation in returns, but whether they explain cross-sectional differences in return predictability. A pricing model that uses portfolio returns as factors may appear to explain a substantial fraction of the predictable variation in returns, even when this predictability is due to collective data mining.

Overall, thus, there are a number of reasons why the empirical results of stock market predictability should be regarded with caution. Stock market predictability is still a very broad and active research topic, and it is impossible to provide a complete survey in just a few pages. The main difficulty with understanding the rather large literature on predictability is the sheer variety of test procedures that have been proposed. The strength and usefulness of many results are still hotly debated – consistently evaluating their significance is thus no easy task.

2 Review of Current Research

This section surveys four recent papers in more depth: Valkanov (2003), Ferson, Sarkissian, and Simin (2003), Ang and Bekaert (2003), and Goyal and Welch (2003). They all critically address the statistical methodologies and conclusions in the stock market predictability literature and are therefore representative for the above-mentioned statistical view. They show that findings against the constant expected excess return hypothesis based on standard statistical inference can appear much more significant than they really are, and hence are not real, least of all out of sample.

Valkanov (2003): Long-horizon Regressions: Theoretical Results and Applications

It has already been noted that a definite, analytical answer to the question of whether long- or short-horizon regressions have greater power and/or whether the more significant long-horizon results are a mere product of size distortion, has not yet been satisfactorily provided (Campbell, 2001, is probably the most rigorous attempt). In contrast to the studies based on simulation methods, which either suggest power gains (e.g., Hodrick, 1992) or show size distortions (e.g., Nelson and Kim, 1993; Goetzmann and Jorion, 1993), but
never systematically analyze the econometric properties of the statistics, Valkanov (2003) addresses this question analytically. Using asymptotic arguments, he shows that the $t$-statistics in long-horizon regressions do not converge to well-defined distributions with adequate power and size. In some cases, moreover, the ordinary least squares estimator is not consistent and the $R^2$ is an inadequate measure of the goodness of fit. Indeed, he demonstrates that long-horizon regressions will always produce “significant” results in finite samples, whether or not there is a structural relation between the underlying variables. Accordingly, these findings can explain the tendency of long-horizon regressions to find significant results where previous short-term approaches find none (or little).

To understand this conclusion, recall that the long-horizon return is a rolling sum of the original (stationary) return series. Yet, in a rolling summation of a series integrated of order zero (i.e., $I(0)$), the new long-horizon variable behaves asymptotically as a (non-stationary) series integrated of order one (i.e., $I(1)$). Such persistent stochastic behaviour will be observed whenever the regressor, the regressand, or both are obtained by summing over a non-trivial fraction of the sample. Based on this insight, Valkanov uses the Functional Central Limit Theorem (FCLT) (e.g., Stock, 1994) to derive accurate approximations of the small-sample distributions of the OLS estimator of the slope coefficient, its $t$-statistic and the coefficient of determination $R^2$. These analytical results yield exact rates of convergence or divergence and permit to modify the statistics in order to conduct correctly sized tests. They are basically broad enough to accommodate predictive variables that have a near unit root, and that the recent work of Lanne (2002) and Torous, Valkanov, and Yan (2002) can be considered as special cases of those. To obtain $t$-statistics that have a well-defined distribution, he then proposes a rescaled $t$-statistic, $t/\sqrt{T}$, whose asymptotic distribution is, although non-normal, easy to simulate. Appropriately rescaled, then, Valkanov demonstrates that long-horizon regressions have a somewhat better power at rejecting alternatives than their short-horizon analogues. Still, in general, a significant $R^2$ in such regressions cannot be interpreted as an indication of a good fit.

Finally, Valkanov applies his framework to explain the empirical and simulation results obtained by the excess return/dividend yield regressions in Fama and French (1988a), Campbell and Shiller (1988b), Hodrick (1992), Goetzmann and Jorion (1993), and Nelson and Kim (1993) for the U.S. stock market. His results stand in stark contrast to those papers. Using the entire sample period, 1927-1999, he cannot reject the null hypothesis of no predictability at usual levels of significance for any of the horizons. Looking at various subsamples, the lack of predictability seems to come mainly from the 1927-1945 and 1982-1999 periods. If there is any evidence of return predictability, it is strongest during the 1946-1980 sample (as in Goyal and Welch, 2003!). Only during that period, the evidence of predictability is statistically significant at the 5% level.

Valkanov’s main conclusion is that when appropriate testing procedures are employed, the evidence of return predictability is not as strong as previously maintained and tests in long-horizon studies should thus be carefully re-evaluated.
Ferson, Sarkissian, and Simin (2003): Spurious Regressions in Financial Economics?  In contrast to Valkanov (2003), the framework of Ferson, Sarkissian, and Simin (2003) is based on short-horizon regressions similar to equations (1) and (2). They focus on two issues.

The first is spurious regression, related to the classic studies of Yule (1926) and Granger and Newbold (1974). These studies warned that spurious relations may be found between the levels of non-stationary time series that are actually independent. However, in predictive regressions, the dependent variable is the (one-period) excess return, which is stationary and not highly persistent. Thus, one might think that spurious regression problems are unlikely. Indeed, when there is no persistence in the true expected return, the spurious regression phenomenon is not a concern, even when the measured predictive variable is highly persistent. This implies that spurious regression is not a problem from the perspective of testing the null hypothesis that expected returns are unpredictable, even if a highly autocorrelated predictive variable is used. Yet, under the alternative hypothesis of stock market predictability, stock market returns may be considered to be the sum of an unobserved time-varying expected return, plus unpredictable noise. If the true underlying expected return is a highly persistent autoregressive time series, Ferson, Sarkissian, and Simin argue that spurious regression can be a serious concern well outside the classic setting of Yule (1926) and Granger and Newbold (1974).

The second issue is data snooping. Ferson, Sarkissian, and Simin show that the spurious regression and data mining effects reinforce each other. If researchers have mined the data for predictive variables that produce high $R^2$'s in predictive regressions, the mining is more likely to uncover the spurious, persistent predictive variables. Indeed, the standard predictive variables in the literature tend to be highly autocorrelated, as expected if they result from a spurious mining process. In the same way, their simulations suggest that many of the predictive regressions in the literature, based on individual predictive variables, may be spurious.

Finally, Ferson, Sarkissian, and Simin critically review the univariate regressions in nine of the major predictability studies (including studies such as Fama and French, 1988a, 1989, and Lettau and Ludvigson, 2001). Accounting for spurious regression bias, they find that 7 of the 17 $t$-statistics and $R^2$s from those studies that would be significant by traditional standard criteria are no longer significant. They therefore call into question the validity of some of the predictive variables identified in the literature. In general, they emphasize that the pattern of evidence for the predictive variables in the literature is similar to what is expected under a spurious mining process with an underlying persistent expected return. In this case one would expect predictive variables to arise, then fail to work out of sample. With fresh data, new predictive variables arise, then fail. The dividend yield rose to prominence in the 80's, but seems to fail to work in post-1990 data (e.g., Goyal and Welch, 2003; Schwert, 2003). With fresh data, new variables appear to work, for example Lettau and Ludvigson’s (2001) consumption-wealth ratio. Ferson, Sarkissian, and Simin’s conclusions are thus twofold. First, one should be concerned that these new variables are
likely to fail out of sample. Second, any stylized facts based on empirically motivated variables and asset pricing tests based on those should be viewed with skepticism!

Ang and Bekaert (2003): Stock Return Predictability: Is it There?
In a rational no-bubble model, dividend-price ratios reflect the expected value (of a basically non-linear function) of future dividend growth and discount rates. Discount rates consist of a risk-free and risk premium component. As indicated in equation (8), the variation of dividend-price ratios can thus be attributed to the variation of expected future excess returns, expected future dividend growth, or expected future interest rates. Ang and Bekaert (2003) is one of the few analyses that is not limited to predictable components in expected future excess returns. They examine the fraction of the variation in dividend-price ratios that reflects the predictability of dividend growth and interest rates as well.

Ang and Bekaert derive a structural model of equity premiums based on dividend-price ratios, price-earnings ratios, and interest rates. After carefully accounting for small sample properties of the standard tests and including joint tests across horizons, they show that at long horizons, excess return predictability by the dividend-price ratio is not statistically significant, not robust across countries and not robust across different sample periods. When they pool their regressions across countries, only the short-term interest rate and the dividend-price ratio jointly predict excess returns — at short horizons and in a bivariate regression framework. In this sense, they conclude that the predictability that has been the focus of most finance research is simply not there. Instead, Ang and Bekaert find that the dividend-price ratio significantly predict interest rates, but, consistent with studies like Cochrane (1992, 1997, 2001, Ch. 20), only weakly predict future dividend growth. Moreover, they show that the most robust predictive variable for excess returns is the short-term interest rate, but it is significant only at short horizons. Finally, in finite samples, using dividends may be problematic because they are often manipulated, smoothed, or set to zero, making them potentially poor indicators of the true value-relevant cash flows in the future (e.g., Campbell, 2000). One obvious way to increase the information set is to use (more variable) earnings. Lamont (1998) argues that the price-earnings ratio has independent predictive power for excess returns in addition to the dividend-price ratio. However, when Ang and Bekaert examine the predictive power of the price-earnings ratio for both returns and dividend growth, they find only weak evidence for Lamont’s (1998) excess return predictability results. Instead, they detect a strong role for the price-earnings ratio as a predictive instrument for future dividend growth.

To examine the size and power properties of their tests, they perform a Monte Carlo analysis. In contrast to previous studies that usually rely on linear or VAR systems that fail to fully capture the non-linear dynamics of dividend-price ratios implied by a dynamic present value model, and ignore the cointegration between dividends and price levels that characterizes rational pricing, Ang
and Bekaert explicitly impose the cointegration relation between dividends and prices and their solution is exact. In addition, because both the return and the dividend-price ratio are non-linear processes in their model and they consider multivariate regressions, the standard upward bias result deriving from the negative correlation between return and dividend-price ratio innovations and the persistence of the latter (as described above) no longer holds.

After all, Ang and Bekaert suggest a re-focus of the predictability debate in four directions. First, their results suggest that predictability is mainly a short-horizon, not a long-horizon, phenomenon. The predictive ability of the dividend-price ratio is best seen in a bivariate regression with short-term interest rates only at short horizons. Second, the strongest predictability comes from the short-term interest rate rather than from the dividend-price ratio. Third, the dividend-price ratio strongly predicts future interest rates. Finally, dividend-price ratios and price-earnings ratios have good predictive power for future dividend growth, but not for excess returns. Hence, a potentially important source of variation in price-earnings and dividend-price ratios is the predictable component in dividend growth. Their results thus generally imply that univariate linear models of expected returns are unlikely to satisfactorily capture all the predictable components in returns. A challenge of future work will thus be to create a present value model with sophisticated dynamics for earnings growth, payout ratios and dividend growth to match Ang and Bekaert’s evidence.

Goyal and Welch (2003): Predicting the Equity Premium with Dividend Ratios

This is probably the most impressive paper. Their findings are not just a matter of quibbling over proper statistical methods to compute well-behaved test statistics. They are much more basic.

Goyal and Welch (2003) suggest a simple, recursive residuals (out-of-sample) graphical approach to evaluate and diagnose the predictive power of the standard equity premium predictive regression given in equation (1). Their approach makes it very easy to understand the relative performance of different forecasting models. Plotting the cumulative sum-squared error from the unconditional model minus the cumulative sum-squared error from the conditional predictive regression model, a positive value indicates that the conditional model has outperformed the unconditional model so far. A positive (negative) slope indicates that the conditional model had a lower (higher) forecasting error than the unconditional model in a given time period. Furthermore, their out-of-sample diagnostic differs from the common in-sample tests in that the predictive regressions are themselves estimated only with then-available data.

In their empirical application based on U.S. dividend yields and dividend-price ratios, Goyal and Welch examine how the dividend-ratio regressions perform when compared against the unconditional equity mean. Both the “conditional dividend-ratio models” and the “unconditional historical equity premium model” (the prevailing simple moving average) are estimated as rolling forecasts to predict the one-year-ahead equity premium.

The results indicate that the presumed predictive ability of the dividend ra-
tios is a mirage, apparent even before the 90's: they practically never seemed to have outperformed the unconditional forecasts. Despite some in-sample forecasting power prior to 1990, Goyal and Welch show that the dividend ratios had poor out-of-sample predictive ability even then. Further, their diagnostic illustrates over what time periods one might imagine finding predictive ability, and makes it immediately obvious that the dividend ratios only had two really good predictive years prior to the 90’s, 1973 and 1974. They thus conclude that the evidence that the equity premium has ever varied predictably with past dividend ratios has always been tenuous. Despite their best attempts, they could not detect any robust out-of-sample predictive ability of the standard dividend-ratio models in any variation. They argue that the primary source of this poor predictive ability is parameter instability: in-sample performance is no guarantee of out-of-sample performance!

Of course, a natural question is why for example Fama and French (1988a) come to different conclusions. The reason is their sample period. Over their period, the slope of the line is sufficiently positive to give the dividend ratios an edge. However, extending the test period forward or backward yields the different conclusions of Goyal and Welch (and others, e.g., Bossaerts and Hillion, 1999; Neely and Weller (1999); Schwert, 2003).

In addition, Goyal and Welch also try numerous variations. But none of these variations impact their conclusion that the out-of-sample performance has always been poor. They try reinvesting the dividends, instead of summing them. Because the dividend ratio has a near unit root, they try changes in dividend ratios. These changes in dividend ratios performed worse in forecasting than the dividend ratios themselves. They try simple returns and yields, instead of log returns and yields. Although annual horizons seem to have been generally agreed to have the least statistical problems, they try predicting on different horizons (monthly, quarterly, multi-yearly). But under no frequency they find the dividend ratio models to outperform at a halfway statistically or economically significant manner. They try different “fixed number of years” estimation windows. The unconditional model typically performs better or as well as the dividend ratio models if 5 or more years are used for parameter estimation. They try forecasting with the Stambaugh (1999) correction for high serial correlation in the dividend ratios. This worsens the out-of-sample performance. Earnings-price, earnings-payout ratios (Lamont, 1998) or more complex measures based on analysts’ forecasts (Lee, Myers, and Swaminathan, 1999) similarly do not appear to predict equity premiums well out of sample. Finally, they try the risk-free rate as predictive variable. They find some in-sample predictive ability on short frequencies (one-month to one-quarter), but little in-sample predictive ability on longer frequencies (one-year). In any case, the out-of-sample predictive ability on annual horizons is considerably worse than the unconditional mean equity premium. Again, they do not believe there is much predictive ability coming from the short-term interest, either. In sum, thus, variations on the specification and variables do not produce instances that would lead one to believe that dividend ratios or other variables can predict equity premiums in a meaningful way. In general, the conditional predictive regression models predict
worse than the prevailing unconditional equity premium.

In any case, Goyal and Welch conclude that there has never been convincing evidence that dividend ratios were ever useful in predicting for investment purposes, even prior to the 90’s! Neither the dividend yield nor the dividend-price ratio had both the in-sample and out-of-sample performance that should have lead one to believe that it could outperform the simple prevailing equity premium average in a statistically or economically significant manner. A naive market-timing trader who just assumed that the equity premium was “like it has always been” would typically have outperformed a trader who employed dividend-ratio predictive regressions!

3 Conclusion and Implications for Asset Management

Recent advances in asset pricing theory and the mounting empirical evidence of stock market predictability seem to have persuaded the majority of researchers to abandon the constant expected returns paradigm. The time variation and predictability of excess returns, labeled as a new fact in finance by Cochrane (1999a), is so widely accepted in the profession that it has generated, with no signs of subsiding, a new wave of conditional asset pricing models and models that try to analyze the implications of return predictability on portfolio decisions (e.g., Brennan, Schwartz, and Lagnado, 1997; Campbell and Viceira, 1999, 2002; Barberis, 2000; Xia, 2001).

Even though, certain aspects of the empirical research on predicting returns remain controversial and should be regarded with caution. Loosely, most of the findings reviewed above suggest that standard statistical inference overreject the null hypothesis of no predictability, so that an increase in sample $R^2$'s at long horizons may has more to do with poorly behaved test statistics than with stock market predictability. Properly adjusting for small sample biases, near unit roots, and other statistical issues associated primarily with long-horizon regressions weakens and often reverses many of the previous inferences. Even worse, current studies such as Bossaerts and Hillion (1999), Goyal and Welch (2003), and Schwert (2003) conclude that the out-of-sample predictive ability of the dividend ratios and other predictive variables is abysmal and, in the words of Schwert (2003), disastrous. Overall, thus, the evidence for stock market predictability is much less transparent than previous work may have suggested!

While the strength and usefulness of these critical results are still hotly debated, they nevertheless raise questions about the implications of these findings for the way academics and practitioners use financial theory.

Conditional Asset Pricing The evidence of time-varying expected returns has obvious implications for the growing literature on conditional asset pricing models. At the academic level, there is an explosion of research on the determinants of time-varying expected returns. Economists are exploiting a great
variety of ideas, from macroeconomic models of real business cycles to more heterodox models of investor psychology. Theoretical research on equilibrium models uses the predictability evidence as a stylized fact to be modeled. In dynamic asset pricing models, for example, countercyclical Sharpe ratios have become a salient empirical fact that must be matched (e.g., Campbell and Cochrane, 1999). However, the long list of studies criticizing the statistical methodologies and, in particular, the poor out-of-sample performance of the predictive regression models, raise questions about the predictive variables’ role in those models. Given the doubtful performance of the simple regression specification in equations (1), (2), and (16), attempting to fit more complicated (but largely non-structural) models such as multiple-beta, conditional APT-type models, might seem a futile exercise. More structural models of equity risk as in Ang and Bekaert (2003) or Engstrom (2003) may help, but, because the stochastic discount factor is not observable likewise, increase the risk of overfitting further still.

Dynamic Asset Allocation Strategies... At a practical level, dynamic asset allocation models are becoming increasingly popular. The conditional risk premium is an essential input for dynamic portfolio optimization. It is thus often argued that the predictability evidence can provide quantitative inputs for investment strategies, be it for “pure market-timing strategies” (equity versus cash), or “sector and/or country rotation strategies”, or both. After all, although the fairly small $R^2$s suggest that predictability represents only a tiny fraction of the variance in stock returns, Kandel and Stambaugh (1996), Fleming, Kirby, and Ostdiek (2001), and, to a lesser degree, Grinold and Kahn’s (2000) “fundamental law of active management”, highlight that even small $R^2$s can potentially lead to optimal portfolios responding by a substantial amount to conditional return forecasts.

Basically, the fact that returns are somewhat predictable modifies the standard portfolio advice in three ways. It potentially introduces (buy-and-hold) horizon effects, it allows (myopic and, better, dynamic) market-timing strategies, and it introduces multiple factors via (intertemporal) hedging demands (if expected returns vary over time, investors may want to hold assets that protect them against this risk; in the case of equities, for example, via dynamic market timing).

Yet, given the doubtful (out-of-sample) performance of the simple regression specification in (1), (2), and (16), should investors really attempt to time equity markets, and if so how aggressive and over what investment horizon (transaction costs!)? To what extent and how should an investment portfolio be tailored to the specific circumstances of an investor? What are the consequences of parameter and model uncertainty for optimal investment strategies?

We do not aim to survey the broad literature on dynamic asset allocation models; this is left for another review. Some of those questions are critically reviewed in Cochrane (1999b). Instead of heavily borrowing from him, we better directly refer to this excellent reference – in the light of the critical findings
discussed above, his rather radical advice seems to be only justifiable!

Interestingly, a recent surge of research has increased attention to parameter uncertainty (estimation risk) and its impact on optimal portfolio choice. Barberis (2000), for example, uses a VAR framework (as in equations (1) and (2)) including the dividend-price ratio and equity returns to calculate expected return and return volatility in a Bayesian setting. In general, the perspective of a Bayesian investor (who uses the sample evidence to update prior beliefs about the regression parameters) seems to be particularly suitable to deal with estimation risk and to weigh the economic versus statistical significance of return predictability. Instead of just assessing the statistical significance of return predictability, it easily allows to directly demonstrate the economic impact on optimal portfolio choice. Economically rather meaningless confidence intervals (or $t$-statistics, etc.) are transformed via expected utility calculations into one optimal asset allocation. While Barberis (2000) does not explicitly calculate utility losses (by certainty equivalents, for example), Kandel and Stambaugh (1996) find that the regression relation can seem weak when described by usual statistical measures, but the current values of the predictive variable(s) can exert a substantial influence on investors’ portfolio decisions. Their approach, however, is restricted to a myopic one-period control problem. Barberis (2000) extends the analysis to the long run and finds that parameter uncertainty plays an important role in portfolio formation and can induce significant (negative) horizon effects that eliminates the (positive) horizon effects caused by return predictability when estimation risk is not taken into account (decreasing conditional volatilities over long investment horizons). Furthermore, he also shows that the aggressiveness of (myopic and dynamic) market timing as well as the intertemporal hedging demand is somewhat, but not completely, eliminated when parameter uncertainty is accounted for. After all, given the doubtful performance of predictive regressions, his choice of an uninformative prior (i.e., i.i.d. returns) is clearly justified. In any case, however, also a Bayesian setting cannot help to judge the out-of-sample performance of a dynamic asset allocation strategy.\footnote{Already Jensen (1978) stressed the importance of trading profitability in assessing market efficiency. He recognized that if anomalous return behavior is not definitive enough for an efficient trader to make money trading on it, then it is not economically significant. Similarly, in the equity premium literature, the outcome of an out-of-sample forecasting exercise is often seen as the most relevant measure for a (economically) successful model.}

Moreover, taking model risk and regime dependence into account, as in Avramov (2002) and Graffund and Nilsson (2003), respectively, only corroborates the conclusions of Barberis (2000). Xia (2001) also reinforces this point by taking into account that uncertainty about the predictive relation affects the optimal portfolio choice through dynamic learning.

\textbf{...and the Asset Management Industry} Cochrane (2001, Ch. 20) argues that the slow movement of the dividend-price ratio means that on a purely statistical basis, return predictability is a very open question. The dividend-price ratio signal has only crossed its mean four times in the 50 years of postwar hist-
tory. Investors thus have to be very patient to profit from a trading rule derived from this – a dynamic market-timing strategy based on a monthly investment horizon, for example, should therefore be viewed with skepticism. He emphasizes that what we really know is that low prices relative to dividends in the 50’s preceded the boom market of the early 60’s; that the low dividend-price ratios of the mid-60’s preceded the the poor returns of the 70’s; that the low price ratios of the mid-70’s preceded the boom market of the 90’s. We really have a once per generation change in expected returns (by the way, where are the business cycles here?). In addition, the last half of the 90’s has seen a historically unprecedented rise in stock prices and price-dividend ratios (or any other ratio). Of course, whether this trend was/is due to structural changes (such as lower interest rates, decreased dividend payouts, or growth expectations), or was/is indicative of market mispricings, is difficult to determine without a more complete valuation model. In any case, the evidence for predictability appears to have weakened in the 90’s; it even “went in the wrong direction” over a very long time period!

Consider the following example. Based on his predictive regression framework, Lamont (1998, p. 1579) claimed that – out of sample – total stock returns over the period from 1995 to 2000 are projected to be one percent less than total Treasury bill returns! If an (active) asset manager had taken this advice seriously and reduced the weight invested in equities accordingly, he would have been out of business probably pretty quickly. Of course, if predictability is real, his portfolio may have a greater mean for a given level of risk over very long horizons, but it will do well and badly at very different times from everyone else’s portfolio. He will often underperform a benchmark, particularly so in the short run. But this is only one side of a fund manager’s business risk. A second component is probably more intriguing. If an asset manager “promotes” a boring “passive” investment strategy without having a fancy story about rational business cycle-related return predictability, for example, he would possibly never come into business at all. Unfortunately, from the (short-term) viewpoint of the asset management industry, the second kind of business risk is likely to be more important than the former, giving rise to the promotion of active market-timing strategies based on stock market predictability (or, in general, based on (other) return anomalies or “investment styles” such as value, growth, small cap, momentum, etc.). However, given the above evidence, it is all but obvious that stock market predictability enhances return opportunities and therefore is really valuable to investors instead of a sophisticated but mere marketing tool for the asset management industry. Investors are presumably better off if fund managers would not extensively vary from a passive market index.

This conclusion is, at least partly, further supported by the following discussion borrowed from Kandel and Stambaugh (1996) and Campbell and Viceira (2002, Ch. 4). They (explicitly) note that based on a partial equilibrium framework (which is standard in the dynamic asset allocation literature), all investors are forced to buy and sell assets at the very same time. With a constant supply of stocks, however, this cannot be consistent with a general equilibrium model. One possible resolution of this difficulty is that the representative investor has
different preferences from those commonly assumed (e.g., power utility), perhaps the habit-formation preferences of Campbell and Cochrane (1999). Models with time-varying risk aversion may solve the complementary problem of finding preferences that make a representative agent content to buy and hold the market in the face of a time-varying equity premium, but under this interpretation, any results on market timing should be used only by investors with constant risk aversion, who cannot be typical of the market as a whole.\footnote{Another way to generate time variation in the equity premium is from time variation in volatility (risk). Schwert (1989b) presents evidence that the conditional volatility of stock returns moves countercyclically along with forecasts of stock returns, and Bollerslev, Engle, and Wooldridge (1988) derive implications of exogenous movements in return volatility for the equity premium. Of course, it is desirable to derive movements in return volatility from underlying fundamentals; accordingly, Abel (1988), Kandel and Stambaugh (1991), Veronesi (1999), and Whitleaw (2000) model heteroskedasticity or time-varying uncertainty about the consumption (dividend) process. A difficulty with these efforts is that the evidence for heteroskedasticity in aggregate consumption is fairly weak. This may suggest that it is worthwhile to build a model in which risk aversion changes over time so that the price of risk, rather than the quantity of risk, is time varying. Several authors have recently argued that trading between investors with different degrees of risk aversion or time preference, possibly in the presence of market frictions or portfolio insurance constraints, can lead to time variation in the market price of risk (Dumas, 1989; Grossman and Zhou, 1996; Wang, 1996; Chan and Kogan, 2002).}

This point is similarly stressed in Cochrane (1999b). He emphasizes that the average investor must hold the market. Market timing can only work if it involves buying stocks when nobody else wants them and selling them when everybody else wants them. Portfolio advice to follow these strategies must fall on deaf ears for the average investor, and a large class of investors must want to head in exactly the other direction. So, at a basic level, return predictability would (and should) not affect the portfolio decisions of average investors! Put it differently, all arguments that return predictability is real are inconsistent with market-timing strategies. If stock market predictability is real, i.e., an equilibrium time-varying reward for holding risk (Bekaert’s (2001) risk view), then the average investor knows about it but does not invest because the extra risk exactly counteracts the extra average return.\footnote{This also raises questions about the way (conditional) performance measurement is conducted and interpreted!} If more than a minuscule fraction of investors are not already at their best allocations, then the market has not reached equilibrium and the premiums will change. If the risk is irrational, then by the time the average investor knows about it, it is gone (no out-of-sample evidence). But an expected return corresponding to an irrational risk premium has the strongest portfolio implications – everyone should do it – but the shortest lifetime. If the average return comes from a behavioral aversion to risk (the behavioral view), it is just as inconsistent with market timing portfolio advice as if it were real. We cannot all be less behavioral than average, just as we cannot all be less exposed to a risk than average. Finally, if return predictability is indeed a statistical fluke due to poor statistical inference (the statistical view), any advice to time the market is questionable anyway.

We would not generally deny that the equity premium is time varying. In any case, it is probably the most promising solution for the excess volatility puzzle.
However, it seems very questionable whether a (business cycle-related) time-varying equity premium can be predicted using simple regression techniques and whether the respective results should be considered as a serious input for (in the end, out-of-sample) dynamic asset allocation strategies. The world of investment opportunities is probably much simpler than indicated by the stock market predictability and dynamic asset allocation literature. For investment purposes, the natural alternative to return predictability is the assumption that excess returns are constant and unpredictable i.i.d. After all, it is already hard enough to estimate and justify the unconditional equity premium (just recall the equity premium puzzle of Mehra and Prescott (1985); also, Merton, 1980). Admittedly, compared with common sense and much industry practice, this seems to be radical advice. Still, this does not mean that intertemporal hedging effects are completely absent. Strategically important arguments such as human capital (labor income), endogenous labor supply, and, in general, life-cycle considerations (asset- and liability management, ALM) may give rise to intertemporal hedging demands as well, even if stock returns themselves are assumed to be i.i.d. (e.g., Samuelson, 1991, 1994; Viceira, 2001; Campbell and Viceira, 2002, for a comprehensive discussion). In any case, a comprehensive consideration of other strategic arguments seems to be much more important to the average investor than quibbling over sophisticated myopic or dynamic market-timing strategies based on the rather weak if not non-existent evidence of stock market predictability.
References


43


45