The Good News and the Bad News about Long-Run Stock Returns

Donald Robertson
Faculty of Economics and Politics
University of Cambridge
Austin Robinson Building
Sidgwick Avenue
Cambridge CB3 9DD
donald.robertson@econ.cam.ac.uk

Stephen Wright
Department of Economics
Birkbeck College
University of London
7-15 Gresse Street
London W1P 2LL
s.wright@bbk.ac.uk

June 25, 2002

Abstract

If single-period stock returns were unpredictable, long-horizon returns would have been significantly riskier than has been observed historically. The good news is that if, as a large body of research suggests, there is even a weak tendency for stationary valuation indicators to predict stock returns, this can explain the relatively low dispersion of long-run returns, and thereby provides some support for the "buy-and-hold" strategy. We illustrate this in a cointegrating vector autoregression, with Tobin's q and the dividend yield, adjusted for new issues and buybacks, as the two cointegrating relations. The degree of bad news depends on the degree of good news. The better is the news on implied long-run variance of returns, the worse the news on expected returns.

Keywords: stock market; valuation ratios; stock return predictability; random walk; buy-and-hold strategy; Tobin's q; dividend yields; cointegrating vector autoregression; interval forecasting.

JEL Classifications: C32, C53, E44, G10, G14.

1. Introduction

A very large proportion of life-cycle consumption smoothing is carried out via the stock market: the proportion of stocks held by a typical US pension fund, for example, has been running at around 70% in recent years.[1] What is the degree of risk associated with the returns on such investments, over such long horizons?

The consensus view of most practitioners, as encapsulated in the"Buy-and-Hold" investment strategy, has probably been expressed most systematically and forcefully in the work of Siegel (Siegel, 1992; 1998), who makes the case for stocks, not just on the familiar grounds of superior returns, but also on the grounds of relative predictability. Comparing returns on stocks, bonds and bills over nearly 200 years of US data, Siegel provides evidence that the degree of riskiness of stock market returns has been barely, if at all, greater than that of bonds and bills, over investment horizons of twenty to thirty years.

Such evidence would be very difficult to reconcile with a world in which stock returns were unpredictable, since the implied range of dispersion at long horizons has been distinctly too narrow. The relatively low historic dispersion of historic returns is however consistent with the very large body of literature that has concluded that there is a degree of predictability in stock market returns, especially over longer horizons (*eg* Poterba & Summers (1988); Breen, Glosten & Jagannathan (1989); Jegadeesh (1990); Bekaert & Hodrick (1992); Ammer & Campbell (1993); Chiang, Liu & Okunev (1995); Pesaran & Timmermann (1995); Campbell & Shiller (1998); Lo and Mackinlay (1999)).[2]

Much of the evidence for predictive power has arisen from the ability of various valuation measures (the dividend yield in particular) to predict (if only weakly) subsequent stock market returns. We show in this paper that, under certain conditions and circumstances, the predictive power of valuation criteria significantly reduces the uncertainty associated with long-run stock market returns. For this to hold, the valuation criterion must imply that the ratio of the stock price to some measure of the "fundamental" be mean-reverting. This ratio must also Granger Cause returns. If stock prices are volatile in the short term, but have some tendency to revert towards a fundamental which may be easier to predict, at least some of the predictability of the fundamental will be passed on to stock prices, and hence in turn to long-horizon returns. Thus the "fundamental" should also be relatively easy to predict, compared to the stock price.

[1] Estimate derived as (corporate equities plus mutual fund shares)/ (total financial assets less miscellaneous assets), for private pension funds, as of end-1997 (Source: Federal Reserve, 1997, Table L119.

[2] For an excellent survey of this field, see Chapters 1,2 and 7 of Campbell, Lo and MacKinlay (1997).

There is a clear link between this property and the related econometric concepts of Granger causality and cointegration, which allows us to draw on the literature initiated by Granger (1986), Engle and Granger (1987) and Johansen (1988), that has systematised the representation of vectors of unit root processes which share one or more common stochastic trends.

Drawing on a new annual dataset for the US nonfinancial corporate sector over the course of the twentieth century (Wright, 2001), we focus on the role of two valuation criteria: the dividend yield (both the standard measure, and and an alternative measure, adjusted for buybacks and new issues), and Tobin's q. Mean reversion of these ratios implies that they can be embedded as cointegrating relations in a vector autoregressive representation that includes real stock prices and dividends. By exploiting Campbell & Shiller's (1988) approximation for the log stock return, we show that confidence intervals in predicting returns over long horizons can be very significantly reduced, compared to those from a simple model in which returns are unpredictable. The implied variance reduction is especially marked when q is one of the cointegrating relations. This is the good news about long-run stock market returns.

The corollary of the good news on the variance of long-horizon returns is some degree of bad news about expected returns. The first element in the bad news is that the mean return implied by the cointegrated systems is distinctly lower than the average of realised returns over the twentieth century . The second element is that at the end of the twentieth century q remained at an historically very high level, implying a high probability of historically poor returns for a prolonged period.

Both the good and the bad news about returns are of course dependent on the validity and predictive power of the valuation criteria used to predict returns. If dividends are adjusted for buybacks and new issues, the implied dividend yield was at an historically normal level at the end of the twentieth century, and hence was forecasting historically normal returns. Thus if q is ignored, the news on expected returns is less bad. There have also been a number of papers in recent years (*eg* Kirby, 1997; Bossaerts & Hillion, 1999; Foster, Smith & Whaley, 1997) that have cast doubt on the earlier evidence of predictability, attributing it to data mining. Without predictability, clearly the best forecast of future returns is simply the historic mean return. However, while these alternative specifications imply less, or no, bad news on expected returns, this is offset by less, or no good news on return variance. They are thus hard to reconcile with the evidence of low dispersion of historic long-horizon returns.

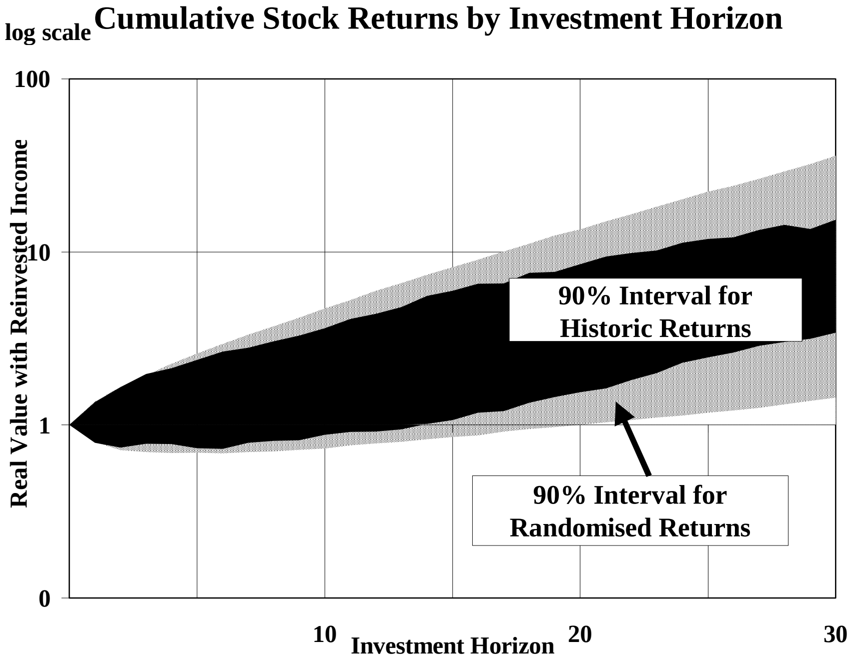

Figure 2.1:

2. The Dispersion of Historic Returns

Figure 2.1 gives an illustration of the difficulties involved in reconciling the dispersion of historic US stock returns with a world in which returns were unpredictable. The inner band shows the range containing 90% of historic values of cumulative real stock returns (constructed from annual returns data for returns over the period1802-1998[3]) at different investment horizons. The outer band shows the same interval if historic annual returns are randomised to destroy any serially correlated structure.[4]

While the two bands give a very similar pattern at horizons up to around five years (indeed, by construction, they coincide at the one year horizon), at longer horizons, they differ significantly. At investment horizons of 20 to 30 years, the dispersion of historic returns is much less than that of randomised returns, with

[3]Data are from Siegel (1994) updated to 1998. We are grateful to Professor Siegel for providing us with the data.

[4]We draw 10,000 returns randomly from the 197 observed annual log returns in order to proxy the population properties of randomised returns. We then construct cumulative returns, $c_t(h) = \Sigma_{i=0}^{h-1} r_{t-i}$, for horizons $h = 1, 2, \ldots, 30$; Figure 2.1 shows the 5th and 95th percentiles of the observed and simulated $\{c_t(h)\}$ at each horizon.

the discrepancy increasing with the investment horizon.

Since we are using only just under two centuries worth of data, this begs the question whether the differences between the two intervals are actually in any sense significant. This is particularly relevant for long horizons, given the small number of non-overlapping long-horizon returns. In Appendix A we investigate, by bootstrapped simulations of artificial samples of the same length, whether the degree of apparent "safety" of observed long-horizon returns (*ie,* the apparently low probability of negative returns) , and their relatively low degree of dispersion could have arisen by chance in a world of randomised annual returns. We conclude that it would not be particularly suprising to have observed at least as great a degree of apparent "safety" of returns; but that the probability of such a low degree of dispersion at relatively long horizons is is much lower (albeit not vanishingly small).

Since the low dispersion of returns appears unlikely to be consistent with unpredictable returns, the corollary is that there must be some form of predictability, that has a particular impact on long-horizon return variance, but a much lower impact at shorter horizons.

3. Valuation Criteria and Return Predictability

We noted in the introduction that a common feature of much recent empirical research is that predictability in stock prices and returns is commonly associated with the use of valuation criteria - most notably the dividend yield. There are two distinct ways to interpret the predictive power of such indicators.

The first implies a fairly literal interpretation of the notion of valuation: *ie* it presumes that stock markets can at times be over- or undervalued, relative to some fundamental, and that this deviation from "fair" value reflects some failure of efficiency (this interpretation was implicit in Shiller's (1981) original findings of "excess volatility"). By implication the associated failure of arbitrage implies the existence of a trading rule which may offer "excess" returns.

The alternative interpretation, in terms of time-varying, but predictable, expected returns rejects such a literal use of the concept of "value". Market efficiency is essentially treated as a maintained hypothesis: hence stock markets are always, by definition, fairly valued, given the information set available to participants in the market. Returns may be predictable, to the extent that expected returns are predictable, but there are no "excess" returns to be earned.[5]

[5] Although these two interpretations are presented as polar extremes, there is no necessary contradiction between the two - indeed they are in many respects complementary. Thus, "noise trader" models of stock market inefficiency such as that of De Long *et al* (1990) usually have an exogenous source of noise, which is normally assumed to be at least partially predictable - such

However, in the empirical analysis of this paper, the differences between these alternative interpretations are unimportant: in both cases the power of valuation criteria rests in the nature of the predictability of stock prices and returns which they confer. Both imply a rejection of the random walk representation of cumulative returns; both share a clear link with the econometric concepts of Granger causality and cointegration.

A simple example helps to illustrate this point, and bring out the crucial empirical characteristics of any useful valuation criterion.[6]

We suppose initially that returns are measured solely by log changes in the real stock price P.[7], and that some valuation criterion, V is proposed, that can be expressed as the ratio of P to F, an indicator of what we shall refer to (purely by convention) as the "fundamental". Suppose, further that the two underlying series can be given the following log-linear first order error correction representation (with lower case letters representing logs of underlying series):

$$\Delta p_t = \alpha_o + \alpha_1 \ v_{t-1} + \varepsilon_t^p \tag{3.1}$$

$$\Delta f_t = \beta_o + \beta_1 \ v_{t-1} + \varepsilon_t^f \tag{3.2}$$

where ε_t^p and ε_t^f are white noise error processes.

In such a representation, if the valuation criterion is to offer any predictive power, $v \, (\equiv p_{t-1} - f_{t-1})$ must be the single cointegrating relation between two unit root processes, p and f. A necessary condition, therefore, for V to be a useful valuation criterion, is that it must be a mean-reverting process.[8] This is however not a sufficient condition. Stationarity of v, and hence cointegration of p and f, could in principle be possible in a random walk stock market, if α_1 equalled zero. In this case, in our simplifed example, the stock price would be an independent random walk, and there would be one way Granger Causality from the stock price

---

noise could in principle easily stem from time-varying expected returns. On the other hand, as Grossman and Stiglitz (1980) demonstrated, assuming full market efficiency as a maintained hypothesis begs the question of the nature of the (approximate) arbitrage which maintains efficiency. Indeed, market efficiency requires that individual stock prices, at least, be on occasion mis-valued in order to provide sufficient incentive for arbitrageurs.

[6] The argument that follows is in essence very similar to that of Campbell and Shiller (1988;1998).

[7] Our econometric work does not make this assumption. Approximating returns by log stock price changes alone neglects a term in the log dividend-price ratio in Campbell & Shiller's (1988) approximation (see equation 4.1 in Section 4 below).

[8] By mean reversion here we wish to capture two key features: first, that point forecasts converge, as the forecast horizon lengthens, to a fixed value (or a close neighbourhood thereof); and second, that forecast variance tends to a finite limit (or grows so slowly as to be indistinguishable from this).

to the "fundamental". V would simply be an indicator of future changes in F, and would not be a useful valuation criterion. The second crucial condition for any useful criterion must therefore be that it must Granger-Cause returns - in the above representation, α_1 must be non-zero.[9] The necessary link between the use of any valuation criterion and the predictability of returns is therefore clear-cut.

There is a third, less obvious characteristic that is required if the predictive power of a valuation criterion is to be useful over long horizons. In our simplified example, forecasting the stock price at long horizons ultimately reduces to forecasting the fundamental. The fundamental should therefore itself also be relatively easy to forecast over long horizons. We discuss this characteristic in more detail in Section 9 below.

Given the limitations of empirical evidence, and the dangers of data mining, however, it is also very important that we should be able to justify any valuation criterion on theoretical grounds. Hence, before we examine the empirical properties of our valuation criteria, we first consider their theoretical basis - with particular reference to the issue of whether we would expect them to meet the necessary condition, discussed above, that they be mean-reverting.

4. Theoretical Properties of Tobin's q and Dividend Yields

4.1. The Dividend Yield

Following Campbell and Shiller (1988) we can derive a log-linear approximation for the log stock return (again, letting lower case letters denote logs):

$$\begin{aligned} r_{t+1} &\equiv \ln(1+R_{t+1}) \equiv \ln\left(\frac{P_{t+1}+D_{t+1}/E_{t+1}}{P_t}\right) \\ &\approx \delta + \Delta p_{t+1} + (1-\rho)\left[d_{t+1} - e_{t+1} - p_{t+1} - (\overline{d-e-p})\right] \end{aligned} \tag{4.1}$$

where P_t is the stock price; D_t is total dividend payments; E_t is the number of shares; $\rho = \frac{1}{1+\exp(\overline{d-p})}$ and $\delta = \ln\left(1+\exp(\overline{d-e-p})\right)$. Solving in terms of $d_t - e_t - p_t$, and iterating forwards, subject to the transversality condition $\lim_{i\to\infty}\rho^i(d_{t+i} - e_{t+i} - p_{t+i}) = 0$, we can write:

$$(d_t - e_t - p_t) \approx \frac{\varphi}{1-\rho} + \sum_{i=1}^{\infty}\rho^{i-1}[r_{t+i} - \Delta(d_{t+i} - e_{t+i})] \tag{4.2}$$

where $\varphi = \delta - (1-\rho)(\overline{d-e-p})$. This formulation has been the motivation for a large body of literature that examines the ability of the dividend yield to

[9] Note that there is no requirement that causality go only one way: β_1 can also be non-zero.

predict future returns. The conclusion of this literature has generally been that the dividend yield does have some, rather weak predictive power for future returns, especially over longer horizons.

This predictive power is however predicated on the assumption that the dividend yield is mean-reverting. Removing subscripts to denote steady-state values, we have:

$$(d - e - p) \approx \frac{\varphi + r - (\Delta d - \Delta e)}{1 - \rho} \tag{4.3}$$

which is simply a log-linearised version of the familiar Gordon Growth formula.[10]. Thus for the dividend yield to be mean-reverting requires that both returns and the growth rate of dividends per share be mean-reverting.

The steady-state return will have a stable mean if its underlying determinants have in turn also been stable. Standard general equilibrium results (eg Kiley, 2000) typically imply that:

$$r = \theta + \sigma^{-1} g_c + \eta \tag{4.4}$$

where θ is the subjective rate of pure time preference, σ is the intertemporal elasticity of substitution, g_c is the growth rate of consumption per capita, and η is the equity premium. Thus stability of returns requires stability of the growth rate, of the underlying intertemporal parameters, and of the (still puzzling[11]) equity premium.

Historic evidence from long samples (*eg*, the dataset of Siegel (1998) used above) suggests that there is some reason to believe that stock returns do indeed have a stable mean[12], although even such long runs of data do not provide sufficient evidence to have great confidence in this conclusion. Fama & French (2001) also discuss the difficulties of assessing whether the equity premium component of underlying returns have been stable in the postwar period. Despite these problems, it is usually assumed in the predictability literature that stock returns are

[10] We distinguish deliberately here between historic means (which are observable magnitudes, and, at any given point in time, fixed), and steady-state values (which are unobservable, but which may in principle vary if there are changes in structural parameters). For small changes in steady-state values, the log-linearisation around historic means will still be valid.

[11] Since Mehra and Prescott (1985) it has been established that, if aggregate consumption is a good proxy for the consumption of the representative investor, η should be very small for plausible values of σ, and its observed large historic value is thus inexplicable in the standard model. See Kocherlakota(1996) for a discussion of more recent work, and the conclusion that the puzzle persists.

[12] For example, Smithers & Wright, 2000, Chapter 17 examine the issue using Siegel's (1998) data on US stock returns since 1802. The evidence of stability of the safe return, and hence the equity premium, is however much weaker.

mean-reverting, and we work on the same assumption.

However, if standard empirical measures of dividends are used, it is more problematic to assume that the other determinant of the steady-state dividend yield in (4.3), the growth rate of dividends per share, will be stable. While it is to be expected that total dividend payments must ultimately grow at the same rate as the economy (and hence consumption), a permanent shift in Δe, the rate at which new shares are issued (or bought back), can in principle shift permanently the growth rate of dividends per share, without violating any transversality condition (since any given flow of dividends could be consistent with an arbitrary growth rates of dividends per share).

Any such shift will imply a shift in the steady-state dividend yield, even if all other underlying parameters are stable. Given the well-known high degree of sensitivity of the Gordon Growth formula to shifts in its determinants[13] (note that $1-\rho$, the term in the denominator of (4.3), is small) even a relatively modest shift in the rate of new issues or buybacks can imply a large corresponding shift in the log dividend yield. Nor is this a merely theoretical consideration - as we shall see in Section 5.2.

4.2. The Adjusted Dividend Yield

An alternative measure of the dividend yield, that is not subject to this problem, is also more consistent with the fundamental valuation model of Miller and Modigliani (1962). Following Mehra (1988), Wright(2001), and Robertson and Wright (2002b) we can define a measure of dividends, adjusted for new issues and buybacks, as:

$$\widetilde{D}_t = D_t - N_t \tag{4.5}$$

where N_t is the value of new issues minus buybacks. Thus $\widetilde{D}_t$, adjusted dividends, measures the total net cashflow between the corporate sector and equity holders. In a counterfactual world where all of this net cashflow had been paid out as dividends, net new issues would, by construction, have been zero in all periods, hence we can define a counterfactual adjusted share price, $\widetilde{P}_t$, consistent with a counterfactual constant number of shares, $\widetilde{E}$, that we can normalise for convenience to unity. Thus $\widetilde{P}_t$, on this definition, is identical to stock market value:

$$\widetilde{P}_t \equiv P_t E_t \tag{4.6}$$

We can also define the associated return as:

[13] See, for example, Heaton and Lucas (1999).

$$\widetilde{r}_{t+1} \equiv \ln(1+\widetilde{R}_{t+1}) \equiv \ln\left(\frac{\widetilde{P}_{t+1}+\widetilde{D}_{t+1}}{\widetilde{P}_t}\right) \tag{4.7}$$

which is simply the return earned by a representative investor who owned the entire stock market, since the numerator and denominator measure such an investor's wealth at the end of periods $t+1$ and t respectively. Wright(2001) shows that this return is trivially different from the standard definition, as given in (4.1).[14]

Thus we can also derive an equivalent expression for the steady-state adjusted dividend yield, as:

$$\left(\widetilde{d}-\widetilde{p}\right) \approx \frac{\widetilde{\varphi}+\widetilde{r}-\Delta\widetilde{d}}{1-\widetilde{\rho}} \tag{4.8}$$

where $\widetilde{\varphi}$ and $\widetilde{\rho}$ are defined consistently with (4.1).[15] This expression differs from the equivalent expression for the unadjusted log dividend yield in (4.3) both by the adjustment to dividends on the left-hand side, and by the removal of any impact of Δe, the rate of net new issues, from the right-hand side. Standard transversality conditions would imply that the growth rate of adjusted dividends must also ultimately be pinned down by the growth of the economy,[16] and hence consumption (indeed the arguments are probably stronger than for unadjusted dividends), thus the steady-state adjusted dividend yield is ultimately determined by the same set of structural parameters that determine the equilibrium return, and, unlike the unadjusted yield, by no others.

4.3. Tobin's *q*

In standard fashion, we define Tobin's q[17] by:

$$Q_t = (= \exp(q_t)) = \frac{P_t.E_t + L_t}{K_t} \tag{4.9}$$

[14] It differs by a cross-product term in two ratios that are each close to zero.

[15] Since the sample mean of $\widetilde{d}_t - \widetilde{p}_t$ in our sample is very close to that of $d_t - e_t - p_t$, (see Section 5.2) the implied linearisation coefficients are virtually identical to those in (4.1), and $\widetilde{r}$ is indistinguishable from r.

[16] Indeed the case is stronger for adjusted dividends, since this is a net cashflow measure. We could in principle imagine a world in which both dividend payments and net new issues exploded relative to GDP, but the difference between the two did not (gross vs net exports being an analogy).

[17] We spend most of the time in this paper inhabiting a log-linear world. We follow standard practice in letting lower case letter letters denote natural logarithms. Hence, for the sake of precision it should be noted that whenever we refer to q we are referring to the log of the levels ratio Q, since most of our results relate to the log series.

where P_t and E_t are as defined above (thus P_tE_t is the value of the stock market), L_t is the market value of corporate liabilities, and K_t is the corporate capital stock.[18]

In contrast to the dividend yield, Tobin's q does not originate as a valuation criterion for the stock market, but rather as an analytical tool for the study of the investment process. After the initial impetus in Brainard & Tobin (1968), the concept of q was developed by Hayashi (1982), Abel & Blanchard (1983,1986) amongst others, who established the link between Tobin's q and marginal q, most commonly interpreted as the derivative of the representative firm's value function with respect to capital. A key feature of marginal q is that its steady state value is likely to be invariant to most, if not all behavioural parameters in the economy,[19] and we would therefore expect it to be mean-reverting. Robertson and Wright (2002a) show that, while measured Tobin's q may differ from marginal q for a range of reasons, there are strong theoretical reasons to expect that, in the absence of unit root measurement error in capital, the difference between Tobin's q and log marginal q will also be mean-reverting (though not necessarily to a mean of zero). Hence Tobin's q, as the sum of two mean-reverting processes, will itself be mean-reverting. Note also that the stability of the mean value of Tobin's q does not depend upon stability of any of the behavioural parameters that determine the mean dividend yield.

Since in our empirical analysis we work in a standard log-linear framework, rather than work with the exact definition of q given in (4.9), we work instead with the log-linearised approximation,(expanding around the mean log level of leverage[20]) as:

$$q_t \approx \xi + (1-\zeta)(p_t + e_t - k_t) + \zeta(l_t - k_t) \tag{4.10}$$

where $\xi = \ln(1 + \exp(\overline{l-p-e}))$, and $\zeta = \exp(\overline{l-p-e})/(1+\exp(\overline{l-p-e})))$.

By substituting from (4.10) into the forward sum representation of the adjusted dividend yield (as in 4.2)), an equivalent forward sum can be derived[21] for q as

[18] We follow standard practice, in our empirical measures, in measuring liabilites in net terms; hence we can ignore corporate financial assets.

[19] Apart from Jensen's Inequality covariance terms, the mean value of log marginal q depends solely on the technology of capital installation. If, as is commonly assumed (see, for example, Kiley, 2000; Cummins & Bond, 2000), the mean marginal cost of installing capital is zero, marginal q will have a mean of unity and will thus be invariant to all behavioural parameters in the economy.

[20] Although we log-linearise q around the sample mean of $l - p - e$, we do not need the true mean to be constant to apply the approximation - it just becomes increasingly inacccurate over long samples. In our sample the squared correlation coefficient between the log-linearised approximation for q and the log of q itself is 0.998, so we do not regard this as a major problem.

[21] For fuller details of the derivation, see Robertson and Wright (*op cit*). An equivalent expression can also be derived in terms of unadjusted dividends.

(ignoring constants):

$$q_t \approx \sum_{i=1}^{\infty} \widetilde{\rho}^{i-1} \left\{ \Delta k_{t+i} + (1-\widetilde{\rho})[(1-\zeta)(\widetilde{d}_{t+i} - k_{t+i}) + \zeta(l_{t+i} - k_{t+i})] - \zeta \Delta l_{t+i} - (1-\zeta) r_{t+i} \right\} \tag{4.11}$$

The empirical investment literature has almost invariably found that measured Tobin's q has very weak predictive power for investment. But the implication of (4.11) is that, given mean reversion, q must predict something else on the right-hand side. Our particular interest in this paper is q's predictive power for returns.

Thus, while the analytical foundations of q in the investment literature are crucial in understanding its tendency to mean reversion, the forward sum in (4.11) implies that its failure as a predictor of investment raises the possibility of its use as a valuation criterion.

5. Data

5.1. Data Sources and Construction

Sources and methodology for all data used in the rest of this paper are provided in detail in Wright (2001). All data come from a new dataset described therein, that relates to the total nonfinancial US corporate sector (rather than the more commonly used subset of quoted companies) over the sample 1900-2000, using data from the Federal Reserve's Flow of Funds Tables, and the BEA National Income and Product Accounts and Tangible Assets data where these exist, and such historical sources as are available in earlier periods. Wright (*op cit*) shows that where series are directly comparable with equivalent series for quoted companies, they generally have similar properties; but have the advantage that they can be related directly to other national accounts and flow of funds data (including, as prime examples, capital stock and net issue/buyback data).

Precise definitions of the particular series used in this paper, in terms of the underlying dataset, are provided in Appendix B.[22]

5.2. Univariate Properties of Dividend Yields and of Tobin's q

Figures 5.1 and 5.2 allow a comparison of the univariate properties of the two alternative measures of the dividend yield with those of Tobin's q. To ease the comparison, both dividend yields are plotted in log reciprocal form (ie, as log price-dividend ratios), and the range of the horizontal scales of both charts is the same in log terms.

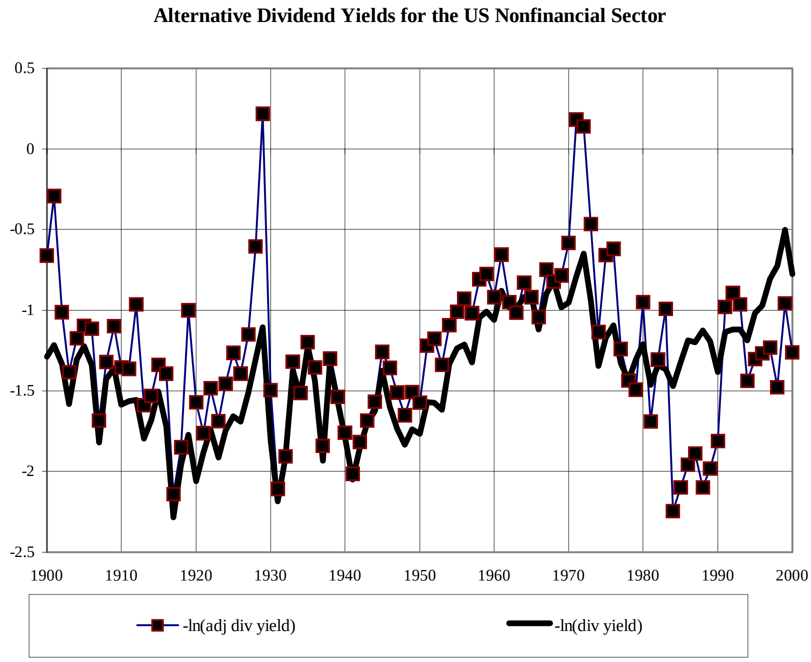

Figure 5.1:

Figure 5.1 shows that, while, as noted above, the unadjusted and adjusted dividend yields have very similar means[23], they have at times distinctively different properties. In roughly the first two thirds of the century, the adjustment reflected distinct surges in new issues at certain periods (most strikingly in 1929, and in the early 1970s, which lowered the implied adjusted yield significantly (hence raised the price/dividend ratio shown in the chart), while in other periods (most notably the 1930s and early 1940s) new issues essentially collapsed to zero, such that the two yields were nearly identical. However, in the last two decades of the century there was a distinct shift, with the adjustment switching sign, as firms engaged in significant levels of buyback activity, that more than offset minimal levels of new issues.

The chart also shows that, while the adjusted yield is distinctly more volatile than the unadjusted yield (and, also, as the comparison with Figure 5.2 shows, than q), it appears to have a more stable mean.[24] The unadjusted yield showed

[22] The underlying data can be downloaded from www.econ.bbk.ac.uk/faculty/wright.

[23] Of 3.97% and 3.51% respectively, with the lower mean of the adjusted yield reflecting the tendency to net new issues in most of the century.

[24] The volatility of the implied dividends series also appears more stable throughout the sam-

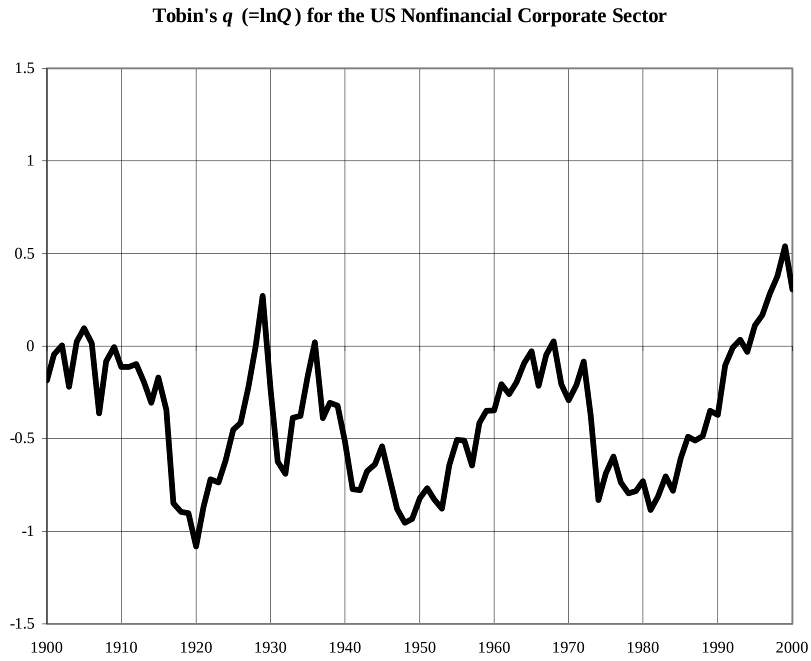

Figure 5.2:

a distinct downward drift in the latter part of the sample (an upward drift in the price/dividend ratio in the chart), in part reflecting the pattern of new issues/buybacks, and very much in line with the analysis of Section 4.

The downward drift in the unadjusted dividend yield in the latter part of the sample was however not entirely explicable in these terms, but also reflected the surge in the stock market during the course of the 1990s, which is evident both in the unadjusted yield and, as Figure 5.2 shows, in q.[25] Strikingly, however, it is not evident in the adjusted dividend yield, which ended the century more or less at its mean, since adjusted dividends grew as rapidly as the stock market during the 1990s, due to the strength of buybacks (indeed, Figure 5.1 shows that adjusted dividends grew distinctly faster than the stock market in the 1980s).

Figure 5.2 shows that for most of the century, Tobin's q was below zero, implying that the underlying ratio in levels was typically less than unity. On the face of it this might appear somewhat surprising, given that it is usually assumed that the levels ratio should have an equilibrium value of unity. We discuss this issue

ple, in contrast to the unadjusted series, which was distinctly less volatile in the second half of the twentieth century.

[25] As also in price-earnings multiples: see Shiller, 2000.

at some length in Robertson and Wright, (2002a) and conclude that it can be explained, either in terms of systematic over-statement of capital by statisticians, or (more speculatively) by informational differences between firms and stock market investors, that may lead the stock return to be more highly correlated with the market's stochastic discount factor than the representative firm's internal return, leading, in effect, to it being valued at a discount relative to its underlying assets.

The two charts provide some visual evidence of mean reversion of q and the adjusted dividend yield, with distinctly weaker evidence for the unadjusted yield. Clearly, however, the experience of the last decade of the sample took both q and the unadjusted yield well outside their previous historical norms.[26] This is largely borne out by unit root tests, shown in Table 1. As might be expected, however, given the well-known difficulties in testing for the presence of unit roots (particularly against an alternative of a stationary but very persistent alternative) test results are not overwhelmingly conclusive at conventional probability levels, for either q or the unadjusted dividend yield. If data for the last ten years of the century are excluded, the evidence is stronger. The rejection of a unit root in the adjusted dividend yield is distinctly more marked, though here the comparison with the other two series may be somewhat distorted by the considerably greater volatility of the adjusted dividend yield, as revealed by Figures 5.1 and 5.2, since unit root tests reject less frequently if there is a low signal to noise ratio (Harvey & Jaeger, 1991).

Table 1 Unit Root Tests for Alternative Valuation Criteria

		Tobin's q	Unadjusted Dividend Yield	Adjusted Dividend Yield
1900-1990	ADF	-3.12**(4)	-2.16(2)	-2.70*(2)
	PP	-2.84*	-3.42**	-4.30***
1900-2000	ADF	-2.71*(4)	-1.45(2)	-3.36**(2)
	PP	-2.46	-2.85*	-4.91***

Figures in parentheses after ADF statistics show number of lagged difference terms in ADF regression, chosen by Akaike Information Criterion

* (**) [***] Rejection at 10% (5%) [1%] significance levels

Overall, however the univariate tests, combined with our theoretical priors, do appear to point to the presumption of mean reversion of q and the adjusted dividend yield; but less so in the case of the unadjusted dividend yield.

[26]It should be noted that the behaviour of returns during this period was equally extraordinary. Realised returns during the 1990s over a range of investment horizons were at or beyond the limits of historical experience (*ie*, in terms of Figure 2.1, cumulative returns went well outside the 90% historic interval at a range of horizons, from 5 up to 20 years). By implication, since the adjusted dividend yield remained in a normal range, the growth of adjusted dividends was also historically exceptional.

6. Tobin's q and the Dividend Yield in a Cointegrating VAR

6.1. The Cointegrating VAR Representation

Given that both q and the dividend yield represent ratios between nonstationary series, and our interest in Granger Causality relations, it is natural to embed these ratios as cointegrating relations in representations of four variables: stock price, dividends, capital and liabilities (the latter is included due to its role in the log-linearised version of Tobin's q, as given in (4.10)).

In order to compare the properties of the two alternative measures of the dividend yield we work with two parallel systems, in the vectors $\mathbf{x}_{ut}$ and $\mathbf{x}_{at}$, defined by:

$$\mathbf{x}_{ut} = \begin{bmatrix} p_t \\ d_t - e_t \\ k_t - e_t \\ l_t - e_t \end{bmatrix}; \quad \mathbf{x}_{at} = \begin{bmatrix} \widetilde{p}_t \\ \widetilde{d}_t \\ k_t \\ l_t \end{bmatrix}; \tag{6.1}$$

where $\mathbf{x}_{ut}$ is defined in per share terms using the unadjusted share price and dividends, and $\mathbf{x}_{at}$ in adjusted terms (*ie*, assuming a constant equity issue, normalised to unity in levels, as in (4.6)) using the adjusted share price and dividends.[27]

Standard lag order testing procedures suggest a VAR(2) representation. Unit root tests, unsurprisingly, do not reject the hypothesis that all four series contain a unit root,[28] but are difference-stationary. We therefore estimate all systems with the general form:

$$(\boldsymbol{\Delta}\mathbf{x}_t - \mathbf{g}) = \boldsymbol{\Phi}(\boldsymbol{\Delta}\mathbf{x}_{t-1} - \mathbf{g}) + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\kappa}) + \boldsymbol{\varepsilon}_t \tag{6.2}$$

where $\mathbf{g}$ is a vector of growth rates, $\boldsymbol{\Phi}$ is a full rank (4×4) matrix of coefficients, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $4 \times r$, where r is the number of cointegrating relations; and $\boldsymbol{\kappa}$ is a vector of r cointegrating constants (mean values of the r cointegrating relations).

In general, cointegration implies that $\mathbf{g}$ must satisfy $\boldsymbol{\beta}'\mathbf{g} = \mathbf{0}$, but we also test the additional restriction

$$\mathbf{g} = g\boldsymbol{\iota} \tag{6.3}$$

where $\boldsymbol{\iota}$ is a vector of ones and g a scalar, so that a common deterministic growth rate is imposed, ruling out deterministic bubbles in any of the underlying ratios.

[27] For a direct comparison of systems using unadjusted and adjusted prices and dividends, see Robertson and Wright (2002b), in which we test the restriction that dividends, new issues and buybacks can be treated as equivalent in their predictive power for prices and returns: thus the adjustment required to work in terms of $\mathbf{x}_{at}$ is econometrically acceptable.

[28] Details of both sets of tests can be obtained from the authors on request.

6.2. Testing for q and the Dividend Yield as Cointegrating Relations

In Robertson and Wright (2002a) we show the results of tests for the rank, r, of the system, hence the number of cointegrating relations. While $r = 0$ is clearly rejected, the data do not give a clear signal as to whether $r > 1$: while the hypothesis that there is just one cointegrating relation cannot be rejected at conventional levels, test statistics are at marginal levels of statistical significance, and model selection criteria such as Akaike's point to there being at least two.

We therefore here investigate the properties of two versions, where $\boldsymbol{\beta}$, the matrix of cointegrating relations, is restricted to be consistent with one or both of the valuation ratios.[29] For the $r = 2$ case, we test:

$$\boldsymbol{\beta}_2' = \begin{pmatrix} 1-\zeta & 0 & -1 & \zeta \\ 0 & 1-\zeta & -1 & \zeta \end{pmatrix} \tag{6.4}$$

where the first row imposes (log-linearised) q as a cointegrating relation. The second row imposes the necessary relation between dividends, debt and capital, such that the dividend yield is also a cointegrating relation. Thus if we define $cr_d \equiv (1-\zeta)(d_t - k_t) + \zeta(l_t - k_t)$, the cointegrating relation implied by the second row of $\boldsymbol{\beta}_2$, then $d - p = \frac{1}{1-\zeta}(cr_d - q))$.[30]

We also investigate the $r = 1$ case, in which q is *not* a cointegrating relation, but the dividend yield is (which, given (4.1), is required for returns to be stationary).[31] Thus we investigate:

$$\boldsymbol{\beta}_1' = \begin{pmatrix} -1 & 1 & 0 & 0 \end{pmatrix} \tag{6.5}$$

Table 2 shows that, in line with our reasoning of Section 4, the data easily accept the imposition of the restrictions in the adjusted dataset, but strongly reject on the unadjusted dataset. The table also shows that the additional restrictions imposing the common growth rate have a minimal impact on test statistics.

[29] In Robertson and Wright (*op cit*) we also examine the properties of a system with leverage $(l - p)$ as the additional cointegrating relation, but the evidence for this third relation is weak, as are theoretical priors.

[30] In the forward sum representation of q, (4.11), cr_d is the term in square brackets on the right-hand side. If returns, Δk_t and Δl_t and q_t are all stationary, this term must also be stationary and thus represents a cointegrating relationship.

[31] Thus we do not investigate the case where q is the sole cointegrating relation, since this would imply nonstationary returns.

Table 2 Likelihood Ratio Tests of Restrictions on β and g					
System with Unadjusted Dividends ($\mathbf{x}_{ut}$)					
Sample:		1902-1990		1902-2000	
	d.o.f.	LR	p-value	LR	p-value
$\boldsymbol{\beta}_2$	4	17.09	.002	20.04	.000
$\boldsymbol{\beta}_2$, g	5	17.36	.004	20.20	.001
$\boldsymbol{\beta}_1$	3	9.49	.023	15.69	.001
$\boldsymbol{\beta}_1$, g	5	9.84	.080	15.85	.007
System with Adjusted Dividends ($\mathbf{x}_{at}$)					
$\boldsymbol{\beta}_2$	4	4.59	.332	9.73	.045
$\boldsymbol{\beta}_2$, g	5	4.66	.459	9.76	.082
$\boldsymbol{\beta}_1$	3	2.13	.547	3.98	.264
$\boldsymbol{\beta}_1$, g	5	2.22	.817	4.19	.523

Given the failure of restrictions on the unadjusted data, the remainder of our analysis is carried out solely on the adjusted dataset.

6.3. Tests of Granger Causality

By the Granger Representation Theorem, the existence of cointegration must imply Granger Causality in at least one direction. As noted above, in Section 3, this in itself need not conflict with unpredictable returns if causality only runs to, and not from the "fundamental" (or, in the $r = 2$ case, the two fundamentals). While, in our simplified example of Section 3 this implied no causality to stock price changes, we are really interested in the issue of causality to returns. Using (4.1), returns are a log-linear combination of prices and dividends. Causality tests can therefore be derived as restrictions on parameters of the VAR.[32] Table 3 shows that, using the adjusted dataset, the hypothesis of no causality to returns is strongly rejected for both cointegrating relations in the $r = 2$ case, and for the single cointegrating relation in the $r = 1$ case. Thus both q and the adjusted dividend yield satisfy the second of the two conditions set out in Section 3.

Table 3 Granger Causality Tests					
System with Adjusted Dividends ($\mathbf{x}_{at}$)					
		1902-1990		1902-2000	
Rank	Hypothesis	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
2	$\mathrm{H}_0 : q \nRightarrow r$	17.47	0.000	13.98	0.000
2	$\mathrm{H}_0 : cr_d \nRightarrow r$	5.77	0.020	8.44	0.004
1	$\mathrm{H}_0 : \widetilde{d} - \widetilde{p} \nRightarrow r$	125.36	0.00	124.40	0.000

[32] See Robertson & Wright (2002a) for further details.

7. The Good News About Long-Run Stock Returns

A significant element of the existing predictability literature has been a focus on the power of the dividend yield and other valuation indicators to predict returns at relatively long horizons. The econometric basis for this finding is not actually especially strong, given the relatively small number of non-overlapping observations of long-period returns; nonetheless Figure 2.1 provides quite strong circumstantial evidence for some form of long-horizon predictability in the relatively low dispersion of long-horizon returns. Rather than estimating long-horizon return regressions directly, we derive confidence intervals of long-horizon return forecasts from alternative versions of the cointegrating VAR, where $h-$period ahead forecasts are derived by rolling the one-step ahead VAR forecasts forward recursively h times.[33]

Any such confidence intervals are conditional upon an assumed structure of $\boldsymbol{\beta}$, the matrix of cointegrating coefficients[34]. They are also conditional upon starting values for the forecast - both in terms of the resulting point forecasts at the centre of the bands, and in terms of the impact of parameter uncertainty.

To bring out the impact of different specifications on dispersion, rather than point forecasts, Figure 7.1 compares three implied forecast confidence intervals for cumulative returns, at horizons up to thirty years, on the assumption that the system is forecasting from steady state (*ie*, with all lagged growth rates and cointegrating relations set to their estimated steady-state values). It thus proxies the range of uncertainty associated with return forecasts from the system under "normal" circumstances.Since the underlying system is $I(1)$, the confidence intervals are increasing in the forecast horizon; but at distinctly different rates.

As a basis for comparison, the widest interval in Figure 7.1 shows the degree of uncertainty associated with cumulative predicted return forecasts if returns were assumed to be unpredictable. While, on the basis both of Figure 2.1 and the predictability literature, this interval would clearly be too wide, it is useful both as a benchmark, and in the light of the recent revisionist literature cited above, that has begun to cast doubt on some of the earlier results on predictability.[35]

[33]Confidence intervals for cumulative returns are derived by augmenting the estimated VAR with the Campbell-Shiller approximation for the log stock return, defining log cumulative returns, cr_t such that $\Delta cr_t = r_t$. We parameterise the approximation around the estimated mean dividend yield from the VAR. For further details, see Appendix C.

[34]The confidence intervals allow for the impact of parameter uncertainty in estimates of $\boldsymbol{\alpha}$ the matrix of adjustment coefficients, g, the common growth rate, and $\boldsymbol{\kappa}$, the vector of cointegrating constants. Hence we also allow for uncertainty in the steady-state return. However, in all the systems we examine, r, the rank of the system, and the individual elements of $\boldsymbol{\beta}$ are imposed; we do not therefore allow for rank uncertainty, or parameter uncertainty in cointegrating coefficients.

[35]Note that we generate all the intervals in Figures 7.1 and 8.1, including the interval assuming unpredictable returns, on the assumption of normality of the underlying disturbances. This is in

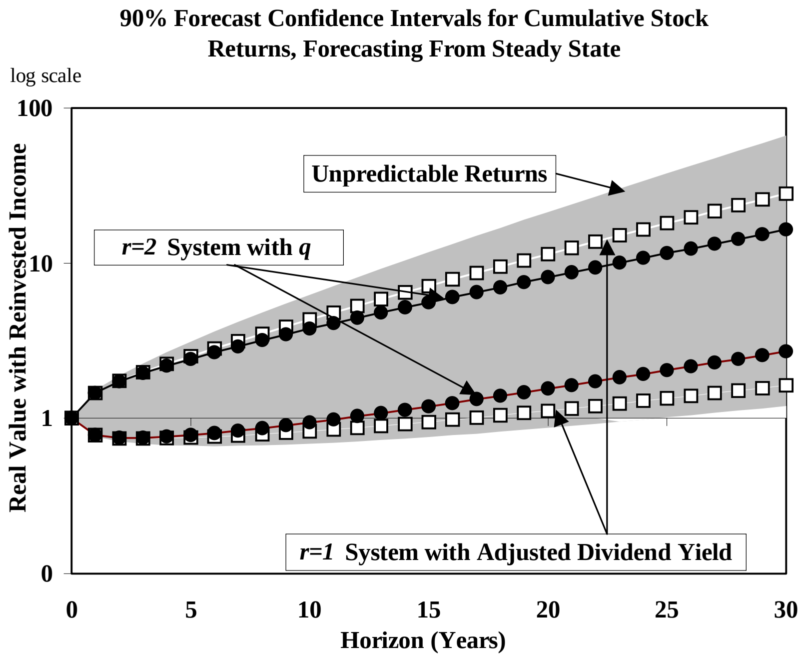

Figure 7.1:

Figure 7.1 shows that, as would be expected, including cointegrating relations reduces the width of forecast intervals for returns compared to the unpredictable returns benchmark; with the degree of variance reduction increasing with the forecast horizon. It thus captures one of the key features of historic returns shown in Figure 2.1. It is notable that there is a distinctly greater reduction in the $r = 2$ system, which includes q, than in the $r = 1$ system, with the adjusted dividend yield only. We discuss this issue further in Section 9 below.

This, then, is the good news about long-run stock returns. Over long horizons, return uncertainty in the cointegrated systems is significantly lower than it would be if returns were unpredictable. The variance at different horizons appears considerably more in line both with the historical evidence shown in Figure 2.1, and with the "Buy-and-Hold" view of investment practitioners - most markedly so when q is included as one of the cointegrating relations.

contrast to the interval in Figure 2.1 which is derived by randomising the observed distribution of annual returns. The two charts make clear that the resulting differences are minor, with the interval in Figure 7.1 being somewhat wider, largely due to parameter uncertainty.

8. The Bad News About Long-Run Stock Returns

The corollary of the good news on the variance of long-horizon returns in the cointegrated systems (and thus the width of the confidence intervals in Figure 7.1) is some degree of bad news on expected returns (the mid-points of the intervals). There are two elements in this bad news: the first is evident in both cointegrated systems; the second only in the $r = 2$ system including q.

The first piece of bad news is that the mean return implied by the cointegrated systems is distinctly lower than the average of realised returns over the twentieth century, which, in a world of unpredictable returns would be the best available forecast of future returns. In the cointegrated systems the estimated mean return is given by:

$$\widehat{\overline{r}} = \widehat{\delta} + \widehat{g}$$

where $\widehat{g}$ is the estimated common growth rate of the four variables in the system, and $\widehat{\delta}$ is the estimated constant term in the Campbell-Shiller approximation (4.1). We calculate $\widehat{\delta}$ in terms of the estimated mean dividend yield in the system.[36] Table 4 shows that in both systems, the resulting estimate of the steady-state return is distinctly lower than the realised mean return over the twentieth century. This difference is largely explicable in terms of the behaviour of returns at the end of the twentieth century. The strength of returns in the 1990s quite significantly raised the estimated mean return for the century as a whole; but had a much more limited impact on the estimated mean returns in the cointegrating systems. The system estimates are also significantly better determined, and imply that there is a correspondingly very low probability that the true mean return is higher than the twentieth century average return.

Table 4. Alternative Estimates of Mean Log Stock Return ($\widehat{\overline{r}}$)

Sample 1900-2000	Unpredictable Returns	$r = 1$ System	$r = 2$ System
Point Estimate	0.0730	0.0661	0.0653
Standard Error*	0.0208	0.0037	0.0042
$\Pr(\overline{r} > .0730)$	0.5	0.031	0.033

*Derived by delta method approximation for the cointegrated systems, using adjusted dataset.

[36] Thus we can allow for parameter uncertainty in both $\widehat{\delta}$ and $\widehat{g}$. From $\widehat{\delta} = \ln\left(1 + \exp(\widehat{\overline{d - e - p}})\right)$. where the estimated mean log dividend yield, $\widehat{\overline{d - e - p}}$, is is the estimated cointegrating constant in the $r = 1$ system, and is $\frac{1}{1-\zeta}(\widehat{\boldsymbol{\kappa}}_2 - \widehat{\boldsymbol{\kappa}}_1)$ in the $r = 2$ system.

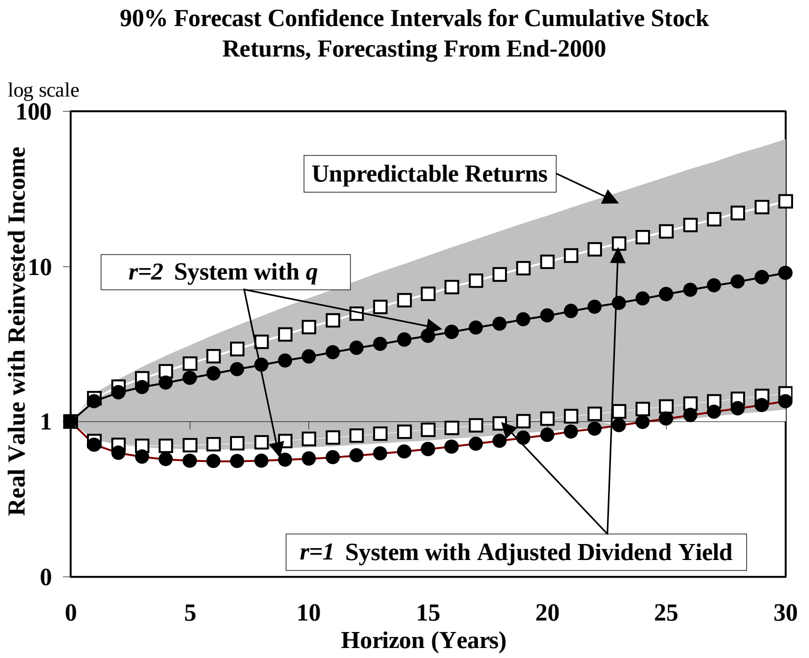

Figure 8.1:

The second, and more significant element of bad news about expected returns arises when forecasts are carried out, not from a position of steady state, but from actual values at end-2000. In the $r = 2$ system, including q, the system was a very long way from its estimated steady state, since, as noted in Section 5.2, q was at historically exceptional levels. As a result, point forecasts from the $r = 2$ system were of a prolonged period of weak returns.

Figure 8.1 shows that as a result, forecasting from end-2000, the 90% confidence interval for future returns was shifted down very significantly, compared to the pattern shown in Figure 7.1. In contrast, the interval assuming unpredictable returns is of course identical to that in Figure 7.1; while that for the $r = 1$ system is little changed, since the sole cointegrating relation, the adjusted dividend yield, was close to its historic mean at end-2000. Thus this element of the bad news is essentially related to the behaviour of q.

9. The Good News and the Bad News About Long-Run Stock Returns

Clearly the nature and extent of good or bad news on expected returns depends on the econometric specification. But, crucially, these differences have a counterpart in the nature and extent of the bad or good news on return variance. Hence some element of both good and bad news is inevitable, whichever specification is closest to the truth.

Thus, Figure 7.1 showed that the $r = 2$ system, including q, implied the narrowest interval for returns, with a particular impact at long horizons. This pattern also roughly corresponds to the observed historic dispersion of returns, shown in Figure 2.1. But the corollary of this distinctly good news, arising from q's predictive power for returns in general, is that there is distinctly bad news when q is at extremely high levels, as it was at end-2000. Furthermore Table 4 showed that there would have been some element of bad news on expected returns, compared to realised returns over the twentieth century, even if the system had been at steady state.

At the opposite extreme, if returns are in fact unpredictable, there is no bad news on expected returns, but there is distinctly bad news on return variance, since it must imply that the historical evidence of low dispersion of returns is a fluke. The Monte Carlo analysis of Appendix A suggests that there was a low probability of this; but that it cannot be entirely ruled out.

In between these two extremes, the $r = 1$ system, with the adjusted dividend yield as the sole cointegrating relation, implies less bad news on expected returns, as of end-2000 (the only element of bad news being the lower estimated steady-state return compared to the average over the twentieth century). But Figure 7.1 showed that it also implied distinctly less good news on return variance.

The explanation for this lies in the relative volatility of the alternative "fundamentals" in the two cointegrating relations: the capital stock versus adjusted dividends. It thus relates to the third characteristic of a useful valuation criterion, as set out in Section 3. In our cointegrated systems stock prices are pulled towards other elements in the system, the "fundamentals". The impact of this in reducing forecast uncertainty at long horizons depends crucially on how forecastable these measures of "fundamentals" are. Figure 5.1 showed that adjusted dividends are considerably more volatile than unadjusted dividends; which are in turn more volatile than the capital stock. This ranking is transmitted to forecast uncertainty in total returns. However, in the $r = 1$ case, with the adjusted dividend yield as the single cointegrating relation, the forecasting advantage arising from the relative predictability of the capital stock disappears, and by implication, long-horizon returns are distinctly more uncertain.

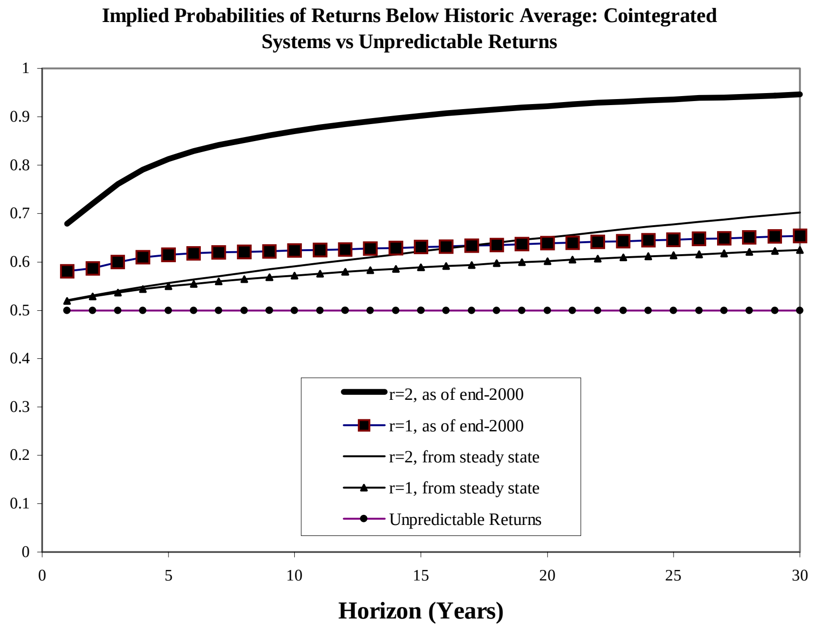

Figure 9.1:

The contrast between these different combinations of good and bad news on returns and return variance can be brought out by comparing the implied probabilities of either weak (defined as less historic average) or negative returns at different horizons, in the three specifications. These are shown in Figures 9.1 and 9.2, which show probabilities when forecasting both from an initial steady state, and from actual end-2000 values.

Figure 9.1 shows that, whereas unpredictable returns give an equal chance of returns below or above the twentieth century average, the cointegrated systems always imply a distinctly higher probability that returns will be below that average, even when forecasting from steady state. For the $r = 1$ case, the difference in starting values has a relatively modest impact; whereas in the $r = 2$ system including q, it has a dramatic effect, with a very high probability that returns at long horizons will be below the twentieth century average.

Figure 9.2 brings out the link between the good or bad news on return variability with the associated bad or good news on expected returns. The lowest of the lines on the chart shows the probability of negative returns in the $r = 2$ case when forecasting from steady state. The predictive power of q means that the probability declines quite rapidly with the forecast horizon, falling below 5%

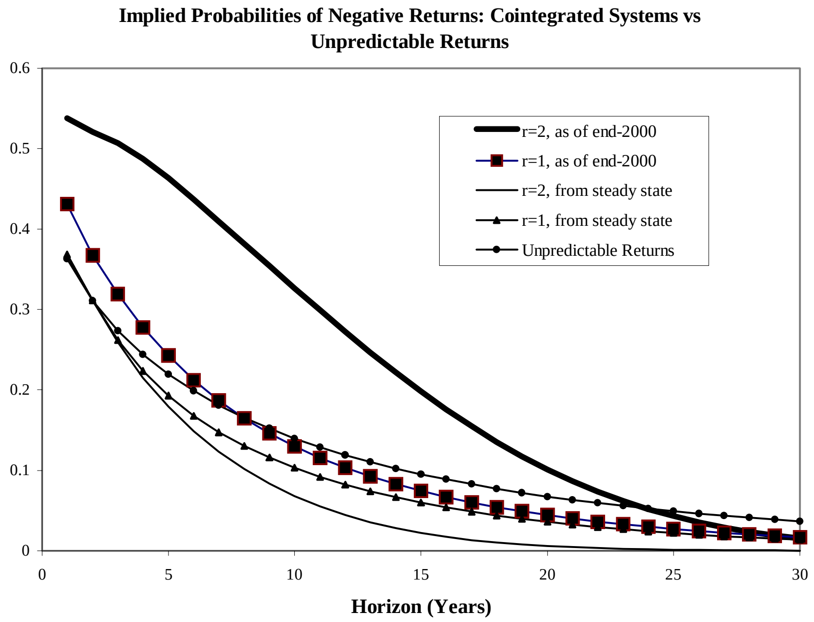

Figure 9.2:

at around twelve years, and becoming trivially small at horizons above 20 years. This is very much in line with the observed distribution of historic returns at different horizons, shown in Figure 2.1.

The highest of the lines on the chart shows, however, that when forecasting from end-2000 values the extremely high level of q implied much higher probabilities of negative returns than in average circumstances, with still quite significant risks of negative returns at horizons of as long as ten or even twenty years. This was in an environment when alternative assets were yielding strongly positive returns - at the time indexed US Treasury bonds, for example, yielded a virtually safe 4%.[37]

The pattern shown by the other lines on the chart illustrates the nature of the link between the good and bad news. Both the $r = 1$ system and the unpredictable returns specification imply distinctly lower probabilities of negative returns at shorter horizons compared to the $r = 2$ system forecasting from end-2000 (since the last observation has little or no impact on the pattern of forecast returns). At short horizons, the probabilities are similar to those of the q-based system, when

[37] The $r = 2$ system implied that the probability of under-performing such a bond was around two thirds at a ten year horizon, and still around one half at a thirty year horizon.

forecasting from steady state. But as the horizon lengthens, the probabilities decline more slowly, such that the risk of negative returns remains significant at even fairly long horizons. Thus the lack of bad news on expected returns is offset by lack of good news on return dispersion.

10. Conclusions

If stock returns were unpredictable, uncertainty about future returns would be so great as to be inconsistent either with historical experience, or with claims that stock returns are relatively safe over long horizons. The good news about long-run stock returns is that if, as a growing body of research suggests, there is even a weak tendency for stationary valuation indicators to predict future stock prices, long-run returns are distinctly more predictable than the unpredictable returns model would imply.

We started by asking what general empirical characteristics any useful (by which we mean long-run variance reducing) valuation criterion should have. There are three such key characteristics: mean reversion, Granger-causality to returns, and relative predictability of the fundamental. Tobin's q appears to possess all three characteristics, as well as having attractive theoretical properties. Standard measures of the dividend yield appear to fail on the first count. If the dividend yield is adjusted for buybacks and new issues, it easily satisfies the first two requirements, but does less well on the third.

The bad news about long-run stock returns is a corollary of the good news. The better is the news on implied long-run variance of returns, the worse the news on expected returns, especially given the still historically high level of q at the end of 2000.

Appendices

A. Assessing the Probability of the Observed "Safety" and Dispersion of Historic Horizon Returns, given Randomised Returns.

We wish to investigate the probability of two key features of the observed historical dataset occurring, given randomised real returns.

One feature of the historic sample visible in Figure 2.1 is the apparent degree of "safety" of returns, defined as a low proportion of negative returns beyond a certain horizon. As Figure 2.1 showed, the 5th percentile of historic returns at different horizons (the lower bound of the interval shown in that chart) crosses zero at a fourteen year investment horizon. Thus, beyond this horizon, the probability of a negative return appears to have been less than 5%.[38] In contrast, the simulated population percentile assuming randomised returns, shown in the same chart, does not cross zero until around twenty years; and the divergence between the two lines increases with the horizon. However, given the relatively small number of observations, we wish to assess the probability that at least the same degree of apparent "safety" could have arisen by chance in any given sample of the same length, drawn randomly from a population of randomised returns.

We simulate 10,000 artificial samples of annual returns, each with 197 observations (the same number of observations as in the data), generated as random draws from the historical distribution of annual log real returns over the sample 1802-1998.[39] We then estimate:

$$
\begin{aligned}
p_1(h) &= \Pr\left[Q\left(0.05, \{c_t(\sigma, h)\}\right) > Q\left(0.05, \{c_t(d, h)\}\right)\right] \\
& \text{ where} \\
Q\left(k, \{c_t(s, h)\}\right) &= \text{ the } k \times 100 \text{ percentile of real returns at horizon } h\text{, in sample } s;\; s = d, \sigma \\
c_t(s, h) &= \Sigma_{i=0}^{h-1} r_{t-i}(s), \text{ is the cumulative return over horizon } h \text{ in sample } s \\
\{r_t(d)\} &= \text{ the sequence of annual returns in the data sample} \\
\{r_t(\sigma)\} &= \text{ any randomised sequence of 197 annual real returns}
\end{aligned}
$$

Thus $p_1(h)$ is the probability of any randomised sample of 197 returns implying at least as great a degree of apparent "safety", at horizon h, as in the observed

[38] Indeed, as Siegel (*op cit*) notes, in the entire sample there have never been negative real returns at horizons from twenty years onwards.

[39] The *Gauss* program used to generate the simulations (and also to generate the intervals shown in Figure 2.1) is available from the authors on request.

data sample. Figure A.1 shows that this probability is far from trivial (in the range 10% to 30%) at all horizons, though lower at longer horizons. Thus, even in a world of unpredictable returns, there exists a reasonable probability of observing the apparent degree of safety that US stocks have displayed over.

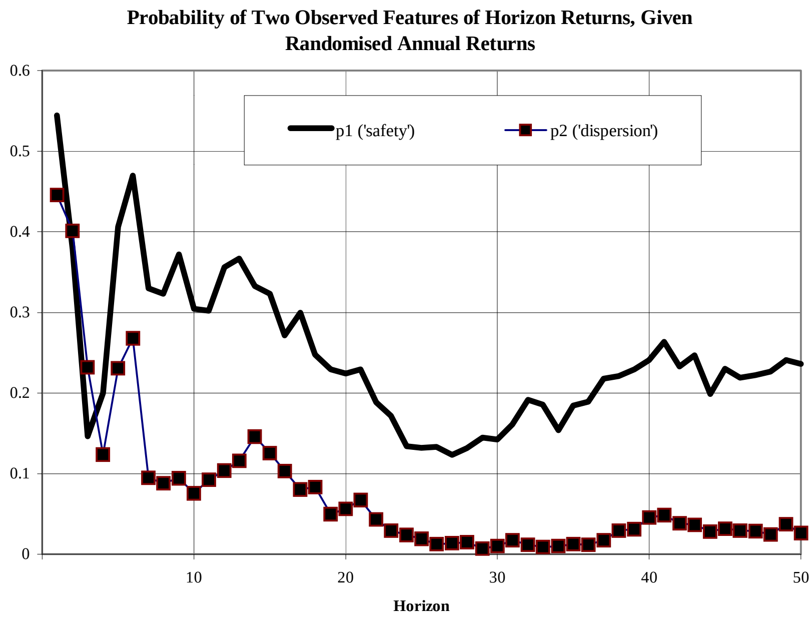

Figure A.1:

The measure $p_1(h)$ is affected not purely by the dispersion of returns in any given sample, but also by variations in the means of the generated samples. An alternative measure of dispersion focusses solely on the width of the interval in Figure 2.1, rather than its position, thus effectively ignoring any impact of the sample mean return. To assess this, we estimate, using the same simulated samples of annual returns:

$$\begin{aligned} p_2(h) &= \Pr\left[I\left(\{c_t(\sigma,h)\}\right) < I\left(\{c_t(d,h)\}\right)\right] \\ &\quad \text{where} \\ I\left(\{c_t(s,h)\}\right) &= Q\left(0.95,\{c_t(s,h)\}\right) - Q\left(0.05,\{c_t(s,h)\}\right), \; s=\sigma,d \end{aligned}$$

Thus $p_2(h)$ is the probability of any randomised sample of 197 annual returns having a narrower 90% interval of returns, at horizon h, than the data. Figure A.1

shows that this probability is distinctly lower at all except the shortest horizons, although at no horizon is it vanishingly small (the probability of a lesser degree of dispersion 30 year returns than in the data sample is just under 1%).

The simulations thus suggest that, while the low frequency of negative long-horizon returns could have arisen relatively easily by chance in a world of unpredictable returns, the low observed degree of dispersion of long-horizon returns could only have occured with a small (though not vanishingly small) degree of probability.

B. Data Definitions for the Cointegrating VAR

All data are taken from the dataset described in Wright (2001). We refer interested readers to that paper for information on sources and definitions of underlying data. The table below gives definitions in terms of variable codes given out in the appendix to the same paper.

Variable	Unadjusted Dataset	Variable	Adjusted Dataset
P	SP/PCEY/E	$\widetilde{P}$	SP/PCEY
D/E	DIV/PCEY/E	$\widetilde{D}$	(DIV - NI)/PCEY
K/E	KBEA/PCEY/E	K	KBEA/PCEY
L/E	NLMBEA/PCEY/E	L	NLMBEA/PCEY
Q	(MV+NLMBEA)/KBEA		

where:
SP = Nonfinancial Share Price, 1945=1
PCEY= End-year CPI[40]
E = MV/SP
where MV = Market Value of Nonfinancial Equities, $bns
DIV = Nonfinancial dividends, $bns
NI = New Issues, net of Buybacks, $bns
KBEA = Capital Stock[41]

[40] We deflate by the end-year CPI because this is the most reliable and consistent price index to span the entire period (it is also arguably most appropriate when we approximate real returns). Results using a (less consistently defined) measure of the nonfinancial GDP deflator (PGDP, in Wright's (2001) dataset) are however very similar, and are available from the authors on request. Note that all series must of necessity be deflated by the same price index if underlying cointegrating relations q, $d-p-e$ (or $\widetilde{d}-\widetilde{p}$) are to be invariant to the method of deflation. As a result, implied changes in the "real" capital stock correspond, effectively, to changes in the real opportunity cost of the value of the capital stock, in terms of foregone consumption.

[41] We use a consistent measure of capital, taken (where available) direct from the BEA, rather than the series from the Fed's flow of funds, in which there is a distinct discontinuity from 1990 onwards(see Wright, 2001 for further discussion).

NLMBEA = Net Liabilities[42]

C. Forecast Confidence Intervals For Cumulative Returns in the Cointegrated System[43]

Our cointegrating VAR representation in (6.2)

$$(\boldsymbol{\Delta}\mathbf{x}_t - \mathbf{g}) = \boldsymbol{\Phi}(\boldsymbol{\Delta}\mathbf{x}_{t-1} - \mathbf{g}) + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\kappa}) + \boldsymbol{\varepsilon}_t$$

can be reparameterised as:

$$\boldsymbol{\Delta}\mathbf{x}_t = \boldsymbol{\Psi} + \boldsymbol{\Phi}\boldsymbol{\Delta}\mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{C.1}$$

with

$$\boldsymbol{\Psi} = (\mathbf{I} - \boldsymbol{\Phi})\mathbf{g} - \alpha\boldsymbol{\kappa}$$

We assume $\boldsymbol{\varepsilon}_t \sim NIID(\mathbf{0}, \boldsymbol{\Sigma})$.

Using the Campbell-Shiller approximation for the log stock return in (4.1), we can define the cumulative return, cr_t at time t by:

$$\begin{aligned} \Delta cr_t &= r_t \\ &= \varphi + \boldsymbol{\eta}'\Delta x_t + \boldsymbol{\lambda}'\boldsymbol{\beta}'\mathbf{x}_{t-1} \end{aligned} \tag{C.2}$$

where

$$\boldsymbol{\eta}' = \begin{pmatrix} \rho & 1-\rho & 0 & 0 \end{pmatrix}$$

and $\boldsymbol{\lambda}'$ is an $(1 \times r)$ vector that selects $(1-\rho)(d_t - p_t)$ from the set of cointegrating vectors.[44]

We can augment the system in (C.1) with (C.2) for cr_t, and rewrite the augmented VAR as:

$$\begin{bmatrix} 1 & -\varphi & -\boldsymbol{\eta}' & -\boldsymbol{\lambda}' \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta cr_t \\ 1 \\ \Delta\mathbf{x}_t \\ \boldsymbol{\beta}'\mathbf{x}_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \boldsymbol{\Psi} & \boldsymbol{\Phi} + \boldsymbol{\alpha}\boldsymbol{\beta}' & \alpha \\ 0 & 0 & \boldsymbol{\beta}' & I \end{bmatrix} \begin{bmatrix} \Delta cr_{t-1} \\ 1 \\ \Delta\mathbf{x}_{t-1} \\ \beta'\mathbf{x}_{t-2} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{\varepsilon}_t \\ 0 \end{bmatrix}$$

[42] Adjusted to ensure consistency with both Fed net worth figures and BEA capital stock figures. The adjustment only applies after 1989. See Wright (2001) for further discussion.

[43] The *Gauss* program that implements the techniques described in this appendix, and was used to generate Figures 7.1 and 8.1, is available on request from the authors.

[44] In the $r = 1$ case, $\boldsymbol{\lambda} = 1 - \rho$; in the $r = 2$ case, $\boldsymbol{\lambda} = \left(\frac{1-\rho}{1-\zeta}\right) \begin{bmatrix} -1 & 1 \end{bmatrix}$.

implying:

$$\mathbf{z}_t = \mathbf{D}\mathbf{z}_{t-1} + \mathbf{v}_t \tag{C.3}$$

with $\mathbf{z}_t = \left[\begin{array}{cccc} \Delta cr_t & 1 & \Delta \mathbf{x}_t & \boldsymbol{\beta}'\mathbf{x}_{t-1} \end{array} \right]'$, $\mathbf{v}_t = \left[\begin{array}{cccc} \boldsymbol{\eta}'\boldsymbol{\varepsilon}_t & 0 & \boldsymbol{\varepsilon}_t & \mathbf{0} \end{array} \right]'$ and

$$\mathbf{D} = \left[\begin{array}{cccc} 0 & \varphi + \boldsymbol{\eta}'\boldsymbol{\Psi} & \boldsymbol{\eta}'(\boldsymbol{\Phi} + \boldsymbol{\alpha}\boldsymbol{\beta}') + \boldsymbol{\lambda}'\boldsymbol{\beta}' & \boldsymbol{\eta}'\boldsymbol{\alpha} + \boldsymbol{\lambda}' \\ 0 & 1 & 0 & 0 \\ 0 & \boldsymbol{\Psi} & \boldsymbol{\Phi} + \boldsymbol{\alpha}\boldsymbol{\beta}' & \boldsymbol{\alpha} \\ 0 & 0 & \boldsymbol{\beta}' & \mathbf{I} \end{array} \right] \tag{C.4}$$

The parameters of the system to be estimated are $\boldsymbol{\alpha}, \boldsymbol{\Phi}, \boldsymbol{\kappa}$ and the common growth rate g (since, as noted in the main text, we impose $\mathbf{g} = g\boldsymbol{\iota}$ where $\boldsymbol{\iota}$ is a vector of 1s), with covariance matrix Σ_δ. We can also define φ, the constant term in the Campbell-Shiller approximation, to be consistent with the estimated steady-state dividend yield in the system, as:

$$\varphi = \ln\left(1 + \exp(\frac{\boldsymbol{\lambda}'\boldsymbol{\kappa}}{(1-\rho)})\right) - (1-\rho)\frac{\lambda'\boldsymbol{\kappa}}{(1-\rho)} \tag{C.5}$$

To forecast cumulative returns from the terminal observation of the sample, T, we can write:

$$cr_{T+h} - cr_T = \sum_{i=1}^{h} \Delta cr_{T+i} = \sum_{i=1}^{h} \mathbf{J}\mathbf{z}_{T+i} = \mathbf{J}\sum_{i=1}^{h}\left(\mathbf{D}^i\mathbf{z}_T + \sum_{j=0}^{i-1}\mathbf{D}^j\mathbf{v}_{T+i-j}\right) \tag{C.6}$$

with selection vector $\mathbf{J} = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \end{array} \right]$.

Write $cr_{T+h|T}$ for the forecast of cr_{T+h} made at T if all the parameters were known, and $\widehat{cr}_{T+h|T}$ the forecast where the parameters are estimated on the basis of information available at T. Then we can partition the error in forecasting cumulative returns from T by writing

$$cr_{T+h} - \widehat{cr}_{T+h|T} = \left(cr_{T+h} - cr_{T+h|T}\right) + \left(cr_{T+h|T} - \widehat{cr}_{T+h|T}\right) \tag{C.7}$$

The two right hand side terms are orthogonal by the assumption that the innovations, $\boldsymbol{\varepsilon}_t$, are serially independent. The first term captures future uncertainty and the second parameter uncertainty.

The first term in (C.7) is:

$$\left(cr_{T+h} - cr_{T+h|T}\right) = \mathbf{J}\sum_{i=1}^{h}\left(\sum_{j=0}^{i-1}\mathbf{D}^j\mathbf{v}_{T+i-j}\right) \tag{C.8}$$

hence future uncertainty is given by:

$$Var\left(cr_{T+h} - cr_{T+h|T}\right) = \sum_{i=1}^{h} \mathbf{J}\left(\sum_{j=0}^{i-1}\mathbf{D}^j\right)\boldsymbol{\Sigma}_{\mathbf{v}}\left(\sum_{j=0}^{i-1}\mathbf{D}^j\right)\mathbf{J}' \tag{C.9}$$

with $\boldsymbol{\Sigma}_{\mathbf{v}}$ the covariance matrix of $\mathbf{v}_t$ which is

$$\begin{bmatrix} \boldsymbol{\eta}'\boldsymbol{\Sigma}\boldsymbol{\eta} & 0 & \boldsymbol{\eta}'\boldsymbol{\Sigma} & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} \\ \boldsymbol{\Sigma}\boldsymbol{\eta} & 0 & \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{C.10}$$

The second term in (C.7), that part of the forecast error due to parameter uncertainty, is

$$cr_{T+h|T} - \widehat{cr}_{T+h|T} = \mathbf{J}\sum_{i=1}^{h}(\mathbf{D}^i - \widehat{\mathbf{D}}^i)\mathbf{z}_T \tag{C.11}$$

The parameters in $\mathbf{D}$ can be collected in the vector $\boldsymbol{\delta} = (g, \boldsymbol{\kappa}, vec(\boldsymbol{\alpha}), vec(\boldsymbol{\Phi}))'$. The estimated parameter vector $\widehat{\boldsymbol{\delta}}$ has estimated covariance matrix $\boldsymbol{\Sigma}_{\delta}$, hence we can approximate by the delta method

$$(cr_{T+h|T} - \widehat{cr}_{T+h|T}) \sim N\left[0, \frac{\partial cr_{T+h|T}}{\partial \boldsymbol{\delta}'}\boldsymbol{\Sigma}_{\delta}\frac{\partial cr_{T+h|T}}{\partial \boldsymbol{\delta}'}'\right]$$

Which can be calculated using

$$\begin{aligned}
\frac{\partial cr_{T+h|T}}{\partial \boldsymbol{\delta}'} &= \frac{\partial vec(\mathbf{J}\sum_{i=1}^{h}\mathbf{D}^i\mathbf{z}_T)}{\partial \boldsymbol{\delta}'} \\
&= \sum_{i=1}^{h}\frac{\partial vec(\mathbf{J}\mathbf{D}^i\mathbf{z}_T)}{\partial \boldsymbol{\delta}'} \\
&= \sum_{i=1}^{h}(\mathbf{z}_T' \otimes \mathbf{J})\frac{\partial vec(\mathbf{D}^i)}{\partial \boldsymbol{\delta}'} \\
&= \sum_{i=1}^{h}(\mathbf{z}_T' \otimes \mathbf{J})\left(\sum_{j=0}^{i-1}(\mathbf{D}')^{i-1-j} \otimes \mathbf{D}^j\right)\frac{\partial vec\mathbf{D}}{\partial \boldsymbol{\delta}'}
\end{aligned}$$

References

[1] Abel, A and Blanchard, O (1983), "An Intertemporal Model of Saving and Investment" *Econometrica* 51(3) pp 675-692

[2] Abel, A and Blanchard, O (1986), "The Present Value of Profits and Cyclical Movements in Investment" *Econometrica* 54(2) pp 249-273

[3] Bekaert, G & Hodrick, R (1992) "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets" *Journal of Finance* 47(2),pages 467-509.

[4] Brainard, W C & Tobin, J (1968), "Pitfalls in Financial Model Building", *American Economic Review*, Vol LVIII, No 2

[5] Breen, W, Glosten, L & Jagannathan, R (1989) "Economic Significance of Predictable Variations in Stock Index Returns" *Journal of Finance*; 44(5), pages 1177-89.

[6] Bossaerts, P and Hillion, P (1999), "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn", *Review of Financial Studies,* Summer 1999, vol 12, no. 2, pp 405-28

[7] Campbell, J & Ammer, J (1993) "What Moves the Stock and Bond Markets? A Variance Decomposition for Long Term Asset Returns", *Journal of Finance* 48(1), pages 3-37.

[8] Campbell, J Y & Shiller R J (1988), "Interpreting Cointegrated Models", *Journal of Economic Dynamics and Control* 12(2/3) pages 505-22.

[9] Campbell J Y and Shiller R J (1998) "Valuation Ratios and the Long-Run Stock Market Outlook", *Journal of Portfolio Management*, Fall 1998

[10] Campbell, J, Lo, A J and MacKinlay, A C (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ

[11] Chiang, R, Liu, P & Okunev, J (1995),"Modelling Mean Reversion of Asset Prices towards Their Fundamental Value", *Journal of Banking and Finance* 19(8), pages 1327-40.

[12] Chopra, N, Lakonishok, J & Ritter, J (1992), "Measuring Abnormal Performance: Do Stocks Overreact?", *Journal of Financial Economics* 31(2),pages 235-68.

[13] DeBondt, W, Thaler, R (1985) "Does the Stock Market Overreact?", *Journal of Finance*; 40(3), July 1985, pages 793-805.

[14] De Long, J B, Shleifer, A, Summers, L H & Waldmann, R J (1990) "Noise Trader Risk in Financial Markets", *Journal of Political Economy*, Vol 98, no 4, pp 703-738.

[15] Engle, R.F. and Granger, C.W.J. (1987) "Co-integration and error correction: representation, estimation and testing", *Econometrica*, vol 55, pp 251-76.

[16] Fama, E and French, K (2001) "The Equity Premium", University of Chicago Graduate School of Business *Working Paper*

[17] Foster, F D, Smith, T and Whaley, (1997) R "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2", *Journal of Finance,* Vol LII No 2.

[18] Granger, C. W. J. (1986) "Developments in the study of cointegrated economic variables", *Oxford Bulletin of Economics and Statistics*, vol. 48, pp 213-28.

[19] Grossman, S, Stiglitz, J (1980), "On the Impossibility of Informationally Efficient Markets", *American Economic Review*, 70(3), June 1980, pages 393-408.

[20] Harvey, AC, and Jaeger, A. (1991). "Detrending, stylized facts and the business cycle." *Journal of Applied Econometrics*, 8: pages 231-247.

[21] Hayashi, F(1982), "Tobin's Marginal q and Average q: A Neoclassical Interpretation", *Econometrica*, 50(1), Jan. 1982, pages 213-24.

[22] Heaton, J C & Lucas, D J (1999), "Stock Prices and Fundamentals", in B Bernanke & J Rotemberg (eds) *1999 Macroeconomics Annual*, Cambridge, MIT Press

[23] Jegadeesh, N (1990), "Evidence of Predictable Behavior of Security Returns", *Journal of Finance* 45(3), pages 881-98.

[24] Johansen, S (1988) "Statistical analysis of cointegration vectors" *Journal of Economic Dynamics and Control*, vol 12, pp.231-54.

[25] Kirby, C (1997), "Measuring the Predictable Variations in Stock and Bond Returns", *Review of Financial Studies,* Fall 1997 Vol 10, no. 3, pp 5-630

[26] Kocherlakota, N (1996), "The Equity Premium: It's Still a Puzzle", *Journal of Economic Literature*, Vol 34, pp 42-71.

[27] Mehra, R (1998) "On the volatility of stock prices: an exercise in quantitative theory" *International Journal of Systems Science*, Vol 29, 11, pp 1203-1211

[28] R Mehra and E Prescott (1985), "The Equity Premium: a puzzle", *Journal of Monetary Economics*, March 1985, pp 145-161.

[29] Miller, M H and Modigliani, F, (1961), "Dividend policy, growth, and the valuation of shares", *Journal of Business*, 34, pp 411-433.

[30] Pesaran, M H & Timmermann,A F (1995) "Predictability of Stock Returns: Robustness and Economic Significance" *Journal of Finance*; 50(4), pp 1201-28

[31] Poterba, J M and Summers, L H (1988), "Mean Reversion in Stock Prices: Evidence and Implications." *Journal of Financial Economics* vol 6 (4) pp 333 64

[32] Robertson, D and Wright, S (2002a), "What does q Predict?", *Mimeo*

[33] Robertson, D and Wright, S (2002b), "The (Near-) Equivalence of Dividends, New Issues and Buybacks in Predicting Aggregate Stock Market Returns" *Mimeo*

[34] Shiller, R (1981) "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review,* LXXI, pp 421-3698

[35] Shiller, R (2000) *Irrational Exuberance*, Princeton University Press

[36] Siegel, J (1992) "The Real Rate of Interest from 1800-1990: A Study of the U.S. and the U.K.", *Journal of Monetary Economics* 29(2), pages 227-52.

[37] Siegel, J (1998) *Stocks for the Long Run*, McGraw-Hill

[38] Smithers, A and Wright, S (2000) *Valuing Wall Street* McGraw-Hill

[39] Wright, S (2001) "Measures of Stock Market Value and Returns for the US Nonfinancial Corporate Sector, 1900-2000", *Mimeo*