The Limits to Stock Return Predictability*

Donald Robertson and Stephen Wright†

September 10, 2009

Abstract

We examine predictive return regressions from a new angle. We ask what observable univariate properties of returns tell us about the "predictive space" that defines the true predictive model: the triplet $\left(\lambda, R_x^2, \rho\right)$, where λ is the predictor's persistence, R_x^2 is the predictive R-squared, and ρ is the "Stambaugh Correlation" (between innovations in the predictive system). When returns are nearly white noise, and the variance ratio slopes downwards, the predictive space can be tightly constrained. Data on real annual US stock returns suggest limited scope for even the best possible predictive regression to out-predict the univariate representation, particularly over long horizons.

Keywords: Predictive return regressions, variance ratio, fundamental and non-fundamental univariate representations.

*We are grateful for comments from Kenneth French, Andrew Harvey, Alan Timmerman and Ron Smith.
†Faculty of Economics, University of Cambridge, dr10011@econ.cam.ac.uk and Dept of Economics, Birkbeck College, University of London, s.wright@bbk.ac.uk

1 Introduction

A perennial problem in empirical finance is the question of whether it is possible to predict stock market returns using some predictor variable. In addition to the question of what predictor variable to use (the list of potential predictors includes, *inter alia,* dividend yields, interest rates, book-market values, price-earnings ratios), there are econometric questions regarding the appropriate distributional theory for inference (Stambaugh 1999, Campbell and Yogo 2006, Ang and Bekaert 2006 and many others); whether any underlying relationships are stable enough to allow useful predictability, or may simply arise from data mining (Pesaran and Timmerman 1995, Timmermann and Paye 2006, Ferson et al 2003, Goyal and Welch 2003, Campbell and Thomson 2008, Cochrane 2008a); and whether observable predictors may be at best imperfect proxies for the true predictor (Pastor & Stambaugh, 2009). Finally there is a closely related literature that addresses differences between one period ahead and long horizon regressions (see for example Campbell and Viceira 2002; Cochrane, 2008a; Boudoukh et al, 2008).

In this paper we examine predictive return regressions from a new angle. It is well-known that when one time series predicts another the properties of the predictive system (including the univariate properties of the predictor variable) determine the univariate properties of the predicted variable (in this case returns).[1] But we can also view the process in reverse. In this paper we ask what the observable univariate properties of returns tell us about the "predictive space" that defines the true predictive model: the triplet (λ, R_x^2, ρ) where λ is the predictor's persistence, R_x^2 is the predictive R^2, and ρ is the "Stambaugh Correlation" (between the innovations to the predictor autoregression and those in the predictive return regression).[2]

If these insights depended on tight estimates of parameters in an ARMA representation our analysis would not be of much practical value. But in fact the reverse is the case. The very fact that ARMA representations of returns fit so *poorly* is informative about the nature of the predictive space for predictive models in general. But a second feature of returns data

[1] Note that we are assuming that the predictor is, in Pastor & Stambaugh's (2009) terminology, a "perfect predictor", ie one that captures "true" expected returns, and hence, up to a white noise expectational error, also generates the actual data for returns.

[2] In this paper we focus on a widely used predictive regression framework that can be reduced to just three parameters. Some of our key results also extend to more general predictive models.

also has potentially very significant implications for the predictive space: whether the variance ratio of long-horizon returns slopes downwards. The academic literature that directly tests for this latter feature has not yielded unanimous conclusions (contrast, for example, the orginal evidence presented by Poterba and Summers 1988 with the revisionist approach of, *eg,* Kim et al 1991. But we would argue that the implications of what we term "variance compression"[3] are worth considering, first, because it is usually taken for granted (whether explicitly or implicitly) by investment practioners as the basis for the buy-and-hold strategy (for the classic explicit statement of this view see Siegel, 1998); and second, because the analysis we present in this paper will lead us to argue that it is also *im*plicit in a much wider range of literature that assumes predictability of returns, especially over long horizons (for example, Campbell & Viceira 2002, Cochrane, 2008a).

Since a declining variance ratio is a univariate property it cannot co-exist with returns being *completely* unpredictable from their own past, although we show that there is no necessary contradiction between a quite significantly declining variance ratio at long investor horizons and a very weak degree of short-term univariate predictability. However, the combination of these two features *does* have significant implications for the predictive space that contains all logically possible true predictive regressions. We show that the predictive space for stock returns can quite easily contract to such an extent that there is little, if any, scope for predictive regressions to out-predict the univariate representation, particularly at long horizons.

Recent stock market movements have been a reminder of the continuing significance of this issue. Many predictors of stock returns originate as valuation indicators. In the late 1990s most were signalling that the market was "over-valued", in the very broad sense of the word[4] that such indicators were predicting weak returns (see, for example, Campbell & Shiller, 1998;Shiller, 2000; Fama & French, 2002). In the more recent past, as markets have weakened sharply, the issue has arisen of when they become sufficiently "cheap" to offer a good prospect of unusually strong returns. Most such indicators can be interpreted as the ratio of the stock price to some

[3]For reasons noted below we prefer this term to the more commonly used "mean reversion", which a) is not always consistently defined by different authors; and b) we argue can be a misnomer.

[4]That is, without putting any necessary interepretation in terms of market effficiency: contrast the perspectives of Shiller, 2000 and Fama & French, 2002, for example.

estimate of the underlying fundamental. Our analysis suggests that even if the best possible measure of the fundamental were found, its price ratio might barely predict any better than simply using the history of returns to forecast future returns.

The paper proceeds as follows.

We first, in Section 2, briefly summarise the evidence for the two key empirical features of stock returns that motivate our analysis: a low univariate R^2 and a declining variance ratio. We then show, in Section 3, the links between the standard predictive regression framework and the univariate representation of returns.

Our key results are in Section 4, where we show how observable univariate properties of returns can restrict the predictive space. We show that we show that we can derive lower and upper bounds for the predictive R^2 of the true predictor that depend solely on univariate properties.[5] We then show that a declining variance ratio for long-horizon returns implies that the Stambaugh Correlation, ρ, is in general bounded away from zero, and for a plausible range of ARMA parameters can be close to unity. A high Stambaugh Correlation is usually treated as a nuisance that complicates inference; we show that it is an intrinsic characteristic of the true predictor of stock returns if there is variance compression.[6] Finally we show a further feature that arises from these restrictions on the predictive space: the correlation between the true predictor and a "pseudo predictor", derived solely from the history of returns, is also bounded away from zero. It approaches unity as the predictive space contracts; however even when the predictive space is quite tightly constrained this correlation can still be some way below unity.

In Section 5 we provide an empirical illustration. Point estimates derived from estimated ARMA representations of returns suggest that the predictive space is very tightly constrained.

[5] The sole case in which this result lacks any content arises if the predictor has persistence $\lambda = 0$, and returns are white noise. In this limiting case the predictor may predict as poorly as the ARMA representation (ie, have an R^2 of zero) or (in the limit) perfectly, with an R^2 of unity (albeit in the latter case only if by divine dispensation, if the predictor this year is simply next year's return). But in all other cases the bounds are non-trivial, in the sense that either upper or lower bound, or both, lie strictly between zero and unity.

[6] This feature is closely related both to Pastor & Stambaugh's (2009) analysis of the correlation between expected and unexpected returns and the present value based analysis of Cochrane (2008). However while these authors base their arguments on a priori reasoning; our analysis is based solely on the observable phenomenon of variance compression. Note that our conceptualisation of the true predictor as a price-fundamental ratio, coupled with variance compression, means that innovations will be of the same sign as innovations to returns, and hence of opposite sign to innovations to *expected* returns. Our ρ, is therefore of opposite sign, but otherwise identical, to the correlation between expected and unexpected returns in Pastor & Stambaugh (2008).

The estimate of the best possible predictive R^2 of any possible predictor is around 10%, and of the minimum absolute Stambaugh Correlation is at least 0.9. The implied lower bound for the correlation between the best possible predictor and the "pseudo predictor" is lower, but still a long way from zero. These estimates therefore suggest limited scope for even the best possible predictor of returns to out-perform the univariate representation.

Given the imprecision of ARMA estimation, we acknowledge that a quite wide range of near-white noise processes could also be generating the return series, for some of which the predictive space is less constrained. However, a key requirement for a significantly *less* constrained predictive space is that the true predictor must have quite low persistence. Our analysis therefore suggests two simple pre-tests for potential predictors of stock returns: they should not look too similar to the "pseudo predictor" that summarises the history of returns; and they should not be too persistent. It is notable that few, if any, commonly used predictors of stock returns match up to either of these criteria.

Finally, in Section 6 we note a further implication of our analysis for the the predictability literature. The strength of long-horizon return predictability is driven by a combination of predictor persistence and a declining variance ratio; but it is exactly these features that lead to a tightly restricted predictive space. It therefore follows that significant long-horizon return predictability, if it exists, must be close to being a univariate phenomenon.

In Section 7 we draw conclusions and implications of our analysis; appendices provide algebraic derivations and proofs.

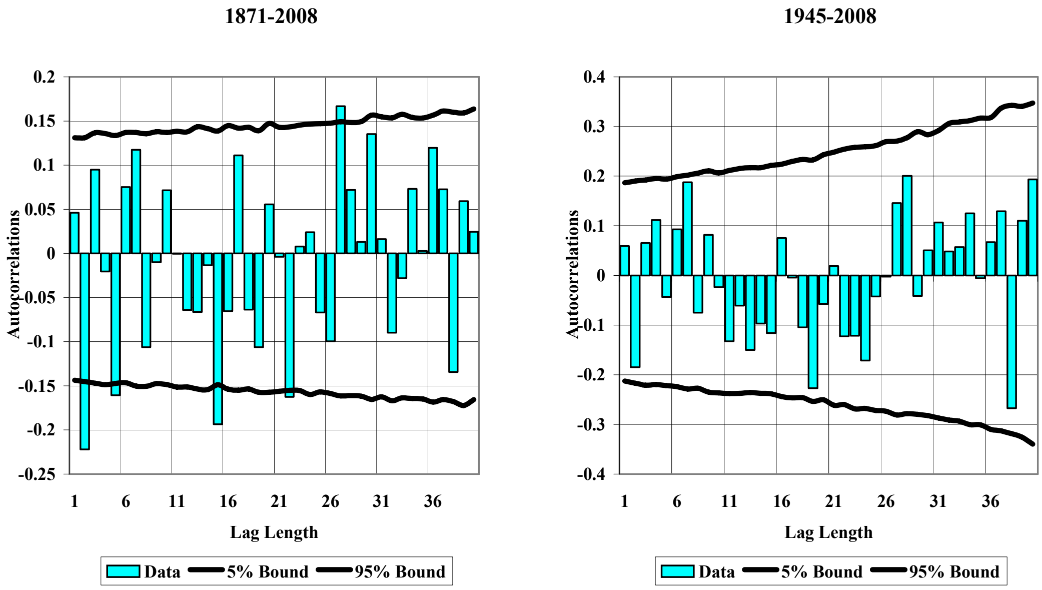
2 Univariate features of returns

We first briefly summarise the observable features of returns that we shall draw on in the rest of the paper.

2.1 Returns are near-white noise

In Figure 1 we show the autocorrelation function for real annual stock returns in the United States over two samples, 1871-2008 and 1945-2008. The first spans the full available dataset on a reasonably consistent basis for a broad based US stock market measure (the Cowles (1938) industrial index from 1871-1925, and the S&P 500 thereafter).[7] The shorter sample allows for the possibility that return properties may have changed in the postwar era (consistent with the claims discussed below by Kim et al, 1991).

Figure 1. The Autocorrelation Function of Real US Stock Returns



We also show bootstrapped 5% and 95% bounds when returns are resampled with replacement to destroy any possible temporal dependence. In the full sample this illustrates that, while autocorrelations are generally very small in absolute terms, a subset are individually marginally significant against the null of white noise; the same applies for the standard Ljung-Box Q portmanteau test at some horizons. However even these apparent rejections of the white noise null are subject to a well-known data mining critique, if we focus only on a relatively small number of rejections. In Table 1 we show simulated p-values for the largest absolute autocorrelation over a different range of lag lengths up to some maximum, over the two different samples, and for the most significant rejection on the Ljung-Box test, both under the null of white noise. This shows that even white noise processes will appear to have significant autocorrelations at *some* lag length with quite high probability; with the probability increasing with the total number of autocorrelations considered. Thus on the basis of standard analysis of autocorrelations, returns

[7] We have also extended this series backwards to 1801 using Siegel's (1998) dataset, as in Pastor & Stambaugh (2008, 2009). Results are very similar.

appear to be very close to white noise even over the full sample. In the postwar sample, there is even less reason to reject the white noise null.

Of course, as is equally well-known, tests of the white noise null will have very low power against an alternative that the true process is close to, but is not quite white noise. But for our purposes the distinction is not of any great importance. We shall show below that even if we allow returns to deviate from white noise by estimating ARMA(1,1) representations (which appears to be quite adequate to remove any serial correlation structure in the resulting residuals) the resulting representations have very low R^2s.

Thus the first key (and probably uncontentious) feature that informs our analysis is that returns are, at best, barely predictable in terms of their own past.

Table 1. Bootstrapped p-values under the white noise null

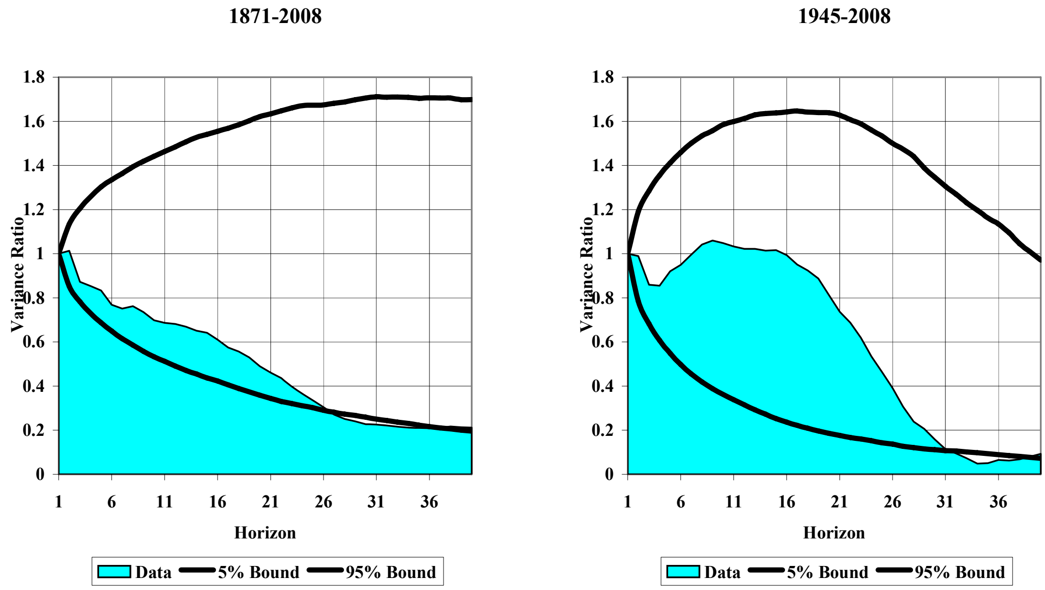
	max(Absolute Autocorrelation)		min (p-value on Q test)		min (Variance Ratio)	
$h_{\max}$	1871-2008	1945-2008	1871-2008	1945-2008	1871-2008	1945-2008
10	0.089	0.797	0.095	0.643	0.225	0.648
20	0.196	0.835	0.142	0.762	0.165	0.761
30	0.314	0.967	0.193	0.796	0.032	0.155
40	0.452	0.973	0.262	0.554	0.074	0.030

Notes to Table 1 We simulate the white noise null by resampling with replacement from the empirical distribution of real annual stock returns, in 10,000 repetitions. The first two columns of Table 1 show the bootstrapped probability of a larger value than in the data for the maximum autocorrelation from 1 to $h_{\max}$, under the white noise null. For Columns 3 and 4, we carry out Ljung-Box Q tests of the joint significance of autocorrelations from 1 to h, in the data, and in each replication, then find the minimum nominal p-value over h=1 to $h_{\max}$: the table shows the probability, across all replications, of observing a lower minimum nominal p-value than in the data. In Columns 5 and 6 we show the bootstrapped probability across all replications, of a lower minimum variance ratio than in the data over horizons 1 to $h_{\max}$.

2.2 The variance ratio slopes downwards

In Figure 2 we show the sample variance ratio for real annual stock returns at horizon h, $VR(h) = Var(\sum_{i=1}^{h} r_{t+i}) / (Var(r_t).h)$ for horizons 1 to 40.[8]

Figure 2. The Variance Ratio of Real US Stock Returns



The first panel shows clearly, over the long sample 1871-2008, the pattern identified by Poterba & Summers, (1988). The sample variance ratio declines nearly monotonically as the horizon increases until around $h = 30$, at which point it appears to level out at a value of around 0.2: thus indicating a reduction in volatility for long-horizon returns that is, in economic terms, highly significant, compared to the white noise benchmark. This pattern has been widely used to argue that investment in stock portfolios is relatively less risky at long horizons.[9] We also show simulated 5% and 95% bounds for the sample variance ratio under the bootstrapped white noise null. The observed pattern does not differ much from white noise at short horizons; but appears increasingly different as the horizon lengthens. While the data mining critique again

[8] We do *not* include the small sample adjustment proposed by Cochrane (1988) and others. Given our focus on simulated results, where the variance ratio is calculated in the same way in both data and simulations, any adjustment is unnecessary. Under the white noise null the unadjusted sample variance ratio is biased downwards; however under alternatives where returns are near-white noise such as the ARMA(1,1) we analyse below, we show that the unadjusted sample variance ratio appears to be close to unbiased.

[9] See, for example, Siegel, 1998; Campbell & Viceira, 1999. Note that we are referring here to true unconditional variance, rather than (conditional) "predictive variance", that allows for parameter uncertainty. Pastor & Stambaugh (2008) show that the horizon profile of predictive variance can difffer quite significantly from that of true variance.

argues against placing too much weight on individual horizons, the third and fourth column of Table 1 shows that if we focus on the minimum variance ratio across all horizons up to a given maximum horizon, the longer the horizon, the lower is the probability of observing such a low value under the white noise null.[10]

The second panel of Figure 2 shows that if we calculate the variance ratio only over the postwar period there is no systematic tendency to decline until $h = 20$ - a result consistent with the estimates in Kim et al (1991). However, for longer horizons the decline is quite marked, and, as Table 1 shows, statistically significant against a white noise null, even allowing for the rather limited number of degrees of freedom.[11] We shall show below that there are also clear *in*direct measures of a declining variance ratio that persist into the postwar era.

There is no necessary contradiction between our weak rejection of the white noise null for the autocorrelation function and the stronger results for the variance ratio, since the latter relates to a long weighted average of autocorrelations.[12] In principle "variance compression"[13] can be both quite significant, and consistent with a very limited degree of short-term predictability. This is indeed what appears to be the case in the data.

It should also be stressed that the probability that *both* these features would appear in the data would be *very* small under the white noise null. Figure 3 illustrates for the full sample. We resample 138 observations of the real stock return to simulate the white noise null. Figure

[10]Pastor & Stambaugh (2008) Figure 10 shows an almost identical pattern using a longer sample, starting in 1802, which implies an even stronger rejection of the white noise null.

[11]The downward bias noted in footnote 8, which is quite severe in such a relatively short sample, is very evident in the simulated upper and lower bounds.

[12]From Cochrane's (1988) orginal analysis showed, we have $VR(h) = 1 + 2\sum_{j=1}^{h-1}\left(\frac{h-j}{h}\right) corr(r_t, r_{t-h})$.

[13]This feature is often referred to as mean reversion (following Poterba & Summers, (1988)), but we avoid this term deliberately, first, because this usage is not universal (cf Pastor & Stambaugh, 2008); and second, because it is a somewhat confusing misnomer. Poterba & Summers define mean reversion as "stock prices (or cumulative returns) have a mean-reverting transitory component". Following Beveridge & Nelson (1981) we can write any general ARMA(p, q) univariate representation of returns as

$$r_t = a(L)\varepsilon_t = a(1)\varepsilon_t + a^*(L)(1 - L)\varepsilon_t$$

with the second term defining the mean-reverting transitory component in cumulative returns $= a^*(L)\,\varepsilon_t$. Such a term will be present for *any* stationary univariate representation where returns have some serial correlation structure, but not all such representations will have a downward sloping variance ratio. It is straightforward to show that $a(1) < 1$ is a sufficient condition for the variance ratio to slope downwards. Since $a(1) + a^*(0) = 1$, in this case the transitory component will be positively correlated with returns, whereas for $a(1) > 1$, which implies that the variance ratio slopes upwards, it will be negatively correlated. But in both cases the transitory component will be mean-reverting (cf Kim et al (1991) who refer to the latter case as "mean aversion").

3 is then a scatter plot of the minimum variance ratio, over horizons 1 to 40, against the sample R^2 for an ARMA(1,1) representation of returns, for each replication. The crossing point of the two lines on the chart shows the values observed in the data.

Figure 3. Univariate Predictability and a Declining Variance Ratio: Data versus White Noise

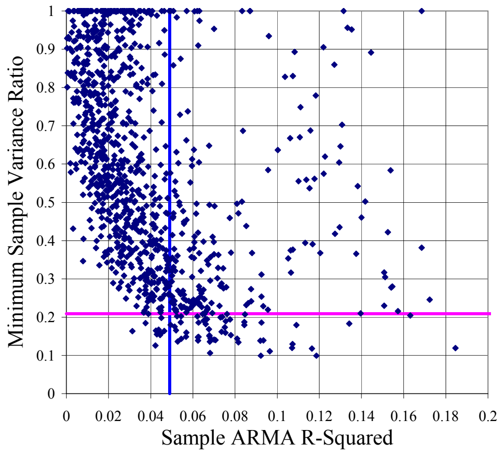


Figure 3 shows that the bulk of replications would have a low ARMA R^2, but with considerable spread: 24% of the sample estimates of the ARMA R^2 would be above the value in the data (very much in line with the evidence on the autocorrelations shown in Table 1). The majority of simulations would generate a minimum variance ratio well above that in the data; the points below the horizontal line correspond to the 7.4% probability given in the bottom row of Table 1. But, most strikingly, samples in which the variance ratio does appear to slope significantly downwards are almost always also samples in which the ARMA model appears to predict distinctly better than in the data: only 1.4% of replications generated combinations in the bottom left quadrant, ie, with both a lower R^2 *and* a lower minimum variance ratio than in the data.[14]

[14] For the postwar sample only 1% of replications lie in this quadrant.

3 The predictive regression framework

3.1 The general system

Consider the system used by Stambaugh (1999) and many others in the analysis of predictive return regressions

$$
\begin{aligned}
r_t &= -\beta_x x_{t-1} + u_t & (1) \\
x_t &= \lambda x_{t-1} + v_t & (2)
\end{aligned}
$$

where the first equation captures the degree of predictability of some variable r_t, typically stock returns or excess returns over some interval, in terms of a predictor variable x_{t-1}, and the second describes the autocorrelation of the predictor variable. We assume $0 \leq \lambda < 1$, so that both r_t and x_t are stationary.[15] We put no restrictions on the innovations u_t and v_t other than that they be (jointly) serially uncorrelated mean zero with finite variance. We assume all data are de-meaned for simplicity, hence neglect constants.

Equation (1) is quite general, since x_{t-1} may in principle be some weighting of a set of variables with predictive power for r_t and the error term may capture a range of nonlinearities. Equation (2) is distinctly more restrictive, but, since Stambaugh (1999) has been widely used in the literature and, again, allowing for exotic errors, can still encompass a wide range of models (including for example two state Markov switching models, Hamilton 1989).[16]

Substituting from (2) into (1) we derive the reduced form process for r_t, which is an ARMA(1,1):[17]

$$r_t = \lambda r_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} = \left(\frac{1 - \theta L}{1 - \lambda L} \right) \varepsilon_t \tag{3}$$

[15] Most of our results generalise to, but are complicated by, $\lambda < 0$; however we regard this as empirically less likely to be of interest.

[16] Apart from differences in notation, our predictive framework is also identical to, eg, Cochrane (2008); Campbell, Lo and Mackinlay (1997), Chapter 7 and Pastor & Stambaugh (2009) (in the the latter context, x_t would be characterised as a "perfect predictor" -ie, one that captures all available information relevant to expected returns).

[17] By letting x be a vector process, with an AR matrix with p distinct roots, we can generalise up to an ARMA(p, p) representation of r_t. We note below that some of our results below still apply in this much more general case, but the ARMA$(1, 1)$ representation considered here has the dual advantage that it can be related much more readily to the standard predictive regression framework, and yields relatively simple analytical results.

where L is the lag operator, such that $Lx_t = x_{t-1}$; ε_t is a serially uncorrelated innovation; and as long as u_t and v_t are less than perfectly correlated, we can choose the "fundamental" solution for the MA parameter that has $\theta \in (-1, 1)$, so the representation is strictly invertible in terms of the history of r_t.[18] If $\theta = \lambda$ the AR and MA components cancel, and r_t will be white noise.

We first note that in the ARMA representation the properties of r_t are entirely determined, up to a scaling factor, by the pair (λ, θ). The properties of the underlying predictive system (1) and (2) can in turn be characterised by the three unit-free parameters (λ, ρ, R_x^2) where $\rho = \sigma_{uv}/(\sigma_u \sigma_v)$ is the Stambaugh Correlation, and $R_x^2 = 1 - \sigma_u^2/\sigma_r^2$ is the R^2 in the predictive regression. We shall refer to the triplet (λ, ρ, R_x^2) as the "predictive space".

The autoregressive coefficient of the predictor variable translates directly to the AR coefficient of the reduced form (3). For the case of the MA parameter θ things are more complicated. In Appendix A we show that, subject to an innocuous normalisation on the sign of β_x, θ depends on all three parameters that define the predictive space,

$$\theta = \theta\left(\lambda, \rho, R_x^2\right) \tag{4}$$

We shall show that the two univariate properties summarised in Section 2 mean that the predictive space can be quite tightly constrained. In so doing it will be helpful to make reference to two important benchmark cases that we shall show determine the nature of these restrictions.

3.2 "Pseudo Predictor" Representations

In this section we define two limiting cases of the predictive system in (1) and (2), both of which can be derived directly from the properties of the ARMA representation. We shall then go on to show, in Section 4, that these limiting cases provide benchmarks that allow us to set limits on the "predictive space" that contains *all* possible predictive systems of the form in (1) and (2).

[18] See Appendix A.

3.2.1 The fundamental pseudo predictor

We can rewrite the ARMA(1,1) representation in (3) as a predictive system of the same general form as (1) and (2):[19]

$$r_t = -\beta_f x_{t-1}^f + \varepsilon_t \tag{5}$$

$$x_t^f = \lambda x_{t-1}^f + \varepsilon_t \tag{6}$$

where $\beta_f = \theta - \lambda$, and we refer to the predictor variable, x_t^f, as the "fundamental pseudo predictor". It has the same AR(1) form as the true predictor variable, but with innovations identical to those in the predictive regression, hence the Stambaugh Correlation is precisely unity. It will generate identical predictions to the fundamental ARMA representation in (3), and will therefore have the same predictive R^2, which we show in Appendix B is given by

$$R_f^2(\lambda, \theta) \equiv 1 - \frac{\sigma_\varepsilon^2}{\sigma_r^2} = \frac{(\theta - \lambda)^2}{1 - \lambda^2 + (\theta - \lambda)^2} \tag{7}$$

Note that, using (6) and (3) we can also write

$$x_t^f = \frac{r_t}{1 - \theta L} = \sum_{i=0}^{\infty} \theta^i r_{t-i} \tag{8}$$

so the fundamental pseudo predictor is simply an exponentially weighted moving average of r_t.[20]

[19] See Appendix A.

[20] An alternative interpretation of x_t^f is, up to a scaling factor, as the optimal estimate of the true predictor x_t given the information set $\{r_i\}_{i=-\infty}^{t}$. Cochrane (2008b) refers to this as the "Observable State Space Representation"; it is also a special case of Hansen & Sargent's (2005) "Innovations Representation".

3.2.2 The non-fundamental pseudo predictor

For every fundamental ARMA(1,1) representation with $\theta \neq 0$[21] there is an associated "non-fundamental" representation, given by

$$r_t = \left(\frac{1-\theta^{-1}L}{1-\lambda L}\right)\eta_t \tag{9}$$

with $\sigma_\eta^2 = \theta^2\sigma_\varepsilon^2$. This representation generates an identical autocorrelation structure for returns to that of the fundamental representation, but, as is well known (see for example Hamilton, 1994, pp 66-67), the non-fundamental innovations, η_t cannot be recovered from the history of r_t, hence the non-fundamental representation does not represent a viable predictive model. As a result, with a few exceptions (Lippi & Reichlin 1994; Hansen & Sargent, 2005; Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson, 2007) non-fundamental representations have received relatively little attention.

To see why η_t cannot be recovered from the data, note that if we attempt to solve (9) for η_t we have

$$\eta_t = \left(\frac{1-\lambda L}{1-\theta^{-1}L}\right) r_t = \sum_{i=0}^{\infty}\theta^{-i}\left[r_{t-i} - \lambda r_{t-i-1}\right]$$

given that $\left|\theta^{-1}\right| > 1$ the sum does not converge, hence the representation in (9) is not invertible in terms of the history of r_t. However, if (strictly hypothetically) we had data on current and *future* values of r_t, we could write

$$\eta_t = \left(\frac{1-\lambda L}{1-\theta^{-1}L}\right) r_t = -\theta F\left(\frac{1-\lambda L}{1-\theta F}\right) r_t = -\theta\sum_{i=1}^{\infty}\theta^{i}\left[r_{t+i} - \lambda r_{t+i-1}\right] \tag{10}$$

where F is the forward shift operator, such that $Fx_t = L^{-1}x_t = x_{t+1}$, and in this case the sum does converge. Thus the non-fundamental ARMA representation *does* have an invertible representation, but only in terms of current and future values of r_t, making it valueless as a predictive model.

Of course, if we already knew the entire future of r_t, we would not need a predictive model

[21] In Appendix D we discuss the special case $\theta = 0$, for which the representation in (9) is undefined, and which we need to deal with separately, but this does not affect any of our results.

at all, therefore there would be no point in constructing a series for η_t. But the reverse is not the case. In general, even if we did have data on η_t (perhaps by some divine dispensation) this would *not* reveal the entire future of r_t, but rather a particular linear combination of future values. Thus while (as we show below) the non-fundamental representation would, if we had the history of η_t, predict better than the fundamental representation, it would not predict perfectly.

While it may seem somewhat peculiar to take an interest in a predictive model that is so manifestly non-viable, it turns out that it provides us with an extremely useful benchmark. And it does so because, while we will never be able to observe η_t in practice, we do know the predictive *properties* of the non-fundamental representation, even if we cannot actually use it to predict, since these can be inferred directly from the properties of the fundamental representation.[22]

As noted above, the equivalence of the two representations must imply that, for $\theta \neq \pm 1$, η_t has lower variance than ε_t, the fundamental innovation (since $\sigma_\eta^2 = \theta^2 \sigma_\varepsilon^2$), hence, if we did have data on η_t, the non-fundamental representation would predict strictly better than the fundamental representation. Its (strictly notional) predictive R^2 can be derived by replacing θ with θ^{-1} in (7), giving

$$R_n^2(\lambda, \theta) \equiv 1 - \frac{\theta^2 \sigma_\varepsilon^2}{\sigma_r^2} = \frac{(1 - \theta\lambda)^2}{1 - \lambda^2 + (\theta - \lambda)^2} > R_f^2(\lambda, \theta)\,; \text{ for } \theta \in (-1.1) \tag{11}$$

As for the fundamental ARMA representation, we can again reverse-engineer a representation of the same general form as (1) and (2), and write

$$\begin{aligned} r_t &= -\beta_n x_{t-1}^n + \eta_t & (12) \\ x_t^n &= \lambda x_{t-1}^n + \eta_t & (13) \end{aligned}$$

[22] In Lippi & Reichlin's (1994) terminology the non-fundamental representation in (9) is a "basic" non-fundamental representation, in that it is of the same order as the observable fundamental representation. There is in principle an infinity of "non-basic" non-fundamental representations of arbitrary higher order, since any white noise innovation can always be given a non-fundamental representation: ie, we could write $\eta_t = \left(1 - \phi^{-1}L\right)(1 - \phi L)^{-1} \omega_t$, with $\sigma_\omega^2 = \phi^2 \sigma_\eta^2$, and in principle then find a non-fundamental representation of ω_t, and so on ad infinitum. But nothing in the data tells us anything about ϕ, and hence about ω_t, hence we can infer nothing from the data about the properties of such non-basic representations.

with $\beta_n = \theta^{-1} - \lambda$, where x_t^n, the "non-fundamental pseudo predictor" has the same innovations, and the same predictive R^2 for r_t, as the non-fundamental ARMA representation, and again has a Stambaugh Correlation of precisely unity.

3.3 The variance ratio in the ARMA(1,1) reduced form

We have already referred, in our discussion of univariate properties in the data, to the variance ratio at horizon h, as originally defined for the general case by Cochrane (1988) as

$$VR(h) = \frac{1}{h}\frac{Var(\sum_{i=1}^{h} r_{t+i})}{Var(r_t)} \tag{14}$$

It is straightforward to show[23] that, for the ARMA(1,1) process (3), $VR(h)$ is monotonic in h and that

$$VR(h)\begin{cases} < 1 \Leftrightarrow \theta > \lambda \\ > 1 \Leftrightarrow \theta < \lambda \end{cases} ;\forall\ h > 1 \tag{15}$$

We shall also make use of the limiting value of the variance ratio, which in the ARMA(1,1) can be expressed as

$$V = \lim_{h\to\infty} VR(h) = (1 - R_f^2)\left(\frac{1-\theta}{1-\lambda}\right)^2 \tag{16}$$

where, given the monotonicity of the $VR(h)$ in h, we also have $V < 1 \Leftrightarrow VR(h) < 1\ \forall h > 1$.

4 The Predictive Space for Stock Returns

4.1 Bounds on the ARMA(1,1) coefficients, θ and λ

In Section 2 we discussed two univariate properties of real stock returns in the data: first that they were near-white noise (hence barely predictable in the short term); and second that there appears to be quite strong evidence of "variance compression" (ie, a declining variance ratio). It is straightforward to show that these two features of the data constrain quite tightly the possible values of λ and θ in the ARMA representation. We shall subsequently see that this in

[23]See Appendix C.

turn will place quite significant restrictions on the predictive space.

Using the ARMA(1,1) framework outlined above we show, in Figure 4, contours in (θ, λ) space of equal R_f^2 and of equal V (the limiting value of the variance ratio).[24] The top panel of Figure 4 shows contours for $V = 0.4$ and $R_f^2 = 0.05$. The shaded area then gives the admissible set of (θ, λ) that generate values of V no greater than 0.4 and R_f^2 of no more than 0.05. Lower values of V push the V-contour up and to the left, while lower values of R_f^2 move the R_f^2-contours towards the 45 degree line, thus reducing the admissible (θ, λ) area. The second panel of Figure 4 illustrates: the shaded area is now the (θ, λ) combinations consistent with V no greater than 0.2, and R_f^2 below 0.025: the permissible space for both ARMA parameters now becomes very tightly constrained: λ must be quite close to unity, and θ must be even closer.[25]

We showed in Section 3 that the ARMA representation inherits the AR parameter λ from the true predictor. Figure 4 shows that the requirement that λ be large arises naturally from the univariate properties of returns. Virtually all observable predictors of stock prices (most notably valuation ratios like the price-dividend ratio or the price-earnings ratio) have this characteristic.[26] But the analysis illustrated by Figure 4 shows that, for sufficiently strong variance compression, and sufficiently weak short-term univariate predictability, the same must apply for *any* logically possible predictor.

We now go on to show that these univariate features can put significant restrictions on the predictive space of the underlying model that generates them.

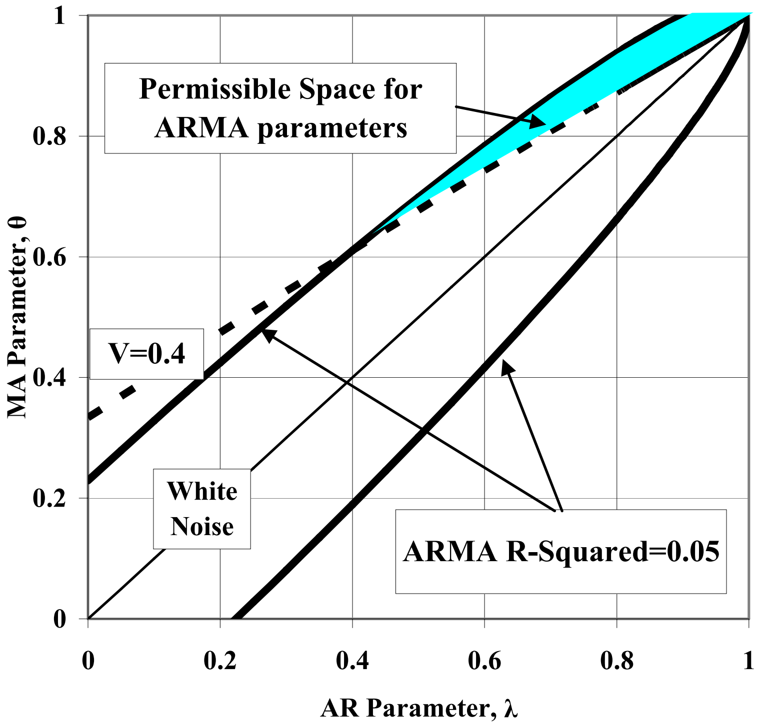
[24]For given values of R_f^2 and V, we solve (7) and (16) for θ in terms of λ. The former gives two solutions for θ, symmetric around the 45 degree line.

[25]The numbers used for R_f^2 and V in Figure 4 are illustrative, but are quite consistent with the evidence illustrated in Figures 1 to 3. Sample estimates of R_f^2 are, if λ is high, subject to severe Stambaugh (1999) bias. Simulation evidence shows that even in a sample as long as the 1871-2008 period discussed in Section 2 a true R_f^2 of 0.025 would result in a mean sample estimate at least twice as large, thus consistent with what we observe in the data. For large λ and θ we can also have $VR(h)$ well above the limiting value V even for horizons as long as those shown in Figure 2.

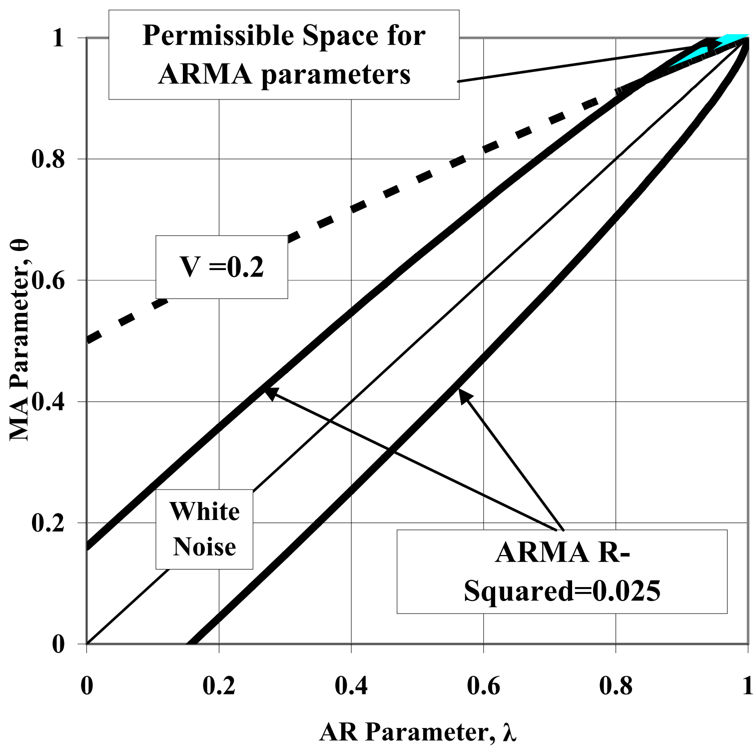
[26]On annual data, the AR(1) coefficients for the dividend yield and the cyclically adjusted P/E multiple, for example, are 0.92, 0.93.

Figure 4
The Permissible Space for ARMA Parameters for Stock Returns

Limiting variance ratio, V≤0.4; ARMA R-Squared ≤ 0.05



Limiting variance ratio, V≤0.2; ARMA R-Squared ≤ 0.025



4.2 Bounds for the one-period-ahead predictive R^2

Proposition 1 *For a fundamental ARMA(1,1) representation of returns which is a reduced form of a predictive regression (1) and a predictor autoregression (2) the one-period-ahead R^2 of the predictive regression, R_x^2, satisfies*

$$R_f^2(\lambda, \theta) \leq R_x^2 \leq R_n^2(\lambda, \theta) \tag{17}$$

where R_f^2 and R_n^2 are as defined in (7) and (11).

Proof. See Appendix D. ∎

The lower bound for R_x^2 is the predictive R^2 of the fundamental ARMA representation, or, equivalently, of the "fundamental pseudo predictor" defined in Section 3.2.1. As such it is quite easy to interpret. As long as the true predictor provides some predictive information for r_t beyond that contained in the history of r_t itself (ie, if $\beta_x \neq 0,\ \rho \in (-1, 1)$) it must have a strictly higher predictive R^2; only if $\beta_x = 0$, or in the special case of the fundamental pseudo predictor, is the lower bound attained.[27]

The upper bound for R_x^2 is the predictive R^2 of the *non*-fundamental ARMA representation, or equivalently of its associated pseudo predictor, defined in Section 3.2.2. The intuition for this result can be related to our earlier discussion of the properties of the non-fundamental representation. We showed in Section 3.2.2 that the non-fundamental innovation η_t can be expressed, in (10), as a linear combination of current and future returns: so we know already that it must have some predictive power beyond that already in the history of returns. But the result in Proposition 1 is distinctly stronger: it shows that the non-fundamental pseudo predictor in period t is the best *possible* predictor of r_{t+1} consistent with its observable univariate properties.[28]

[27] We noted in Section 3.2.1, the alternative interpretation of the fundamental pseudo predictor, x_t^f as the optimal estimate of the true predictor x_t given the information set $\{r_i\}_{i=-\infty}^{t}$. If we have data on x_t, rather than its estimate, we must be able to predict better, except in the special case that $x_t = x_t^f$.

[28] Given the ARMA(1,1) property of returns we know that the true predictor must be an AR(1). For reasons discussed in footnote 22 there may in principle be multiple predictors, of arbitrary ARMA order, that predict arbitrarily better. But the data tell us absolutely nothing about such predictors. Equivalently, the non-fundamental ARMA representations associated with such predictors are, in Lippi & Reichlin's (1994) terminology, "non-basic".

We shall show that when the ARMA parameters, θ and λ, lie within the permissible range illustrated in Figure 4, for a given degree of variance compression and low short-term predictibability then the allowable range of R_x^2 given by Proposition 1 can become quite small.

We have already noted that if $\theta = \lambda$ returns are white noise. This arises trivially if $\beta_x = 0$. But there is also a more interesting special case:

Remark *(Predictable White Noise) If $R_x^2 > 0$ but $\theta\left(\lambda, \rho, R_x^2\right) = \lambda$, the inequality in (17) reduces to*

$$0 < R_x^2 \leq 1 - \lambda^2$$

We discuss the properties of this special case in more detail in Section 4.5 below.

4.3 Bounds for ρ for predictor variables

Our focus thus far has been on just two of the elements in the predictive space, namely λ and R_x^2. But a further important feature of Proposition 1 is that both the upper and lower bounds arise in limiting cases of the predictive system (in (5) and (6), and in (12) and (13)) for which the Stambaugh Correlation, ρ, is precisely unity. We now examine intermediate cases in which the innovations are not perfectly correlated.

We have from (4) that the ARMA coefficients θ and λ are linked to the predictive R_x^2 and the Stambaugh Correlation between the innovations, ρ, by $\theta = \theta\left(\lambda, \rho, R_x^2\right)$. Thus for a given (θ, λ) pair there is a contour of possible values of (R_x^2, ρ) consistent with the ARMA representation. Again it turns out that the univariate properties of r_t impose limits on the possible values of ρ.

Proposition 2 *Consider a fundamental ARMA$(1,1)$ univariate representation (3) which is a reduced form of a predictive regression (1) and a predictor autoregression (2). For $0 < \lambda < \theta$ (ie the variance ratio slopes downwards and the predictor has positive persistence) the Stambaugh correlation ρ satisfies*

$$|\rho| \geq \rho_{\min}\left(\lambda, \theta\right) > 0$$

Proof. See Appendix E. ■

Figure 5 illustrates the link betweeen Propositions 1 and 2. We graph the contours in (R_x^2, ρ) space for a range of different representations of the return process. To simplify the presentation we constrain ρ to be non-negative (which we can always ensure is the case by an appropriate rescaling of the data for x_t).[29]

Figure 5. The Predictive Space for Stock Returns: R_x^2 and ρ

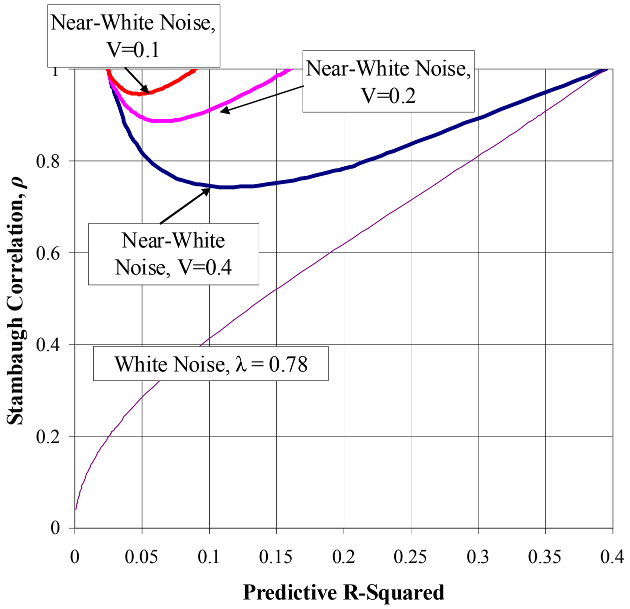
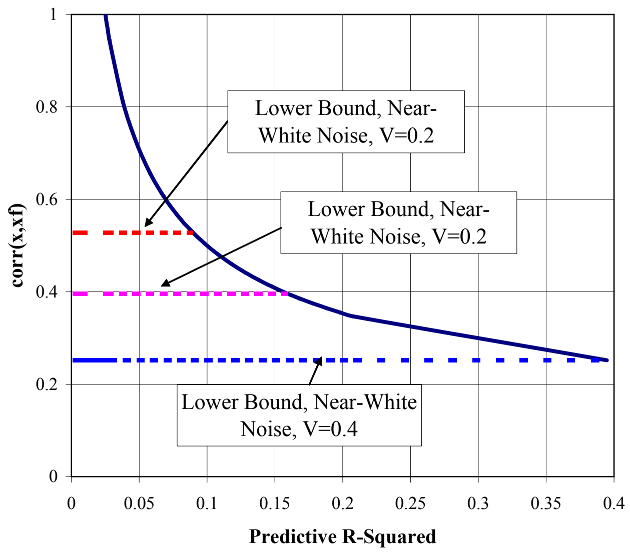


Figure 6. The correlation between the true predictor and the fundamental pseudo predictor



As a benchmark for comparison, the lowest contour line shows combinations of the two parameters consistent with the special case of predictable white noise noted in the previous

[29] As noted in the introduction, ρ will be positive if x_t is expressed as a log ratio of price to fundamental, and hence has innovations of the opposite sign to those to expected returns.

section, when the predictor is quite strongly persistent ($\theta = \lambda = 0.78$). The better the predictive model, the higher the associated Stambaugh Correlation, ρ, must be. While ρ can take any value in $[0, 1]$, the lower bound for ρ will only be attained with $R_x^2 = 0$. Thus even in the white noise case any *useful* predictor with positive λ must also have non-zero ρ.

The remaining contour lines represent a range of near-white noise processes, all with the same univariate R^2 ($R_f^2 = 0.025$) but with progressively stronger degrees of variance compression (ie, lower values of V). For a given degree of short-term predictability, this correponds to a progressive reduction in the upper bound for R_x^2. Since $\rho = 1$ at both upper and lower bounds, the range of possible values of ρ is progressively reduced, hence $\rho_{\min}$ in Proposition 2 becomes progressively closer to unity.This feature of our results sheds light on a significant feature of the the empirical literature on predictive regressions. In most of this literature a high value of the Stambaugh Correlation is usually treated simply as a nuisance that complicates inference. Our results show that when returns have declining variance ratios (or even if they are purely white noise) it is an intrinsic feature of the true predictor of returns.[30]

4.4 How different are predictor variables from the history of returns?

We have shown in the previous section that, given observable univariate properties of returns, the Stambaugh Correlation is likely to be close to unity in absolute value (ie, innovations to the predictor variable will be strongly correlated with innovations in the predictive regression). We also know that, by construction, the fundamental pseudo predictor, which from (8) is simply a weighted average of past returns, has a Stambaugh Correlation of precisely unity. It might therefore seem that any predictor must resemble the pseudo predictor quite closely. In fact, while this may be the case for certain univariate processes, the correlation between the true predictor and the fundamental pseudo predictor can in principle cover a distinctly wider range than the Stambaugh Correlation, as the following proposition shows:

[30] Compare the related arguments of Cochrane, 2008 and Pastor & Stambaugh, 2009, discussed in footnote 6. To complicate matters, in Robertson & Wright, 2009 we show that it also an endemic feature of predictors that are actually redundant once we correctly condition on the history of returns, thus making it hard to distinguish between true and redundant predictors, especially if returns exhibit significant variance compression.

Proposition 3 *If x_t is the true predictor in the predictive regression (1), with predictive R-squared R_x^2, and x_t^f is the fundamental pseudo predictor, which, from (8) can be constructed from the history of returns, then*

$$corr(x_t, x_t^f)^2 = \frac{R_f^2}{R_x^2} \geq \frac{R_f^2}{R_n^2} = \frac{1}{\theta^2}\left(\frac{\theta - \lambda}{\theta^{-1} - \lambda}\right)^2$$

where $R_f^2(\lambda, \theta)$ and $R_n^2(\lambda, \theta)$ are the upper and lower bounds given in Proposition 1.

Proof. See Appendix F. ■

By inspection of the relationship in Proposition 3, it is evident that in the limiting case of white noise returns ($\theta = \lambda$) the correlation is precisely zero, since $R_f^2 = 0$. Indeed it is of the essence of a white noise process that it its own history is entirely uninformative about its own future values, and hence it must be uninformative about any predictor of its future values.

For near-white noise processes the correlation is non-zero, and the proposition shows that the better the true predictor predicts, the less similar it will be to the fundamental pseudo predictor. But the upper bound on R_x^2 given in Proposition 1 implies a *lower* bound on the correlation in Proposition 3: hence the narrower is the range of possible values of R_x^2, the more similar the true predictor must be to the fundamental pseudo predictor. The lower bound in Proposition 3 is determined by the relative predictive power of the fundamental vs non-fundamental representations.

Figure 6 (below Figure 5) illustrates, for the three near-white noise processes already illustrated in Figure 5. Since all three have the same value of R_f^2, the univariate R^2, the relationship between $\text{corr}\left(x_t, x_t^f\right)$ and R_x^2 given in Proposition 3 is identical for all three processes. If the true predictor predicts barely any better than the univariate representation, it will very closely resemble it, but the better it predicts the weaker this resemblance will be. The only impact of greater variance compression (a lower value of V) will be that, since this reduces the upper bound for R_x^2, it must *increase* the lower bound for $\text{corr}\left(x_t, x_t^f\right)$ (the lower bounds for each of the three processes are shown as dotted lines).[31]

[31] Figure 6 is placed directly below Figure 5 to illustrate that, for each process in Figure 6, the lower bound corresponds to the point in Figure 5 where the Stambaugh correlation for that process hits unity.

However, Figure 6 illustrates that even when there is very significant variance compression (as θ approaches unity) there is still scope for the true predictor to look quite *dis*similar to the fundamental pseudo predictor. This suggests a simple pre-test when looking for predictor variables for stock returns: we should seek those that do *not* simply look like the history of returns.

4.5 A special case: predictable white noise

We have already noted, in our discussion of Proposition 1, that the special case in which returns are entirely unpredictable from their own past does not rule out predictability from some other predictor variable. This case is worth considering not just as a benchmark for comparison, but also because it is an implication (whether implicit or explicit) of a range of revisionist investigations of return predictability. Some of these (eg, Goyal & Welch, 2003) have concluded that there is simply no return predictability at all of any kind (ie, $\beta_x = 0$) ; others (eg Kim et al, 1991) have concluded that the true variance ratio does not differ significantly from unity at any horizon, which must imply directly that there may be little or no univariate predictability (ie, the more general white noise case $\theta = \lambda$). Even defenders of return predictability such as Campbell & Viceira (2002), Cochrane (2005, Chapter 20) have acknowledged the possibility that there may be no univariate predictability.

Of course if returns are white noise, we have no way of inferring anything directly from the history of returns about the values of the ARMA parameters, except that they must be equal. But this still tells us something about the predictive space: that it depends on a single parameter, λ, the persistence of the true predictor, since the white noise property means that the predictive space must always satisfy $\theta\left(\lambda, \rho, R_x^2\right) = \lambda$. We noted in Section 4.1 that most observable predictors of stock returns are strongly persistent. In Table 2 we show that the maximum possible predictive R^2, as given by Proposition 1, declines as λ increases. For strongly persistent predictors the scope for return predictability, from *any* possible predictor of a white noise return process, is therefore quite limited.

Table 2. The predictive space if stock returns are white noise

λ	0.0	0.25	0.5	0.6	0.7	0.8	0.9	0.95
$\max\left(R_x^2\right) = R_n^2\left(\lambda,\lambda\right)$	1	0.94	0.75	0.64	0.51	0.36	0.19	0.10
$\rho \vert R_x^2 = 0.05; \beta_x > 0$	0	0.06	0.13	0.17	0.22	0.31	0.47	0.70

Notes to Table 2 In line 1 of the Table we show the maximum predictive R^2 for a white noise process, using Proposition 1, which implies $\max\left(R_x^2\right) = 1 - \lambda^2$. Line 2 shows the required value of the Stambaugh correlation ρ, for a given value of R_x^2 using equation (39) setting $\theta = \lambda$. We constrain ρ to lie in $[0,1]$ by normalisation of the data for x_t such that $\beta_x > 0$.

At the other extreme, Table 2 also highlights a further special case of a white noise predictor of white noise returns (ie, $\lambda = \theta = 0$). This is the sole case for which the inequality in Proposition 1 is devoid of content, since it reduces to the condition that $R_x^2 \in [0,1]$. In this case the predictor has (trivially) the same ARMA order as returns (ie, they are both ARMA$(0,0)$), which therefore nests the case: $x_t = r_{t+1} \Rightarrow R_x^2 = 1$. Thus in this case the absolute upper bound for R_x^2 can be attained, at least in logic, if not (in the absence of divine dispensation or time travel) in practice.[32]

As noted in our discussion of Proposition 2, the predictable white noise case means that, subject to our normalisation of x_t, the Stambaugh Correlation, ρ, can in principle live anywhere in $[0,1]$; but a *useful* predictor with $\lambda, \beta_x > 0$ must have a positive Stambaugh Correlation, and the better it predicts the higher ρ must be (since the limiting case of the best possible predictor is the non-fundamental pseudo predictor defined in Section 3.2.2, with $\rho = 1$). Furthermore, for any given value of R_x^2, ρ is also increasing in λ, because a higher value of λ brings down the upper bound at which ρ equals unity. The bottom row of Table 2 illustrates this relationship.

This necessary link between ρ, R_x^2 and λ in the case of predictable white noise returns casts another interesting light on the predictability literature. As noted above, valuation ratios such as the price-dividends and price-earnings ratios have frequently been proposed as predictors of returns. In Robertson & Wright (2009) we show that a range of such predictors all have AR(1) parameters in the neighbourhood of 0.9; with estimated Stambaugh Correlations also around this value, or in some cases, even closer to unity. In contrast the bottom row of Table 2 shows

[32] Note that the key condition here is actually $\theta = 0$, which, from (37), always implies $\max\left(R_x^2\right) = R_n^2 = 1$.

that for white noise returns, and $R_x^2 = 0.05$ (a figure not out of line with those found in the return predictability literature) the required value of ρ for $\lambda = 0.9$ is very much lower than this. An immediate conclusion that follows is that it would not be possible to claim simultaneously that any one of these predictors is the true predictor, *and* that returns are white noise. We shall see in the next section that higher values of ρ *are* more consistent with a predictor of a return process with a declining variance ratio, but in that case the univariate predictability that necessarily follows from this provides an alternative benchmark against which to compare such predictors. In Robertson & Wright (2009) we conclude that none of these commonly used predictors can be distinguished in the data from the pseudo predictor consistent with this univariate predictability.

5 The predictive space for real US stock returns 1871-2008: some empirical estimates

In estimating the limits to the predictive space consistent with the observed history of returns examined at the start of the paper, we should note at the outset that, given the near-white noise properties of returns, no method of estimation can be expected to yield well-determined results. Nor do we wish to pin ourselves down to any assumption that the univariate representation has been stable, and of the restrictive ARMA(1,1) form, over the entire sample of returns.[33] Our estimates in this section are thus largely illustrative.

Even in the absence of any empirical estimates, it should be noted that, simply by allowing for the possibility that returns may be near-white noise with a declining variance ratio, it follows straightforwardly that, for any given degree of predictor persistence, the predictive space must contract relative to the white noise case. Variance compression requires $\theta > \lambda$. This raises R_f^2, the fit of the fundamental ARMA representation, above zero, but at the same time *decreases*

[33] It is quite possible that there may have been structural shifts in the ARMA parameters as well as both the volatility and unconditional mean of real returns. However, there is a non-trivial caveat: the *extent* of any shifts in the mean return cannot have been too large. If such shifts are treated as stochastic, in, for example, a Markov Switching Model, then, following Hamilton (1989), the reduced form process has a higher order ARMA representation. For sufficiently strong shifts the declining variance ratio observed in the data would be ruled out.

the maximum possible predictive R^2 (that of the non-fundamental representation, $R^2_{n.}$) thus contracting the space that R^2_x can feasibly inhabit. At the same time, from Propositions 2 and 3, increasing variance compression raises towards unity the lower bounds on both the absolute Stambaugh Correlation and the correlation between the predictor and the fundamental pseudo predictor. Thus on the basis of a priori reasoning alone we know that greater is the degree of variance compression, the more the predictive space must contract.

Our starting point is simply to estimate the ARMA representation. There are obvious caveats: the near-white noise property means that the AR and MA components are very close to cancellation, and thus, as is well known, both λ and θ are likely to be poorly estimated, and subject to significant small-sample (essentially Stambaugh, 1999) bias. There is however an important cross-check on our results, in the spirit of Cochrane, 1988. We showed in Section 3.3 that in the ARMA(1,1) there is a direct correspondence between the sign of $\theta - \lambda$ and the slope of the variance ratio. It is also straightforward to show[34] that, if $\theta > \lambda$, the *rate* at which the variance ratio slopes downwards is determined solely by the magnitude of $V(\lambda, \theta)$ (the limiting variance ratio) and λ. In principle direct measurement of the variance ratio could, for some processes, yield very different answers from that implied by ARMA estimates;[35] but in both the long annual sample 1871-2008 and (with caveats) the shorter postwar sample 1945-2008, the results are reassuringly similar.

Figure 7 illustrates. We estimate ARMA(1,1) representations of returns in both samples. In terms of the expectations derived from our analysis thus far the point estimates are certainly in the right ballpark: for the full sample we have $\widehat{\lambda} = 0.860$ and $\widehat{\theta} = 0.977$, and in the postwar sample we have $\widehat{\lambda} = 0.89$ $\widehat{\theta} = 0.95$, thus in both samples the point estimates are consistent with variance compression,[36] but they are somewhat closer together in the postwar sample, and hence returns are somewhat closer to white noise. Figure 7 shows that if we treat the ARMA estimates as equal to their true population values, the results are not in conflict with the evidence from direct measurement of the variance ratio.

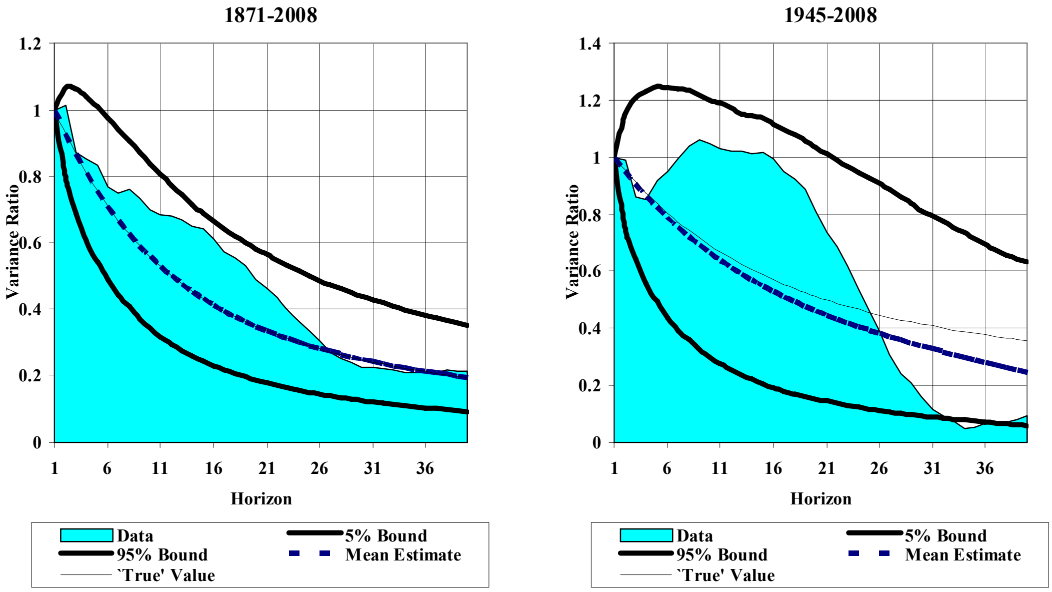
[34] See Appendix C.

[35] See, for example, the comparison between the very different implications of the variance ratio and ARMA representations of GNP growth in Cochrane, 1988.

[36] In Appendix C we show that they are also consistent with a *conditional* variance ratio for the true predictor below unity for all $h > 1$, (*cf* Pastor & Stambaugh, 2008).

Figure 7. The Variance Ratio:

Data vs Implications of ARMA Estimates



Notes to Figure 7 We show the variance ratio in the data, as in Figure 2. The 5% and 95% bounds and mean estimates are simulated in 10,000 replications using the estimated ARMA model as the data generating process. The two panels also show the calculated true value of the variance ratio, as given by (34) in Appendix C, on the same assumptions.otes to Table 1. We simulate the white noise null by resampling with replacement from the empirical distribution of real annual stock returns, in 10,000 repetitions.

In the full sample this consistency is particularly marked. The implied "true" horizon variance ratio matches the sample variance ratio well, particularly at longer investor horizons; and even when the two profiles differ somewhat at shorter horizons, the deviation is well within the range of sampling variation.[37]

In the post-war sample, while the ARMA estimates are quite similar to those estimated over the full sample, they are less consistent with direct measurement of the variance ratio. But the differences are not in general statistically significant. Given the short sample, if the estimated ARMA parameters were truly generating the returns data, the range of sampling variation of the variance ratio would be quite wide, especially at short horizons, hence the lack of any decline for horizons up to around 15 years (as noted by Kim et al, 1991) would not of

[37] Note that if this is the true data generating process the extent of sampling variation in the directly measured variance ratio is very much lower than in the white noise case, additionally there is essentially no small sample bias (in the left-hand panel of Figure 7 the mean estimate and the "true" value are indistinguishable).

itself be particularly significant. Indeed the only statistically significant contrast between the two approaches is at very long horizons, when the observed variance ratio actually breaches the lower 5% bound consistent with the ARMA estimates being the true model. However, given the range of uncertainty in both approaches, it is fairly obvious that it would take only a very limited amount of data mining to find an ARMA representation that was consistent both with the direct ARMA estimates and the evidence of the variance ratio, over both samples. Any such representation would have a high value of λ, and $\theta > \lambda$.

Given the mutual consistency of the two approaches (particularly in the long sample) we have no obvious reason, in terms of the variance ratio evidence at least, to object to the ARMA estimates. In Table 3 we therefore take these estimates at face value, and use them to calculate the implied constraints on the predictive space, using estimates from both long and short samples.[38]

The implied value of R_f^2 the univariate R^2, which, from Proposition 1, provides the lower bound for the predictive R^2 of the true predictor is around 5% in the long sample. This is reasonably consistent with the sample estimate (if anything, given the known impact of Stambaugh Bias, we might expect the sample value to be rather higher). In the postwar sample, as noted above, returns appear closer to white noise, hence the implied true R_f^2 is distinctly closer to zero.[39]

For the upper bound, $\max\left(R_x^2\right) = R_n^2$, the notional R^2 of the non-fundamental ARMA representation, we have, of course, no cross-check from the data, but we can calculate it directly from the estimated values of λ and θ if we treat them as the true parameters. This calculation implies that, in both samples, the best possible predictor of stock returns would have an R^2 of around 10%: thus in terms of predictive R^2 the predictive space is quite narrow. The implied space for the Stambaugh Correlation is even more tightly constrained: the point estimate of $\rho_{\min}$, as defined in Proposition 2, is very close to unity, particularly for full sample estimates.

[38] We do not report standard errors, because in this region of the parameter space they are likely to be highly misleading.

[39] Note that in our theoretical analysis we focussed on the true R^2, which is a function of the true values of λ and θ. In any given finite sample if we use the formula for the population value to calculate $R_f^2\left(\widehat{\lambda}, \widehat{\theta}\right)$ the result of this calculation need not be the same as the sample R-squared calculated from the ARMA estimation; in practice, however, simulation evidence shows that the two figures are usually quite close.

Table 3 Point Estimates of Limits on the Predictive Space for US Stock Returns Implied by ARMA Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	$\widehat{\lambda}$	$\widehat{\theta}$	$\min\left(R_x^2\right)$	$\max\left(R_x^2\right)$	$\rho_{\min}$	$\min\left[corr\left(x_t, x_t^f\right)\right]$
1871-2008	0.860	0.977	0.052	0.095	0.986	0.739
1945-2008	0.891	0.955	0.020	0.106	0.908	0.429

Notes to Table 3 Columns (1) and (2) show the estimated autoregressive (λ) and moving average (θ) parameters in estimated ARMA(1,1) representations of returns over the given samples. Columns (3) and (4) give the implied upper and lower bounds for the predictive R-squared from Proposition 1, given by (column 3) $\min\left(R_x^2\right) = R_f^2\left(\widehat{\lambda}, \widehat{\theta}\right)$ and (column 4) $\max(R_x^2) = R_n^2\left(\widehat{\lambda}, \widehat{\theta}\right)$. Column (5) gives the implied lower bound, $\rho_{\min}\left(\widehat{\lambda}, \widehat{\theta}\right)$ for the Stambaugh Correlation from Proposition 2. Column (6) gives the lower bound for the correlation between the true predictor and the pseudo predictor, as given by Proposition 3, as $\left(R_f^2\left(\widehat{\lambda}, \widehat{\theta}\right) / R_n^2\left(\widehat{\lambda}, \widehat{\theta}\right)\right)^{1/2}$.

In the final column of the table we calculate the implied lower bound for the correlation between the true predictor and the fundamental pseudo predictor. It is noticeable that, despite the apparently very limited predictive space for R_x^2 and ρ, the true predictor can still in principle look reasonably different from the pseudo predictor - particularly so if we use the postwar ARMA estimates.[40] Nonetheless the clear implication of Table 3 is that point estimates consistent with the data suggest only very little space for any predictor to out-predict the univariate representation.

Of course, given the known problems in estimating ARMA representations with near-cancellation of AR and MA roots, the figures in Table 3 should only be treated as illustrative. We certainly would not wish to state categorically that the true predictive space must be as narrow as the ARMA estimates suggest. Given sampling variation, the history of returns is in principle consistent with a range of true data generating processes, some of which have a distinctly less constrained predictive space. However, if we wish to argue that the predictive space *is* less constrained, we show in the next section that this has important implications for

[40] This reflects the fact that, while the difference between R_f^2 and R_n^2 is small, it is the *ratio* that determines the lower bound for the correlation, and in the postwar sample in particular this ratio is quite high, despite the low value of R_n^2 itself.

another aspect of return predictability on which we have not yet touched: namely long-horizon predictability.

6 The predictive space and long-horizon return predictability

We have stressed already that the estimates in Table 3 are only illustrative. Given the range of sampling variation of ARMA estimates we might quite easily derive point estimates of a similar order of magnitude to those in Table 3 for a quite wide range of white and near-white noise processes, albeit subject to the following considerations:

- First, and fairly obviously, the true process cannot be very far from white noise. In terms of Figure 4, the true values of θ and λ must lie within the quite narrow range given by the R^2 contour lines. Hence we would reject any data generating process for returns for which θ was very far from λ.

- Additionally we have strong grounds to reject near-white noise processes with variance *expansion* rather than compression (ie, with $\theta < \lambda$, and hence an upward sloping variance ratio), given the quite strong rejection of the white noise null by sample variance ratio data, discussed in Section 2. Even if the true variance ratio had only a modest upward slope, the probability of observing the low values observed in the data rapidly wouuld be vanishingly small.

- On the other hand we know that the rejection of the strict white noise case is at best only marginal. Hence the data also do not reject values of θ and λ for which the variance ratio only slopes down very modestly (ie, for which V is less than, but quite close to unity)

While these considerations rule out quite a wide range of (λ, θ) combinations, it is evident that the data *do* admit representations that lie roughly between the 45° line and the upper R^2 contour in Figure 4. Since this includes representations in which λ and θ are both close to zero, this means that the true predictive space could be considerably less constrained than the point estimates in Table 3 suggest. For true processes sufficiently close to white noise, and with sufficiently low λ, we showed in Section 4.5, Table 2, that the upper bound for R_x^2 could in principle be close to unity, and the lower bound for ρ could be close to (though not below) zero.

However, while this is a logical possibility, if we *did* wish to argue that the predictive space were significantly wider, we could only do so by simultaneously assuming that another often-assumed characteristic of stock returns is absent, namely long-horizon predictability. Any predictive regression system has an associated profile for the horizon R^2, given by $R_x^2(h)$, which is the R^2 in predicting the average return over h periods from period t, $r_{t,t+h} \equiv h^{-1}\sum_{i=1}^{h} r_{t+h}$. In the literature[41] it is commonly found that the $R_x^2(h)$ profile peaks at long investor horizons. For example, Cochrane (2008a) reports regression estimates that imply a peak at horizons of 20 years or more.[42]

A convenient way to summarise the characteristics of the horizon profile is to define the horizon, h^* at which the horizon R^2 is at its maximum value. In Appendix G we show that the system in (1) and (2) has the convenient property that the horizon R^2 profile for the true predictor is simply a scaling of the univariate horizon profile, $R_f^2(h)$. It follows that we have

$$h^* \equiv \arg\max\left(R_x^2(h)\right) = \arg\max\left(R_f^2(h)\right) = h^*(\lambda, \theta) \tag{18}$$

so that, while the true predictor could in principle have a very different value of $R_x^2(h^*)$ from the univariate representation, the horizon at which this maximum occurs is identical to that for the equivalent univariate R^2 profile, $R_f^2(h)$.[43]

Table 4 illustrates the link between horizon predictability and different (λ, θ) pairs in the true process for returns. The shaded area shows values for which the one-period ahead univariate R^2 is less than 5%, hence any data-consistent representation of returns must lie roughly in this

[41] See for example Campbell and Viceira 2002; Cochrane, 2008; Boudoukh et al, 2008

[42] We use Cochrane's Table 6, p 1561, direct and indirect coefficient estimates for the unweighted sum of returns $\sum_{i=1}^{h} r_{t+i}$; we then calculate the implied R-squared profile using the formulae in Appendix G, which are consistent with Cochrane's regression framework.

[43] The expression in (18) holds for any value of $\theta \neq \lambda$, hence for any return process that is arbitrarily close to white noise. For the case $\theta = \lambda$ we can however still derive h^*, by using the horizon profile of the non-fundamental pseudo predictor, $R_n^2(h)$.

area.

The table shows that for quite a wide range of (λ, θ) pairs close to zero, there are no horizon effects at all (ie h^* is unity). Only as we move north-eastwards, for relatively high values of of both parameters, do horizon effects become more significant. In order to match values of h^* of 20 or above we need to assume either very high values of λ (with strict white noise returns), or, for (somewhat) more modest values of λ, we need to assume $\theta > \lambda$, and hence variance compression (Table 4 shows that this accentuates horizon effects). But we have seen already that is precisely for these parameter values that the predictive space becomes very tightly constrained. Putting it another way, if there *is* strong long-horizon predictability, it must be close to being a univariate phenomenon.

Table 4. The optimal horizon for the horizon R^2 of the true predictor of ARMA(1,1) returns.

		AR Parameter, λ							
		0	**0.2**	**0.4**	**0.6**	**0.8**	**0.9**	**0.95**	**0.98**
MA Parameter, θ	**0.98**	1	5	8	13	22	33	46	**62**
	0.95	1	4	6	9	14	20	**24**	29
	0.9	1	3	5	7	9	**12**	14	15
	0.8	1	2	3	4	**6**	7	7	7
	0.6	1	2	2	**2**	3	3	3	3
	0.4	1	1	**1**	2	2	2	2	2
	0.2	1	**1**	1	1	1	1	1	1
	0	**1**	1	1	1	1	1	1	1

Notes to Table 4 The table shows the value of h^*, as defined in equation (18) for different values of θ and λ in the ARMA(1,1) representation of returns in (3). The figures for h^* shown in the table are calculated numerically using the general formula for the horizon R-squared profile in (42) in Appendix G. The shaded area shows (λ, θ) pairs with $R_f^2 \leq 0.05$, bolded figures are the special case of predictable white noise.

7 Conclusions

This paper shows that the univariate properties of stock returns can be used to infer restrictions on the nature of any true predictor variable for stock returns, because these univariate properties depend in specific ways on the parameters that characterise the predictive system, which we call the predictive space. The argument does not hinge particularly on whether we are able to estimate a univariate ARMA model precisely, because merely knowing that returns are near-white noise and have a downward sloping variance ratio provides sufficient information to restrict strongly the predictive space.

Our results have three strong implications for the return predictability literature. First, if predictor variables are persistent then it may be that the predictive space contracts to such an extent that no predictor variable will predict very much better than a “pseudo predictor”, that is itself simply a weighted average of past returns. Second, the converse also applies, that is if we seek predictor variables that have meaningful predictive ability, they will have to be dissimilar to weighted past returns and will likely have lower persistence than most existing predictor variables. Third, strong horizon effects imply a tightly constrained predictive space, leaving little scope for the true predictor to outperform the univariate representation.

Appendix

A The ARMA(1,1) Reduced Form.

A.1 Normalisations

In what follows we assume that by an appropriate scaling of the data for the true predictor, x_t, we can ensure $\beta_x \geq 0$. To ensure that both fundamental and non-fundamental pseudo predictors satisfy this sign convention, in what follows we also re-define each by

$$\begin{aligned} x_t^f &= \frac{\rho_f \varepsilon_t}{1-\lambda L} \\ \beta_f &= |\theta - \lambda| \\ \rho_f &= 1;\ \theta \geq \lambda; \\ &= -1;\ \theta < \lambda \end{aligned} \tag{19}$$

$$\begin{aligned} x_t^n &= \frac{\rho_n \eta_t}{1-\lambda L} \\ \beta_n &= |\theta^{-1} - \lambda| \\ \rho_n &= sign\left(\theta^{-1} - \lambda\right) \end{aligned} \tag{20}$$

An intuitive rationale for this normalisation is that in our benchmark case of variance compression, $\beta_f > 0$, hence for both x_t^f and x_t^n we have, $u_t = v_t$ in (1) and (2) hence the Stambaugh Correlation ρ is unity. In this benchmark case, given the assumption that β_x is also positive, the true predictor x_t can be interpreted as some "valuation ratio" with the stock price in the numerator, hence of opposite sign to $E_t r_{t+1}$. As a result ρ is also positive, and thus of opposite sign to its equivalent in Pastor & Stambaugh, 2009, 2008.

While we restrict ourselves to this benchmark case for Proposition 2, the sign convention above allows for the logical possibility of variance expansion ($\theta < \lambda$), for which the proofs of the remaining propositions remain valid.

A.2 Derivation of reduced form

Start from

$$\begin{aligned} r_t &= -\beta_x x_{t-1} + u_t \\ x_t &= \lambda x_{t-1} + v_t \end{aligned}$$

where $\begin{pmatrix} u_t \\ v_t \end{pmatrix}$ are jointly serially uncorrelated mean zero with covariance matrix $\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$. Solving gives us the ARMA reduced form for r_t

$$r_t = \lambda r_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

where

$$\varepsilon_t - \theta \varepsilon_{t-1} = u_t - \lambda u_{t-1} - \beta_x v_{t-1}$$

Where θ must satisfy the moment condition

$$\begin{aligned} \frac{-\theta}{1+\theta^2} &= \frac{cov\left(-\beta_x v_{t-1} + (1-\lambda L)u_t, -\beta_x v_{t-2} + (1-\lambda L)u_{t-1}\right)}{var\left(-\beta_x v_{t-1} + (1-\lambda L)u_t\right)} \\ &= \frac{-\left(\lambda \sigma_u^2 + \beta_x \sigma_{vu}\right)}{\beta_x^2 \sigma_v^2 + \sigma_u^2(1+\lambda^2) + 2\lambda \beta_x \sigma_{vu}} = \frac{-\left(\lambda + \beta_x \rho s\right)}{1 + \lambda^2 + \beta_x^2 s^2 + 2\lambda \rho \beta_x s} \end{aligned}$$

where $\rho = \sigma_{vu}/(\sigma_u \sigma_v)$; $s = \sigma_v/\sigma_u$.

Note also that if R_x^2 is the R^2 from the predictive regression then

$$R_x^2 = \frac{\beta_x^2 \sigma_x^2}{\sigma_r^2} = \frac{\beta_x^2 s^2}{\beta_x^2 s^2 + 1 - \lambda^2}$$

implying

$$\beta_x^2 s^2 = F^2 \text{ where } F(R_x^2, \lambda) = \sqrt{\left(1-\lambda^2\right) \frac{R_x^2}{1-R_x^2}} \tag{21}$$

hence the moment condition defining θ can be written as

$$\frac{\theta}{1+\theta^2} = \kappa(\lambda, \rho, R_x^2) \tag{22}$$

$$\text{where } \kappa(\lambda, \rho, R_x^2) = \frac{\lambda + \rho F(R_x^2, \lambda)}{1 + \lambda^2 + F(R_x^2, \lambda)^2 + 2\lambda\rho F(R_x^2, \lambda)} \tag{23}$$

A real solution for $\theta \in (-1, 1)$ requires that $\kappa \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Note first that we have

$$\partial\kappa/\partial\rho = \frac{F\left(1 + F^2 - \lambda^2\right)}{\left(1 + \lambda^2 + F^2 + 2\lambda\rho F\right)^2} > 0, \quad \lambda \in (0, 1), R_x^2 \in (0, 1) \tag{24}$$

Thus, for $\rho \in (-1, 1)$, we have

$$\begin{aligned} \kappa\left(\lambda, \rho, R_x^2\right) &\in \left(\kappa\left(\lambda, -1, R_x^2\right), \kappa\left(\lambda, 1, R_x^2\right)\right) \\ &\in \left(g\left(\lambda - F\right), g\left(\lambda + F\right)\right) \end{aligned}$$

where

$$g(z) = \frac{z}{1 + z^2} \in \left(-\frac{1}{2}, \frac{1}{2}\right) \tag{25}$$

hence we do indeed have $\kappa \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Given this, we know that

$$\frac{\partial\theta}{\partial\kappa} = \frac{1}{2}\left(\frac{1 - \sqrt{1 - 4\kappa^2}}{\kappa^2\sqrt{1 - 4\kappa^2}}\right) \geq 0 \tag{26}$$

which in turn gives us

$$\frac{\partial\theta}{\partial\rho} = \frac{\partial\theta}{\partial\kappa}\frac{\partial\kappa}{\partial\rho} \geq 0 \tag{27}$$

which we shall exploit later.

The MA parameter in the fundamental representation is then given by[44]

$$\theta\left(\lambda, \rho, R_x^2\right) = \frac{1 - \left(1 - 4\kappa\left(\lambda, \rho, R_x^2\right)^2\right)^{\frac{1}{2}}}{2\kappa\left(\lambda, \rho, R_x^2\right)} \tag{28}$$

[44] The other solution to (23) gives the nonfundamental representation.

A.3 Pseudo Predictor Representations

To derive the representation in terms of the fundamental pseudo predictor in (5) and (6), write the ARMA as

$$r_t = \left(\frac{1-\theta L}{1-\lambda L}\right)\varepsilon_t = \varepsilon_t + \frac{(\lambda L - \theta L)\varepsilon_t}{1-\lambda L} = \varepsilon_t + (\lambda - \theta)\frac{\varepsilon_{t-1}}{1-\lambda L}$$

if we then define the fundamental pseudo predictor and β_f as in (19), and substitute, we can write the system as in (5) and (6) which is nested within the general system (1) and (2). Analogous substitutions can be used to derive (12) and (13).

B The Fundamental ARMA R^2

We have

$$R_f^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_r^2}$$

We can use the Yule-Walker equations to derive

$$\sigma_r^2 = \left(\frac{1-\lambda^2+(\theta-\lambda)^2}{1-\lambda^2}\right)\sigma_\varepsilon^2$$

hence

$$R_f^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_r^2} = \frac{(\theta-\lambda)^2}{1-\lambda^2+(\theta-\lambda)^2} \tag{29}$$

C The Variance Ratio and ARMA Parameters

C.1 The Unconditional Variance Ratio

The standard definition of the unconditional variance ratio given in Cochrane (1988) is

$$VR(h) = 1 + 2\sum_{j=1}^{h-1}\left(\frac{h-j}{h}\right) corr(j) \quad h = 1, 2, \ldots \tag{30}$$

where $corr(h) = corr(r_t, r_{t-h})$. In the ARMA(1,1) we have, using the Yule-Walker equations,

$$\begin{aligned} corr(1) &= \frac{cov(r_t, r_{t-1})}{var(r_t)} = -\left(\frac{(\theta-\lambda)(1-\theta\lambda)}{1-\lambda^2+(\theta-\lambda)^2}\right) && (31) \\ corr(j) &= corr(1)\lambda^{j-1}; \quad j > 1 && (32) \end{aligned}$$

hence substituting and evaluating the sum we have

$$VR(h) = 1 + 2.\frac{corr(1)}{1-\lambda}\left(1 - \frac{1-\lambda^h}{h(1-\lambda)}\right) \tag{33}$$

and note that $\frac{1-\lambda^h}{h}$ is decreasing in h for $h > 1$ so $VR(h)$ decreases or increases monotonically from $VR(1) = 1$.

As $h \to \infty$ we have the limiting variance ratio

$$\lim_{h\to\infty} VR(h) = 1 + 2.\frac{corr(1)}{1-\lambda} = 1 - \frac{2}{1-\lambda}\left(\frac{(\theta-\lambda)(1-\theta\lambda)}{1-\lambda^2+(\theta-\lambda)^2}\right)$$

noting that

$$1 - R_f^2 = \frac{1-\lambda^2}{1-\lambda^2+(\theta-\lambda)^2} = \frac{(1-\lambda)(1+\lambda)}{1-\lambda^2+(\theta-\lambda)^2}$$

and writing

$$\begin{aligned} V = \lim_{h\to\infty} VR(h) &= \tfrac{1}{1-\lambda}\left(\tfrac{(1-\lambda)\left(1-\lambda^2+(\theta-\lambda)^2\right)-2(\theta-\lambda)(1-\theta\lambda)}{1-\lambda^2+(\theta-\lambda)^2}\right) \\ &= \tfrac{\left(1-R_f^2\right)\left((1-\lambda)\left(1-\lambda^2+(\theta-\lambda)^2\right)-2(\theta-\lambda)(1-\theta\lambda)\right)}{(1-\lambda)^2(1+\lambda)} \end{aligned}$$

which simplifies to

$$V = (1 - R_f^2)\left(\frac{1-\theta}{1-\lambda}\right)^2$$

Now by inspection

$$\theta > \lambda \Rightarrow V < 1$$

Conversely

$$\begin{aligned}
V &< 1 \Rightarrow \\
\left(\frac{1-\lambda^2}{1-\lambda^2+(\theta-\lambda)^2}\right)(1-\theta)^2 &< (1-\lambda)^2 \\
0 &< \lambda^2-\lambda-\lambda\theta^2+\lambda^2\theta^2-\theta\lambda+\lambda^2\theta-\lambda^3\theta+\theta \\
0 &< (1-\lambda^3)\theta+\left(1+\theta+\theta^2\right)(\lambda^2-\lambda) \\
\left(1+\theta+\theta^2\right)\lambda(1-\lambda) &< (1-\lambda^3)\theta \\
\frac{\lambda(1-\lambda)}{1-\lambda^3} &< \frac{\theta}{1+\theta+\theta^2}=\frac{\theta(1-\theta)}{1-\theta^3} \\
f(\lambda) &< f(\theta)
\end{aligned}$$

with $f(z)=\frac{z(1-z)}{1-z^3}$. Now $f(.)$ is a monotone increasing function since

$$f'(z)=\frac{1-z^2}{\left(1+z+z^2\right)^2}>0 \text{ for } |z|<1$$

hence

$$V<1 \Leftrightarrow \theta>\lambda$$

and this plus monotonicity of $VR(h)$ gives the result in (15).

Finally note that we have

$$VR(h)=V+(1-V)\sqrt{\frac{H\left(h,\lambda\right)}{h}} \tag{34}$$

where

$$H(h,\lambda)=\frac{1}{h}\left(\frac{1-\lambda^h}{1-\lambda}\right)^2$$

C.2 The Conditional Variance Ratio

By definition, we have, for any predictor x_t, using the definition in Section 6,

$$R_x^2\left(h\right)=1-\frac{\sigma_t^2\left(h\right)}{\sigma^2\left(h\right)}=1-\frac{\sigma_t^2\left(h\right)/\sigma^2\left(1\right)}{VR\left(h\right)}$$

where $\sigma_t^2(h)$ is the conditional variance of the h-period average return, and $\sigma^2(h)$ its unconditional variance. Hence,

$$\frac{\sigma_t^2(h)}{\sigma^2(1)} = \left(1 - R_x^2(h)\right) VR(h)$$

and since, $\sigma_t^2(1) = (1 - R_x^2)\,\sigma^2(1)$ the conditional variance ratio (defined as in Pastor & Stambaugh, 2008) is

$$\begin{aligned} VR_t(h) &= \frac{\sigma_t^2(h)}{\sigma_t^2(1)} = S(h)\,VR(h) \\ \text{where } S(h) &= \frac{1 - R_x^2(h)}{1 - R_x^2} \end{aligned}$$

hence note that

$$VR_t(\infty) = \frac{1}{1 - R_x^2} VR(\infty) = \frac{V(\lambda, \theta)}{1 - R_x^2}$$

the better the predictor, the higher is the asymptote for the conditional ratio, relative to that of the unconditional ratio, defined in (16).

For any predictor with a hump-shaped horizon profile (see Section 6 and Appendix G) $S(h)$ is initially downward sloping in h, reaches a minimum at $h = h^* = \arg\max\left(R_x^2(h)\right)$ (as in (18)) and then slopes upwards. If there is no hump $S(h)$ simply slopes upwards. A sufficient condition for $VR_t(h)$ to be always less than unity is therefore if $VR(h)$ and $VR_t(\infty)$ are both less than unity. The latter requires $R_x^2 < 1 - V$, but, from Proposition 1, we have $R_x^2 \leq R_n^2(\lambda, \theta)$, hence a sufficient condition for the conditional variance ratio for the true predictor to be always less than unity is

$$R_n^2(\lambda, \theta) < 1 - V(\lambda, \theta)$$

which solves, using (37) and (16), to give the very simple condition

$$\theta > \frac{1}{2 - \lambda} = \lambda + \frac{(1 - \lambda)^2}{2 - \lambda} > \lambda \tag{35}$$

hence this condition is somewhat (but only marginally) stronger than the condition $\theta > \lambda \Rightarrow VR(h) < 1 \forall h$ (see 15), hence both conditions are satisfied.

The condition in (35), which implies only a modestly downward sloping profile for $VR(h)$, is easily satisfied by both ARMA representations of real US stock returns used in Section 5.

D Proof of Proposition 1

We wish to establish the inequality

$$R_f^2(\lambda,\theta) \leq R_x^2 \leq R_n^2(\lambda,\theta)$$

D.1 Relation of R_x^2 to R_f^2

The first inequality is straightforward. Using the derivation of the ARMA(1,1) representation in Appendix A we have, for $-1<\rho<1$, $\beta_x \neq 0$,

$$\begin{aligned}
\varepsilon_t &= \frac{1}{1-\theta L}\left[-\beta_x v_{t-1} + (1-\lambda L)u_t\right] \\
&= u_t - \lambda u_{t-1} - \beta_x v_{t-1} + \theta\varepsilon_{t-1} \\
&= u_t + \psi_{t-1}
\end{aligned}$$

hence

$$var(\varepsilon_t) = var(u_t) + var(\psi_t) > var(u_t)$$

since $cov(u_t, \psi_{t-1}) = 0$, hence

$$\begin{aligned}
1 - \frac{\sigma_\varepsilon^2}{\sigma_r^2} &< 1 - \frac{\sigma_u^2}{\sigma_r^2} \\
R_f^2 &< R_x^2
\end{aligned}$$

For the limiting case of the fundamental pseudo predictor representation in (5) and (6), we have $\sigma_u^2 = \sigma_\varepsilon^2 \Rightarrow R_x^2 = R_f^2$; and for $\beta_x = 0$, trivially, we have $R_x^2 = R_f^2 = 0$. so for the general case we have

$$R_f^2 \leq R_x^2$$

D.2 Relation of R_x^2 to R_n^2

For $\theta \neq 0$ we have the non-fundamental representation

$$r_t = \left(\frac{1 - \theta^{-1} L}{1 - \lambda L} \right) \eta_t \tag{36}$$

where η_t is the non-fundamental innovation, and we know (Hamilton, 1994, pp 66-67)

$$\sigma_\eta^2 = \theta^2 \sigma_\varepsilon^2$$

hence

$$R_n^2 = R_f^2 + \left(1 - \theta^2\right) \frac{\sigma_\varepsilon^2}{\sigma_r^2} = R_f^2 + \left(1 - \theta^2\right) \left(1 - R_f^2\right)$$

which, after substituting from (29) gives

$$R_n^2 = \frac{(1 - \theta\lambda)^2}{1 - \lambda^2 + (\theta - \lambda)^2} \tag{37}$$

which can also be derived directly from (29) by substituting θ^{-1} for θ.

Note that for $\theta = 0$ the non-fundamental representation (9) is undefined. As $\theta \to 0$ we have $\sigma_\eta^2 = \theta \sigma_\varepsilon^2 \to 0$, but, expanding (10),

$$E_t r_{t+1} | x_t^n = - \left(\theta^{-1} - \lambda\right) \left[-\theta r_{t+1} - \theta^2 r_{t+2} + ...\right]$$

so we have $\lim_{\theta \to 0} E_t r_{t+1} | x_t^n = r_{t+1}$, consistent with setting $\theta = 0$, giving $R_n^2 = 1$ in (37). For $\theta = \pm 1$ the non-fundamental and fundamental representations coincide but the moving average representations do not converge either backwards or forwards. However the formula in (37) is valid for any $\theta \in [-1, 1]$

We wish to establish the inequality

$$G\left(\lambda, \rho, R_x^2\right) \equiv R_n^2 \left(\lambda, \theta \left(\lambda, \rho, R_x^2\right)\right) - R_x^2 \geq 0$$

While G depends in principle on the triplet (λ, ρ, R_x^2) we shall analyse its properties for a given (λ, R_x^2) pair; we shall show that the result holds for any (λ, R_x^2) within their allowable ranges. Note that for this proof we do not require λ to be positive.

Thus we can write $G = G(\theta(\rho))$. From (27) we also have $\partial\theta/\partial\rho > 0$, hence we can write

$$\begin{aligned} G &= G(\theta)\,; \quad \theta \in [\theta_{\min}, \theta_{\max}] \\ &\text{where} \\ \theta_{\min} &= \theta\left(\lambda, -1, R_x^2\right); \quad \theta_{\max} = \theta\left(\lambda, 1, R_x^2\right) \end{aligned}$$

and we have

$$\begin{aligned} G'(\theta) &= \frac{\partial R_n^2}{\partial\theta} = -2\theta\left(\frac{\left(1-\lambda^2\right)(1-\theta\lambda)}{1-\lambda^2+(\theta-\lambda)^2}\right) \\ &\Rightarrow \; sign\left(G'(\theta)\right) = -sign(\theta) \end{aligned} \tag{38}$$

There are four cases:

Case 1: $\theta_{\min} > 0$: For this case,we have $G(\theta_{\max}) = 0$, since at this point x is the non-fundamental pseudo predictor with $R_x^2 = R_n^2$, and, from (20) we have $\rho = \rho_n = sign\left(\theta^{-1} - \lambda\right) = 1$. From (38) we also have $G' < 0$ hence $G \geq 0$.

Case 2: $\theta_{\max} < 0$: For this case,we have $G(\theta_{\min}) = 0$, since at this point x is again the non-fundamental pseudo predictor with $R_x^2 = R_n^2$, but with $\rho = sign\left(\theta^{-1} - \lambda\right) = -1$. From (38) we have $G' > 0$ hence $G \geq 0$.

Case 3: $\theta_{\min} < 0 < \theta_{\max}$: For this case,we have $G(\theta_{\min}) = G(\theta_{\max}) = 0$, $G'(\theta_{\min}) > 0$; $G'(\theta_{\max}) < 0$, and, from (38) G has a single turning point at $\theta = 0$, hence again we have $G \geq 0$.

Case 4: $\theta_{\min} = \theta_{\max} = 0$: This is the limiting case in which $R_x^2 = R_n^2 = 1$ (ie, as discussed in Section 4.5, $x_t = r_{t+1}$), giving $G = 0$.

Since the inequality holds in Cases 1 to 4 for any θ and λ, by implication it also holds for any λ and R_x^2. We have thus established the right-hand inequality in (17) for all possible cases, completing the proof.■

E Proof of Proposition 2

Any given values of θ and λ must imply a condition on κ (as defined in (23)) of the form

$$\kappa(\lambda, \rho, R_x^2) = \frac{\theta}{1+\theta^2}$$

For given values of θ and λ this can be taken to imply a restriction on ρ, the correlation between the two underlying innovations, which solves to give

$$\rho(\theta, \lambda, R_x^2) = \frac{(\theta-\lambda)(1-\theta\lambda) + F(\lambda, R_x^2)^2\theta}{\left(1-\lambda^2+(\theta-\lambda)^2\right)F(\lambda, R_x^2)}; \quad \rho \in (-1,1) \tag{39}$$

where $F(\lambda, R_x^2)$ is as defined in (21). If the solved value for ρ lies outside this range, the triplet (θ, λ, R_x^2) is not feasible (ie, taking θ and λ as given, R_x^2 does not satisfy the inequality condition (17)).

The first order condition yields a unique stationary point:

$$\frac{\partial\rho(\theta, \lambda, R_x^2)}{\partial R_x^2} = 0 \Rightarrow R_x^2 = \frac{(\theta-\lambda)(1-\theta\lambda)}{\theta-\lambda+\theta(1-\theta\lambda)}$$

which after substituting into (39) yields a real solution if

$$(\theta-\lambda)\,\theta > 0$$

which is satisfied for $\theta > \lambda$, given $\lambda > 0$, as in the Proposition. The second-order condition confirms that this then yields the minimum value

$$\rho_{\min} = sign\,(\theta-\lambda)\,\frac{2\sqrt{(\theta-\lambda)(1-\theta\lambda)\,\theta}}{\left(1-\lambda^2+(\theta-\lambda)^2\right)} > 0.\blacksquare$$

F Proof of Proposition 3

Consider the regression

$$r_t = -\beta_x x_{t-1} - \beta_f x_{t-1}^f + u_t$$

where x_t^f is the fundamental pseudo predictor and x_t the predictor variable. Then given the true model (1) that generates the data, it must be that $\beta_f = 0$. Treating x_{t-1} as an omitted variable in the regression

$$r_t = -\beta_f x_{t-1}^f + w_t$$

then $w_t = -\beta_x x_t + u_t$.

Using the formula for omitted variable bias we can obtain population values via moment conditions as

$$-\beta_f = \frac{Cov\left(x^f, r\right)}{Var\left(x^f\right)} = \frac{Cov\left(x^f, -\beta_x x + u\right)}{Var\left(x^f\right)} = -\beta_x \frac{Cov\left(x^f, x\right)}{\sqrt{Var\left(x^f\right) Var\left(x\right)}} \frac{\sqrt{Var\left(x\right)}}{\sqrt{Var\left(x^f\right)}}$$

$$= -\beta_x corr(x^f, x) \frac{\sigma_x}{\sigma_{x^f}}$$

hence

$$\beta_f \frac{\sigma_{x^f}}{\sigma_r} = corr(x^f, x) \beta_x \frac{\sigma_x}{\sigma_r}$$

Squaring both sides we get

$$\left(\beta_f \frac{\sigma_{x^f}}{\sigma_r}\right)^2 = \left(\beta_x \frac{\sigma_x}{\sigma_r}\right)^2 corr(x^f, x)^2$$

or

$$R_f^2 = corr(x^f, x)^2 R_x^2$$

which rearranges to give the expression in the Proposition■

G The horizon R^2

Using (1) and (2), we have

$$E_t r_{t+h} = -\beta_x E_t x_{t+h-1} = -\beta_x \lambda^{h-1} x_t$$

$$E_t \left(\frac{1}{h} \sum_{i=1}^{h} r_{t+i}\right) = -\frac{\beta_x}{h} \sum_{i=1}^{h} \lambda^{h-1} x_t = -\beta_h x_t$$

where

$$\beta_h = \frac{\beta_x}{h}\left(\frac{1-\lambda^h}{1-\lambda}\right) = \beta_x\sqrt{\frac{H(h,\lambda)}{h}} \tag{40}$$

so the horizon coefficient profile is just a scaling of the variance ratio profile in (34).

We also know, from the definition of the variance ratio,

$$var\left(\frac{1}{h}\sum_{i=1}^{h} r_{t+i}\right) = \frac{1}{h^2}h.\sigma_r^2 VR(h) = \frac{\sigma_r^2}{h}VR(h) \tag{41}$$

hence

$$R_x^2(h) = \frac{\beta_h^2\sigma_x^2}{\frac{\sigma_r^2}{h}VR(h)} = \beta_x^2\frac{\sigma_x^2}{\sigma_r^2}\frac{H(h;\lambda)}{VR(h)} = R_x^2\frac{H(h;\lambda)}{VR(h)} \tag{42}$$

Note that $H(h;\lambda)$ and $VR(h)$ depend only on ARMA parameters and h. Since we could also derive an equivalent horizon profile, $R_f^2(h)$ for the fundamental pseudo predictor, by replacing R_x^2 with R_f^2, the horizon profile for the true predictor is a scaling of the univariate horizon profile and hence has the same turning point, $h^*(\lambda,\theta)$.

References

Ang, A and G Bekaert (2007) "Stock Return Predictability: Is it There?" *Review of Financial Studies* 20(3):651-707

Beveridge, S. and C R Nelson. (1981) "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle". *Journal of Monetary Econonomics*, 7.pp151-174.

Boudoukh J, Richardson M and R Whitelaw (2008) "The Myth of Long Horizon Predictability" *Review of Financial Studies* 21(4):1577-1605

Campbell J Y and R J Shiller (1998) "Valuation Ratios and the Long-Run Stock Market Outlook" *Journal of Portfolio Management*, 24, pp11–26

Campbell, J Y, and S Thompson (2008) "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?", *Review of Financial Studies* 2008 21(4):1509-1531

Campbell, J Y., and L M. Viceira, (2002), *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press)

Campbell, J. Y., and M. Yogo, (2006), “Efficient Tests of Stock Return Predictability,” *Journal of Financial Economics*, 81(1), 27–60.

Cochrane, J.H (1988) “How big is the random walk in GDP?. ” *Journal of Political Economy*, v. 96, n. 5, p. 893-92

Cochrane J H (1994) “Permanent and Transitory Components of GNP and Stock Prices ” *The Quarterly Journal of Economics* Vol. 109, No. 1, pp. 241-265

Cochrane J H (2005), *Asset Pricing,* Princeton, NJ

Cochrane J H (2008a) “The Dog that Did Not Bark: A Defense of Return Predictability” *Review of Financial Studies* 2008, 1533-1575.

Cochrane, J H (2008b) “State-Space vs VAR Models for Stock Returns” *Mimeo*

Cowles, A C (1938) *Common-Stock Indices* Principia Press, Bloomington Indiana

Fama E F and K R French (2002) “The Equity Premium”, *Journal of Finance* 62, pp637-659.

Fernandez-Villaverde, J, Rubio-Ramirez, F, Sargent, T and Watson, M (2007) "ABCs (and Ds) of Understanding VARs", *American Economic Review*, pp 1021-1026

Ferson W E Simin T and S Sarkissian (2003) “Spurious regressions in Financial Economics?”, *Journal of Finance* 58, 1393-1414 (August)

Goyal, A., and I. Welch (2003). “Predicting the equity premium with dividend ratios”, *Management Science* 49, 639-654.

Hamilton J D (1989) “A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle” *Econometrica* 57:357-84

Hamilton J D (1994) *Time Series Analysis* 1994 Princeton University Press

Hansen, L and Sargent, T (2005), *Recursive Models of Dynamic Linear Economies*

Kim M J C R Nelson and R Startz (1991) "Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence" *The Review of Economic Studies*, Vol. 58, No. 3, pp. 515-528

Lippi, M and L Reichlin (1994) "VAR analysis, nonfundamental representations, Blaschke matrices", *Journal of Econometrics*, 63 pp 307-325

Pesaran M.H., and A. Timmerman (1995): "Predictability of Stock Returns: Robustness and Economic Significance" *Journal of Finance*, 50, 1201-1228.

Pastor L and R F Stambaugh (2008) "Are Stocks Really Less Volatile in the Long Run?" *Working Paper*

Pastor L and R F Stambaugh (2009) "Predictive Systems: Living with Imperfect Predictors" *Journal of Finance* 64(4) p1583-1628

Poterba, J.M. and L H Summers, (1988). "Mean reversion in stock prices: evidence and implications". *Journal of.Financial Economics*. 22, 27–59.

Robertson D and S Wright (2009) "Stambaugh Correlations and Redundant Predictors" *Working Paper*

Shiller R J (2000) *Irrational Exuberance* Princeton University Press

Siegel J J (1998) *Stocks for the Long Run* McGraw-Hill

Stambaugh R F (1999) "Predictive Regressions", *Journal of Financial Economics* 54: 375–421.

Timmermann, A, and B. Paye, 2006, "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.