Testing for Redundant Predictor Variables*

Donald Robertson
University of Cambridge

Stephen Wright
Birkbeck College, University of London

February 2, 2009

Abstract

We consider inference in a widely used predictive model. "Stambaugh Bias" arises when innovations to the predictor variable are correlated with those in the predictive regression. We show that such correlations will arise if the predictor is actually redundant, but simply proxies univariate predictability. In these circumstances even bias-corrected conventional tests of redundance will be severely distorted. We propose tests that distinguish well between redundant predictors and the "true" predictor. Empirical results suggest that several widely used predictors of stock returns are redundant, and that univariate predictability provides most of the explanatory power of the consumption/GNP ratio for output growth.

Keywords: Predictive regressions, Stambaugh bias, ARMA models, cointegrating models, Granger Causality.

*We thank John Campbell, John Cochrane, James Davidson, Andrew Harvey, Richard Smith, Ron Smith, Peter Tinsley, Melvyn Weeks, and participants at seminars in Cambridge and Exeter for helpful comments and advice.

1 Introduction

There is an extensive literature, mainly in empirical finance, that focusses on inference difficulties in a predictive framework of the form

$$y_t = \beta_z z_{t-1} + u_{zt} \quad (1)$$

$$z_t = \lambda_z z_{t-1} + v_{zt} \quad (2)$$

where the first equation captures the degree of predictability of some variable, y_t in terms of some predictor variable z_t, while the second models the predictor as an AR(1) process. In the finance literature y_t is usually some measure of returns or excess returns, while z_t is frequently some indicator of value, such as the price-dividend ratio; but the framework is in principle much more general.[1] Usually we are interested in testing $H_0 : \beta_z = 0$, .ie, that z_{t-1} is predictively redundant for y_t. Stambaugh (1999) noted that if the predictor variable is persistent and the two innovations are correlated, conventional OLS will lead to estimates of β_z that are biased in finite samples. Other authors have also noted the risks of various forms of data mining - whether in choosing from a list of possible regressors (eg, Ferson et al,2003) or by sample selection (eg Goyal & Welch, 2003), while yet others have noted that Stambaugh Bias also affects long-horizon regressions (Boudhoukh et al,2006).[2]

It is now standard to subject predictive regressions to various corrections for bias, structural stability and data mining. Only when the null $H_0 : \beta_z = 0$ can be convincingly rejected after these corrections is a predictor variable deemed to have any genuine statistical significance.

In this paper we show that even these stringent test procedures may not be enough. Predictor variables can be redundant even if the standard null $\beta_z = 0$ (which we shall henceforth refer to as the "Stambaugh Null"[3]) is rejected.

[1] The predictive framework can also, for example, be interpreted as a reparameterised cointegrating VAR(1), where $z_t \equiv s_t - \theta q_t$ is some candidate cointegrating relation between two non-stationary variables, and the first equation is one of the implied error correction equations (eg, $y_t \equiv \Delta s_t$).

[2] If the system is interpreted as in the previous footnote as a reparameterised cointegrating VAR, then tests on $\widehat{\beta}_z$ are equivalent to tests of no Granger Causality, taking cointegration as given (ie, assuming z_t is stationary). Interestingly the impact of Stambaugh Bias on such tests does not appear to have received any attention in the cointegration literature.

[3] Stambaugh's original paper focused on a more general case of small samples where in principle the true

Under the Stambaugh Null, $\beta_z = 0$ implies $y_t = u_{zt}$ which is usually assumed to be white noise. Under our proposed generalisation of the Stambaugh Null we allow it to follow an ARMA(1,1) process. This is not an arbitrary assumption. If, for example, there exists some other underlying but possibly unobserved true predictor variable, x_t, in a predictive framework of the same form as (1) and (2) but with $\beta_x \neq 0$, then the reduced form for y_t will be an ARMA(1,1). Under this null, if a predictive regression of the form of (1) is estimated by OLS, conditioning only on z_t, then even if z_t is a redundant predictor, it may be sufficiently correlated with the predictions from the ARMA representation to reject the Stambaugh Null in its usual form, even after correcting for biases in the usual way. Furthermore, if z_t has emerged from a predictive "horse race" we would positively *expect* to find a significant "Stambaugh Correlation" between the innovations in (1) and (2).

We suggest a simple test procedure for the generalised Stambaugh Null that z_t is a redundant predictor. Our tests exploit the property that under the null that z_t is a redundant predictor, its *marginal* predictive power must disappear, once we condition correctly upon the univariate prediction.[4] Our tests appear to have reasonable power against the obvious alternative, that $z_t = x_t$, i.e., is in fact the true predictor, despite the fact that the true and a redundant predictor are likely to have very similar characteristics.

We provide two empirical illustrations. We show that tests of our generalised version of the Stambaugh Null are particularly relevant when testing for predictability of stock returns. We examine four candidate predictors that have been used in past research, several of which appear to reject the standard Stambaugh Null even using bias-corrected test procedures. However, most signally fail to reject our generalised null - leading to the conclusion that such indicators are doing little or nothing more than proxy univariate properties. We also show that the test procedure potentially has wider relevance to tests for cointegrating relationships, which we illustrate with a test based on Cochrane's (1994) use of the consumption-GNP ratio as a

value of β_z may be non-zero, but in practice almost all applications of his analysis have focussed on the null of no predictability.

[4] In so doing we are simply carrying out a test of the null of no Granger Causality from z_t to y_t in the general form originally specified in Granger (1969). However we note below that a key difference in our test procedure is that, in contrast to Granger Causality tests as conventionally implemented (of which the usual Stambaugh Null is a special case), our test procedure allows y_t to be an ARMA process of the same order under both H_0 and H_1. We also discuss extensions to allow for higher order ARMA processes.

predictor for GNP growth.

The paper is structured as follows. In Section 2 we derive the relationships between the predicted process and both true and redundant predictors. In Section 3 we describe our proposed test procedures. In Section 4 we provide two empirical examples, and Section 5 concludes. Appendices provide technical details.

2 Univariate Characteristics and Predictor Variables

2.1 The true predictive regression and its ARMA(1,1) reduced form

We assume that the true process for y_t can be represented by a predictive regression in terms of a true (and possibly unobservable) predictor, x_t, and an autoregressive representation of the true predictor[5]

$$y_t = \beta_x x_{t-1} + u_{xt} \tag{3}$$

$$x_t = \lambda x_{t-1} + v_{xt} \tag{4}$$

where $\begin{bmatrix} u_t & v_t \end{bmatrix}'$ is a vector of serially uncorrelated shocks with $\rho = corr\left(u_{xt}, v_{xt}\right) \in (-1, 1)$. We assume $0 \leq \lambda < 1$,[6] and neglect constants for simplicity.

While the system in (3) and (4) has the same structure as the system in terms of z_t in (1) and (2), we stress that for the true predictor, x_t, we have $\beta_x \neq 0$ by assumption.

Substituting from (4) into (3) we can derive the reduced form univariate process for y_t, which is an ARMA(1,1):

$$y_t = \left(\frac{1-\theta L}{1-\lambda L}\right)\varepsilon_t \tag{5}$$

where L is the lag operator (defined by $Lx_t = x_{t-1}$); ε_t is a composite univariate innovation

[5] In many contexts (most notably if y_t is some measure of asset returns) the white noise property of u_{xt} arises naturally from some combination of rational expectations and efficient markets. The restrictive nature of the model lies in the model of x_t, the predictor, rather than the predictive regression itself: ie that efficient forecasts of y_t can be expressed in terms of a single AR(1) state variable.

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

that is white noise conditional upon the history $\{y_i\}_0^{t-1}$. As long as u_{xt} and v_{xt}, the innovations in (3) and (4) are less than perfectly correlated, we can always derive a representation that is strictly invertible (or "fundamental"), hence we assume $\theta \in (-1, 1)$. If $\theta = \lambda$ the AR and MA components cancel, and there is no univariate predictability. Note that this special case, in which the reduced form for y_t is white noise, can arise even when $\beta_x \neq 0$.[7]

Equation (5) shows that the autoregressive coefficient of the predictor variable translates directly to the AR coefficient of the reduced form. For the MA parameter things are more complicated: in Section 2.4 we show how it is determined by the characteristics of the underlying predictive system (3) and (4).

The system (3) and (4) and its reduced form in (5) are both representations of the true process for y_t, which differ only by the nature of the conditioning information set.[8]

2.2 A "pseudo predictor"

Starting from the ARMA(1,1) representation in (5) we can reverse-engineer a useful limiting representation of the same form as (3) and (4). Define the "fundamental pseudo predictor", denoted x_t^f, by

$$x_t^f = \frac{\varepsilon_t}{1 - \lambda L} \tag{6}$$

which, if we set $\beta_f = \lambda - \theta$, and substitute into (3) and (4) (implying $u_{xt} = v_{xt} = \varepsilon_t$) gives us back (5). The predicted value of y_t from the ARMA representation is thus simply a scaling of x_{t-1}^f.

Note that, using (6) and (5) the fundamental pseudo predictor can also be derived as an

[7] This feature of the reduced form is well-known: see for example Campbell, Lo and Mackinlay (1995)

[8] While the ARMA(1,1) representation might seem overly restrictive we can easily generalise at least to an ARMA$(p, 1)$, which arises if (3) includes $p - 1$ lags of y_t.This can be reduced to an ARMA(1,1) representation by quasi-differencing and re-defining the dependent variable (see our second empirical example in Section 4.2). The general case with multiple predictor variables, that can in turn be represented by a VAR(1) in q variables, cannot be reduced below an ARMA(q, q) : hence the key issue is how many predictor variables (or in the cointegrating case, how many cointegrating relations) there are.

infinite moving average of the predicted process itself[9]

$$x_t^f = \frac{y_t}{1-\theta L} \tag{7}$$

The pseudo predictor will have a predictive R-squared for y_t identical to the ARMA R-squared, and, by inspection of (6), a Stambaugh correlation of precisely unity ($\rho = \rho_f = 1$).[10]

2.3 A redundant predictor

Consider the following general definition of a redundant predictor:

Definition 1 *A redundant predictor for some variable y_t contains no information of predictive value that is not already in the history of the predicted variable itself, $\{y_i\}_{i=0}^{t-1}$.*

Note that this is simply a statement that z_t does not Granger-Cause y_t, in the original very general sense of Granger (1969). However, in practical applications the null of No Granger Causality is almost always represented in the more restrictive sense of predictive redundance conditional upon a finite order autoregressive representation of y_t (and possibly a set of other variables).[11] When the reduced form for y_t is an ARMA representation, as in (5), this distinction can be very important, as we shall show below.

Supppose that z_t is a redundant predictor in the sense of Definition 1. We can reinterpret the predictive regression (1) as a least squares linear projection of y_t on z_{t-1} and a constant term. The resulting innovation process u_{zt} will be orthogonal to z_t but will not be white noise (although it may well be virtually indistinguishable from white noise in the data). If we are prepared to assume that z_t is correctly specified as an AR(1) we can derive the following properties:

[9] The irreducible ARMA(q,q) model discussed in the previous footnote can also always be represented in a similar way, with q pseudo-predictors defined, analogously to (7), by $x_{k,t}^f = (1-\theta_k L)^{-1} y_t, k = 1..q$ Thus the number of underlying predictors, which in turn determines the order of the MA component (and hence the number of pseudo predictors) is the crucial limiting factor.

[10] The fundamental pseudo-predictor in period t is also, up to a scaling factor and initial condition, identical to the optimal estimate of the true predictor variable using the Kalman Filter, when the information set consists only of the history $\{y_i\}_{i=0}^t$.

[11] See for example the discussion in Hamilton, 1994.

Proposition 1 *Assume that z_t follows the AR(1) process in (2).*

a) A sufficient condition for z_{t-1} to be redundant for y_t by Definition 1 is that v_{zt}, the innovation to z_t can be written as

$$v_{zt} = \psi_\varepsilon \varepsilon_t + \omega_t \tag{8}$$

where ε_t is the innovation in the ARMA reduced form (5), and ω_t is a white noise process that satisfies $E(\omega_t \varepsilon_{t-i}) = E(\omega_t u_{xt-i}) = E(\omega_t v_{xt-i}) = 0$, $\forall i$, and ψ_ε is some constant;

b) Let $R_z^2 \equiv 1 - var(u_{zt})/var(y_t)$ be the predictive R-squared from a linear projection of y_t on z_{t-1}, of the form in (1); and let $\rho_z = corr(u_{zt}, v_{zt})$ be the associated Stambaugh Correlation. If z_t is a redundant predictor by Definition 1, with innovations as in (8), then

$$\frac{R_z^2}{1 - R_z^2} \leq \rho_z^2 \frac{R_f^2}{1 - R_f^2} \tag{9}$$

where $R_f^2 \equiv 1 - \sigma_\varepsilon^2/\sigma_y^2$ is the predictive R-squared for the fundamental ARMA representation of y_t in (5). The expression for R_z^2 in (9) holds with equality if $\lambda_z = \lambda$, and the implied upper bound $R_z^2 = R_f^2$ is attainable for some z_t since $\rho_z \in [-1, 1]$.

Proof. See Appendix A. ■

The characteristics in Proposition 1 define a set of redundant AR(1) predictors with relatively simple and empirically relevant properties.[12] The specification of the innovation in (8) means that z_t is clearly redundant by Definition 1 (since any representation of the history $\{y_i\}_{i=0}^t$ has an equivalent represention in terms of $\{\varepsilon_i\}_0^t$). For any specification that allowed the innovation v_{zt} to be independently correlated with either current or lagged values of the true innovations u_{xt} and v_{xt}, z_t would not be redundant.

Part b) of the Proposition shows that even a redundant predictor can still appear to have significant predictive power in a predictive regression of the form in (1). Such a linear projection on a redundant predictor will have a degree of predictive power which, first, depends on how

[12] We show in the appendix that the complementary set of redundant AR(1) predictors that do not satisfy all of the conditions in part a) of the proposition (because $E(\omega_t \varepsilon_{t-i}) \neq 0$ for some i, or because ω_t is not white noise, or both) will either have a predictive R-squared that is strictly bounded below R_f^2, or will have innovations containing a process ω_t that must satisfy very tight restrictions such that v_{zt} remains white noise (which we require for z_t to be AR(1)).

well the ARMA representation itself predicts (ie, on how high R_f^2 is) and second, on how good a proxy the predictor is for the "pseudo predictor" defined in Section 2.2. This is in turn captured by two elements. The first is the Stambaugh correlation. In the Appendix we show that this is very closely related to the correlation between the innovations to the redundant predictor, in (8), and those to the ARMA representation in (5) (indeed we show that for y_t reasonably close to white noise, $\rho_z \approx \text{corr}(v_{zt}, \varepsilon_t)$). The second is the extent to which the the persistence of the redundant predictor matches the AR(1) parameter in the ARMA representation: we show that the closer λ_z is to λ, the better the prediction.

Proposition 1 provides a useful insight into the characteristics of observable predictors. We frequently observe high Stambaugh correlations in the data (particularly, but by no means exclusively, in the literature on predictability of returns[13]). In the existing empirical finance literature this characteristic is usually simply treated as a nuisance that complicates inference.[14] However, Proposition 1 suggests that it may be much more than this. Consider the case where a given predictor variable has been the result of a wider search of candidate predictors (cf Ferson, Simin and Sarkissian, 2003; Sullivan, Timmerman and White, 1999) by way of some form of data-mining horse race procedure. For a redundant predictor to survive this procedure, a high Stambaugh correlation is a *necessary* characteristic, since, from (9), the higher is the Stambaugh correlation, the higher is R_z^2. Furthermore, from (9), data-mining econometricians will also have a greater tendency to single out redundant predictors with AR(1) parameters as close as possible to the true predictor, since, for a given Stambaugh Correlation, this will also push up R_z^2. Indeed, if we follow the logic of data mining through to its ultimate conclusion, there is a straightforward corollary to Proposition 1:

Corollary 1 *Assume that data mining econometricians run horse races between redundant predictors with the characteristics given by Proposition 1, in which the sole objective is to maximise the predictive R-squared, R_z^2 in the predictive regression (1). If there is no constraint on the number of such potential indicators or on the manner in which they are constructed, the*

[13] In cointegrating models, as in footnote 1, a high Stambaugh Correlation will tend to follow naturally from the definition of the cointegrating relation, eg if $y_t = \Delta s_t$; $z_t = s_t - \theta q_t$, then the Stambaugh correlation will be high if innovations to Δs_t have high variance compared to, and are not strongly correlated with, those to Δq_t. This is essentially the case analysed in Cochrane (2008), pp 1548-9.

[14] An important exception is Cochrane (2008).

redundant predictor with the best track record will, up to a scaling factor, be arbitrarily close to the "pseudo predictor" defined in (6), with a Stambaugh correlation arbitrarily close to unity in absolute value.

This corollary is similar in spirit to the claim that an infinite number of monkeys typing for an infinite amount of time will almost surely type the complete works of Shakespeare:[15] here we are in effect modelling an indirect method of ARMA estimation by monkeys. The difference is that the monkey typists' behaviour is purely random, while our monkey econometricians could in principle be replaced by computer programs with relatively straightforward data-mining algorithms, with a well-defined objective function. And our empirical results will show that Corollary 1 appears to come quite close to explaining the nature of some predictors of stock market returns.[16]

2.4 Characteristics of the true predictor

In our framework the null hypothesis is that z_t is a redundant predictor. But for any test procedure to have practical value it needs to have power against the natural alternative, that z_t is in fact the true predictor, x_t.[17] It turns out that for a range of parameter values that appear empirically relevant (particularly, but not exclusively, for predictive return regressions) the characteristics of the true predictor may be superficially similar to those of a redundant predictor.

We have already noted that the AR parameter in the ARMA representation (5) corresponds directly to that of the true predictor. In Appendix B we show that, subject to an innocuous normalisation of the data for x_t that fixes the sign of β_x, the MA parameter θ can be expressed

[15] Strictly speaking, one monkey will do, but will take infinitely longer.

[16] The data-mining econometrician monkeys in Corollary 1 are of course unsophisticated in that they ignore Stambaugh Bias. If the monkeys' objective criterion included a penalty for Stambaugh Bias we would not expect to find Stambaugh Correlations of precisely unity. However, as our first empirical examples show, some predictors of stock returns get very close.

[17] We could in principle consider intermediate cases, for example where z_t is not the true predictor, but has predictive power, even conditioning upon the ARMA prediction, because it provides a noisy signal of x_t. However in this case even a non-redundant predictor would lead to a mis-specified predictive regression since the innovations would not be white noise. The cleaner alternative is therefore $H_1 : z_t = x_t$

in terms of three unit-free characteristics of the predictive model:

$$\theta = \theta\left(\lambda, \rho, R_x^2\right) \tag{10}$$

where $\rho = corr(u_{xt}, v_{xt})$ is the Stambaugh Correlation for the true predictor x_t in (3) and (4) and $R_x^2 = 1 - var\left(u_{xt}\right)/var\left(y_t\right)$ is its predictive R-squared. By using this relationship in reverse, the ARMA representation can tell us a lot about the nature of the true predictor. The following proposition summarises results discussed in more detail in Robertson and Wright (2009):

Proposition 2 *For the ARMA(1,1) univariate representation (5) which is a reduced form of a predictive regression (3) and a predictor autoregression (4), both defined in terms of the true predictor,* x_t,

a) The one-period-ahead R-squared of the predictive regression, R_x^2 *satisfies the inequality*

$$R_f^2 \leq R_x^2 \leq R_n^2 \tag{11}$$

where $R_f^2\left(\lambda, \theta\right) = \left(\theta - \lambda\right)^2\left(1 - \lambda^2 + \left(\theta - \lambda\right)^2\right)^{-1}$ *and* $R_n^2\left(\lambda, \theta\right) \equiv R_f^2\left(\lambda, \theta^{-1}\right)$ *are the predictive R-squareds of the "fundamental" and "non-fundamental" ARMA representations;*

b) If $\theta > \lambda > 0$ *(which is a sufficient condition for the horizon variance ratio, as defined in Cochrane (1988) to be less than unity at all horizons), then* $\rho = corr\left(u_{xt}, v_{xt}\right)$*, the Stambaugh correlation for the true predictor, satisfies*

$$|\rho| \geq \rho_{\min}\left(\lambda, \theta\right) > 0 \tag{12}$$

Proof. See Appendices B to D. ■

Taken as a whole, Proposition 2 implies significant constraints on the "predictive space" that the true predictor can inhabit.

Part a) shows that the ARMA parameters can be used to derive upper and lower bounds for the true predictive R-squared.

The lower bound is easy to interpret: the true predictor provides predictive information for y_t beyond that contained in the history of y_t itself (ie, it is not redundant). It must therefore have a higher predictive R-squared than the ARMA representation.

The upper bound is more novel, but also has a reasonably intuitive explanation. As is well known (see for example Hamilton, 1994, pp 66-67) the ARMA representation in (5), which is defined to be "fundamental" by constraining θ to lie in $(-1, 1)$, has an alternative representation that replaces θ with θ^{-1}, with innovations, η_t, with strictly lower variance than those in the fundamental representation, ie $\sigma_\eta^2 = \theta^2 \sigma_\varepsilon^2$. The non-fundamental innovation η_t cannot be recovered from the history of y_t but *can* be recovered as a linear combinations of future values of y_t. Intuitively, since the non-fundamental representation also conditions on the history of y_t, the increase in the information set automatically implies $\sigma_\eta^2 < \sigma_\varepsilon^2$, hence $R_n^2 \geq R_f^2$. The stronger result, that R_x^2 must lie within these two bounds, means that the information set provided by the non-fundamental representation dominates that of any possible AR(1) predictor. Equivalently, if we define a pseudo predictor in the same way as in (6), but in terms of η_t, the innovations to the non-fundamental representation, this "non-fundamental pseudo predictor" in period $t-1$ is the best possible AR(1) predictor of y_t, consistent with the ARMA representation.

In Robertson & Wright (2009) we show that the upper and lower bounds in (11) constrain R_x^2 non-trivially (ie to a space strictly within $(0, 1)$) except in the case where $\theta = \lambda = 0$ (which implies $R_f^2 = 0, R_n^2 = 1$) and can for some parameter values be quite restrictive.

Part b) of Proposition 2 shows that, if we restrict ourselves to a particular subset of ARMA representations for which the variance ratio slopes downwards (a feature that, for example, Cochrane (1988; 1994) asserts holds for both stock returns and GNP growth[18]) there is a further

[18] In the context of asset returns this is often referred to as mean reversion (following Poterba & Summers, (1988)), but this 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

restriction on the space that the true predictor can inhabit: its "Stambaugh Correlation" must be bounded away from zero. Indeed, it is possible to show that for a wide range of values satisfying $\theta > \lambda > 0$, $\rho_{\min}$ is quite close to unity. For such processes, Stambaugh bias will therefore be endemic for any true predictor just as it is for redundant predictors (albeit for very different reasons).[19]

The constraint that the Stambaugh correlation for the true predictor must lie close to unity suggests a superficial similarity both to the redundant predictor discussed in Section 2.2 (which we have seen must have a high Stambaugh correlation if it is to be a good proxy for the "fundamental pseudo predictor") and to the pseudo predictor itself, which has a Stambaugh Correlation of precisely unity. However, there is a very important difference. It is straightforward to show that the correlation between the true predictor and the pseudo predictor satisfies

$$corr\left(x_t, x_t^f\right)^2 = \frac{R_f^2}{R_x^2} \tag{13}$$

Thus, while the redundant predictor only has predictive power to the extent that it is correlated with the pseudo predictor, the true predictor is *less* correlated with the pseudo predictor, the *better* it predicts.[20]

2.5 A geometric comparison of predictor characteristics

Figure 1 provides a geometric comparison of the key characteristics of both true and redundant predictors.

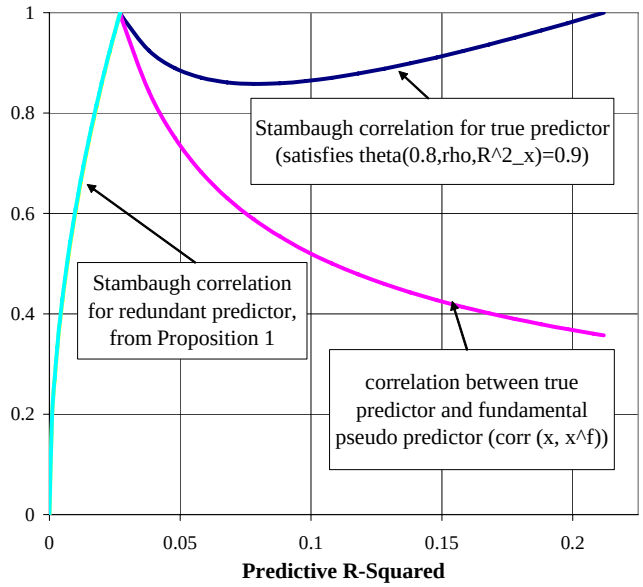
component will be mean-reverting (cf Kim et al (1991) who refer to the latter case as "mean aversion").

[19] It is possible to show that Part (b) of the proposition also holds for representations with $\theta < \lambda < 0$. We do not at present see this as empirically very interesting; but it does imply an even greater prevalence of Stambaugh Bias.

[20] See Appendix E.

Figure 1: Predictor Characteristics and Univariate Properties

An illustration for an ARMA(1,1) with $\lambda = 0.8$; $\theta = 0.9$.



As noted in discussion of equation (10), for any given (λ, θ) combination the true predictor can be represented by a (λ, ρ, R_x^2) combination, where ρ is the Stambaugh Correlation of the true predictor. To simplify matters we normalise ρ to lie between zero and one (which we can always do by changing the sign of x_t). The upper line in Figure 1 plots combinations of ρ and R_x^2 consistent with an ARMA representation with $\lambda = 0.8$, $\theta = 0.9$.[21] With this combination of parameters y_t is a near-white noise process but with a horizon variance ratio, $VR(h)$, that declines monotonically from unity to the asymptote $VR(\infty) = 0.24$. The lower and upper bounds for R_x^2 in Proposition 2 are given by the extreme points of this curve. At the far left is the limiting case where the true predictor is the (fundamental) pseudo predictor. The Stambaugh Correlation is unity and $R_x^2 = R_f^2 = 0.027$: ie, there is barely any univariate predictability. At the far right is the limiting case of the "non-fundamental pseudo predictor". The Stambaugh Correlation is again unity and $R_x^2 = R_n^2 = 0.219$: ie, this limiting predictor (which from Proposition 2, is the best possible AR(1) predictor consistent with the ARMA) predicts substantially better than the fundamental ARMA representation. If the true predictor lies anywhere between these two points it must have a very high Stambaugh Correlation: the low point of the curve, corresponding to $\rho_{\min}$ in part b) of Proposition 2, is at $\rho_{\min} = 0.858$.

[21] The formula for this line is given in Appendix D, equation (40).

The lower curve plots $\text{corr}\left(x_t, x_t^f\right)$, the correlation between the true predictor and the (fundamental) "pseudo predictor", using the expression in (13). The curve shows that a true predictor that is interestingly different from the pseudo predictor (because it predicts better by a reasonable margin) should also look fairly different (we shall use this as an informal diagnostic in our empirics).

The same diagram can also be used to plot out the relationship between R_z^2, the predictive R-squared from a linear projection of y_t on z_{t-1}, and its Stambaugh Correlation ρ_z, if z_t is a redundant predictor. The curve is drawn on the simplifying assumption that $\lambda_z = \lambda$, thus we simply use (9) in Proposition 1, which holds with equality.[22] As noted in Section 2.3, for a redundant predictor, a high Stambaugh Correlation is a requirement for it to have any chance of surviving horse races - hence we would expect to find it fairly high up on this curve. It is therefore likely that the true predictor and a redundant predictor will have very similar Stambaugh Correlations. However the true predictor must predict strictly better than the univariate representation, and will therefore have marginal predictive power even conditioning upon $\{y_i\}_{i=0}^t$ (this is the basis for our statistical tests).

While for a true predictor, we saw that, from (13) the better it predicts, the more dissimilar it is from the pseudo predictor, for a redundant predictor the reverse is the case: the better it predicts, the more *similar* it must be to the pseudo predictor. Indeed, for the case shown in the diagram, where $\lambda_z = \lambda$, it is straightforward to show that we have $corr\left(z_t, x_t^f\right) = corr\left(v_{zt}, \varepsilon_t\right) \approx \rho_z$.

Thus, despite, their superficial similarities, the true and redundant predictors differ in important ways that offer scope to discriminate between them. Even in the case illustrated in Figure 1, where y_t is close to being white noise, while the "predictive space" is significantly constrained, the true predictor may nonetheless differ significantly from the pseudo predictor (the correlation between the two, given by (13) has a minimum value of $\sqrt{R_f^2/R_n^2} = 0.36$), and will have a significantly higher predictive R-squared. It should therefore also, *a fortiori,* be readily distinguishable from a redundant predictor.

[22] For general λ_z, the curve can be interpreted as a lower bound for the Stambaugh correlation, for a given value of R_z^2.

3 Testing for predictive redundance

3.1 A general framework for testing the null of predictive redundance.

We wish to test the null hypothesis that z_t is a redundant AR(1) predictor with the characteristics in Proposition 1, given that the true process for y_t is (5) (or equivalently, the predictive system (3) and (4) in terms of the true predictor, x_t). We consider tests of this null against the alternative, $H_1 : z_t = x_t$. A key feature of this testing framework is that y_t is ARMA(1,1) under both the null and the alternative.

In contrast, in the standard testing framework a predictive regression of the form in (1) and the predictor autoregression (2) are analysed on the assumption that u_{zt} is white noise, hence under the "Stambaugh Null" $H_0 : \beta_z = 0$, y_t is also white noise. The alternative hypothesis in the standard framework is the same as in ours, namely $H_1 : z_t = x_t$, hence y_t is ARMA(1,1) in reduced form. Thus for tests of the standard Stambaugh Null (as indeed in any conventional test of Granger Causality) the order of the ARMA representation differs between the null and the alternative.

Under our null, z_{t-1} may appear to have statistically significant predictive power for y_t, however it is redundant once we condition on the history $\{y_i\}_{i=0}^{t-1}$. Both null and alternative can be represented in a multivariate predictive regression which includes the (fundamental) pseudo predictor, as defined in (6), as an additional regressor,

$$y_t = \gamma_1 x_{t-1}^f + \gamma_2 z_{t-1} + \xi_t \tag{14}$$

then the null of predictive redundance is $H_0 : \gamma_2 = 0$ (and $\xi_t = \varepsilon_t$) and, for the alternative that z_t is the true predictor, x_t we have H_1:$\gamma_1 = 0$ (and $\xi_t = u_{xt}$).

Note that our generalised null nests the standard Stambaugh Null as a special case if the true predictor has the characteristics (λ, ρ, R_x^2) such that $\theta(\lambda, \rho, R_x^2) = \lambda$, and hence y_t is white noise. In this case both γ_1 and γ_2 are zero under the null. This does not rule out the existence of a true predictor variable, x_t with a predictive R^2 that may in principle be anywhere in the

range given by (11); but under the null z_{t-1} is as useless in predicting y_t as the pseudo predictor, x^f_{t-1}, hence the standard Stambaugh null applies.

The practical obstacle to the implementation of a test procedure based on (14) is that it requires an estimate of the fundamental pseudo predictor, x^f_t, or equivalently of the one-step-ahead prediction from the ARMA representation. We consider three tests that use alternative approaches to this problem. We first describe the tests, and then, in Section 3.3, discuss their sampling properties.

3.1.1 The indirect pseudo predictor approach with $\lambda_z = \lambda$ (RP_1)

In order to include an estimate of the pseudo predictor in a multivariate predictive regression of the general form (14), we require an estimate of the MA parameter θ. But direct estimation of ARMA parameters is known to be problematic, particularly when θ is close to the AR parameter, λ (in the limit, for $\theta = \lambda$, the parameters are not identified). However, our first test procedure circumvents these difficulties by modifying the null to include a further restriction, that allows us to derive an *in*direct estimate of θ, solely from the properties of the predictive system in (1) and (2). It turns out that, under this modified null, the resulting indirect estimate has distinctly better sampling properties than the direct estimate.

From the properties of the function given in (10), θ must satisfy

$$\theta\left(\lambda, \rho, R^2_x\right) = \theta\left(\lambda, 1, R^2_f\right) \tag{15}$$

ie, θ must be consistent both with the properties of the true predictor, and, by definition, with those of the pseudo predictor, which has a Stambaugh Correlation of unity and the same predictive R-squared as the ARMA representation. Under the joint null that z_t is redundant *and* that $\lambda = \lambda_z$,[23] the predictive system in (1) and (2) allows us to derive estimates of both λ and R^2_f, which in turn, using the right-hand side of (15) allow us to derive an indirect estimate of θ.

[23] From Corollary 1 this is a natural restriction on the null model if the predictor has arisen from a process of data mining. This does however have implications for the choice of critical values (see discussion in Section 4 below).

The estimate of λ is straightforward since we can set $\lambda = \widehat{\lambda}_z$. But we can also exploit the fact that under the restricted null the inequality between R_z^2, ρ_z and R_f^2 in (9) in Proposition 1 holds with equality. Thus if we have estimates of R_z^2 and ρ_z from the system in (1) and (2) then we can derive an implied value of R_f^2, which coupled with an estimate of λ_z, is sufficient to derive an estimate of θ. We refer to this estimate as $\widehat{\theta}_z$.[24]

Given this indirect estimate of θ, an estimate of the pseudo predictor $\widehat{x}_t^f$ consistent with the null can then be derived using (7), conditional upon some initial estimate $\widehat{x}_0^f$ which for simplicity can be set to its unconditional mean of zero. The null that z_t is a redundant predictor can then be tested by a simple F-test of the hypothesis that γ_2 is zero in the multivariate predictive regression (14), with x_t^f replaced by $\widehat{x}_t^f$. We refer to this test statistic as RP_1(**R**edundant **P**redictor$_1$).

The seemingly convoluted nature of this indirect approach can be readily justified by sampling properties. We show in Appendix F that, under this restricted version of our null, the implied indirect estimate, $\widehat{\theta}_z$ has distinctly lower sampling variation than the direct ARMA estimate (particularly when the true value of λ and θ are very close), and hence we have a better estimate of the pseudo predictor in deriving the test statistic. However a clear disadvantage is the usual joint null problem, that we may reject the null, even when z_t is truly redundant, because λ_z differs from λ. Our other two tests do not suffer from this joint null problem, but have the offsetting disadvantage that they rely on ARMA estimation.

3.1.2 A two stage ARMA-based approach with $\lambda_z \neq \lambda$ (RP_2)

We noted in Section 2.2 that the one-step-ahead prediction of y_t from the ARMA representation is simply a scaling of the pseudo predictor. The reverse is also, straightforwardly, the case, and our second test exploits this equivalence. In the first stage of the test procedure we estimate an ARMA(1,1) representation of y_t. We then take the one-step-ahead predictions from the estimated ARMA model and include them as a proxy for the pseudo predictor in a multivariate

[24] We give a precise description of how $\widehat{\theta}_z$ is calculated in Appendix F, where we also consider the impact of small sample biases, and discuss sampling properties.

predictive regression of the same form as (14).[25] We then again conduct an F-test for the marginal predictive significance of z_{t-1} in this regression, and denote this test statistic RP_2

3.1.3 A one stage ARMAX-based approach with $\lambda_z \neq \lambda$ (RP_3)

A more direct approach is to estimate an equation of the same form as the standard predictive regression (1), but allowing the error term u_{zt} to follow an ARMA(1,1). Under the null that z_t is a redundant predictor, once we allow this specification of the error term then we again have $H_0 : \beta_z = 0$, since under the null the process for u_{zt} can capture the reduced form for y_t itself as in (5). The test statistic RP_3 is the likelihood ratio statistic for this restriction against the unrestricted ARMAX(1,1) representation of y_t.[26]

3.2 A link with the standard Stambaugh Bias problem

Consider the multivariate predictive regression (14). As already noted, we can consider our tests as alternative ways of producing an estimate of the univariate innovations, and hence of the true pseudo predictor, x_t^f. Suppose that we actually had data on the true pseudo predictor. Then, using (6), it follows that, by setting $\gamma_1 = \lambda - \theta$ in (14),we would have $y_t - \gamma_1 x_t^f = \varepsilon_t$, the true univariate innovation. If this were the case then (14) could be re-written as

$$\varepsilon_t = \gamma_2 z_{t-1} + \xi_t$$

where, under the null that z_t is predictively redundant, $\gamma_2 = 0$ and $\xi_t = \varepsilon_t$; but innovations to z_t are correlated with ε_t (recall that from Proposition 1, they *must* be, since otherwise z_t would not appear to have any predictive power). By inspection this equation has exactly the same form as the predictive regression (1) under the standard "Stambaugh Null" that $\beta_z = 0$. Hence in this case the only form of small sample biases we would have to worry about would be those identified by Stambaugh. Given that, as we have already noted, a redundant predictor

[25] Note that this approach automatically circumvents the problem of an initial value for the pseudo predictor (required for the RP_1 test) since maximum likelihood estimation of the ARMA model will generate an estimate of $\widehat{\varepsilon}_0$ by backcasting.

[26] Note that in principle the methodology of both the ARMA based tests, RP_2 and RP_3 could be easily extended to more general ARMA(p, q) processes.

is likely to have a high Stambaugh Correlation, we would expect that, at a minimum, our test statistics would be distorted by Stambaugh bias in small samples.[27]

However, in practice we must form an estimate of the pseudo predictor (or, equivalently, of the one-step-ahead predictions from the univariate representation). To the extent that these estimates are imperfect, we shall introduce additional sources of small sample distortions. Thus the distribution of all three test statistics will in principle depend on both the true ARMA parameters, as well as on the Stambaugh correlation and the persistence of the predictor variable. These latter two parameters are likely to be reasonably well-estimated in the data, but estimates of λ and θ (on which both RP_2 and RP_3 rely) are much more problematic, particularly when y_t is close to being white noise. For this reason we also examine the distribution of the test statistic RP_1 under the more general null, $\lambda_z \neq \lambda$, despite the fact that in this more general case it is mis-specified.

3.3 Monte Carlo Evidence

Table 1 provides some background and motivation for the Monte Carlo evidence we present for our three proposed tests. We first illustrate the difficulties the standard test procedure may encounter when the true predicted process is, an ARMA(1,1).

Panel A of Table 1 shows the implied population R-squared, R_z^2, from a least squares linear projection of the form (1) when the predictor, z_t is redundant by Definition 1. The Stambaugh correlation, ρ_z is set at 0.9. For simplicity the figures in the table assume that the AR(1) parameter of the redundant predictor is equal to that of the true predictor, ie $\lambda_z = \lambda$, since, from Proposition 1, for a given value of the Stambaugh correlation, this implies the maximum apparent predictive power for a redundant predictor. Given this assumption, (9) in Proposition 1 holds with equality, and R_z^2 is to a reasonable approximation just a scaling-down of R_f^2, the

[27] Note that Stambaugh Bias will also frequently affect estimates of β_x in the true predictive regression, if $\theta > \lambda$ (from Proposition 2b), hence may affect power. It will also be a chronic problem for ARMA estimates, since, as noted in Section 2.2, if the ARMA representation is reparameterised in terms of the pseudo predictor, the associated Stambaugh correlation is precisely unity. In the special case that $\theta = 0$ this gives the familiar downward bias in $\widehat{\beta}_f = \widehat{\lambda}$; more generally $\widehat{\beta}_f = \widehat{\lambda - \theta}$ is downward biased (which we verify in our simulations). Hence if y_t has a flat or downward-sloping variance ratio (ie is "mean-reverting"), so $\theta > \lambda$, then $\widehat{\beta}_f$ is biased away from zero.

ARMA R-squared. The table shows that the key determinant of R_z^2 is therefore how far the two key ARMA parameters, λ and θ, are from the diagonal. Along the diagonal the predicted process, y_t, is white noise, so the true values of both R_f^2 and R_z^2 are precisely zero. In the cells next to the diagonal, where λ is usually quite close to θ, there is only weak univariate predictability, and hence even weaker predictability from the redundant predictor (since its predictive power stems solely from its correlation with the ARMA predictions). However, further away from the diagonal the redundant predictor has much stronger predictive power, reflecting the (even stronger) degree of univariate predictability.

Panel B of Table 1 shows the simulated size[28] of a 1-sided t-test on the OLS estimate of β_z in (1), at a nominal 5% level. The size distortion along the diagonal corresponds precisely to the bias originally noted by Stambaugh (1999). Moving down the diagonal, as λ, and hence λ_z increases, the bias increases.[29] But the size distortion along the diagonal due to Stambaugh bias is dwarfed by the size distortion away from the diagonal, due to the correlation of the redundant predictor with the "pseudo predictor" that captures univariate predictability. Even for cells next to the diagonal, where, as the top panel of the table shows, there is a very modest degree of true predictability, a redundant predictor will nonetheless, for most values of λ and θ, appear to have significant predictive power in the majority of replications. Given the much more modest size distortion due to pure Stambaugh bias (ie when the predicted process is white noise) as shown on the diagonal, the null of no predictability is likely to be rejected frequently even after correcting for Stambaugh bias. The table therefore clearly illustrates the difficulties of inference if we do not allow for univariate predictability.

Tables 2 to 4 provide estimates of the size of our three proposed test statistics. All three tables show the rejection rate at a nominal 5% level under the null hypothesis that z_t is a redundant predictor, for the three test statistics,[30] for a range of values of λ and θ. As in Table 1 we set ρ_z, the Stambaugh Correlation of the redundant predictor, to 0.9. For the AR parameter λ_z of the redundant predictor, we consider three different variants. In Table 2 we

[28] For details of simulation methodology see Appendix G.

[29] The size distortion along the diagonal increases for lower values of T: For example, for $\lambda_z = \lambda = \theta = 0.95$ the simulated size increases to 10% and 14% for $T = 100$ and $T = 50$ respectively.

[30] For comparability between the three tests we show the size of a two-sided test. We have also carried out simulations of the size of 1-sided t-test versions of the first two tests but the results are very similar.

assume that, for any (λ, θ) pair the true model is consistent with the restricted null underpinning our first test statistic, RP_1, and hence set $\lambda_z = \lambda$. In Table 3 we assume that it is systematically lower ($\lambda_z = \frac{\lambda}{2}$), while in Table 4 we assume it is systematically higher $\left(\lambda_z = \lambda + \frac{1-\lambda}{2}\right)$.[31]

The most notable feature of these tables is that, while all three tests clearly display size distortions, these are typically much more modest, and vary much less with θ and λ than do the size distortions associated with tests on the simple predictive regression, illustrated in the bottom panel of Table 1. The comparison with the standard testing framework is also, for all three tests, most favourable in the neighbourhood of the diagonal, where y_t is close to, but not quite, white noise.

In terms of a comparison between our three proposed tests, the key features worth noting in the tables are:

- Under the restricted null that $\lambda_z = \lambda$, size distortions for the first test, RP_1 (as described in Section 3.1.1) that exploits the indirect estimate of the MA parameter, θ, in deriving an estimate of the pseudo predictor, are very modest except well away from the diagonal. To the extent that there are size distortions, these are largely due to Stambaugh bias (for reasons outlined in Section 3.2) , and thus can in principle be allowed for fairly easily. But it is also striking that even when λ_z is not equal to λ, as shown in Tables 3 and 4, and thus the test statistic is mis-specified, the size distortions remain fairly modest, and are again largely explicable by Stambaugh bias. The link with Stambaugh bias is most easily seen along the diagonal, where the size distortion is systematically larger in the top panel of Table 4 (since λ_z is systematically higher), and systematically less in the top panel of Table 3 (since λ_z is systematically lower). Away from the diagonal the size distortions become much more significant, particularly in Tables 3 and 4, but univariate predictability in such cases is likely to be so readily detectable that these are much less likely to be relevant.

- The two-stage ARMA-based test RP_2 (as described in Section 3.1.2) fairly systematically

[31] We would argue that it is reasonable to assume some relationship between λ_z and λ, given that, as noted in Proposition 1, R_z^2 is decreasing in the absolute difference between the two AR parameters - hence redundant predictors with very different AR parameters from the true predictor are much less likely to appear significant.

under-rejects the null in the neighbourhood of the diagonal. We presume that this is due to a generated regressor problem more than offsetting Stambaugh bias. In contrast the ARMAX-based test, RP_3 (as described in Section 3.1.3) systematically over-rejects the null However in both cases the size distortion is sufficiently stable across population parameters, particularly in the neighbourhood of the diagonal, that simple adjustments to critical values, or simulation of p-values based on estimates of the population parameters should allow reasonably accurate inference (we provide an illustration of the latter approach in our empirical examples).

Table 5 provides an indicator of the power of the three tests. It shows rejection rates at a nominal 5% level under the alternative hypothesis that $z_t = x_t$, the true predictor. To quantify the alternative hypothesis we need to make some assumption about where the true predictor sits within the "predictive space" summarised in Proposition 2. Part a) of the Proposition shows that x_t must always predict strictly better than the fundamental pseudo predictor (or equivalently, than the ARMA representation), but clearly the more marginal is the improvement in predictive power, the less often we will be able to reject the null that x_t is redundant. The figures in Table 5 assume that the predictive R-squared of the true predictor, R_x^2, is a fixed linear weighting of the upper and lower bounds $R_f^2(\lambda, \theta)$ and $R_n^2(\lambda, \theta)$ given in Proposition 2. We set $R_x^2 = 0.25R_f^2 + 0.75R_n^2$, such that the true predictor is distinctly closer to the fundamental pseudo predictor than it is to the non-fundamental pseudo predictor. For reference, the table also shows, in the bottom two panels, the implied value of R_x^2 and of the Stambaugh Correlation of the true predictor, ρ. The lowest panel shows that for a wide range of values of λ and θ the true predictor will have a Stambaugh Correlation quite close to unity in absolute value - hence Stambaugh bias will affect small sample estimation even of the true predictive regression, and in many cases the true predictor may superficially resemble a redundant predictor.

For most values of θ and λ all three tests correctly reject the null of redundance with high probability (note that the comparison between the tests is complicated by the differences in size distortions - the relevant values of which are shown in Table 2[32]). Only for values of $\theta > \lambda$

[32] While ideally we should calculate size-corrected power this is by no means straightforward given the degree to which true size depends on unknown parameters, as illustrated in Tables 1 to 4.

and both close to unity do rejection rates fall off.[33]

4 Empirical Applications

In Table 6 we apply our proposed test procedure to some empirical examples.

4.1 Predicting Real Annual US Stock Returns

The first four columns of Table 6 summarise results for a range of alternative predictors of real stock returns, over a long sample of annual data:[34] the dividend yield; the P/E multiple (cyclically adjusted using ten year average earnings as in Campbell & Shiller, 1988); Tobin's q; and an alternative Miller-Modigliani-consistent "cashflow" yield.[35]

In Panel A we show the two key characteristics of each of the predictors. All are strongly persistent, and most have high Stambaugh Correlations. The analysis of Sections 2.3 and 2.4 showed that high Stambaugh Correlations may be a feature of both redundant predictors and the true predictor. The conventional approach to testing does not convincingly distriminate between these two explanations. The bottom row of Panel B shows that nominal p-values for a t-test on $\widehat{\beta}_z$ in a predictive regression of the form (1), estimated by OLS, reject the "Stambaugh Null" $H_0 : \beta_z = 0$ at conventional significance levels for three out of the four indicators. Even if we correct for Stambaugh bias by simulating bootstrapped p-values under the null that returns are white noise (as shown in the bottom row of Panel D) two out of the four still appear to

[33] In such cases, the "predictive space" defined by Proposition 2 contracts: both upper and lower bounds for R_x^2 approach zero, and, as shown in the bottom panel of Table 5, ρ, the true Stambaugh Correlation, tends to unity. Thus in these limiting cases the true predictor becomes harder to distinguish from the fundamental pseudo predictor or from redundant predictors. We discuss this limiting case in more detail in Robertson & Wright, 2009.

[34] All data are taken from the dataset described in Wright (2004), updated to end-2007. P/E and dividend yield data (both from spliced Cowles/S&P 500 data, as in Shiller, 2000) are available on the same basis from 1871 onwards, but for comparability we align the samples for these predictors with those for the other two, which are only available from 1900 onwards.

[35] The literature on the dividend yield is massive. See Campbell, Lo and Mackinlay, (1995) for a survey of the early literature; Goyal & Welch (2003) as an example of the recent critique, and Cochrane (2008); Campbell & Thompson (2007) for responses. The use of a cyclically adjusted P/E dates at least as far back as Graham & Dodd (1934) but more recently was revived by Campbell & Shiller (1998) and Carroll (2008) as a tool for long-horizon forecasting. See also Lamont (1998) on the unadjusted P/E. On Tobin's q (and the closely related book-to-market ratio), see Smithers & Wright, 2000; Robertson & Wright, 1998; Vuolteenaho, 1999. On the "cashflow" yield, see Robertson & Wright, 2006; Boudhoukh et al, 2007.

have quite strongly significant predictive power.[36]

In Panel B we show nominal p-values for the three test statistics that we have proposed as a means of weeding out redundant predictors. For three out of the four predictors, the dividend yield, the P/E and q, we cannot reject the null that all three are redundant predictors of stock returns, once we condition on the history of returns. This result can be read off straightforwardly from the nominal p-values for all three of our proposed tests, shown in Panel B; this conclusion is unaffected when bootstrapped (Panels C and D) p-values are used. Furthermore we do not reject the restricted version of the null, such that $\lambda_z = \lambda$, so we can rely simply on our first test, RP_1, that uses an indirect estimate of the pseudo predictor. We thus cannot reject the null that the apparent predictive power of all three indicators simply reflects their correlation with the pseudo predictor, as evidenced by their high Stambaugh Correlations.

In the case of the dividend yield, we would have arrived at the same conclusion simply by looking at the predictive regression, since even the nominal p-value on β_z (shown in Table 6, Panel B) suggests insignificance.[37]

It is noteworthy that the very high Stambaugh correlation of 0.98 for the cyclically adjusted P/E shows a striking similarity with the result we proposed in Corollary 1, suggesting that the the cyclical adjustment of earnings is required simply to boost the Stambaugh Correlation, and hence the apparent significance of this predictor. Results with unsmoothed earnings lower the Stambaugh correlation significantly, but simultaneously greatly weaken the apparent predictive power of this predictor. Essentially, the P/E multiple can only be turned into anything resembling a useful predictor of stock returns by eliminating any of its independent informational content.

Results for the fourth indicator, the cashflow yield, are less clear-cut. There is a strong rejection of redundance on our first test, RP_1. However, given the restricted nature of the null underlying this test, the rejection could in principle reflect a standard joint null problem, arising

[36] Bootstrapping methodology is described in Appendix G.

[37] Cochrane (2008) argues strongly that this negative result cannot be viewed in isolation, and that return predictability should be inferred from the joint properties of a system that exploits the present value relation beween the dividend yield, returns and future dividend growth. Since the dividend yield does not predict dividend growth, he argues that it must predict returns. Our results do not conflict with this conclusion. We simply argue that, in predicting returns, the dividend yield is proxying the pseudo predictor.

from the restriction that $\lambda_z = \lambda$. We therefore need also to look at our two remaining tests, RP_2 and RP_3, that do not impose this additional restriction. Our simulation results showed that size distortions were larger for these tests, hence we focus on bootstrapped p-values: for both tests these indicate more marginal rejections. A further caveat arises from our discussion of data mining in Section 2.3. To the extent that a predictor variable is chosen on the basis of horse races between predictive regressions, then this prior filtering of the data means that there can be significant distortions to p-values (cf Ferson, Simin and Sarkissian, 2003; Sullivan, Timmerman and White, 1999). Arguably therefore the results for this predictor should be regarded as more marginal even than suggested by the bootstrapped p-values shown in Panel C.[38]

Figure 2 provides some insight into the results. It shows each of the four predictor variables over the common sample period from 1900 to 2007, along with estimates of the fundamental pseudo predictor, constructed using the formula in (7) from the history of returns,[39] and the indirect estimate of $\widehat{\theta}_z$ derived from the properties of each of the predictor variables as described in Section 3.1.1. Unsurprisingly the estimated pseudo predictors are very similar in all four panels, since all are long moving averages of the same return process, and the estimates of θ are all quite similar (and all quite close to unity).

The charts show that for both the price-earnings ratio and Tobin's q the correlation with the pseudo predictor is very strong (around 0.85 for both predictors). This is consistent with the analysis of Figure 1 in Section 2.5 in which we noted that for a redundant predictor this correlation will approximately equal the Stambaugh Correlation.[40] For the dividend yield the

[38] We are thus undermining our own claim, in Robertson & Wright (2006), that the cashflow yield is a significant predictor of stock returns. The basis for this claim was essentially the p-value of 0.002 on β_z under the "Stambaugh Null" that y_t is white noise, as shown in the bottom line of Panel D of Table 6. We showed in our earlier paper that this rejection of the Stambaugh Null was robust to sample changes and to a range of different simulation techniques for p-values. The evidence of the tests in Table 6 makes these results look distinctly more marginal. However, in defence of our past selves, the cashflow yield is at least to some extent proof against the data mining critique since it has a stronger basis in economic theory than the conventional dividend yield, and thus was picked as a predictor for this reason, rather than on the basis of its predictive power.

[39] Initial values are set to the unconditional mean, however, since we have almost a full century of data on returns before the start of the sample used in the charts, the impact of the initial value is trivially small.

[40] For both predictors Table 6 shows that the latter is distinctly higher, but this may be due to sampling differences in estimating the correlation with the pseudo predictor. For two strongly serially correlated processes a given sample provides much less information on the true correlation than for two white noise processes.

correlation is lower (0.68) but Table 6 shows that so is the Stambaugh Correlation and the degree of apparent predictive power: thus it seems reasonable to conclude that the dividend yield is simply a poorer proxy of the pseudo predictor than are the P/E or q.

Figure 2. Predictors and Pseudo Predictors for US Stock Returns

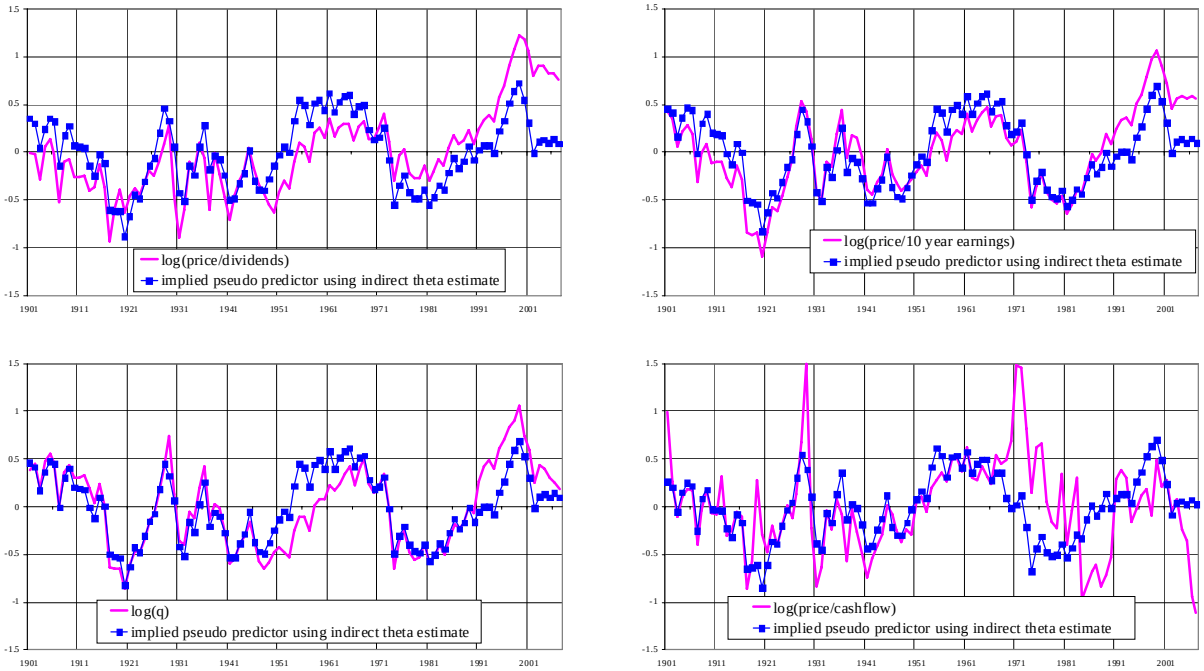


Figure 2 also shows that the correlation between the price-dividend ratio and the pseudo predictor is much higher if the sample is truncated to end in the mid 1980s. In this sample the apparent statistical significance of the dividend yield is also distinctly greater - consistent with the conclusion of Goyal and Welch (*op cit*) that the evidence of predictability from the dividend yield is an artefact of sample selection. Our results suggest the interpretation that this simply reflects the fact that during this sample the dividend yield was a better proxy for the pseudo predictor.[41]

In contrast Figure 2 shows that the cashflow yield has a much lower correlation with the pseudo predictor (0.46) despite (as Table 6 showed) having stronger predictive power for returns. This is again broadly consistent with our discussion of Figure 1, in which we noted that the true predictor may have a quite low correlation with the pseudo predictor. The marginal nature of

[41] If we carry out our three tests over these truncated samples, we still fail to reject the null that it is a redundant predictor.

the results for this variable in Table 6 mean that we certainly cannot be sure that the cashflow yield is the true predictor, but it does at least suggest that it may be a reasonable proxy for it.

It should be stressed that the evidence presented in both Table 6 and Figure 2 does not *rely* on the assumption that stock returns have a significant degree of univariate predictability. Simulated p-values are also shown in Table 6, Panel D, under the null that returns are white noise. For our three proposed test statistics the associated p-values are typically quite similar, and the conclusions to be drawn from them are unaltered. This is in marked contrast with simulated p-values for $\widehat{\beta}_z$, which quite strongly reject the white noise null for one predictor (q) and marginally so for another (the P/E). The reconciliation of these two results is straightforward: these two predictors do have a degree of predictive power that we would be very unlikely to observe if returns were pure white noise. But our results suggest that this is simply because they are proxies for the pseudo predictor which captures univariate predictability. For these two indicators (and all the more so for the dividend yield) the evidence of predictability of real stock returns (such as it is) is thus almost entirely univariate in nature.[42]

4.2 Predicting quarterly GNP growth

In our second example, shown in the final column of Table 6, we examine a somewhat simplified version of Cochrane's (1994) predictive equation for quarterly GNP growth using the ratio of consumption to GNP as a predictor variable. Cochrane used this predictive equation (in the context of a vector autoregression) to draw the conclusion that consumption provides a good estimate of the permanent component of GNP. This example does not fit so readily into our ARMA(1,1) framework, hence the formal statistical tests can only be viewed as illustrative. Nonetheless the exercise yields some interesting insights.

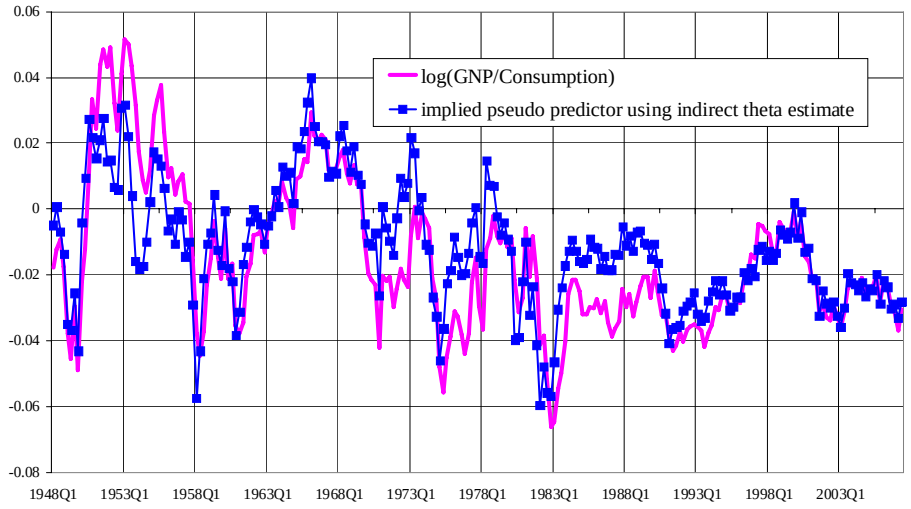
Cochrane's original regression equation included two lags of both consumption growth and GNP growth as additional regressors. We can allow for the terms in lagged GNP growth by by modifying the dependent variable to be in quasi-differenced form, but we can only match our very simple predictive regression framework by discarding the terms in lagged consumption

[42]It is noteworthy that pseudo predictors actually predict better, in-sample, than both the P/E and the dividend yield.

growth.[43] Thus our restricted predictive regression specification fits less well than Cochrane's (both updated to include all available data). But the significance of the lagged cointegrating term, the log ratio of GNP to consumption, is very similar, as is the actual size of the coefficient.

Within this rather restrictive framework we find some fairly clear-cut conclusions. First, the GNP/consumption ratio is emphatically *not* redundant: it does contain significant predictive information, independent of the history of GNP growth, with strong rejections of the null on all three of our tests.

Figure 3: The Consumption/GDP Ratio and its Associated Pseudo Predictor



However, perhaps the most interesting feature of this second set of results is arguably to be found if we again examine the properties of the resulting pseudo predictor (derived solely from the history of GNP growth). While the tests in Table 6 suggest strongly that there is independent predictive value in the GNP/consumption ratio, Figure 3 shows that the predictor and its associated pseudo predictor are very similar indeed: ie, the "great ratio" is very similar to a simple long weighted moving average of GNP growth. We would be surprised if any

[43] To be precise, we estimate an equation of the form

$$\Delta \ln GNP_t = \widehat{\alpha}_1 \Delta \ln GNP_{t-1} + \widehat{\alpha}_2 \Delta \ln GNP_{t-2} - \widehat{\beta} x_{t-1} + \widehat{u}_t$$

where $x_t = \log(GNP_t/CONS_t)$ and then define

$$y_t = \Delta \ln GNP_t - \widehat{\alpha}_1 \Delta \ln GNP_{t-1} - \widehat{\alpha}_2 \Delta \ln GNP_{t-2}$$

Note that our framework also over-simplifies the dynamics of x_t, the cointegrating relation.

applied macroeconomist, however familiar with the data, would be able to say which series was which without the aid of the legend. Thus although the cointegrating framework does provide statistically significant predictive power, in quantitative terms this improvement is quite marginal. Univariate properties appear to be the dominant element in predictability of GNP growth.[44]

5 Conclusions

We have examined a very simple predictive model which is widely used in empirical finance, and which also captures key features of cointegrating systems. When candidate predictor variables have high "Stambaugh Correlations" it is well-known (in empirical finance at least) that Stambaugh bias matters and may lead to over-rejection of the null that it is a redundant predictor (which is in turn a simple test of Granger Causality[45]). But our results show that, wherever Stambaugh Correlations are high, Stambaugh bias actually matters much less than taking proper account of univariate predictability (which in turn may well be a key explanation of *why* the Stambaugh Correlation is high). A key feature of our proposed test procedure, in contrast to conventional tests of Granger Causality, is that the predicted process has the same ARMA representation under both the null and alternative hypotheses.

Our results do not, however, *depend* on there being significant ARMA features: p-values for our tests are relatively invariant to the true ARMA parameters in the neighbourhood of the white noise case, in contrast to those for a standard predictive regression which are highly sensitive to ARMA parameters in this neighbourhood.

Our results have strong implications for the literature on predictability of returns. Our redundant predictors provide a pretty consistent picture that a model of returns that is nearly

[44] We have noted that our test procedure in this last example is constrained by our assumption of an ARMA(1,1) representation. But we have also noted that our RW_2 and RW_3 tests are in principle easy to generalise. As an alternative to the constrained test procedure reported in Table 6, we have also carried out equivalent versions of RW_1 and RW_2 on the assumption of a higher order ARMA process for $\Delta \ln GNP$ itself (up to third order in both AR and MA components). These yield very similar results: namely a strong rejection of predictor redundance for the consumption/GNP ratio, but coupled with a distinctly modest increase in R^2 over the univariate representation. For these generalised tests we have however had to rely on the assumption that the true size of the tests is reasonably well captured by their nominal size.

[45] Although the issue seems to have been neglected in more general applications of Granger Causality testing in cointegrating systems.

but not quite white noise matches long sample properties. This representation has a slowly declining variance ratio which may well imply strong long horizon predictability. But our results suggest strongly that any such long-horizon predictability, if it does exist, is entirely univariate in nature.[46] The dividend yield, the P/E and Tobin's q are at best imperfect proxies for the pseudo predictor associated with this representation. After taking account of univariate properties, the only predictor of returns with even marginal significance is the cashflow yield. This is less persistent than the other indicators but this does not necessarily contradict the above representation, since it might just be an imperfect, low persistence and noisy proxy for the "true" predictor.

While our applications primarily focussed on stock returns,they are potentially of considerably wider applicability. All of our return predictors can be (and have been in the literature) viewed as cointegrating relations. The data show that most, or possibly all of these predictors, can be treated as redundant - ie that multivariate results are actually just proxying univariate properties. Our results in this paper on the consumption/GNP ratio suggest that, while it does have statistically significant predictive power for output growth, it is pretty marginal in quantitative terms relative to a univariate benchmark. We suspect that other evidence derived from Granger Causality tests in a cointegrating framework should be revisited in the same spirit.

[46]Note that the evidence for a declining variance ratio for stock returns originally found over long historical samples by Cochrane (1994) has been disputed by Kim, Nelson and Startz (1991) on the basis of more recent data. The power of direct evidence on the variance ratio over such short samples is however known to be quite low: their postwar results would also be consistent with a high (λ, θ), declining variance ratio ARMA(1,1).

Appendix

Note: for convenience of referees we include fairly lengthy derivations and proofs. We would envisage that the published version of the appendix could be significantly shorter.

A Redundant Predictors

A.1 The general innovation specification

Assume that z_t is a redundant predictor by Definition 1. By an innocuous re-scaling of z_t, let the innovation to the AR(1) process in (2) be given by

$$\begin{aligned} v_{zt} &= \gamma_\varepsilon \varepsilon_t + \omega_t \\ &= \rho_{\varepsilon v_z}\varepsilon_t + \sigma_\varepsilon \sqrt{1-\rho^2_{\varepsilon v_z}}\, q_t \end{aligned} \tag{16}$$

where $Eq_t = 0, Eq_t^2 = 1$. This gives the useful normalisation

$$\begin{aligned} \sigma_{v_z} &= \sigma_\varepsilon; \\ corr\left(v_{zt}, \varepsilon_t\right) &= \rho_{\varepsilon v_z} \end{aligned}$$

Note that, while the AR(1) property of z_t requires v_{zt} to be white noise this need not necessarily be the case for ω_t. We consider this more general case below, but first consider the benchmark case where ω_t is white noise, and is also orthogonal to all lags of ε_t, as assumed in Proposition 1.

A.2 Proof of Proposition 1: The predictive R-squared if ω_t is white noise orthogonal to all lags of ε_t.

Given the normalisation of the innovation variance we have, for the general case,

$$\beta_z = \frac{cov\left(y_{t+1}, z_t\right)}{var\left(z_t\right)} = \frac{cov\left(y_{t+1}, z_t\right)}{\sigma_\varepsilon^2/\left(1-\lambda_z^2\right)}$$

and, given the orthogonality assumption we have, using (5),

$$\begin{aligned} cov\left(y_{t+1}, z_t\right) &= cov\left(\varepsilon_{t+1} - (\theta-\lambda)\frac{\varepsilon_t}{1-\lambda L}, \frac{v_{zt}}{1-\lambda_z L}\right) \\ &= -(\theta-\lambda)\rho_{\varepsilon v_z} cov\left(\frac{\varepsilon_t}{1-\lambda L}, \frac{\varepsilon_t}{1-\lambda_z L}\right) \\ &= -(\theta-\lambda)\rho_{\varepsilon v_z} cov\left[\left(1+\lambda L+\lambda^2 L^2+...\right)\varepsilon_t, \left(1+\lambda_z L+\lambda_z^2 L^2+...\right)\varepsilon_t\right] \\ &= \frac{-(\theta-\lambda)\rho_{\varepsilon v_z}\sigma_\varepsilon^2}{1-\lambda\lambda_z} \end{aligned}$$

hence we have, for the general case

$$\beta_z = -(\theta-\lambda)\rho_{\varepsilon v_z}\left(\frac{1-\lambda_z^2}{1-\lambda\lambda_z}\right)$$

and hence

$$R_z^2 = \frac{\beta_z^2\sigma_z^2}{\sigma_y^2} = (\theta-\lambda)^2\rho_{\varepsilon v_z}^2\left(\frac{1-\lambda_z^2}{1-\lambda\lambda_z}\right)^2\frac{\sigma_\varepsilon^2}{\left(1-\lambda_z^2\right)}\frac{1}{\sigma_y^2} \tag{17}$$

but we have, using the formulae in Proposition 2,

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

hence we have

$$\frac{\sigma_\varepsilon^2}{\sigma_y^2} = 1 - R_f^2 \text{ and } (\theta-\lambda)^2\left(1-R_f^2\right) = R_f^2\left(1-\lambda^2\right)$$

so that substituting into (17) we can write

$$R_z^2 = \rho_{\varepsilon v_z}^2 R_f^2 g(\lambda, \lambda_z) \tag{18}$$

where

$$g(\lambda, \lambda_z) = \frac{\left(1-\lambda^2\right)\left(1-\lambda_z^2\right)}{\left(1-\lambda\lambda_z\right)^2} \tag{19}$$

The expression in (18) is defined in terms of $\rho_{\varepsilon v_z} = corr(\varepsilon_t, v_{zt})$. To show the link with the Stambaugh Correlation, note that we have

$$u_{zt} = y_t - \beta_z z_{t-1} = \left(\frac{1-\theta L}{1-\lambda L}\right)\varepsilon_t - \beta_z \left(\frac{\rho_{\varepsilon v_z}\varepsilon_{t-1} + \sigma_\varepsilon\sqrt{1-\rho_{\varepsilon v_z}^2}q_{t-1}}{1-\lambda_z L}\right)$$

hence

$$\rho_z \equiv corr(u_{zt}, v_{zt}) = \frac{\rho_{\varepsilon v_z}\sigma_\varepsilon^2}{\sigma_{u_z}\sigma_\varepsilon} = \rho_{\varepsilon v_z}\frac{\sigma_\varepsilon}{\sigma_{u_z}} = \rho_{\varepsilon v_z}\sqrt{\frac{1-R_f^2}{1-R_z^2}} \tag{20}$$

which allows us to substitute into (18), giving

$$R_z^2 = \left(\frac{1-R_z^2}{1-R_f^2}\right)\rho_z^2 R_f^2 g(\lambda, \lambda_z) \leq \left(\frac{1-R_z^2}{1-R_f^2}\right)\rho_z^2 R_f^2 \tag{21}$$

since $g(.)$ has a maximum value of unity at $\lambda_z = \lambda$. Equivalently, as in the Proposition

$$\frac{R_z^2}{1-R_z^2} \leq \rho_z^2\left(\frac{R_f^2}{1-R_f^2}\right) \tag{22}$$

which in turn implies an equivalent upper bound on R_z^2 itself since $f(x) = x/(1-x)$ is a strictly increasing function. By inspection of (20) ρ_z lies within $[-1, 1]$, hence for $\rho_z = 1$ we have $R_z^2 = R_f^2$.■

A.3 More general processes for ω_t

Consider the more general processes for v_{zt} and ω_t

$$v_{zt} = \gamma_0\varepsilon_t + \omega_t; \quad E(v_{zt}\omega_t) = 0$$

$$\omega_t = \sum_{i=1}^{\infty} \gamma_i \varepsilon_{t-i} + \xi_t; \quad E\left(\xi_t \varepsilon_{t-i}\right) = 0 \ \forall \ i \geq 0$$

$$\xi_t = \sum_{i=0}^{\infty} \pi_i s_{t-i}$$

where s_t is white noise hence we require

$$\begin{aligned} E\left(\xi_t \varepsilon_{t-i}\right) &= \pi_i E\left(s_{t-i}\varepsilon_{t-i}\right) = 0 \ \forall \ i > 0 \\ &\Rightarrow \quad E\left(s_{t-i}\varepsilon_{t-i}\right) = 0 \ \forall \ i \end{aligned}$$

hence if v_{zt} is white noise, which we require for AR(1)ness of z_t, we require

$$\begin{aligned} E\left(v_{zt}v_{zt-k}\right) &= E\left[\left(\sum_{i=0}^{\infty}\gamma_i\varepsilon_{t-i} + \sum_{i=0}^{\infty}\pi_i s_{t-i}\right)\left(\sum_{i=0}^{\infty}\gamma_i\varepsilon_{t-k-i} + \sum_{i=0}^{\infty}\pi_i s_{t-k-i}\right)\right] \\ &= \sigma_\varepsilon^2\left[\gamma_k\gamma_0 + \gamma_{k+1}\gamma_1 + ...\right] + \sigma_s^2\left[\pi_k\pi_0 + \pi_{k+1}\pi_1 + ...\right] \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty}\gamma_{k+i}\gamma_i + \sigma_s^2\sum_{i=0}^{\infty}\pi_{k+i}\pi_i = 0 \forall k > 0 \end{aligned} \tag{23}$$

It is evident that if we set $\gamma_i = 0 \forall i > 0$ then if we also set $\pi_i = 0 \forall i > 0$ then the condition is satisfied for all k. This is the benchmark case analysed in Proposition 1, where ω_t is white noise uncorrelated with all lags of ε_t.

We next consider two cases that satisfy the autocovariance condition in (23) for more general processes

A.3.1 Special case: ω_t is white noise but $E\left(\omega_t\varepsilon_{t-j}\right) \neq$ for some $j > 0$

For this condition to hold we may allow $\gamma_j \neq 0$ for some $j > 0$. But by inspection of (23) this requires $\gamma_i = 0$ for $i \neq j$ (which in turn implies $\gamma_0 = 0$ - hence a zero contemporaneous correlation of v_{zt} and ε_t) and an equivalent restriction on the π_i (which will be satisfied if, eg $\pi_i = 0 \ \forall i > 0$). Then we can write, subject to a normalisation

$$v_{zt} = \gamma_j \varepsilon_{t-j} + \omega_t$$

which is a white noise process, as is $\omega_t = s_t$. From this specification it follows that z_t will be a scaling of x^f_{t-j}, plus an AR(1) error. But it also follows that we must have

$$R^2_z \leq \lambda^{2j} R^2_f$$

(where the upper bound is attained when $\gamma_j = 1$, $s_t = 0$), thus the higher is j the lower the upper bound on the predictive R-squared.

A.3.2 A more general case: ω_t not white noise

In this more general case, while ω_t may not be white noise, the autocovariance condition (23) puts a very tight restriction on the nature of the two underlying polynomials $\gamma(L)$ and $\pi(L)$, such that v_{zt} is white noise. Any non-zero γ_i put corresponding restrictions on the π_i, which in turn increases the noise element in v_{zt}, which in turn must lower the predictive R-squared. Whilst we have as yet not been able to establish any general implications of such a process, we suspect that most such processes will as a result have low R-squareds. Some processes are also entirely ruled out (eg $\gamma(L)$ and $\pi(L)$ cannot both be finite order ARs).

B The ARMA(1,1) Reduced Form

B.1 Normalisations

In what follows we assume that by an appropriate scaling of the data for the true predictor, x_t, we can ensure $\beta_x > 0$. We also suppress x-subscripts so that all parameters without a subscript are assumed to relate to x.

To ensure that both fundamental and non-fundamental pseudo predictors satisfy this sign convention, in what follows we also re-define each by setting

$$\begin{aligned} x^f_t &= sign(\theta - \lambda) \frac{\varepsilon_t}{1 - \lambda L} \\ \beta &= |\theta - \lambda| \Rightarrow \rho_f = sign(\theta - \lambda) \end{aligned} \tag{24}$$

and

$$\begin{aligned} x_t^n &= sign\left(\theta^{-1}-\lambda\right)\frac{\eta_t}{1-\lambda L} \\ \beta &= \left|\theta^{-1}-\lambda\right| \Rightarrow \rho_n = sign\left(\theta^{-1}-\lambda\right) \end{aligned} \tag{25}$$

where η_t is the non-fundamental innovation.

B.2 Derivation of reduced form

Substituting from (4) into (3) we have

$$y_t = \frac{\xi_t}{(1-\lambda L)} \tag{26}$$

where

$$\xi_t = -\beta v_{t-1} + (1-\lambda L)u_t \tag{27}$$

which is a moving average error defined in terms of the two underlying innovations. We can represent this in terms of a composite univariate innovation, ε_t, which satisfies

$$\xi_t = (1-\theta L)\,\varepsilon_t$$

where θ satisfies the moment condition

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

where $\rho = \sigma_{vu}/(\sigma_u\sigma_v)$, $s = \sigma_v/\sigma_u$, and R_x^2 is the R-squared from the predictive regression (3). Note also that

$$R_x^2 = \frac{\beta^2\sigma_x^2}{\sigma_y^2} = \frac{\beta^2 s^2}{\beta^2 s^2 + 1 - \lambda^2}$$

implying

$$\beta^2 s^2 = F^2$$

$$\text{where } F(R_x^2, \lambda) = \sqrt{\left(1-\lambda^2\right)\frac{R_x^2}{1-R_x^2}} \tag{28}$$

hence the moment condition defining θ can be written as

$$\frac{\theta}{1+\theta^2} = \kappa$$

$$\text{where } \kappa(\lambda, \rho, R_x^2) = \frac{\lambda + \rho F(R_x^2, \lambda)}{1 + \lambda^2 + F(R_x^2, \lambda)^2 + 2\lambda\rho F(R_x^2, \lambda)} \tag{29}$$

which solves to give the MA parameter in the fundamental representation[47]

$$\theta\left(\lambda, \rho, R_x^2\right) = \frac{1 - (1 - 4\kappa\left(\lambda, \rho, R_x^2\right)^2)^{\frac{1}{2}}}{2\kappa\left(\lambda, \rho, R_x^2\right)} \tag{30}$$

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

$$\partial\kappa/\partial\rho = \frac{F\left(1 + F^2 - \lambda^2\right)}{\left(1 + \lambda^2 + F^2 + 2\lambda\rho F\right)^2} > 0 \tag{31}$$

for $\lambda \in [0, 1), R_x^2 \in (0, 1)$. Thus, since $\rho \in (-1, 1)$, we know that

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

where

$$g(x) = \frac{x}{1+x^2} \in \left(-\frac{1}{2}, \frac{1}{2}\right) \tag{32}$$

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

$$\frac{\partial\theta}{\partial\kappa} = \frac{1}{2}\left(\frac{1 - \sqrt{1 - 4\kappa^2}}{\sqrt{1 - 4\kappa^2}\kappa^2}\right) \geq 0 \tag{33}$$

[47] The other solution to (29) gives the nonfundamental representation.

which in turn gives us

$$\frac{\partial \theta}{\partial \rho} = \frac{\partial \theta}{\partial \kappa}\frac{\partial \kappa}{\partial \rho} \geq 0 \tag{34}$$

which we shall exploit in the proof of Proposition 2, part a).

C Proof of Proposition 2, part a)

We wish to establish the inequality

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

C.0.1 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 B we have

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

hence

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

since $cov(u_t, \psi_{t-1}) = 0$. Hence for $1 < \rho < 1$ we always have

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

For the fundamental pseudo predictor case, if $\theta > \lambda$ we have $u_t = \varepsilon_t = v_t$ (hence $\rho = 1$) so $\psi_{t-1} = (\theta - \lambda - \beta)\varepsilon_{t-1} = 0$ and if $\theta < \lambda$ we have $u_t = \varepsilon_t = -v_t$ (hence $\rho = -1$), so $\psi_{t-1} = (\theta - \lambda + \beta)\varepsilon_{t-1} = 0$. Hence for the limiting case of the fundamental pseudo predictor

we always have $\sigma_u^2 = \sigma_\varepsilon^2 \Rightarrow R_f^2 = R_x^2$ so for the general case we have

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

We can also use the Yule-Walker equations to derive

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

hence

$$R_f^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2} = \frac{(\theta - \lambda)^2}{1 - \lambda^2 + (\theta - \lambda)^2} \tag{35}$$

C.0.2 Relation of R_x^2 to R_n^2

We have the non-fundamental representation

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

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

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

which, after substituting from (35) gives

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

which can also be derived directly from (35) by substituting θ^{-1} for θ, as in the Proposition.

We wish to establish the weak inequality

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

While H 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 hold 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 $H = H(\theta(\rho))$. From (34) we also have $\partial\theta/\partial\rho > 0$, hence we can write

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

and we have

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

There are three cases:

Case 1: $\theta_{\min} > 0$; For this case,we have $H(\theta_{\max}) = 0$, since at this point x is the non-fundamental pseudo predictor in (25) with $R_x^2 = R_n^2$, $\rho = \rho_n = 1$. From (38) we also have $H' < 0$ hence $H \geq 0$.

Case 2: $\theta_{\max} < 0$; For this case,we have $H(\theta_{\min}) = 0$, since at this point x is again the non-fundamental pseudo predictor in (25) with $R_x^2 = R_n^2$, but with $\rho = \rho_n = -1$. From (38) we have $H' > 0$ hence $H \geq 0$.

Case 3: $\theta_{\min} < 0 < \theta_{\max}$; For this case,we have $H(\theta_{\min}) = H(\theta_{\max}) = 0$, $H'(\theta_{\min}) > 0$; $H'(\theta_{\max}) < 0$, and, from (38) H has a single turning point at $\theta = 0$,[48] hence again we have $H \geq 0$.

[48]Not that for the case $\theta\left(\lambda, \rho, R_x^2\right) = 0$ the non-fundamental representation (36) is undefined. At the limit the non-fundamental pseudo-predictor has zero variance but if for the general case we substitute for y_t in (25) we can solve forward to derive

$$x_t^n = -sign\left(\theta^{-1} - \lambda\right)\frac{\theta}{1-\theta L^{-1}}y_{t+1} \tag{39}$$

which we can expand to give

$$E_{t-1}y_t|x_{t-1}^n = -\left(\theta^{-1} - \lambda\right)\left[-\theta y_t - \theta^2 y_{t+1} + ...\right]$$

which gives $\lim_{\theta\to 0} E_{t-1}y_t|x_{t-1}^n = y_t$, hence $\lim_{\theta\to 0} R_n^2 = 1$.

Since we have shown that these results hold for any θ and λ, by implication they also hold for any λ and R_x^2.

We have thus established the right-hand inequality in (11) for all three cases thus completing the proof of part a) of Proposition 2.■

D Proof of Proposition 2, part b)

Any given values of θ and λ must imply a condition on κ (as defined in (29)) 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

$$\rho(R_x^2) = \frac{(\theta-\lambda)(1-\theta\lambda) + F(\lambda, R_x^2)^2\theta}{\left(1-\lambda^2+(\theta-\lambda)^2\right)F(\lambda, R_x^2)}; \quad \rho \in (-1,1) \tag{40}$$

where $F(\lambda, R_x^2)$, as defined in (28). If the solved value for ρ lies outside this range, the triplet (θ, λ, R_x^2) is not feasible.

The first order condition yields a unique stationary point:

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

which after substituting into (40) yields a real solution if

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

which is satisfied for $\theta > \lambda$, given $\lambda > 0$. The second-order condition confirms that for this range of parameter values this yields the minimum value

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

E The correlation between the true predictor and the fundamental pseudo predictor.

Consider the regression

$$y_t = \gamma_x x_{t-1} + \gamma_f x_{t-1}^f + w_t$$

where x_t^f is the fundamental pseudo predictor and x_t the true predictor. Then it must be that $\gamma_f = 0$; $\gamma_x = \beta_x$ (from (3)). Treating x_{t-1} as an omitted variable, if we estimate a predictive regression in terms of the pseudo predictor as define in (6), then using the formula for omitted variable bias we have

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

which implies

$$\beta_f \frac{\sigma_{x^f}}{\sigma_y} = \beta_x \text{corr}\left(x_t^f, x_t\right) \frac{\sigma_x}{\sigma_y}$$

which, after squaring both sides gives

$$R_f^2 = \text{corr}\left(x_t^f, x_t\right)^2 R_x^2$$

which in turn gives (13) in the main text. Note that the squared correlation in (13) is bounded below by the ratio R_f^2/R_n^2 since, from part a) of Proposition 2, R_x^2 is bounded above by R_n^2, the predictive R-squared of the non-fundamental pseudo predictor.

F Properties of $\widehat{\theta}_z$, used in the indirect pseudo predictor based test, RP_1

F.1 Derivation

To derive $\widehat{\theta}_z$ as defined in Section 3.1.1 we need estimates of λ and R_f^2. The first is easy given the additional restriction $\lambda_z = \lambda$. It also allows us to write (21), using $g(\lambda, \lambda_z) = 1$, as

$$f\left(R_z^2\right) = \rho_z^2 f\left(R_f^2\right) \tag{41}$$

where $f(x) = x/(1-x)$, so for a given Stambaugh correlation, inverting (41), we have

$$R_f^2 = \frac{f\left(R_z^2\right)}{\rho_z^2 + f\left(R_z^2\right)} \tag{42}$$

Hence we can write

$$\theta\left(\lambda, 1, R_f^2\right) = \theta\left(\lambda_z, 1, \frac{f\left(R_z^2\right)}{\rho_z^2 + f\left(R_z^2\right)}\right) = \theta_z\left(\lambda_z, \rho_z, R_z^2\right). \tag{43}$$

Under the restricted null the final expression defines a functional relationship that holds exactly in terms of population parameters. For purposes of estimation we define $\widehat{\theta}_z = \theta_z\left(\widehat{\lambda}_z, \widehat{\rho}_z, \widehat{R}_z^2\right)$.

F.2 Sampling Properties

Table A1 provides a comparison of the sampling properties of $\widehat{\theta}_z$ and the ARMA estimate of θ, under the joint null that z_t is predictively redundant and $\lambda_z = \lambda$, as for our proposed test statistic RP_1.

The first two panels show the sampling properties of the ARMA estimate for a range of values of the true population parameters, λ and θ, on the assumption (used in previous tables) that the Stambaugh Correlation $\rho_z = 0.9$. The top panel shows that the ARMA estimate displays non-trivial bias for virtually all population values of λ and θ, with severe bias near the diagonal (the white noise case), the second panel shows that there is an equivalent increase

in dispersion. We have in fact arguably somewhat *under*stated the problems with the ARMA estimates, since for each replication the estimation is actually carried out twice: once without starting values; and once using starting values (for convenience given by $\widehat{\lambda}_z$ and $\widehat{\theta}_z$) to reflect the prior that the true values are both positive. The program chooses the estimate with the highest value of the estimated log likelihood. Both the bias and the wide dispersion in part reflect the fact that, even exploiting these starting values, a high proportion of estimated values of θ are negative.

The lower two panels of Table A1 provide equivalent simulation evidence for our indirect estimate, $\widehat{\theta}_z$. Both bias and dispersion are dramatically lower. Given the nonlinearity of the expression derived in the previous section it is perhaps surprising that there is so little bias. However, while it is well known that OLS estimates of λ are downward-biased in small samples, at the same time R_z^2 is upward biased (due to Stambaugh Bias) For most values of the true parameters these two biases appear to offset. In principle a more sophisticated attempt at bias correction could be applied.

The low dispersion of the indirect estimate $\widehat{\theta}_z$ means in turn that the resulting estimated pseudo predictor used in deriving the test statistic RP_1 is, under the null, very close to being the true pseudo predictor, which helps to explain why size distortions for this test statistic are so low.

Lest we appear to be getting the econometric equivalent of something for nothing, it should be stressed that this indirect method of estimating the MA parameter from the properties of the redundant variable hinges crucially upon the assumption that $\lambda_z = \lambda$. Hence if the joint null underlying RP_1 is rejected this may be because z_t is not redundant, but it could also be because the assumption that $\lambda_z = \lambda$ is incorrect (which, for sufficient differences between λ_z and λ will imply that the indirect estimate $\widehat{\theta}_z$ may be severely biased).

G Simulation Methodology

G.1 Monte Carlos

The input parameters for each simulation are λ, θ, ρ_z and λ_z and a weight, μ, such that $R_x^2 = (1-\mu) R_f^2(\lambda,\theta) + \mu R_n^2(\lambda,\theta)$, where R_f^2 and R_n^2 are as defined in Proposition 2. By inverting the formula for $\theta(\lambda, \rho, R_x^2)$ in (10) this yields a value of ρ, the true Stambaugh Correlation. We then simulate underlying joint normal white noise innovations u_{xt} v_{xt} and ω_t (in equations (3), (4) and (8) respectively with the appropriate correlations. This in turn generates processes for x_t, y_t ε_t (using $\varepsilon_t = (1-\theta L)^{-1}[(1-\lambda L) u_t - \beta v_{t-1}]$) and z_t. Note that for Tables 1 to 4 and A1 we could equally well simulate ε_t and generate y_t from the ARMA$(1,1)$, but for Table 5 we need to generate the data from the underlying model. For each replication we simulate 100 initial observations before estimation to approximate the unconditional distribution.

G.2 Bootstrapped p-Values

For bootstrapped p-values in Table 6, we use different methods of bootstrapping depending on the test statistic and the null model, as follows:

In **Panel C**, for test statistics RP_2, RP_3 and $t(\beta_z)$ we estimate an ARMA$(1,1)$ representation of the dependent variable (as in (5)) and an AR(1) representation of the predictor (as in (2)) and store the residuals $\{\widehat{\varepsilon}_t, \widehat{v}_{zt}\}_{t=1}^T$ and the estimates of the parameters $\left(\widehat{\theta}, \widehat{\lambda}, \widehat{\lambda}_z, \widehat{\rho}_z\right)$. For RP_1 we estimate the predictive regression and the predictor autoregression, use the properties thereof to derive an estimate of $\widehat{\theta}_z$ as outlined in Appendix F and hence derive an estimate of the pseudo predictor $\widehat{x}_t^f$ (these are shown in Figure 2). We then estimate a predictive regression of the same form as (3) in terms of the estimated pseudo predictor, which under the null gives an estimated series for ε_t. We again store the residuals $\{\widehat{\varepsilon}_t, \widehat{v}_{zt}\}_{t=1}^T$ and the estimates of the parameters $\left(\widehat{\theta}, \widehat{\lambda}, \widehat{\lambda}_z, \widehat{\rho}_z\right)$ (where the first two of these are indirect estimates, setting $\theta = \theta_z\left(\widehat{\lambda}_z, \widehat{\rho}_z, \widehat{R}_z^2\right)$ and $\widehat{\lambda} = \widehat{\lambda}_z$).

In **Panel D**, for all test statistics we assume that y_t is white noise, hence simply set $\widehat{\varepsilon}_t = y_t$. We again store the residuals $\{\widehat{\varepsilon}_t, \widehat{v}_{zt}\}_{t=1}^T$ and the estimates of the parameters $\left(\widehat{\theta}, \widehat{\lambda}, \widehat{\lambda}_z, \widehat{\rho}_z\right)$

(where under the white noise null we can arbitrarily set $\widehat{\lambda} = \widehat{\theta} = \widehat{\lambda}_z$).

To simulate p-values we re-sample (using 5000 replications) from the relevant sets of estimated residuals and simulate as described in the previous section using estimated values of the input parameters, except that here we generate y_t directly from the ARMA representation (since we do not need to make any assumption on the nature of the true predictor).

References

Beveridge, S. and C.R. Nelson (1981), "A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle"," *Journal of Monetary Economics*, 7(2), 151-174.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts, (2007), "On the importance of measuring payout yield: Implications for empirical asset pricing," *Journal of Finance* 62:2, 877-916.

Boudoukh J, Richardson M and R Whitelaw (2006)"The Myth of Long Horizon Predictability" forthcoming: *The Review of Financial Studies*

Campbell J Y, A W Lo and A C Mackinlay (1995) *The Econometrics of Financial Markets* Princeton University Press

Campbell, J Y. and Shiller, R J (1998), Valuation Ratios and the Long-Run Stock Market Outlook," *Journal of Portfolio Management* 24: 11–26

Campbell, J Y, and Thompson, S (2007) "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" , forthcoming *Review of Financial Studies*

Campbell J Y & L M. Viceira, (1999). "Consumption And Portfolio Decisions When Expected Returns Are Time Varying," *The Quarterly Journal of Economics*, MIT Press, vol. 114(2), pages 433-495, May

Carroll C (2008) "Recent Stock Declines: Panic or the Purge of 'Irrational Exuberance'"?, *The Economist's Voice*, November 2008, www.bepress.com/ev

Christoffersen, P and Diebold, F (1998) "Cointegration and Long-Horizon Forecasting," *Journal of Business and Economic Statistics*, 16, 450-458,

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 (1997), Where is the Market Going? Uncertain Facts and Novel Theories, *Economic Perspectives,* XXI (6)

Cochrane J H (2007) The Dog that Did Not Bark: A Defense of Return Predictability *Review of Financial Studies* (2007)

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.

Graham, B and Dodd, D (1934) *Securities Analysis*. New York: McGraw-Hill.

Granger, C (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica,* Vol. 37, No. 3 (Aug., 1969), pp. 424-438

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

Kim, M.J., C.R. Nelson and R. Startz, 1991. Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence, *Review of Economic Studies*.

Lamont, O (1998), "Earnings and expected returns", *Journal of Finance* 53, 1563–1587.

Poterba, J. and Summers, L, 1988. Mean Reversion in Stock Returns: Evidence and Implications. *Journal of Financial Economics.*

Robertson D and S Wright (1998) The Good News and the Bad News about Long-Run Stock Returns, University of Cambridge *Working Paper*

Robertson D and S Wright (2006) Dividends, Total Cash Flow to Shareholders, and Predictive Return Regressions *Review of Economics and Statistics*, vol. 88, issue 1, pages 91-99

Robertson D and S Wright (2009) Predictor Variables and Predictive Regressions, *Working Paper*

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

Smithers, A and Wright, S (2000) *Valuing Wall Street,* McGraw Hill, New York

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

Sullivan, R., A. Timmerman and H. White (1999). Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. *The Journal of Finance* 54, 1647-1691

Vuolteenaho, T (1999) "Understanding the aggregate book-to-market ratio", *Working Paper*

Wright S (2004) "Measures of Stock Market Value and Returns for the US Nonfinancial Corporate Sector, 1900-2002" *Review of Income and Wealth*, 50 (4) pp 561-584

Table 1. OLS-Based Tests of a Redundant AR(1) Predictor of an ARMA(1,1) process

Panel A: Population R-Squared of Redundant Predictor, ρ_z=0.9							
		θ					
		0	0.5	0.7	0.8	0.9	0.95
λ	0	**0.000**	0.168	0.284	0.341	0.396	0.422
	0.5	0.213	**0.000**	0.041	0.089	0.147	0.179
	0.7	0.438	0.060	**0.000**	0.016	0.060	0.090
	0.8	0.590	0.168	0.022	**0.000**	0.022	0.048
	0.9	0.775	0.406	0.146	0.041	**0.000**	0.011
	0.95	0.882	0.627	0.342	0.157	0.020	**0.000**
Panel B: Size of 1-sided t-test on OLS estimate of β_z, at notional 5% level,T=200, when z_t is a redundant predictor							
		θ					
		0	0.5	0.7	0.8	0.9	0.95
λ	0	**0.040**	1.000	1.000	1.000	1.000	1.000
	0.5	1.000	**0.054**	0.881	0.999	1.000	1.000
	0.7	1.000	0.856	**0.067**	0.495	0.996	1.000
	0.8	1.000	0.994	0.426	**0.065**	0.668	0.978
	0.9	1.000	1.000	0.969	0.594	**0.072**	0.397
	0.95	1.000	1.000	0.999	0.918	0.305	**0.080**

Table 1 assumes that the predicted process is ARMA(1,1): $y_t = (1-\theta L)/(1-\lambda L)\varepsilon_t$; and z_t is a redundant AR(1) predictor with $\lambda_z=\lambda$ and Stambaugh correlation $\rho_z=\text{corr}(u_{zt}, v_{zt})=0.9$ in the predictive system (1) and (2). Panel A gives the value of R_z^2 in Proposition 1 (where the inequality holds precisely since $\lambda_z=\lambda$). Panel B shows the simulated size of a *t*-test on $\hat{\beta}_z$ in equation (1) in 1000 replications.

Table 2. Simulated size of three tests of the null that z is a redundant predictor

$\lambda_z = \lambda$

		θ					
RP₁	λ	0	0.5	0.7	0.8	0.9	0.95
	0	**0.065**	0.055	0.073	0.097	0.200	0.290
	0.5	0.077	**0.060**	0.060	0.059	0.092	0.139
	0.7	0.092	0.060	**0.077**	0.073	0.064	0.092
	0.8	0.106	0.083	0.060	**0.085**	0.067	0.065
	0.9	0.187	0.095	0.081	0.079	**0.107**	0.062
	0.95	0.329	0.171	0.129	0.113	0.116	**0.112**
RP₂							
	0	**0.018**	0.041	0.050	0.054	0.060	0.076
	0.5	0.054	**0.017**	0.044	0.041	0.054	0.063
	0.7	0.063	0.043	**0.029**	0.042	0.061	0.071
	0.8	0.072	0.051	0.022	**0.030**	0.057	0.069
	0.9	0.082	0.073	0.059	0.042	**0.036**	0.057
	0.95	0.097	0.094	0.092	0.075	0.036	**0.038**
RP₃							
	0	**0.131**	0.120	0.165	0.229	0.150	0.117
	0.5	0.100	**0.107**	0.205	0.232	0.189	0.124
	0.7	0.099	0.145	**0.116**	0.168	0.191	0.132
	0.8	0.106	0.135	0.129	**0.112**	0.143	0.115
	0.9	0.127	0.113	0.167	0.169	**0.075**	0.091
	0.95	0.163	0.132	0.163	0.198	0.137	**0.082**

Table 2 shows the simulated size, in 1000 replications, of the three tests that z_t is a redundant predictor of an ARMA(1,1) process, as described in Sections 3.1.1 to 3.1.3. The processes for y_t and z_t are as for Table 1. *T=200*

Table 3. Simulated size of three tests of the null that z is a redundant predictor

$\lambda_z = \lambda/2$

		θ					
RP1	λ	0	0.5	0.7	0.8	0.9	0.95
	0	**0.065**	0.055	0.073	0.097	0.2	0.29
	0.5	0.081	**0.062**	0.052	0.07	0.201	0.388
	0.7	0.114	0.072	**0.063**	0.061	0.139	0.311
	0.8	0.176	0.098	0.073	**0.063**	0.076	0.196
	0.9	0.298	0.512	0.213	0.089	**0.059**	0.074
	0.95	0.437	0.873	0.711	0.376	0.073	**0.069**
RP2							
	0	**0.018**	0.041	0.05	0.054	0.06	0.076
	0.5	0.034	**0.014**	0.037	0.035	0.044	0.068
	0.7	0.01	0.029	**0.017**	0.032	0.045	0.074
	0.8	0.004	0.012	0.023	**0.02**	0.039	0.064
	0.9	0.004	0.002	0.01	0.02	**0.019**	0.035
	0.95	0.005	0	0.005	0.013	0.017	**0.026**
RP3							
	0	**0.131**	0.12	0.165	0.229	0.15	0.117
	0.5	0.087	**0.118**	0.18	0.205	0.185	0.106
	0.7	0.071	0.125	**0.112**	0.158	0.157	0.094
	0.8	0.071	0.085	0.115	**0.102**	0.141	0.096
	0.9	0.065	0.066	0.08	0.108	**0.105**	0.097
	0.95	0.053	0.056	0.065	0.083	0.112	**0.117**

Table 3 shows the simulated size, in 1000 replications, of the three tests that z_t is a redundant predictor of an ARMA(1,1) process, as described in Sections 3.1.1 to 3.1.3. The processes for y_t and z_t are as for Table 1, but with $\lambda_z=\lambda/2$. *T=200*

Table 4. Simulated size of three tests of the null that z is a redundant predictor

$$\lambda_z = \lambda+(1-\lambda)/2$$

		θ					
RP1	λ	0	0.5	0.7	0.8	0.9	0.95
	0	**0.06**	0.348	0.743	0.071	0.333	0.709
	0.5	0.074	**0.081**	0.109	0.311	0.127	0.22
	0.7	0.07	0.064	**0.095**	0.092	0.211	0.121
	0.8	0.072	0.087	0.069	**0.11**	0.103	0.112
	0.9	0.094	0.111	0.101	0.096	**0.12**	0.112
	0.95	0.148	0.152	0.143	0.136	0.142	**0.131**
RP2							
	0	**0.018**	0.016	0.019	0.027	0.04	0.057
	0.5	0.033	**0.029**	0.036	0.032	0.046	0.059
	0.7	0.05	0.042	**0.037**	0.04	0.057	0.055
	0.8	0.06	0.04	0.029	**0.041**	0.054	0.065
	0.9	0.076	0.069	0.06	0.043	**0.039**	0.07
	0.95	0.108	0.098	0.086	0.078	0.04	**0.05**
RP3							
	0	**0.104**	0.115	0.156	0.219	0.163	0.104
	0.5	0.072	**0.127**	0.193	0.212	0.163	0.099
	0.7	0.082	0.088	**0.102**	0.147	0.144	0.119
	0.8	0.087	0.07	0.099	**0.085**	0.126	0.158
	0.9	0.082	0.078	0.094	0.106	**0.068**	0.138
	0.95	0.076	0.076	0.086	0.119	0.101	**0.115**

Table 3 shows the simulated size, in 1000 replications, of the three tests that z_t is a redundant predictor of an ARMA(1,1) process, as described in Sections 3.1.1 to 3.1.3. The processes for y_t and z_t are as for Table 1, but with $\lambda_z=\lambda + (1-\lambda)/2$. *T=200*

Table 5 Simulated rejection rates of the three tests under H_1: $z_t = x_t$.

$$R_x^2 = 0.25R_f^2 + 0.75R_n^2$$

		Θ					
RP1	λ	0	0.5	0.7	0.8	0.9	0.95
	0	**1**	1	1	0.998	0.93	0.786
	0.5	1	**1**	1	0.998	0.901	0.662
	0.7	1	1	**0.999**	0.991	0.886	0.609
	0.8	1	1	0.997	**0.991**	0.861	0.558
	0.9	1	1	1	0.99	**0.875**	0.548
	0.95	1	1	1	0.993	0.879	**0.593**
RP2							
	0	**1**	1	1	0.992	0.772	0.304
	0.5	1	**1**	0.998	0.995	0.763	0.324
	0.7	1	1	**1**	0.985	0.754	0.318
	0.8	1	1	0.996	**0.987**	0.752	0.334
	0.9	1	1	0.997	0.987	**0.799**	0.4
	0.95	1	1	0.998	0.985	0.809	**0.475**
RP3							
	0	**0.981**	0.999	0.998	0.988	0.895	0.718
	0.5	0.999	**1**	0.999	0.991	0.906	0.738
	0.7	0.997	1	**1**	0.99	0.893	0.744
	0.8	0.998	1	0.995	**0.978**	0.887	0.734
	0.9	0.999	1	0.998	0.989	**0.85**	0.683
	0.95	0.996	1	0.998	0.988	0.874	**0.634**
Memo: R-Squared of True Predictor (R_x^2)							
	0	**0.250**	0.350	0.414	0.445	0.474	0.487
	0.5	0.438	**0.188**	0.172	0.188	0.215	0.232
	0.7	0.618	0.247	**0.128**	0.108	0.117	0.131
	0.8	0.730	0.350	0.151	**0.090**	0.073	0.082
	0.9	0.858	0.559	0.279	0.136	**0.048**	0.037
	0.95	0.927	0.736	0.468	0.261	0.071	**0.024**
Memo: Stambaugh Correlation of True Predictor (ρ)							
	0	**0.000**	0.839	0.954	0.982	0.996	0.999
	0.5	-0.655	**0.277**	0.767	0.911	0.982	0.996
	0.7	-0.771	-0.207	**0.375**	0.723	0.943	0.988
	0.8	-0.811	-0.419	0.009	**0.419**	0.859	0.970
	0.9	-0.842	-0.580	-0.360	-0.121	**0.461**	0.862
	0.95	-0.855	-0.642	-0.504	-0.382	-0.059	**0.481**

Table 5 shows the rejection rate at a nominal 5% size, in 1000 replications, of the three tests that z_t is a redundant predictor, as described in Sections 3.1.1 to 3.1.3, under the alternative hypothesis H_1: $z_t=x_t$. The true predictor x_t is assumed to have a predictive R-squared.given by a fixed linear weighting of the upper and lower bounds given in Proposition 2: $R_x^2 = 0.25R_f^2 + 0.75R_n^2$, where both upper and lower bounds are functions of the ARMA parameters alone. *T=200*. The bottom two panels show, for reference, the implied values of the R-Squared and the Stambaugh Correlation for the true predictor, consistent with Proposition 2.

Table 6. Tests of Predictor Redundance: Some Empirical Examples

	Predicted Variable				
	Real stock returns				Quasi-differenced log GNP growth
	Predictor				
	log(price / dividend)	log(price / 10 year earnings)	log(Tobin's q)	log(price / total cash transfers)	log(GNP / consumption)
Sample	1901-2007	1901-2007	1901-2007	1901-2007	1948:01-2007:02
T	107	107	107	107	238
Panel A. Predictor Characteristics					
Stambaugh Correlation ($\widehat{\rho_z}$)	0.835	0.983	0.914	0.552	0.816
Predictor AR(1) parameter ($\widehat{\lambda_z}$)*	0.922	0.928	0.905	0.694	0.943
Panel B. Nominal P-Values					
RP_1	1.000	0.888	0.440	0.011	0.003
RP_2	0.898	0.885	0.567	0.068	0.001
RP_3	0.872	0.255	0.132	0.014	0.000
$t(\beta_z)$	*0.141*	*0.034*	*0.009*	*0.002*	*0.000*
Panel C: Bootstrapped P-Values, y=ARMA(1,1)**					
RP_1	0.9996	0.8984	0.6882	0.0178	0.0048
RP_2	0.9218	0.9456	0.659	0.0994	0.0122
RP_3	0.9038	0.8368	0.6876	0.0578	0.0022
$t(\beta_z)$	0.9474	0.9184	0.6088	0.0308	0.0364
Panel D: Bootstrapped P-Values, y =white noise					
RP_1	0.9994	0.9078	0.5286	0.0164	0.0044
RP_2	0.8898	0.8546	0.5222	0.0706	0.0006
RP_3	0.8524	0.2516	0.1712	0.0362	0.0000
$t(\beta_z))$	0.1838	0.0634	0.0148	0.0026	0.0002

* AR(1) estimates include bias-correction. ** Bootstrapped p-values for *RP1* set $\lambda=\lambda_z$; $\theta=\theta_z(.)$ as in equation (15); bootstrapped p values for other tests use direct ARMA estimates of λ and θ . See Appendix C for further detail

Table A1 Sampling Properties of ARMA vs Indirect Estimates of θ

		Θ					
	λ	0	0.5	0.7	0.8	0.9	0.95
Mean ARMA estimate of θ							
	0	**0.027**	0.532	0.729	0.830	0.935	0.977
	0.5	-0.031	**0.293**	0.735	0.840	0.936	0.977
	0.7	-0.020	0.403	**0.417**	0.748	0.934	0.975
	0.8	-0.017	0.465	0.467	**0.513**	0.871	0.963
	0.9	-0.013	0.483	0.656	0.611	**0.590**	0.832
	0.95	-0.011	0.488	0.681	0.755	0.612	**0.613**
Standard Deviation of ARMA estimate							
	0	**0.626**	0.136	0.080	0.064	0.046	0.024
	0.5	0.157	**0.655**	0.235	0.104	0.055	0.027
	0.7	0.109	0.312	**0.662**	0.396	0.135	0.066
	0.8	0.095	0.159	0.501	**0.652**	0.274	0.157
	0.9	0.085	0.096	0.181	0.435	**0.637**	0.455
	0.95	0.080	0.080	0.091	0.193	0.573	**0.642**
Mean of Indirect Estimate (θ_z)							
	0	**0.061**	0.500	0.700	0.801	0.901	0.948
	0.5	0.001	**0.546**	0.700	0.800	0.900	0.950
	0.7	0.001	0.501	**0.734**	0.801	0.900	0.950
	0.8	0.001	0.502	0.696	**0.825**	0.900	0.950
	0.9	0.001	0.502	0.701	0.798	**0.914**	0.950
	0.95	0.001	0.501	0.701	0.801	0.890	**0.958**
Standard Deviation of Indirect Estimate (θ_z)							
	0	**0.081**	0.032	0.032	0.033	0.034	0.031
	0.5	0.036	**0.065**	0.027	0.026	0.025	0.024
	0.7	0.036	0.032	**0.051**	0.022	0.021	0.021
	0.8	0.037	0.033	0.030	**0.040**	0.019	0.018
	0.9	0.042	0.036	0.030	0.025	**0.025**	0.014
	0.95	0.061	0.043	0.035	0.028	0.029	**0.016**

Table A1 compares, for different values of the population parameters θ and λ, sampling properties of the ARMA estimate of θ and the indirect estimate, $\hat{\theta}_z = \theta_z\left(\hat{\lambda}_z, \widehat{\rho_z}, \hat{R}_z^2\right)$ derived from the properties of the predictive regression (1) and the predictor autoregression (2) under the joint null that $\lambda_z=\lambda$ and z_t is redundant (see Section 3.1.1 and Appendix B). Results are shown for 1000 replications, with ρ_z=0.9, T=200.To allow for the prior that λ and θ are both positive, ARMA estimates in each replication use as starting values estimates of λ_z and θ_z but discard these if zero starting values yield a higher value of the estimated log likelihood.